

# Nonparametric Lag Selection for Additive Models Based on the Smooth Backfitting Estimator

Zheng-Feng Guo<sup>1</sup>, Ling Yan Cao<sup>2</sup>, Ying He<sup>3</sup>

<sup>1</sup>Vanderbilt University, Nashville, USA

<sup>2</sup>University of Maryland, Maryland, USA

<sup>3</sup>Ninewell, Inc., Bethesda, Maryland, USA

E-mail: [zhengfeng.guo@vanderbilt.edu](mailto:zhengfeng.guo@vanderbilt.edu); [lycao@math.umd.edu](mailto:lycao@math.umd.edu); [yheumd@yahoo.com](mailto:yheumd@yahoo.com)

Received May 22, 2011; revised June 15, 2011; accepted June 22, 2011

## Abstract

This paper proposes a nonparametric FPE-like procedure based on the smooth backfitting estimator when the additive structure is a priori known. This procedure can be expected to perform well because of its well-known finite sample performance of the smooth backfitting estimator. Consistency of our procedure is established under very general conditions, including heteroskedasticity.

**Keywords:** Lag Selection, FPE, Consistency, Additive Models

## 1. Introduction

With the wide application of nonparametric techniques in the time series literature, many nonparametric lag selection criteria based on kernel smoothing methods have been proposed; such as nonparametric FPE (Tjostheim and Auestad, 1994 [1] and Tschernig and Yang, 2000 [2]) and cross validation (Cheng and Tong, 1992 [3]). Under very general assumptions, Tschernig and Yang (2000) [2] show the asymptotic equivalence of cross validation and nonparametric FPE and the consistency of the latter procedure originally proposed by Tjostheim and Auestad (1994) [1]. Unfortunately, despite the desirable asymptotic property of the FPE procedure, Tschernig and Yang (2000) [2] point out that overfitting models are selected too often when the sample size is small.

Until recently, Guo and Shintani (2011) [4] impose the additivity assumption and propose a consistent FPE-like lag selection procedure based on the marginal integration method by Linton and Nelson (1995) [5]. In contrast to the unrestricted FPE procedure without the additivity assumption, the additive nonparametric FPE-like procedure performs reasonably well in small samples and generally outperforms the unrestricted FPE due to the reduction of overfitting.

As is discussed in the conclusion of Guo and Shintani (2011) [4], the better finite sample properties of the backfitting method over marginal integration have been reported in simulation studies (e.g., Sperlich, Linton and Hardle, 1999 [6]). The possibility of developing a more

effective lag selection procedure based on the smooth backfitting remains to be studied. In this paper, we close this gap and propose a consistent FPE-like procedure based on the smooth backfitting estimator. We provide the conditions required for the consistency. In contrast to the FPE-like procedure by Guo and Shintani (2011) [4], our procedure can be expected to perform better in the finite sample because of its well-known desirable properties.

The remainder of the paper is organized as follows. In section 2, we introduce the model and discuss the asymptotic properties of our procedure. Section 3 discusses the implementation and the consistency of the criterion.

## 2. The Nonparametric FPE for Additive Models

In this paper, we consider the problem of selecting the combination of lag  $S = \{i_1, i_2, \dots, i_m\}$ , where  $i_j > i_k$  for  $j > k$ , in an additive AR model with the form of

$$Y_t = c + \sum_{i \in S} f_i(Y_{t-i}) + \sigma(X_t)\xi_t$$

for  $t = 1, \dots, n$ , where  $X_t = (Y_{t-i_1}, Y_{t-i_2}, \dots, Y_{t-i_m})$ . Since the convergence rate of additive regression estimators does not depend on the dimension of the model, we do not impose any restriction on  $m$  ( $\leq i_m$ ). Below are our main assumptions.

### Assumptions.

- For some integer  $M > i_m$ , the vector process  $X_{M,t} =$

$(Y_{t-1}, Y_{t-2}, \dots, Y_{t-M})$  is strictly stationary and  $\beta$ -mixing with  $\beta(n) \leq c_0 n^{-2(2+\delta)/\delta}$  for some  $\delta > 0$  and  $c_0 > 0$ .

- The stationary distribution of the process  $X_{M,t}$  has a continuous differentiable density  $\mu(X_M)$ .
- The autoregression function for  $i \in S$  is twice continuously differentiable while  $\sigma(\cdot)$  is continuous and positive on the support of  $\mu(\cdot)$ .
- $\xi_t$  is a sequence of i.i.d random variables with  $E(\xi_t) = 0$ ,  $E(\xi_t^2) = 0$  and a finite fourth moment.
- The support of the weight function  $\omega(\cdot)$  is compact with nonempty interior. The function  $\omega(\cdot)$  is continuous, nonnegative and  $\mu(x_M) > 0$  for  $x_M \in \text{supp}(\omega)$ .
- The kernel-based nonparametric additive regression estimator  $\hat{f}_i(x_i)$  for  $i \in S$  converges to  $f_i(x_i)$  at the one-dimensional rate.

