



Title	Local polynomial modeling and bandwidth selection for time-varying linear models
Author(s)	Chan, SC; Zhang, ZG
Citation	The 7th International Conference on Information, Communications and Signal Processing (ICICS 2009), Macau, China, 8-10 December 2009. In Proceedings of the International Conference on Information, Communications and Signal Processing, 2009, p. 1-5
Issued Date	2009
URL	http://hdl.handle.net/10722/126154
Rights	International Conference on Information, Communications and Signal Processing. Copyright © IEEE.

Local Polynomial Modeling and Bandwidth Selection for Time-Varying Linear Models

S. C. Chan and Z. G. Zhang

Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
{scchan, zgzhang}@eee.hku.hk

Abstract—This paper proposes a local polynomial modeling approach and bandwidth selection algorithm for estimating time-varying linear models (TVLM). The time-varying coefficients of a TVLM are modeled locally by polynomials and estimated using least-squares estimation with a kernel having a certain bandwidth or support. Asymptotic behavior of the proposed estimator is established and it shows that there exists an optimal local bandwidth which minimizes the weighted mean squared error (MSE). A data-driven variable bandwidth selection method is also proposed to estimate this optimal bandwidth. Simulation results show that the proposed LPM method with adaptive bandwidth selection outperforms conventional TVLM identification methods in a large variety of testing conditions.

Index Terms—local polynomial modeling, time-varying linear model, least-squares, bandwidth selection

I. INTRODUCTION

The time-varying linear model (TVLM) is a widely used model to characterize a dynamic system:

$$Y(t) = \sum_{k=1}^L a(k,t)X(k,t) + \sigma(t)\varepsilon(t), \quad (1)$$

where $Y(t)$, $X(k,t)$ and $a(k,t)$, are respectively the system output, input and time-varying response, which are assumed to be functions of time variable t , L is the order of the model, $\varepsilon(t)$ is a zero mean white Gaussian process with unit variance, and the conditional variance of additive noise given $t=t_0$ is $\sigma^2(t_0)$. In the general case of time-varying linear regression model, $Y(t)$, $X(k,t)$ and $a(k,t)$ are respectively the observations, explanatory variables and time-varying regression coefficients. Numerous methods have been proposed to identify the TVLM of (1), i.e., to estimate the time-varying coefficients $a(k,t)$, and they can be generally divided into three categories [1]: adaptive filtering/Kalman filtering (AF/KF), basis functions (BF), and weighted least-squares (WLS) methods. Adaptive filtering methods, such as least mean squares (LMS) and recursive least-squares (RLS), estimate the coefficients recursively to meet a performance requirement. Most AF methods make use of past data samples for estimation and the converging speed is therefore limited [2]. The Kalman filtering method is an optimal recursive estimator in a minimum mean-square error sense, given the prior information of the stochastic model of system dynamics and noise. However, such prior knowledge is often vague in practice, so the estimation accuracy of Kalman filtering is sometimes limited [1]. The BF method assumes an explicit deterministic model of the coefficient variations, and it approximates the time-varying coefficients by a linear combination of known basis functions of time. The performance of the BF method is greatly dependent on the

selection of basis functions, which is not always accessible. The WLS method employs kernels or windows to assign larger weights to local data and smaller weights to remote data, and the time-varying coefficients are estimated by minimizing a weighted sum of squared estimation errors. The window size or kernel bandwidth has to be carefully chosen to compromise between estimation accuracy (variance) and modeling error (bias). However, automatic data-driven bandwidth selection for WLS is very difficult. Consequently, WLS is infrequently used in practice [1].

In this paper, a local polynomial modeling (LPM) approach is proposed to model and estimate the time-varying coefficients of a TVLM. More precisely, the time-varying coefficients are modeled locally by a set of polynomials with a kernel having a certain bandwidth [2]-[5]. Consequently, the estimation of time-varying coefficients is reduced to the estimation of polynomial coefficients, which can be easily performed using the least-squares (LS) technique. To establish the asymptotic behaviors of the proposed estimator, new asymptotic expressions for their bias and variance are derived. Both are functions of the kernel bandwidth and we show that there exists an optimal bandwidth which minimizes the MSE. An important advantage of the LPM method is that the kernel bandwidth can be determined locally and adaptively to minimize the MSE at each time instant. Since the optimal expressions involve quantities which are in general unknown, new approximation algorithm is proposed to estimate the bias, variance and MSE. This allows the optimal bandwidth to be determined by minimizing over the approximate MSE over a discrete bandwidth set. A novel pilot bandwidth estimation to compute the approximate MSE using an intersection of confidence intervals (ICI) method [6]-[8] is also proposed. The performance of the LPM method was evaluated using various types of simulated TVLMs, and the results showed that the proposed method can achieve more accurate estimates than conventional TVLM identification methods.

