May, 2009

Nonlinear Forecasting Using Nonparametric Transfer Function Models

Jun Liu, *Georgia Southern University*

Nonlinear Forecasting Using Nonparametric Transfer Function Models

Jun M. Liu
Georgia Southern University
Department of Finance & Quantitative Analysis
Statesboro, GA 30460
USA
jliu@georgiasouthern.edu

Abstract: The focus of this paper is using nonparametric transfer function models in forecasting. Nonparametric smoothing methods are used to model the relationship between variables (the transfer function) and the noise is modeled as an Autoregressive Moving Average (ARMA) process. The transfer function is estimated jointly with the ARMA parameters. Nonparametric smoothing methods are flexible thus can be used to model highly nonlinear relationships between variables. In this paper polynomial splines are used to model the transfer function. Modeling noise term as an ARMA process removes the serial correlation so the transfer function can be estimated efficiently. As a result, the nonparametric transfer function model can generate accurate forecasts when the transfer function is highly nonlinear with unknown functional form. The proposed polynomial splines-based estimator is also highly computationally efficient. The performance of nonparametric transfer function models is demonstrated in this paper by forecasting river flow based on temperature and precipitation. A comparison of the results show that the performance of this model is better than some widely accepted benchmark models.

Key–Words: Nonparametric smoothing, Time series, Forecast

1 Introduction

In this paper we consider a new method to model relationships between 'output' and 'input' time series. Exploring relationships between variables has been a constant interest of researchers, and extensive research has been conducted in this area. For example, the linear transfer function model (Box and Jenkins, 1976) has been extensively used in practice and proven successful in many practical areas. However, in practice we often encounter nonlinear relationships that cannot be well approximated by linear models. Consequently, nonlinear parametric models are introduced (Chen and Tsay, 1996; Tong, 1990; Haggan and Ozaki, 1981; Engel, 1982; Bollerslev, 1986). One problem with nonlinear parametric model is, beyond the linear domain there are infinitely many candidate nonlinear functions, so it is usually difficult to justify the explicit parametric functional forms a priori. To avoid the subjectivity in selecting the parametric models, researchers adopt the principle of "letting the data speak for themselves" and use nonparametric smoothing methods to model nonlinear time series (Robinson, 1983; Auestad and Tjøstheim, 1990; Lewis and Stevens, 1991; Masry, 1996a&b; Fan and Gilbels, 1996; Smith, Wong, and Kohn, 1998; Aydin and Omay, 2006;). To overcome the 'curse of dimensionality', various specially structured nonparametric models have been proposed, including the functional-coefficient autoregressive (FAR) model (Chen and Tsay, 1993a; Cai, Fan and Yao, 2000), the nonlinear additive autoregressive model (Chen and Tsay, 1993b), the adaptive functional-coefficient model (Ichimura, 1993; Xia and Li, 1999; Fan, Yao and Cai, 2003), the single index model (e.g., Härdle, Hall, and Ichimura, 1993; Carroll, Fan, Gijbels, and Wand, 1997; Newey and Stoker, 1993; Heckman, Ichimura, Smith, and Todd, 1998; Xia, Tong, Li, and Zhu, 2002) and the partially linear models (Härdle, Liang and Gao, 2000). The literature about nonlinear and nonparametric time series analysis is extensive, reviews can be found in Tjøstheim (1994), Härdle, Lütkepohl and Chen (1997) and Fan and Yao (2003).

In this paper we consider the following relationship between two time series:

$$Y_t = f(X_t) + e_t, \tag{1}$$

where $f(\cdot)$ is an unknown and smooth function, $\{X_t, e_t\}$ are jointly strictly stationary. Recently Xiao, Linton, Carroll and Mammen (2003), Su and Ullah (2006), and Liu, Chen and Yao (2008) developed methods to estimate the transfer function efficiently. In their studies local polynomial is used to model the transfer function $f(\cdot)$. They showed that by modeling the serial correlation in the noise, $f(\cdot)$ can be estimated at the usual rate of convergence as if e_t is iid.

The above methods differs mainly in the treatment of the noise e_t. Xiao, et al. (2003) assumes the noise is a general linear process and approximates it by an AR process whose order is allowed to grow to infinity with the sample size. Su and Ullah (2006) assumes the noise is a finite-order nonparametric AR process. Liu, Chen and Yao (2008) models the noise explicitly with an ARMA model. The above methods are all computationally intensive because of the use of local polynomial. As a result they may be difficult to apply in certain practical situations, for example, it may require a very long time to generate multiple step ahead forecast by simulation.

Another drawback of the local polynomial-based estimators is that they are difficult to apply when the noise $\{e_t\}$ is non-stationary, which is common in practice. When the noise is non-stationary, local polynomial estimators assuming independent noise (the "conventional" estimators) are no longer consistent. The consistency of the "conventional" estimator plays a key role in the local-polynomial based estimators, without it, these estimators can no longer be used. One possible solution to handle non-stationarity is to take differences until the noise becomes stationary (e.g., Box and Jenkins 1976), however taking difference in the local polynomial estimator makes the estimation very difficult, if not impossible. New estimators are needed to handle non-stationary noise. In this paper polynomial spline is used to model the transfer function. The explicit functional form of polynomial splines not only significantly simplifies the estimation, it also makes the extension to non-stationary noise straightforward. As a result, the noise is allowed to follow an Autoregressive Integrated Moving Average (ARIMA) process.

