The Annals of Statistics
1994, Vol. 22, No. 1, 495–514

IDENTIFICATION OF NONLINEAR TIME SERIES FROM FIRST ORDER CUMULATIVE CHARACTERISTICS

BY IAN W. MCKEAGUE[1] AND MEI-JIE ZHANG[1,2]

Florida State University and Medical College of Wisconsin

A new approach to the problem of identifying a nonlinear time series model is considered. We argue that *cumulative* lagged conditional mean and variance functions are the appropriate "signatures" of a nonlinear time series for the purpose of model identification, being analogous to cumulative distribution functions or cumulative hazard functions in iid models. We introduce estimators of the cumulative lagged conditional mean and variance functions and study their asymptotic properties. A goodness-of-fit test for parametric time series models is also developed.

1. Introduction. Currently, one of the most challenging problems in nonlinear time series analysis is to identify the class of 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), page 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 $\gamma(x) = \mathrm{var}(X_t|X_{t-1} = x)$ 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 $\widehat{\Lambda}$ and $\widehat{\Gamma}$, are obtained by integrating Tukey regressograms for λ and γ. 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 λ and γ are only useful pointwise. We advocate $\widehat{\Lambda}$ and $\widehat{\Gamma}$ as natural "signatures" of a time series in preference to estimates of λ and γ.

Received August 1991; revised October 1992.

[1]Research partially supported by the Air Force Office of Scientific Research under Grant AFOSR91-0048.

[2]Research partially supported by National Cancer Institute Grant 1 RO1 CA54706-01.

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.

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

We also consider the problem of testing whether the conditional mean function λ has a specific parametric form. Klimko and Nelson (1978) developed consistency and asymptotic distribution results for the conditional least square estimator $\widetilde{\theta}$ of θ for the parametric model $\lambda(x) = g(\theta, x)$, where g is a known function and θ is an unknown parameter. We construct a goodness-of-fit test for this model based on a comparison of $\widehat{\Lambda}$ and $\widetilde{\Lambda} = \int_a^{\cdot} g(\widetilde{\theta}, x)\,dx$. Here $\widetilde{\Lambda}$ is the natural estimator of Λ under the parametric model. We obtain a functional limit theorem for the process $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$ and use it to derive a chi-squared test. A particular application is a test for linearity of λ. Our test is more powerful than Robinson's [(1983), page 193] test based on estimates of λ at finitely many points, and it is preferable to tests constructed by arranging the linear model to be nested within various larger parametric models [see Tong, (1990), Section 5.2]—such tests are sensitive only to restricted classes of alternatives.

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 $\widetilde{\Lambda}$ is closely related to an estimator introduced by McKeague and Utikal (1990b). Martingale techniques play an important role here, as they do in survival analysis.

Our asymptotic distribution results for $\widehat{\Lambda}$ and $\widehat{\Gamma}$ 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 Λ and Γ. Assume that the conditional mean and variance of X_t given $X_0, X_1, \ldots, X_{t-1}$ only depend on X_{t-1}. This property holds, for example, if $\{X_t\}$ is a Markov chain. In particular, it holds in the important case of a nonlinear autoregressive process

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

where $\{\varepsilon_t\}$ are iid with zero mean and unit variance and $\gamma = \sigma^2$. If the distribution of ε_0 is symmetric, then process (1.1) is characterized by the triple $(\lambda, \gamma$, distribution of $\varepsilon_0)$. For us it is λ and γ that are of primary interest.

We restrict attention to estimation of Λ and Γ on a fixed interval $[a, b]$. It is assumed throughout that $\{X_t\}$ is stationary with a marginal density denoted f. The regressogram estimators $\widehat{\lambda}$ and $\widehat{\gamma}$ are defined as follows. Let $\mathcal{I}_1, \ldots, \mathcal{I}_{d_n}$ 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

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

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

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

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

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

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

Although more-sophisticated estimators (e.g., kernel) will outperform regressograms in estimating λ and γ, there appears little to be gained from using them in $\widehat{\Lambda}$ and $\widehat{\Gamma}$ since integration has such a strong smoothing effect. In fact we shall see that $\widehat{\Lambda}$ and $\widehat{\Gamma}$ converge at rate $O_P(n^{-1/2})$, a rate that does not even involve the bin width w_n. Kernel estimators would be useful, however, in extending our approach to higher-order models (see Section 4). In practice, some care needs to be taken in choosing the interval $[a, b]$ and the bins to ensure that the regressogram estimates are not too unstable. Based on Monte Carlo studies, we find that for good results the bin widths should be of comparable size (we have taken them to be of equal length w_n merely for simplicity), and there should be at least five observations per bin.

