Modeling Conditional Variance Functions Using Nonparametric Transfer Function Models

Jun M. Liu*

Abstract
Estimating conditional variance functions is of great importance in practice. A nonparametric method is proposed to estimate conditional variance functions with correlated noise. In this method, polynomial splines are used to approximate the transfer function and the conditional variance function, while the noise is assumed to follow an Autoregressive-Moving Average (ARMA) process. It is shown via simulations that the estimators proposed in this paper possess the "oracle" property, i.e., any one of the three components in this model (the ARMA parameters, the conditional variance function, and the transfer function) can be estimated as if the other two components are known. Additionally, it is shown that for time series data, it is necessary to model the serial correlation in the noise to achieve optimal efficiency in the nonparametric estimation of both the transfer function and the conditional variance function. By using polynomial splines, this method is not only flexible but also computationally efficient compared with other nonparametric smoothing methods. The asymptotic properties of the estimators are discussed. The usefulness of this model is illustrated through a real data example.

Key Words: Regression, nonparametric/semiparametric smoothing, Time series analysis, Financial statistics

1. Introduction

Time-varying conditional variance is an important feature of time series data. In many applications, the assumption that the conditional variance is time-invariant does not hold. In such cases, the time-dependence of the conditional variance function must be taken into consideration in order to obtain correct inferences. The conditional variance function itself sometimes is of interest, for example, in finance, volatility plays a curial role in asset pricing. The modeling of conditional variance has gained much attention in the past a few decades, and some parametric models have been developed, for examples, the ARCH model (Engle 1982) , the GARCH model (Bollerslev 1986)Tsay (1987) , the CIR model by Cox, Ingersoll and Ross (1985) , the CKLS model by Chan, Karolyi, Longstaff and Sanders (1992) , Cao and Tsay (1992) Harvey, Ruiz and Shephard (1994) , Alizadeh, Brandt and Diebold (2002) , Anderson, Bollerslev, Diebold and Labys (2001a & b) , and Bai, Russell, and Tiao (2003) . Parametric models require strong assumptions about the functional forms of the unknown functions to be estimated, however these assumptions are sometimes difficult to justify, especially when the underlying functions are highly nonlinear. On the other hand, nonparametric smoothing methods take a data-driven approach to explore the appropriate functional forms of the underlying functions, therefore require very little assumption regarding the functional forms. Because of their data-driven nature, nonparametric smoothing methods are very flexible and are suitable to explore complex, highly nonlinear relationships. Nonparametric smoothing methods are typically computationally intensive, but the recent advances in computer technology made it more realistic to use them to model conditional variance functions. Some of the representative work in this area include Conley, Hansen, Luttmer, and Scheinkman (1997) , Fan and Yao (1998) Dahl and

*Georgia Southern University, PO BOx 8136, Statesboro, GA 30460. Tel: (912)478-5085, Email: jliu@georgiasouthern.edu

Levine (2006) , Yuan and Wahba (2004) . In this paper, we assume that the conditional variance function and the transfer function are unknown, smooth functions of the independent variable(s). Polynomial splines are used to approximate both the transfer function and the conditional variance function. The main motivation to use polynomial splines is they are computationally more efficient than other nonparametric smoothing methods such as the local polynomial, while retaining the typical flexibility of nonparametric smoothing methods. The paper is organized as follows: the model and the estimation procedures are described in Section 2. Extensive simulation scenarios are used to study the properties of the proposed estimators, some of the simulation results are presented in Section 3. The observations made in the simulations lead to a discussion of the properties of the proposed estimators in Section 4. The model is applied on a real-life example and the results are summarized in Section 5. Section 6 concludes the paper with a brief summary and discussions.

2. The Model and Estimation Procedures

The problem considered in this paper is represented by the following model

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

where $f(\cdot)$ and $\sigma(\cdot)$ are unknown, smooth functions. $\{X\}$ and $\{e_t\}$ are jointly stationary. Our main objective is to estimate the conditional variance function $\sigma^2(\cdot)$. The innovation process $\{e_t\}$ is assumed to follow a stationary *Autoregressive-Moving Average* process of orders p and q (henceforth the ARMA(p, q) process, see Box and Jenkins (1976) , i.e.,

$$e_t - \sum_{i=1}^{p} \phi_i e_{t-i} = \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$

where E$(\varepsilon_t|X_t = x) = 0$, Var$(\varepsilon_t|X_t = x) = 1$. Let $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$, and $\theta(L) = 1 - \sum_{j=1}^{q} \theta_j L^j$, where L is the lag operator defined as $L^i Y_t = Y_{t-i}$, model (1) can be re-parameterized as

$$\frac{\phi(L)}{\theta(L)}\Big[Y_t - f(X_t)\Big] = \sigma(X_t)\varepsilon_t.$$

$f(\cdot)$, $\sigma(\cdot)$, and the ARMA parameters can be estimated jointly by minimizing the objective function below

$$\sum_{t=1}^{n} \Big\{ \frac{\phi(L)}{\theta(L)} \Big[\frac{Y_t - f(X_t)}{\sigma(X_t)} \Big] \Big\}^2. \tag{2}$$

Let $r(x)$ denote the "pre-whitened" partial residuals

$$r(x) = \frac{\phi(L)}{\theta(L)}[Y - f(x)],$$

a residual-based estimator for $\sigma^2(x)$ can be derived based on the relationship

$$E[r^2(X)|X = x] = \sigma^2(x).$$

A common issue with many variance estimator is that sometimes the estimates can be negative. To ensure the positivity of the estimate of $\sigma^2(\cdot)$, we adopt an idea similar to Yuan and Wahba (2004) and let $\sigma^2(x) = \exp[g(x)]$.

In this paper, we use polynomial splines to approximate $f(x)$ and $g(x)$. 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*). Precisely, a polynomial spline of degree $d \geq 0$ 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 d 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 $\mathcal{X}$. Given knot sequence $\boldsymbol{\lambda}$ and degree d, the collection of spline functions forms a function space spanned by a set of basis functions. One of the commonly used basis functions is the *truncated power basis*, which is the set of functions $\{1, x, \cdots, x^d, (x-\lambda_1)_+^d, \cdots, (x-\lambda_k)_+^d\}$, where $(x)_+^d \equiv (x_+)^d$, $x_+ = x$ if $x \geq 0$ and $x_+ = 0$ if $x < 0$. The dimension of the spline function space is $K = d + k + 1$. When applied on real-life problems, regression models using the truncated power basis functions as regressors are easy to interpret because the coefficients usually bear practical meanings, however, such models can suffer from multicollinearity. The computer solutions of such models can be numerically unstable, especially when the degree of polynomial is high. Another popular basis function is the *B-spline basis*. It has nice theoretical properties (see for examples, Deboor 2001, Schumaker (1981)), so it is often used to derive the asymptotic properties of estimators. The B-spline basis also tends to be more stable numerically when implemented in computer programs. Obviously, the numeric results do not depend on the choice of the basis functions. Polynomial splines allow the underlying function to have different polynomial forms in different regions of the support, so they are very flexible and are suitable to explore complex, highly nonlinear relationships without explicit assumptions about the functional forms a priori. Polynomial splines are also highly efficient computationally, because once the knots are determined, the estimation can be carried out as one standard least squares regression. In contrast, the local polynomial smoothing method requires least squares procedures to be performed at many focal points, consequently takes much longer to calculate the solution.

Denote a set of basis functions as $\{B_j(\cdot)\}_{j=1}^K$, approximate $f(\cdot)$ and $g(\cdot)$ with polynomial splines, specifically,

$$f(x) \approx \sum_{i=1}^{K_1} \alpha_i B_{1i}(x), \quad g(x) \approx \sum_{i=1}^{K_2} \beta_i B_{2i}(x),$$

where $\{B_{1i}(\cdot)\}_{i=1}^{K_1}$ and $\{B_{2i}(\cdot)\}_{i=1}^{K_2}$ are the B-spline basis of the estimation spaces for $f(\cdot)$ and $g(\cdot)$, respectively, K_1 and K_2 are the corresponding dimensions of the spline spaces, $K_j = k_j + d_j + 1$ ($j = 1$, 2), where k_j is the number of interior knots, and d_j is the degree of the polynomial. With the above approximations, (2) can be rewritten as

$$\sum_{t=1}^{n} \frac{\{[\phi(L)\theta(L)^{-1}][Y_t - \sum_{i=1}^{K_1} \alpha_i B_{1i}(X_t)]\}^2}{\exp[\sum_{i=1}^{K_2} \beta_i B_{2i}(X_t)]}. \tag{3}$$

To solve this problem, we consider an iterative estimation procedure which is convenient for the discussion of the properties of the estimators. Let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{K_1})^\tau$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{K_2})^\tau$, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)^\tau$, and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)^\tau$. A brief description of the estimation procedure is given below.

