2004 IEEE Intelligent Transportation Systems Conference
Washington, D.C., USA, October 3-6, 2004

Interval Prediction for Traffic Time series Using Local Linear Predictor

Hongyu Sun, Chunming Zhang, and Bin Ran

Abstract—**This paper addresses the issue of the interval forecasting (constructing prediction intervals for future observations) of the traffic data time series using one of local polynomial nonparametric models – the local linear predictor.**

Two methods are proposed and compared. One is based on the theoretical formulation of the asymptotic prediction intervals and another is an empirical procedure using bootstrap, both for the local linear predictor. Finally, a case study using real-world traffic data will be presented for both approaches, along with the results compared with each other. The results coincide with expectations and have validated the proposed methods.

I. Introduction

TRAFFIC forecasting systems could be improved significantly by the development of interval prediction at a given probability (confidence level) in addition to point prediction. Not only can the provision of interval prediction increase the user comfort by reducing error risk associated with the information, it can also be used to assess the predictor forwards (not afterwards) for model selection and preemption in an adaptive or cooperative setting. A reduction in prediction variance (that is, smaller prediction interval) can increase the transportation system reliability and quality by reducing travelers' uncertainty thus increasing their comfort.

Short-term prediction by means of regression consists of two steps[1]: The first step uses training data (historical and/or real-time) to approximate the conditional mean regression function between inputs (data at one or more time instants) and the output (data at future time instants, with respect to the input data). Once this function is established, the second step uses it to estimate future data relative to current inputs. Usually, the value of the conditional mean function for the given input data is used as the point prediction. Prediction intervals often can be constructed if the prediction of mean, variance and bias can be calculated. The derivatives of the mean function can be used to obtain bias.

The regression model can be parametric (such as linear regression, AR model, Kalman filter, multifractal) or nonparametric (such as neural network, local polynomial, wavelet, chaos). It is known that local polynomial models are advantageous over other nonparametric models such as neural networks, in that they have explicit and solid math formulation, real-time computing, and potential for adaptive and parallel implementation. Such properties are usually only available for parametric models which, however, due to their requirement for assumptions normally unavailable in real-world data, tend to be eclipsed by nonparametric models due to their lower accuracy.

Estimating bias, variance and prediction interval is not new for most parametric models. This is especially true for the linear regression predictor and for Gaussian data for which strategies have been well established. Although such measurements are in increasing demand in transportation application, the usually high complexity of nonparametric models has led to little research in this area.

All these results can be derived for the local polynomial regression by extending the results from global models to local models. For example, the prediction interval in linear models can be applied to local linear models [2]. A closed-form expression of asymptotic estimation for prediction intervals will be elaborated in this paper for the local linear traffic predictor proposed in [1].

On the other hand, such derived expressions based on certain assumptions may not perform very well for real-world traffic data. For example, the mean function for traffic prediction is usually not smooth enough to have the second derivative. In that case, the bias cannot be estimated using the equations. Also, the residual error distribution for the small sample is often unknown instead of being assumed as normal or t-distribution. In this context, the more general bootstrap method may be proposed.

Bootstrap is a simple resampling procedure which generates samples by randomly resampling the original training set with replacements [3-5]. The idea borrows the spirit similar to Monte Carlo simulation. This paper will

Manuscript received February 29, 2004.

H. Sun is with the University of Wisconsin at Madison, 1415 Engineering Drive, Madison, WI 53706 USA. (phone: 608-294-9312; e-mail: sun@cae.wisc.edu).

C. Zhang is with the Statistics Department, University of Wisconsin at Madison, Madison, WI 53706 USA. (e-mail: cmzhang@stat. wisc.edu).

B. Ran is with the Civil Engineering Department, University of Wisconsin at Madison, Madison, WI 53706 USA. (e-mail: bran@wisc.edu).

propose a method of applying bootstrap to time series prediction when implemented by the local linear model.

While it could be quite straightforward to compute the well-established prediction intervals for parametric traffic prediction models such as in [6], few concrete research results on interval prediction seem to be presented so far for nonparametric traffic predictors. One related literature found so far, is a presentation by Rilett [7] in which the LOESS (a software program for smoothing multivariate scattered data by locally weighted least square criteria) was used to analyze historical traffic patterns such as mean, variance and confidence limit. A recently published paper by Lint [8] explored two methods to assign confidence intervals to the outcome of a neural network based freeway travel time predictor, both using bootstrap methods. However, confidence intervals are not prediction intervals and no explicit formulas for confidence intervals or prediction intervals for that model are available. As neural network models do not have rigorous and explicit mathematical formulation available. It is noted that another bootstrap procedure that is plausibly similar for prediction interval was described in [9]. However, the prediction interval in that paper refers to the interval of prediction error which was calculated after the observations in order to verify the accuracy of the forecast. Instead, the prediction interval in this paper is constructed before observations arrive. Therefore, it seems that our paper is the first presented work that addresses the prediction intervals for nonparametric traffic prediction both theoretically and empirically.

