Semiparametric Forecasting of Nonlinear Temporal Processes

Joshua Patrick* Jane Harvill†

Abstract

The curse of dimensionality has been problematic in the application of nonparametric and semi-parametric regression techniques to high-dimensional time series data. Spline-backfitted local linear (SBLL) and spline-backfitted kernel (SBK) estimators have been successful in addressing this problem, and provide computationally efficient estimators. Moreover, under fairly weak conditions, the estimators are point-wise asymptotically normal. Little work has been conducted in investigating the properties of forecasts using models estimated via SBLL or SBK methods. We propose a method for SBLL and SBK forecasting, and investigate the properties of those forecasts. For illustration, we apply the forecasting methods to irradiance data collected from a solar power plant in Sacramento, California provided by Sandia Research Laboratories.

Key Words: Spline-backfitting, Nonlinear time series, Forecasting, Oracle estimation, Solar energy

1. Introduction

Nonlinear models are often more appropriate than linear models for fitting some time series data. Some non-linear dynamic systems include limit cycles, jump phenomenon, time irreversibility, and amplitude-frequency models (Tong 1990). The class of nonlinear time series models is extremely large and include models such as threshold autoregressive (TAR), exponential autoregressive (EXPAR), bilinear, smooth threshold autoregressive (STAR), and generalized autoregressive conditional heteroscedasticity (GARCH) models. A useful structure that can reduce the size of the class of nonlinear models is the functional-coefficient autoregressive (FCAR) model.

Härdle, Lütkepohl, and Chen (1997) give a summary of nonparametric techniques that have been found useful in nonlinear time series analysis. However, problems such as the "curse of dimensionality" arise when fitting nonlinear models with nonparametric methods. Cai, Fan, and Yao (2000) used a local linear regression technique to estimate FCAR models. Their method is flexible and reduces the curse of dimensionality. Wang and Yang (2007) and Wang and Yang (2009) have developed methods known as spline-backfitted kernel smoothing (SBK) and spline-backfitted local linear smoothing (SBLL) that outperform the local linear method for high dimensional data. However neither of these papers extend these methods to forecasting. In this paper, we apply the SBK and SBLL methods to FCAR models and propose forecasting methods using the SBK and SBLL estimators. We apply these methods to solar irradiance data.

The paper is structured as follows: Section 2 describes the FCAR model and estimation using the SBK and SBLL methods. Preliminary simulations for estimating the FCAR model and the irradiance data application are shown in Section 3. The forecasting methods are proposed in Section 4, and we conclude with a discussion.

*Baylor University, Department of Statistical Science, Waco, TX
†Baylor University, Department of Statistical Science, Waco, TX

2. SBK and SBLL Estimation of FCAR Models

We first introduce the the SBK method for the i.i.d. case. For the sample $\{Y_i,\ X_{i1},\ldots,\ X_{id}\}_{i=1}^{n}$ the additive model is

$$Y_i = m\left(X_i\right) + \sigma\left(X_i\right)\varepsilon_i \tag{1}$$

where

$$m\left(X_i\right) = c + \sum_{\alpha=1}^{d} m_\alpha\left(x_{i\alpha}\right),$$

$$E\left[m_\alpha\left(X_\alpha\right)\right] \equiv 0,\quad \alpha = 1,\ldots,d,$$

and $\varepsilon_i \sim N\left(0,1\right)$. The goal is to estimate the d unknown component functions $m_\alpha\left(X_\alpha\right)_{\alpha=1}^{d}$ based on the sample. Linton (1997) used one step kernel backfitting based on marginal integration and Huang and Yang (2004) used polynomial spline estimation to estimate the component functions. The marginal integration method is computationally expensive and the polynomial spline method lacks the asymptotic properties of kernel smoothing. Wang and Yang (2009) have shown the SBK method is computationally expedient while having asymptotic properties allowing for point-wise confidence intervals and confidence bands.