The rest of the paper is organized as follows. In Section II, the local polynomial modeling for time-varying linear models is introduced. Section III is devoted to the asymptotic analysis of the LPM estimator. The adaptive bandwidth selection method for LPM is developed in Section IV. Simulation results and comparisons to conventional methods are presented in Section V. Finally, conclusions are drawn in Section VI.

II. LOCAL POLYNOMIAL MODELING OF TVLM

The LPM method models k -th coefficient $a(k,t)$ of the TVLM of (1) at $t=t_0$ locally as a p -th order polynomial [2]-[5]:

$$a(k,t) \approx \sum_{j=0}^p \frac{1}{j!} a^{(j)}(k,t_0)(t_i - t_0)^j. \quad (2)$$

These polynomials can be estimated locally by minimizing

This study was partially supported by a grant from the Research Grants Council of the Hong Kong SAR, China.

a WLS criterion as follows:

$$\min_{\beta} \sum_{i=1}^n [Y_i - \sum_{k=0}^{L-1} \sum_{j=0}^p \beta_j(k)(t_i - t_0)^j X(k, t_i)]^2 K_h(t_i - t_0), \quad (3)$$

where $Y_i = Y(t_i)$, n is the data length, $\beta_j(k) = a^{(j)}(k, t_0) / j!$, and $K_h(t_i - t_0) = \frac{1}{h} K(\frac{1}{h}(t_i - t_0))$ is a weight function which controls the bandwidth h and hence the number of neighboring samples around t_0 used to estimate $\beta_j(k)$. The weight function or kernel $K_h(\cdot)$ is a scaled version of a basis kernel function $K(\cdot)$ by a factor of h . Note, $\beta_j(k)$ is a function of t_0 and for notation simplicity we have dropped this dependence in subsequent text. The LPM estimator can be viewed as the combination of the WLS method and the BF method with polynomial basis functions. When $p=0$, LPM is reduced to the conventional WLS method. An important advantage of the proposed LPM method is that the bias and variance can be analytically derived, which paths the way to the solution of the key problem of automatic data-driven adaptive bandwidth selection.

Eq (3) can be written more compactly in matrix form as

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W} (\mathbf{y} - \mathbf{X}\beta), \quad (4)$$

where $\mathbf{y} = [Y_1, \dots, Y_n]^T \in \mathbf{R}^n$, $\mathbf{W} = \text{diag}\{K_h(t_i - t_0)\} \in \mathbf{R}^{n \times n}$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^T & (t_1 - t_0)\mathbf{X}_1^T & \cdots & (t_1 - t_0)^p \mathbf{X}_1^T \\ \mathbf{X}_2^T & (t_2 - t_0)\mathbf{X}_2^T & \cdots & (t_2 - t_0)^p \mathbf{X}_2^T \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{X}_n^T & (t_n - t_0)\mathbf{X}_n^T & \cdots & (t_n - t_0)^p \mathbf{X}_n^T \end{pmatrix} \in \mathbf{R}^{n \times (p+1)L} \quad \text{with}$$

$$\mathbf{X}_i = [X(1, t_i), X(2, t_i), \dots, X(L, t_i)]^T, \quad \text{and}$$

$$\beta = [\beta_0^T, \dots, \beta_p^T]^T \in \mathbf{R}^{(p+1)L} \quad \text{with} \quad \beta_j = [\beta_j(1), \dots, \beta_j(L)]^T.$$

The LS solution to (4) is given by

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}. \quad (5)$$

We next derive new expressions for the asymptotic bias and variance of the LS estimator of (5). This allows us to estimate the optimal bandwidth parameter h to minimize the MSE.

III. ASYMPTOTIC ANALYSIS OF LPM ESTIMATOR

The conditional bias based on $\mathbb{X} = \{\mathbf{X}, \mathbf{y}\}$ can be obtained by taking the expectation of (5):

$$\begin{aligned} E(\hat{\beta} | \mathbb{X}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \cdot \mathbf{m} \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \cdot (\mathbf{m} - \mathbf{X}\beta) + \beta \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \cdot \mathbf{r} + \beta, \end{aligned} \quad (6)$$

where $\mathbf{m} = E(\mathbf{y} | \mathbb{X}) = [m(t_1), \dots, m(t_n)]^T$, β is the true parameter, and $\mathbf{r} = \mathbf{m} - \mathbf{X}\beta$ is the residual vector of the local polynomial approximation. Thus, the conditional bias is

$$\text{Bias}(\hat{\beta} | \mathbb{X}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{r}. \quad (7)$$

From definition, the conditional covariance of (5) is given by

$$\begin{aligned} &E(\{\hat{\beta} - E(\hat{\beta})\} \{\hat{\beta} - E(\hat{\beta})\}^T | \mathbb{X}) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} E((\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})^T | \mathbb{X}) \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \Sigma_{rr} \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \Sigma \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}, \end{aligned} \quad (8)$$

where $\Sigma = \mathbf{W} \Sigma_{rr} \mathbf{W} = \text{diag}\{K_h^2(t_i - t_0) \sigma^2(t_i)\}$ and $\Sigma_{rr} = \text{diag}\{\sigma^2(t_i)\}$.