By modeling the transfer function $f(\cdot)$ nonparametrically, the model is flexible therefore can be used to model highly nonlinear relationships of unknown functional forms. By modeling $\{e_t\}$ explicitly, the autocorrelation in the data is removed so $f(\cdot)$ can be estimated efficiently. Additionally, the explicit correlation structure can be used to improve the forecasting performance.

This paper is organized as follows. In section 2, the model is introduced and a short introduction of polynomial spline is included. The estimation method is introduced in section 3, the estimator when the noise is non-stationary is also introduced in this section. Section 4 illustrates the finite-sample performance of the estimator through simulation. The proposed procedures are applied to forecast river flow and the results are presented in section 5. Section 6 contains summary and discussion.

2 The model

In this paper we make the assumption that $\{e_t\}$ in model (1) follows a strictly stationary AR(p) process, $e_t = \sum_{i=1}^{p} \phi_i e_{t-i} + \varepsilon_t$. So model (1) can be rewritten as

$$Y_t = f(X_t) + \frac{\varepsilon_t}{1 - \sum_{i=1}^{p} \phi_i B^i}\ , \qquad (2)$$

where B is the back-shift operator, $B^i X_t = X_{t-i}$, $\{\varepsilon_t\}$ is a sequence of independent random variables with mean 0 and standard deviation σ. Note that the assumption of AR(p) noise is mainly for the convenience of discussion, the idea presented in this paper can be extended to more general structures of the noise, such as the ARMA(p, q) model. We also assume that $\{X_t\}$ and $\{\varepsilon_t\}$ are independent. Our interest is in estimating both $f(\cdot)$ and the AR parameters.

In this paper we use polynomials to model the transfer function f. Polynomial splines are piecewise polynomials defined on disjoint partitions of the support of X, with the pieces joining smoothly at a set of interior points (the *knots*). More precisely, a polynomial spline of degree $m \geq 0$ defined on an interval $\mathcal{X}$ with knot sequence $\boldsymbol{\lambda} = \{\lambda_0, \lambda_1, \cdots, \lambda_{k+1}\}$ ($\lambda_0 < \lambda_1 < \cdots < \lambda_{k+1}$) is a function consisting of pieces of polynomials of degree m on each of the intervals $[\lambda_i, \lambda_{i+1}),\ i = 0, \cdots, k$, and $[\lambda_k, \lambda_{k+1}]$ (where λ_0 and λ_{k+1} are the end points of $\mathcal{X}$), connected smoothly at the knots. Given knot sequence $\boldsymbol{\lambda}$ and degree m, the collection of spline functions form a function space spanned by basis functions. Commonly used basis functions include the well-known truncated power basis, which is the set of functions $\{1, x, \cdots, x^m, (x - \lambda_1)_+^m, \cdots, (x - \lambda_k)_+^m\}$, where $(x)_+^m \equiv (x_+)^m$, the dimension of the spline function space is given by $K = m + k + 1$. B-spline is often used to develop the asymptotic properties because of its nice theoretical properties (for details please see de Boor, 2001; Schumaker, 1981), but the result does not depend on the choice of the basis functions. Polynomial splines have the flexibility of nonparametric smoothers because they allow the function to be different polynomial in different intervals. On the other hand, the explicit functional form of polynomial splines makes them very computationally efficient. As a result, they have been studied extensively (e.g., Huang 2003; Huang and Shen 2004; Wang and Yang 2007; Wang and Yang 2009; Phokharatkul, Kamnuanchai, Kimpan, and Phaiboon, 2006; Gáti and Bede, 2006). Denote a set of basis functions as $\{B_j(\cdot)\}_{j=1}^{K}$, using the polynomial spline to approximate the transfer function $f(\cdot)$ in (2), $f(X_t) \approx \sum_{i=1}^{K} a_i B_i(X_t)$, after "pre-whitening" the noise e_t, we have the following regression model

$$Y_t \approx \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{K} a_j \Big[B_j(X_t) - \sum_{i=1}^{p} \phi_i B_j(X_{t-i}) \Big] + \varepsilon_t \ , \quad (3)$$

the estimation of the unknown parameters are carried out by solving the following optimization problem

$$\arg_{a_j,\phi} \min \sum_{t=1}^{n} \Big\{ Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} - \sum_{j=1}^{K} a_j \Big[B_j(X_t) - \sum_{i=1}^{p} \phi_i B_j(X_{t-i}) \Big] \Big\}^2 . \quad (4)$$

3 Estimation methodology

The optimization of (4) can be carried out by standard nonlinear optimization methods. In this paper we consider the following iterative estimation algorithm, which is equivalent to the commonly used nonlinear optimization methods, such as the Gauss-Newton method. This algorithm allows us to investigate the estimators individually, which makes the discussion of the asymptotic properties more convenient.

1. Obtain preliminary estimates $\widetilde{a}_i, \ i = 1, \cdots, K$, which are the solutions of

$$\arg_{a_i} \min \sum_{t=1}^{n} \Big\{ Y_t - \sum_{i=1}^{K} a_i B_i(X_t) \Big\}^2 ,$$

   a preliminary estimate of f is given by $\widetilde{f}(x) = \sum_{i=1}^{K} \widetilde{a}_i B_i(x)$.