2.1 Oracle smoothers

The computational and asymptotic properties of the SBK method are based on the idea of oracle smoothing. For model (1), if the last $d-1$ of the component functions were known, then we can define a new variable

$$Y_{i,1} = Y_i - c - \sum_{\alpha=2}^{d} m_\alpha\left(X_{i,\alpha}\right).$$

The variable $Y_{i,1}$ is said to be known by "oracle." We can now use $Y_{i,1}$ to regress on the numerical variable $X_{i,1}$ to estimate the only unknown function $m_1\left(X_1\right)$. If $\{m_\alpha\}_{\alpha=2}^{d}$ is known by "oracle," define the kernel "oracle smoother" as

$$\tilde{m}_1\left(x_1\right) = \frac{\sum_{i=1}^{n} K_h\left(X_{i,1}-x_1\right)Y_{i,1}}{\sum_{i=1}^{n} K_h\left(X_{i,1}-x_1\right)}$$

where $h \sim n^{-1/5}$ is the bandwidth and K is the kernel function to determine how to assign the weights. The same process can be used for each of the d component functions where the remaining $d-1$ component functions are used to define the oracle response variable $Y_{i,\alpha}$.

Bosq (1998) has shown that the oracle smoother has the asymptotic properties

$$\sup_{x\in[h,1-h]}\left|\tilde{m}_\alpha\left(x\right) - m_\alpha\left(x\right)\right| = o_p\left(n^{-2/5}\log n\right)$$

and

$$\sqrt{nh}\left\{\tilde{m}_\alpha\left(x\right) - m_\alpha\left(x\right) - b_\alpha\left(x\right)h^2\right\} \xrightarrow{D} N\left\{0, v_\alpha^2\left(x\right)\right\}$$

where

$$b_\alpha\left(x\right) = \int u^2 K\left(u\right)du\left\{m_\alpha''\left(x\right)f_\alpha\left(x\right)/2 + m_\alpha'\left(x\right)f_\alpha'\left(x\right)\right\}f_\alpha^{-1}\left(x\right)$$

and

$$v_\alpha^2\left(x\right) = \int K^2\left(u\right)du E\left[\sigma^2 x\left(X_1,\ldots,X_d\right)|X_\alpha = x\right]f_\alpha^{-1}\left(x\right).$$

Note that the $b_\alpha(x)$ term requires the second moment of the kernel K to exist.

Use of the oracle smoother requires us to know the $d-1$ functions used to construct the oracle response variable. The SBK method estimates these functions with an under-smoothed constant spline. Without loss of generality, let each X_α be defined on the compact set [0, 1]. Pre-select an integer $N_n \sim n^{2/5}\log(n)$ and define the indicator function of the $(N+1)$ equally-spaced subintervals on $[0,\ 1]$

$$I_{J,\alpha}(x_\alpha) = \begin{cases} 1 & JH \le x_\alpha < (J+1)H \\ 0 & \text{otherwise,} \end{cases}, \quad H = (N_n+1)^{-1},\ J = 0,\ 1,\ \ldots,\ N.$$

The spline estimator of $m(x)$ is

$$\hat{m}(x) = \hat{\lambda}_0 + \sum_{\alpha=1}^{d}\sum_{J=1}^{N}\hat{\lambda}_{J,\alpha} I_{J,\alpha}(x_\alpha)$$

where the coefficients $\left(\hat{\lambda}_0,\ \hat{\lambda}_{1,1},\ \ldots,\ \hat{\lambda}_{N,d}\right)$ minimize

$$\sum_{i=1}^{n}\left\{Y_i - \lambda_0 - \sum_{\alpha=1}^{d}\sum_{J=1}^{N}\lambda_{J,\alpha} I_{J,\alpha}(x_\alpha)\right\}^2.$$

We pre-estimate $\{m_\alpha(x_\alpha)\}_{\alpha=1}^{d}$ by its pilot estimate $\left\{\hat{m}_\alpha(x_\alpha)\right\}_{\alpha=1}^{d}$ with an under-smoothed constant spline. Using these pilot estimates, we construct the pseudo-responses

$$\hat{Y}_{i,\alpha} = Y_i - \hat{c} - \sum_{\beta\neq\alpha}\hat{m}_\beta\left(x_{i,\beta}\right).$$

For the pseudo-data $\left\{\hat{Y}_{i,\alpha},\, X_{i,\alpha}\right\}_{i=1}^{n}$, the SBK estimate is

$$\hat{m}_\alpha^*(x_\alpha) = \frac{\sum_{i=1}^{n} K_h\left(X_{i,\alpha} - x_\alpha\right)\hat{Y}_{i,\alpha}}{\sum_{i=1}^{n} K_h\left(X_{i,\alpha} - x_\alpha\right)}.$$

Wang and Yang (2009) have shown that $\hat{m}_\alpha^*(x_\alpha)$ has the asymptotic properties of the oracle smoother.