Asymptotic expressions for the conditional bias and variance are now derived. Let

$$\mathbf{S}_n^* = \mathbf{X}^T \Sigma \mathbf{X}, \quad (9)$$

$$\mathbf{S}_n = \mathbf{X}^T \mathbf{W} \mathbf{X}. \quad (10)$$

Each element $(\mathbf{S}_n)_{jL+q, lL+m} = s_{n, j+l, q, m}$ ($0 \leq j, l \leq p$ and $1 \leq q, m \leq L$) in the matrix $\mathbf{S}_n \in \mathbb{R}^{(p+1)L \times (p+1)L}$ is

$$s_{n, k, q, m} = \sum_{i=1}^n X(q, t_i) X(m, t_i) K_h(t_i - t_0) (t_i - t_0)^k. \quad (11)$$

Similarly, each element $(\mathbf{S}_n^*)_{jL+q, lL+m} = s_{n, j+l, q, m}^*$ ($0 \leq j, l \leq p$ and $1 \leq q, m \leq L$), in the matrix $\mathbf{S}_n^* \in \mathbb{R}^{(p+1)L \times (p+1)L}$ is

$$s_{n, j+l, q, m}^* = \sum_{i=1}^n X(q, t_i) X(m, t_i) K_h^2(t_i - t_0) (t_i - t_0)^k \sigma^2(t_0). \quad (12)$$

The conditional variance is thus given by $\mathbf{S}_n^{-1} \mathbf{S}_n^* \mathbf{S}_n^{-1}$.

Since the observations are independent and identically distributed (i.i.d.), by the central limit theorem (CLT), we have

$$\begin{aligned} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X(q, t_i) X(m, t_i) K_h(t_i - t_0) (t_i - t_0)^k - \mu_S \right) \\ \xrightarrow{D} \mathfrak{N}(0, \sigma_S^2), \end{aligned} \quad (13)$$

where $\mu_S = E[X(q, t_i) X(m, t_i) K_h(t_i - t_0) (t_i - t_0)^k]$ and $\sigma_S^2 = \text{var}[X(q, t_i) X(m, t_i) K_h(t_i - t_0) (t_i - t_0)^k]$ are respectively the sample mean and variance. By the definition of μ_S , one gets

$$\mu_S = \frac{1}{T} \int X(q, t) X(m, t) K(\frac{1}{h}(t - t_0)) (t - t_0)^k f(t) dt, \quad (14)$$

where $f(t)$ is the sampling density function at t . Using the substitution, $t - t_0 = h\tau$, (14) becomes

$$\begin{aligned} \mu_S &= h^k \int X(q, t_0 + h\tau) X(m, t_0 + h\tau) \\ &\quad \cdot K(\tau) \tau^k f(t_0 + h\tau) d\tau. \end{aligned} \quad (15)$$

We are interested in a Taylor series expansion in term of the bandwidth parameter h . Hence, we assume that $h \rightarrow 0$, while $nh \rightarrow \infty$, i.e. the number of samples is still very large that the CLT is applicable. Using the Taylor series expansion of $f(t_0 + h\tau)$, (15) becomes

$$\begin{aligned} \mu_S &= h^k \int X(q, t_0 + h\tau) X(m, t_0 + h\tau) K(\tau) \tau^k (f(t_0) + O(h)\tau) d\tau \\ &= h^k f(t_0) \mu_{k, r_{mm}(q, m)}(t_0) \{1 + o(1)\} \end{aligned} \quad (16)$$

where $\mu_{k, r_{mm}(q, m)}(t_0) = \int X(q, t_0 + h\tau) X(m, t_0 + h\tau) K(\tau) \tau^k d\tau$ is the local weighted correlation between $X(q, t_0)$ and $X(m, t_0)$ with weighting $K(\tau) \tau^k$.

Assuming that $X(q, t_0 + h\tau)$ is a smooth signal, i.e., it is differentiable to a certain degree, we further have

$$\begin{aligned} X(q, t_0 + h\tau) &= X(q, t_0) + (h\tau) X'(q, t_0) + \cdots \\ &= X(q, t_0) \{1 + o(1)\}. \end{aligned} \quad (17)$$

Similarly, we have

$$X(m, t_0 + h\tau) = X(m, t_0) \{1 + o(1)\}. \quad (18)$$

Accordingly,

$$\mu_S = h^k f(t_0) r_{X, q, m}(t_0) \mu_k \{1 + o(1)\}, \quad (19)$$

where $r_{X, q, m}(t_0) = X(q, t_0) X(m, t_0)$ and $\mu_k = \int \tau^k K(\tau) d\tau$.