We now state the main results of this section, giving the asymptotic distributions of $\widehat{\Lambda}$ and $\widehat{\Gamma}$. It is assumed throughout that λ, γ and f are twice differentiable (although this can be relaxed to some extent). We also need the following.

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 does not vanish on $[a, b]$;

(A3) $\operatorname{var}[\widehat{f}(x)] = o(w_n)$ and $\operatorname{var}[\widehat{\delta}(x)] = o(1/n)$ uniformly over $x \in [a, b]$, where

$$\widehat{\delta}(x) = (nw_n)^{-1}\sum_{t=1}^{n} I\{X_{t-1} \in \mathcal{I}_x\}(X_{t-1} - x).$$

THEOREM 2.1. *Suppose that Condition* A *holds,* $nw_n^2 \to \infty$ *and* $nw_n^4 \to 0$ *as* $n \to \infty$. *Then* $\sqrt{n}(\widehat{\Lambda} - \Lambda)$ *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* $EX_0^{16} < \infty$. *Then* $\sqrt{n}(\widehat{\Gamma} - \Gamma)$ *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* $\nu(x) = \operatorname{var}([X_t - \lambda(x)]^2 | X_{t-1} = x)$ *is assumed to be Lipschitz.*

2.1. *Checking condition* (A3). A sufficient condition for a stationary Markov process $\{X_t\}$ to satisfy (A3) is that it is strong mixing with a geometric mixing rate and $nw_n^2 \to \infty$. This can be seen using arguments similar to Auestad and Tjøstheim [(1990), pages 680, 681] which show that, under the geometric mixing rate, $\operatorname{var}[\widehat{f}(x)] = O(1/(nw_n))$ and $\operatorname{var}[\widehat{\delta}(x)] = O(w_n/n)$ uniformly over $[a,b]$ provided that f is bounded there. Note that the variance of $I\{X_{t-1} \in \mathcal{I}_x\}(X_{t-1} - x)$ is of order $O(w_n^3)$, which leads to the faster rate of convergence for $\operatorname{var}[\widehat{\delta}(x)]$. 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). A sufficient (but by no means necessary) condition for geometric ergodicity of the nonlinear autoregressive process (1.1) is that λ and σ are bounded on compact sets and there exists a constant C such that

$$\sup_{|x|>C} (|\lambda(x)|/|x|) < 1, \qquad \sup_{|x|>C} |\sigma(x)| < \infty.$$

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

$$\operatorname{var}[\widehat{f}(x)] = O\left(\frac{\beta_n}{n}\right) + O\left(\frac{1}{nw_n}\right),$$

uniformly in x. Similar calculations show that

$$\operatorname{var}[\widehat{\delta}(x)] = O\left(\frac{w_n^2 \beta_n}{n}\right) + O\left(\frac{w_n}{n}\right),$$

uniformly in x. Hence, if $\beta_n = O(d_n)$ and $nw_n^2 \to \infty$, then conditions (A3) holds (recall that $d_n = 1/w_n$ is the number of bins). 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.

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

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

where c_α is the upper α quantile of the distribution of $\sup_{t\in[0,1/2]} |B^0(t)|$ and B^0 is a Brownian bridge [see Anderson and Borgan (1985), page 114]. Tables for c_α can be found in Hall and Wellner (1980). A confidence band for Γ can be obtained in a similar way.

2.3. *Testing simple hypotheses.* A test of the simple hypotheses, $\lambda = \lambda_0$ and $\gamma = \gamma_0$, where λ_0 and γ_0 are specified, can be made by checking whether the above confidence bands contain Λ_0 and Γ_0. A rather different approach has been taken by Diebolt (1990), who developed a test based on a piecewise constant version of

$$\sqrt{n} \int_a^{\cdot} \widehat{f}(x)(\widehat{\lambda}(x) - \lambda_0(x))\, dx.$$

In the special case of the autoregressive model (1.1), Diebolt obtained a functional limit theorem for the above process, and a similar one designed to test $\gamma = \gamma_0$ (γ_0 specified) when λ is known.

2.4. *Testing for a difference between two conditional mean functions.* Consider the "two-sample problem" of testing whether two independent time series have identical conditional mean functions λ. Denote the various functions, sample sizes, estimators and so on associated with the two series by using a subscript 1 or 2, as in λ_j, $j = 1,2$. Let $n = n_1 + n_2$. Then, if $n_j/n \to p_j > 0$, for $j = 1,2$, and the conditions of Theorem 2.1 are satisfied for the two series, $\sqrt{n}(\widehat{\Lambda}_1 - \widehat{\Lambda}_2)$ 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 $[a,b]$. 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 λ belongs to a given parametric family $\{g(\theta,\cdot): \theta \in \Theta\}$. Here g is a known deterministic function, and Θ is a closed, bounded subset of $\mathbb{R}^p$. Our test is based on a functional limit theorem for $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$, where $\widetilde{\Lambda}(z) = \int_a^z g(\widetilde{\theta},x)\,dx$ and $\widetilde{\theta}$ is the conditional least squares estimator $\widetilde{\theta} = \arg\min_{\theta\in\Theta} \sum_{t=1}^n (X_t - g(\theta,X_{t-1}))^2$.