2.2 Estimation for the FCAR model

For time series data, the functional coefficient model has the form

$$X_t = a_1(X_{t-d})X_{t-1} + \cdots + a_p(X_{t-d})X_{t-p} + \sigma(X_{t-d})\varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent random variables with mean of zero and variance of one. The variable X_{t-d} is called the delay variable and $a_1(\cdot), \ldots, a_p(\cdot)$ are the unknown coefficient functions. Model (2) can be denoted by FCAR(p, d). Letting

$$Y_i = X_t, X_{t1} = X_{t-1}, \cdots, X_{tp} = X_{t-p}, U_t = X_{t-d},$$

we can write (2) as

$$Y = a_1(U)X_1 + \cdots + a_p(U)X_p + \sigma(U)\varepsilon_t. \tag{3}$$

Cai et al. (2000) estimated the functions $a_j(\cdot)$'s in (3) using local linear regression. Assuming $a_j(\cdot)$ has a continuous second derivative, we can approximate $a_j(\cdot)$ locally at u_0 by a linear function

$$a_j(u) \approx a_j + b_j(u - u_0).$$

The local linear estimator is defined as $\hat{a}_j(u_0) = \hat{a}_j$, where $\left\{\left(\hat{a}_j, \hat{b}_j\right)\right\}$ minimizes the sum of weighted squares

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

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and h is a bandwidth. The estimates for $a_j(\cdot)$ are

$$\hat{a}_j(u_0) = \sum_{k=1}^{n} K_{n,j}(X_k, U_k - u_0)\, Y_k$$

where

$$K_{n,j}(x, u) = e_{j,p}^T\left(\tilde{X}^T W \tilde{X}\right)\begin{pmatrix} x \\ ux \end{pmatrix} K_h(h),$$

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

Using the spline-backfitted smoothing method, we can obtain pseudo responses

$$\hat{Y}_\alpha = Y - \sum_{\beta \neq \alpha} \hat{m}_\beta\left(U, X_\beta\right)$$

where $m_\beta\left(U, X_\beta\right) = a_\beta(U)\, X_\beta$ and $\hat{m}_\beta\left(U, X_\beta\right)$ is estimated using constant spline smoothing. The SBK estimator for $a_j(U)$ is

$$\hat{a}_j(u_0) = \sum_{k=1}^{n} K_{n,j}(X_k, U_k - u_0)\, \hat{Y}_{j,k}.$$

Due to preestimating the coefficient functions with under-smooth splines, we expect the SBK method will perform better in mean square error (MSE) than the local linear regression method of Cai et al. (2000), particularly for large p.

3. Preliminary Simulations and Irradiance Data Example

We compare the SBK method to the local linear regression (LL) method of Cai et al. (2000) through some preliminary simulations. These simulations examine the performance of the two methods on an exponential autoregressive (EXPAR) model. The EXPAR model was chosen for comparison with the results in Cai et al. (2000). We then fit a FCAR model to solar irradiance data obtained from Sandia National Laboratories.

3.1 Simulation results

We simulated data from two EXPAR models and compared the MSE of the SBK and LL methods. For the first example, consider the EXPAR model

$$x_t = a_1(x_{t-1})\, x_{t-1} + a_2(x_{t-1})\, x_{t-2} + \varepsilon_t \tag{4}$$

where $a_1(u) = 0.138 + (0.316 + 0.982u)\exp\left(-3.89u^2\right)$, $a_2 = -0.437(u) - (0.659 + 1.26u\exp\left(-3.89u^2\right)$, and $\{\varepsilon_t\} \sim N\left(0, 0.2^2\right)$. We simulated this model 500 times for series lengths of 100, 250, 500, and 1000. For each iteration, we calculated the mean square error (MSE) for both the SBK and the LL methods. Figure 1 shows the boxplots of the MSE's of the fit of $a_1(x_{t-1})$ for both methods. As expected, the SBK method has much smaller MSE