1. Obtain $\check{\alpha}_i$ by minimizing

$$\sum_{t=1}^{n} \left\{ Y_t - \sum_{i=1}^{K_1} \alpha_i B_{1i}(X_t) \right\}^2,$$

the preliminary estimate is $\check{f}(x) = \sum_{i=1}^{K_1} \check{\alpha}_i B_{1i}(x)$.

2. Obtain preliminary estimate $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\theta}}$ by minimizing

$$\sum_{t=1}^{n} \left\{ \phi(L)\theta(L)^{-1}[Y_t - \check{f}(X_t)] \right\}^2 ,$$

let $r(X_t) = \widehat{\phi}(L)\widehat{\theta}(L)^{-1}[Y_t - \check{f}(X_t)]$.

3. Obtain $\widehat{\beta}_i$ by minimizing

$$\sum_{t=1}^{n} \left[\log(r^2(X_t)) - \sum_{i=1}^{K_2} \beta_i B_{2i}(X_t) \right]^2 ,$$

$\widehat{\sigma}(x) = \sqrt{\exp[\sum_{i=1}^{K_2} \widehat{\beta}_i B_{2i}(x)]}$.

4. Obtain $\widehat{\alpha}_i$ by minimizing

$$\sum_{t=1}^{n} \left\{ \frac{\widehat{\phi}(L)}{\widehat{\theta}(L)} \left[\frac{Y_t - \sum_{i=1}^{K_1} \alpha_i B_{1i}(X_t)}{\widehat{\sigma}(X_t)} \right] \right\}^2,$$

$\widehat{f}(x) = \sum_{i=1}^{K_1} \widehat{\alpha}_i B_{1i}(x)$.

5. Obtain $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\theta}}$ by minimizing

$$\sum_{t=1}^{n} \left\{ \frac{\phi(L)}{\theta(L)} \left[\frac{Y_t - \widehat{f}(X_t)}{\widehat{\sigma}(X_t)} \right] \right\}^2.$$

Steps 3-5 may be iterated to improve the finite sample performance.

3. Simulations

Extensive simulations are used to investigate the behavior of the procedure proposed in this paper. Due to the limit of space, only selected cases are reported here. In reference to model (1), data are generated using the following models:

$$\begin{aligned} f(x) &= \sin(4x) + \cos(2x), \\ \sigma(x) &= 0.4\exp(-2x^2) + 0.2, \end{aligned}$$

the dependent variable X_t is generated from the following AR(1) process:

$$X_t = 0.3X_{t-1} + a_t, \quad a_t \sim \mathrm{N}(0, .5^2),$$

the innovation process $\{e_t\}$ is generated from an ARMA(1,1) model

$$e_t - \phi e_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim N(0,1),$$

the parameters of this model are selected combinations of ϕ =-.8, -.5, -.2, .2, .5, .8, and θ = -.8, -.5, -.2, .2, .5, .8, one obvious requirement is $\phi \neq \theta$ so that there is no common factor on the two sides of the equation. The sample sizes used in the simulations are 200, 500, and 1000. Four hundred replications are used in each case.

A key issue in polynomial spline regression is the number and location of knots. The degree of the polynomial, although not as crucial as the knots, should also be selected to improve the performance of the model. In the simulations we adopt an exhaustive search method similar to the idea of Fan and Gilbels . Specifically, for each degree of the polynomial $d_i \in \{0, 1, 2, 3\}$, k_i $(i = 1, 2)$ interior knots are placed on the percentile points so that there are equal number of observations between adjacent knots. k_i is in the range $\left[\lfloor .5k_0 \rfloor, \lceil (\min(5k_0, n/4) \rceil\right]$, where $k_0 = n^{1/5}$ is the theoretically optimal order. The combination of d_i and k_i $(i = 1, 2)$ that minimizes the Bayesian Information Criteria (BIC) is selected. In the simulation it is found that the Akaike Information Criteria (AIC) tends to over-fit the model.

The proposed estimation procedure is used to estimate $f(x)$, $\sigma(x)$ and the ARMA parameters. The results when $\{e_t\}$ follows a pure AR(1), a pure MA(1) and the more general ARMA(1,1) process are summarized in Tables 1, 2, and 3 respectively. The criterion used to evaluate the performance of the nonparametric estimators $\widehat{\sigma}(\cdot)$ and $\widehat{f}(\cdot)$ is the mean absolute error (MAE), for example, for $\widehat{\sigma}(x)$,

$$\text{MAE}_{\widehat{\sigma}} = \frac{1}{n} \sum_{t=1}^{n} |\sigma(X_t) - \widehat{\sigma}(X_t)|,$$

$\text{MAE}_{\widehat{f}}$ is defined the same way. To put the performance of the proposed estimators in perspective, the mean absolute error (MAE) of the proposed estimators $\widehat{f}$ and $\widehat{\sigma}$ are compared with those of the corresponding "idealized" estimators $\widetilde{f}$ and $\widetilde{\sigma}$. $\widetilde{\sigma}(x)$ is the same polynomial splines estimator as $\widehat{\sigma}(x)$, except calculated with known $f(x)$, ϕ and θ; similarly, $\widetilde{f}(x)$ is the same estimator as $\widehat{f}(x)$, except calculated with known $\sigma(x)$, ϕ and θ. The sample mean and sample standard deviation of the proposed estimator $\widehat{\phi}$ and $\widehat{\theta}$ are also compared with their "idealized" counterparts $\widetilde{\phi}$ and $\widetilde{\theta}$, which are calculated with known $f(x)$ and known $\sigma(x)$. Needless to say, with all other components known, $\widetilde{f}(x)$ and $\widetilde{\sigma}(x)$ are univariate polynomial spline estimators, $\widetilde{\phi}$ and $\widetilde{\theta}$ are standard ARMA estimators with observed $\{e_t\}$.

In Tables 1, 2, and 3 below, the MAE of the proposed estimators $\widehat{f}(\cdot)$ and $\widehat{\sigma}(\cdot)$ are averaged over the 400 replications and used as the baseline. The average MAE of the "idealized" estimators $\widetilde{f}(\cdot)$ and $\widetilde{\sigma}(\cdot)$ is divided by the average MAE of the corresponding proposed estimator ($\widehat{f}$ and $\widehat{\sigma}$) as a measure of relative efficiency. The relative efficiency measures for $\widehat{f}(\cdot)$ and $\widehat{\sigma}(\cdot)$ are denoted as Eff_f and Eff_σ, respectively, in Tables 1-3. The means and standard deviations of $\widehat{\phi}$, $\widehat{\theta}$, as well as those of the "idealized" $\widetilde{\phi}$ and $\widetilde{\theta}$ over the 400 replications, are included in the tables under "mean" and "std".

The serial correlation is an important feature in time series data and must be taken into consideration in nonparametric smoothing. In nonparametric smoothing, if the serial correlation in the data is ignored, the nonparametric estimator will not be efficient. On the other hand, if the serial correlation is modeled appropriately, nonparametric estimation can achieve the optimal efficiency as if the data is iid (Liu, Chen, and Yao 2010). The serial correlation contains useful information that can improve forecasting performance. The explicit ARMA structure used in this paper for the noise offers a parsimonious way to utilize this information In order to show the effect of removing serial correlation in spline estimation, we consider a "naive" version of estimators which are the same polynomial splines estimators as $\widehat{f}(x)$ and $\widehat{\sigma}(x)$ except the serial correlation in the noise is ignored, i.e., assuming e_t is iid. The relative efficiency of the "naive" estimators, measured as the ratio of their average MAE to those of the corresponding proposed estimators are included in Tables 1, 2, and 3 under "Effn_f" and "Effn_σ".

