Statistica Sinica **16**(2006), 683-696

INFERENCE FOR TIME SERIES AND STOCHASTIC PROCESSES

Ngai Hang Chan

Chinese University of Hong Kong

Abstract: This paper reviews Ching-Zong Wei's research in time series and stochastic processes. It gives an overview of Ching-Zong Wei's contributions in the general area of time series and related fields and describes the impact of Wei's work on subsequent research.

Key words and phrases: Asymptotic inference, model selection, nonstationary time series, prediction principle and strong consistency.

1. Introduction

Ching-Zong Wei has made many important contributions to statistics and other areas, ranging from sequential analysis and stochastic approximation to time series and systems control. This article mainly focuses on his contributions to the area of time series and related fields. Wei received his formal training in statistics from Columbia University. Naturally, his research interest reflects the strong tradition of Columbia University. Wei's contributions in time series manifest in the typical fashion of a classical statistician. Wei possessed a unique ability of applying powerful probabilistic tools to tackle convoluted problems from which elegant and profound solutions were developed. Because of his talent, some people consider Wei more of a probabilist than a statistician. As will be demonstrated in the following section, Wei's contributions have many deep and far-reaching impacts that lie well beyond probability; he was always strongly motivated by applications.

This paper is organized as follows. In Subsections 2.1 to 2.3, we review Wei's work according to the categories of estimation, inference and model selection. Here, we also provide a road map for certain portions of the extensive literature on these topics, and the impact of Wei's work in related areas. Wei's contributions are not restricted to linear time series. In Section 3, his contributions to nonlinear time series and other areas such as branching processes are discussed. Section 3 also contains ideas on new developments and some concluding remarks.

2. Inference

Wei's research in time series can be classified into three stages: estimation, inference and model selection. Throughout these stages, there are some underlying common themes. The first one, perhaps the most important, is the issue of on-line recursion. As will be seen in the following subsections, one of the guiding principles in his work is to construct an on-line inferential framework. The second theme of Wei's work is the order of the information contained in the model. Such an idea, coupled with local martingale convergence theory, constitute most of the building blocks of his work. But Wei's contributions do not cease at the mathematical level. He has always advocated the importance of furnishing "life" to the theory, which means that one has to articulate the essence of the theory in statistical or non-mathematical terms. These themes are reflected throughout his work.

2.1. Estimation

The seminal paper of Lai, Robbins and Wei (1978) considered the strong consistency property of the least squares estimate of a multiple regression model. This result extended the earlier work of Anderson and Taylor (1979) and constituted the turning point of the study of strong consistency in the regression literature. Since then, Wei and his co-authors in a series of paper (see Lai, Robbins and Wei (1978), Chen, Lai and Wei (1981), Lai and Wei (1982a, 1982b, 1982c)) extended strong consistency in various directions. A detailed account on these developments, in connection with stochastic approximation, is given in Lai (2003).

One of the major breakthroughs in the consistency problem was given in Lai and Wei (1982a), where a minimal sufficient condition for the least squares estimate of a stochastic regression model to be strongly consistent was provided. Specifically, consider the stochastic regression model

$$y_n = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_n + \epsilon_n, \ n = 1, 2, \ldots, \tag{1}$$

where the ϵ_n are unobservable errors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ are unknown parameters and y_n is the observable response corresponding to the design vector $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{np})^{\mathrm{T}}$. Then

$$\boldsymbol{b}_n = (\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}})^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \tag{2}$$

is the least squares estimate (LSE) of the unknown parameter vector $\boldsymbol{\beta}$ based on the observations $\boldsymbol{x}_1, y_1, \ldots, \boldsymbol{x}_n, y_n$. Herein, the unobservable sequence $\{\epsilon_n\}$ is assumed to be a martingale difference sequence with respect to an increasing sequence of sigma fields $\mathcal{F}_n$ satisfying a Lyapunov condition

$$\sup_n E(|\epsilon_n|^{\gamma} | \mathcal{F}_{n-1}) < \infty \text{ almost surely (a.s.) for some } \gamma > 2. \tag{3}$$

By assuming that the design vector at stage n is adaptable, i.e., $\boldsymbol{x}_n$ is $\mathcal{F}_{n-1}$ measurable, Lai and Wei (1982a), proved the following.

Theorem 2.1. $\boldsymbol{b}_n \to \boldsymbol{\beta}$ *a.s. if*

$$\lambda_{\min}(n) \to \infty \text{ a.s. and } \log \lambda_{\max}(n) = o(\lambda_{\min}(n)) \text{ a.s.}, \tag{4}$$

where $\lambda_{\min}(n)$ *and* $\lambda_{\max}(n)$ *denote respectively the minimum and the maximum eigenvalues of the design matrix* $\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}$ *at stage* n.