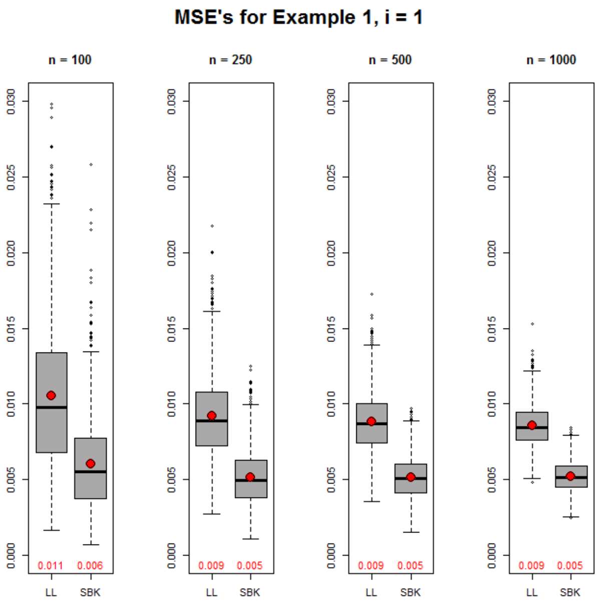

Figure 1: Boxplots of the MSE for the SBK and LL fits of $a_1\left(x_{t-1}\right)$ in (4).

for smaller series lengths. As the series lengths increase, the MSE for the LL method gets closer to the SBK method. These results show that the SBK method does not suffer from the "curse of dimensionality" as much as the LL method.

For the second example, consider the the model

$$x_t = \sum_{i=1}^{6} a_i\left(x_{t-1}\right) x_{t-i} + \varepsilon_t \tag{5}$$

where $a_1(u) = 0.1 + (0.3+0.5u)\exp\left(-3.89u^2\right)$, $a_2(u) - (0.6+u)\exp\left(-3.89u^2\right)$, $a_3(u) = 0.5 + (0.7+0.9u)\exp\left(-3.89u^2\right)$, $a_4(u) = -0.3 - (0.5+0.9u)\exp\left(-3.89u^2\right)$, $a_5(u) = 0.2 + (0.5+u)\exp\left(-3.89u^2\right)$, $a_6(u) = -0.1 - (0.3+0.5u)\exp\left(-3.89u^2\right)$, and $\{\varepsilon_t\} \sim N\left(0, 0.2^2\right)$. Figure 2 shows the boxplot of the fit for $a_2\left(x_{t-1}\right)$. The results are similar to the results in Figure 1. The SBK method performs much better for smaller series lengths. For this example, the MSE's for the LL method do not get as close to the SBK method. Larger series lengths would likely show the MSE's getting closer since this example has 6 functional coefficients.

Another advantage of the SBK method is the small computing time required to fit the model. For comparison, the computing time was recorded for each method for each iteration. The simulations were conducted in R on a PC with an Intel Xeon 2.49 GHz processor and 4.0 GB RAM. Figure 3 shows the mean computing time per iteration for both examples. Clearly, the time required for the SBK method is much smaller than for the LL method.

3.2 Solar irradiance example

3.2.1 Data set description

Power system output of a photovoltaic (PV) plant is affected by variation in solar resources. These resources are known as solar irradiance. PV cells are set up in some type of arrangement to simulate the amount of irradiance measured for a utility-scale PV plant. These cell arrangements can be in an array layout with all of the cells close together or the cells can

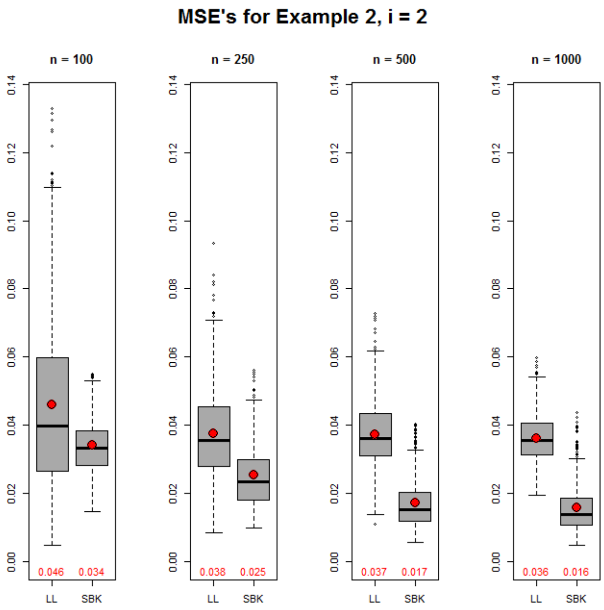

Figure 2: Boxplots of the MSE for the SBK and LL fits of $a_2\ (x_{t-1})$ in (5).