This article is organized as follows. Section II and III will describe the closed-form equation and the bootstrap scheme of interval prediction for the local linear regression model. Section IV will be devoted to numerical study. Discussion and future research directions will be provided in Section V.

II. Asymptotic Prediction Interval

This section will, first, briefly review the mean prediction of the local linear model and address the variance prediction. Then the estimator bias and variance will be introduced. Finally, interval prediction will be derived.

Given the observations $\{(\mathbf{X}_i^T, Y_i)$: $i = 1, \ldots, n\}$ of the multivariate covariate $\mathbf{X}$ and a univariate response Y, the relationship between $\mathbf{X}$ and Y can be modeled as:

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon, \quad (1)$$

where $\mathbf{X}$ and ε are not necessarily independent, ε is the additive error term with $E(\varepsilon|\mathbf{X}) = 0$ (2)

and $\mathrm{Var}(\varepsilon|\mathbf{X}) = 1$. (3)

Here n is the number of the observations,

$$\mathbf{X} = (X_1, \ldots, X_d)^T \quad (4)$$

and $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ (5)

with d the dimension of $\mathbf{X}$.

It is of interest to estimate the mean regression function

$$m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x}) \quad (6)$$

and the possible heteroscedasticity (conditional variance function) $\sigma^2(\mathbf{x}) = \mathrm{Var}(Y|\mathbf{X} = \mathbf{x})$, (7)

where $\mathbf{x}^T = (x_1, .., x_d)$ (8)

is a point in $\mathscr{R}^d$.

Once the estimated mean regression function (denoted as $\hat{m}(\ )$) is obtained, the fitted regression is used as a mechanism for prediction of response values. That is, if the prediction of Y at $\mathbf{X} = \mathbf{x}$ is denoted as $\hat{y}(\mathbf{x})$, then

$$\hat{y}(\mathbf{x}) = \hat{m}(\mathbf{x}). \quad (9)$$

A. Mean Prediction

A local polynomial model is formed much as a Taylor series model, a function in the neighborhood of a query point x. In the local linear model, the Taylor expansion terms up to the first (linear) order are used to make the local approximation. That is, the function m is estimated $m(\mathbf{X}) \approx m(\mathbf{x}) + \mathbf{g}^T(\mathbf{X} - \mathbf{x})$, (10)

where $\mathbf{g} = (\beta_1, \ldots, \beta_d)^T$. For the convenience of a matrix expression, redefine the vectors taking into account the constant term. Write $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)^T$, (11)

where $\beta_0 = m(\mathbf{x})$, (12)

and $\tilde{\mathbf{X}} = (1, (\mathbf{X} - \mathbf{x})^T)^T$, (13)

Then $m(\mathbf{X}) \approx \boldsymbol{\beta}\tilde{\mathbf{X}}$ (14)

The observations $\{(\mathbf{X}_i^T, Y_i)$: $i = 1, \ldots, n\}$ are used as training data to estimate $\boldsymbol{\beta}$. The weighted least square criterion is used to obtain the fit [10].

The estimation of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}_d\boldsymbol{\beta})^T \mathbf{W} (\mathbf{y} - \mathbf{X}_d\boldsymbol{\beta}) \quad (15)$$

is $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_d)^T = (\mathbf{X}_d^T\mathbf{W}\mathbf{X}_d)^{-1}\mathbf{X}_d^T\mathbf{W}\mathbf{y}$, (16)

where

$$\mathbf{X}_d = \begin{pmatrix} 1 & X_{11} - x_1 & \cdots & X_{1d} - x_d \\ 1 & X_{21} - x_1 & \cdots & X_{2d} - x_d \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} - x_1 & \cdots & X_{nd} - x_d \end{pmatrix}, \quad (17)$$

$$\mathbf{W} = \mathrm{diag}\{K_B(\mathbf{X}_i - \mathbf{x})\}, \quad (18)$$

$$\mathbf{y} = (Y_1, \ldots, Y_n)^T \quad (19)$$

and $K_B(\mathbf{u}) = \frac{1}{|B|}K(B^{-1}\mathbf{u})$, (20)

where $K(\bullet)$ is a multivariate probability density function (weighting kernel function) with mean zero and the covariance matrix of $\mu_2(K)\mathbf{I}_d$, with $\mathbf{I}_d$ the $d \times d$ identity matrix. $\boldsymbol{B}$ is called bandwidth matrix and $|\boldsymbol{B}|$ denotes its determinant. In this study, K is chosen as a Gaussian function and $\boldsymbol{B} = h\mathbf{I}_d$, where h is called bandwidth. That is,