Lai and Wei (1982a) also showed that, without further assumptions, condition (4) is the best possible one in the sense that they constructed an example in which $\lambda_{\min}(n) \to \infty$ a.s. and $\log \lambda_{\max}(n)/\lambda_{\min}(n)$ converges to a positive random variable, but $P(\boldsymbol{b}_n \to \boldsymbol{\beta}) = 0$.

Condition (4) is quite intriguing and fails to hold in many applications in stochastic approximation and systems control, see for example Lai and Wei (1987). In view of this deficiency, Wei (1985) conducted a refined analysis of the problem and came up with the result that if there exists a linear transformation $\boldsymbol{A}$ such that the transformed design vectors $\boldsymbol{z}_n = \boldsymbol{A}\boldsymbol{x}_n$ satisfy

$$\liminf_{n\to\infty} \Big(\boldsymbol{D}_n^{-1} \sum_{i=1}^n \boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{D}_n^{-1}\Big) > 0 \text{ where } \boldsymbol{D}_n = \Big\{\mathrm{diag}\Big(\sum_{i=1}^n \boldsymbol{z}_i {\boldsymbol{z}_i}^{\mathrm{T}}\Big)\Big\}^{1/2}, \tag{5}$$

then $\boldsymbol{b}_n$ is still strongly consistent. The following result ensued.

Theorem 2.2. *Suppose that in the regression model* (1), *condition* (3) *holds. Assume that* $\lambda_{\min}(n) \to \infty$ *and*

$$\log \lambda_{\max}(n)^{2\delta} = o(\lambda_{\min}(n)) \text{ a.s. for some } \delta > (\min(\gamma, 4))^{-1}. \tag{6}$$

Suppose that there exists a nonsingular matrix $\boldsymbol{A}$ *such that the random vectors* $\boldsymbol{z}_n = \boldsymbol{A}\boldsymbol{x}_n$ *satisfy* (5). *Then* $\boldsymbol{b}_n \to \boldsymbol{\beta}$ *a.s.*

If the transformed design vectors satisfy an additional assumption (see Wei (1985)), then (6) can further be weakened to the condition

$$\lambda_{\min}(n) \to \infty \text{ a.s. and } \log\log \lambda_{\max}(n) = o(\lambda_{\min}(n)) \text{ a.s. }, \tag{7}$$

in which $\boldsymbol{b}_n$ is still strongly consistent.

As evidenced in this development, Wei made significant contributions to the estimation of stochastic regression models. One of the key elements to the estimation problem is to assess the order of magnitude of the design matrix $\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}$. To understand this order, Wei made use of the recursive nature of the least squares estimate from which a martingale transform structure was uncovered. By analyzing this structure in detail, Wei was able to decompose this quantity into different components and to apply the local martingale convergence result of Chow to evaluate the order of the design matrix. It turns out that this consideration has an impact far beyond estimation.

Because of its relevance to systems control, different extensions of Wei's result have been carried out. Lai and Wei applied the consistency idea to study the properties of self-tuning regulators and adaptive controllers (see Lai and Wei (1986) and Lai and Wei (1987)). The monograph of Guo (1990) provides a succinct review of various strong consistency issues in systems control engineering. In time series, Lai and Wei (1982c and 1985) established the strong consistency of the least squares estimate for a general AR model, irrespective to the location of its characteristic roots, while Tiao and Tsay (1983) considered the weak consistency property of the least squares estimate for a nonstationary ARMA model. Recently, Nielsen (2005) extended Lai and Wei's result to the general VAR model with deterministic components.

2.2. Limiting distributions

Back in 1958, White (1958) amongst others, first showed that the limiting distribution of the least squares estimate of the autoregressive coefficient of a nonstationary AR(1) model, i.e., when the autoregressive coefficient equals to one, is a functional of a stochastic integral of a standard Brownian motion. It turns out that this model has a strong bearing in the econometric literature in testing whether a time series is a random walk, the so-called unit root testing problem. After Dickey and Fuller (1979) established the form of this limiting distribution as a ratio of sums of i.i.d. random variables, the unit root testing problem became a topical issue in econometrics. Numerous articles were written on this and an elementary survey on this literature was given in Stock and Watson (1988).