2. For given $a_j, \ j = 1, \cdots, K$, obtain $\widehat{\phi}_1, \cdots, \widehat{\phi}_p$ by solving

$$\arg_{\phi_i} \min \sum_{t=1}^{n} \Big\{ Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} - \sum_{j=1}^{K} a_j \Big[B_j(X_t) - \sum_{i=1}^{p} \phi_i B_j(X_{t-i}) \Big] \Big\}^2$$

3. For given $\phi_1, \cdots, \phi_p$, obtain $\widehat{a_j}, \ j = 1, \cdots, K$ by solving

$$\arg_{a_j} \min \sum_{t=1}^{n} \Big\{ Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} - \sum_{j=1}^{K} a_j \Big[B_j(X_t) - \sum_{i=1}^{p} \phi_i B_j(X_{t-i}) \Big] \Big\}^2$$

It can be easily seen that the above algorithm is guaranteed to converge. When the noise follows the more general ARMA process, the convergence is no longer guaranteed but our experience in the simulation indicates the convergence can be expected in most of the cases. The terminating values of $\widehat{\phi}$ and $\widehat{a_i}$ provide the final estimates, specifically, $\widehat{f}(x) = \sum_{j=1}^{K} \widehat{a_j} B_j(x)$ is the final estimate of $f(x)$. e_t is assumed to follow a stationary ARMA process, so it is a mixing process, and we can expect that the preliminary estimate $\widetilde{f}$ to be consistent. The ARMA parameters can be estimated with the parametric rate. By modeling the serial correlation in e_t, the transfer function f can be estimated as if e_t is iid and the asymptotic results established in Huang (2003) continue to hold.

3.1 Non-stationary Noise

The estimation procedure above can be extended easily to handle non-stationary noise. Here we focus on the unit root case and model the noise $\{e_t\}$ as an autoregressive integrated moving average (ARIMA) process $\phi(B)\nabla^d e_t = \theta(B)\varepsilon_t$, where $\nabla^d = (1 - B)^d$ is the d-th power of the difference operator. When $\{e_t\}$ is nonstationary, it is no longer a mixing process so standard nonparametric smoothing results no longer apply. As a result, the existing local polynomial-based estimation procedures designed for stationary $\{e_t\}$ (e.g., Xiao et al., 2003; Su and Ullah, 2006; and Liu et al., 2008) can no longer be used. In time series analysis, one commonly adopted method to deal with unit root is to take differences so that the resulting series becomes stationary and standard estimation procedures can be applied (Box and Jenkins, 1976). Unfortunately, this idea does not work well with the local polynomial-based estimation procedures. To illustrate, consider the following model in which $\{e_t\}$ follows a random walk process, $Y_t = f(X_t) + e_t, \ e_t = e_{t-1} + \varepsilon_t$, after the first-order difference, we have $Y_t - Y_{t-1} = f(X_t) - f(X_{t-1}) + \varepsilon_t$. f can be estimated by solving the following optimization problem

$$\inf_f \sum_{t=1}^{n} \Big\{ Y_t - f(X_t) - \Big[Y_{t-1} - f(X_{t-1}) \Big] \Big\}^2 .$$

However, the two f's must be restricted to be the same during the estimation. This task is very difficult, if not impossible, for local polynomial-based estimators. On the other hand, because of the explicit parametric form of polynomial splines (given the order of the polynomial and knot sequence), to impose the same restrictions on $f(\cdot)$ one only needs to restrict the coefficients to be the same. For identifiability we must assume $E(f(\cdot)) = 0$. Let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)^\tau$ and $\boldsymbol{\beta} = (\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q)^\tau$, the estimation is carried out by solving the following optimization problem,

$$\arg_{\boldsymbol{\alpha},\boldsymbol{\beta}} \min \sum_{t=1}^{n} \left\{ \frac{\phi(B)\nabla^d}{\theta(B)} \left[Y_t - \sum_{i=1}^{K} \alpha_i B_i(X_t) \right] \right\}^2.$$

After differencing the noise becomes an ARMA(p, q) process and the estimation can be carried out using previous results. Thus polynomial spline provides a viable solution to the problem of modeling non-stationary noise.

4 Numeric Properties

Simulation is conducted to illustrate the finite sample properties of the proposed estimator. In the simulation below, we use $f(x) = x + 2\exp(-16x^2)$, X_t is generated from an AR(1) process, $X_t = 0.3X_{t-1} + a_t$, $a_t \sim \mathrm{N}(0, 0.5^2)$, and e_t is generated from an AR(1) model with different values of ϕ: $e_t = \phi e_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathrm{N}(0, 0.5^2)$. For the ease of implementation the truncated power basis is used. In the simulation we used a cubic spline, i.e., $m = 3$. The knots are placed at the percentile points such that there are equal number of observations between two adjacent knots. The number of knots is set as a multiple of the theoretically optimal number $[n^{1/2m+3}]$. Three sample sizes (200, 500, and 1000) are used and the simulations are run 200 replications. In the simulation the mean squared errors

$$\mathrm{MSE} = \frac{1}{n} \sum_{t=1}^{n} \left\{ f(X_t) - \widehat{f}(X_t) \right\}^2,$$

of the proposed estimator are averaged over the replications and compared to that of the “conventional” estimator in which e_t is assumed to be iid. The simulation results are summarized in Table 1 below. In this table MSE_1 is the average mean squared error of conventional estimator assuming iid noise, and MSE_2 is the average MSE of the proposed estimator. RLMSE=MSE_2/MSE_1 is the relative MSE which shows the gain in efficiency in estimating f. To illustrate the results, a histogram of $\widehat{\phi}$ and a plot of $\widehat{f}$ are given in Figure 1.