$K(\mathbf{u}) = e^{-(dis(\mathbf{u}))^2}$ (21)

The distance function $dis(\bullet)$ used in this study is the Euclidean distance. $dis(\mathbf{u}) = \sqrt{\mathbf{u}^T\mathbf{u}}$ (22)

Thus $\hat{m}(\mathbf{x}) = \hat{\beta}_0$, (23)

and $(\frac{\partial \hat{m}}{\partial x_j})(\mathbf{x}) = \hat{\beta}_j$, $j = 1, \ldots, d$. (24)

The prediction value $\hat{y}$ is equal to $\hat{\beta}_0$. That is,

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 = \mathbf{p_x}^T\mathbf{y} = \sum_{i=1}^{n} p_i(\mathbf{x})Y_i \quad (25)$$

where $\mathbf{q} = (1, 0, \ldots, 0)^T$. (26)

The vector $\mathbf{p_x}$, also written as $\mathbf{p}(\mathbf{x})$, will be useful for calculating the bias and variance of the local model.

$$\mathbf{p_x} = \mathbf{p}(\mathbf{x}) = \left(\mathbf{q}^T\left(\mathbf{X}_d^T\mathbf{W}\mathbf{X}_d\right)^{-1}\mathbf{X}_d^T\mathbf{W}\right)^T \quad (27)$$

It is easy to find that

$$\mathbf{p}_\mathbf{x}^T\mathbf{p_x} = \mathbf{q}^T\left(\mathbf{X}_d^T\mathbf{W}\mathbf{X}_d\right)^{-1}\mathbf{q} \quad (28)$$

and $\Sigma p_i(\mathbf{x}) = 1$. (29)

B. Variance Prediction

[11] derived an estimate of $\sigma^2(\mathbf{x})$, the local noise variance, that is, the variance for the traffic data in our study. A local linear regression produces residuals at all training points. The weighted residual $r_i(\mathbf{x})$ is given by:

$r_i(\mathbf{x}) = w_i(\mathbf{x})\mathbf{X}_i^T\beta(\mathbf{x}) - w_i(\mathbf{x})Y_i$ (30)

where $w_i(\mathbf{x}) = \sqrt{K(d(\mathbf{X}_i, \mathbf{x}))}$. (31)

The training criteria is to minimize the weighted sum of the squared errors C(x):

$$C(\mathbf{x}) = \sum_i r_i^2(\mathbf{x}) = (\mathbf{y} - \mathbf{X}_d\boldsymbol{\beta})^T\mathbf{W}(\mathbf{y} - \mathbf{X}_d\boldsymbol{\beta}) \quad (32)$$

A reasonable estimator for the local value of the noise variance is $\hat{\sigma}^2(\mathbf{x}) = C(\mathbf{x})/n_{LL}(\mathbf{x})$ (33)

where n_{LL} is a measure of the number of data points:

$$n_{LL}(\mathbf{x}) = \sum_{i=1}^{n} w_i^2 = \sum_{i=1}^{n} K\left(\frac{d(X_i, \mathbf{x})}{h}\right) \quad (34)$$

The bias of the estimate $\hat{\sigma}^2(\mathbf{x})$ can be reduced by taking into account the number of parameters in the local linear regression:

$s^2(\mathbf{x}) = (\sum r_i^2(\mathbf{x}))/(n_{LL}(\mathbf{x}) - p_{LL}(\mathbf{x}))$ (35)

where $p_{LL}(\mathbf{x})$ is a measure of the local number of the free parameters in the local model:

$$p_{LL}(\mathbf{x}) = \sum_i w_i^4 X_i^T(\mathbf{X}_d^T\mathbf{W}\mathbf{X}_d)^{-1}X_i \quad (36)$$

C. Estimator Variance and Bias

For any estimator $\hat{m}(\mathbf{x})$, define

$\text{Bias}(\hat{m}(\mathbf{x})) = E\{\hat{m}(\mathbf{x}) \mid \mathbf{X} = \mathbf{x}\} - m(\mathbf{x})$ (37)