The variance σ_S^2 can be derived similarly as

$$\begin{aligned}\sigma_S^2 &= E[X^2(q,t)X^2(m,t)K_h^2(t-t_0)(t-t_0)^{2k}] \\ &= h^{2k-1} \int r_{X,q,m}^2(t_0) f(t_0) K^2(\tau) \tau^{2k} \{1+o(1)\} d\tau \\ &= h^{2k-1} f(t_0) r_{X,q,m}^2(t_0) \nu_k \{1+o(1)\},\end{aligned}\quad (20)$$

where $\nu_k = \int \tau^k K^2(\tau) d\tau$. Combining (13), (19) and (20) gives

$$\begin{aligned}s_{n,k,q,m} &= n\mu_S + \sqrt{nh} O_p(\sqrt{\sigma_S^2}) \\ &= nh^k \{f(t_0) r_{X,q,m}(t_0) \mu_k \{1+o(1)\} \\ &\quad + O_p(\sqrt{(nh)^{-1} \nu_{2k} f(t_0) r_{X,q,m}^2(t_0) \{1+o(1)\}})\} \\ &= nh^k f(t_0) r_{X,q,m}(t_0) \mu_k \{1+o_p(1)\}\end{aligned}\quad (21)$$

Using (21) and (11), one gets

$$\begin{aligned}(\mathbf{S}_n)_{jL+q, jL+m} &= s_{n,j+1,q,m} \\ &= nf(t_0) h^{(j+1)} X(q, t_0) X(m, t_0) \mu_{j+1} \{1+o_p(1)\},\end{aligned}\quad (22)$$

$$\mathbf{S}_n = nf(t_0) \mathbf{H}' \mathbf{S}' \mathbf{H}' \{1+o_p(1)\}.\quad (23)$$

where $\mathbf{H}' = \mathbf{U}\mathbf{H}$, $\mathbf{H} = \text{diag}(1, h, \dots, h^p) \otimes \mathbf{I}_L$, $\mathbf{U} = \mathbf{I}_{p+1} \otimes \text{diag}(X(1, t_0), \dots, X(L, t_0))$, $\mathbf{S}' = (\mathbf{S} \otimes \mathbf{I}_L)$ (\mathbf{I}_L is a matrix of all ones) and $(\mathbf{S})_{0 \leq j, l \leq p} = \mu_{j+l}$.

Using similar arguments, we have

$$\mathbf{S}_n^* = nf(t_0) h^{-1} \sigma^2(t_0) \mathbf{H}' \mathbf{S}'^* \mathbf{H}' \{1+o_p(1)\},\quad (24)$$

where $\mathbf{S}'^* = \mathbf{S}' \otimes \mathbf{I}_L$ and $(\mathbf{S}'^*)_{0 \leq j, l \leq p} = \nu_{j+l}$.

Finally, using Slutsky's theorem, we get the variance of $\hat{\boldsymbol{\beta}}$ as

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\beta}} | \mathbb{X}) &= \mathbf{S}_n^{-1} \mathbf{S}_n^* \mathbf{S}_n^{-1} \\ &= \frac{\sigma^2(t_0)}{nf(t_0)h} \mathbf{H}'^{-1} \mathbf{S}'^{-1} \mathbf{S}'^* \mathbf{S}'^{-1} \mathbf{H}'^{-1} \{1+o_p(1)\}.\end{aligned}\quad (25)$$

Applying the Taylor expansion of the residual $\mathbf{r} = \mathbf{m} - \mathbf{X}\boldsymbol{\beta}$ around $\hat{\boldsymbol{\beta}}$ to the conditional bias, one gets

$$\begin{aligned}\text{Bias}(\hat{\boldsymbol{\beta}} | \mathbb{X}) &= (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}\mathbf{r} = \mathbf{S}_n^{-1} \mathbf{X}^T \mathbf{W}\mathbf{r} \\ &= \mathbf{S}_n^{-1} \mathbf{X}^T \mathbf{W}[\mathbf{x}_{p+1} \boldsymbol{\beta}_{p+1} + o_p(\mathbf{x}_{p+1})] \\ &= \mathbf{S}_n^{-1} \{\mathbf{c}_n \boldsymbol{\beta}_{p+1} + o_p(nh^{p+1})\},\end{aligned}\quad (26)$$

where

$\mathbf{x}_{p+1} = [(t_1 - t_0)^{p+1} \mathbf{X}_1, (t_2 - t_0)^{p+1} \mathbf{X}_2, \dots, (t_n - t_0)^{p+1} \mathbf{X}_n]^T$ with $\mathbf{X}_i = [X(1, t_i), X(2, t_i), \dots, X(L, t_i)]^T$, and $\mathbf{c}_n = \mathbf{X}^T \mathbf{W}\mathbf{x}_{p+1}$.

