


## REPORT DOCUMENTATION PAGE

2

1a. R <b>AD-A239 822</b> 		1b. RESTRICTIVE MARKINGS <b>E D</b> 1991													
2a. S		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.													
2b. .															
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Department of Statistics Technical Report No. M-853		5. MONITORING ORGANIZATION REPORT NUMBER(S) <b>ARO 27868.10-MA</b>													
6a. NAME OF PERFORMING ORGANIZATION Florida State University	6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office													
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics B-167 Florida State University Tallahassee, FL 32306-3033		7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211													
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office	8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER <b>DAAL03-86-G-0204</b>													
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211		10. SOURCE OF FUNDING NUMBERS <table border="1"><tr><td>PROGRAM ELEMENT NO.</td><td>PROJECT NO.</td><td>TASK NO.</td><td>WORK UNIT ACCESSION NO.</td></tr><tr><td></td><td></td><td></td><td></td></tr></table>		PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.								
PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.												
11. TITLE (Include Security Classification) Identification of nonlinear times series from first order cumulative characteristics															
12. PERSONAL AUTHOR(S) Ian W. McKeague and Mei-Jie Zhang															
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) August, 1991	15. PAGE COUNT 24												
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.															
17. COSATI CODES <table border="1"><tr><td>FIELD</td><td>GROUP</td><td>SUB-GROUP</td></tr><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr></table>		FIELD	GROUP	SUB-GROUP										18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Stationary time series, Markov processes, goodness of fit tests, martingale central limit theorem, nonparametric estimation.	
FIELD	GROUP	SUB-GROUP													
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>We consider the problem of identifying the class of time series model to which a series belongs based on observation of part of the series. Techniques of nonparametric estimation have been applied to this problem by various authors using kernel estimates of the one-step lagged conditional mean and variance functions. We study cumulative versions of Tukey regressogram estimators of such functions. These are more stable than estimates of the mean and variance functions themselves and can be used to construct confidence bands. Goodness-of-fit tests for specific parametric models are also developed.</p>															
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified													
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL												

91 8 19 073

# Identification of nonlinear times series from first order cumulative characteristics

Ian W. McKeague<sup>1</sup> and Mei-Jie Zhang<sup>2</sup>

*Department of Statistics  
Florida State University, Tallahassee, Florida 32306-3033*

FSU Technical Report No. M-853  
USARO Technical Report No. D-121  
AFOSR Technical Report No. 91-261

August, 1991

AMS 1991 subject classifications. Primary 62M10; Secondary 62G07, 62G02, 62M02.

Key words and phrases. Stationary time series, Markov processes, goodness of fit tests, martingale central limit theorem, nonparametric estimation.

---

<sup>1</sup> Research supported by Army Research Office Grant DAA03-90-G0103.

<sup>2</sup> Research supported by the Air Force Office of Scientific Research under Grant AFOSR91-0048.

91 8 19 073

18 91-08230



## Abstract

We consider the problem of identifying the class of time series model to which a series belongs based on observation of part of the series. Techniques of nonparametric estimation have been applied to this problem by various authors using kernel estimates of the one-step lagged conditional mean and variance functions. We study cumulative versions of Tukey regressogram estimators of such functions. These are more stable than estimates of the mean and variance functions themselves and can be used to construct confidence bands. Goodness-of-fit tests for specific parametric models are also developed.



Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

## 1. Introduction

Currently one of the most challenging problems in nonlinear time series analysis is to identify the class of time series model to which a series  $\{X_t\}$  belongs based on observation of part of the series,  $\{X_t, t = 0, 1, \dots, n\}$ . Techniques of nonparametric estimation have been applied to this problem by Robinson (1983), who studied the large sample properties of kernel estimators of lagged conditional means  $E(X_t|X_{t-j})$  and  $E(X_t|X_{t-j}, X_{t-k})$  for various  $j$  and  $k$  values. Such estimators are useful for detecting nonlinearities graphically, see Tong (1990, p. 12). This approach has been further developed by Auestad and Tjøstheim (1990) who focused on kernel estimates of the one-step lagged conditional mean and variance functions  $\lambda(x) = E(X_t|X_{t-1} = x)$  and  $\gamma(x) = \text{var}(X_t|X_{t-1} = x)$  for the purpose of identifying common nonlinear models such as threshold (Tong, 1983) and exponential autoregressive (Ozaki, 1980).

In the present paper we introduce an approach to this problem based on estimation of *cumulative* versions of the conditional mean and variance functions,  $\Lambda(\cdot) = \int_a^\cdot \lambda(x) dx$  and  $\Gamma(\cdot) = \int_a^\cdot \gamma(x) dx$ , where  $a$  is an appropriately chosen point in the state space. These estimators, denoted  $\hat{\Lambda}$  and  $\hat{\Gamma}$ , are obtained by integrating Tukey regressograms for  $\lambda$  and  $\gamma$ . The reason for considering cumulative versions of the conditional mean and variance is that it is possible to derive functional limit theorems, whereas available asymptotic results for kernel or regressogram estimators of  $\lambda$  and  $\gamma$  are only useful pointwise. We advocate  $\hat{\Lambda}$  and  $\hat{\Gamma}$  as natural 'signatures' of a time-series in preference to estimates of  $\lambda$  and  $\gamma$ .

We derive functional limit theorems for  $\hat{\Lambda}$  and  $\hat{\Gamma}$  under conditions that can be readily checked when  $\{X_t\}$  is a Markov chain. These results can be used to construct confidence bands, which are more helpful than confidence intervals in assessing plots. This is the chief benefit from estimating *cumulative* conditional means and variances rather than  $\lambda$  and  $\gamma$  themselves. Another benefit is that  $\hat{\Lambda}$  and  $\hat{\Gamma}$  are relatively insensitive to variations in bandwidth compared to the kernel or regressogram estimators.

We also consider the problem of testing whether the regression function  $\lambda$  has a specific parametric form. Klimko and Nelson (1978) developed consistency and asymptotic distribution results for the conditional least square estimator  $\tilde{\theta}$  of  $\theta$  for the parametric model  $\lambda(x) = g(\theta, x)$ , where  $g$  is a known function and  $\theta$  is an unknown parameter. We construct a goodness-of-fit test for this model based on a comparison of  $\hat{\Lambda}$  and a smoothed version of  $\int_a^\cdot g(\tilde{\theta}, x) dx$ , denoted  $\tilde{\Lambda}$ . Here  $\tilde{\Lambda}$  is

the natural estimator of  $\Lambda$  under the parametric model. We derive a functional limit theorem for the process  $\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})$ . As a particular application we give a test for linearity of  $\lambda$ . Robinson (1983) has given a test for linearity at finitely many locations; other formal tests for linearity are parametric—constructed by arranging the linear model to be nested within various larger parametric models, see Tong (1990, Section 5.2).

There are some connections between the present paper and cumulative hazard function estimation in survival analysis, see the survey articles of Andersen and Borgan (1985) and McKeague and Utikal (1990a). In fact  $\hat{\Lambda}$  is closely related to an estimator introduced by McKeague and Utikal (1990b). Martingale techniques play an important role here, as they do survival analysis.

Our asymptotic distribution results for  $\hat{\Lambda}$  and  $\hat{\Gamma}$  are given in Section 2. The goodness-of-fit test for parametric submodels is discussed in Section 3. We indicate how our results can be extended to lags of higher order in Section 4. The results of a simulation study and some applications to real data are presented in Section 5. Proofs are given in Section 6.