From Tables 1, 2 and 3, we can see while the average MAE of the "idealized" estimators tend to be smaller than those of the proposed estimators, which is expected, but overall the difference is quite small. As expected, on average the proposed ARMA estimators have larger bias and standard deviation than the "idealized" estimators, but the difference is

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

ϕ	n	MAE$_{\widehat{f}}$	Eff$_{\widetilde{f}}$	Effn$_f$	MAE$_{\widehat{\sigma}}$	Eff$_{\widetilde{\sigma}}$	Effn$_\sigma$	mean($\widehat{\phi}$)	std($\widehat{\phi}$)	mean($\widetilde{\phi}$)	std($\widetilde{\phi}$)
	200	0.0783	1.1561	1.3460	0.0957	0.8396	3.1657	-0.7635	0.0485	-0.7717	0.0477
-0.8	500	0.0588	1.1539	1.2251	0.0726	0.8775	4.2643	-0.7753	0.0292	-0.7823	0.0333
	1000	0.0517	1.1128	1.1719	0.0704	0.8311	4.4788	-0.7818	0.0319	-0.7893	0.0184
	200	0.0729	0.9951	1.0825	0.0592	0.9253	1.2416	-0.4883	0.0662	-0.4894	0.0650
-0.5	500	0.0564	1.0110	1.0602	0.0439	0.9096	1.7296	-0.4871	0.0446	-0.4923	0.0419
	1000	0.0453	1.0449	1.0555	0.0320	0.9323	2.2723	-0.4921	0.0279	-0.4963	0.0288
	200	0.0777	0.9492	0.9817	0.0530	0.9679	0.9502	-0.1995	0.0749	-0.2007	0.0736
-0.2	500	0.0569	0.9607	0.9850	0.0386	0.9462	1.1401	-0.1940	0.0465	-0.1983	0.0468
	1000	0.0459	0.9764	0.9874	0.0279	0.9538	1.3964	-0.1963	0.0316	-0.1995	0.0309
	200	0.0816	0.9722	0.9821	0.0559	0.9301	0.9408	0.1847	0.0754	0.1890	0.0683
0.2	500	0.0597	0.9568	0.9880	0.0445	0.8310	0.9938	0.1890	0.0465	0.1942	0.0446
	1000	0.0470	0.9675	0.9877	0.0265	0.9740	1.4255	0.1944	0.0313	0.1973	0.0308
	200	0.0939	0.9976	1.0548	0.0597	0.9551	1.2794	0.4829	0.0619	0.4882	0.0607
0.5	500	0.0679	1.0029	1.0296	0.0450	0.8745	1.6084	0.4817	0.0409	0.4878	0.0399
	1000	0.0547	0.9991	1.0308	0.0310	0.9715	2.2892	0.4894	0.0287	0.4937	0.0286
	200	0.1621	1.0263	1.0876	0.0887	0.8873	3.1496	0.7480	0.0683	0.7530	0.0506
0.8	500	0.1111	0.9934	1.0608	0.0762	0.8672	4.0292	0.7734	0.0329	0.7792	0.0283
	1000	0.0841	0.9841	1.0403	0.0672	0.8660	4.5694	0.7791	0.0199	0.7852	0.0177

Table 2: Simulation results when $\{e_t\}$ follows MA(1) processes

θ	n	MAE$_{\widehat{f}}$	Eff$_{\widetilde{f}}$	Effn$_f$	MAE$_{\widehat{\sigma}}$	Eff$_{\widetilde{\sigma}}$	Effn$_\sigma$	mean($\widehat{\theta}$)	std($\widehat{\theta}$)	mean($\widetilde{\theta}$)	std($\widetilde{\theta}$)
	200	0.0683	1.0773	1.2046	0.0781	0.9279	1.6327	-0.7264	0.0748	-0.7402	0.0731
-0.8	500	0.0526	1.0869	1.1820	0.0649	0.9668	2.0019	-0.7306	0.0464	-0.7589	0.0478
	1000	0.0444	1.1037	1.1569	0.0580	0.9538	2.2703	-0.7434	0.0299	-0.7699	0.0333
	200	0.0708	1.0106	1.0849	0.0570	0.9894	1.0859	-0.4965	0.0725	-0.4934	0.0724
-0.5	500	0.0550	1.0084	1.0630	0.0424	0.9034	1.4620	-0.4844	0.0434	-0.4923	0.0433
	1000	0.0456	1.0141	1.0615	0.0328	0.8948	1.7264	-0.4899	0.0311	-0.4981	0.0312
	200	0.0740	0.9635	0.9867	0.0523	0.9551	0.9499	-0.2035	0.0813	-0.2043	0.0760
-0.2	500	0.0565	0.9591	0.9804	0.0383	0.9504	1.1433	-0.2006	0.0458	-0.2040	0.0428
	1000	0.0451	0.9697	0.9901	0.0261	1.0080	1.4571	-0.1985	0.0316	-0.2016	0.0309
	200	0.0810	0.9492	0.9801	0.0546	0.9560	0.9414	0.1866	0.0765	0.1916	0.0699
0.2	500	0.0595	0.9653	0.9788	0.0383	0.9459	1.1407	0.1947	0.0471	0.1981	0.0447
	1000	0.0477	0.9702	0.9792	0.0268	0.9546	1.4005	0.1951	0.0328	0.1981	0.0323
	200	0.0856	0.9961	1.0558	0.0624	0.8821	1.0004	0.4774	0.0787	0.4860	0.0671
0.5	500	0.0611	1.0043	1.0476	0.0436	0.9142	1.3586	0.4835	0.0445	0.4904	0.0482
	1000	0.0488	1.0104	1.0384	0.0314	0.8978	1.7491	0.4845	0.0292	0.4935	0.0279
	200	0.0875	1.0316	1.1317	0.0847	0.8938	1.4780	0.7183	0.0798	0.7279	0.0813
0.8	500	0.0643	1.0490	1.1270	0.0665	0.9209	1.9400	0.7264	0.0490	0.7496	0.0578
	1000	0.0521	1.0441	1.1047	0.0605	0.9193	2.1841	0.7429	0.0299	0.7670	0.0304

quite small and diminishing with the increase of the sample size. Similar observations are made in other simulation scenarios not reported here. As a visual example, the MAEs of the proposed estimators $\widehat{f}(\cdot)$ and $\widehat{\sigma}(\cdot)$ in the 400 replications are plotted against those of the "idealized" estimators in Figure 1, with a 45-degree line passing though the origin added to show the pattern. This figure shows that the proposed estimators perform similarly as their respective "idealized" counterparts, indicating that using the proposed method, the conditional variance function and the transfer function can be estimated as if other components of the model are known by some "oracle". This graph is made for $n = 500$ and e_t follows an AR(1) process with $\phi = 0.5$, but similar pattern is observed in other cases.

The histogram of $\widehat{\phi}$ in the left panel of Figure 2 shows that the sampling distribution of $\widehat{\phi}$ is close to a normal distribution centered at ϕ, the true AR parameter. The right panel of