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 the true parameter value θ_0, which is assumed to be in the interior of Θ.

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

CONDITION B.

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

$$V = E\left[g'(\theta_0,X_0)^{\otimes 2}\right],$$
$$S = E\left[g'(\theta_0,X_0)^{\otimes 2}\gamma(X_0)\right]$$

are positive definite.

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

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* $nw_n^2 \to \infty$ *and* $nw_n^4 \to 0$, *then* $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$ *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 a Brownian motion.*

A chi-squared goodness-of-fit test for the parametric model is now easily constructed. Let $\mathcal{J}_1, \ldots, \mathcal{J}_q$ be a partition of $[a,b]$ consisting of intervals. Denote the increment of $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$ over $\mathcal{J}_j$ by Δ_j. It can be checked that $\Delta = (\Delta_j)$ converges in distribution to a Gaussian random vector with mean zero and covariance matrix having rlth 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 $\widehat{G}$ be the natural estimate of this covariance matrix obtained by replacing the unknown θ, f and γ by their estimates. Then, under the parametric model, the Wald test statistic $\Delta^T \widehat{G}^{-1} \Delta$ has a limiting chi-squared distribution with q degrees of freedom, provided that the limiting covariance matrix of Δ is of full rank and the conditions of Theorem 3.2 hold. A test for a parametric model of γ can be developed in a similar way.

4. Extension to higher order lags. It is possible in principle to extend our approach to higher order lagged conditional means and variances, but this will be curbed in practice by the "curse of dimensionality"—the data become sparser at an exponential rate as the dimension increases. However, there are some extremely long time series in some fields in which nonlinearity is of interest and the approach may be helpful. We briefly indicate how to handle the case of a secondorder lagged conditional mean $\lambda(x,y) = E(X_t | X_{t-1} = x, X_{t-2} = y)$. This mostly amounts to a recasting of our original notation.

Write $\mathbf{X}_t = (X_t, X_{t-1})$ and assume that the conditional mean and variance of X_t given $X_0, X_1, \ldots, X_{t-1}$ are $\lambda(\mathbf{X}_{t-1})$ and $\gamma(\mathbf{X}_{t-1})$, respectively. Set $\widehat{\Lambda} = \int_a^{\cdot} \int_a^{\cdot} \widehat{\lambda}\, dx\, dy$, where $\widehat{\lambda}$ is the regressogram

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

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

$$\widehat{f}(x,y) = \left(n w_n^2\right)^{-1} \sum_{t=2}^{n} I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\}$$

is the histogram estimate of the density of $\mathbf{X}_1$.

In order to obtain the asymptotic distribution of $\widehat{\Lambda}$ we need new versions of conditions (A2) and (A3). In (A2) f_t becomes the joint density of $\mathbf{X}_1$ and $\mathbf{X}_t$. In (A3) we need the stronger condition $\operatorname{var}[\widehat{f}(x,y)] = o(w_n^2)$ and also $\operatorname{var}[\widehat{\delta}_x(x,y)] = o(1/n)$, where

$$\widehat{\delta}_x(x,y) = \left(n w_n^2\right)^{-1} \sum_{t=2}^{n} I\{\mathbf{X}_{t-1} \in \mathcal{I}_{xy}\}(X_{t-1} - x),$$

and similarly for $\widehat{\delta}_y$, in which $X_{t-1} - x$ is replaced by $X_{t-2} - y$. The use of a kernel estimator in place of the regressogram $\widehat{\lambda}$ may be desirable here because of the faster rate of convergence needed for $\operatorname{var}[\widehat{f}]$.

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 second order version of Condition* A *holds, $nw_n^2 \to \infty$ and $nw_n^4 \to 0$. Then $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$ converges in distribution in $C[a,b]^2$ to a two-parameter Gaussian martingale with mean zero and variance function given by $\int_a^{\cdot}\int_a^{\cdot} \gamma/f \, dx \, dy$.*

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 $\sqrt{n}(\widehat{\Lambda} - \widetilde{\Lambda})$ 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 a Brownian sheet.*

5. Numerical results and examples

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

Model 1. linear autoregressive, $X_t = 0.8X_{t-1} + \varepsilon_t$;
Model 2. threshold autoregressive,

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

Model 3. exponential autoregressive,

$$X_t = \{0.8 - 1.1 \exp(-50X_{t-1}^2)\}X_{t-1} + \varepsilon_t.$$

Here ε_t is Gaussian white noise with mean zero and standard deviation 0.1. Auestad and Tjøstheim checked geometric ergodicity and stationarity for these examples. Inspection of the proofs of our results shows that the differentiability conditions imposed on λ (f and γ) can be weakened to hold piecewise, so they also apply to the threshold model.