## 2. Estimation of $\Lambda$ and $\Gamma$

Assume that the conditional mean and variance of  $X_t$  given  $X_0, X_1, \dots, X_{t-1}$  only depend on  $X_{t-1}$ . This property holds, for example, if  $\{X_t\}$  is a Markov chain. In particular, an important example is the nonlinear autoregressive process

$$X_t = \lambda(X_{t-1}) + \sigma(X_{t-1})\epsilon_t, \quad (1.1)$$

where  $\{\epsilon_t\}$  are iid with zero-mean and unit variance and  $\gamma = \sigma^2$ . In this case the time series is characterized by the triplet  $(\lambda, \gamma, \text{distribution of } \epsilon_0)$ . We are primarily interested in  $\lambda$  and  $\gamma$ . It is assumed throughout that  $\{X_t\}$  is stationary with a marginal density denoted  $f$ .

We restrict attention to estimation of  $\Lambda$  and  $\Gamma$  on a fixed interval  $[a, b]$ . The regressogram estimators  $\hat{\lambda}$  and  $\hat{\gamma}$  are defined as follows. Let  $\mathcal{I}_1, \dots, \mathcal{I}_{d_n}$  be a partition of  $[a, b]$  made up of intervals of equal length  $w_n$ , the bins of the regressogram, and denote  $\mathcal{I}_x = \mathcal{I}_j$  for  $x \in \mathcal{I}_j$ . Set

$$\hat{\lambda}(x) = (nw_n \hat{f}(x))^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} X_t,$$

$$\hat{\gamma}(x) = (nw_n \hat{f}(x))^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} (X_t - \hat{\lambda}(x))^2,$$

where  $\hat{f}$  is the histogram estimator of  $f$  given by

$$\hat{f}(x) = (nw_n)^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\},$$

and  $I(\cdot)$  is the indicator function. Regressogram estimators were introduced by Tukey (1961) and have been studied recently by Diebolt (1990).

Introduce the estimators

$$\hat{\Lambda}(\cdot) = \int_a^\cdot \hat{\lambda}(x) dx \quad \text{and} \quad \hat{\Gamma}(\cdot) = \int_a^\cdot \hat{\gamma}(x) dx.$$

Although it is possible to use the more sophisticated kernel estimators to yield better estimates of  $\lambda$  and  $\gamma$ , there is little to be gained from using them in  $\hat{\Lambda}$  and  $\hat{\Gamma}$ , which are less sensitive to variations in  $\hat{\lambda}$  and  $\hat{\gamma}$ . We prefer the regressogram estimators due to their computational simplicity. In practice, care needs to be taken in choosing the interval  $[a, b]$  and the bins to ensure that the regressogram estimates are not too unstable. For good results, the binwidths should be of comparable size (we have taken them to be of equal size merely to simplify the notation), and there should be at least 5 observations per bin.

Ideally, in order to carry out inference on  $\lambda$ , using a confidence band for  $\Lambda$  say, we would like to find the limiting distribution of  $\sqrt{n}(\hat{\Lambda} - \Lambda)$ . However, for technical reasons we are only able to obtain a satisfactory weak convergence theory when  $\Lambda$  is replaced by the smoothed version of  $\Lambda$  given by  $\Lambda^*(z) = \int_a^z \lambda^*(x) dx$ , where

$$\lambda^*(x) = \int_{\mathcal{I}_x} f^*(u) \lambda(u) du \Big/ \int_{\mathcal{I}_x} f^*(u) du$$

and  $f^*$  is the histogram estimator of  $f$  determined by a finer partition of  $[a, b]$  consisting of intervals of equal length  $w_n^*$ .

We regard  $\Lambda^*$  as a 'surrogate' for  $\Lambda$ , which is reasonable since  $\Lambda^*$  converges uniformly in probability to  $\Lambda$ . However  $\sqrt{n}(\Lambda^* - \Lambda)$  may not be asymptotically negligible; see the remark following the proof of Theorem 2.1. If it is (for example if  $\lambda$  is piecewise constant over  $\mathcal{I}_1, \dots, \mathcal{I}_{d_n}$  for some  $n$ ) then  $\Lambda^*$  is not needed and we can deal with  $\Lambda$  directly. Similar comments can be made concerning  $\hat{\Gamma}$ , with  $\Gamma^*$  defined in a similar way to  $\Lambda^*$ .

We now proceed to state the main results of this section, giving the asymptotic distributions of  $\hat{\Lambda}$  and  $\hat{\Gamma}$ . It is assumed throughout that  $\lambda$  and  $\gamma$  are Lipschitz. We also need:

*Condition A*

- (A1)  $EX_0^8 < \infty$ .
- (A2)  $(X_0, X_t)$  has a bounded joint density  $f_t$  for all  $t \geq 1$ , and the marginal density  $f$  is continuous and does not vanish on  $[a, b]$ .
- (A3)  $\sup_{x \in [a, b]} \text{var}[\hat{f}(x)] = o(w_n)$ .

**THEOREM 2.1.** *Suppose that Condition A holds,  $nw_n \rightarrow \infty$ ,  $nw_n^4 \rightarrow 0$  and  $w_n^* \sim w_n^2$  as  $n \rightarrow \infty$ . Then  $\sqrt{n}(\hat{\Lambda} - \Lambda^*)$  converges in distribution in  $C[a, b]$  to a continuous Gaussian martingale with mean zero and variance function*

$$H(z) = \int_a^z \frac{\gamma(x)}{f(x)} dx.$$

**THEOREM 2.2.** *Suppose that the hypotheses of Theorem 2.1 hold, except that  $nw_n^2 \rightarrow \infty$  and  $EX_0^{16} < \infty$ . Then  $\sqrt{n}(\hat{\Gamma} - \Gamma^*)$  converges in distribution in  $C[a, b]$  to a continuous Gaussian martingale with mean zero and variance function  $\int_a^\cdot \nu/f dx$ , where  $\nu(x) = \text{var}([X_t - \lambda(x)]^2 | X_{t-1} = x)$  and  $\nu$  is assumed to be Lipschitz.*

*Checking Condition (A3):* A large class of stationary Markov processes  $\{X_t\}$  that satisfy Condition (A3) is described by Auestad and Tjøstheim (1990), who show (pp. 680, 681) that strong mixing with a geometric mixing rate implies  $\text{var}[\hat{f}(x)] \sim O((nw_n)^{-1})$  uniformly over  $[a, b]$  provided that  $f$  is bounded there. Thus (A3) holds under this mixing condition if  $nw_n^2 \rightarrow \infty$ . In a particular example it will be easier to check geometric ergodicity (Nummelin, 1984), which implies strong mixing with a geometric mixing rate. Geometric ergodicity is in turn implied by a readily checkable condition of Tweedie (1983).

Another way of checking Condition (A3), which is not restricted to Markov processes, is to verify a mixing condition of Castellana and Leadbetter (1986, Theorem 3.3). They considered the following *dependence index sequence*

$$\beta_n = \sup_{x, y \in [a, b]} \sum_{t=1}^n |f_t(x, y) - f(x)f(y)|$$

and showed that

$$\text{var}(\hat{f}(x)) = O\left(\frac{\beta_n}{n}\right) + O\left(\frac{1}{nw_n}\right),$$

uniformly in  $x$ . Hence, if  $\beta_n = O(d_n)$  and  $nw_n^2 \rightarrow \infty$ , then Condition (A3) holds. The moment condition (A1) can probably be weakened, but it makes the results easier to prove.

We now mention some possible applications of these results.

*Confidence bands:* Condition (A3) implies that  $\hat{f}$  is uniformly consistent (see the remark at the beginning of Section 6). Thus, using Theorem 2.2, it can be shown that  $\hat{H}(\cdot) = \int_a \hat{\gamma}/\hat{f} dx$  is a uniformly consistent estimator of  $H$ . Then, by Theorem 2.1, an asymptotic  $100(1 - \alpha)\%$  confidence band for  $\Lambda^*$  is given by

$$\hat{\Lambda}(x) \pm c_\alpha n^{-1/2} \hat{H}(b)^{1/2} \left(1 + \frac{\hat{H}(x)}{\hat{H}(b)}\right) \quad x \in [a, b],$$

where  $c_\alpha$  is the upper  $\alpha$  quantile of the distribution of  $\sup_{t \in [0, 1/2]} |B^0(t)|$  and  $B^0$  is the Brownian bridge process, see Andersen and Borgen (1985, p. 114). Tables for  $c_\alpha$  can be found in Hall and Wellner (1980). A confidence band for  $\Gamma^*$  can be obtained in a similar way.