From the above results, we have the following observations:

1. The relative MSE in most of the cases is less than one, which shows that modeling the serial correlation in the noise can improve the efficiency in estimating f.

Table 1: Simulation results when $\{e_t\}$ follows AR(1) models

ϕ	n	mean($\widehat{\phi}$)	$s_{\widehat{\phi}}$	MSE_1	MSE_2	RLMSE
	200	-.7959	.0409	.0389	.0095	.2433
-.8	500	-.8016	.0264	.0151	.0036	.2388
	1000	-.8016	.0206	.0083	.0019	.2238
	200	-.5171	.0608	.0193	.0122	.6305
-.5	500	-.5064	.0365	.0079	.0051	.6437
	1000	-.5045	.0271	.0042	.0026	.6043
	200	-.2075	.0799	.0159	.0152	.9531
-.2	500	-.2081	.0449	.0067	.0062	.9298
	1000	-.2032	.0315	.0035	.0033	.9429
	200	-.0084	.0802	.0163	.0170	1.040
0	500	-.0013	.0487	.0067	.0068	1.007
	1000	.0000	.0331	.0033	.0034	1.006
	200	.1905	.0753	.0184	.0176	.9605
.2	500	.2015	.0461	.0072	.0068	.9435
	1000	.1997	.0322	.0039	.0037	.9383
	200	.5066	.0649	.0262	.0180	.6846
.5	500	.4988	.0388	.0111	.0075	.6769
	1000	.4989	.0296	.0059	.0040	.6717
	200	.7955	.0414	.0796	.0454	.5703
.8	500	.7999	.0259	.0320	.0169	.5270
	1000	.7966	.0211	.0166	.0085	.5082

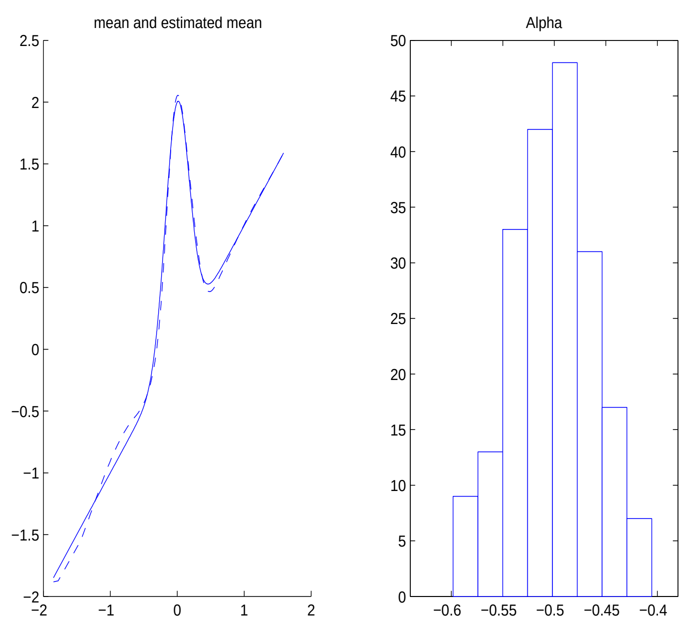

Figure 1: $\phi = -0.5$, n=500. Left panel: histogram of $\widehat{\phi}$, right panel: f (solid) and $\widehat{f}$ (dashed) in a typical simulation.

2. The relative MSE is closely related to the strength of correlation in e_t, the stronger the correlation, the smaller the RLMSE, hence the larger the gain in efficiency.

3. The histogram in Figure 1 shows the sampling distribution of $\widehat{\phi}$ is close to a normal distribution.

4. The performance of the estimators improves with the sample size.

Similar observations are made in Liu et.al. (2008), but the polynomial splines used in this paper greatly simplifies the estimation. For example, in Liu et. al. (2008), with a sample size of 400, it normally takes several minutes to complete the estimation, while the same estimate only takes about ten seconds with the spline-based approach proposed in this paper. As a result, much larger sample sizes become more affordable. Computational efficiency is of special importance in time series forecasting, because time series under consideration are typically long, especially in financial data. Being able to generate forecasts in a timely manner when sample size is large has a direct impact in the applicability of the model.

An important application of the proposed method is to estimate nonparametric transfer function models with non-stationary noise. To study the finite sample properties under such situations, we consider a simple random walk process for the noise, i.e., $e_t = e_{t-1} + \varepsilon_t$. Note that in this case because of the differencing in the estimation, the constant term is not estimable so f can only be estimated up to a constant. In practice the constant may be estimated using the mean of the observations Y_t, this is also the approach we adopt in the simulation below. Similar comparison of the average mean squared errors between the proposed and the conventional estimators is made and the results are summarized in Table 2 below. The estimated transfer functions in six simulations are plotted in Figure 2.

