Functional-Coefficient Autoregressive Models — **Source link** ⧉

Rong Chen, Ruey S. Tsay

Institutions: Texas A&M University, University of Chicago

Related papers:

- Functional-Coefficient Regression Models for Nonlinear Time Series

- Varying-Coefficient Models

- Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model

- Statistical estimation in varying coefficient models

- Non-linear time series. A dynamical system approach

Functional coefficient autoregressive models: estimation and tests of hypotheses

Rong Chen[1]

Department of Statistics

Texas A&M University

SUMMARY

In this paper we study nonparametric estimation and hypothesis testing procedures for the functional coefficient AR (FAR) models of the form $X_t = f_1(X_{t-d})X_{t-1} + \ldots + f_p(X_{t-d})X_{t-p} + \varepsilon_t$, first proposed by Chen and Tsay (1993). As a direct generalization of the linear AR model, the FAR model is a rich class of models that includes many successful parametric nonlinear time series models such as the threshold AR models of Tong (1983), exponential AR models of Haggan and Ozaki (1978) and many others. We propose a local linear estimation procedure for estimating the coefficient functions and study its asymptotic properties. In addition, we propose two testing procedures. The first one tests whether all the coefficient functions are constant (i.e. whether the process is linear). The second one tests if all the coefficient functions are continuous, (i.e. if any threshold type of nonlinearity presents in the process). Some simulation results are presented.

Key Words': Continuity test; Linearity test; Local linear estimation; Nonparametric estimation; One sided kernel; Threshold Model;

1. INTRODUCTION

Nonlinear time series analysis has been one of the major areas of research in time series for more than two decades now. Many nonlinear parametric models such as the threshold AR (TAR) model of Tong (1983, 1990), the exponential AR (EXPAR) model of Haggan and Ozaki (1978) and the smooth transition AR (STAR) model of Granger and Teräsvirta (1993) and Teräsvirta (1994) have been proposed and successfully applied in many important real problems. Tong (1990) and Priestley (1988) provided many foundations of parametric nonlinear time sereis analysis. A more recent review of the subject can be found in Tjøsteim (1994).

It is noted that although in many applications background knowledge can often shed lights on finding an appropriate model, other applications lack such knowledge and often require trial-and-error type of model selection procedures. To overcome the subjectivity in model selection, Chen and Tsay (1993) proposed a class of models referred to as functional coefficient (FAR) models which assumes the form of

$$X_t = f_1(X_{t-d})X_{t-1} + \ldots + f_p(X_{t-d})X_{t-p} + \varepsilon_t. \tag{1}$$

where ε_t is white noise with finite variance σ^2 and is independent of X_s for all $s < t$. It is a direct extension of the linear AR model, but allows the coefficients varying according to a threshold variable X_{t-d}. They suggested using nonparametric procedures to determine the functions in the model, hence allowing 'data to speak for themselves' regarding the model to be used. It is noted that many of the successful parametric nonlinear models belong to the FAR family. For example, if the functions $f_i(x)$ in (1) are step functions $f_i(x) = a_i + b_i I(x \geq c)$, we have the TAR model. When $f_i(x) = a_i + b_i e^{-\gamma x^2}$, the model becomes an EXPAR model. STAR and many other models also belong to this class. Hence, nonparametric determination of the functional forms in model (1) may provide objective guild-lines on choosing an appropriate parametric model. It also allows researchers to develop new models that are useful in their applications by specifying a parametric form for the coefficient functions based on the nonparametric estimates. In addition, nonparametric estimators can also be the final solution to the problem on hand.

Nonparametric procedures have been used extensively in time series analysis. Györfi et al (1989), Tjøsteim (1994), Härdle, Lütkepohl and Chen (1997) and Hart (1996) have given selective reviews on this topic. These procedures borrow many nonparametric procedures developed in regression context into time series analysis.

In this paper, we concentrate on three aspects of the FAR models. First, in section 2, we propose a local linear estimator for estimating the coefficient functions nonparametrically. It is similar

to the running window procedure proposed by Chen and Tsay (1993), though we use Kernel weight functions. We systematically study the asymptotic properties of the estimator. Note that this procedure is slightly different from local polynomial curve estimation procedures of Cleveland and Devlin (1988), Fan and Gijbels (1996) and Tsybakov (1986). Here, we are interested in estimating the coefficient functions. Hastie and Tibshirani (1993) have proposed similar estimation procedures in regression context for 'varying coefficient models', which is similar to the FAR model.

Second, in section 3, we develop a procedure to test if the coefficient functions are constant functions. It is basically a linearity test since when all the coefficient functions are constant, the FAR model becomes a linear AR model. There are many linearity tests available in the literature. For example. Keenan (1985), Tsay (1986) and Luukkonen, Saikkonen and Teräsvirta (1988) proposed different forms of Lagrange multiplier type of tests. Chan and Tong (1986) and Tsay (1989) considered testing threshold type of nonlinearity. Nonparametrically, Hjellvik and Tjøstheim (1995, 1996) and Hjellvik, Yao and Tjøstheim (1997) developed linearity tests by comparing nonparametric and linear estimates of $E[X_t \mid X_{t-k}]$. Here we attack this problem within the FAR model framework.

Third, in section 4, we develop yet another testing procedure, to detect if there are any discontinuous points in the coefficient functions. This is of interest due to the fact that all threshold type of models have jump points in the coefficient functions. Since the class of threshold models is one of the most important and widely used classes of nonlinear time series models, it is certainly important to be able to detect if there is any threshold type of nonlinearity when one uses FAR models as a tool for model selection. The test is also of interest when the nonparametric estimate is treated as a final solution of the problem. Most of the nonparametric estimators are designed to estimate continuous functions. They are not consistent at discontinuous points. In finite samples, they tend to have large bias in the neighborhood of the discontinuous points. Hence it is important to detect the existence of jump points, in order to select suitable nonparametric estimators. The proposed testing procedure is based on nonparametric estimation of the coefficient functions with one-sided kernels and the fact that at a discontinuous point, estimates with left-side kernels and right-side kernels are significantly different while at continuous points they are not.

2. Nonparametric estimation of the FAR model

We begin this section by mentioning that, for $d > p$, we may also include an intercept coefficient function in model (1), resulting in

$$X_t = f_0(X_{t-d}) + f_1(X_{t-d})X_{t-1} + \ldots + f_p(X_{t-d})X_{t-p} + \varepsilon_t.$$

3

However, when $d \leq p$, such a term will create ambiguity. In what follows, we will only consider model (1) without the f_0 term.

We propose the following local linear estimator to estimate the functions $\boldsymbol{f}(x) = (f_1(x), \ldots, f_p(x))'$ nonparametrically. Let

$$\widehat{\boldsymbol{f}}(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=\ell+1}^{n} (X_t - \boldsymbol{X}_t'\boldsymbol{\beta})^2 K_h(X_{t-d} - x), \tag{2}$$

where $\boldsymbol{X}_t = (X_{t-1}, \cdots, X_{t-p})'$ and $K_h(u) = h^{-1}K(u/h)$ where K is a kernel function, h is the bandwidth and $\ell = \max\{d, p\}$. It is easily seen that

$$\hat{\boldsymbol{f}}(x) = (\boldsymbol{X}'W_x\boldsymbol{X})^{-1}\boldsymbol{X}'W_x\boldsymbol{Y},$$

where $\boldsymbol{X} = [\boldsymbol{X}_{\ell+1} : \cdots : \boldsymbol{X}_n]'$, $\boldsymbol{Y} = (X_{\ell+1}, \cdots, X_n)'$ and W_x is a diagonal matrix with the diagonal elements being $K_h(X_{t-d} - x)$ for $t = \ell+1, \ldots, n$.