Table 3: Simulation results when $\{e_t\}$ follows ARMA(1,1) processes

ϕ	θ	n	MAE$_{\widehat{f}}$	Eff$_{\widetilde{f}}$	Effn$_f$	MAE$_{\widehat{\sigma}}$	Eff$_{\widetilde{\sigma}}$	Effn$_\sigma$	mean$(\widehat{\phi})$	std$(\widehat{\phi})$	mean$(\widehat{\theta})$	std$(\widehat{\theta})$
		200	0.0719	1.1216	1.2910	0.0887	0.9105	2.7090	-0.4846	0.0841	0.4064	0.1239
-0.5	0.5	500	0.0574	1.1105	1.2045	0.0805	0.8935	3.0985	-0.4985	0.0502	0.4154	0.0666
		1000	0.0488	1.1120	1.1689	0.0717	0.9085	3.4562	-0.4976	0.0354	0.4356	0.0448
		200	0.0781	1.1303	1.4052	0.1543	1.2347	2.4103	-0.4861	0.0800	0.5350	0.1289
-0.5	0.8	500	0.0547	1.1938	1.3290	0.1439	1.3194	2.6110	-0.4899	0.0477	0.5920	0.0753
		1000	0.0506	1.1591	1.2387	0.1476	1.2962	2.5591	-0.4948	0.0344	0.6036	0.0547
		200	0.0744	0.9899	1.0358	0.0561	0.9515	0.9940	-0.1612	0.2165	0.2323	0.2259
-0.2	0.2	500	0.0553	0.9956	1.0219	0.0421	0.9299	1.1842	-0.1868	0.1235	0.2053	0.1277
		1000	0.0444	0.9933	1.0239	0.0286	0.9786	1.4971	-0.1963	0.0781	0.1969	0.0785
		200	0.0704	1.0547	1.1450	0.0706	0.8574	1.4977	-0.1869	0.1289	0.4768	0.1314
-0.2	0.5	500	0.0554	1.0383	1.1060	0.0532	0.9024	2.0139	-0.1914	0.0746	0.4702	0.0703
		1000	0.0449	1.0566	1.1045	0.0401	0.9010	2.6664	-0.2003	0.0514	0.4751	0.0503
		200	0.0656	1.1503	1.3275	0.1096	0.8987	1.7720	-0.1824	0.0987	0.6776	0.1287
-0.2	0.8	500	0.0535	1.1060	1.2271	0.0826	1.1054	2.4355	-0.1938	0.0578	0.6916	0.0609
		1000	0.0462	1.1079	1.1810	0.0845	1.0465	2.3820	-0.1995	0.0423	0.7016	0.0420
		200	0.0978	1.0060	1.1518	0.0948	1.0438	2.0067	0.1844	0.0921	-0.6677	0.1107
0.2	-0.8	500	0.0707	1.0395	1.1438	0.0887	1.0293	2.2245	0.1968	0.0575	-0.6782	0.0616
		1000	0.0558	1.0591	1.1311	0.0912	0.9691	2.1914	0.1954	0.0392	-0.7061	0.0406
		200	0.0864	1.0254	1.1017	0.0674	0.9193	1.5075	0.1896	0.1384	-0.4562	0.1435
0.2	-0.5	500	0.0660	1.0137	1.0867	0.0532	0.8412	1.9838	0.1933	0.0788	-0.4715	0.0716
		1000	0.0537	1.0020	1.0706	0.0417	0.8608	2.5587	0.2028	0.0513	-0.4713	0.0483
		200	0.0850	0.9735	1.0159	0.0546	0.9638	0.9741	0.1661	0.2264	-0.2101	0.2234
0.2	-0.2	500	0.0622	0.9720	1.0030	0.0396	0.9079	1.2641	0.2035	0.1194	-0.1847	0.1243
		1000	0.0478	1.0054	1.0257	0.0279	1.0012	1.5335	0.1964	0.0816	-0.1958	0.0834
		200	0.1182	1.0081	1.1225	0.0964	0.8892	2.4511	0.4888	0.0927	-0.3907	0.1238
0.5	-0.5	500	0.0782	1.0325	1.1210	0.0787	0.8971	3.0604	0.4907	0.0471	-0.4211	0.0657
		1000	0.0632	1.0446	1.1174	0.0720	0.9035	3.3999	0.4940	0.0353	-0.4371	0.0454
		200	0.0996	1.0136	1.0875	0.0697	0.9085	1.7586	0.4781	0.1015	-0.1812	0.1223
0.5	-0.2	500	0.0721	1.0149	1.0728	0.0509	0.8990	2.5316	0.4930	0.0670	-0.1752	0.0768
		1000	0.0570	1.0101	1.0668	0.0402	0.9162	3.2859	0.4966	0.0422	-0.1800	0.0526

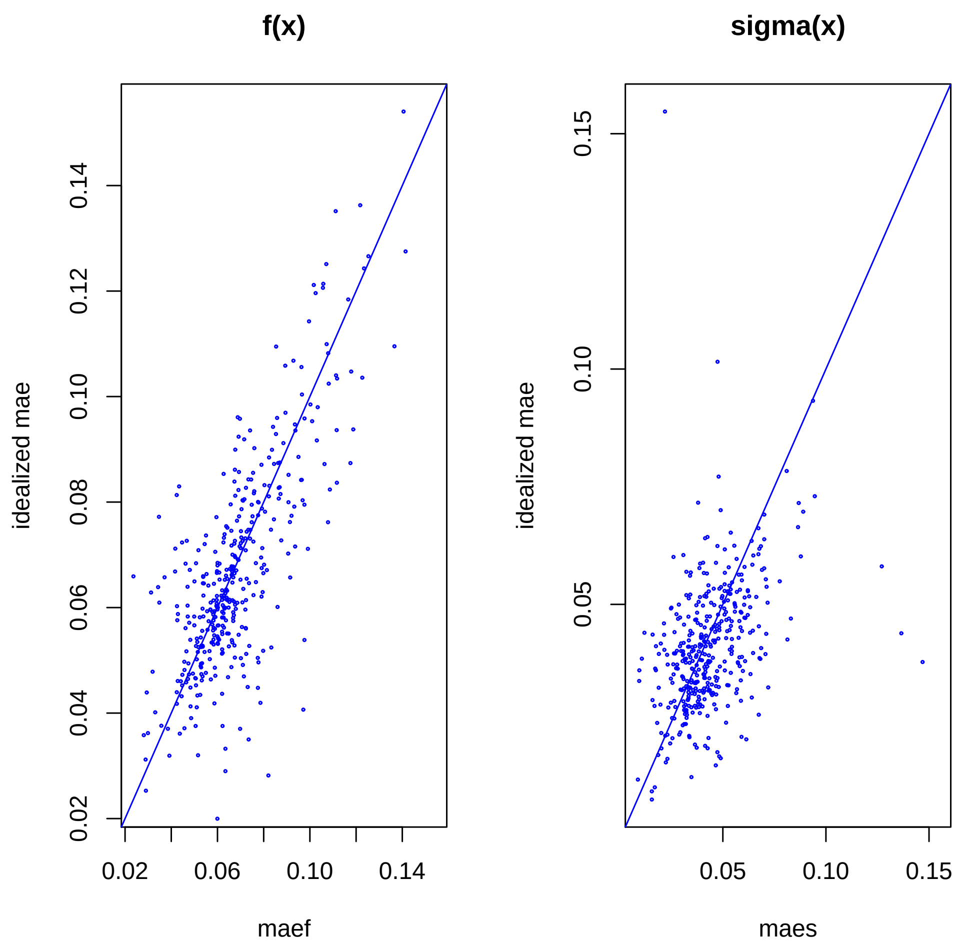

Figure 1: Left: MAE of $\widehat{f}(\cdot)$ vs. $\widetilde{f}(\cdot)$. Right: MAE of $\widehat{\sigma}(\cdot)$ vs. $\widetilde{\sigma}(\cdot)$. $\phi = 0.5$, n=500.

the same figure shows $\widehat{\phi}$ against the "idealized" $\widetilde{\phi}$ in 400 replications, we again see that $\widehat{\phi}$ behaviors similarly to the "idealized " ARMA estimator which is estimated with observed $\{e_t\}$, i.e., with known f and σ. Figure 3 shows $\widehat{f}$ and $\widehat{\sigma}$ in a typical simulation when $n = 500$ and e_t follows an MA(1) process with $\theta = 0.5$. From Tables 1-3, it is obvious that the performance of the estimators improves with the increase of sample size. These observations are similar to those made by Fan and Yao (1998) in which local polynomial smoothers are used to model $f(\cdot)$ and $\sigma(\cdot)$. This leads us to conjecture that the estimators proposed in this paper have the "oracle" properties similar to those observed by Fan and Yao (1998), i.e., the conditional variance function $\sigma^2(\cdot)$ can be estimated as if the other components of the model $f(\cdot)$, ϕ and θ, are known by an "oracle". Similarly, the transfer function $f(\cdot)$ can be estimated as if $\sigma(\cdot)$, ϕ and θ are known, and the ARMA estimators has the same asymptotic normal distribution as a standard univariate ARMA estimator.

We also see that the "naive" estimators have larger average MAE than the proposed estimators, which indicates the information contained in the serial correlation of the noise should be utilized to improve the efficiency in nonparametric estimation. The gain in efficiency in estimating $f(x)$ and $\sigma(x)$ is larger when the serial correlation is strong (e.g., when ϕ is high in absolute value in AR(1) cases).