We restricted estimation of Λ to the interval $[-0.3, 0.3]$. The bin width was taken as $w_n = 0.05$ [as in Auestad and Tjøstheim (1990), who plotted point estimates of λ for these three models]. Inspecting the plots of $\widehat{\Lambda}$ 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 conditional mean 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

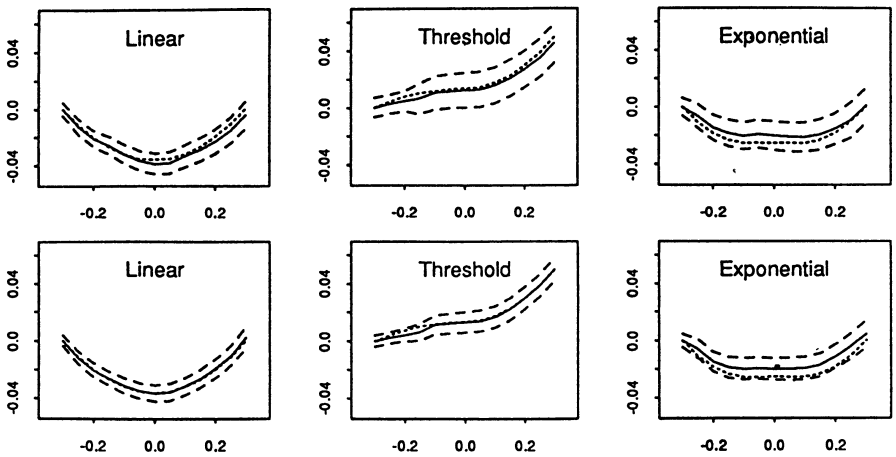

FIG. 1. 95% *confidence bands (dashed lines) for* Λ; solid-lines, Λ; *dotted lines*, Λ; *first row*, $n = 250$; *second row*, $n = 500$.

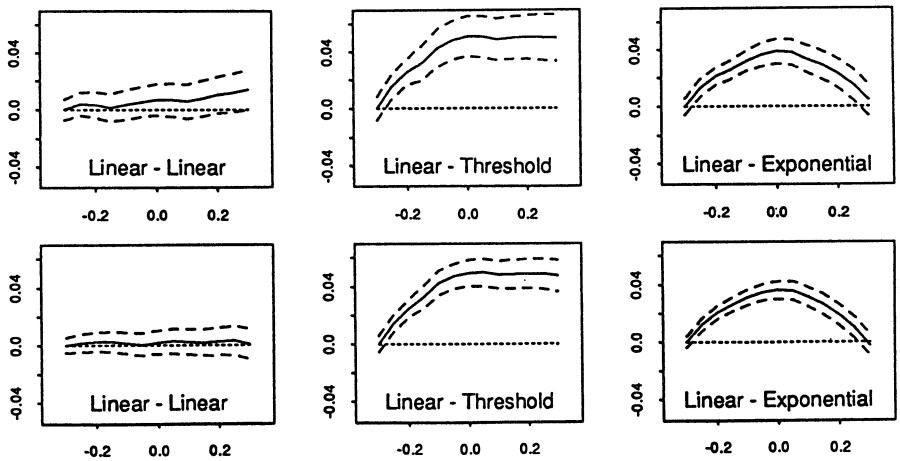

FIG. 2. 95% *confidence bands for* $\Lambda_1 - \Lambda_2$; *first row*, $n = 250$; *second row*, $n = 500$.

TABLE 1
Observed levels and powers of goodness-of-fit test for linear autoregressive model at nominal level of 5%; bin width $w_n = 0.05$; $q = 4$

	Sample size					
Observed series	**100**	**250**	**500**	**1000**	**1500**	**2500**
Linear	0.3096	0.0728	0.0590	0.0628	0.0588	0.0740
Threshold	0.9690	0.9996	1.0000	1.0000	1.0000	1.0000
Exponential	0.6762	0.9498	0.9992	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.

using the linear model and the zero function is contained within the band, so our test would correctly conclude that the conditional mean functions are identical. In the other plots, the zero function is well outside the bands and the test correctly concludes that the conditional mean functions are different.