The asymptotic properties of the above estimator can be summarized in the following theorem. The theorem concerns only the continuous points. We will study the case of discontinuous coefficient functions in section 4.

Define $\mu_2 = \int u^2 K(u)du$ and $K_2^2 = \int K^2(u)du$. Let $p_{i,j,d}$ be the joint stationary density of the triple $(X_{t-i}, X_{t-j}, X_{t-d})$ and $p(x)$ be the stationary marginal density of X_t.

THEOREM 0.1 *Let x be a continuous point of the coefficient functions $f_1, \ldots, f_p$. Under assumptions (A1) to (A8) in the appendix, we have*

$$n^{2/5}(\hat{\boldsymbol{f}}(x) - \boldsymbol{f}(x) - \beta^2 \boldsymbol{b}(x)) \xrightarrow{\mathcal{D}} N_p(0, \beta^{-1}\sigma^2 K_2^2 \boldsymbol{A}^{-1}(x)),$$

where $\boldsymbol{A}(x) = p(x)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d} = x]$ and $\boldsymbol{b}(x) = \mu_2 \boldsymbol{A}^{-1}(x)\boldsymbol{B}(x)$ where $\boldsymbol{B}(x)$ is a vector with i-th element being

$$\sum_{j=1}^{p} \int uv \left\{ \frac{1}{2} f_j''(x)p_{i,j,d}(u,v,x) + f_j'(x)p_{i,j,d}'(u,v,x) \right\} dudv,$$

with $f_j'(x)$ and $f_j''(x)$ being the first and second derivative of $f_j(x)$, respectively, and $p_{i,j,d}'$ being the partial derivative with respect to the third argument.

The proof of the theorem is given in the appendix. We note that the asymptotic result is similar to that of kernel estimation of a response curve. It can be easily extended to resemble that of local polynomial estimation of a response curve using the following estimator:

$$\hat{\boldsymbol{f}}(x) = \arg_{\boldsymbol{\beta}} \min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{t=\ell+1}^{n} (X_t - \boldsymbol{X}_t'\boldsymbol{\beta} - \boldsymbol{Y}_t'\boldsymbol{\gamma})K_h(X_{t-d} - x),$$

4

with $\boldsymbol{X}_t$ as defined before and $\boldsymbol{Y}_t = ((X_{t-1}-x)^2, \ldots, (X_{t-p}-x)^2, \ldots, (X_{t-1}-x)^k, \ldots, (X_{t-p}-x)^k)'$. It should entertain many nice properties of the local polynomial estimator, though the derivation of the asymptotic distribution becomes more complicated and tedious. In this paper we restrict ourselves to estimator (2).

The theorem shows that the estimator has the rate of convergence of one dimensional smoothing. As a consequence, estimation of the response surface $E[X_t \mid \boldsymbol{X}_t = \boldsymbol{x}]$ will have the same rate, hence does not suffer the curse of dimensionality as that in direct p-dimensional estimation of the surface. This advantage is due to the special structure of the model, which serves as a dimension reduction tool. Specifically, let $\boldsymbol{x} = (x_1, \ldots, x_p)'$. The conditional mean function $m(\boldsymbol{x}) = E[X_t \mid \boldsymbol{X}_t = \boldsymbol{x}] = \boldsymbol{x}'\boldsymbol{f}(x_d)$ can be estimated by $\hat{m}(\boldsymbol{x}) = \boldsymbol{x}'\hat{\boldsymbol{f}}(x_d)$ and we have

Corollary 0.1 *Under conditions of theorem 0.1, we have*

$$n^{2/5}(\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) - \beta^2 \boldsymbol{x}'\boldsymbol{b}(x_d)) \xrightarrow{\mathcal{D}} N(0, \beta^{-1}\sigma^2 K_2^2 \boldsymbol{x}'\boldsymbol{A}^{-1}(x_d)\boldsymbol{x}).$$

Automatic bandwidth selection procedure is always one of the key ingredients in practical implementation of nonparametric procedures. There are many approaches such as the cross-validation approach of Härdle and Vieu (1992) and Cheng and Tong (1992) in time series, the plug-in approach of Sheather (1983, 1986), Ruppert, Sheather, and Wand (1995), Park and Marron (1990) and many others in regression. It is somewhat difficult to use the plug-in approach here since the bias term involves the partial derivative of a three dimensional density, which is not easy to estimate. Hence we suggest to use the cross-validation procedure through the response surface estimation. Specifically, define

$$cv(h) = \sum_{i=\ell+1}^{n} (X_i - \sum_{j=1}^{p} \hat{f}_{h,j}^{(-i)}(X_{i-d})X_{i-j})^2 w(X_{i-d}),$$

where $\hat{f}_{h,j}^{(-i)}, j = 1, \ldots, p$ is that in (2) without the $t = i$ term in the summation and w is a weight function with a compact support.

Consider a second order EXPAR model (3),

$$(\text{EXPAR}(2)) \quad X_t = (0.5 - 1.1e^{-50X_{t-1}^2})X_{t-1} + (0.3 - 0.5e^{-50X_{t-1}^2})X_{t-2} + 0.2\varepsilon_t. \tag{3}$$

Figure 1 shows the estimation results of a simulated series from (3) with $\varepsilon_t \sim N(0,1)$ and 400 samples, using the optimal cross-validation bandwidth with the quartic kernel function $K(u) = 0.9375(1 - u^2)^2 I(|u| \le 1)$. The solid lines are the true function.
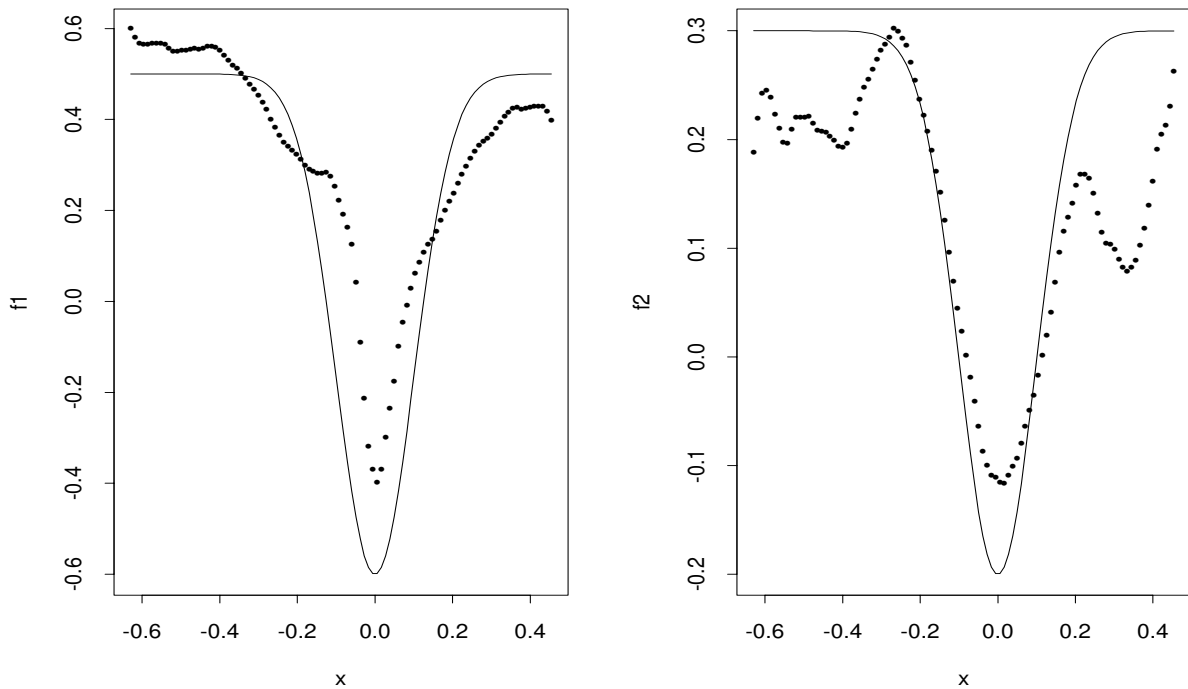
Figure 1: Function estimation of a simulated series from model (3). The solid lines are the true functions

Figure 2 shows the cross-validation curves of 5 series generated from the above process. We can see that the procedure is reasonablely robust .