*Testing simple hypotheses:* A test of the simple hypotheses,  $\lambda = \lambda_0$  and  $\gamma = \gamma_0$ , where  $\lambda_0$  and  $\gamma_0$  are given, can be made by checking whether the above confidence bands contain  $\Lambda_0^*$  and  $\Gamma_0^*$ . A rather different approach has been taken by Diebolt (1990), who developed a test based on a piecewise constant version of

$$\sqrt{n} \left( \int_a \hat{f}(x) \hat{\lambda}(x) dx - \int_a \hat{f}(x) \lambda_0(x) dx \right).$$

Diebolt obtained a functional limit theorem for this process, and a similar one designed to test  $\gamma = \gamma_0$ , where  $\gamma_0$  is given and  $\lambda$  is known, in the special case of model (1.1).

*Testing for a difference between two regression functions:* Consider the “two-sample problem” of testing whether two independent time series have identical regression functions  $\lambda$ . Denote the various functions, sample sizes estimators etc. associated with the two series by using a subscript 1 or 2, as in  $\lambda_j$ ,  $j = 1, 2$ . Let  $n = n_1 + n_2$ .



Then, if  $n_j/n \rightarrow p_j$  for  $j = 1, 2$ , and the conditions of Theorem 2.1 are satisfied for the two series,  $\sqrt{n}(\hat{\Lambda}_1 - \hat{\Lambda}_2)$  converges in distribution in  $C[a, b]$  to a continuous Gaussian martingale with mean zero and variance function

$$p_1^{-1} \int_a^\cdot \frac{\gamma_1(x)}{f_1(x)} dx + p_2^{-1} \int_a^\cdot \frac{\gamma_2(x)}{f_2(x)} dx,$$

provided that  $\lambda_1 = \lambda_2$  on  $[a, b]$  and  $\sqrt{n}(\Lambda_1^* - \Lambda_2^*)$  converges uniformly in probability to zero. The latter condition holds if the common  $\lambda$  is piecewise constant, as mentioned earlier. Confidence bands for  $\Lambda_1^* - \Lambda_2^*$  are constructed as above. Some plots of such bands are given in Section 5.

### 3. Goodness-of-fit tests for parametric models

In this section we consider the problem of testing whether  $\lambda$  belongs to a parametric family  $\{g(\theta, \cdot) : \theta \in \Theta\}$  of regression functions. Here  $g$  is a known deterministic function, and  $\Theta$  is a closed, bounded subset of  $\mathbf{R}^p$ . Our test is based on a functional limit theorem for  $\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})$ , where  $\tilde{\Lambda}(z) = \int_a^z \tilde{\lambda}(x) dx$ ,

$$\tilde{\lambda}(x) = \int_{\mathcal{I}_x} f^*(u) g(\tilde{\theta}, u) du \Big/ \int_{\mathcal{I}_x} f^*(u) du$$

and  $\tilde{\theta}$  is the conditional least squares estimator minimizing  $\sum_{t=1}^n (X_t - g(\theta, X_{t-1}))^2$ .

First we state a version of the consistency and asymptotic normality result of Klimko and Nelson (1978) that is adapted to our present setting, taking the opportunity to simplify their approach a little. We assume that  $\{X_t\}$  is an ergodic process and  $E(X_1 - g(\theta, X_0))^2$  has a unique minimum at a point  $\theta_0$  in the interior of  $\Theta$ .

For a matrix  $Y$  and a vector  $y$ , denote  $\|Y\| = \sup_{i,j} |Y_{ij}|$ ,  $\|y\| = \sup_i |y_i|$ , and  $y^{\otimes 2} = yy^T$ . It is assumed that  $g(\theta, x)$  is twice differentiable w.r.t.  $\theta$  and the corresponding derivatives are denoted  $g'$  and  $g''$ .

*Condition B*

- (B1) There exists a function  $J$  such that  $\|g''(\theta, x) - g''(\zeta, x)\| \leq J(x)\delta(\theta - \zeta)$ , where  $J(X_0)$  has a finite second moment, and  $\lim_{\alpha \rightarrow 0} \delta(\alpha) = 0$ .
- (B2) There exists a function  $K$  such that  $\|g''(\theta, x)\| \leq K(x)$ , where  $K(X_0)$  has a finite fourth moment.

(B3)  $g(\theta_0, X_0)$  and  $\gamma(X_0)$  have finite second moments, and all the components of  $g'(\theta_0, X_0)$  have a finite fourth moment.

(B4) The matrices

$$\begin{aligned} V &= E[g'(\theta_0, X_0)^{\otimes 2}], \\ S &= E[g'(\theta_0, X_0)^{\otimes 2} \gamma(X_0)] \end{aligned}$$

are positive definite.

THEOREM 3.1. Under Condition B,  $\tilde{\theta} \xrightarrow{a.s.} \theta_0$  and  $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, V^{-1}SV)$ .

We now state the main result of this section.

THEOREM 3.2. Suppose that Conditions A and B hold and  $\lambda(\cdot) = g(\theta_0, \cdot)$ . If  $nw_n \rightarrow \infty$  and  $nw_n^4 \rightarrow 0$ , then  $\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})$  converges in distribution in  $C[a, b]$  to

$$\int_a^\cdot \sqrt{\gamma(x)/f(x)} dW(x) - \psi(\cdot) \int_{-\infty}^\infty g'(\theta_0, x) \sqrt{\gamma(x)f(x)} dW(x),$$

where

$$\psi(z) = \int_a^z g'(\theta_0, x)^T dx V^{-1},$$

and  $W$  is the Wiener process extended to the whole real line.

A chi-squared goodness-of-fit test for the parametric model is now easily constructed. Let  $\mathcal{J}_1, \dots, \mathcal{J}_L$  be a partition of  $[a, b]$  consisting of intervals. Denote the increment of  $\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})$  over  $\mathcal{J}_j$  by  $\Delta_j$ . It can be checked that  $\Delta = (\Delta_j)$  converges in distribution to a Gaussian random vector with mean zero and covariance matrix having  $rl$ th entry

$$H(\mathcal{J}_r \cap \mathcal{J}_l) + \psi(\mathcal{J}_r)S\psi(\mathcal{J}_l)^T - \psi(\mathcal{J}_r)H_1(\mathcal{J}_l) - \psi(\mathcal{J}_l)H_1(\mathcal{J}_r),$$

where  $H$  is defined in Theorem 2.1 and

$$H_1(z) = \int_a^z g'(\theta_0, x)\gamma(x) dx.$$

Let  $\hat{G}$  be the natural estimate of this covariance matrix obtained by replacing the unknown  $\theta$ ,  $f$ , and  $\gamma$  by their estimates. Then the Wald test statistic  $\Delta^T \hat{G}^{-1} \Delta$

has a limiting  $\chi_q^2$  distribution under the parametric model, where  $q$  is the rank of the limiting covariance matrix of  $\Delta$ . A test for a parametric model of  $\gamma$  can be developed in a similar way.

#### 4. Extension to higher order lags

It is possible to extend our results to higher order lagged conditional means, but it would be unreasonable to use more than second order lags in practice because of the “curse of dimensionality”—the data becomes sparser at an exponential rate as the dimension increases. We briefly indicate how to handle the second order lagged conditional mean  $\lambda(x, y) = E(X_t | X_{t-1} = x, X_{t-2} = y)$ . This mostly amounts to just a reinterpretation of our original notation.