Having built the groundwork in estimation, the next stage of Wei's work concerned the issue of asymptotic inference. He was interested in the limiting distributions of the least squares estimates of the parameters of a general nonstationary autoregressive model. Chan and Wei (1987) first considered the AR(1) model when the autoregressive coefficient converges to one asymptotically. Instead of a Brownian motion, Chan and Wei (1987) showed that the limiting distribution of the least squares estimate of a nearly nonstationary AR(1) model converges weakly to a functional of an Ornstein-Uhlenbeck process. Specifically, consider a first-order autoregressive process

$$y_{t,n} = \beta_n y_{t-1,n} + \epsilon_t, \quad t = 1, \cdots, n, \tag{8}$$

where $\beta_n = 1 - \gamma/n$, γ is a real number, $y_{0,n} = 0$ for all n, and $\{\epsilon_t\}$ is a martingale difference sequence satisfying (3). This is known as the nearly nonstationary or near-integrated (see Phillips (1987)) time series. If $\gamma = 0$, (8) reduces to the traditional unit root model.

Theorem 2.3. *Let the time series* $\{y_{t,n}\}$ *follow* (8) *with the innovation sequence* $\{\epsilon_t\}$ *satisfying* (3). *Let the least squares estimate of* β_n *be* $b_n = (\sum_{t=1}^n y_{t-1,n}\epsilon_t)/(\sum_{t=1}^n y_{t-1,n}^2)$. *Then as* $n \to \infty$,

$$n(b_n - \beta_n) \to_D \mathcal{L}(\gamma) := \frac{\int_0^1 X(t)\, dX(t)}{\int_0^1 X^2(t)\, dt},$$

where $\to_D$ *denotes convergence in distribution and* $X(t)$ *is the Ornstein-Uhlenbeck process satisfying the diffusion equation*

$$dX(t) = -\gamma X(t)\, dt + dW(t),$$

$X(0) = 0$ *and* $W(t)$ *is a standard Brownian motion.*

This particular result encompasses the unit-root case of White (1958) when $\gamma = 0$. In this case, the autoregressive coefficient $\beta_n = 1$ and $\mathcal{L}(0) = \int_0^1 W(t)dW(t) / \int_0^1 W^2(t)dt$, which is the limitng distribution of the Dickey-Fuller statistic. Using reproducing kernels, Chan (1988) further developed this limiting form as sums of i.i.d. random variables. The near-integrated notion was formulated with reference to the work of LeCam about limiting experiments in terms of contiguous alternatives. This idea was later explored by Jeganathan (1991, 1995), who generalized the near-integrated notion to a general AR(p) case and introduced the idea of Local Asymptotic Brownian Functional in studying optimality issues. In a spectral setting, Dahlhaus (1985) considered both tapered and non-tapered Yule-Walker estimates for near-integrated models. Since then, numerous extensions have been carried out by econometricians and statisticians alike.

Chan and Wei (1988) considered the limiting distributions of the least squares estimate of a general nonstationary AR(p) model when the characteristic roots lie on or outside the unit circle, each of which may have different multiplicities. This was the first comprehensive treatment of the LSE for a general nonstationary AR(p) model, and it was shown in Chan and Wei (1988) that the locations of the roots of the time series played an important role in characterizing the limiting distributions. Specifically, they considered a general nonstationary AR(p) model

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \epsilon_t. \tag{9}$$

In (9), the autoregressive polynomial $\beta(z) = (1 - \beta_1 z - \cdots - \beta_p z^p)$ has roots lying on or outside the unit circle. That is,

$$\beta(z) = (1-z)^a (1+z)^b \prod_{k=1}^{\ell} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z), \tag{10}$$

where a, b, ℓ, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is a polynomial of order $r = p - (a + b + 2d_1 + \cdots + 2d_k)$ that has all roots outside the unit disk.

When the underlying model is stationary with all roots lying outside the unit circle, classical central limit theorem type results can be obtained. But when the roots are of unit modulus, it turns out that the asymptotic distributions are characterized in terms of iterated integrals of Brownian motions. The key idea in obtaining these results lies in analyzing the order of magnitude of the observed Fisher's information matrix. Note that the least squares estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ can be expressed as

$$\boldsymbol{b}_n = \Big(\sum_{t=1}^{n} \boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}{}^{\mathrm{T}}\Big)^{-1} \sum_{t=1}^{n} \boldsymbol{y}_{t-1} y_t, \tag{11}$$

where $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})^{\mathrm{T}}$ and $\boldsymbol{y}_0 = (0, \ldots, 0)^{\mathrm{T}}$. Similar to the estimation problem, different characteristic roots carry different information. By transforming the original nonstationary AR model into components according to their characteristic roots, Chan and Wei (1988) was able to derive the precise form of the limiting distributions. During the course of this investigation, they also obtained an important result about the weak convergence of stochastic integrals (Theorem 2.4 in Chan and Wei (1988)) which is of indepedent interest and has many applications in different areas, see for example Kurtz and Protter (1991). In addition, Chan and Wei also showed that different components are asymptotically uncorrelated and, as a result, a joint limiting law can be established. Specifically, using the notations of Chan and Wei (1988), the following theorem was established.