Table 1 gives observed levels and powers of the chi-squared goodness-of-fit test for the linear autoregressive model $X_t = \theta X_{t-1} + \varepsilon_t$, when the time series is generated by each model. At small sample sizes (less than 250), the covariance matrix estimator $\widehat{G}$ sometimes failed to be positive definite and the chi-squared statistic value was negative. The percentage of negative chi-squared statistics was 8 and 0.2% for sample sizes of 100 and 250 with the linear model; 29 and 6% with the threshold model; 25 and 6% with the exponential model. We rejected the linear model when the chi-squared statistic was negative. This is reasonable since $\widehat{G}$ 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-western 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

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

where $\varepsilon_t \sim N(0, 0.046)$. 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 (having three parameters) using $d_n = 5, 6, \ldots, 10$, and 4 (2 by 2) and 9 (3 by

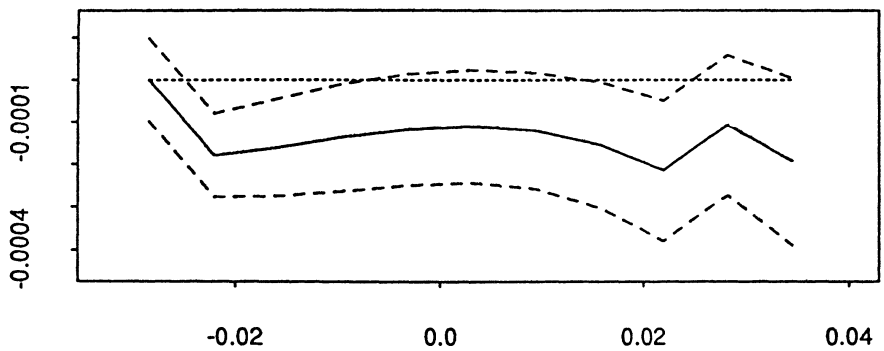

FIG. 3. *95% confidence band for* $\Lambda_1 - \Lambda_2$, *IBM stock price data*; $d_n = 10$; Λ_1 = *period I*, Λ_2 = *period II*.

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. In order to apply the two-sample test we need to assume that the two subseries are independent (or approximately so); this is not unreasonable since periods I and II are well separated in time. Figure 3 gives a plot of the difference between the estimates of the cumulative conditional mean 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 one-step lagged conditional means for the two periods differ significantly from one another.

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

$$E\left(\sup_{x\in[a,b]} |\widehat{f}(x) - f^{\flat}(x)|\right)^2 \le \sum_{j=1}^{d_n} \operatorname{var}[\widehat{f}(x_j)] = d_n o(w_n) \to 0,$$

and, by stationarity, $f^{\flat}(x) \equiv E\widehat{f}(x) = w_n^{-1} \int_{\mathcal{I}_x} f(u)\, du \to f(x)$, uniformly on $[a,b]$. Note that uniform consistency of $\widehat{f}$ and (A2) allow us to divide by $\widehat{f}$ (which we often do) when proving the asymptotic results. Also note that $\xi_t \equiv X_t - \lambda(X_{t-1})$ is a martingale difference with respect to the natural filtration $\mathcal{F}_t =$

$\sigma(X_0, \ldots, X_t)$.

PROOF OF THEOREM 2.1. We begin by splitting $\widehat{\lambda}$ into two parts, the first of which turns out to be close to λ and the second of which involves the martingale differences ξ_t:

$$\widehat{\lambda}(x) = \frac{\widehat{\alpha}(x)}{\widehat{f}(x)} + [nw_n\widehat{f}(x)]^{-1}\sum_{t=1}^{n} I\{X_{t-1} \in \mathcal{I}_x\}\xi_t,$$

where

$$\widehat{\alpha}(x) = [nw_n]^{-1}\sum_{t=1}^{n} I\{X_{t-1} \in \mathcal{I}_x\}\lambda(X_{t-1}).$$

Upon integrating,

$$\sqrt{n}(\widehat{\Lambda} - \Lambda)(z) = \overline{M(n,\cdot)}(z) + R_1(z) + R_2(z),$$

where

$$M(k,z) = \frac{1}{\sqrt{n}}\sum_{j=1}^{L(z)} f^{\flat}(x_j)^{-1}\sum_{t=1}^{k} I\{X_{t-1} \in \mathcal{I}_j\}\xi_t, \qquad k = 1, \ldots, n,$$

$$R_1(z) = \frac{1}{\sqrt{n}w_n}\int_a^z \left[\frac{f^{\flat}(x) - \widehat{f}(x)}{\widehat{f}(x)f^{\flat}(x)}\right]\sum_{t=1}^{n} I\{X_{t-1} \in \mathcal{I}_x\}\xi_t\, dx,$$