Denote  $\mathbf{X}_t = (X_t, X_{t-1})$  and assume that the conditional mean and variance of  $X_t$  given  $X_0, X_1, \dots, X_{t-1}$  are  $\lambda(\mathbf{X}_{t-1})$  and  $\gamma(\mathbf{X}_{t-1})$  respectively. The regressogram estimator of  $\lambda$  is

$$\hat{\lambda}(x, y) = (nw_n^2 \hat{f}(x, y))^{-1} \sum_{t=2}^n I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\}$$

where  $\mathcal{I}_{xy} = \mathcal{I}_x \times \mathcal{I}_y$  and

$$\hat{f}(x, y) = (nw_n^2)^{-1} \sum_{t=2}^n I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\}.$$

Here  $\hat{f}$  is a histogram estimate of the density of  $\mathbf{X}_1$ .

In order to obtain the asymptotic distribution of  $\hat{\Lambda} = \int_a \int_a \hat{\lambda} dx dy$  we need to extend Conditions (A2) and (A3). In Condition (A2),  $f_t$  is now the joint density of  $\mathbf{X}_1$  and  $\mathbf{X}_t$ . The rate in (A3) is now  $o(w_n^2)$ . Castellana and Leadbetter’s (1986) dependence index sequence  $\beta_n$  can be extended in the same fashion. If  $\beta_n = O(d_n^2)$  and  $nw_n^4 \rightarrow \infty$ , then the extended version of Condition (A3) holds.

The functions  $f^*$ ,  $\lambda^*$  and  $\tilde{\lambda}$  are defined much as before, except using a partition of  $[a, b]^2$  consisting of squares with sides of length  $w_n^*$ , and integrals over  $\mathcal{I}_{xy}$ . Let  $C[a, b]^2$  denote the space of continuous functions on  $[a, b]^2$  provided with the supremum norm. Our earlier results now extend as follows.

**THEOREM 4.1.** *Suppose that the extended version of Condition A holds,  $nw_n^2 \rightarrow \infty$ ,  $nw_n^{9/2} \rightarrow 0$ , and  $w_n^* \sim w_n^{9/4}$ . Then  $\sqrt{n}(\hat{\Lambda} - \Lambda^*)$  converges in distribution in  $C[a, b]^2$*

to a two-parameter Gaussian martingale with zero mean and variance function  $\int_a^\cdot \int_a^\cdot \gamma/f \, dx dy$ .

**THEOREM 4.2.** Suppose that the hypotheses of Theorem 4.1 and the extended version of Condition B hold and  $\lambda = g(\theta_0, \cdot, \cdot)$ . Then  $\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})$  converges in distribution in  $C[a, b]^2$  to a process which has the same form as the limiting process in Theorem 3.2 except that the integrals are with respect to the Wiener sheet extended to  $\mathbf{R}^2$ .

## 5. Numerical results and examples

**5.1. Simulation study:** We have carried out simulations using three model examples taken from Auestad and Tjøstheim (1990):

*Model 1:* linear autoregressive,  $X_t = 0.8X_{t-1} + \epsilon_t$ ;

*Model 2:* threshold autoregressive,

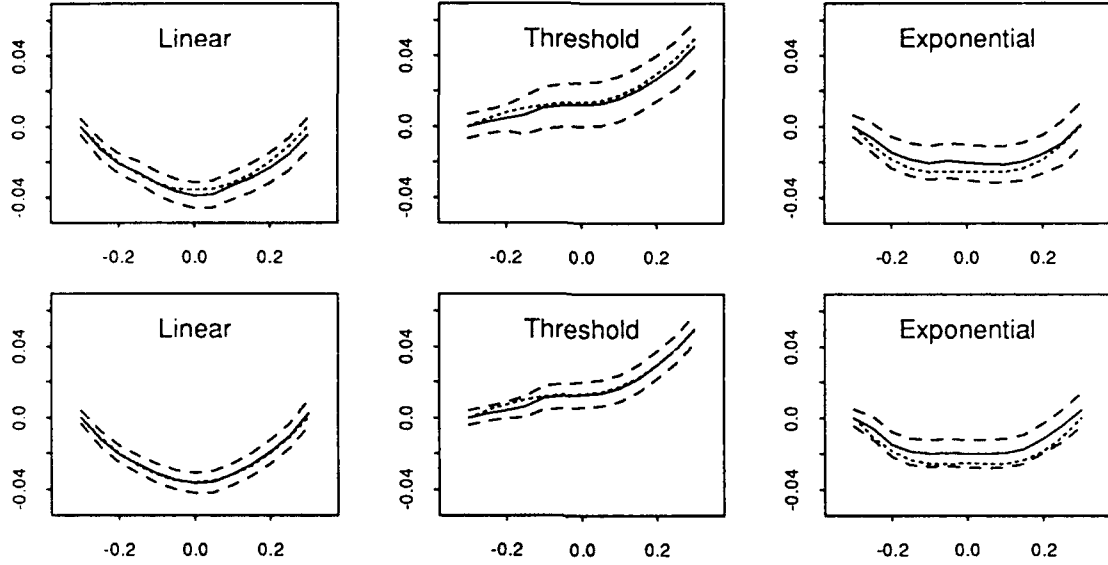
$$X_t = \begin{cases} -0.3X_{t-1} + \epsilon_t, & \text{if } X_{t-1} \leq 0, \\ 0.8X_{t-1} + \epsilon_t, & \text{if } X_{t-1} > 0; \end{cases}$$

*Model 3:* exponential autoregressive,  $X_t = \{0.8 - 1.1 \exp(-50X_{t-1}^2)\}X_{t-1} + \epsilon_t$ .

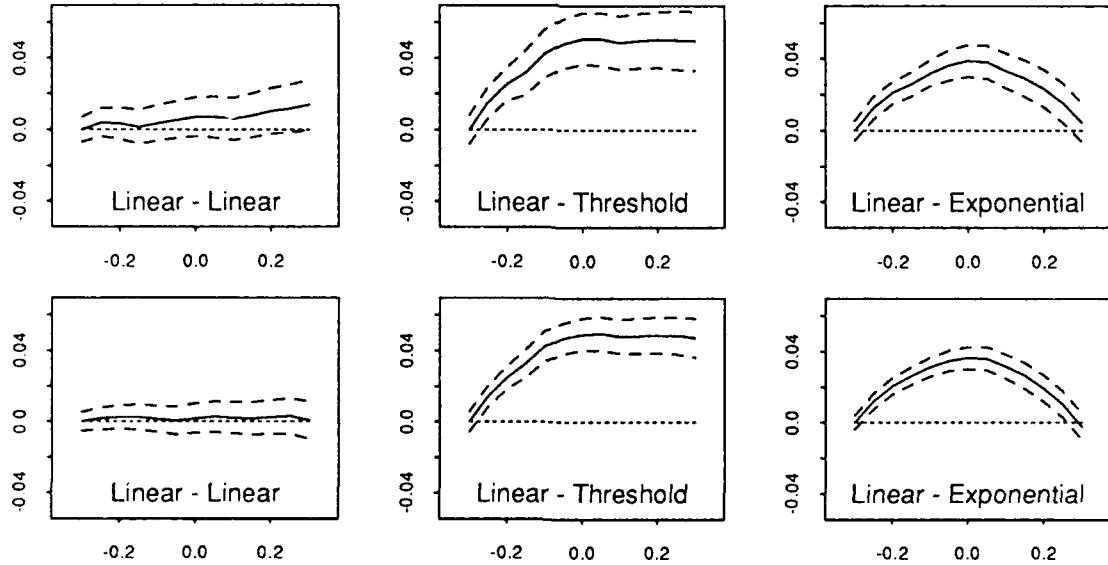
Here  $\epsilon_t$  is Gaussian white noise with mean zero and standard deviation 0.1. Auestad and Tjøstheim (1990) checked geometric ergodicity and stationarity for these examples.