Using (22) and (23), (26) becomes

$$\begin{aligned}\text{Bias}(\hat{\boldsymbol{\beta}} | \mathbb{X}) &= \frac{1}{nf(t_0)} \mathbf{H}'^{-1} \mathbf{S}'^{-1} \mathbf{H}'^{-1} \mathbf{c}_n \boldsymbol{\beta}_{p+1} \{1+o_p(1)\} \\ &= \mathbf{H}'^{-1} \mathbf{S}'^{-1} \mathbf{c}_p \boldsymbol{\beta}_{p+1} h^{p+1} \{1+o_p(1)\}.\end{aligned}\quad (27)$$

where $\mathbf{c}_p = \mathbf{H}'^{-1} \mathbf{c}_n$.

Since $a^{(j)}(k, t_0) = j! \beta_j(k) = j! \boldsymbol{\beta}(jL+k)$, the asymptotic conditional bias and variance of $\hat{a}^{(j)}(k, t_0)$ are respectively

$$\text{Bias}(\hat{a}^{(j)}(k, t_0) | \mathbb{X}) = \frac{[e_v^T \mathbf{S}'^{-1} \mathbf{c}_p \mathbf{a}^{(p+1)}(t_0)] j! h^{p+1-j}}{(p+1)! X(k, t_0)} + o_p(h^{p+1-j}),\quad (28)$$

$$\text{Var}(\hat{a}^{(j)}(k, t_0) | \mathbb{X}) = \frac{[e_v^T \mathbf{S}'^{-1} \mathbf{S}'^* \mathbf{S}'^{-1} e_v] j!^2 \sigma^2(t_0)}{nf(t_0) h^{2j+1} X^2(k, t_0)} + o_p\left(\frac{1}{nh^{2j+1}}\right),\quad (29)$$

where $v = jL+k$, $e_v = (0, \dots, 0, 1, 0, \dots, 0)^T$ with 1 on the v -th position and zero elsewhere.

It can be seen from (28) and (29) that, with the increase of h , the bias will increase while the variance will decrease. Hence, there exists a locally optimal bandwidth $h^{opt}(k, t_0)$ for estimating $a^{(j)}(k, t_0)$, and $h^{opt}(k, t_0)$ should minimize the MSE as follows:

$$\text{MSE}(\hat{a}^{(j)}(k, t_0) | \mathbb{X}) = [\text{Bias}(\hat{a}^{(j)}(k, t_0) | \mathbb{X})]^2 + \text{Var}(\hat{a}^{(j)}(k, t_0) | \mathbb{X}).\quad (30)$$

The optimal bandwidth can be obtained by setting the derivative of (30) with respect to h to zero. However, some quantities in (28) and (29) are difficult to be calculated, which makes the optimal bandwidth difficult to be estimated accurately. Instead of computing an optimal bandwidth in an analytical form, we proposed an empirical method to select the optimal bandwidth from a finite set of possible bandwidths.

IV. ADAPTIVE BANDWIDTH SELECTION FOR LPM

A. Approximated Bias and Variance

Although the bias and variance cannot be directly computed because of the unknown quantities, good finite sample approximations of the bias and variance, which are highly desirable, can still be derived. By approximating the MSE values from a set of possible bandwidths, an approximate optimal bandwidth can be obtained by minimizing the MSE within the bandwidth set. Towards this end, we need the following approximation methods to determine the bias, variance, and MSE.

The conditional bias, which contains the unknown residual $\mathbf{r} = \mathbf{m} - \mathbf{X}\boldsymbol{\beta}$, can be estimated using a Taylor's expansion with an order $p+p_{ex}$:

$$\hat{\mathbf{b}}(t_0) = (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}\boldsymbol{\tau},\quad (31)$$

where $\boldsymbol{\tau}$ is an $n \times 1$ vector with $\sum_{\zeta=1}^{p_{ex}} \sum_{k=1}^L \beta_{p+\zeta}(k, t_0) (t_i - t_0)^{p+\zeta} X(t_i - k)$

as its i -th ($i=1, 2, \dots, n$) element. The quantities $\beta_{p+\zeta}$ can be estimated by fitting a polynomial of degree $p+p_{ex}$. The excess order p_{ex} is generally chosen as $p_{ex}=2$ because this selection would reduce the computational costs and lead to a bandwidth selector which is not far from being \sqrt{n} -consistent, [2], [4]. However, we still need a pilot bandwidth h^* in the $(p+p_{ex})$ -th order LPM, which will be discussed later.

Next, suppose local homoscedasticity, the conditional variance can be estimated as:

$$\hat{V}(t_0) = (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^2 \mathbf{X} (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \sigma^2(t_0).\quad (32)$$

The noise variance $\sigma^2(t_0)$ is estimated as the normalized weighted residual sum of squares [4]:

$$\hat{\sigma}^2(t_0) = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 K_{h^*}(T_i - t_0)}{\text{tr}\{\mathbf{W}^* - \mathbf{W}^* \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{W}^* \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{W}^*\}},\quad (33)$$

where \mathbf{X}^* and \mathbf{W}^* are respectively the design matrix and weighting matrix in the $(p+p_{ex})$ -th order LPM using pilot bandwidth h^* .