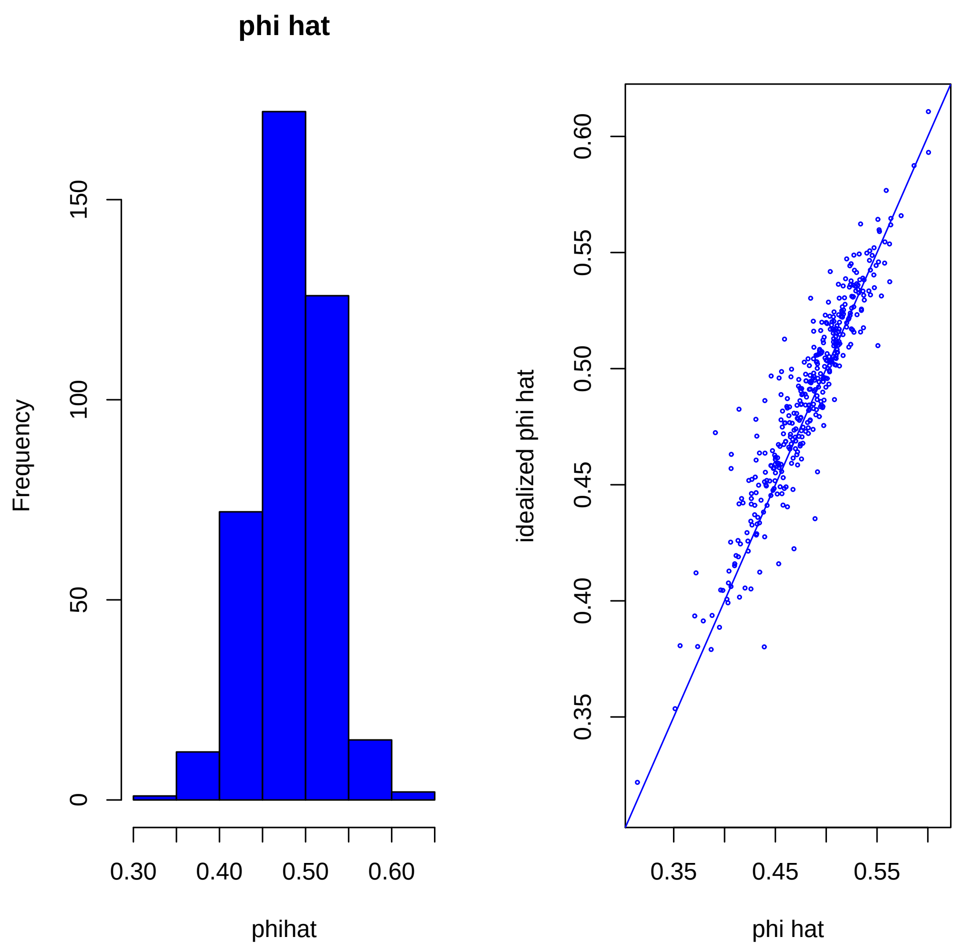

Figure 2: Left: Histogram of $\widetilde{\phi}$, Right: $\widehat{\phi}$ vs $\widetilde{\phi}$. $\phi = 0.5, n = 500$.

4. The Properties of the Estimators

The following assumptions are needed in the discussion of the properties of the proposed estimators. Most of them are standard assumptions in the literature.

(A1) $f(\cdot)$ and $\sigma(\cdot)$ are q-smooth [1].

(A2) The density of X, p_X is absolutely continuous and bounded away from zero and infinity.

(A3) $\sigma(\cdot)$ is bounded away from zero and infinity.

(A4) The knot sequence $\boldsymbol{\kappa} = \{\kappa_0 \leq \kappa_1 < \cdots < \kappa_K \leq \kappa_{K+1})$ has bounded mesh ratio, that is, for $j = 1, \cdots, K$, $\frac{\max(\kappa_{j+1}-\kappa_j)}{\min(\kappa_{j+1}-\kappa_j)} \leq c$. The number of interior knots k_i satisfies $k \sim n^{1/(2p+1)}$.

[1] A function $f(\cdot)$ is $q-$smooth ($q = r+\beta$) if it is r-times continuously differentiable on $\mathcal{X}$ and $f^{(r)}$ satisfies a Hölder condition with exponent β, i.e., there exists a positive integer γ such that $|f^{(r)}(x_1) - f^{(r)}(x_2)| \leq \gamma |x_1 - x_2|^{\beta}$, for $x_1, x_2 \in \mathcal{X},\ 0 < \beta \leq 1$.

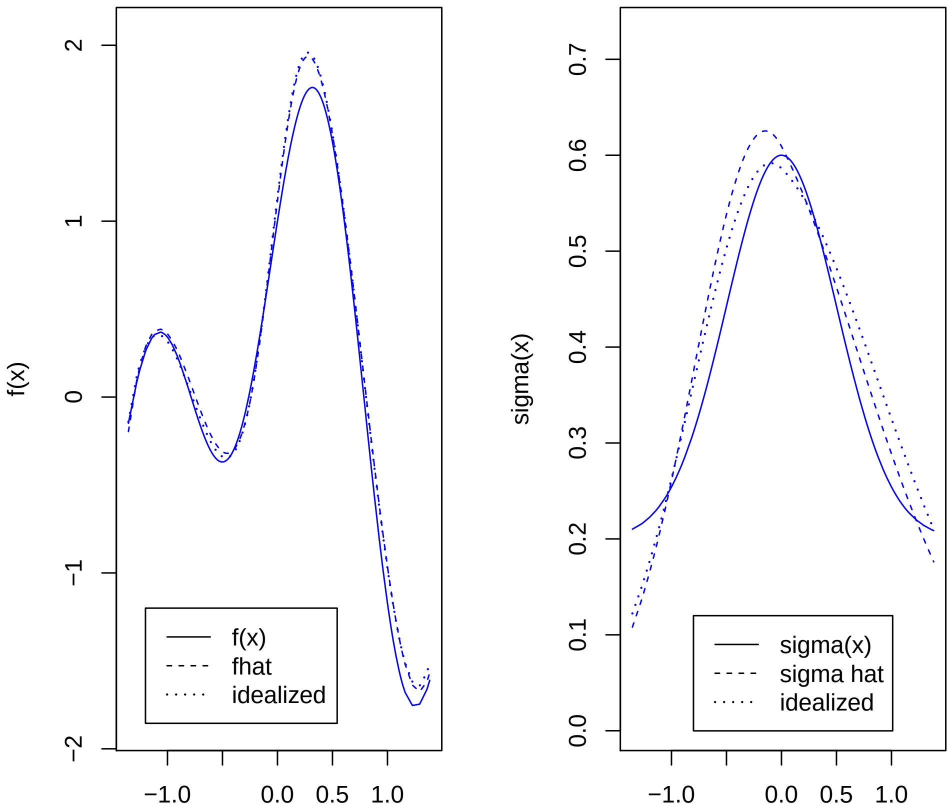

Figure 3: $\widehat{f}(\cdot)$ and $\widehat{\sigma}(\cdot)$ in a typical simulation, $\theta = 0.5$, n=500.

(A5) The process $\{Y_t, X_t\}$ is jointly α-mixing in the sense that the $\alpha-$mixing coefficient $\alpha(k) \leq c\rho^k$ for constants $c > 0$, $0 < \rho < 1$, where

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

where $\mathcal{F}_i^j$ is the σ-algebra generated by $\{X_i, \cdots, X_j\}$ for $i \leq j$.