3. TESTING LINEARITY IN FAR MODELS

When all the coefficient functions are constant functions, an FAR model becomes a linear AR model. In this section we develop a testing procedure to determine if the process is linear.

Let $\hat{\boldsymbol{f}}(x)$ be that in (2) and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)$ be the Yule-Walker estimator of a linear AR(p) model. We define the following statistic to test linearity:

$$T = \frac{1}{n} \sum_{t=\ell+1}^{n} \boldsymbol{d}_t' \boldsymbol{d}_t w(X_{t-d}), \tag{4}$$

where $\boldsymbol{d}_t = (\boldsymbol{X}' W_t \boldsymbol{X})(\hat{\boldsymbol{f}}(X_{t-d}) - \hat{\boldsymbol{\theta}})$ with W_t being a diagonal matrix with diagonal element being $K(X_{i-d} - X_{t-d})$ for $i = \ell + 1, \ldots, n$. The weight function w has a compact support, designed
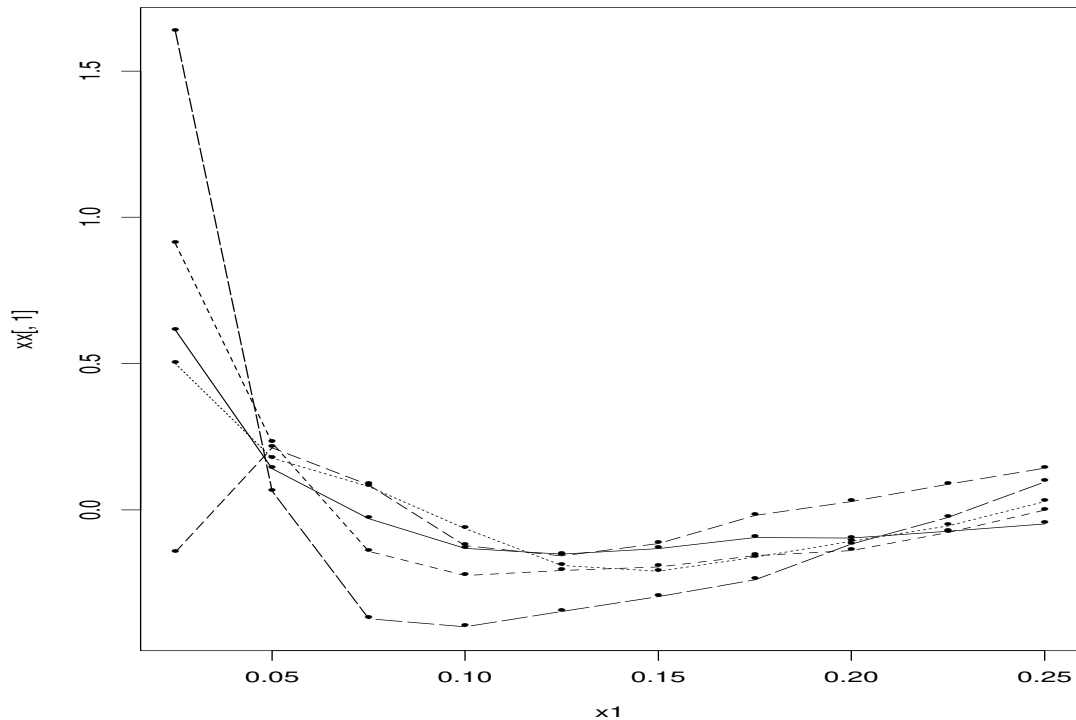
6

Figure 2: Cross-validation curves for 5 simulated series from model (3)

to reduce the boundary effects on the test statistic. Note that traditionally one would use $\boldsymbol{d}_t = \hat{\boldsymbol{f}}(X_{t-d}) - \hat{\boldsymbol{\theta}}$ in (4). The use of $\boldsymbol{d}_t$ in its current form is purely for the simplicity and weaker conditions in obtaining the asymptotic distribution of T.

We have the following theorem for the asymptotic distribution of the test statistic T.

THEOREM 0.2 *Under conditions (A2), (B1) and (B2) listed in the appendix, and the null hypothesis that $f_j(x) = \theta_j$ for $j = 1, \ldots, p$, with all the roots of $z^p - \theta_1 z^{p-1} - \ldots - \theta_p = 0$ inside the unit circle, we have*

$$nh^{1/2}T \xrightarrow{\mathcal{D}} N(h^{-1/2}a_0, s_0^2),$$

where

$$a_0 = 2K_2^2\sigma^2 E[\sum_{k=1}^{p} X_{t-k}^2 p(X_{t-d})w(X_{t-d})],$$

7

and

$$s_o^2 = \sigma^4 \int K(u)K(v)K(u-z)K(v-z)dudvdz \int s^4(x)p^2(x)w^2(x)dx,$$

where $s^2(x) = E[(\sum_{k=1}^p X_{t-k})^2 \mid X_{t-d} = x]$.

THEOREM 0.3 *Under the conditions (A1) to (A4), (B1) and (B2) listed in the appendix, and the alternative hypothesis that at least one of the coefficient functions $f_j(x)$ are not constant, then*

$$T \xrightarrow{\mathcal{D}} N(a_1, s_1^2/n),$$

where

$$a_1 = E[(\boldsymbol{f}(X_t) - \boldsymbol{\theta})'\boldsymbol{A}(X_t)\boldsymbol{A}(X_t)(\boldsymbol{f}(X_t) - \boldsymbol{\theta})w(X_t)],$$

and

$$s_1^2 = Var[(\boldsymbol{f}(X_t) - \boldsymbol{\theta})'\boldsymbol{A}(X_t)\boldsymbol{A}(X_t)(\boldsymbol{f}(X_t) - \boldsymbol{\theta})w(X_t)],$$

where $\boldsymbol{A}(x) = p(x)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d} = x]$, and $\boldsymbol{\theta}$ is the coefficient of the best linear prediction of X_t given $X_{t-1}, \ldots, X_{t-p}$.

The theorem shows that as $nh \to \infty$, T goes to zero in probability under the null hypothesis. Hence, large value of the statistic indicates departure from linearity. It also shows that under the null hypothesis, $nh^{1/2}T$ is asymptotic normal with finite variance, but the mean goes to infinity. This type of results were observed by Härdle and Mammen (1993), Hjvellik et al. (1997) in similar problems. The proof of the theorems basically follows similar proofs in Yoshihara (1976) and Hjvellik et al. (1997). First we obtain the Hoeffding's decomposition of the test statistic. Then a martingale central limit theorem is used on the resulting U-statistic. The proof is tedious and is omitted here.

Although Theorem 0.2 can be used to obtain asymptotic level of the test statistic, it is noted by many researchers (e.g. Skaug and Tjøstheim 1993, Hjellvik and Tjøstheim 1995, 1996) that in finite samples, the asymptotic level does not perform well in most cases. Hence, for practical purposes, we suggest to use bootstrap procedures.