We restricted estimation of  $\Lambda$  to the interval  $[-0.3, 0.3]$ . The binwidth was taken as  $w_n = 0.05$  (same as Auestad and Tjøstheim, who plotted point estimates of  $\lambda$  for these three models). Inspecting the plots of  $\hat{\Lambda}$  in Figure 1, we find that the three models are easily distinguishable, even for sample size as low as 250. The parabolic shape of the linear autoregressive model, and the ‘squashed’ parabola of the exponential autoregressive are especially distinct.

Figure 2 shows plots of differences between the estimates of the cumulative regression functions in the two sample problem, for various pairs of the above models. In the first plot in each row, the two series are generated using the linear model and the zero function is contained within the band, so our test would correctly conclude that the regression functions are identical. In the other plots, the zero function is well outside the bands and the test correctly concludes that the regression functions are different.



**Figure 1.**  $\hat{\Lambda}$  with 95% confidence bands; solid lines,  $\hat{\Lambda}$ ; dotted lines,  $\Lambda$ ; dashed lines, confidence bands; first row,  $n = 250$ ; second row,  $n = 500$ .



**Figure 2.**  $\hat{\Lambda}_1 - \hat{\Lambda}_2$  with 95% confidence bands; first row,  $n = 250$ ; second row,  $n = 500$ .

**Table 1.** Observed Levels and Powers of Goodness-of-Fit Test for Linear Autoregressive Model at Nominal Level of 5%; binwidth,  $w_n = 0.05$ ;  $L = 4$ .

Observed Series	Sample Size					
	100	250	500	1000	1500	2500
Linear	0.0974	0.0604	0.0518	0.0496	0.0426	0.0496
Threshold	0.8674	0.9938	0.9994	1.0000	1.0000	1.0000
Exponential	0.4836	0.9376	0.9996	1.0000	1.0000	1.0000

NOTE: The data were generated using the Gaussian random number generator of Marsaglia and Tsang (1984). The number of samples in each run was 5000.

Table 1 gives observed levels and powers of the chi-squared goodness-of-fit test for the linear autoregressive model  $X_t = \theta X_{t-1} + \epsilon_t$ , when the time series is generated by each model. At small sample sizes (less than 250), the covariance matrix estimator  $\hat{G}$  sometimes failed to be positive definite and the chi-squared statistic value was negative. The percentage of negative chi-squared statistics was 2.4% and 0.1% for sample sizes of 100 and 250 with the linear model; 7% and 3% with the threshold model; 3.6% and 0.94% with the exponential model. We rejected the linear model when the chi-squared statistic was negative. This is reasonable since  $\hat{G}$  is consistent under the null hypothesis so that a negative chi-squared statistic is evidence in favor of the alternative. The observed levels are very close to their nominal 5% values and the powers are close to 100% (except for  $n = 100$ ) under the threshold and exponential models.

**5.2. Canadian lynx data:** The classic Canadian lynx data set consists of the annual numbers of Canadian lynx trapped in the Mackenzie River district of North-west Canada for the period 1821–1934. Various parametric time series models have been proposed to fit these data, see Tong (1990) for an extensive review. Moran (1953) fitted a second order linear autoregressive model, after first transforming by  $\log_{10}$ , to obtain

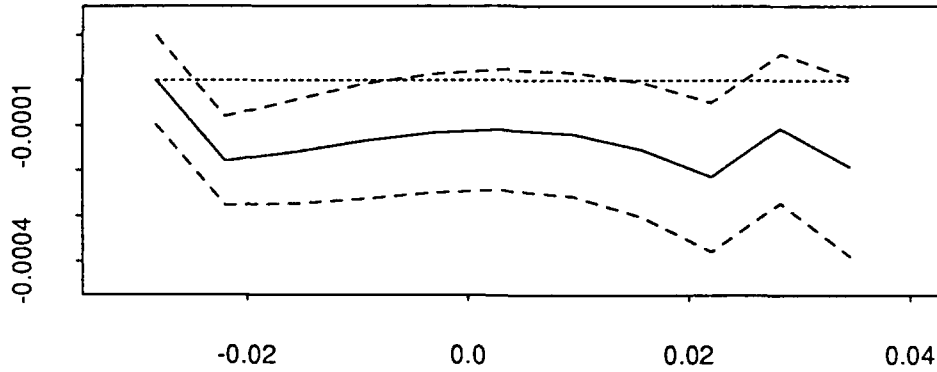
$$X_t = 1.05 + 1.41X_{t-1} - 0.77X_{t-2} + \epsilon_t$$

where  $\epsilon_t \sim \text{iid}(0, 0.04591)$ . However, many authors, including Bartlett (1954), Hannan (1960), Campbell and Walker (1977) and Tong (1977), have judged this model to be inadequate compared with some other parametric models.

We carried out our goodness-of-fit test for the second order linear model (hav-

ing three parameters) using  $d_n = 5, 6, \dots, 10$ , and 4 (2 by 2) and 9 (3 by 3) degrees of freedom. The bins were arranged to cover the whole range of the data and to contain, as closely as possible, equal numbers of data points. All our tests indicated an extremely strong departure from the linear model.

**5.3. IBM stock price data:** Consider the set of IBM daily closing stock prices from late 1959 to mid 1960 (period I) and mid 1961 to early 1962 (period II) given in Tong (1990). The daily relative change in price appears to be stationary and is used in place of the raw data. Tong (1990) tested for linearity and decided that period I is linear and period II is nonlinear. Figure 3 gives a plot of the difference between the estimates of the cumulative regression functions in the two periods, along with the 95% confidence band, using  $d_n = 10$ . The confidence band does not contain the zero function, so we conclude that the regression functions for the two periods differ significantly from one another. Our chi-squared test with  $d_n = 8, 10$  and 12, and degrees of freedom  $L = 2$  and 4, gave the same result.



**Figure 3.**  $\hat{\Lambda}_1 - \hat{\Lambda}_2$  with 95% confidence band for IBM stock price data;  $d_n = 10$ ;  $\hat{\Lambda}_1 =$  period I,  $\hat{\Lambda}_2 =$  period II.

## 6. Proofs

Recall that the intervals  $\mathcal{I}_j$  partition  $[a, b]$ . We write them explicitly as  $\mathcal{I}_j = (x_{j-1}, x_j]$ ,  $j = 1, \dots, d_n$ . In what follows we need  $\hat{f}$  to be uniformly consistent for  $f$  on  $[a, b]$ . This holds under Conditions (A2) and (A3) since

$$E(\sup_{x \in [a, b]} |\hat{f}(x) - f^\dagger(x)|)^2 \leq \sum_{j=1}^{d_n} \text{var}[\hat{f}(x_j)] = d_n o(w_n) \rightarrow 0,$$

and, by stationarity,  $f^\dagger(x) \equiv E\hat{f}(x) = w_n^{-1} \int_{\mathcal{I}_x} f(u) du \rightarrow f(x)$  uniformly on  $[a, b]$ . Also note that  $\xi_t \equiv X_t - \lambda(X_{t-1})$  is a martingale difference with respect to the natural filtration  $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$ .