$\text{Var}(\hat{m}(\mathbf{x})) = \text{Var}\{\hat{m}(\mathbf{x}) \mid \mathbf{X} = \mathbf{x}\}$ (38)

Mean Squared Error $\text{MSE}(\mathbf{x}) = \text{Bias}^2(\mathbf{x}) + \text{Var}(\mathbf{x})$ (39)

In order to develop the intervals, the parameter $\text{Var}(\hat{y}(\mathbf{x}))$ must be determined. A standard error $s_{\hat{y}(\mathbf{x})}$ of $\hat{y}(\mathbf{x})$ can be interpreted as the standard error of the estimator of mean response, conditional on **x**. The notion standard error evokes the image of precision or variation. In this case, it reflects the variation of $\hat{y}$ at **x**, if repeated regressions were conducted, based on the same **X**-levels and new observations on Y each time.

The variance and bias of the multivariate local linear estimator are shown below as given by Atkeson et al. [11].

$E(\hat{y}(\mathbf{x})) = m(\mathbf{x}) + \mathbf{p_x}^T(\mathbf{m} - \mathbf{X}_d\boldsymbol{\beta}) = m(\mathbf{x}) + \mathbf{p_x}^T\mathbf{t}$ (40)

where $\mathbf{m} = [m(\mathbf{X}_1), \ldots, m(\mathbf{X}_n)]^T$ (41)

and $\mathbf{t} = \mathbf{m} - \mathbf{X}_d\boldsymbol{\beta}$. (42)

$\text{Var}(\hat{y}(\mathbf{x})) = \text{Var}(\hat{m}(\mathbf{x})) = \sigma^2(\mathbf{x})\,\mathbf{p_x}^T\mathbf{p_x}$ (43)

From (40), it is easy to get

$\text{Bias}(\hat{y}(\mathbf{x})) = \mathbf{p_x}^T(\mathbf{m} - \mathbf{X}_d\boldsymbol{\beta})$ (44)

If the estimator $\sigma^2(\mathbf{x})$ is substituted by $s^2(\mathbf{x})$, from (43) the standard error of prediction can be defined as

$$s_{\hat{y}(\mathbf{x})} = s(\mathbf{x})\sqrt{\mathbf{p}_\mathbf{x}^T\mathbf{p_x}} \quad (45)$$

Using Taylor's expansion of $m(\mathbf{X})$ [10],

$m(\mathbf{X}) = \mathbf{X}_d\boldsymbol{\beta}$ + higher terms of $(\mathbf{X} - \mathbf{x})$ (46)

Denote $\tau_i = m(\mathbf{X}_i) - \mathbf{X}_d\boldsymbol{\beta}$ = higher terms of $(\mathbf{X}_i - \mathbf{x})$ and $\boldsymbol{\tau} = [\tau_1\ \tau_2 \ldots \tau_n]^T$, the estimation of $\text{Bias}(\hat{y}(\mathbf{x}))$ is

$$\widehat{Bias}(\hat{y}(\mathbf{x})) = \mathbf{p_x}^T\boldsymbol{\tau} \quad (47)$$

D. Asymptotic Prediction Interval

To derive confidence intervals requires the distribution of the error. Here, the error is assumed normal, $\varepsilon \sim N(0,1)$. From (1), $Y \sim N(m(\mathbf{x}), \sigma(\mathbf{x}))$. If $m(\mathbf{x})$ is linear in **x**, the local linear estimator is unbiased [11]. That is,

$\text{Bias}(\hat{y}(\mathbf{x})) = E(\hat{y}(\mathbf{x})) - y_{true}(\mathbf{x}) = 0$. (48)

The following section will first give the derivation of the prediction interval for the unbiased case and then will discuss the biased case. Under the condition of normal errors, $\hat{y}(\mathbf{x})$ is normal, and a confidence interval at the $100(1-\alpha)\%$ confidence level for $E(Y \mid \mathbf{x})$ can be written

$$\hat{y}(\mathbf{x}) \pm t_{\alpha/2, n-2}\, s_{\hat{y}(\mathbf{x})} \quad (49)$$

The expression in (49) is, indeed, that of a confidence interval and is not to be confused with the prediction level on a new response observation at $\mathbf{X} = \mathbf{x}$. The latter reflects bounds in which the analysts can realistically expect an observation of y at $\mathbf{X} = \mathbf{x}$ to fall.