Specifically, first we obtain residuals

$$\hat{e}_t = X_t - \sum_{i=1}^p \hat{\phi}_i X_{t-i},$$

where $\hat{\phi}_i$, $i = 1, \ldots, p$ are the Yule-Walker estimates of a linear AR(p) model fitted to the data. Then we create bootstrap versions of the process

$$X_t^* = \sum_{i=1}^p \hat{\phi}_i X_{t-i}^* + e_t^*,$$

α	AR(2)	AR(1)	EXPAR(2)	TAR	STAR	EXPAR(1)
0.10	10	4	96	100	94	40
0.05	8	4	94	100	90	24
0.01	2	2	84	90	74	10

Table 1: Percentage of rejection of the linearity test

for $t = \ell + 1, \ldots, n$, where e_t^* are independently sampled from $\{\hat{e}_{\ell+1}, \ldots, \hat{e}_n\}$ with replacement and $X_t^* = X_t$ for $t = 1, \ldots, \ell$. Then a bootstrap value of test statistic T^* is obtained by replacing X_t^* in the place of X_t in calculating the test statistic (4). The bootstrap null distribution can be then obtained.

In Table 1 we present a small scale simulation for checking the performance of the proposed tests. In addition to model (3), we include five other models

$$\text{(AR(2))} \qquad X_t = 0.6X_{t-1} - 0.3X_{t-2} + \varepsilon_t, \qquad (5)$$

$$\text{(AR(1))} \qquad X_t = 0.5X_{t-1} + \varepsilon_t, \qquad (6)$$

$$\text{(TAR)} \quad X_t = (0.4 - 1.0I(X_{t-1} > 0))X_{t-1} + (-0.8 + 1.0I(X_{t-1} > 0))X_{t-2} + \varepsilon_t. \qquad (7)$$

$$\text{(STAR)} \qquad X_t = \left(0.5 - \frac{1.1}{1+e^{-2X_{t-1}}}\right)X_{t-1} + \left(0.3 - \frac{0.5}{1+e^{-2X_{t-1}}}\right)X_{t-2} + \varepsilon_t, \qquad (8)$$

$$\text{(EXPAR(1))} \qquad X_t = 0.5 - 1.1e^{-50X_{t-1}^2} + 0.2\varepsilon_t, \qquad (9)$$

For each model we generated 50 series of size 400. We use 50 bootstrap replications to obtain the bootstrap null distribution. For each model, we use a common bandwidth obtained by averaging 5 cross-validation bandwidths of five simulated samples. Table 1 presents the percentage of rejection of the null hypothesis under three different α levels. From the table we can see that the proposed testing procedure works reasonablely well.

4. Testing threshold type of discontinuity

First we define local linear estimates of the coefficient functions using one-sided kernels. Let $\boldsymbol{f}^+(x) = \lim_{\delta \to 0+} \boldsymbol{f}(x + \delta)$ and $\boldsymbol{f}^-(x) = \lim_{\delta \to 0-} \boldsymbol{f}(x + \delta)$. Define

$$\hat{\boldsymbol{f}}^+(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=\ell+1}^{n} (X_t - \boldsymbol{X}_t'\boldsymbol{\beta})^2 K_h^+(X_{t-d} - x),$$

and

$$\hat{\boldsymbol{f}}^{-}(x) = \arg\min_{\boldsymbol{\beta}} \sum_{t=\ell+1}^{n} (X_t - \boldsymbol{X}_t'\boldsymbol{\beta})^2 K_h^{-}(X_{t-d} - x),$$

where $K_h^{+}(u) = 2K_h(u)I(u \geq 0)$ and $K_h^{-}(u) = 2K_h(u)I(u \leq 0)$ where K is a symmetric kernel function with bounded support and $\int K(u)du = 1$. Let $\mu_1^{+} = \int uK^{+}(u)du$, $\mu_1^{-} = \int uK^{-}(u)du$.

We have the following theorem:

THEOREM 0.4 *Under conditions (A1) to (A4) and (C1) and (C2) listed in the appendix, we have*

(i) $\hat{\boldsymbol{f}}^{+}(x)$ and $\hat{\boldsymbol{f}}^{-}(x)$ are asymptotically uncorrelated.

(ii)

$$n^{1/3}(\hat{\boldsymbol{f}}^{+}(x) - \boldsymbol{f}^{+}(x) - \beta\boldsymbol{b}^{+}(x)) \xrightarrow{\mathcal{D}} N(0, \beta^{-1}\sigma^2 K_2^2 \boldsymbol{A}^{-1}(x)),$$

and

$$n^{1/3}(\hat{\boldsymbol{f}}^{-}(x) - \boldsymbol{f}^{-}(x) - \beta\boldsymbol{b}^{-}(x)) \xrightarrow{\mathcal{D}} N(0, \beta^{-1}\sigma^2 K_2^2 \boldsymbol{A}^{-1}(x)),$$

where $\boldsymbol{A}(x) = p(x)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d} = x]$ and $\boldsymbol{b}^{+}(x) = \boldsymbol{A}^{-1}(x)\boldsymbol{B}^{+}(x)$, $\boldsymbol{b}^{-}(x) = \boldsymbol{A}^{-1}(x)\boldsymbol{B}^{-}(x)$, with $\boldsymbol{B}^{+}(x)$ and $\boldsymbol{B}^{-}(x)$ being vectors with i-th element being

$$\mu_1^{+} \sum_{j=1}^{p} f_j^{'(+)}(x)E[X_{t-i}X_{t-j} \mid X_{t-d} = x] \quad and \quad \mu_1^{-} \sum_{j=1}^{p} f_j^{'(-)}(x)E[X_{t-i}X_{t-j} \mid X_{t-d} = x],$$

respectively, where $f_j^{'(+)}$ and $f_j^{'(-)}$ are the left and right derivatives of the function f_j at point x.

The proof of the theorem is similar to that of Theorem 0.1. A brief discussion is given in the appendix. Note that the convergence rate is lower than that of the two sided estimator. Similar results were obtained by Cline and Hart (1991) for density estimation.

In Figure 3 we present the estimated coefficient functions using one-sided and two-sided kernels from a simulated TAR(2) series of (7). The sample size used is 400. Again, the quartic kernel is used, with bandwidth $h = 1.5$ for one-sided kernels and $h = 0.75$ for the two-sided kernel. We can see that, away from the discontinuous point ($x = 0$), both one-sided and two-sided estimates work well. Note that, in the TAR case, away from the discontinuous point, the functions are constant, hence there is no bias in those estimates. Thus there are not much differences between one-sided and two-sided estimates. Around the discontinuous point, $\hat{\boldsymbol{f}}^{+}$ is consistent right of the point, $\hat{\boldsymbol{f}}^{-}$ is consistent left of the point and the two-sided estimate is not consistent. In Figure 4 we plotted $\hat{f}_1^{+}(x) - \hat{f}_1^{-}(x)$ and $\hat{f}_2^{+}(x) - \hat{f}_2^{-}(x)$. We can see that around the discontinuity point $x = 0$, the

differences between the two estimated functions are the largest. Hence, we suggest to use the following statistic to test threshold type of nonlinearity:

$$T = \sup_{x \in \mathcal{D}} \max_{j=1,\dots,p} |\hat{f}_j^+(x) - \hat{f}_j^-(x)|$$

