The Annals of Statistics
1990, Vol 18, No 4, 1729–1756

ESTIMATION OF NONSTATIONARY ARMAX MODELS BASED ON THE HANNAN–RISSANEN METHOD

BY DAWEI HUANG AND LEI GUO

Peking University and Australian National University

We consider in this paper the estimation problems for both orders and coefficients of linear feedback control systems, described by ARMAX models. The estimation algorithms are inspired by the Hannan–Rissanen method used for the estimation of stationary ARMA models, while the convergence analyses are based on limit theorems for both double array martingales and nonnegative supermartingales, and on techniques of stochastic Lyapunov functions. Traditionally used assumptions, such as the *strictly positive real condition* and the *requirement of known upper bounds for true orders*, are not imposed here

1. Introduction. One of the basic issues in statistical sciences is how to choose models to fit observations. The observations are objective, yet the models are generally ideal. Thus, the models considered have been more realistic and hence usually more complicated.

In this paper, we consider the linear feedback control systems described by the ARMAX model

$$A(z)y_t = B(z)u_t + C(z)w_t, \qquad t \geq 0, \tag{1.1}$$

where y_t, u_t and w_t are, respectively, the m-, l-, and m-dimensional system output, input and noise sequences, with initial values $\{y_i, u_j, w_k, \ -p_0 \leq i \leq -1, \ -q_0 \leq j \leq -1, \ -r_0 \leq k \leq 1\}$ $A(z)$, $B(z)$ and $C(z)$ are *unknown* matrix polynomials in backwards-shift operator z

$$A(z) = I + A_1 z + \cdots + A_{p_0} z^{p_0}, \qquad p_0 \geq 0, \tag{1.2}$$

$$B(z) = B_1 z + B_2 z^2 + \cdots + B_{q_0} z^{q_0}, \qquad q_0 \geq 0, \tag{1.3}$$

$$C(z) = I + C_1 z + \cdots + C_{r_0} z^{r_0}, \qquad r_0 \geq 0, \tag{1.4}$$

where p_0, q_0 and r_0 are the *unknown* true orders ($A_{p_0} \neq 0, B_{q_0} \neq 0, C_{r_0} \neq 0$).

Such an ARMAX model though may not be uniquely defined in the multivariable case [see, e.g., Hannan and Deistler (1988), Section 2.7], it is not critical for a portion of results we shall investigate. However, for estimating the orders (p_0, q_0, r_0) and the parameters $\{A_i, B_j, C_k, \ i = 1, \dots, p_0, \ j = 1, \dots, q_0, \ k = 1, \dots, r_0\}$, we will need the following identifiability condition:

(1.5) (z), $B(z)$ and $C(z)$ have no common left factor and A_{p_0}, B_{q_0} and C_{r_0} are all of row full rank.

Received July 1988; revised March 1989.
AMS 1980 *subject classifications.* Primary 62M10, 93E12; secondary 60F15.
Key words and phrases. ARMAX models, nonstationary, least squares, order estimation, martingales, feedback, positive real.

It should be noted that in the scalar variable case the rank condition is automatically satisfied.

In this paper, we shall also need the usual minimum phase condition on the noise model, i.e.,

$$\det C(z) \neq 0, \qquad |z| \leq 1. \tag{1.6}$$

The above ARMAX models have been studied in at least two different areas: time series analysis and adaptive estimation and control. The special case of $q_0 = 0$ corresponds to the standard ARMA model in the time series analysis. For stationary cases ($\det A(z) \neq 0, |z| \leq 1$), Hannan and Rissanen (1982) have proposed a three step procedure to estimate both the orders and coefficients of the ARMA model. The first step was to estimate the innovation process $\{w_t\}$ by increasing lag autoregressions. The second step was to estimate the coefficients and orders by observations $\{y_t\}$ and the innovation estimates obtained from the first step. The third step was to obtain efficient estimates for the coefficients by use of the estimates of the second step. The first and second steps originated from Durbin (1961) for the case of known orders p_0 and r_0. However, rigorous theoretical analysis has been carried out only since the work of Hannan and Rissanen (1982). In the analysis of increasing lag autoregressions (the first step), some kind of uniform convergence rate for autocovariances [An, Chen and Hannan (1982)] or autocorrelations [Hannan and Kavalieris (1983)] is needed. This leads to the consideration of asymptotic behaviors of sequences of the form

$$\max_{1 \leq k \leq h_n} \max_{1 \leq i \leq n} \left| \sum_{j=0}^{i} w_j^\tau y_{j-k} \right|, \tag{1.7}$$