Theorem 2.4. *Assume that* $\{y_t\}$ *follows* (9) *with the characteristic polynomial satisfying* (10) *and the innovation sequence* $\{\epsilon_t\}$ *satisfying* (3). *Then as* $n \to \infty$,

$$Q^{\mathrm{T}} G_n^{\mathrm{T}}(\boldsymbol{b}_n - \boldsymbol{\beta}) \to_D \ ((F^{-1}\boldsymbol{\xi})^{\mathrm{T}}, (\tilde{F}^{-1}\boldsymbol{\eta})^{\mathrm{T}}, (H_1^{-1}\boldsymbol{\zeta}_1)^{\mathrm{T}}, \ldots, (H_\ell^{-1}\boldsymbol{\zeta}_\ell)^{\mathrm{T}}, N^{\mathrm{T}})^{\mathrm{T}},$$

where $(F, \boldsymbol{\xi}), (\tilde{F}, \boldsymbol{\eta}), (H_1, \boldsymbol{\zeta}_1), \ldots, (H_\ell, \boldsymbol{\zeta}_\ell), N, G_n, Q$ *are independent and defined in equations* (3.2), (3.3) *and Theorem* 2.2 *of Chan and Wei* (1988).

This result of Chan and Wei paved the way to the analysis of nonstationary processes and, since then, numerous extensions have been conducted. Jeganathan (1991) generalized this idea to the near-integrated situations where the limiting distributions of the LSE are expressed in terms of iterated integrals of Ornstein-Uhlenbeck processes. When the underlying model has long-memory, Chan and Terrin (1995) extended this result to functionals of fractional Brownian motions, while Ling and Li (1998) considered the case when the innovations are modeled by GARCH processes. Extensions of this result to vector AR processes are given in Tsay and Tiao (1990) and to processes with deterministic trends in Chan (1989). On the econometric front, Theorem 2.4 provides a fundamental tool in analyzing co-integrated systems. A comprehensive review on co-integrated vector autoregressions is given in Johansen (1995).

Beyond limiting distributions, Wei was also interested in residual analysis. Lee and Wei (1999) considered the stochastic regression model (1)

$$y_{nt} = \boldsymbol{\beta}_n^{\mathrm{T}} \boldsymbol{x}_{nt} + r_{nt} + \epsilon_{nt}, \quad 1 \le t \le n, \tag{12}$$

where $\boldsymbol{\beta}_n$ are unknown parameters, $\boldsymbol{x}_{nt}$ are observable random vectors and r_{nt} are random variables which they called "model bias." This model can be considered an extension of (1) as it encompasses both stochastic regressions and autoregressive time series. Let $\boldsymbol{b}_n$ denote the least squares estimate of $\boldsymbol{\beta}_n$ by regressing y on $\boldsymbol{x}$ ignoring r, and let the residual be defined as $\tilde{\epsilon}_{nt} = y_{nt} - \boldsymbol{b}_n^{\mathrm{T}} x_{nt}$. Consider the residual empirical process

$$\hat{Y}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} [I(H_n(\tilde{\epsilon}_{nt}) \le u) - u], \tag{13}$$

where H_n is the underlying distribution of $\{\epsilon_{nt}\}$. Under certain regularity conditions on H_n and the growth rates of the orders of the model, Lee and Wei (1999) showed that for a Gaussian stationary AR(∞) model, under the setting of a null hypothesis $K_0 : H(\cdot) = \Phi(\cdot)$ and a contiguous sequence of alternatives $K_n : H_n(\cdot) = (1 - \gamma/\sqrt{n})\Phi(\cdot) + (\gamma/\sqrt{n})H(\cdot)$, where H is a distribution function with mean zero and variance one, the following theorem holds.