where $\mathcal{D}$ is a compact interval of interest. Since one-sided kernels have severe boundary effects, in our simulation studies we choose $\mathcal{D}$ be the interval between 20 and 80 percentile of the data range, for samples of size of 400.
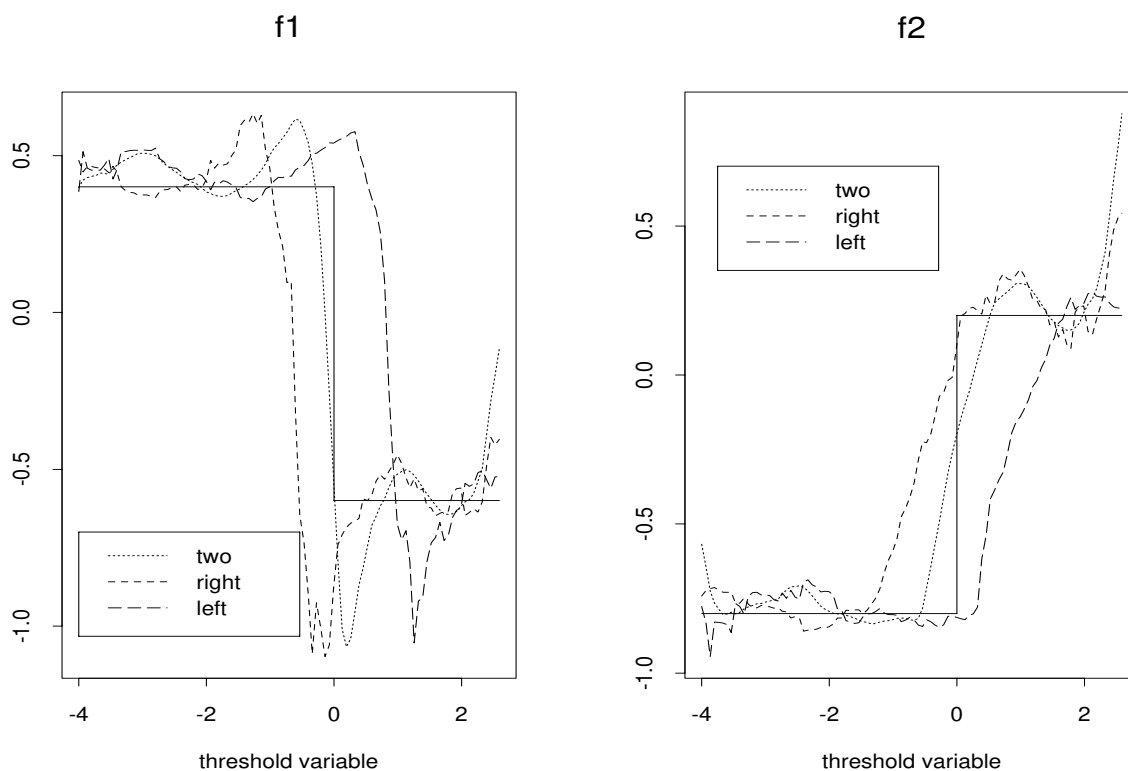

Figure 3: Function estimates using two-sided, left and right sided kernels from a simulated series of model (7). The solid lines are the true functions.

THEOREM 0.5 *Under conditions (A1) to (A3), (C1) to (C4)and the null hypothesis that the functions are all continuous, we have*
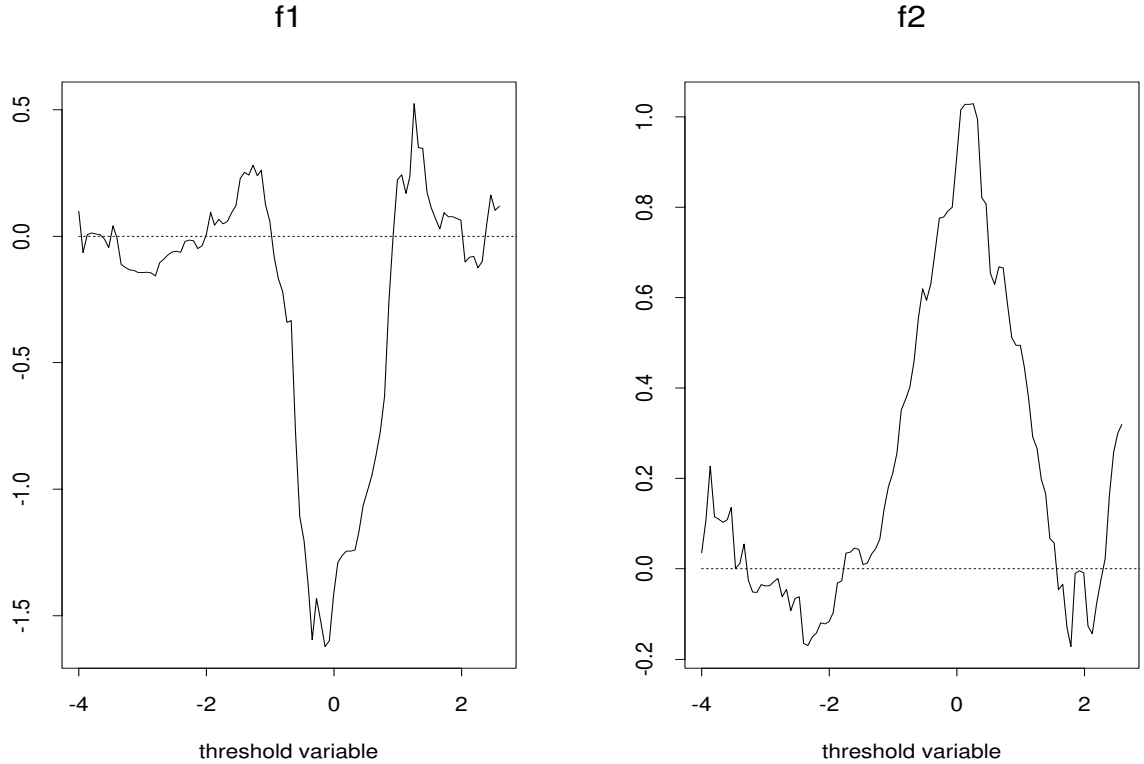
$$T \longrightarrow 0 \quad \text{in probability,}$$

Figure 4: The difference of function estimates using left and right side kernels from a simulated series of model (7)

.

and if there is a discontinuity point in $\mathcal{D}$, then

$$T \longrightarrow \max_{j=1,\dots,p} \sup_{x \in \mathcal{D}} |f_j^+(x) - f_j^-(x)| > 0 \quad \text{in probability.}$$

The above theorem shows that under the null hypothesis the test statistic goes to zero in probability. Hence a large value of the statistic indicates tendency of departing form the null. If the null hypothesis is rejected, the function $\max_{j=1,\dots,p} |\hat{f}_j^+(x) - \hat{f}_j^-(x)|$ can be used to estimate the location of the threshold.

The asymptotic distribution of the test statistic is very difficult to obtain and may not be useful in practice, as the case in section 3. So again we use bootstrap approaches. However, there are several difficulties. Note that, in order to construct bootstrapped version of the data under the null hypothesis (that the functions are continuous in the interval of consideration), we must

estimate the coefficient functions nonparametrically (e.g. using two sided kernels as in section 2.) However, it is a local estimator hence cannot be used outside the data range. In addition, it suffers the boundary effects. So if ones tries to construct a bootstrap version of the time series using $X_t^* = \sum_{i=1}^p \hat{f}_i(X_{t-d}^*)X_{t-i}^* + \varepsilon_t^*$ where ε_t^* is sampled from the residuals $\hat{e}_t = X_t - \sum_{i=1}^p \hat{f}_i(X_{t-d})X_{t-i}$, not only the residual distribution is not correct (due to the boundary effects), but also the generated threshold variable X_{t-d}^* may be out of the range of reliable estimate of f_i. In order to overcome these two difficulties, we convert the problem to a regression setting. Specifically, we fix the original design matrix using the original data and bootstrap only the response. This is slightly different from our time series setting, but under strong mixing condition, the effect will be minimal. In addition, to reduce the boundary effect, we only bootstrap the observation with X_{t-d} within 10 and 90 percentile of the data range. For data outside the range, we always use the original observation. We also only resample from the residuals obtained within the same region.