The standard error of prediction, given by (45), is used in constructing a confidence interval on the mean response. However, it is not appropriate for establishing any form of inference on a future single observation. Suppose the mean response at a fixed $\mathbf{X} = \mathbf{x}$ is not of interest. Rather, one is interested in some type of bound on a single response observation at $\mathbf{x}$. Consider a single observation at $\mathbf{X} = \mathbf{x}$ denoted symbolically by $y_{new}(\mathbf{x})$, independently of $\hat{y}(\mathbf{x})$. A prediction interval on y can be constructed by beginning with $y_{new}(\mathbf{x}) - \hat{y}(\mathbf{x})$.

Note that Var $(y_{new}(\mathbf{x}) - \hat{y}(\mathbf{x})) = \sigma^2(\mathbf{x})\left(1 + \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}\right)$. This reflects both the additive noise in sampling at the new point ($\sigma^2(\mathbf{x})$) and the prediction error of the estimator ($\sigma^2(\mathbf{x})\, \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}$).

Under the assumption (48),

$E[y_{new}(\mathbf{x}) - \hat{y}(\mathbf{x})] = m(\mathbf{x}) - E(\hat{m}(\mathbf{x})) = -\text{Bias}(\hat{y}(\mathbf{x})) = 0$,

then, $(y_{new}(\mathbf{x}) - \hat{y}(\mathbf{x})) / (\sigma(\mathbf{x})\sqrt{1 + \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}}) \sim N(0,1)$ under the normal theory assumptions. In many nonparametric regression situations, there may be only a few local data points in the neighborhood of $\mathbf{x}$ and the asymptotic normality is not accurate enough. Therefore [11] replace σ by s and $(y_{new}(\mathbf{x}) - \hat{y}(\mathbf{x})) / (s(\mathbf{x})\sqrt{1 + \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}}) \sim t_{n-2}$ (50)

From (50) a probability bound or prediction interval can be placed on y_{new}, i.e., an interval in which y_{new} is contained with a fixed probability ($1-\alpha$).

This prediction interval is given by

$$\hat{y}(\mathbf{x}) \pm t_{\alpha/2, n-2}\, s(\mathbf{x})\sqrt{1 + \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}}$$

This expression of the prediction intervals is independent of the output values of the training data Y_i, and reflects how well the data is distributed in the input space (see (27)).

When the bias is not zero, however, the variance only reflects the difference between the prediction and the mean prediction, and not the difference between the prediction and the true value, which requires knowledge of the predictor's bias. Only when the local model structure is correct will the bias be zero.

Under certain regularity conditions, extending the univariate case [10], it can be shown that asymptotically

$$(\hat{y}(\mathbf{x}) - \widehat{Bias}(\hat{y}(\mathbf{x}))) / (s_{\hat{y}(\mathbf{x})}) \rightarrow N(0,1)$$

Therefore, the prediction interval can be estimated as

$$\hat{y}(\mathbf{x}) - \widehat{Bias}(\hat{y}(\mathbf{x})) \pm t_{\alpha/2, n-2}\, s(\mathbf{x})\sqrt{1 + \mathbf{p}_{\mathbf{x}}^T \mathbf{p}_{\mathbf{x}}} \quad (51)$$

Considering the t-distribution may not be valid for our data, the bootstrap method is proposed.

III. Bootstrap Prediction Interval

The bootstrap is a method for estimating the distribution of an estimator or a test statistic by resampling one's data or a model estimated from the data. The bootstrap principle is that the distribution of (resampled – sample), which can be computed directly from data, approximates the distribution of (sample – true). Often, the bootstrap provides approximations that are more accurate than those of the first-order asymptotic theory.

Bootstrap is a popular method despite its disadvantage of being time consuming. In terms of obtaining prediction intervals, it could be applied to many prediction models and needs few assumptions. Bootstrap can provide a reliable solution and it is easy to implement when the asymptotic equations are not available or not valid. This may occur due to a small sample size, or the limitations set by the problem characteristics such as smoothness of the mean function.

From an original sample $\Psi_n = (Y_1, Y_2, \ldots, Y_n) \overset{iid}{\sim} F$, draw a new sample of n observations among the original sample with replacements, each observation having the same probability of being drawn ($= 1/n$). A bootstrap sample is often denoted $\Psi_n^* = (Y_1^*, Y_2^*, \ldots, Y_n^*) \overset{iid}{\sim} F_n$ where F_n is the empirical distribution. The behavior of a random variable $\hat{\theta} = \theta(\Psi_n, F)$ can be studied by considering B new values obtained through computation of B bootstrap samples. An approximation of the distribution of the estimate $\hat{\theta} = \theta(\Psi_n, F)$ is provided by the distribution of $\hat{\theta}^{*b} = \theta(\Psi_n^{*b}, F_n)$, $b = 1, \ldots, B$.