$$R_2(z) = \sqrt{n}\int_a^z \left[\frac{\widehat{\alpha}(x)}{\widehat{f}(x)} - \lambda(x)\right] dx,$$

and, given a function ϕ defined on $[a,b]$, $\bar{\phi}$ is the piecewise linear approximation to ϕ that agrees with ϕ at each x_j. Here, $L(z)$ is the integer part of $(z-a)/w_u$ and $M(k,z)$ is defined to be zero when $L(z) = 0$. To complete the proof, we need to show that the remainder terms R_1 and R_2 converge uniformly in probability to zero and $M(n,\cdot) \longrightarrow_{\mathcal{D}} m$, where m denotes the Gaussian martingale given in the statement of the theorem; Lemma 4.1 of McKeague (1988) then implies that $\overline{M(n,\cdot)} \longrightarrow_{\mathcal{D}} m$.

Now $M(\cdot,z)$ is an $\mathcal{F}_t$ martingale for each fixed z. We shall use the martingale central limit theorem [see, e.g., Theorem A.2 of Aalen (1977)] 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

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

Next we check the Lindeberg condition:

$$L_n \equiv \frac{1}{n}\sum_{j=1}^{L(z)} f^\flat(x_j)^{-2}\sum_{t=1}^{n} I\{X_{t-1}\in\mathcal{I}_j\} \times E\left\{\xi_t^2 I\left(\frac{I\{X_{t-1}\in\mathcal{I}_j\}|\xi_t|}{\sqrt{n}f^\flat(x_j)}>\varepsilon\right)\Big|\mathcal{F}_{t-1}\right\}$$

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

$$\left\{E(\xi_t^4|\mathcal{F}_{t-1})\right\}^{1/2}\left\{[\sqrt{n}\varepsilon f^\flat(x_j)]^{-2}E(I\{X_{t-1}\in\mathcal{I}_j\}\xi_t^2|\mathcal{F}_{t-1})\right\}^{1/2} = O\left(\frac{1}{\sqrt{n}}\right)\{E(\xi_t^4|\mathcal{F}_{t-1})\}^{1/2}I\{X_{t-1}\in\mathcal{I}_j\}\gamma(X_{t-1})^{1/2}.$$

Now (A1), stationarity of $\{X_t\}$ and λ Lipschitz imply that $\sup_t E\xi_t^8 < \infty$, so again using the Cauchy–Schwarz inequality, (A2), boundedness of f and γ and $nw_n^2 \to \infty$, we have

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

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

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

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

$$E|M(n,y)-M(n,x)|^4 \le C(y-x)^2 + o(1). \tag{6.1}$$

Using Rosenthal's inequality [Hall and Heyde (1980), page 23], the left-hand side of (6.1) is bounded by

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

where the summation over j runs from $L(x)+1$ to $L(y)$. By (A2), the first term of (6.2) is bounded by

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

and the second term of (6.2) is bounded by

$$\begin{aligned} &O\left(\frac{1}{n^2}\right)\sum_j\sum_{t=1}^n\big(EI\{X_{t-1}\in\mathcal{I}_j\}\big)^{1/2}\big(E\xi_t^8\big)^{1/2}\\ &\quad\le O\left(\frac{1}{n\sqrt{w_n}}\right)(y-x)\longrightarrow 0, \end{aligned}$$

since $nw_n^2\to\infty$. So (6.1) holds.

Next we show that R_1 converges uniformly in probability to zero. Since $\widehat{f}$ is a uniformly consistent estimator of f, which is bounded away from zero on $[a,b]$, it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{d_n}\big|\widehat{f}(x_j)-f^{\flat}(x_j)\big|\left|\sum_{t=1}^n I\{X_{t-1}\in\mathcal{I}_j\}\xi_t\right|\longrightarrow_P 0. \tag{6.3}$$

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

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

proving (6.3).

Finally, we need to show that R_2 converges uniformly in probability to zero. First note that Taylor expanding λ about x inside each term of $\widehat{\alpha}$ gives $\widehat{\alpha}=\lambda\widehat{f}+\lambda'\widehat{\delta}+O_P(w_n^2)$, where λ' is the first derivative of λ. Thus

$$\frac{\widehat{\alpha}}{\widehat{f}}-\lambda=\frac{\lambda'\big(\widehat{\delta}-\delta^{\flat}\big)}{\widehat{f}}+\frac{\lambda'\delta^{\flat}}{\widehat{f}}+O_P\big(w_n^2\big),$$

where $\delta^{\flat}(x)\equiv E\widehat{\delta}(x)=w_n^{-1}\int_{\mathcal{I}_x}(u-x)f(u)\,du$. We need to show that the integral from a to z of each term on the rhs is of order $O_P(n^{-\frac{1}{2}})$ uniformly in z. For the first term,