We performed a small scale simulation to check the performance of the proposed testing method. In addition to models (5), (3), (8) and (7), we also tried the following four models.

$$(\text{STAR-2}) \quad X_t = (0.5 - \frac{1.1}{1 + e^{X_{t-1}}})X_{t-1} + (0.3 - \frac{0.5}{1 + e^{X_{t-1}}})X_{t-2} + \varepsilon_t, \tag{10}$$

$$(\text{TAR-2}) \quad X_t = (0.4 - 0.5I(X_{t-1} > 0))X_{t-1} + (-0.8 + 0.5I(X_{t-1} > 0))X_{t-2} + \varepsilon_t. \tag{11}$$

$$(\text{Tsin}) \quad X_t = (1 - 2I(X_{t-1} > 0))\cos(0.5\pi X_{t-1})X_{t-1} +$$
$$+(-0.5 + I(X_{t-1} > 0))\cos(\pi x_{t-1})x_{t-2} + 0.2\varepsilon_t \tag{12}$$

$$(\text{Tsin-2}) \quad X_t = (0.5 - I(X_{t-1} > 0))\cos(0.5\pi X_{t-1})X_{t-1} +$$
$$+(-0.2 + 0.4I(X_{t-1} > 0))\cos(\pi X_{t-1})X_{t-2} + 0.2\varepsilon_t \tag{13}$$

Table 2 shows the percentage of rejections of the null hypothesis that the coefficient functions are continuous, in 50 simulated samples, each of sample size 400. The differences between TAR and TAR-2, Tsin and Tsin-2 are the jump size. We can see that, with smaller jump size, (TAR2 and Tsin-2), the power of the test is samller, as expected.

<div align="center">APPENDIX</div>

First we list all the necessary assumptions:

(A1) The process is geometrically ergodic. A set of sufficient ergodic conditions for the FAR model can be found in Chen and Tsay (1993) and Cline and Pu (1995).

α	AR(2)	EXPAR(2)	STAR	STAR-2	TAR	TAR-2	TSIN	TSIN-2
0.10	14	6	6	2	96	30	48	40
0.05	6	2	4	2	92	24	38	34
0.01	2	0	2	0	84	12	22	18

Table 2: Percentage of rejections of the continuity test

(A2) The kernel function K is a positive, compactly supported bounded function, with $\int K(u)du = 1$ and $\int uK(u)du = 0$. And $|K(x_1) - K(x_2)| < c|x_1 - x_2|$ for all x_1 and x_2 in its support.

(A3) The density of the stationary distribution exists and is bounded.

(A4) The matrix $\boldsymbol{A}(y) = p(y)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d} = y]$ is of full rank. $\boldsymbol{A}(y)$ and $\boldsymbol{A}^{-1}(y)$ is bounded element-wise in a neighborhood of x.

(A5) Let $p_{i,j,d}$ be the joint density of $(X_{t-i}, X_{t-j}, X_{t-d})$. We assume $p_{i,j,d}$ has Holder continuous first partial derivative with respect to the third argument.

(A6) The second derivative of the coefficient functions exists and are Holder continuous.

(A7) The term
$$\int uv \left\{ \frac{1}{2}f_j''(y)p_{i,j,d}(u,v,y) + f_j'(y)p_{i,j,d}'(u,v,y) \right\} dudv$$
is bounded in a neighborhood of x for all $1 \leq i, j \leq p$.

(A8) $h = \beta n^{-1/5}$, $\beta > 0$.

(B1) The joint density of distinct elements of $\{X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4}, X_{t_5}, X_{t_6}, X_{t_7}, X_{t_8}, X_{t_9}\}$ is continuous and bounded by a constant independent of t_i, $i = 1, \ldots, 9$.

(B2) As $n \to \infty$, then $h \to 0$ and $nh^{(2+4\delta)/(1+\delta)}/\log n \to \infty$.

(C1) $f_j^{'(+)}$ and $f_j^{'(-)}$ exist and Holder continuous in $(x, x + \delta)$ and $(x - \delta, x)$, respectively.

(C2) $h = \beta n^{-1/3}$, for $\beta > 0$.

(C3) The matrix $\boldsymbol{A}(y) = p(y)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d} = y]$ is of full rank. $\boldsymbol{A}(y)$ and $\boldsymbol{A}^{-1}(y)$ is bounded uniformly in a compact intereval $\mathcal{D}$ if interest.

(C4) $f_j^{'(+)}$ and $f_j^{'(-)}$ exist and Holder continuous in $\mathcal{D}$.

We need the following lemmas.

Lemma 0.1 *(Liptser and Shirjaev (1980) Corollary 6). Denote $\mathcal{F}_k$ be a σ-field. Let for every $n > 0$, the sequence $(\eta_{nk}, \mathcal{F}_k)$ be a square integrable martingale difference, i.e.*

$$E(\eta_{nk} \mid \mathcal{F}_{k-1}) = 0, \quad E(\eta_{nk}^2) < \infty, \quad 1 \le k \le n,$$

and let $\sum_{i=1}^n E(\eta_{nk}^2) = 1$, for any $n \ge n_0 \ge 0$. The conditions

$$\sum_{k=1}^n E(\eta_{nk}^2 \mid \mathcal{F}_{k-1}) \xrightarrow{P} 1, \quad as \quad n \to \infty;$$

$$\sum_{k=1}^n E(\eta_{nk}^2 I(|\eta_{nk}| > \epsilon \mid \mathcal{F}_{k-1}) \xrightarrow{p} 0, \quad as \quad n \to \infty,$$

for $\epsilon > 0$ are necessary and sufficient for convergence

$$\sum_{k=1}^n \eta_{nk} \xrightarrow{\mathcal{D}} N(0, 1).$$

Lemma 0.2 *Let p be the stationary density of X_t. Under conditions (A1) (geometric ergodic) we have*

$$n^{-1} \sum_{t=\ell+1}^n X_{t-i} X_{t-j} K_h(X_{t-d} - x) - E[X_{t-i}X_{t-j} \mid X_{t-d} = x]p(x) = o_p(1),$$

and

$$n^{-1} \sum_{t=\ell+1}^n E[X_{t-i} X_{t-j} K_h(X_{t-d} - x)] - E[X_{t-i}X_{t-j} \mid X_{t-d} = x]p(x) = o(1).$$

Proof: By an ergodic theorem, we have

$$n^{-1} \sum_{t=\ell+1}^n X_{t-i} X_{t-j} K_h(X_{t-d} - x) - E[X_{t-i}X_{t-j} K_h(X_{t-d} - x)] = o_p(1).$$

Let $p_{i,j,d}$ be the joint density of $(X_{t-i}, X_{t-j}, X_{t-d})$. We have

$$
\begin{aligned}
E[X_{t-i}X_{t-j} K_h(X_{t-d} - x)] &= \int uv K(w) p_{i,j,d}(u, v, x + hw) du\, dv\, dw \\
&= \int uv p_{i,j,d}(u, v, x) du\, dv\, dw (1 + o(1)) \\
&= p(x) E[X_{t-i}X_{t-j} \mid X_{t-d} = x](1 + o(1)).
\end{aligned}
$$

Proof of Theorem 0.1: Let $n^* = n - \ell$. Let $\boldsymbol{\varepsilon} = (\varepsilon_{t+1}, \ldots, \varepsilon_n)'$. Then

$$\hat{\boldsymbol{f}}(x) - \boldsymbol{f}(x) = (\boldsymbol{X}'W_x\boldsymbol{X})^{-1}\boldsymbol{X}'W_x[\boldsymbol{Y} - \boldsymbol{\varepsilon} - \boldsymbol{X}\boldsymbol{f}(x)] + (\boldsymbol{X}'W_x\boldsymbol{X})^{-1}\boldsymbol{X}'W_x\boldsymbol{\varepsilon} = I_1 + I_2, \quad \text{say.}$$

First we work with $(n^*)^{-1}\boldsymbol{X}'W_x\boldsymbol{X}$. The (i,j)-th element of $(n^*)^{-1}\boldsymbol{X}'W\boldsymbol{X}$ is

$$\frac{1}{n^*}\sum_{t=\ell+1}^{n}X_{t-i}X_{t-j}K_h(X_{t-d}-x) = p(x)E[X_{t-i}X_{t-j} \mid X_{t-d}=x](1+o_p(1)),$$

by Lemma 0.2. Hence $(n^*)^{-1}\boldsymbol{X}'W_x\boldsymbol{X} = \boldsymbol{A}(x)(1+o_p(1))$, where $\boldsymbol{A}(x) = p(x)E[\boldsymbol{X}_t\boldsymbol{X}_t' \mid X_{t-d}=x]$.