Define $||f||_\infty \equiv \sup_{x \in \mathcal{X}} f(x)$. Let $\mathcal{S}_n$ be the space of polynomial splines, which is the estimation space. Usually f is not in $\mathcal{S}_n$ and we use functions in $\mathcal{S}_n$ to approximation unknown f. In order to obtain good approximation we allow the dimension of $\mathcal{S}_n$ to grow with the sample size n. For notational convenience we suppress the dependence of $\mathcal{S}_n$ on n in the subsequent discussion. For any integrable function f on $\mathcal{X}$, define $\mathrm{E}_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$, the empirical inner product is defined by $\langle f_1, f_2 \rangle_n = \mathrm{E}_n(f_1 f_2)$, and the empirical norm is defined as $||f||_n^2 = \langle f, f \rangle_n$. The theoretical inner product and norm are defined as $\langle f_1, f_2 \rangle = \mathrm{E}(f_1 f_2)$, and $||f||^2 = \langle f, f \rangle$, respectively. The least squares estimate $\widehat{f}$ is an orthogonal projection of the observations Y on the estimation space $\mathcal{S}$, with respect to the empirical inner product. Denote such an orthogonal projection operator by P_n, we have $\widehat{f} = P_n Y = P_n f + P_n E$, where $E = (e_1, e_2, \cdots, e_n)^\tau$. From the definitions of expectation and orthogonal projection, it can be easily shown that $\bar{f} \equiv \mathrm{E}(\widehat{f}|X_1, \cdots, X_n) =$

$P_n f$, therefore we have the following decomposition

$$\widehat{f}(x) - f(x) = \widehat{f}(x) - \bar{f}(x) + \bar{f}(x) - f(x),$$

in subsequent discussions $\widehat{f}(x) - \bar{f}(x)$ is referred to as the variance term and $\bar{f}(x) - f(x)$ the bias term. To show the dependence on sample size n, we use K_n to denote the dimension of the polynomial spline. In the following we will discuss the asymptotic normality of the variance term and the order of the bias term. Let $\sigma_e^2 = \mathrm{Var}(e_t)$, under our stationary ARMA assumption, $\sigma_e^2 < \infty$.

Let $\boldsymbol{\gamma} = (\boldsymbol{\phi}^\tau, \boldsymbol{\theta}^\tau)^\tau$, let $\widehat{\boldsymbol{\gamma}}$ be the proposed ARMA estimator and $\widetilde{\boldsymbol{\gamma}}$ the "idealized" ARMA estimator when e_t is observable. Assume $\boldsymbol{\phi}$ meets the stationarity condition and $\boldsymbol{\theta}$ meets the invertibility condition. Then as $n \to \infty$,

$$\sqrt{n}\big(\widehat{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\gamma}}\big) = o_p(1).$$

Hence $\widehat{\boldsymbol{\gamma}}$ shares the same asymptotic distribution of $\widetilde{\boldsymbol{\gamma}}$,

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{D} N\Big(0,\ \sigma^2 \mathbf{V}(\boldsymbol{\gamma})^{-1}\Big), \text{where}$$

$$\mathbf{V}(\boldsymbol{\gamma}) = \mathrm{E}\begin{bmatrix} \mathbf{U}_1\mathbf{U}_1^\tau & \mathbf{U}_1\mathbf{V}_1^\tau \\ \mathbf{V}_1\mathbf{U}_1^\tau & \mathbf{V}_1\mathbf{V}_1^\tau \end{bmatrix},$$

$\mathbf{U}_t = (U_t, U_{t-1}, \cdots, U_{t+1-p})^\tau$, $\mathbf{V}_t = (V_t, V_{t-1}, \cdots, V_{t+1-q})^\tau$. $\phi(L)U_t = \varepsilon_{1t}$ and $\theta(L)V_t = \varepsilon_{2t}$ (Brock and Davis (1987)).

The above result shows that $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = O_p(\sqrt{n})$, faster than the nonparametric rate in $\widehat{f}(\cdot)$. Based on this we can treat $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$ in the derivation. Replace $\widehat{\boldsymbol{\gamma}}$ with $\boldsymbol{\gamma}$, Y_t is pre-whitened and used in the next step of the estimation:

$$\frac{\phi(L)}{\theta(L)} Y_t = \frac{\phi(L)}{\theta(L)} f(X_t) + \varepsilon_t,$$

with preliminary estimate $\check{f}(\cdot)$, use the π weights $\phi(L)\theta(L)^{-1} = 1 + \sum_{i=1}^{\infty} \pi_i L^i$, it is easy to see that the response variable in the spline estimation next stage is

$$\begin{aligned} \widetilde{Y}_t &= \frac{\phi(L)}{\theta(L)} Y_t - \sum_{i=1}^{\infty} \pi_i \check{f}(X_{t-i}) \\ &= f(X_t) + \varepsilon_t + \sum_{i=1}^{\infty} \pi_i \Big[f(X_{t-i}) - \check{f}(X_{t-i}) \Big], \end{aligned}$$

and $\widehat{f}(\cdot)$ is the orthogonal projection of $\widetilde{Y}$ onto polynomial splines space S_n. $\widetilde{Y}_t$ is made up of the "pre-whitened" observation $f(X_t) + \varepsilon_t$, and the error resulting from the preliminary estimation $\sum_{i=1}^{\infty} \pi_i \Big[f(X_{t-i}) - \check{f}(X_{t-i}) \Big]$. From the stationarity assumption, we can show the error term is negligible in the estimation, and $\widehat{f}(\cdot)$ has the same limiting distribution as if $\{e_t\}$ is iid. The results below are essentially the same results in Huang (2003) for cross-sectional data, our simulation results indicates that the proposed estimators $\widehat{f}(\cdot)$ and $\sigma(\cdot)$ behave similarly to the corresponding "idealized" estimators, which are univariate polynomial splines estimators applied on time series data with serial correlation removed, therefore these results continue to hold in the estimation of $f(\cdot)$ and $\sigma(\cdot)$. Let $\Phi(\cdot)$ be the distribution function of the standard normal distribution. The following establishes the asymptotic normality of the variance term $\widehat{f} - \bar{f}$, where $\bar{f} = P_n f$ as before.

Conjecture 1 *Assume* $\lim_{\lambda\to\infty} E\left(\varepsilon^2 I_{\{|\varepsilon|>\lambda\}}|X=x\right)=0$, *then for* $x\in\mathcal{X}$ *and* $t\in\mathcal{R}$,

$$P\Big(\widehat{f}(x)-\bar{f}(x)\le t\sqrt{\mathit{Var}[\widehat{f}(x)|X_1,\cdots,X_n]}\Big)-\Phi(t)=o_p(1),$$

consequently

$$\frac{\widehat{f}(x)-\bar{f}(x)}{\sqrt{\mathit{Var}(\widehat{f}(x)|X_1,\cdots,X_n)}}\xrightarrow{D}N(0,1).$$

From the above result, we can derive that

$$\sup_{x\in\mathcal{X}}|\widehat{f}(x)-\bar{f}(x)|=O_p(\sqrt{K_n/n}). \tag{4}$$

The following result, which is similar to Theorem 5.2 of Huang (2003) , shows that when f is q-smooth, the bias term of $\widetilde{f}(x)$ is negligible comparing with the variance term.

Conjecture 2 *Assume that the marginal density function* p_X *is bounded away from zero and infinity. If* $\lim_{n\to\infty} K_n/n^{2q+1}=\infty$ *and* $\lim_{n\to\infty} K_n\log n/n=0$, *then*

$$\sup_{x\in\mathcal{X}}\Big|\frac{\bar{f}(x)-f(x)}{\sqrt{Var(\widehat{f}(x)|X_1,\cdots,X_n)}}\Big|=o_p(1)$$

It is shown in the approximation theory (e.g., Huang 2003) that the bias term $|\bar{f}(x)-f(x)|$ caused by using regression spline to approximate f is controlled by $\inf_{f^*\in\mathcal{S}}||f-f^*||_\infty$, the best rate in L_∞ norm for approximating f with a regression spline function. With a B-spline basis, using results in approximation theory we can show the following result

$$\sup_{x\in\mathcal{X}}|\bar{f}(x)-f(x)|=O_p(K_n^{-q}). \tag{5}$$

Using (4) and (5), balancing the variance and the squared bias, we can see that if $K_n=O_p(n^{\frac{1}{2q+1}})$ the optimal estimation rate $O_p(n^{-\frac{2q}{2q+1}})$ (Stone (1982)) can be obtained. The above results show that after removing the correlation in the noise, f can be estimated with the same rate as if e_t is iid.