$$\sup_z\left|\int_a^z\frac{\lambda'\big(\widehat{\delta}-\delta^{\flat}\big)}{\widehat{f}}\,dx\right|\le O_P(1)\int_a^b\big|\widehat{\delta}-\delta^{\flat}\big|\,dx=o_P\big(n^{-1/2}\big),$$

since $\widehat{f}$ is uniformly consistent for f (which is bounded away from zero) and

$$E\int_a^b |\widehat{\delta} - \delta^\flat|\, dx \le \int_a^b (\operatorname{var}[\widehat{\delta}(x)])^{1/2}\, dx = o(n^{-1/2})$$

by condition (A3). For the second term, Taylor expanding λ' and f about the midpoint of each interval $\mathcal{I}_j$, noting that $\widehat{f}$ is constant over each interval and $\int_{\mathcal{I}_j}\int_{\mathcal{I}_j}(u-x)\, du\, dx = 0$ leads to

$$\int_a^z \frac{\lambda'\delta^\flat}{\widehat{f}}\, dx = O_P(w_n^2), \tag{6.4}$$

uniformly in z, which is of order $o_P(n^{-\frac{1}{2}})$ since $nw_n^4 \to 0$. □

PROOF OF THEOREM 2.2. Since λ and γ are Lipschitz, and $\widehat{f}$ uniformly converges in probability to f, which is bounded away from 0,

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

Let $\tau_t = \xi_t^2 - \gamma(X_{t-1})$ and define $\widehat{\eta}(x)$ to be $\widehat{\alpha}(x)$ in the proof of Theorem 2.1, but with λ replaced by γ. Then, expanding the above expression from $\widehat{\gamma}$, we obtain

$$\begin{aligned}\sqrt{n}(\widehat{\Gamma}-\Gamma)(z) = &\frac{1}{\sqrt{n}w_n}\int_a^z \frac{\sum_t I\{X_{t-1}\in\mathcal{I}_x\}\tau_t}{\widehat{f}(x)}\, dx\\ &+\sqrt{n}\int_a^z\left[\frac{\widehat{\eta}(x)}{\widehat{f}(x)} - \gamma(x)\right] dx\\ &+O_P\left(\frac{1}{n^{3/2}w_n^2}\right)\int_a^z\left[\sum_{t=1}^n I\{X_{t-1}\in\mathcal{I}_x\}\xi_t\right]^2 dx\\ &+O_P\left(\frac{1}{\sqrt{n}}\right)\int_a^z\sum_{t=1}^n I\{X_{t-1}\in\mathcal{I}_x\}\xi_t\, dx + O_P\left(\sqrt{n}w_n^2\right),\end{aligned} \tag{6.5}$$

uniformly in z. The first term in (6.5) has the same form as the leading term in the decomposition of $\sqrt{n}(\widehat{\Lambda}-\Lambda)$ except that τ_t replaces ξ_t. Note that τ_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 first term of (6.5) converges to the desired limiting distribution by the proof of Theorem 2.1. The second term in (6.5) has the same form as R_2 in the proof of Theorem 2.1, so it converges uniformly in probability to zero. The third term in (6.5) is uniformly bounded

by

$$O_P\Big(\frac{1}{n^{3/2}w_n}\Big)\sum_{j=1}^{d_n}\Big[\sum_{t=1}^{n} I\{X_{t-1}\in\mathcal{I}_j\}\xi_t\Big]^2$$
$$= O_P\Big(\frac{1}{n^{3/2}w_n^2}\Big)O_P(nw_n) = O_P\Big(\frac{1}{\sqrt{nw_n^2}}\Big) \longrightarrow_P 0,$$

since $nw_n^2 \to \infty$. The fourth term in (6.5) is uniformly bounded by

$$O_P\Big(\frac{w_n}{\sqrt{n}}\Big)\sum_{j=1}^{d_n}\Big|\sum_{t=1}^{n} I\{X_{t-1}\in\mathcal{I}_j\}\xi_t\Big| \le O_P\Big(\frac{1}{\sqrt{n}}\Big)O_P(\sqrt{nw_n}) \longrightarrow_P 0.$$

This completes the proof. □

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

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

By condition (B1), we have that

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

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

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

where C is finite almost surely. It follows that $\{n^{-1}Q_n(\cdot)\}$ is equicontinuous. Again by the ergodic theorem, $n^{-1}Q_n(\theta) \longrightarrow q(\theta)$ $(< \infty)$, a.s., 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 Θ. Thus $n^{-1}Q_n(\cdot)$ converges uniformly to $q(\cdot)$ on Θ almost surely. Since $q(\theta)$ has a unique minimum at $\theta_0 \in \Theta$, and $\widetilde{\theta}$ minimizes $Q_n(\theta)$, we conclude that $\widetilde{\theta}$ is consistent.