A. Bootstrap Bias

In general, let θ be a parameter and $\hat{\theta}$ an estimate. Let $\hat{\theta}^*$ be the bootstrap estimate calculated in the same way as $\hat{\theta}$. Then the bootstrap assessment of the bias is

Bias = Mean of $(\hat{\theta}^*)$ - $\hat{\theta}$. The bias-corrected estimate of θ is then $\bar{\theta} = \hat{\theta}$ - Bias = $2\hat{\theta}$ - Mean of $(\hat{\theta}^*)$

B. Bootstrap Prediction Interval

A residual-based bootstrap with bias correction is proposed to compute the prediction interval based on the percentile method [3].

Denote the bootstrap distribution of $\hat{\theta}^*$ by $G_n^*(t) = P_{F_n}(\hat{\theta}^* \le t)$, approximated by $\hat{G}_n^*(t) = \#\{\hat{\theta}^* \le t\} / B$

The percentile method takes the $1-2\alpha$ confidence interval for θ as being $[\hat{G}_n^{*-1}(\alpha), \hat{G}_n^{*-1}(1-\alpha)]$. Theoretically this is equivalent to the replacement of the unknown distribution

$G(t,F) = P_{F_n}(\hat{\theta}^* \le t)$ by the estimate $G(t, F_n)$.

The bootstrap interval prediction procedure can be divided into three steps:

1. Given training data $\{(\mathbf{X}_i^T, Y_i)$: $i = 1, \ldots, n\}$ of size n (n = 14 for our case study), fit the local linear model $m(\mathbf{X})$ and calculate the corresponding residuals $\hat{\varepsilon}_i = Y_i - \hat{y}_i = Y_i - \hat{m}(\mathbf{X}_i)$, $i = 1, \ldots, n$. Since Eε =0 and Var ε = 1 are assumed by our model ((2) and (3)), $\hat{\varepsilon}_i$ needs to be divided by the square root of Var($\hat{y}(\mathbf{x})$) (45) before standardization to avoid a system error in the bootstrap. The standardization includes centering by subtracting the average [12], so $\tilde{\varepsilon}_i$ = $\hat{\varepsilon}_i - \frac{1}{n}\sum_{k=1}^{k}\hat{\varepsilon}_k$, $k = 1, \ldots, n$.

2. Then, draw B bootstrap errors $\{\varepsilon_i^*(b)$, $i = 1, \ldots, n$; $b = 1, \ldots, B\}$ each of size n with replacement from the sample distribution given by the centered residuals. Finally B bootstrap outputs are formed as $Y_i^*(b) = \hat{y}_i + \varepsilon_i^*(b)$ to get B bootstrap training datasets (X_i^T, $Y_i^*(b)$), $i = 1, \ldots, n$; $b = 1, \ldots, B$. To each bootstrap dataset a local linear model is fitted as $\hat{m}^{(b)}(\mathbf{X})$ and the prediction $\hat{m}^{(b)}(\boldsymbol{x})$ for the testing data query point $\boldsymbol{x}$ is computed. Bias could be estimated by using average of $\hat{m}^{(b)}(\boldsymbol{x})$. A bias corrected prediction $\hat{y}^{(b)}(\boldsymbol{x}) = 2\,\hat{m}^{(b)}(\boldsymbol{x}) - (1/B)\sum_{b=1}^{B}\hat{m}^{(b)}(x)$.

3. The prediction interval for $\hat{y}(\boldsymbol{x})$ with the confidence level of 100(1- α) percent is obtained as [$\hat{y}(\boldsymbol{x})^{*(\psi)}$, $\hat{y}(\boldsymbol{x})^{*(1-\psi)}$], where $\hat{y}(\boldsymbol{x})^{*(\psi)}$ is the 100ψ-th percentile of the bootstrap distribution $\{\hat{y}^{(b)}(\mathbf{x})\}(b = 1, \ldots, B)$ and $\psi = 0.5\alpha$.

IV. Numerical Study

The detailed process for obtaining and preparing the data, the preliminary data analysis and experimental design is basically the same as in the study in [1] for the point prediction performance of the local linear predictor. the selected road segment for study is US-290 from the cross street Sam Houston toll way to the cross street Fairbanks based on Houston's US-290 Northwest freeway eastbound traffic time(speed) data collected from February 2002 to July 2002 every five minutes. The differences between this numerical study from that study are pointed out as below.