PROOF OF THEOREM 2.1. First observe that  $\hat{f}(x) = w_n^{-1} \int_{\mathcal{I}_x} f^* du$ . Since  $\lambda$  is Lipschitz and  $\hat{f}$  converges uniformly in probability to  $f$ , which is bounded away from zero, we have

$$\begin{aligned} \hat{\lambda}(x) &= [nw_n \hat{f}(x) w_n^*]^{-1} \int_{\mathcal{I}_x} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x^*\} [\lambda(X_{t-1}) + \xi_t] du \\ &= \lambda^*(x) + O_P(w_n^*) + [nw_n \hat{f}(x)]^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \xi_t. \end{aligned}$$

Hence, by  $w_n^* \sim w_n^2$  and  $nw_n^4 \rightarrow 0$ ,

$$\begin{aligned} \sqrt{n}(\hat{\Lambda} - \Lambda^*)(z) &= \frac{1}{\sqrt{n}w_n} \int_a^z \frac{\sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \xi_t}{\hat{f}(x)} dx + o_P(1) \\ &= \overline{M(n, \cdot)}(z) + R(z) + o_P(1) \end{aligned}$$

uniformly in  $z$ , where

$$\begin{aligned} M(k, z) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{[zd_n]} f^\dagger(x_j)^{-1} \sum_{t=1}^k I\{X_{t-1} \in \mathcal{I}_j\} \xi_t, \quad k = 1, \dots, n \\ R(z) &= \frac{1}{\sqrt{n}w_n} \int_a^z \left[ \frac{f^\dagger(x) - \hat{f}(x)}{\hat{f}(x)f^\dagger(x)} \right] \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \xi_t dx, \end{aligned}$$

and, given a function  $\phi$  defined on  $[a, b]$ ,  $\bar{\phi}$  is the piecewise linear approximation to  $\phi$  that agrees with  $\phi$  at each  $x_j$ . Here  $[\cdot]$  denotes the integer part and  $M(k, z)$  is defined to be zero when  $[zd_n] = 0$ . To complete the proof we need to show that the remainder term  $R$  converges uniformly in probability to zero and  $M(n, \cdot) \xrightarrow{\mathcal{D}} m$ , where  $m$  denotes the Gaussian martingale given in the statement of the theorem, for Lemma 4.1 of McKeague (1988) then implies that  $\overline{M(n, \cdot)} \xrightarrow{\mathcal{D}} m$ .

Now  $M(\cdot, z)$  is an  $\mathcal{F}_t$  martingale for each fixed  $z$ . We shall use the martingale central limit theorem (see Theorem A.2 of Aalen (1977), for instance) to show that all finite dimensional distributions of  $M(n, \cdot)$  converge to those of  $m$ . The



predictable variation process of  $M(\cdot, z)$  evaluated at  $k = n$  is given by

$$\begin{aligned}\langle M(\cdot, z) \rangle_n &= \frac{1}{n} \sum_{j=1}^{[zd_n]} f^\dagger(x_j)^{-2} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} \gamma(X_{t-1}) \\ &= \int_a^z \frac{[\gamma(x) + O(w_n)] \hat{f}(x)}{f^\dagger(x)^2} dx + o_P(1) \xrightarrow{P} H(z).\end{aligned}$$

Next, we check the Lindeberg condition that

$$L_n \equiv \frac{1}{n} \sum_{j=1}^{[zd_n]} f^\dagger(x_j)^{-2} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} E\left\{ \xi_t^2 I\left(\frac{I\{X_{t-1} \in \mathcal{I}_j\} |\xi_t|}{\sqrt{n} f^\dagger(x_j)} > \epsilon\right) \middle| \mathcal{F}_{t-1} \right\}$$

converges in probability to zero for all  $\epsilon > 0$ . By the conditional Cauchy-Schwarz and Chebyshev inequalities, and since  $f^\dagger$  is bounded away from zero on  $[a, b]$ , the conditional expectation in  $L_n$  is bounded above by

$$\begin{aligned}& \{E(\xi_t^4 | \mathcal{F}_{t-1})\}^{\frac{1}{2}} \{[\sqrt{n} \epsilon f^\dagger(x_j)]^{-2} E(I\{X_{t-1} \in \mathcal{I}_j\} \xi_t^2 | \mathcal{F}_{t-1})\}^{\frac{1}{2}} \\ &= O\left(\frac{1}{\sqrt{n}}\right) \{E(\xi_t^4 | \mathcal{F}_{t-1})\}^{\frac{1}{2}} I\{X_{t-1} \in \mathcal{I}_j\} \gamma(X_{t-1})^{\frac{1}{2}}.\end{aligned}$$

Now (A1), stationarity of  $\{X_t\}$  and  $\lambda$  Lipschitz imply that  $\sup_t E \xi_t^8 < \infty$ , so again using the Cauchy-Schwarz inequality, (A2), boundedness of  $f$  and  $\gamma$ , and  $nw_n \rightarrow \infty$ , we have

$$E(L_n) \leq O\left(\frac{1}{n\sqrt{n}}\right) \sum_{j=1}^{d_n} \sum_{t=1}^n (E I\{X_{t-1} \in \mathcal{I}_j\})^{\frac{1}{2}} = O\left(\frac{1}{\sqrt{nw_n}}\right) \rightarrow 0,$$

so the Lindeberg condition holds. By the martingale central limit theorem, the one dimensional distributions of  $M(n, \cdot)$  converge to those of  $m$ . The above argument readily extends to all finite dimensional distributions of  $M(n, \cdot)$  using the fact that increments of  $M(\cdot, z)$  over disjoint intervals in  $z$  are orthogonal martingales.

The next step is to show that  $\{M(n, \cdot) : n \geq 1\}$  is tight in  $D[a, b]$ . By a slight extension of Theorem 15.6 of Billingsley (1968), it suffices to show that

$$E|M(n, y) - M(n, x)|^2 |M(n, z) - M(n, y)|^2 \leq C(z - x)^2 + o(1)$$

for  $a \leq x \leq y \leq z \leq b$ , where  $C$  is a generic positive constant. Indeed, by the Cauchy-Schwarz inequality it suffices to show that

$$E|M(n, y) - M(n, x)|^4 \leq C(y - x)^2 + o(1). \quad (6.1)$$

Using Rosenthal's inequality (Hall and Heyde, 1980, p. 23), the left hand side of (6.1) is bounded by

$$CE \left[ \sum_j \sum_{t=1}^n E \left( \left( \frac{I\{X_{t-1} \in \mathcal{I}_j\} \xi_t}{\sqrt{n} f^\dagger(x_j)} \right)^2 \middle| \mathcal{F}_{t-1} \right) \right]^2 + C \sum_j \sum_{t=1}^n E \left( \frac{I\{X_{t-1} \in \mathcal{I}_j\} \xi_t}{\sqrt{n} f^\dagger(x_j)} \right)^4, \quad (6.2)$$

where the summation over  $j$  runs from  $[xd_n] + 1$  to  $[yd_n]$ . By (A2), the first term of (6.2) is bounded by

$$\begin{aligned} & O\left(\frac{1}{n^2}\right) \sum_j \sum_{t=1}^n E(I\{X_{t-1} \in \mathcal{I}_j\}) + O\left(\frac{1}{n^2}\right) \sum_{j,h} \sum_{s \neq t} E(I\{X_{s-1} \in \mathcal{I}_j, X_{t-1} \in \mathcal{I}_h\}) \\ & = O\left(\frac{1}{n}\right)(y-x) + O(1)(y-x)^2 \leq C(y-x)^2 + o(1), \end{aligned}$$

and the second term of (6.2) is bounded by

$$O\left(\frac{1}{n^2}\right) \sum_j \sum_{t=1}^n (EI\{X_{t-1} \in \mathcal{I}_j\})^{\frac{1}{2}} (E\xi_t^8)^{\frac{1}{2}} \leq O\left(\frac{1}{n\sqrt{w_n}}\right)(y-x) = o(1),$$

since  $nw_n \rightarrow \infty$ . So (6.1) holds.