The MSE of the v -th entry of $\hat{\boldsymbol{\beta}}$ at a given point t_0 is

$$\text{MSE}(v, t_0) = \hat{b}^2(v, t_0) + \hat{V}(v, t_0),\quad (34)$$

where $\hat{b}(v, t_0)$ is the v -th element of $\hat{\mathbf{b}}(t_0)$ and $\hat{V}(v, t_0)$ is the v -th diagonal element of $\hat{\mathbf{V}}(t_0)$. Note that we only aim to estimate the coefficients $\hat{a}(k, t_0) = \hat{a}^{(0)}(k, t_0) = \boldsymbol{\beta}(k, t_0)$, although their derivatives can be estimated as well. Now, the remaining problem is how to select a good pilot bandwidth h^* for estimating \mathbf{X}^* , \mathbf{W}^* , and the higher derivatives $\hat{\boldsymbol{\beta}}_{p+\zeta}$.

B. Pilot Bandwidth Selection (the ICI method)

The intersection of confidence interval (ICI) method is employed in this paper to select the pilot bandwidth h^* . The ICI method is an empirical adaptive bandwidth selection method proposed by Goldenshluger and Nemirovski [6], and it has been systematically applied to various areas, including local polynomial regression, image processing, and time-frequency analysis, for selecting the locally adaptive bandwidth [7]-[9]. Due to page limitation, the theoretical background and details of the ICI method are omitted, and more details can be found in [7]-[9]. Here, we only introduce the algorithm of the ICI method.

Given a set of bandwidth parameters in an ascending order,

$$\tilde{\mathbf{H}} = \{h_j \mathbf{I} \mid h_j = h_0 h_a^j, j=1, \dots, J\}, \quad (35)$$

where $h_a > 1$ is a step factor, $h_0 > 0$ is the base bandwidth, and J is the size of the bandwidth set, the ICI method determines the optimal bandwidth by comparing the confidence intervals of the estimates with different bandwidths in the bandwidth set.

Consider a series of confidence intervals $D_j = [L_j, U_j]$ from the estimated $\hat{a}(k, t_0)$ with different values of bandwidth h_j from the set $\tilde{\mathbf{H}}$:

$$U_j = \hat{a}(k, t_0; h_j) + \kappa \cdot \text{Std}(\hat{a}(k, t_0; h_j)), \quad (36)$$

$$L_j = \hat{a}(k, t_0; h_j) - \kappa \cdot \text{Std}(\hat{a}(k, t_0; h_j)), \quad (37)$$

where the standard deviation $\text{Std}(\hat{a}(k, t_0; h_j))$ is the square root of $\hat{V}_k(k, t_0; h_j)$, κ is a threshold parameter used to adjust the width of the confidence interval and it can be chosen as the one that minimizes the cross-validation criterion [7].

The ICI bandwidth selection method computes and examines the following quantities from the confident intervals in order to detect this sudden change:

$$\begin{aligned} \bar{L}_j &= \max[\bar{L}_{j-1}, L_j], \text{ for } j=1, 2, \dots, J, \\ \underline{U}_j &= \min[\underline{U}_{j-1}, U_j], \text{ for } j=1, 2, \dots, J, \\ \bar{L}_0 &= \underline{U}_0 = 0. \end{aligned} \quad (38)$$

It can be seen that \bar{L}_j is the largest upper bound of the confidence interval for bandwidth evaluated up to h_j , while \underline{U}_j is the corresponding lower bound. The largest value of these j for which $\underline{U}_j \geq \bar{L}_j$, denoted by j^+ , gives the desirable bandwidth $h_{ICI}(k, t)$, because above which the confidence intervals no longer intersect with each other.

It should be noted that the ICI method itself is an adaptive bandwidth selection method, but it can also be employed to estimate pilot bandwidth for approximating the MSE of (34).

C. Adaptive Variable Bandwidth Selection

Suppose we have used the ICI method to obtain the adaptive bandwidths $h_{ICI}(k, t)$ for each k -th coefficient at each time instant t_0 . Then, $h_{ICI}(k, t)$ will be used as pilot bandwidths $h^*(k, t)$ to compute \mathbf{X}^* , \mathbf{W}^* , and $\hat{\boldsymbol{\beta}}_{p+\zeta}$. Next, with pilot estimates, we can calculate a series of bias, variance, and MSE values of the LPM estimators using different bandwidths in the set $\tilde{\mathbf{H}}$. Finally, the optimal bandwidth is the bandwidth having the minimum MSE.

Due to the large variability of the MSE estimation [4], it is better for the proposed adaptive bandwidth selector to estimate the optimal bandwidth for a subinterval, not for each individual t_0 . That is, the whole interval of data samples is first split up into a set of non-overlapping subintervals, say I_l .