where $\{h_n\}$ is a nondecreasing sequence of integers [e.g., Hannan and Kavalieris (1984); Huang (1987)]. It is usually the case that for any fixed k, $\{w_j^\tau y_{j-k}, j \geq 0\}$ constitutes a martingale difference sequence and so the summation part in (1.7) is a special form of double array martingales. It is worth noting that the study of (1.7) is also a crucial step for the order estimation problems when the upper bounds for the true orders are not available. By using estimations for sequences of the form (1.7), Hannan, et al. [e.g., Hannan and Kavalieris (1984); Hannan and Deistler (1988)] extend their results for stationary ARMA models to stationary ARMAX ones. In these works, their main interests are in open-loop identifications, since they require that the system (1.1) is stable in structure or open-loop stable (i.e., $\det A(z) \neq 0, |z| \leq 1$) and that the input sequence $\{u_t\}$ and the noise sequence $\{w_t\}$ are either stationary correlated or independent. This later assumption may exclude the application of their results to general feedback control systems, because any real feedback controller depends essentially on the system output and hence the driven noise and is generally nonstationary.

For this reason, the above mentioned stationary and independency assumptions on $\{u_t\}$ are usually not imposed in the area of adaptive estimation and control. The simple case of $r_0 = 0$ was first studied [e.g., Astrom (1968); Ljung

(1976); Moore (1978)] where an identifiability condition on the observations, i.e., the "Persistence of Excitation" (PE) condition was required. This condition was later relaxed by Chen (1982) and Lai and Wei (1982a). Particularly, Lai and Wei (1982a) obtained the weakest possible convergence conditions for least squares (LS) estimates in some sense. Unfortunately, straightforward extensions of these results from the special case of $r_0 = 0$ to the general $r_0 \geq 0$ cases are hardly possible without further assumptions on the noise model besides (1.6). Indeed, in the $r_0 \geq 0$ cases, most of the existing adaptive estimation and control algorithms need some kind of strictly positive real (SPR) conditions on the noise model in the convergence analysis [e.g., Ljung and Soderstrom (1983); Goodwin and Sin (1984)]. In particular, for the convergence of the standard extended least squares (ELS) algorithm, it is required that [e.g., Ljung (1977); Solo (1979), Chen (1982); Lai and Wei (1986); Chen and Guo (1986)]

$$(1.8) \qquad C^{-1}(e^{i\lambda}) + C^{-\tau}(e^{-i\lambda}) - I > 0, \qquad \forall \lambda \in [0, 2\pi]$$

Qualitatively, this condition means that the system noise $\{C(z)w_t\}$ is not too *colored*. Indeed, it can be shown that (1.8) implies the minimum phase condition (1.6) and $\|[C_1, \dots, C_{r_0}]\| < 1$ (see the Appendix), where the norm for a matrix X is defined as $\{\lambda_{\max}(XX^*)\}^{1/2}$ and $\lambda_{\max}(\cdot)(\lambda_{\min}(\cdot))$ denotes the maximum (minimum) eigenvalue of the corresponding matrix. Thus, in the scalar case, (1.8) implies $\sum_{i=1}^{r_0}[c_i]^2 < 1$. And so, for example, it is immediately seen that the minimum phase polynomial $C(z) = (1 - 0.3z)(1 - 0.4z)(1 - 0.5z)$ does not satisfy (1.8). It is also known that if the SPR condition (1.8) fails, the ELS algorithm generally does not converge [e.g., Ljung and Soderstrom (1983)]. Many efforts have recently been devoted to relax the SPR condition in adaptive estimation and control. However, all of these contributions either need some extra a priori information on $C(z)$ besides (1.6), or cannot be applied to feedback control systems [see Guo and Huang (1988) and the references therein].

Perhaps, the main reason why the standard ELS algorithm requires a SPR condition for its convergence is that the *innovation estimate is generated by itself*! By using the results in Guo, Huang and Hannan (1990), we have recently proposed a two-step method to obtain strongly consistent parameter estimates for ARMAX models without the SPR condition [Guo and Huang (1989)]. This method is similar to the Hannan–Rissanen method, because it also estimates the innovations at the first step and estimates the coefficients at the second step. However, as is seen from Guo and Huang (1989), the convergence analysis for feedback control cases appears to be completely different from those for the traditional stationary case. One of the reasons is that the standard notions of autocovariances and autocorrelations, which are so important in the stationary case, become useless in the general feedback control cases. Nevertheless, some precise convergence results can still be established when some knowledge about the true order (p_0, q_0, r_0) is available [Guo and Huang (1989)]. The key idea used in Guo and Huang (1989) is that: Although in the general feedback control cases the estimates for the autoregressive parameters in the first step may not converge, the innovation estimates have some very desirable properties according to the theory in Guo, Huang and Hannan (1990). Thus in the second step, the innovation estimates of the first step can still be successfully used in getting consistent parameter estimates.