It only remains to show that  $R$  converges uniformly in probability to 0. Since  $\hat{f}$  is a uniformly consistent estimator of  $f$ , which is bounded away from zero on  $[a, b]$ , it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{d_n} |\hat{f}(x_j) - f^\dagger(x_j)| \left| \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} \xi_t \right| \xrightarrow{P} 0. \quad (6.3)$$

By the Cauchy-Schwarz inequality and (A3), the expectation of (6.3) is bounded by

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{j=1}^{d_n} (\text{var}[\hat{f}(x_j)])^{\frac{1}{2}} \left\{ \sum_{t=1}^n E(I\{X_{t-1} \in \mathcal{I}_j\} \xi_t)^2 \right\}^{\frac{1}{2}} \\ & = \frac{d_n}{\sqrt{n}} o(\sqrt{w_n}) O(\sqrt{nw_n}) \rightarrow 0, \end{aligned}$$

as required.  $\square$

*Remark:* To show the uniform consistency of  $\hat{f}$ , we used Condition (A3), which can be readily checked under  $nw_n^2 \rightarrow \infty$ . However, under  $nw_n^2 \rightarrow \infty$  we are unable to

show that  $\sqrt{n}(\Lambda^* - \Lambda)$  tends uniformly in probability to zero since  $\lambda^* - \lambda$  is at best of order  $O(w_n)$ . Thus we have not been able to obtain an asymptotic distribution result for  $\sqrt{n}(\hat{\Lambda} - \Lambda)$  in general.

PROOF OF THEOREM 2.2. Define  $\tau_t = \xi_t^2 - \gamma(X_{t-1})$ . Since  $\lambda$  and  $\gamma$  are Lipschitz,  $\hat{f}$  uniformly converges in probability to  $f$  which is bounded away from 0,

$$\begin{aligned}\hat{\gamma}(x) &= [nw_n \hat{f}(x)]^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \left[ \xi_t - \frac{\sum_i I\{X_{i-1} \in \mathcal{I}_x\} \xi_i}{\sum_i I\{X_{i-1} \in \mathcal{I}_x\}} + O_P(w_n) \right]^2 \\ \gamma^*(x) &= [nw_n \hat{f}(x)]^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \gamma(X_{t-1}) + O_P(w_n^*).\end{aligned}$$

Noting that  $\sqrt{n}w_n^* \sim (nw_n^4)^{1/2} \rightarrow 0$ , we have

$$\begin{aligned}\sqrt{n}(\hat{\Gamma} - \Gamma^*)(z) &= \frac{1}{\sqrt{n}w_n} \int_a^z \frac{\sum_t I\{X_{t-1} \in \mathcal{I}_x\} \tau_t}{\hat{f}(x)} dx \\ &\quad + O_P\left(\frac{1}{n^{3/2}w_n^2}\right) \int_a^z \left[ \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \xi_t \right]^2 dx \\ &\quad + O_P\left(\frac{1}{\sqrt{n}}\right) \int_a^z \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} \xi_t dx + o_P(1)\end{aligned} \tag{6.4}$$

uniformly in  $z$ . The second term in (6.4) is uniformly bounded by

$$\begin{aligned}O_P\left(\frac{1}{n^{3/2}w_n}\right) \sum_{j=1}^{d_n} \left[ \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} \xi_t \right]^2 \\ = O_P\left(\frac{1}{n^{3/2}w_n^2}\right) O_P(nw_n) = O_P\left(\frac{1}{\sqrt{n}w_n^2}\right) = o_P(1),\end{aligned}$$

since  $nw_n^2 \rightarrow \infty$ . The third term in (6.4) is uniformly bounded by

$$O_P\left(\frac{w_n}{\sqrt{n}}\right) \sum_{j=1}^{d_n} \left| \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} \xi_t \right| \leq O_P\left(\frac{1}{\sqrt{n}}\right) O_P(\sqrt{n}w_n) = o_P(1).$$

Hence,  $\sqrt{n}(\hat{\Gamma} - \Gamma^*)$  has the same form as  $\sqrt{n}(\hat{\Lambda} - \Lambda^*)$  except that  $\tau_t$  replaces  $\xi_t$ . Note that  $\tau_t$  is a martingale difference and  $E(\tau_t^2 | X_{t-1} = x) = \nu(x)$ . Also, the condition

$EX_0^{16} < \infty$  implies that  $\sup_t E(\tau_t^8) < \infty$ . Therefore, the result follows by the proof of Theorem 2.1.  $\square$

PROOF OF THEOREM 3.1. Define  $Q_n(\theta) = \sum_{t=1}^n (X_t - g(\theta, X_{t-1}))^2$  and  $q(\theta) = E(X_1 - g(\theta, X_0))^2$ . Note that

$$\begin{aligned} \frac{1}{n}(Q_n(\theta) - Q_n(\zeta)) &= \frac{2}{n} \sum_{t=1}^n X_t [g(\theta, X_{t-1}) - g(\zeta, X_{t-1})] \\ &\quad + \frac{1}{n} \sum_{t=1}^n [g(\theta, X_{t-1}) + g(\zeta, X_{t-1})][g(\theta, X_{t-1}) - g(\zeta, X_{t-1})]. \end{aligned}$$

By Condition (B1), we have that

$$|g(\theta, x) - g(\zeta, x)| \leq [CK(x) + \|g'(\theta_0, x)\|] \|\theta - \zeta\|.$$

Hence, under the moment conditions in (B1) and (B3), and the ergodic theorem,

$$\frac{1}{n} |Q_n(\theta) - Q_n(\zeta)| \leq C \|\theta - \zeta\|,$$

where  $C$  is finite almost surely. It follows that  $\{n^{-1}Q_n(\cdot)\}$  is equicontinuous. Again by the ergodic theorem,  $n^{-1}Q_n(\theta) \xrightarrow{\text{a.s.}} q(\theta)$  ( $< \infty$ ), which implies that  $\{n^{-1}Q_n(\cdot)\}$  is pointwise bounded almost surely. It follows by the Arzela-Ascoli theorem that this family of functions is almost surely relatively compact in the space of continuous functions on  $\Theta$ . Thus  $n^{-1}Q_n(\cdot)$  converges uniformly to  $q(\cdot)$  on  $\Theta$  almost surely. Since  $q(\theta)$  has a unique minimum at  $\theta_0 \in \Theta$ , and  $\tilde{\theta}$  minimizes  $Q_n(\theta)$ , we conclude that  $\tilde{\theta}$  is consistent.

Next, Taylor expanding  $Q'_n$  about  $\theta_0$ , we can write

$$\sqrt{n}(\tilde{\theta} - \theta_0) = U_n/V_n(\theta^*),$$

where  $U_n = U_n^{(n)}$ ,

$$U_k^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^k g'(\theta_0, X_{t-1}) \xi_t, \quad k = 1, \dots, n; \quad V_n(\theta) = \frac{1}{2n} Q_n''(\theta),$$

and  $\theta^*$  is on the line joining  $\theta_0$  and  $\tilde{\theta}$ . Since  $U_k^{(n)}$  is a martingale in  $k$ , the martingale central limit theorem can be used to show that  $U_n \xrightarrow{\mathcal{D}} N(0, S)$  under the moment

conditions in (B3). To complete the proof we need to show that  $V_n(\theta^*) \xrightarrow{P} V$ . Routine algebra gives that

$$\begin{aligned} V_n(\theta^*) &= \frac{1}{n} \sum_{t=1}^n g'(\theta_0, X_{t-1})^{\otimes 2} \\ &+ \frac{1}{n} \sum_{t=1}^n [g'(\theta^*, X_{t-1})^{\otimes 2} - g'(\theta_0, X_{t-1})^{\otimes 2}] \\ &+ \frac{1}{n} \sum_{t=1}^n [g(\theta^*, X_{t-1}) - g(\theta_0, X_{t-1})] g''(\theta^*, X_{t-1}) \\ &- \frac{1}{n} \sum_{t=1}^n \xi_t g''(\theta_0, X_{t-1}) \\ &+ \frac{1}{n} \sum_{t=1}^n [X_t - g(\theta_0, X_{t-1})] [g''(\theta_0, X_{t-1}) - g''(\theta^*, X_{t-1})]. \end{aligned}$$

By (B3) and the ergodic theorem, the first term converges to  $V$  almost surely. Using,  $\theta^* \xrightarrow{\text{a.s.}} \theta_0$ , Conditions (B1)–(B3) and the ergodic theorem it can be shown that the second, third and last terms above converge almost surely to zero. A strong law of large numbers, see Hall and Heyde (1980, Theorem 2.19), (B2) and (B3), and the martingale difference property of  $\xi_t$ , give that the fourth term also converges almost surely to zero. We conclude that  $V_n(\theta^*) \xrightarrow{\text{a.s.}} V$ .  $\square$