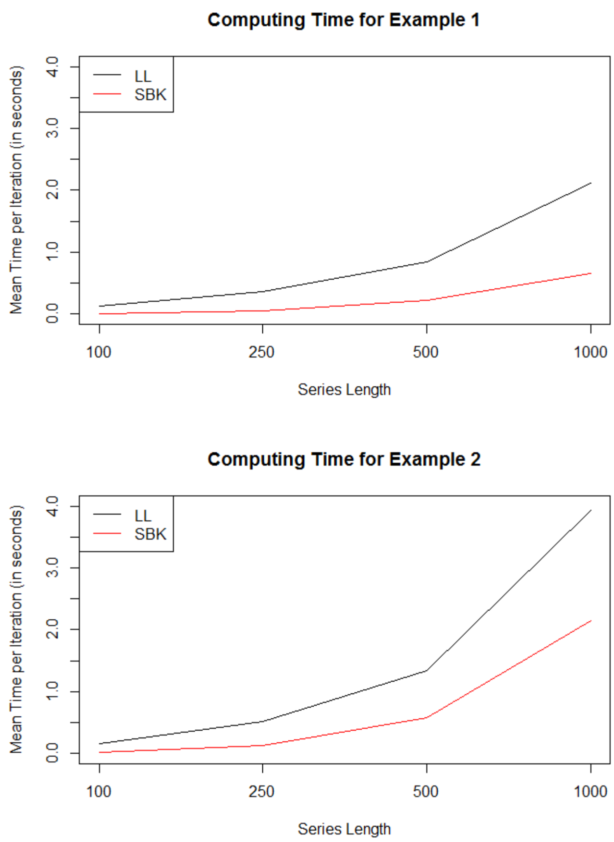

Figure 3: Mean computing time per iteration for example 1 (top) and example 2 (bottom).

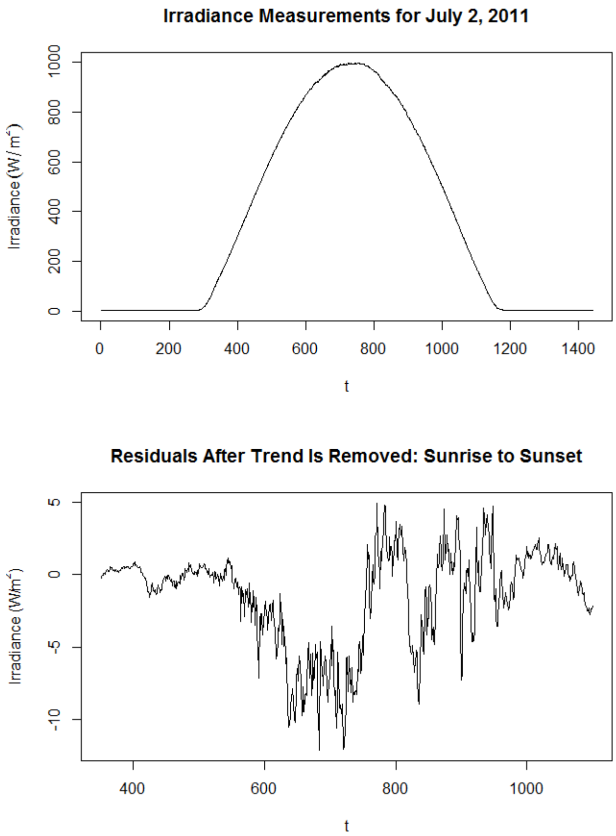

Figure 4: (Top) Irradiance measurements for sensor 20 on July 2, 2011. (Bottom) Residuals after the diurnal trend is removed for sunrise to sunset.

be spread out over a large area. Predicting the irradiance measurements from these cells is key in the development of utility-scale PV plant models. A statistical model is needed to aggregate the irradiance that can be calibrated with measurements.

The Sacramento irradiance data set contains 66 separate irradiance sensors spread out over a 2300 square kilometer area. Each sensor records an irradiance measurement at every minute of the day. Figure 4 shows the irradiance measurements (in W/m^2) for sensor 20 on July 2, 2011. We used this day since it is a clear day and we fit a FCAR model using the SBK method.