Next, Taylor expanding Q_n' about θ_0, we can write

$$\sqrt{n}(\widetilde{\theta} - \theta_0) = V_n(\theta^*)^{-1}U_n,$$

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

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

and θ^* is on the line joining θ_0 and $\widetilde{\theta}$. Since $U_k^{(n)}$ is a martingale in k, the martingale central limit theorem can be used to show that $U_n \longrightarrow_{\mathcal{D}} N(0,S)$ under the moment conditions in (B3). To complete the proof, we need to show that $V_n(\theta^*) \longrightarrow_P V$. Routine algebra gives that

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

By (B3) and the ergodic theorem, the first term converges to V almost surely. Using $\theta^* \longrightarrow \theta_0$ a.s., 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 ξ_t, give that the fourth term also converges almost surely to zero. We conclude that $V_n(\theta^*) \longrightarrow V$ a.s. □

PROOF OF THEOREM 3.2. By Taylor expanding $g(\cdot, u)$ about θ_0 for each fixed u,

$$\sqrt{n}(\widetilde{\Lambda} - \Lambda)(z) = \left(\int_a^z g'(\theta_x^*, x)\, dx\right)^T \sqrt{n}(\widetilde{\theta} - \theta_0),$$

where θ_x^* lies on the line joining θ_0 and $\widetilde{\theta}$. Since $\widetilde{\theta}$ is a consistent estimator of θ_0, and g' is continuous,

$$\int_a^z g'(\theta_x^*, x)\, dx \longrightarrow_P \int_a^z g'(\theta_0, x)\, dx.$$

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

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

uniformly in z. By a $D[a,b] \times \mathbb{R}$ version of McKeague [(1988), Lemma 4.1] 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 $M(n,\cdot) \longrightarrow_{\mathcal{D}} m$ and $U_n \longrightarrow_{\mathcal{D}} U_\infty$. 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

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

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. □

PROOF OF THEOREM 4.1. The proof runs along the lines of the proof of Theorem 2.1, except that $\mathbf{X}_t$ replaces $X_t, \mathcal{I}_{xy}$ replaces $\mathcal{I}_x$, double integral (summation) replaces single integral (summation) and $\overline{\phi}$ is the piecewise linear approximation to ϕ determined by cells $\mathcal{I}_{xy}$. Note that $nw_n^2 \to \infty$ 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 (1968) given in Bickel and Wichura (1971). The analogue of (6.4) is

$$\int_a^{z_1}\int_a^{z_2} \frac{\lambda_x' E\widehat{\delta}_x}{\widehat{f}}\, dx\, dy = O_P(w_n^2) \tag{6.6}$$

(uniformly in z_1, z_2) along with a similar property for λ_y' and $E\widehat{\delta}_y$. Here λ_x' and λ_y' are the partial derivatives of λ with respect to x and y; (6.6) is obtained by Taylor expanding λ_x' and f about the midpoint of each cell. □

PROOF OF THEOREM 4.2. The proof is similar to that 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 stationaires. *Publ. Inst. Statist. Univ. Paris* **3** 119–134.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. 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 (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 411–431; 448–468.

CASTELLANA, J. V. and LEADBETTER, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.* **21** 179–193.

DIEBOLT, J. (1990). Testing the functions defining a nonlinear autoregressive time series. *Stochastic Process. Appl.* **36** 85–106.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic, 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. Statist. Comput.* **5** 349–359.

MCKEAGUE, I. W. (1988). A counting process approach to the regression analysis of grouped survival data. *Stochastic 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. *Australian Journal of Zoology* **1** 163–173.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge Univ. Press.

OZAKI, T. (1980). Non-linear time series model for non-linear random vibrations. *J. Appl. Probab.* **17** 84–93.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

TONG, H. (1977). Some comments on the Canadian lynx data—with discussion. *J. Roy Statist. Soc. Ser. A* **140** 432–436; 448–468.

TONG, H. (1983). *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statist.* **21.** Springer, New York.

TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon, Oxford.

TUKEY, J. (1961). Curves as parameters and touch estimation. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 681–684. Univ. California Press, Berkeley.

TWEEDIE, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing theory. In *Papers in Probability, Statistics and Analysis* (J. F. C. Kingman and G. E. H. Reuter, eds.) 260–276. Cambridge Univ. Press.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306-3033

INTERNATIONAL BONE MARROW
TRANSPLANT REGISTRY
MEDICAL COLLEGE OF WISCONSIN
P.O. BOX 26509
MILWAUKEE, WISCONSIN 53226