Table 2: Simulation results when $\{e_t\}$ is a random walk

n	MSE_1	MSE_2	RLMSE
200	.5162	.0206	.0399
500	.4702	.008	.0171
1000	.4512	.0038	.0085
2000	.4407	.0021	.0048

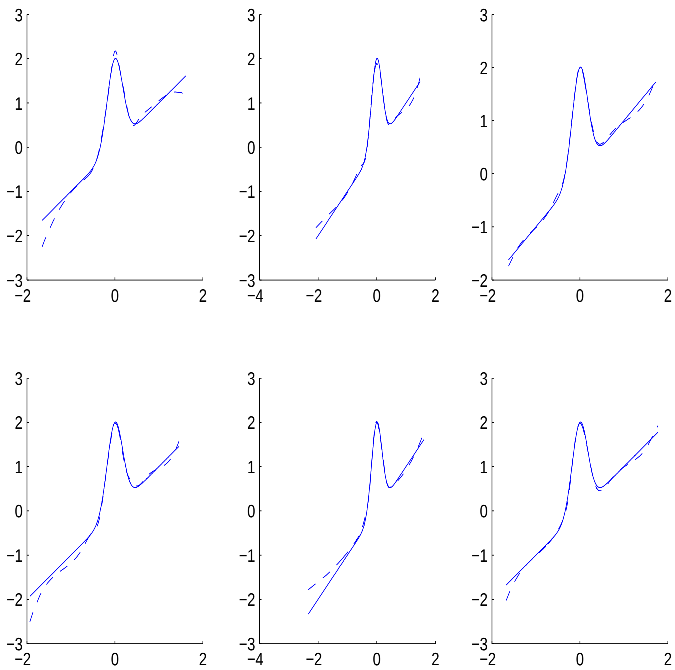

Figure 2: f (solid) and $\hat{f}$ (dashed) in six typical simulations.

5 Forecasting river flow

River flow forecasting is an important issue in many practical areas. For example, accurate and efficient forecast of river flow is important to the safety of river transportation and river structures such as river dams and bridges, it also enables us to manage water resources more efficiently and helps protect the environment. A typical difficulty in river flow modeling is that many factors affect the river flow, usually in a highly nonlinear fashion. Non-linear parametric models have been developed to address this issue (some highly-regarded models include Tong, Thanoon and Gudmundsson, 1985; Chen and Tsay, 1993b). A difficulty in nonlinear parametric models is, with the indefinitely many functions in the nonlinear domain, it is very difficult to identify the best candidate function a priori, therefore there is usually more subjectivity in the model selection. The subjectivity can be avoided by using nonparametric methods, because of the "let the data speak for themselves" property.

In this section we use the proposed nonparametric transfer function model to study the effects of temperature and precipitation on the flow rate of River Jökulsá eystri in Iceland. The data consists of daily records of river flow Y_t (in m^3/sec), temperature X_t (in °C), and precipitation Z_t (in mm) from January 1, 1972 to December 31, 1974. An interesting feature of this river is that there is a glacier in its drainage area, so we expect that the temperature effect on river flow is more than melting snow. Tong et. al. (1985) used the threshold autoregressive model (TAR) to analyze this data set. Chen and Tsay (1993b) used it as an example of the nonlinear additive ARX model (NAARX). For more detailed information of the data, see Tong, et.al. (1985).

To apply the proposed approach, we first check the stationarity of the data. The sample ACF and PACF of Y_t show indications of non-stationarity, (Figure 3), however the result of Augmented Dicky-Fuller (ADF) test rejects the hypothesis that a unit root exists. Similar analysis shows that the exogenous variables X_t and Z_t are stationary. The details of the ADF tests are omitted to save space. To obtain some rough idea about the candidate variables of the model, as an initial step of model identification, we estimated the linear transfer function weights using the linear transfer function method (Liu, 1982). The estimated linear transfer function weights are plotted against the lags in Figure 4, with the 95% confidence band plotted in the dashed lines. The estimated transfer function weights suggest the following variables are good candidates of the model: $\{X_t, X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, Z_t, Z_{t-1}, Z_{t-6}\}$.

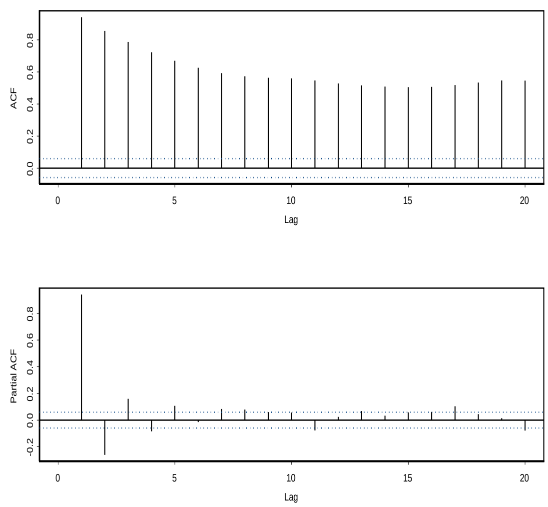

Figure 3: Upper panel: The sample ACF of Y_t. Lower panel: The sample PACF plot of Y_t.