3.2.2 *Fitting the model*

We first remove the diurnal trend and then fit the model

$$X_t = a_1\left(X_{t-1}\right) X_{t-1} + a_2\left(X_{t-1}\right) X_{t-2} + \varepsilon_t \tag{6}$$

to the residuals. Figure 4 shows the residuals after the trend is removed for times between sunrise and sunset. Models with more dimensions were also fitted but the model in (6) had the smallest MSE. Figure 5 shows the fit of the SBK method to the measurements. The plot only shows the time interval from noon to 4 p.m. in order to see the fit more clearly. It appears that the SBK method was able to fit the data very well. Figure 5 also shows the estimate of $a_1\left(X_{t-1}\right)$. Future research will attempt to modify the model in (6) to fit the data on cloudy days.

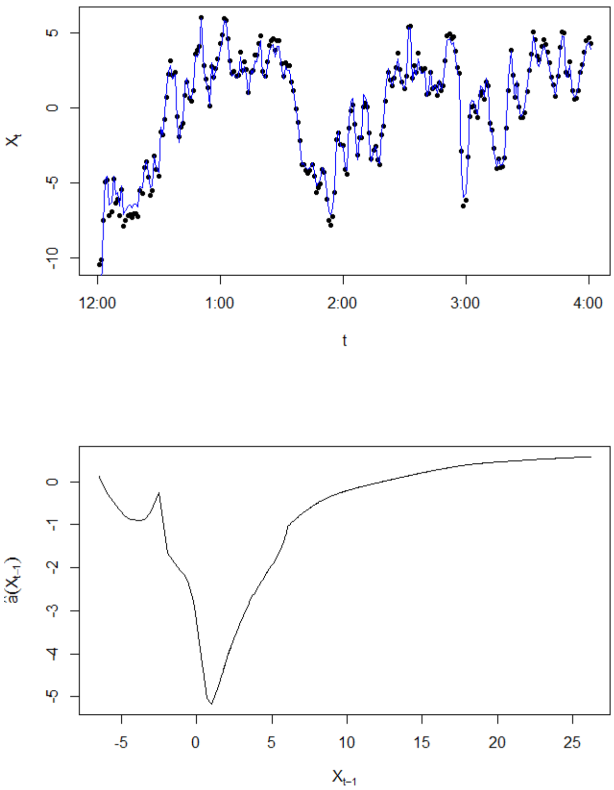

Figure 5: (Top) Fit of the model in (6) using the SBK method (line) to the irradiance data (points) from noon to 4 p.m. (Bottom) The estimate of a_1 (X_{t-1}).

4. Forecasting Methods

The goal in forecasting is to find an estimator of $\mathrm{E}\left[X_{n+M} \mid X_n, \ldots, X_{n-p}\right]$. For FCAR models, we want an estimator of

$$\mathrm{E}\left[X_{n+M} \mid X_n, \ldots, X_{n-p}\right] \quad = \quad \mathrm{E}\left[\sum_{j=1}^{p} f_j(Y_{n+M}) X_{n+M-j} | X_n, \ldots, X_{n-p}\right]. \tag{7}$$

Three methods exist for estimating (7): direct and iterative plug-in methods (Fan and Yao 2003), the bootstrap method (Harvill and Ray 2005), and multistage smoothing (Chen 1996, Harvill and Ray 2005).

4.1 Plug-in predictor

The plug-in predictor estimates the coefficients using within-series values. This estimation is done simply by plugging-in $\hat{X}_{t+M-j}$ into the forecast equation. If ${}_t = X_{t-d}$, then we have

$$\hat{X}_{n+M} = \sum_{j=1}^{p} \hat{f}_j\left(\hat{X}_{n+M-d}\right)\hat{X}_{n+M-j},$$

where $\hat{X}_t = X_t$ if $t \leq n$ and $\hat{f}_j$ are some estimate of f_j. The plug-in predictor ignores the fact that the expectation in (7) is not linear in X_{t+M-j}.