In estimating the additive AR model, we employ the smooth backfitting estimator, which is a useful practical variant of the classical backfitting estimator (see Mammen, Linton and Nielsen, 1999 [7], and Nielsen and Sperlich, 2005 [8]). By using an analogy to the asymptotic FPE of Tschernig and Yang (2000 [2]), the second term in formula (7) of Tjostheim and Auestad (1994) [1] is decomposed as follows

$$\begin{aligned} & E \left[ \sum_{i \in S} f_i(Y_{t-i}) - \sum_{i \in S} \hat{f}_i(Y_{t-i}) \right]^2 \omega(X_{M,t}) \\ &= E \left[ \sum_{i \in S} f_i(Y_{t-i}) - \sum_{i \in S} E \hat{f}_i(Y_{t-i}) + \sum_{i \in S} E \hat{f}_i(Y_{t-i}) - \sum_{i \in S} \hat{f}_i(Y_{t-i}) \right]^2 \omega(X_{M,t}) \\ &= E \left[ (I' + II')^2 \omega(X_{M,t}) \right] \end{aligned}$$

We can show

$$\begin{aligned} & E \left[ (II')^2 \omega(X_{M,t}) \right] \\ &= h^4 c_h^4 \sigma_K^4 \int \left[ \sum_{i \in S} \beta_i(x_i) \right] \cdot \omega(x_M) \mu(x_M) dx_M \end{aligned}$$

with  $c_h$  being the limit of  $n^{-1/5}h$  and

$$\beta_i(x_i) = \sum_{i \in S} \left[ \frac{\nabla f_i(x_i)}{\mu(x)} \frac{\partial \mu(x)}{\partial x_i} + \frac{1}{2} \nabla^2 f_i(x_i) \right]$$

when the local linear estimator is used for  $\hat{f}(x)$ , or with  $\beta_i(x_i) = \nabla^2 f_i(x_i) - \int \nabla^2 f_i(x_i) \mu_j(x_j) dx_j$  when the local constant (Nadaraya-Watson) estimator is used for  $\hat{f}(x)$ .

Similarly, we can show that  $E \left[ (I')^2 \omega(X_{M,t}) \right]$  converges to

$$\frac{1}{nh} \|K\|^2 \int \left[ \sum_{i \in S} \frac{\sigma_i^2(x_i)}{c_h \mu(x_i)} \right] \omega(x_M) \mu(x_M) dx_M$$

where

$$\sigma_i^2(x_i) = \text{var}(Y - f(x) | X_i = x_i).$$

Therefore, we can define AFPE as

$$AFPE = A + \frac{1}{nh} \|K\|_2^2 B + h^4 \frac{\sigma_K^4}{4} C \tag{1}$$

with

$$\begin{aligned} A &= \int \sigma^2(x_M) \omega(x_M) \mu(x_M) dx_M, \\ B &= \|K\|_2^2 \int \left[ \sum_{i \in S} \frac{\sigma_i^2(x_i)}{c_h \mu(x_i)} \right] \omega(x_M) \mu(x_M) dx_M, \\ C &= \int \left[ \sum_{i \in S} r_i(x_i) \right]^2 \omega(x_M) \mu(x_M) dx_M, \end{aligned}$$

where  $\|K\|_2^2 = \int K^2(u) du$ ,  $\sigma_K^2 = \int K(u) u^2 du$  and  $r_i(x_i)$  is the term appears in the asymptotic bias of the estimator  $\hat{f}_i(x)$ . The optimal bandwidth,  $h_{\text{opt}}$  which minimizes equation (1), is given by

$$h_{\text{opt}} = \left\{ \|K\|_2^2 \sigma_K^{-4} B C^{-1} \right\}^{1/5} n^{-1/5}.$$

### 3. Estimation and the Consistency of our Criterion

Our criterion for additive AR models takes the form

$$FPE(S) = \hat{A} + \frac{1}{nh^{(m-1)\eta+1}} 2K(0) \hat{B}$$

where  $\eta \in [0, 1]$ ,

$$\hat{A} = n^{-1} \sum_{t=1}^n \left( Y_t - \sum_{i \in S} \hat{f}_i(Y_{t-i}) \right)^2 \omega(X_{M,t}),$$

and

$$\hat{B} = n^{-1} \sum_{t=1}^n \sum_{i \in S} \frac{(Y_t - \hat{f}_i(Y_{t-i}))^2}{\hat{\mu}(Y_{t-i})} \omega(X_{M,t}).$$

The first term in  $FPE(S)$  corresponds to the measure of regression fit in traditional information criteria for the model selection, while the second term serves as a penalty to avoid overfitting, depending on a tuning parameter  $\eta$ .

We follow Tschernig and Yang (2000) [2] and focus on the case when the optimal bandwidth  $h_{\text{opt}}$  is used for  $\hat{f}(x)$  in  $\hat{A}$ , but any bandwidth of order  $n^{-1/5}$  can be used for  $\hat{f}(x)$  in  $\hat{B}$ . We select the subset  $\hat{S} = \{i'_1, i'_2, \dots, i'_m\} \subseteq \{1, 2, \dots, M\}$  which minimizes  $FPE(S)$  among all possible combinations of  $\{1, 2, \dots, M\}$ . The selected  $\hat{S} = S'$  overfits if  $\hat{S} \supseteq S'$  and  $\hat{S} \neq S'$ ; and underfits if it does not overfit and  $\hat{S} \neq S'$ .