Theorem 2.5. *Under K_n, the residual empirical process $\hat{Y}_n(u)$ defined in* (13) *converges weakly to a Gaussian process Y with mean and covariance*

$$\begin{aligned} \mathrm{E}Y(u) &= -\gamma(u - H \circ \Phi^{-1}(u)) \\ \mathrm{Cov}(Y(u), Y(v)) &= u \wedge v - uv - 0.5\phi(\Phi^{-1}(u))\Phi^{-1}(u)\phi(\Phi^{-1}(v))\Phi^{-1}(v), \end{aligned}$$

where $0 \le u, v \le 1$. Here Φ and ϕ denote the cumulative distribution function and the density function of a standard normal random variable, respectively.

In particular, for nonstationary AR(p) models, following the notations used in Lee and Wei (1999), let (W_1, W_2) be a mean zero two-dimensional Gaussian process with covariance structure such that for all $s, t \in [0, 1]$:

$$\begin{aligned} &\mathrm{Cov}(W_1(s), W_1(t)) = s \wedge t - st, \\ &\mathrm{Cov}(W_2(s), W_2(t)) = s \wedge t, \\ &\mathrm{Cov}(W_1(s), W_2(t)) = (t/\sigma) \int_{-\infty}^{G^{-1}(s)} x \, dG(x), \end{aligned} \tag{14}$$

where G is the distribution function of the innovation sequence $\{\epsilon_t\}$ in (9). The following weak convergence result of the residual empirical processes was established in Lee and Wei (1999).

Theorem 2.6. *Consider a nonstationary AR(p) model satisfying* (9) *with a characteristic root of one with multiplicity $a \geq 1$, and an i.i.d. innovation sequence $\{\epsilon_t\}$ with mean zero, variance $0 < \sigma^2 < \infty$ and continuous distribution function G. Then as $n \to \infty$,*

$$\hat{Y}_n(u) \to_D \; W_1(u) + \sigma(F^{-1}\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{\eta}G^{\mathrm{T}}(G^{-1}(u)),$$

where (W_1, W_2) is the two-dimensional Gaussian process defined in (14); $F_0 = \sigma W_2$, $F_1 = \int_0^1 F_0(s)ds$, $F_j = \int_0^1 F_{j-1}(s)ds$, $j = 2, \ldots, a$; $\boldsymbol{\xi} = (\int_0^1 F_{a-1}(s)dW_2(s), \ldots, \int_0^1 F_0(s)dW_2(s))^{\mathrm{T}}$; $\boldsymbol{\eta} = (F_a(1), \ldots, F_1(1))^{\mathrm{T}}$; *and F is the matrix whose (j, l)th entry is $\sigma_{jl} = \int_0^1 F_{j-1}(s)F_{l-1}(s)ds$.*

This theorem indicates that the residual empirical process for an unstable AR(p) model with roots of one does not converge to a Brownian bridge as in the stable case. As a result, Lee and Wei recommended under such a situation, one should conduct a unit root test before using conventional methods such as the Kolmogorov-Smirnov test.

2.3. Model Selection

Consider the stochastic regression model (1) again. Wei (1987) studied the performance of the least squares estimates in prediction and formulated the so-called predictive principle for model selection. By analyzing the order of the cumulative predictive errors

$$C_n = \sum_{k=1}^{n}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_k - \boldsymbol{b}_{k-1}^{\mathrm{T}}\boldsymbol{x}_k)^2 = \sum_{k=1}^{n}(\hat{\epsilon_k} - \epsilon_k)^2,$$

where $\hat{\epsilon_k} = y_k - \hat{y}_k = y_k - \boldsymbol{b}_{k-1}^{\mathrm{T}}\boldsymbol{x}_k$ is the one-step prediction error, Wei observed that the term C_n plays a crucial role for order selection.

Theorem 2.7. *Consider the regression model* (1) *with $\{\epsilon_n\}$ satisfying assumption* (3). *Assume that*

$$\boldsymbol{x}_n^{\mathrm{T}}\Big(\sum_{k=1}^{n}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}\Big)^{-1}\boldsymbol{x}_n \to v \;\; a.s. \;\; as \; n \to \infty,$$

where v is a nonnegative random variable. Then

$$(1 - v)C_n + \sum_{k=1}^{n}[(\boldsymbol{b}_n - \boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{x}_k]^2 \sim nv\sigma^2 \;\; a.s.$$

on the set $\{1 > v > 0,\ c_n \to \infty\}$ *and*

$$C_n + \sum_{k=1}^{n} [(\boldsymbol{b}_n - \boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{x}_k]^2 \sim \sigma^2 \log\det\Big(\sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}}\Big)\ a.s.$$

on the set $\{v = 0,\ C_n \to \infty,\ \lambda_{\min}(n) \to \infty\}$, *where* $\lambda_{\min}(n)$ *denotes the minimum eigenvalue of the design matrix* $\sum_{k=1}^{n} \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}}$.