PROOF OF THEOREM 3.2. By Taylor expanding  $g(\cdot, u)$  about  $\theta_0$  for each fixed  $u$ ,

$$\sqrt{n}(\tilde{\Lambda} - \Lambda^*)(z) = \left( \int_a^z [w_n \hat{f}(x)]^{-1} \left[ \int_{\mathcal{I}_x} f^*(u) g'(\theta_u^*, u) du \right] dx \right)^T \sqrt{n}(\tilde{\theta} - \theta_0),$$

where  $\theta_u^*$  lies on the line joining  $\theta_0$  and  $\tilde{\theta}$ . Since  $\tilde{\theta}$  is a consistent estimator of  $\theta_0$ , and  $g'$  is continuous,

$$\int_a^z [w_n \hat{f}(x)]^{-1} \left[ \int_{\mathcal{I}_x} f^*(u) g'(\theta_u^*, u) du \right] dx \xrightarrow{P} \int_a^z g'(\theta_0, x) dx.$$

From the proof of Theorem 3.1,  $\sqrt{n}(\tilde{\theta} - \theta_0) = V^{-1}U_n + o_P(1)$ , so using the proof of Theorem 2.1,

$$\sqrt{n}(\hat{\Lambda} - \tilde{\Lambda})(z) = \overline{M(n, \cdot)}(z) - \psi(z)U_n + o_P(1)$$

uniformly in  $z$ . By a  $D[a, b] \times \mathbf{R}$  version of Lemma 4.1 of McKeague (1988), it suffices to show that  $(M(n, \cdot), U_n)$  converges in distribution to  $(m(\cdot), U_\infty)$ , where

$$m(z) = \int_a^z \sqrt{\gamma(x)/f(x)} dW(x),$$

$$U_\infty = \int_{-\infty}^{\infty} g'(\theta_0, x) \sqrt{\gamma(x)f(x)} dW(x).$$

The proofs of Theorems 2.1 and 3.1 give that  $M(n, \cdot) \xrightarrow{\mathcal{D}} m$  and  $U_n \xrightarrow{\mathcal{D}} U_\infty$ . It only remains to show that the finite dimensional distributions of  $(M(n, \cdot), U_n)$  converge to those of  $(m(\cdot), U_\infty)$ . This is done by applying the martingale central limit theorem to the vector-valued martingale consisting of  $U^{(n)}$  and increments of  $M(\cdot, z)$  over disjoint intervals in  $z$ . In particular, note that

$$\begin{aligned} \langle M(\cdot, z), U^{(n)} \rangle_n &= \frac{1}{n} \sum_{j=1}^{\lfloor zd_n \rfloor} f^\dagger(x_j)^{-1} \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_j\} g'(\theta_0, X_{t-1}) \gamma(X_{t-1}) \\ &= \frac{1}{nw_n} \int_a^z \left[ \frac{g'(\theta_0, x) \gamma(x) + O(w_n)}{f^\dagger(x)} \right] \sum_{t=1}^n I\{X_{t-1} \in \mathcal{I}_x\} dx + o_P(1) \\ &\xrightarrow{P} \int_a^z g'(\theta_0, x) \gamma(x) dx = \text{Cov}(m(z), U_\infty). \end{aligned}$$

The Lindeberg conditions involving increments of  $M(\cdot, z)$  have been checked in the proof of Theorem 2.1, and those involving the  $p$  components of  $U^{(n)}$  in the proof of Theorem 3.1.  $\square$

PROOF OF THEOREM 4.1. Since  $\lambda$  is Lipschitz and  $\hat{f}$  uniformly converges in probability to  $f$ , which is bounded away from 0, we have

$$\hat{\lambda}(x, y) = \lambda^*(x, y) + O_P(w_n^*) + [nw_n^2 \hat{f}(x, y)]^{-1} \sum_{t=1}^n I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\} \xi_t.$$

Since  $\sqrt{n}w_n^* \sim (nw_n^{9/2})^{1/2} \rightarrow 0$ ,

$$\sqrt{n}(\hat{\Lambda} - \Lambda^*)(z_1, z_2) = \frac{1}{\sqrt{n}w_n^2} \int_a^{z_1} \int_a^{z_2} \frac{\sum_t I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\} \xi_t}{\hat{f}(x, y)} dx dy + o_P(1).$$

The remainder of the proof is almost identical to the proof of Theorem 2.1 except that  $\mathbf{X}_t$  replaces  $X_t$ ,  $\mathcal{I}_{xy}$  replaces  $\mathcal{I}_x$ ,  $w_n^2$  replaces  $w_n$ , double integral (summation)

replaces single integral (summation), and  $\bar{\phi}$  is the piecewise linear approximation to  $\phi$  determined by cells  $\mathcal{I}_{xy}$ . Note that Condition  $nw_n^2 \rightarrow \infty$  is used in checking the Lindeberg condition, and tightness can be checked by using a two-dimensional time parameter version of Theorem 15.6 of Billingsley (1986) given in Bickel and Wichura (1971). We omit the details.  $\square$

PROOF OF THEOREM 4.2. The proof is similar to the proof of Theorem 3.2 and is omitted.

## References

- Aalen, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. *Z. Wahrsch. verw. Gebiete* **38** 261–277.
- Andersen, P. K. and Borgan, Ø. (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12** 97–158.
- Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* **77** 669–687.
- Bartlett, M. S. (1954). Problèmes de l'analyse spectrale des series temporelles stationnaires. *Publ. Inst. Stat. (Univ. Paris)*, Fasc. **3** 119–134.
- Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.* **42** 1656–1670.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Campbell, M. J. and Walker, A. M. (1977). A survey of statistical work on the Mackenzie River series of annual Canadian lynx trappings for the years 1821–1934, and a new analysis. *J. R. Stat. Soc. Series A* **140** 411–431, discussion 448–468.
- Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stoch. Process. Appl.* **21** 179–193.
- Diebolt, J. (1990). Testing the functions defining a nonlinear autoregressive time series. *Stoch. Process. Appl.* **36** 85–106.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York.
- Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67** 133–143.
- Hannan, E. J. (1960). *Time Series Analysis*. Methuen, London.
- Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6** 629–642.
- Marsaglia, G. and Tsang, W. W. (1984). A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *Siam. J. Sci. Stat. Comput.* **5** 349–359.
- McKeague, I. W. (1988). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process. Appl.* **28** 221–239.
- McKeague, I. W. and Utikal, K. J. (1990a). Stochastic calculus as a tool in survival analysis: a review. *Appl. Math. Comput.* **38** 23–49.
- McKeague, I. W. and Utikal, K. J. (1990b). Identifying nonlinear covariate effects in semimartingale regression models. *Probab. Theory Related Fields* **87** 1–25.
- Moran, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle. I: structure and prediction. *Aust. J. Zool.* **1** 163–173.
- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press.



- Ozaki, T. (1980). Non-linear time series model for non-linear random vibrations. *J. Appl. Prob.* **17** 84-93.
- Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Series Anal.* **4** 185-207.
- Tong, H. (1977). Some comments on the Canadian lynx data—with Discussion. *J. R. Stat. Soc. Series A* **140** 432-436, 448-468.
- Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics **21**. Springer, New York.
- Tong, H. (1990). *Non-linear Time Series: a Dynamical System Approach*. Clarendon.
- Tukey, J. (1961). Curves as parameters and touch estimation. *Proc. 4th Berkeley Symp. Math. Statist. Probab.* **1** 681-684.
- Tweedie, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing theory. In *Papers in Probability, Statistics and Analysis*, Eds. J. F. C. Kingman and G. E. H. Reuter, pp. 260-276. Cambridge University Press.