For the conditional variance function estimator $\widehat{\sigma}(\cdot)$, perform the same decomposition of $\widehat{\sigma}(\cdot)-\sigma(\cdot)=\widehat{\sigma}(\cdot)-\bar{\sigma}(\cdot)+\bar{\sigma}(\cdot)-\sigma(\cdot)$, similar results as above should hold for $\widehat{\sigma}$, i.e., the variance term $\widehat{\sigma}(\cdot)-\bar{\sigma}(\cdot)$ is asymptotically normally distributed, and the bias term $\bar{\sigma}(\cdot)-\sigma(\cdot)$ is of smaller order than the variance term therefore is negligible.

5. An empirical example

To illustrate the proposed approach, we consider the weekly yields of three-month Treasury Bills from January 18, 1954 to December 31, 1999. This data set was previously analyzed by Anderson and Lund (1997) , Gallant and Tauchen (1997) , and Fan and Yao (1998) . A time series plot is given in Figure 4.

To build the model, the Dicky-Fuller ADF test is first applied to test for nonstationarity of Y_t. The test statistic is -2.7538, with a p-value of 0.2592, therefore the null hypothesis of unit root is not rejected. As a result, first order difference is taken to eliminate the unit root. Let $Z_t=Y_t-Y_{t-1}$, we consider the following model

$$Z_t=f(Y_{t-1})+\sigma(Y_{t-1})e_t. \tag{6}$$

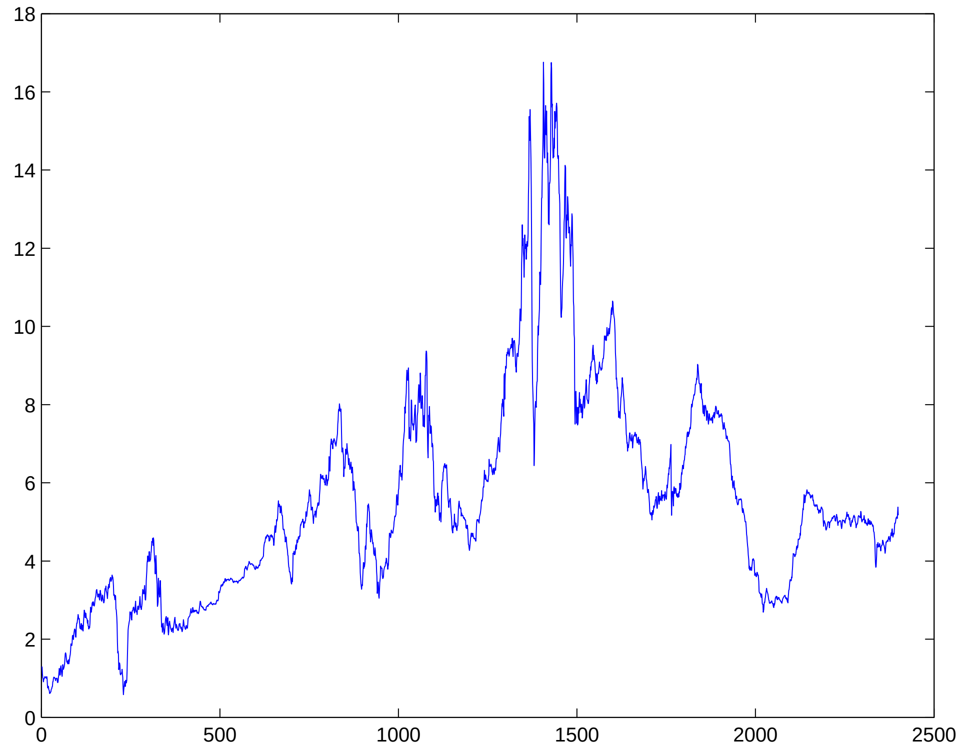

Figure 4: Yields of 3-month T-Bills, 1/8/1954-12/31/1999

To select the polynomial spline models, the degree of the polynomial is selected in the range of 0 through 6. To select the knots k_i, considering the rather large sample size (n=2400), k_i is chosen from the range of $[1, 100]$. At $k_i = 100$, one interior knot is placed on each percentile point, while when $k_i = 1$ the only knot is placed on the median of the "independent variable" Y_{t-1}. The degree of polynomial d_i and number of knots k_i $(i = 1, 2)$ are then selected using BIC. The exhaustive search procedure mentioned before results in a degree-5 spline with 1 knot for $f(\cdot)$, and a linear spline with 2 knots for $\sigma(\cdot)$. As for the innovation process $\{e_t\}$, BIC is used to identify the orders of the ARMA model, the result is an ARMA (1,2) model. Based on these model selection results, the identified model is :

$$Z_t = \sum_{i=1}^{7} \beta_i B_{1i}(Y_{t-1}) + \sqrt{\exp\left[\sum_{j=1}^{4} \alpha_j B_{2j}(Y_{t-1})\right]} \frac{(1-\theta_1 L)(1-\theta_2 L^2)}{1-\phi L} \varepsilon_t.$$

The proposed iterative estimation procedure is used in the estimation, the residual variance is 5.18. The estimated transfer function $f(Y_{t-1})$ is plotted against Y_{t-1} in Y_t is plotted against Y_{t-1} in the left panel of Figure 5, superimposed on the scatter diagram of Y_{t-1} against Y_t. In the right panel of Figure 5, the absolute value of $r(Y_{t-1}) = \frac{1-\phi L}{(1-\theta_1 L)(1-\theta_2 L^2)} \left[Y_t - f(Y_{t-1}) \right]$ is plotted against Y_{t-1}, on which the estimated conditional standard deviation function is superimposed. From both panels of Figure 5, we can see that the variability is rather stable and show small decrease when previous week's yield is less than 4 approximately, after which the variability starts to increase rapidly. The estimated ARMA parameters are given in Table 4. The ACF and PACF of the standardized residual (Figure 6), as well as the Ljung-Box $Q(10)$, indicate the residual is white noise.

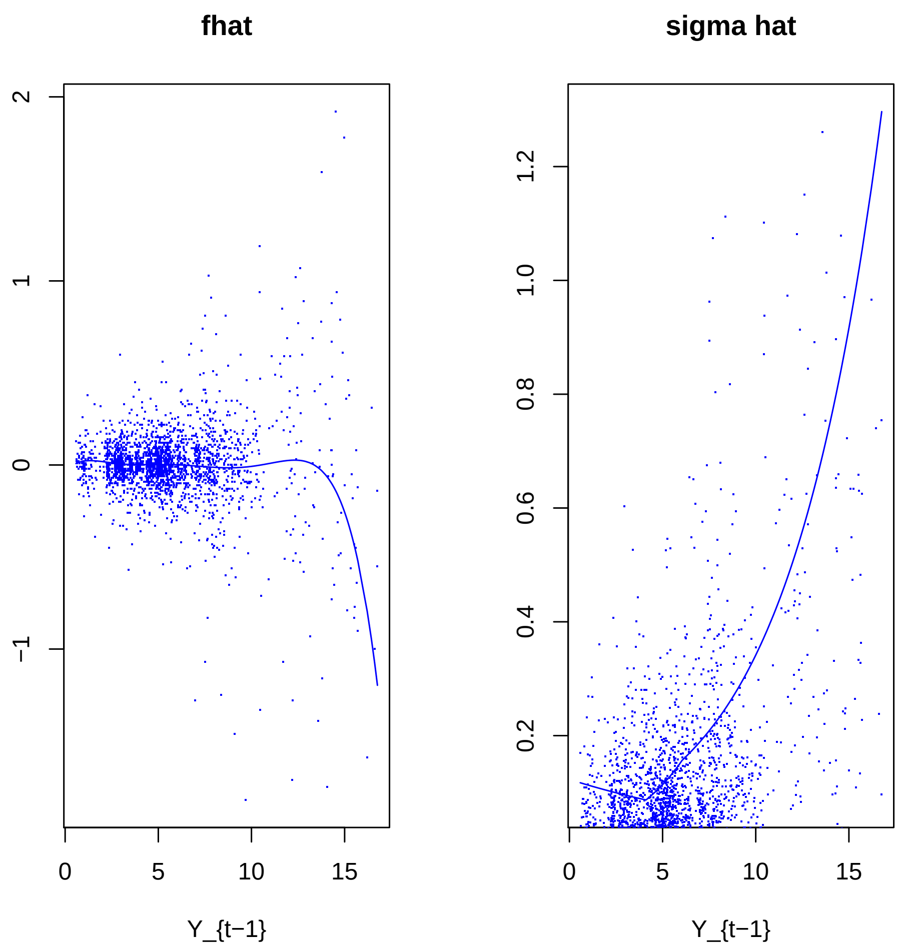

Figure 5: Estimated mean and volatility functions

6. Summary and Discussions

In this paper a nonparametric method is developed to model conditional variance functions in transfer function models. In this method, the conditional variance function and the transfer function are assumed to be smooth, no additional assumption is made about their functional forms. Polynomial splines are used to approximate the conditional variance function and the transfer function due to their flexibility and computational efficiency. The serial correlation in the innovation process is modeled as a stationary ARMA process. Through simulation studies, the proposed estimators are shown to process the "oracle" property, i.e., the conditional variance function estimator performs as if the transfer function and the ARMA parameters are known by some "oracle". Similarly, the transfer function estima-