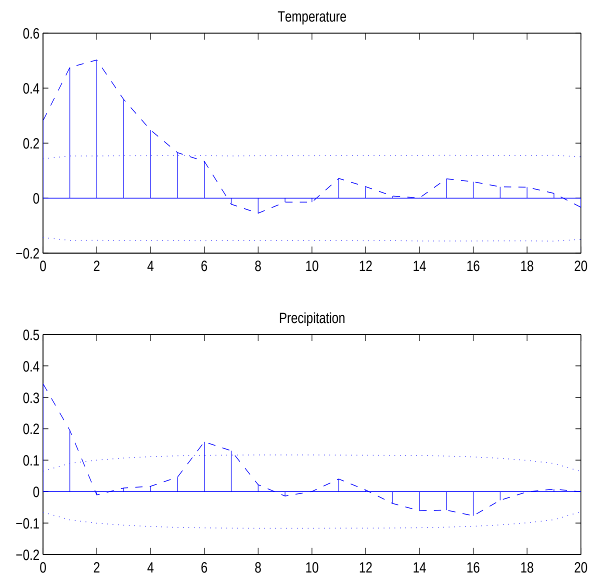

Figure 4: The estimated linear transfer function weights

A common problem with high-dimensional nonparametric smoothing is the *curse of dimensionality*. As the dimension of the model increases, the sample size needed to obtain stable “local” estimate increases exponentially with the sample size. Because of this problem, with the sample sizes typically available in practice, the dimension of nonparametric smoothing models are typically low in practice. Without any restriction, a nonparametric regression model containing all the above variables is difficult to estimate. To overcome this problem, we consider an additive model

$$\begin{aligned} Y_t &= \sum_{i=0}^{4} f_i(X_{t-i}) + g_0(Z_t) + g_1(Z_{t-1}) \\ &+ g_6(Z_{t-6}) + e_t, \end{aligned}$$

where each additive component is approximated by regression spline. The truncated power basis is used in the spline approximation. To simplify discussion, we assume that the orders and the number of knots of the spline functions are the same, thus the model can be written as

$$\begin{aligned} Y_t &= a_0 + \sum_{i=0}^{4}\Big[\sum_{j=1}^{k} a_{ij}X_{t-i}^j + \sum_{r=1}^{m} a_{ir}(X_{t-i}-\lambda_r)_+^k\Big] \\ &+ \sum_{i=0}^{1}\Big[\sum_{j=1}^{k} b_{ij}Z_{t-i}^j + \sum_{r=1}^{m} b_{ir}(Z_{t-i}-\lambda_r)_+^k\Big] + e_t. \end{aligned}$$

We further assume that the knots are placed on the percentile points so that there are equal number of observations between any two adjacent knots. A grid search is conducted to determine the number of interior knots k and spline degree m. In the grid search k is in the interval [1, int($5n^{1/(2m+3)}$)], where the upper limit is a multiple of the theoretical optimum number of knots (Huang, 2003). We consider linear, quadratic and cubic splines, ($m = 1, 2, 3$, respectively). The values k and m that minimizes BIC=$\log(MSE) + \log(n)[1 + (k + m)(d_x + d_z)]/n$ is the optimal number of knots, where d_x and d_z are the number of lags in X_t and Z_t, respectively. The results suggest a linear spline model with $k = 1$. Similar to the case of Y_t, the sample ACF and PACF of the partial residual $\widetilde{e}_t$ show indications of nonstationarity, however the ADF test results again reject the existence of a unit root. As a result, an AR(4) model is selected using the BIC criterion, the sample PACF (Figure 5) also suggests such a model. Based on the

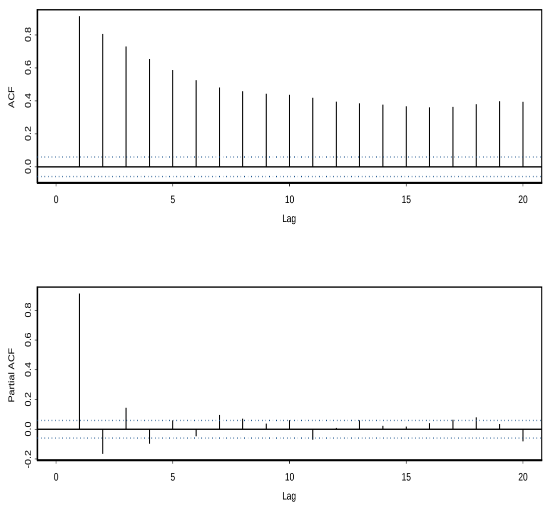

Figure 5: Sample ACF and PACF of the preliminary residual $\widetilde{e}_t$

above preliminary information about the underlying model, we refine the model by selecting the knots locations to minimize the residual sum of squares. The results show that Z_t and Z_{t-1} have such large knots that beyond these knots there are only a few observations, this indicates that their effects are essentially linear, also, X_{t-4} and Z_{t-6} are found to be insignificant. As a result, the model simplifies further to

$$\begin{aligned} Y_t &= c + \sum_{i=0}^{3} \left[a_{i1} X_{t-i} + a_{i2}(X_{t-i} - \lambda_i)_+ \right] \\ &+ \sum_{j=0}^{1} b_j Z_{t-j} + \frac{\varepsilon_t}{1 - \sum_{i=1}^{4} \phi_i B^i}. \end{aligned} \tag{5}$$

The optimized knots are -1.3, 0.5, 0.2, and -0.2 for X_t, X_{t-1}, X_{t-2}, and X_{t-3}, respectively, and the estimated parameters and their stand deviations are given in Table 3.

The sample ACF and PACF in Figure 6 show that the residual series is roughly "white".