First is the performance index or evaluation criterion. Since the main concern of the interval prediction is the predicted bounds instead of the predicted value, therefore, the relative mean error (RME) won't be used as the performance index.

The expected result is that the predicted bounds should include the predicted values and the bootstrap method is better in terms of giving narrower prediction intervals. Thus, the predicted values should fall within the intervals formed by predicted upper bounds and lower bounds. Provided this premise is verified, the narrower prediction intervals give better performance. Also, when comparing

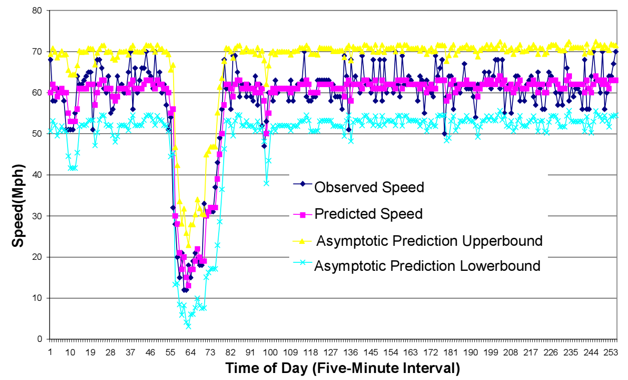

Fig. 1. 95% Prediction Upper and Lower Bounds for One-Day Traffic Time Series, Computed by Asymptotic Equations of the Local Linear Predictor.

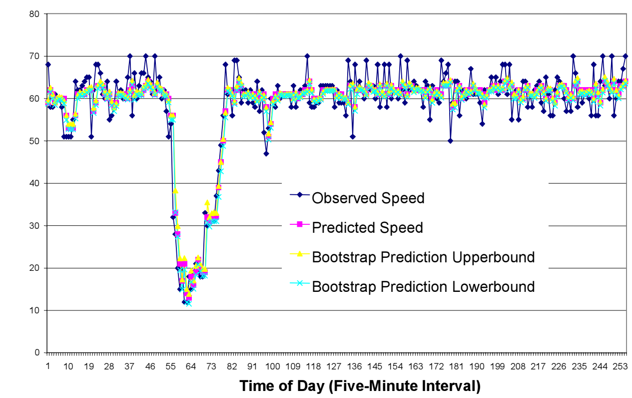

Fig. 2. 95% Prediction Upper and Lower Bounds for One-Day Traffic Time Series, Computed by the Bootstrap Procedure for the Local Linear Predictor.

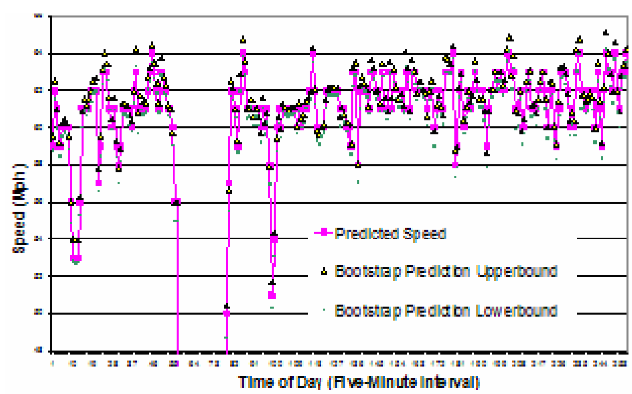

Fig. 3. Zoom-In Results for 95% Prediction Upper and Lower Bounds for One-Day Traffic Time Series, Computed by the Bootstrap Procedure for the Local Linear Predictor.

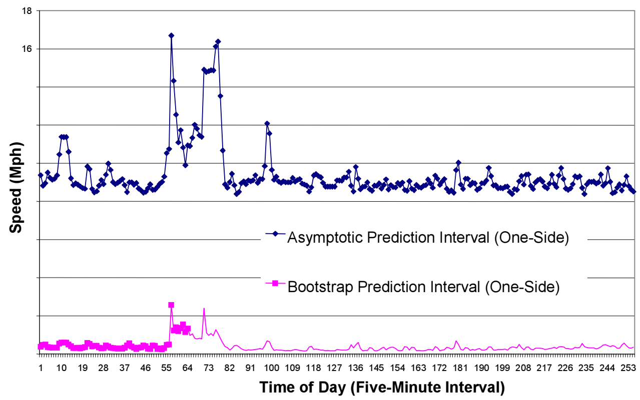

Fig. 4. Comparison of One-Sided Prediction Intervals at 95% Confidence Level for One-Day Traffic Time Series, Computed by Asymptotic Equations and the Bootstrap Procedure for the Local Linear Predictor, Respectively.

different methods of interval prediction, results over multiple experimental runs may not require to be averaged. If consistent results for each run are observed then one-run results may be enough to infer comparison conclusions. In this paper, one-day results are given to illustrate the comparison since all other runs show the same trends.