The proof of this result again relies on the Local Martingale Convergence Theorem and follows the argument of Lai and Wei (1982a). With this result, a new order selection criterion was proposed for a nonstationary AR(p) model. Specifically, Wei (1987) obtained the following result.

Theorem 2.8. *Assume that the autoregressive model* (9) *has roots equal to one or bigger than one in magnitude* (*i.e.,* $b = d_1 = \cdots = d_\ell = 0$) *and assume that* $\beta_p \neq 0$ *for an unknown* p, *but* $r \geq p$ *is given. Let* $\boldsymbol{y}_n = (y_1, \ldots, y_{n-r+1})^{\mathrm{T}}$. *Then*

$$\hat{a}_n = \Big[\log\det\Big(\sum_{k=p}^{n} \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}}\Big) / \log n - r\Big]^{1/2} \to a \ \textit{in probability.}$$

By means of the estimator $\hat{a}_n$, one can determine how many times to difference an integrated time series to achieve stationarity when the exact order p is unknown, but an upper bound r of the order is given. After differencing the integrated series $\hat{a}_n$ times, one can then apply traditional AIC or BIC for order selection. In other words, this theorem can be used to construct a two-step order selection procedure.

With this result, Wei went further on the notion of predictive least squares in model selection. In Wei (1992), he reconsidered (1) and examined the conventional model selection criterion

$$\log \hat{\sigma}_n^2 + \frac{c_n}{n}, \tag{15}$$

where n is the sample size and $\hat{\sigma}_n^2$ is the residual variance after fitting the model based on $\boldsymbol{x}$, and c_n is a nonnegative random variable that measures the complexity of the chosen model, which is proportional to the number of parameters. Common criteria such as the Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC) fall within this setting. Motivated by (15), Wei (1992) introduced the idea of the predictive least squares (PLS) criterion

$$\mathrm{PLS}(\boldsymbol{x}) = \sum_{i=m+1}^{n} (y_i - \boldsymbol{b}_{i-1}^{\mathrm{T}} \boldsymbol{x}_i)^2, \tag{16}$$

and considered four important issues. The first one deals with the interpretation of PLS in which (16) is decomposed into a sum of goodness of fit and a penalty that reflects the model complexity. The second one deals with the strong consistency of PLS for a general model including nonstationary time series. The third one is the asymptotic equivalence of PLS and BIC. Here, regression and time series depart from each other. Wei showed that for a polynomial regression model, BIC is not equivalent to PLS. But for a Gaussian time series model, Wei showed the asymptotic equivalence of PLS and BIC. The last issue deals with modifying the PLS to come up with an efficient criterion for model selection. Here he introduced the so-called Fisher's Information Criterion (FIC) and proved that the FIC can be expressed as

$$\mathrm{FIC}(M) = n\hat{\sigma}_n^2 + \tilde{\sigma}_n^2 \log\det\Big(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}\Big), \tag{17}$$

where M is the model with design vector $\boldsymbol{x}_i$, and $\hat{\sigma}_n^2$ and $\tilde{\sigma}_n^2$ are variance estimators based on the model M and the full model, respectively. For a linear regression model with Gaussian errors, the conditional Fisher's information matrix is simply $\sigma^{-2}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}$, which can be interpreted as the amount of information about the underlying unknown parameter. The FIC expression in (17) replaces the second quantity in the conventional criterion (15), which is proportional to the topological dimension of the selected model as reflected by c_n, by the second quantity of (17), which is proportional to the logarithm of the statistical information that is contained in M as reflected by the conditional Fisher's information matrix. This insight enables Wei to further link up PLS with FIC via

$$\mathrm{PLS} \sim n\hat{\sigma}_n^2 + \sigma^2 \log\det\Big(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}\Big).$$

Replacing σ^2 in the right-hand side by an estimator, PLS is simply FIC. In summary, Wei showed the following.