To put the performance of the proposed NPTF model in perspective, we consider two widely-accepted "benchmark" models that were used to analyze the same data set. The first model is the TAR

Table 3: Parameter Estimates

Parameter	c	a_0	a_{01}	a_{11}	a_{12}
Estimate	29.67	.03	.47	-.11	2.44
StdErr.	.08	.18	.08	.17	.03
Parameter	a_{21}	a_{22}	a_{31}	a_{32}	b_0
Estimate	.05	1.42	.03	.61	.32
StdErr.	2.24	.08	.16	.08	.18
Parameter	b_1	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Estimate	.17	1.15	-.40	.25	-.09
StdErr	.03	.03	.05	.05	.03

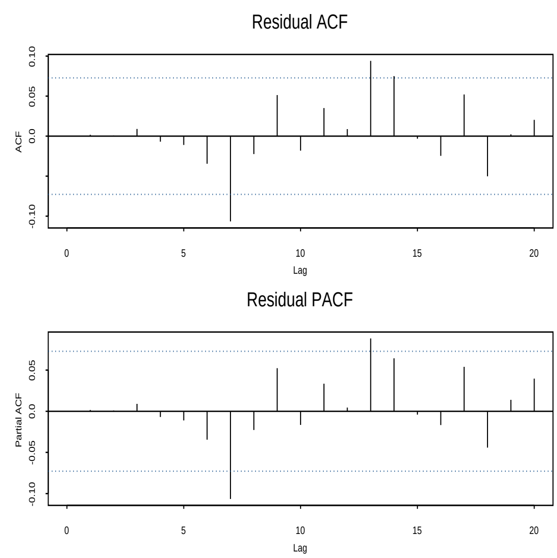

Figure 6: The estimated residual ACF and PACF

model used by Tong, et.al.(1985, page 658),

$$\begin{aligned} Y_t &= c_1 + (1,\cdots,6)Y_t + (0,\cdots,5)Z_t \\ &\quad +(0,\cdots,3)X_t + e_{1t}, \text{ if } X_t \le -2 \\ &= c_2 + (1,\cdots,8)Y_t + (0,\cdots,7)Z_t \\ &\quad +(0,\cdots,7)X_t + e_{2t}, \text{ if } X_t > -2, \quad (6) \end{aligned}$$

where $(1,\cdots,6)Y_t$ means that $Y_{t-1},\cdots,Y_{t-6}$ are included in the model, other terms are similarly defined. The second model is the NARRX model used by Chen and Tsay (1993b, page 963),

$$\begin{aligned} Y_t &= c + \phi_{1,1}Y_{t-1} + \phi_{1,2}Y_{t-1}I(Y_{t-1} \ge c_1) \\ &+ \phi_{1,3}Y_{t-1}I(Y_{t-1} \ge c_2) + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} \\ &+ \phi_4 Y_{t-4} + \beta_1 Z_t + \beta_2 Z_{t-1} + \omega_{1,1}X_{t-1} \\ &+ \omega_{1,2}X_{t-1}I(X_{t-1} \ge c_3) + \omega_{3,1}X_{t-3} \\ &+ \omega_{3,2}X_{t-3}I(X_{t-3} \ge c_4) + \varepsilon_t \quad (7) \end{aligned}$$

The residual variances of the NPTF model, together with the residual variances of the NAARX and the TAR (Chen and Tsay 1993b) are shown in the last row of Table 4. We can see while the within-sample performance of NPTF, NAARX and TAR are similar, the NPTF model has the smallest residual variance in the three models. Although the NPTF model uses two more parameters than the NAARX model, it is still preferred by the AIC criterion. It is interesting to see that the NPTF model (5) and the NAARX model reveals similar features of the underlying process, for example, in both models piecewise linear functions are found to well describe the relationship between temperature and river flow; in both models the precipitation effect is linear. The main difference is that in the NPTF model (5) an AR model is used to model the noise to account for the serial correlation, while in the NAARX model lags of Y_t is used.

To study the forecasting performance of the proposed model, we consider the following post-sample forecast scheme: model (5) is re-estimated using the first two years of data, one-step to 12-step ahead post-sample forecasts are conducted using the data of the third year. This forecasting scheme is similar to the one used in Chen and Tsay (1993b), the main difference is that in Chen and Tsay (1993b), actual observations of X_t, Z_t and their lags are used in the forecasts, while here the forecast values are used. Two simple AR(1) models are found appropriate for this purpose: $X_t = \phi_x X_{t-1} + a_{1t}$ and $Z_t = c_0 + \phi_z Z_{t-1} + a_{2t}$. The mean squared errors (MSE) are calculated and shown in Table 4 under "NPTF". For the purpose of comparison we produce the forecasts using the aforementioned NAARX model (7) and the TAR model (6) under the same setting and report the MSE in Table 4.

Table 4: MSE of the post-sample forecasts

Lead Time	NPTF	NAARX	TAR
1	68.38	70.06	72.84
2	164.38	169.49	183.35
3	229.10	242.17	262.32
4	291.03	303.28	336.04
5	337.66	349.77	390.82
6	366.83	382.34	421.13
7	386.43	406.94	443.02
8	391.70	418.05	457.87
9	400.98	431.81	476.08
10	408.22	439.77	490.63
11	419.27	453.05	506.71
12	440.58	477.45	536.99
In-sample	31.65	33.15	31.74

The results in Table 4 show that the forecast MSE of the nonparametric transfer function (NPTF) model are consistently smaller than those of the NARRX model and the TAR model. In this example, the proposed NPTF model performances well in both within-sample and post-sample, this shows the good potential of the NPTF model in analyzing nonlinear time series data.