Secondly, results using 16-day data instead of 32-day data as the data set are given. Since a smaller data set has less total computation time and it is found that the two data sets give the same comparison results for the two interval prediction approaches.

A resampling size B of 500 is used in the bootstrap procedure. Fig. 1 to 3 show one-day prediction upper-bounds and lower-bounds at a confidence level of 95% computed by asymptotic equations and the bootstrap method, respectively. Fig. 3 is a zoom-in view of Fig. 2 for a clearer display. Fig. 4 compares the one-sided prediction intervals obtained by these two methods.

It is observed that the prediction intervals given by both methods can include the prediction values. This is self-evident for the asymptotic method since the bounds are derived after computing one-sided intervals. So the results validate the proposed bootstrap procedure using the percentile method. On the other hand, the similarity of the trends of both results validates the asymptotic equations.

The average and maximum one-sided prediction interval are approximately $\pm$10Mph and $\pm$16Mph for the asymptotic equation approach, $\pm$0.3Mph and $\pm$2.5Mph for the bootstrap approach. The latter is much smaller than the former thus the bootstrap is better than that of the asymptotic results. This is in accordance with the expected outcome of the experimental design.

From Fig. 4, it is shown that both approaches have a larger interval when entering and leaving peak hours than at other times. But the bootstrap method gives much more stable intervals than the asymptotic method. This also proves the bootstrap is advantageous.

V. CONCLUSIONS

Pioneering work from theoretical methodologies to implementations is presented on achieving more informational traffic forecasting by providing interval predictions. Both the asymptotic equations and an empirical bootstrap method were derived for the local linear interval prediction. The experimental results using a set of real-world data were given and have validated the proposed methodologies. The case study results are consistent with what are expected. That is, both methods are valid and the bootstrap method gives better results.

REFERENCES

[1] Sun, H., Liu, H., Xiao, H., He, R. and Ran, B. "Short-term traffic forecasting using the local linear regression model". *Transportation Research Record 1836*, Journal of the Transportation Research Board, pp143~150, 2003.

[2] Schaal, S., & Atkeson, C. G. "Assessing the quality of learned local models". In J. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in Neural Information Processing Systems* 6 (pp. 160-167). San Mateo, CA: Morgan Kaufmann, 1994.

[3] Enfron, B. and Tibshirani, R. *An introduction to the bootstrap*. Chapman and Hall, 1993.

[4] Freedman, D., Peters, S. "Bootstrapping a regression equation: some empirical results (in theory and methods) ". *Journal of the American Statistical Association*, Vol. 79, No. 385. (Mar., 1984), pp. 97-106.

[5] Peters, S. and Freedman, D. "Better bootstrap confidence intervals: comment (in theory and methods)". *Journal of the American Statistical Association*, Vol. 82, No. 397. (Mar., 1987), pp. 186-187.

[6] Yang, F., Yin, Z., Liu, H., and Ran, B. "Online recursive short-term traffic prediction algorithm". *Transportation Research Record*, Journal of the Transportation Research Board, to be published 2004.

[7] Rilett, L. & Eisele, B. "Investigation of travel time reliability estimates for ITS data". Presentation in May 2001, Texas A&M University and Texas Transportation Institute.

[8] Lint, van, J.W.C. "Confidence intervals for real time freeway travel time prediction". Intelligent Transportation Systems, Proceeding. Volume 2, pp 1353~1458, 12-15 Oct. 2003, Shanghai, China.

[9] Cho., H and Rilett, L. "Forecasting train travel time". TRB 2003 Annual Meeting CD-ROM.

[10] Fan, J. and I. Gijbels. *Local polynomial modelling and its applications*. London: Chapman & Hall, 1996.

[11] Atkeson, C., Moore, A. and Schaal, S. "Locally weighted learning". Artificial Intelligence Review, 11, 1-5, pp.11-73, 1997.

[12] Freedman, D.A. "Bootstrapping regression models". Annals of Statistics, 9, 1218-1228, 1981.