For each subinterval I_l , we determine the bandwidth function $h_{MMSE}(k, I)$ by minimizing the integrated MSE

$$h_{MMSE}(k, I) = \arg \min_h \int_{I_l} \text{MSE}(k, t; h) dt. \quad (39)$$

$h_{MMSE}(k, I)$ can be smoothed again with the length of I_l to yield the local bandwidth at each location t , $h_{MMSE}(k, t)$.

Lastly, $h_{MMSE}(k, t)$ is used in LPM-based WLS to compute the final estimate $\hat{a}(k, t; h_{MMSE}(k, t))$.

V. SIMULATION RESULTS

We first illustrate the superior performance of the proposed LPM with adaptive bandwidth selection method using a specific TVLM with both fast-changing and static coefficients. The order of the TVLM was $L=4$, and the time-varying coefficients were shown in Figure 1 (a). The input data X was generated from a Gaussian process with zero mean and unit variance. The sampling rate is 512 Hz, and the number of data samples was $n=512$. The length of I_l was 32, and there were totally 16 subintervals. A zero mean white Gaussian noise with variance $\sigma^2=0.1$ was added into the test signal so that the signal-to-noise ratio (SNR) was around 10dB. Epanechnikov kernels, $K(h) = \frac{3}{4}(1-|h|^2)_+$, were employed, and the bandwidth set for ICI was $\{1/64, 1/32, 1/16, 1/8\}$. The polynomial order used in LPM is $p=1$ and the excess order for approximating the bias is $p_{ex}=2$. The mean squared deviation (MSD), $MSD = 10 \log_{10} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^L [a(k, t_i) - \hat{a}(k, t_i)]^2 \right\}$, was calculated from true coefficients and was used as the performance measure.

It can be seen clearly from Figure 1 that, in LPM: 1) a small bandwidth can detect fast change of coefficients, but it may lead to large variations for slow-varying coefficients, as shown in Figure 1 (b); 2) a large bandwidth can obtain smooth estimates when the coefficients varied slowly, but it cannot accurately estimate fast-varying coefficients, as shown in Figure 1 (c); 3) adaptive local bandwidth can obtain satisfactory results for the whole data by employing small bandwidths for fast-varying coefficients and large bandwidths for slow-varying coefficients, as shown in Figure 1 (d)&(e). Further, the MSD values of LPM with constant $h=1/64, 1/32, 1/16, 1/8$, are respectively -13.06dB, -15.69dB, -15.11dB, -

10.40dB, while the MSD values with variable h_{ICI} and h_{MMSE} are respectively -17.76dB and -17.91dB. It can be seen that the LPM estimates with h_{MMSE} had the best performance.

Next, we compared the proposed LPM method with other conventional TVLM identification methods, including the RLS, the Kalman filtering, and the BF methods, in different testing situations. In RLS, two forgetting factors, $\lambda = 0.95$ and 0.99 , were tested, respectively. In the Kalman filtering, the state transition matrix and state noise matrix were set as identity matrices, and the covariance matrices of state noise and observation noise were estimated recursively using the algorithms in [10]. In the BF method, polynomial basis functions were employed and the numbers of basis functions was $p_{BF} = 10$. In the stimulated TVLMs, the time-varying coefficients were generated by filtering white Gaussian signals with zero mean and unit variance using low-pass filters. Four cutoff frequencies f_c (normalized by the sampling rate): 0.01, 0.02, 0.05, 0.1, were used to simulate different extents of coefficient variations (system dynamics). Zero mean white Gaussian noises with different variances $\sigma^2 : 0.01, 0.05, 0.1$, were added to simulate different SNR values. Other parameters were the same with those of previous simulations. The figures listed in Table I were averages of 100 independent runs.

We can see from Table I that the proposed LPM method with adaptive bandwidth selection has a better performance than other methods for TVLMs having different dynamics and different noise levels. In addition, we can conclude from Table I that the LPM with MMSE-derived bandwidths outperforms the LPM with ICI-derived bandwidths in most situations.