4.2 Bootstrap predictor

Similar to the plug-in predictor, the bootstrap predictor uses within-series to estimate f_j. Unlike the plug-in predictor, the predicted values are computed as

$$\hat{X}_{n+M} = \sum_{j=1}^{p} \hat{f}_j\left(\hat{X}_{n+M-d}\right)\hat{X}_{n+M-j} + \epsilon^{(b)},$$

where $\epsilon^{(b)}$ is a bootstrapped value of the within-series residuals from the fitted FCAR model. The forecast is obtained for $b = 1, \ldots, B$ and the M-step ahead forecast is the average across all bootstrap replications. Huang and Shen (2004) proposed a similar method using polynomial splines to estimate FCAR coefficients, however they note that if $\hat{X}_{t+M-d}$ falls outside or near boundary of range of X_{t-d}, the estimated functional coefficients can be unreliable.

4.3 Multistage predictor

The multistage predictor is a modification of the plug-in predictor. This predictor updates the coefficient estimates at each step to incorporate information from X_t encoded in the predicted response at time $n + j,\ \ j = 1, 2, \ldots, M - 1$;

$$\hat{X}_{n+M} = \sum_{j=1}^{p} \hat{f}_j^M\left(\hat{X}_{n+M-d}\right)\hat{X}_{n+M-j},$$

where $\hat{X}_t = X_t,\ t \leq n$, and $\hat{f}_j^M$ are values $\hat{a}_j$ minimizing

$$\sum_{t=p+1}^{n+M-1}\left\{X_t - \sum_{j=1}^{p}\left[a_j + b_j(y - y_0)\right]X_{t-j}\right\}K_h(y).$$

4.4 Empirical comparison of forecast methods

Harvill and Ray (2005) conducted a small empirical study using a non-linear, univariate STAR model and a linear vector autoregressive (VAR) model of order 2 for comparison of the three forecast methods. Consider m-step ahead forecasts for $m = 1, \ldots, M = 7$, using series length $n = 250$. For each model, 500 replications were run and the number of bootstrap replications for the bootstrap method was 400. Harvill and Ray (2005) found that the bias increased with m for all three methods. For the bootstrap method, the bias was larger at $m = 1$, but smaller at $m = M$. The bias for the multistage method was less than the bias for the plug-in method. They also found the root meas square prediction error (MSPE) increase with m. For $m = 4, 5$, the plug-in method had smallest and the multistage method had largest root MSPE. At other values of m, the three methods were essentially equivalent.

5. Discussion and Future Work

We have discussed an application of the SBK and SBLL methods to an FCAR model and explored three methods for forecasting that can be used in tandem with the SBK and SBLL methods. Future research will include developing asymptotic results and inferential procedures using SBK and SBLL estimates of the FCAR model. We will also look to extend these methods to vector time series and to estimating covariance structure in spatial data. The preliminary results in this paper show that these methods will be useful in forecasting solar irradiance which will be beneficial to planning utility-scale PV plants. Models that take into account cloud cover will be explored next and fitted using the SBK and SBLL methods. Future simulations will be conducted using a wider array of FCAR models to compare with the results of the EXPAR model in this paper.

REFERENCES

Bosq, D. (1998), *Nonparametric statistics for stochastic processes: estimation and prediction*, 2nd ed., Berlin: Springer-Verlag.

Cai, Z., Fan, J., and Yao, Q. (2000), “Functional-coefficient regression models for nonlinear time series” *J. Amer. Statist. Assoc.*, 95, 941–956.

Chen, R. (1996), “A nonparametric multi-step prediction estimator in Markovian structures” *Statistica Sinica*, 63, 603–615.

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

Harvill, J. and Ray, B. (2005), “A note on multi-step forecasting with functional coefficient autoregressive models” *International Journal of Forecasting*, 21, 717–727.

Härdle, W., Lütkepohl, K., and Chen, R. (1997), “A Review of nonparametric time series analysis” *Int. Statist. Rev.*, 65, 49–72.

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

Huang, J. and Yang, L. (2004), “Identification of nonlinear additive autoregressive models” *J. of the Royal Statist. Soc. Series B*, 66, 463–477.

Linton, O. (1997), “Efficient estimation of additive nonparametric regression models” *Biometricka*, 84, 469–473.

Tong, H. (1990), *Nonlinear time series: a dynamical system approach*, Oxford, UK: Oxford University Press.

Wang, J., and Yang, L. (2009), “Efficient and fast spline-backfitted kernel smoothing of additive regression model” *Annals of the Institute of Statistical Mathematics*, 61, 663–690.

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