Theorem 2.9. *Assume either the stochastic regression model* (1) *holds with certain regularity conditions being satisfied, or that the nonstationary time series model* (9) *holds with characteristic polynomial satisfying* (10) *together with certain regularity conditions* (*see Theorem* 5.1.1 *of Wei* (1992)) *being satisfied. Then the FIC is strongly consistent.*

This theorem has many important consequences. It is related to the order selection criterion studied in Pötscher (1989) and the predictive minimum description length idea used in Rissanen (1986). A similar idea was proposed by Phillips (1995) in a Bayesian setting, where it is known as the posterior information criterion (PIC). A thorough discussion on this subject can be found in the 1995 themed issue of the *Journal of Econometrics* **69** entitled "Bayesian and Classical Econometric Modeling of Time Series." Further discussions about Bayesian unit root inference were given in Kadane, Chan and Wolfson (1996). Ing and Wei (2003, 2005) built on Wei's idea to study the so-called same realization prediction principle for order selection of autoregressive time series. They showed that AIC is asymptotically efficient for same realization prediction problems.

3. Miscellaneous Developments and Conclusion

Wei's contributions to time series are not restricted to the aforementioned domains. In a series of papers, Findley and Wei (1993, 2002) and Findley, Pötscher and Wei (2001, 2004) studied the convergence properties of sample moments for a general class of time series from which consistency of parameter estimates based on minimal multi-step ahead forecast errors were established. Wei has also made important contributions to non-linear modeling. The essence of the probabilistic developments was given in Guo and Wei (1993). In non-Gaussian processes, Wei supervised Karagrigoriou in a thesis that studied the optimality of order selection, see Karagrigoriou (1997).

In addition to time series, Wei has also produced a number of profound results in inference for stochastic processes in general. In Wei and Winnicki (1989, 1990), asymptotic results about critical branching processes were established. Although the form of these results are different from those arising in nonstationary time series, the techniques involved and the nature of the asymptotics are somewhat similar to the nonstationary case. Collectively, Wei considered both fields to be examples of what he called critical phenomena. In the area of stochastic modeling, Hu and Wei (1989) considered the irreversible adaptive allocation rule, and Basak, Hu and Wei (1997) studied the issue of weak convergence of recursions.

The field of time series and stochastic regression have been going through rigorous developments for the last two decades. Wei has provided an arsenal of fundamental contributions in these areas. Time after time, he either came up with the sharpest rate of convergence or he was among the first group of scholars who established the asymptotic validity of certain procedures. This is mainly due to his superb technical insight and his tenacity.

By reviewing his publications, one often comes up with more than the stated results. In fact, there are numerous areas of possible extensions. In the field of nonstationary time series, the recent popularity of long-memory models remains an open field, see Chan and Terrin (1995) and Buchmann and Chan (2005). Equally important is the area of empirical likelihood inference for time series. In addition to MLE, nonparametric procedures like empirical likelihood are gaining popularity and extension to this area is likely to be important, for related literature see Chan and Ling (2006). Another area of importance is inference for infinite variance models. Here, many of the LSE type results are no longer valid and an entirely new asymptotic theory needs to be established, see for example Chan, Peng and Qi (2005). Through asymptotic inference, Wei has greatly broadened the scope of time series and stochastic regression and pushed them to new frontiers with his keen new insight.

Acknowledgement

This research was supported in part by a HKSAR-RGC Grants CUHK400305 and CUHK4043/02P. Insightful comments from Dr. Sangyeol Lee, Dr. Sam Wong and an anonymous referee are gratefully acknowledged.

References

Anderson, T. W. and Taylor, J. (1979). Strong consistency of least squares estimators in dynamic models. *Ann. Statist.* **7**, 484-489.

Basak, G., Hu, I. and Wei, C. Z. (1997). Weak convergence of recursion. *Stochastic Process. Appl.* **68**, 65-82.

Buchmann, B. and Chan, N. H. (2005). Inference for nearly unstable processes under strong depedence. Technical Report, Department of Statistics, Chinese University of Hong Kong, Hong Kong.

Chan, N. H. (1988). The parameter inference for nearly nonstationary time series. *J. Amer. Statist. Assoc.* **83**, 857-862.

Chan, N. H. (1989). Asymptotic inference for unstable autoregressive time series with drifts. *J. Statist. Plann. Inference* **23**, 301-312.

Chan, N. H. and Ling, S. Q. (2006). Empirical likelihood for GARCH models. *Econom. Theory* **22**, 403-428.

Chan, N. H., Peng, L. and Qi, Y. (2005). Quantile inference for nearly nonstationary autoregressive time series with infinite variance. *Statist. Sinica* **15**, 15-28.

Chan, N. H. and Terrin, N. C. (1995). Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Ann. Statist.* **23**, 1662-1683.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15**, 1050-1063.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367-401.

Chen, G. J., Lai, T. L. and Wei, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* **11**, 319-333.