6 Summary and Discussion

In this paper we introduce the regression spline-based nonparametric transfer function model. This method is flexible and ideal for modeling highly nonlinear relationships between time series. Efficient estimation of the transfer function model is achieved by incorporating the correlation structure of the noise in nonparametric smoothing. Compared with the local polynomial-based methods, the explicit functional form of polynomial splines makes estimation much less intensive computationally. The finite sample properties of the estimators are studied through simulation. The proposed model is used to model river flow based on temperature and precipitation and found successful when compared with widely-accepted nonlinear parametric models.

Acknowledgement

The author's research is partially supported by NSF grant DMS-0707082 and COBA summer research grant of Georgia Southern University.

References

Aydin, D. and Omay R.E., 2007, The Smoothing Parameter Selection Problem in Smoothing Spline Regression for Different Data Sets. *WSEAS Transactions on Mathematics* **3** (6), 477–483.

Auestad, B. and D. Tjöstheim, 1990, Identification of nonlinear time series: First order characterization and order estimation. *Biometrika* **77**, 669-687.

Box, G.E.P. and G.M. Jenkins, 1976, *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1st ed.

Cai, Z., J. Fan and Q. Yao, 2000, Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* **95**, 941–956.

Carroll, R.J., J. Fan, I. Gijbels and M.P. Wand, 1997, Generalized partially linear single-index models. *Journal of the American Statistical Association* **92**, 477–489.

Chen, R. and R.S. Tsay, 1996, Nonlinear transfer functions. *Journal of Nonparametric Statistics* **66**, 193-204.

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

Chen, R. and R.S. Tsay, 1993b, Nonlinear additive ARX models, *Journal of the American Statistical Association* **88**, 955-967.

de Boor, C., 2001. A Practical Guide to Splines. Springer, New York.

Fan, J. and I. Gilbels, 1996, *Local Polynomial Modeling and Its Applications*. Chapman and Hall, Suffolk.

Fan, J. and Q. Yao, 2003, *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Fan, J., Q. Yao and Z. Cai, 2003, Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society, Series B* **65**, 57–80.

Gáti, J. and Bede, B. 2006. Variation Diminishing Spline Approximation Of Fuzzy Functions. *WSEAS Transactions on Mathematics* **1**(5), 48–54.

Hädle, W., H. Lütkepohl, and R. Chen, 1997, A review of nonparametric time series analysis. *International Statistical Review* **65**, 49-72.

Härdle, W., P. Hall, and H. Ichimura, 1993, Optimal smoothing in single-index models. *The Annals of Statistics* **21**, 157–178.

Härdle, W., H. Liang, and J. Gao, 2000, *Partially Linear Models.* Physica-Verlag, Heidelberg.

Haggan V. and T. Ozaki, 1981, Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189196.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1998, Characterizing selection bias using experimental data. *Econometrica* **66**, 1017–1098.

Huang, J.Z., 2003, Local Asymptotics For Polynomial Spline Regression. *Ann. Statist.*, **31**, 1600-1635

J.Z. Huang and H. Shen (2004), Functional coefficient regression models for nonlinear time series models: a polynomial spline approach. *Scandinavian Journal of Statistics*, **31**, 515-534

Ichiruma, H., 1993, Semiparametric least-squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.

Liu, J.M., R. Chen and Q. Yao, 2008, Nonparametric Transfer Function Models. *Technical report, Georgia Southern University*.

Masry, E., 1996a, Multivariate local polynomial regression for time series: Uniform consistency and rates. *Journal of Time Series Analysis* **17**, 571-599.

Masry, E., 1996b, Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and Their Applications* **65**, 81-101.

Newey, W.K., and T.M. Stoker, 1993, Efficiency of weighted average derivative estimators and index models. *Econometrica* **61**, 1199-1223.

Ozaki, T., 1985, Statistical identification of storage models with application to stochastic hydrology. *Water Resources Bulletin* **21**, 663-675.

Robinson, P.M., 1983, Nonparametric estimators for time series. *Journal of Time Series Analysis* **4**, 185-207.

Phokharatkul, P., Kamnuanchai, S., Kimpan, C., and Phaiboon, S. 2006. Comparative Study of B-Spline and Cardinal Spline with Genetic Algorithm for Invariant Shape Object Recognition*WSEAS Transactions on Computers* bf 3 (5), 536–544.

Ruckstuhl, A., A.H. Welsh, and R.J. Carroll, 2000, Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica* **10**, 51-71.

Schumaker, L.L., 1981, Spline Functions: Basic Theory. Wiley, New York.

Smith, M., C.M. Wong, and R. Kohn, 1998, Additive nonparametric regression with autocorrelated errors. *Journal of the Royal Statistical Society* **60**, 311-331.

Su, L. and A. Ullah, 2006, More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econometric Theory* **22**, 98-126.

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

Tong, H., 1990, *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, New York.

Tong, H., B. Thanoon, and G. Gudmundsson, 1985, Threshold time series modeling of two Icelandic riverflow systems. *Water Resources Bulletin* **21**, 651–661.

Xiao, Z., O.B. Linton, R.J. Carroll, and E. Mammen, 2003, Model efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of the American Statistical Association* **98**, 980-992.

Wang J. and L. Yang 2009. Polynomial spline confidence bands for regression curves. *Statistica Sinica* **19** (1), 325–342.

Wang, L. and L. Yang 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* **35** (6), 2474–2503.