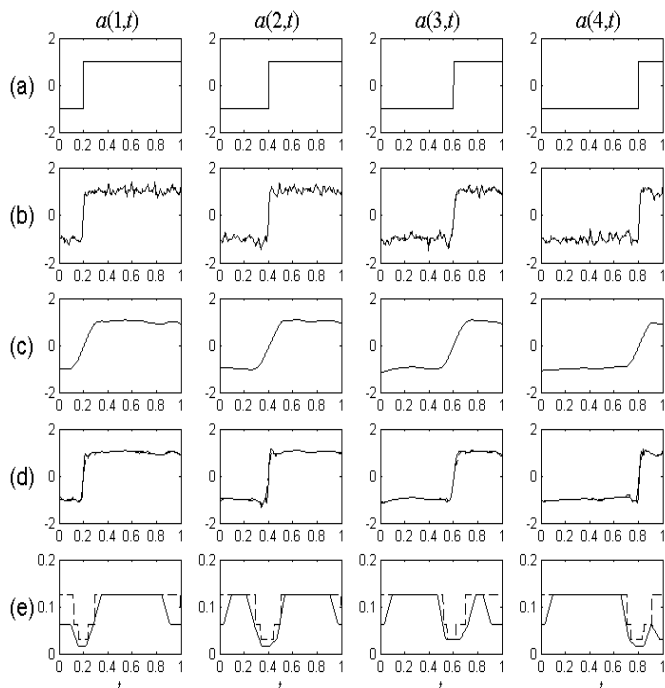


Figure 1. Local polynomial modeling for identification of a specific time-varying linear model: (a) the true coefficients, (b) LPM estimates with a constant small kernel ($h=1/64$), (c) LPM estimates with a constant large kernel ($h=1/8$), (d) LPM estimates with adaptive bandwidths (dash lines: estimates with h_{ICI} , solid lines: estimates with h_{MMSE}), (e) adaptive bandwidths (dash lines: h_{ICI} , solid lines: h_{MMSE})

TABLE I. MSE COMPARISONS BETWEEN DIFFERENT TVLM IDENTIFICATION METHODS (UNIT: dB)

Methods	$f_c=0.01$			$f_c=0.02$		
	$\sigma^2=0.01$	$\sigma^2=0.05$	$\sigma^2=0.1$	$\sigma^2=0.01$	$\sigma^2=0.05$	$\sigma^2=0.1$
BF ($p_{BF}=10$)	-25.73	-24.74	-21.94	-16.39	-16.32	-16.11
RLS ($\lambda=0.95$)	-20.42	-20.95	-18.39	-17.01	-16.38	-15.57
RLS ($\lambda=0.99$)	-18.07	-17.80	-16.33	-14.44	-13.57	-13.15
KF	-21.82	-20.44	-17.95	-16.50	-15.20	-14.60
LPM with h_{ICI}	-29.07	-27.65	-23.03	-23.65	-19.43	-18.29
LPM with h_{MMSE}	-30.44	-28.02	-22.05	-27.38	-21.83	-19.60
Methods	$f_c=0.05$			$f_c=0.1$		
	$\sigma^2=0.01$	$\sigma^2=0.05$	$\sigma^2=0.1$	$\sigma^2=0.01$	$\sigma^2=0.05$	$\sigma^2=0.1$
BF ($p_{BF}=10$)	-9.31	-9.27	-8.91	-5.36	-5.15	-5.46
RLS ($\lambda=0.95$)	-10.16	-9.71	-9.32	-6.07	-5.97	-5.62
RLS ($\lambda=0.99$)	-8.68	-8.11	-7.79	-5.17	-4.67	-4.11
KF	-10.40	-9.98	-9.35	-6.17	-5.95	-5.84
LPM with h_{ICI}	-15.19	-12.18	-10.78	-7.54	-6.71	-6.30
LPM with h_{MMSE}	-20.24	-15.58	-13.38	-10.49	-8.55	-7.65

VI. CONCLUSION

A novel local polynomial modeling method for identification of time-varying linear models was presented. A new data-driven variable bandwidth selection scheme was also developed to minimize the mean squared error. Simulation results showed that the performance of the LPM with adaptive bandwidth selection method is superior to conventional TVLM identification methods in a variety of testing conditions. The proposed LPM method is expected to find applications in communication, bio-signal processing, etc.

REFERENCES

- [1] M. Niedzwiecki, *Identification of Time-varying Processes*, Chechester: John Wiley, 2000.
- [2] D. G. Manolakis, V. K. Ingle, and S. M. Kogan, *Statistical and Adaptive Signal Processing*, New York: McGraw-Hill, 2000.
- [3] J. Fan and I. Gijbels, "Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation," *Stat. Sin.*, vol. 57, pp. 371-394, 1995.
- [4] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications*, London: Chapman and Hall, 1996.
- [5] J. Fan, I. Gijbels, T. C. Hu, and L. S. Huang, "A study of variable bandwidth selection for local polynomial regression," *Stat. Sin.*, vol. 6, pp. 113-127, 1996.
- [6] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Methods Statist.*, vol. 6, no. 2, pp. 135-170, 1997.
- [7] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567-2571, Sept. 1999.
- [8] V. Katkovnik, K. Egiazarian, and J. Astola, *Local Approximation Techniques in Signal and Image Processing*, Bellingham: SPIE Press, Sept. 2006.
- [9] Z. G. Zhang, S. C. Chan, K. L. Ho, and K. C. Ho, "On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data," *J. Signal Process. Syst.*, vol. 52, no. 3, pp. 263-280, Sept. 2008.
- [10] Z. G. Zhang, S. C. Chan, and K. M. Tsui, "A recursive frequency estimator using linear prediction and a Kalman filter-based iterative algorithm," *IEEE Trans. Circuits Syst. II*, vol. 55, no. 6, pp. 576-580, June 2008.