Dahlhaus, R. (1985). *Data Tapers in Time Series Analysis.* Habilitation thesis, Universität-GHS, Essen. Essen, Germany.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427-431.

Findley, D. F. and Wei, C. Z. (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statist. Sinica* **3**, 453-470.

Findley, D. F. and Wei, C. Z. (2002). AIC, overfitting principles and the boundedness of moments of inverse matrices for vector autoregressions and related models. *J. Multivariate Anal.* **83**, 415-450.

Findley, D. F., Pötscher, B. and Wei, C. Z. (2001). Uniform convergence of sample second moments of families of time series arrays. *Ann. Statist.* **29**, 815-838.

Findley, D. F., Pötscher, B. and Wei, C. Z. (2004). Modeling of time series arrays by multistep prediction or likelihood methods. *J. Econometrics* **118**, 151-187.

Guo, L. (1990). *Time-Varying Stochastic Systems: Stability, Estimation and Control.* Jilin Science and Technology Press, Jilin, China.

Guo, M. H. and Wei, C. Z. (1993). A lower bound for expectation of a convex functional. *Statist. Probab. Lett.* **18**, 191-194.

Hu, I. and Wei, C. Z. (1989). Irreversible adaptive allocation rules. *Ann. Statist.* **17**, 801-823.

Ing, C. K. and Wei, C. Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.* **85**, 130-155.

Ing, C. K. and Wei, C. Z. (2005). Order selection for the same-realization prediction in autoregressive processes. *Ann. Statist.* **33**, in press.

Jeganathan, P. (1991). On the asymptotic behavior of least-squares estimators in AR time series with roots neat the unit circle. *Econom. Theory* **7**, 269-306.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econom. Theory* **11**, 818-887.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

Kadane, J. B., Chan, N. H. and Wolfson, L. (1996). Priors for unit root models. *J. Econometrics* **75**, 99-111.

Karagrigoriou, A. (1997). Asymptotic efficiency of the order selection of a non-Gaussian AR process. *Statist. Sinica* **7**, 407-423.

Kurtz, T. G. and Protter, P. (1991). Weak limit thoerems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19**, 1035-1070.

Lai, T. L. (2003). Stochastic approximation. *Ann. Statist.* **31**, 391-406.

Lai, T. L. and Wei, C. Z. (1982a). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10**, 154-166.

Lai, T. L. and Wei, C. Z. (1982b). Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.* **12**, 346-370.

Lai, T. L. and Wei, C. Z. (1982c). Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters. *J. Multivariate Anal.* **13**, 1-23.

Lai, T. L. and Wei, C. Z. (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In *Multivariate Analysis VI*, (Edited by Krishnaiah, P. R.), 375-393. North-Holland, Amsterdam.

Lai, T. L. and Wei, C. Z. (1986). Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Trans. Automat. Control* AC-**31**, 898-906.

Lai, T. L. and Wei, C. Z. (1987). Asymptotically efficient self-tuning regulators. *SIAM J. Control Optim.* **25**, 466-481.

Lai, T. L., Robbins, H. and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. U.S.A.* **75**, 3034-3036.

Lee, S. Y. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.* **27**, 237-261.

Ling, S. and Li, W. K. (1998) Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroskedastic errors. *Ann. Statist.* **26**, 84-125.

Nielsen, B. (2005). Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econom. Theory* **21**, 534-561.

Phillips, P. C. B. (1987). Toward a unified asymptotic theory of autoregressions. *Biometrika* **74**, 535-547.

Phillips, P. C. B. (1995). Bayesian prediction: a response. *J. Econometrics* **69**, 351-365.

Pötscher, B. M. (1989). Model selection under nonstationary autoregressive models and stochastic linear regression models. *Ann. Statist.* **17**, 1257-1274.

Rissanen, J. (1986). Stochastic complexity and modeling. *Ann. Statist.* **14**, 1080-1100.

Stock, J. H. and Watson, M. W. (1988). Variable trends in economic time series. *J. Economic Perspective* **2**, 147-174.

Tiao, G. C. and Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11**, 856-871.

Tsay, R. S. and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Ann. Statist.* **18**, 220-250.

Wei, C. Z. (1985). Asymptotic properties of least squares estimates in stochastic regression models. *Ann. Statist.* **13**, 1498-1508.

Wei, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15**, 1667-1682.

Wei, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20**, 1-42.

Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stochastic Process. Appl.* **31**, 261-282.

Wei, C. Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.* **18**, 1757-1778.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188-1197.

Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong.
E-mail: nhchan@sta.cuhk.edu.hk

(Received April 2005; accepted October 2005)