Table 4: The Estimated ARMA(1,2) Cofficients

Coefficients	ϕ	θ_1	θ_2
Estimates	0.9036	-0.6478	-0.1967
s.e.	0.0492	0.0542	0.0264

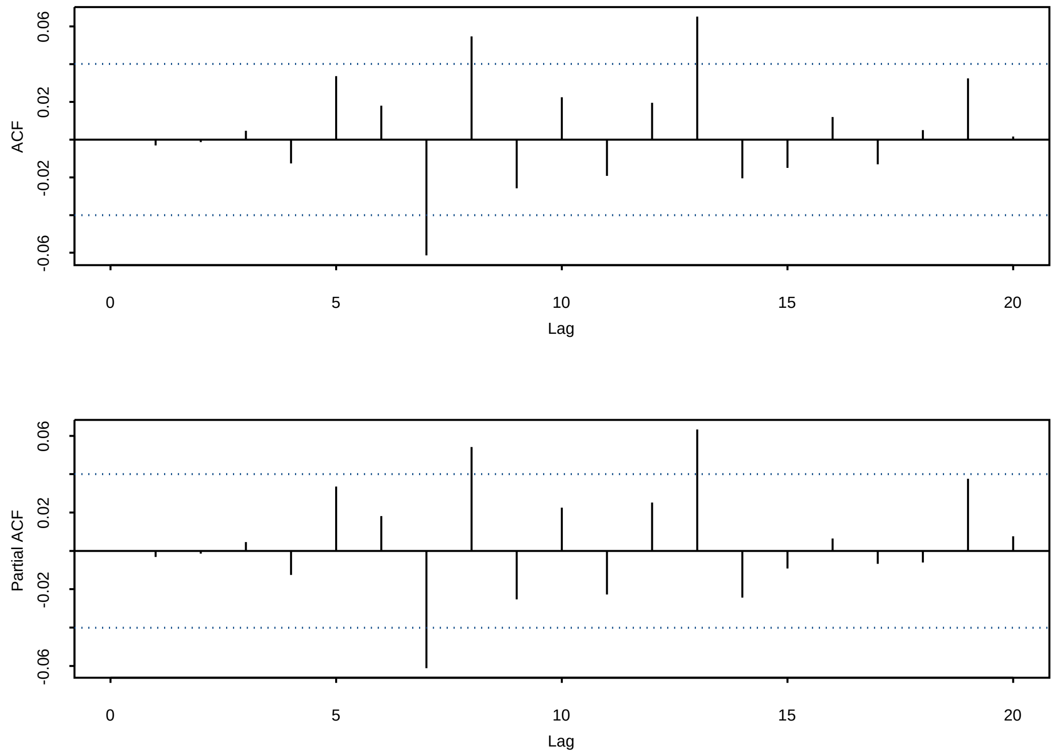

Figure 6: Residual Sample ACF and PACF

tor behaves as if the conditional variance function and the ARMA parameters are known. The ARMA estimators are shown to be approximately normally distributed, with mean and standard deviation close to those of the standard ARMA estimator with observed e_t. Without the need of making explicit assumptions on the forms of the unknown functions, this nonparametric model can be used when the underlying functions are highly nonlinear. By using polynomial splines, this method is very computationally efficient therefore applications with large number of observations can be analyzed with ease. It is also demonstrated that it is necessary to model the serial correlation in the data in order to improve the efficiency in estimating the transfer function and conditional variance function. The usefulness of this model is demonstrated in a classical example in the literature. The proposed method shows promising properties through simulation studies, but the asymptotic properties of the estimators need more rigorous study in the future.

REFERENCES

Alizadeh, S. and Brandt, M. and Diebold, F.X. (2002), "Range-based estimation od schedastic volatility models," *Journal of Finance 57, 1047–1092.*

Andersen, T.G. and Lund, J. (1997), "Estimating continuous-time stochastic volatility models of the short-term interest rate," Journal of Econometrics 77(2), *pp. 343–377.*

Anderson, T.W. and Bollerslev, T. and Diebold, F.X. and Labys, P. (2001), "The distribution of realized exchange rate volatility," Journal of the American Statistical Association 96, *pp. 42–55.*

Anderson, T.W. and Bollerslev, T. and Diebold, F.X. and Labys, P. (2001), "The distribution of realized stock volatility," Journal of Financial Economics 61, *pp. 43–76.*

Bai, X. and Russell, J.R. and Tiao, G.C. (2003), "Kurtosis of GARCH and stochastic volatility models with non-normal innovations," Journal of Econometrics 114, *pp. 349–360.*

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Hetroscedasticy," Journal of Econometrics 31*, pp. 307–327.*

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

Brockwell, P.J. and Davis, R.A. (1987), Time Series: Theory and Methods *(revised). Springer-Verlag: New York.*

Cao, C. and Tsay R.S. (1992), "Nonlinear time series analysis of stock volatilties," Journal of Applied Econometrics 7*, pp. s165–s185.*

Chan, K.C. and Karolyi, G and Longstaff, F. and Sanders A. (1992) "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," Journal of Finance 47*, pp. 1209–1227.*

Conley, T.G. and Hansen, L.P. and Luttmer, E.G.J. and Scheinkman, J.A. (1997) "Short-Term Interest Rates as Subordinated Diffusions," The Review of Financial Studies 10*, pp. 525–577.*

Cox, J.C. and J.E. Ingersoll and S.A. Ross. (1985), "A Theory of the Term Structure of Interest Rates," Econometrica 53*, pp. 385–407.*

Dahl, C.M. and Levine, M. (2006), "Nonparametric Estimation of Volatility Models With Serially Dependent Innovations," Statistics and Probability Letters 76*, pp. 2007–2016.*

de Boor, C. (2001). A Practical Guide to Splines *(revised). Springer-Verlag: New York.*

Engle, F.R. (1982). "Autoregressive Conditional Hetroscedasticy with Estimates of the Variance of the U.K. Inflation," Econometrica 50*, pp. 251–276.*

Fan, J. and Gilbels, I. (1996) Local Polynomial Modelling and Its Applications*. Chapman and Hall: Suffolk.*

Fan, J. and Yao, Q. (1998) "Efficient Estimation of Conditional Variance Function in Stochastic regression," Biometrika 85(3)*, pp. 645–660.*

Gallant, A.R. and Tauchen, G. (1997) "Estimation Of Continuous-Time Models For Stock Returns And Interest Rates," Macroeconomic Dynamics 1*, pp. 135–168.*

Harvey, A.C. and Ruiz, E and Sherhard, N. (1994), "Multivariate stochastic variance models," Revies of Economic Studies 61*, pp. 247–264.*

Huang, J.Z. (2003) "Local Asymptotics For Polynomial Spline Regression," The Annals of Statistics 31*, pp. 1600–1635.*

Liu, J.M. and Chen, R. and Yao, Q. (2010) "Nonparametric Transfer Function Models," Journal of Econometrics 157(1)*, pp. 151–164.*

Stone, C.J. (1982) "Optimal Global Rates of Convergence for Nonparametric Regression," The Annals of Statistics 10(4)*, pp. 1040–1053.*

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

Tsay, R.S. (1987). "Conditional heteroscedastic time series models," Journal of the American Statistical Association 82*, pp. 590–604.*

Yuan, M. and Wahba, G. (2004) "Doubly Penalized Likelihood Estimator in Heteroscedastic Regression," Statistics and Probability Letters 69(1)*, pp. 11–20.*