Second, the i-th element of $(n^*)^{-1}\boldsymbol{X}'W_x(\boldsymbol{Y}-\boldsymbol{\varepsilon}-\boldsymbol{X}\boldsymbol{f}(x))$ is

$$\begin{aligned}
\frac{1}{n^*}&\sum_{t=\ell+1}^{n}X_{t-i}K_h(X_{t-d}-x)\sum_{j=1}^{p}\{f_j(X_{t-d})-f_j(x)\}X_{t-j}\\
&= \sum_{j=1}^{p}\frac{1}{n^*}\sum_{t=\ell+1}^{n}X_{t-i}X_{t-j}K_h(X_{t-d}-x)\{f_j(X_{t-d})-f_j(x)\}\\
&= \sum_{j=1}^{p}\int uvK(w)\{f_j(x+hw)-f_j(x)\}p_{i,j,d}(u,v,x+hw)dudvdw(1+o_p(1))\\
&= \mu_2h^2\sum_{j=1}^{p}\int uv\left\{\frac{1}{2}f''(x)p_{i,j,d}(u,v,x)+f'(x)p'_{i,j,d}(u,v,x)\right\}dudv(1+o_p(1))\\
&= \mu_2h^2B_i(x)(1+o_p(1)) \quad \text{say.}
\end{aligned}$$

The second equality is the result of the ergodic theorem. Let $\boldsymbol{B}(x)=(B_1(x),\ldots,B_p(x))'$, then

$$I_1 = \mu_2h^2\boldsymbol{A}^{-1}(x)\boldsymbol{B}(x)(1+o_p(1)).$$

Now we work with I_2. The i-th element of $(n^*)^{-1}\boldsymbol{X}'W_x\boldsymbol{\varepsilon}$ is $e_i = (n^*)^{-1}\sum_{t=\ell+1}^{n}X_{t-i}K_h(X_{t-d}-x)\varepsilon_t$. We show that e_i is asymptotically normal by checking all the conditions of Lemma 0.1. First, since ε_t is independent of X_s for all $s<t$, we have $E(e_i)=0$. Standard calculation yields

$$s_i^2 = Var(\varepsilon_i) = \frac{1}{n^{*2}}\sigma^2\sum_{t=\ell+1}^{n}E[X_{t-i}^2K_h^2(X_{t-d}-x)] = \frac{1}{n^*h}\sigma^2K_2^2E[X_{t-i}^2 \mid X_{t-d}=x]p(x)(1+o(1)),$$

where $p_{i,d}$ is the joint density of X_{t-i} and X_{t-d}. Define

$$\eta_t = \frac{1}{n^*}\frac{X_{t-i}K_h(X_{t-d}-x)}{s_i}\varepsilon_t.$$

Note that η_t actually depends on i. For brevity, we here work with a fixed i and suppress the index i on η. Let $\mathcal{F}_t$ be the σ-field generated by $(X_1,\ldots,X_t)$. Since ε_t is independent of X_s for all $s<t$, we have

$$\begin{aligned}
E[\eta_t \mid \mathcal{F}_{t-1}] &= 0,\\
E[\eta_t^2] &= E\left[\frac{X_{t-i}^2K_h^2(X_{t-d}-x)}{n^{*2}s_i^2}\varepsilon_t^2\right] < \infty,
\end{aligned}$$

$$\sum_{t=\ell+1}^{n} E[\eta_t^2] = \frac{\sum_{t=\ell+1}^{n} E[X_{t-i}^2 K_h^2(X_{t-d} - x)\varepsilon_t^2]}{n^{*2} s_i^2} = 1,$$

$$\sum_{t=\ell+1}^{n} E[\eta_t^2 \mid \mathcal{F}_{t-1}] = \frac{\sum_{t=\ell+1}^{n} X_{t-i}^2 K_h^2(X_{t-d} - x)\sigma^2}{n^{*2} s_i^2} \to 1 \quad \text{as} \quad n \to \infty.$$

Finally, for any $\epsilon > 0$, we want to show

$$\sum_{t=\ell+1}^{n} E[\eta_t^2 I(|\eta_t| > \epsilon) \mid \mathcal{F}_{t-1}] = \sum_{t=\ell+1}^{n} \frac{X_{t-i}^2 K_h^2(X_{t-d} - x) E(\varepsilon_t^2 I(|\eta_t| > \epsilon) \mid \mathcal{F}_{t-1})}{n^{*2} s_i^2(x)} = o(1).$$

For some constant C_1 and C, we have

$$
\begin{aligned}
E[\varepsilon_t^2 I(|\eta_t| > \epsilon)] &\leq \{E(\varepsilon_t^4) E[I(|\eta_t| > \epsilon)]\}^{1/2} \\
&\leq C_1 \left[P\left(|\varepsilon_t| > \frac{n^* s_i \epsilon}{|X_{t-i}| K_h(X_{t-d} - x)}\right) \right]^{1/2} \\
&\leq C_1 \frac{\sigma |X_{t-i}| K_h(X_{t-d} - x)}{n^* s_i \epsilon}.
\end{aligned}
$$

Hence,

$$\left| \sum_{t=\ell+1}^{n} E[\eta_t^2 I(|\eta_t| > \epsilon) \mid \mathcal{F}_{t-1}] \right| \leq C \frac{\sum_{t=\ell+1}^{n} |X_{t-i}|^3 K_h^3(X_{t-d} - x)\sigma}{n^3 s_i^3 \epsilon} = o_p(1).$$

Then by Lemma 0.1, we have

$$\frac{1}{s_i} \left[\frac{1}{n^*} \sum_{t=\ell+1}^{n} X_{t-i} K_h(X_{t-d} - x)\varepsilon_t \right] \xrightarrow{\mathcal{D}} N(0, 1).$$

Also note that

$$\frac{1}{n^{*2}} E\left[\left(\sum_{t=\ell+1}^{n} X_{t-i} K_h(X_{t-d} - x)\varepsilon_t \right) \left(\sum_{t=\ell+1}^{n} X_{t-j} K_h(X_{t-d} - x)\varepsilon_t \right) \right]$$

$$= \frac{1}{nh^*} \sigma^2 K_2^2 E[X_{t-i} X_{t-j} \mid X_{t-d} = x] p(x)(1 + o(1)).$$

By a Crame-Wold device, it is easy to show that

$$(n^* h)\left(\frac{1}{n^*} X'W\varepsilon \right) \xrightarrow{\mathcal{D}} N_p(0, \sigma^2 K_2^2 \boldsymbol{A}(x)).$$

Hence

$$(n^* h) I_2 \xrightarrow{\mathcal{D}} N_p(0, \sigma^2 K_2^2 \boldsymbol{A}^{-1}(x)).$$

Let $h = \beta n^{-1/5}$, the theorem follows.