The lag selection procedure is consistent if the probability of  $\hat{S} = S$  approaches one as  $n \rightarrow \infty$ .

Theorem 1: Under our assumptions and  $\eta \in [0,1]$ , as  $n \rightarrow \infty$ ,

$$\frac{FPE(S') - A}{FPE(S) - A} \rightarrow +\infty$$

for any overfitting combinations  $S' = \{i'_1, i'_2, \dots, i'_m\}$ .

The overfitting  $FPE(S)$  asymptotically becomes larger than the correctly specified  $FPE(S)$  because the penalty of the former converges at a rate slower than the latter as long as  $\eta > 0$ . It should be noted that  $h'_{opt}$  used for  $FPE(S')$  differs from  $h_{opt}$ . Unlike the unrestricted FPE, however, the convergence rates of two bandwidths are the same for additive models even if the dimensions of the regressors are different, that is why  $\eta = 0$  is not desirable in excluding overfitting models. Following the same argument as in Guo and Shintani (2011) [4], we can easily show that the FPE for underfitting case is larger than that of a correctly fitting model. Then we have:

Theorem 2. Under our assumptions and  $\eta \in [0,1]$ , as  $n \rightarrow \infty$ ,

$$P[\hat{S} = S] \rightarrow 1.$$

#### Remarks

- The consistency of our procedure holds for both local linear and local constant estimators.
- If  $\eta > 0$ , the probability of selecting the correct model converges to one as the sample size increases. If  $\eta = 0$ , our criterion is asymptotically equivalent to the asymptotic FPE.
- While the FPE-like procedure by Guo and Shintani (2011) [4] and our procedure are both consistent, the latter procedure can be expected to perform better in the finite sample because of better finite sample performance of our procedure.

#### 4. Conclusions

The better finite sample properties of the backfitting method over marginal integration have been often reported in many simulation studies. Guo and Shintani (2011) [4] propose a FPE-like procedure based on the marginal integration method due to its simplicity. Our paper proposes a more effective lag selection criterion based on the smooth backfitting estimator. The new criterion can be expected to perform better in the finite sample.

thod over marginal integration have been often reported in many simulation studies. Guo and Shintani (2011) [4] propose a FPE-like procedure based on the marginal integration method due to its simplicity. Our paper proposes a more effective lag selection criterion based on the smooth backfitting estimator. The new criterion can be expected to perform better in the finite sample.

#### 5. References

- [1] D. Tjøstheim and B. Auestad, "Nonparametric Identification of Nonlinear Time Series: Selecting Significant Lags," *Journal of the American Statistical Association*, Vol. 89, No. 428, 1994, pp. 1410-1419. [doi:10.2307/2291003](https://doi.org/10.2307/2291003)
- [2] R. Tschernig and L. Yang, "Nonparametric Lag Selection for Time Series," *Journal of Time Series Analysis*, Vol. 21, No. 4, 2000, pp. 457-585. [doi:10.1111/1467-9892.00193](https://doi.org/10.1111/1467-9892.00193)
- [3] B. Cheng and H. Tong, "On Consistent Nonparametric Order Determination and Chaos," *Journal of the Royal Statistical Society series B (Methodological)*, Vol. 54, No. 2, 1992, pp. 427-449.
- [4] Z. F. Guo and M. Shintani, "Nonparametric Lag Selection for Additive Models," *Economics Letters*, Vol. 2, No. 2, 2011, pp. 131-134. [doi:10.1016/j.econlet.2011.01.014](https://doi.org/10.1016/j.econlet.2011.01.014)
- [5] O. B. Linton and J. P. Nielsen, "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, Vol. 82, No. 1, 1995, pp. 93-100. [doi:10.1093/biomet/82.1.93](https://doi.org/10.1093/biomet/82.1.93)
- [6] S. Sperlich, O. B. Linton and W. Härdle, "Integration and Backfitting Methods in Additive Models-Finite Sample Properties and Comparison," *Test*, Vol. 8, No. 2, 1999, pp. 419-458. [doi:10.1007/BF02595879](https://doi.org/10.1007/BF02595879)
- [7] E. Mammen, O. B. Linton and J. P. Nielsen, "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions," *Annals of Statistics*, Vol. 27, No. 5, 1999, pp. 1443-1490. [doi:10.1214/aos/1017939137](https://doi.org/10.1214/aos/1017939137)
- [8] J. P. Nielsen and S. Sperlich, "Smoothing Backfitting in Practice," *Journal of the Royal Statistical Society Series B*, Vol. 67, No. 1, 2005, pp. 43-61. [doi:10.1111/j.1467-9868.2005.00487.x](https://doi.org/10.1111/j.1467-9868.2005.00487.x)