Proof of Theorem 0.4: The proof is essentially the same as that of Theorem 0.1 with slightly differences in the bias calculation since $\int u K^+(u) du = \mu_1^+ \neq 0$. The asymptotic uncorrelation of

$\hat{\boldsymbol{f}}^+(x)$ and $\hat{\boldsymbol{f}}^-(x)$ is due to the fact that the covariance of the i-th element of $(n^*)^{-1}\boldsymbol{X}'W_x^+\boldsymbol{\varepsilon}$ and the j-th element of $(n^*)^{-1}\boldsymbol{X}'W_x^-\boldsymbol{\varepsilon}$ is

$$\frac{1}{n^{*2}}\sum_{t_1=\ell+1}^{n}\sum_{t_2=\ell+1}^{n}E[X_{t_1-i}X_{t_2-j}K_h^+(X_{t_1-d}-x)K_h^-(X_{t_2-d}-x)\varepsilon_{t_1}\varepsilon_{t_2}].$$

Note that for $t_1 < t_2$, ε_{t_2} is independent of the rest of the terms. Since $E(\varepsilon_{t_2}) = 0$, all the terms with $t_1 \neq t_2$ are zero. On the other hand, for $t_1 = t_2$, $K_h^+(X_{t_1-d}-x)K_h^-(X_{t_1-d}-x) = 0$ by definition. Hence $\hat{f}^+(x)$ and $\hat{f}^-(x)$ are asymptotic uncorrelated.

Proof of Theorem 0.5: With the result of theorem 0.4, it is easy to show, by construction of finite open intervals, that for a compact interval,

$$\sup_{x\in\mathcal{D}}|\hat{f}_j^+(x) - f_j^+(x)| \longrightarrow 0 \quad \text{in prob.} \quad \text{and} \quad \sup_{x\in\mathcal{D}}|\hat{f}_j^-(x) - f_j^-(x)| \longrightarrow 0 \quad \text{in prob.}$$

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

$$\left|\sup_{x\in\mathcal{D}}|\hat{f}_j^+(x) - \hat{f}_j^-(x)| - \sup_{x\in\mathcal{D}}|f_j^+(x) - f_j^-(x)|\right| \leq \sup_{x\in\mathcal{D}}|\hat{f}_j^+(x) - f_j^+(x)| + \sup_{x\in\mathcal{D}}|\hat{f}_j^-(x) - f_j^-(x)|$$
$$\longrightarrow 0 \quad \text{in prob.}$$

Hence,

$$\sup_{x\in\mathcal{D}}|\hat{f}_j^+(x) - \hat{f}_j^-(x)| \longrightarrow \sup_{x\in\mathcal{D}}|f_j^+(x) - f_j^-(x)| \quad \text{in prob.}$$

The theorem follows.

References

Chan, K.S. and Tong, H. (1986), On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 179-190

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

Cheng, B. and Tong, H. (1992), On consist non-parametric order determination and chaos (with discussion), *Journal of the Royal Statistical Society, Series B* **54**, 427-474

Cleveland, W. S. and Devlin, S. J. (1988), Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association* **83**: 596–610.

CLINE, D.B.H. AND HART, J.D. (1991), KERNEL ESTIMATION OF DENSITIES WITH DISCONTI-
NUITIES OR DISCONTINUOUS DERIVATIVES, *Statistics.* **22**, 69-84

CLINE, D.B.H. AND PU, H.H. (1995), GEOMETRIC ERGODICITY OF NONPARAMETRIC NON-
LINEAR TIME SERIES, *Technical report, Dept. of Statistics, Texas A & M Univ.*

FAN, J. AND GIJBELS, I. (1996), *Local Polynomial Modeling and Its Applications*, CHAPMAN
& HALL, UK.

GRANGER, C. AND TERÄSVIRTA, T. (1993), *Modeling Nonlinear Economic Relationships*, OX-
FORD UNIVERSITY PRESS, OXFORD

GYÖRFI, L., HÄRDLE, W., SARDA, P. AND VIEU, P. (1989) *Nonparametric Curve Estimation
from Time Series.* LECTURE NOTES IN STATISTICS 60. SPRINGER-VERLAG, HEIDELBERG.

HAGGAN, V. AND OZAKI, T. (1981), MODELING NONLINEAR VIBRATIONS USING AN AMPLITUDE-
DEPENDENT AUTOREGRESSIVE TIME SERIES MODEL, *Biometrika* **68**: 189–196.

HÄRDLE, W., LÜTKEPOHL, H. AND CHEN, R. (1997), A REVIEW OF NONPARAMETRIC TIME
SERIES ANALYSIS, *International Statistical Review*, **65**, 49-72

HÄRDLE, W. AND MAMMEN, E. (1993), COMPARING NONPARAMETRIC VERSUS PARAMETRIC
REGRESSION FITS. *Annals of Statistics.* **21**, 1926-1947

HÄRDLE, W. AND VIEU, P. (1992), KERNEL REGRESSION SMOOTHING OF TIME SERIES, *Journal
of Time Series Analysis* **13**: 209–232.

HART, J.D. (1996), SOME AUTOMATED METHODS OF SMOOTHING TIME-DEPENDENT DATA. *J.
Nonparametric Statistics*, **6**, 115-142

HASTIE, T. J. AND TIBSHIRANI, R. J. (1993), VARYING-COEFFICIENT MODELS JRSSB, **55**,
757-796

HJELLVIK, V. AND TJØSTHEIM, D. (1995), NONPARAMETRIC TESTS OF LINEARITY FOR TIME
SERIES, *Biometrika*, **82**, 351-368

HJELLVIK, V. AND TJØSTHEIM, D. (1996), NONPARAMETRIC STATISTICS FOR TESTING OF
LINEARITY AND SERIAL INDEPENDENCE. *J. Nonparametrics*, IN PRESS

HJELLVIK, V., YAO, Q. AND TJØSTHEIM, D. (1997), LINEARITY TESING USING LOCAL POLYNOMIAL APPROXIMATION, *Technical report, University of Bergen, Norway.*

KEENAN (1985), D.M. (1985), A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, **72**, 39-44

LIPSTER, R. SH. AND SHIRJAEV, A.N. (1980), A functional central limit theorem for martingales. *Theory of Probability and its Applications*, **25**, 668-668

LUUKKONEN, R.P. SAIKKONEN, P. AND TERÄSVIRTA, T. (1988), Testing linearity against smooth transition autoregression. *Biometrika*, **75**, 491-499

PARK, B.U. AND MARRON, J.S. (1990), Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, **85**, 66-72

PRIESTLEY, M. B. (1988), *Non-linear and Non-stationary Time Series Analysis*, Academic Press, New York.

RUPPERT, D., SHEATHER, S.J. AND WAND, M.P. (1995), An efficient bandwidth selector for local least squares regression. *JASA*, **90**, 1257-1270

SHEATHER, S.J. (1983), A data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics and Data Analysis*, **1**, 229-238

SHEATHER, S.J. (1986), An improved data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics and Data Analysis*, **4**, 61-65

SKAUG, H.J. AND TJØSTHEIM, D. (1993), Non-parametric tests of serial independence. *The M. Priestley Birthday Volume* (ed. T. Subba Rao), pp. 207-229

TERÄSVIRTA, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208-218.

TJØSTHEIM, D. (1994), Nonlinear time series, a selective review, *Scand. J. Statist*, **21**, 97-130.

TONG, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Vol. 21 of *Lecture Notes in Statistics*, Springer, Heidelberg.

TONG, H. (1990), *Nonlinear Time Series Analysis: A Dynamic Approach*, Oxford University Press, Oxford.

TSAY, R.S. (1986), Nonlinearity tests for time series. *Biometrika*, **73**, 461-466

TSAY, R.S. (1989), Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84, 231-240.

TSYBAKOV, A.B. (1986), Robust reconstruction of functions by the local approximation method, *Problems of Information Transmission*, **22**, 133-146

YOSHIHARA, K. (1976), Limiting behavior of U-statistics for stationary absolutely regular process. *Z. Wahr. v. Gebiete*, **35**, 237-252