In this paper, we will continue the above work. As a crucial step, two kinds of innovation estimates are proposed and analyzed in the first place (Theorems 2.1 and 2.2). They are essential improvements and extensions over those in Guo, Huang and Hannan (1990) and Guo and Huang (1989). Then we consider estimation problems for both the unknown orders and coefficients of system (1.1) (Theorem 2.3). In the present results, the standard assumptions such as SPR conditions and a priori known upper bounds for the true orders [e.g., Guo, Chen and Zhang (1989)] are removed.

2. The main results. Since the innovation estimate plays a crucial role in ARMAX model identification, we shall consider it first.

2.1. *Estimation of the innovation process.* Let $\{h_n\}$ be a sequence of nondecreasing positive integers and introduce the following regression vectors for any $n \geq 1$:

$$(2.1)\quad \psi_t(h_n) = \left[y_t^\tau, y_{t-1}^\tau, \dots, y_{t-h_n+1}^\tau, u_t^\tau, u_{t-1}^\tau, \dots, u_{t-h_n+1}^\tau\right]^\tau, \quad 0 \leq t \leq n.$$

The innovation process $\{w_t\}$ can be estimated by either of the following estimates.

1. The "honest estimate" $\{\hat{w}_t(n), 1 \leq t \leq n\}$:

$$(2.2)\quad \hat{w}_t(n) = y_t - \hat{\alpha}_t^\tau(n)\psi_{t-1}(h_n), \qquad 1 \leq t \leq n,$$

$$(2.3)\quad \hat{\alpha}_{t+1}(n) = \hat{\alpha}_t(n) + b_t(n)P_t(n)\psi_t(h_n)\left[y_{t+1} - \psi_t^\tau(h_n)\hat{\alpha}_t(n)\right],$$

$$(2.4)\quad P_{t+1}(n) = P_t(n) - b_t(n)P_t(n)\psi_t(h_n)\psi_t^\tau(h_n)P_t(n),$$

$$b_t(n) = \{1 + \psi_t^\tau(h_n)P_t(n)\psi_t(h_n)\}^{-1},$$

where the initial values $\hat{\alpha}_0(n) = 0$ and $P_0(n) = \beta I$, $\beta > 0$.

2. The "final estimate" $\{\hat{\varepsilon}_t(n), 1 \leq t \leq n\}$:

$$(2.5)\quad [\hat{\varepsilon}_1(n), \hat{\varepsilon}_2(n), \dots, \hat{\varepsilon}_n(n)]^\tau \triangleq Y_n - \Phi_n[\Phi_n^\tau\Phi_n]^{-1}\Phi_n^\tau Y_n,$$

$$(2.6)\quad Y_n = [y_1, y_2, \dots, y_n]^\tau,$$

$$(2.7)\quad \Phi_n = [\psi_0(h_n), \psi_1(h_n), \dots, \psi_{n-1}(h_n)]^\tau.$$

The honest estimate and the final estimate are so named because $\hat{w}_t(n)$ is $\sigma\{y_i, u_i, i \leq t\}$-measurable and $\hat{\varepsilon}_t(n)$ is $\sigma\{y_i, u_i, i \leq n\}$-measurable for any $t \in [1, n]$. The following theorem establishes the asymptotic properties of these two kinds of innovation estimates.

THEOREM 2.1. *For the* ARMAX *model* (1.1), *assume that* $\{w_t, F_t\}$ *is a martingale difference sequence satisfying*

$$\sup_t E\left[\|w_{t+1}\|^2 \middle| F_t\right] < \infty, \qquad \|w_t\| = O(\varphi(t)) \quad a.s., \tag{2.8}$$

where the function $\varphi(\cdot)$ *is positive, deterministic, nondecreasing and satisfies*

$$\sup_k \varphi(e^{k+1})/\varphi(e^k) < \infty. \tag{2.9}$$

Assume further that for some constant b,

$$\sum_{i=0}^{n-1} \left(\|y_i\|^2 + \|u_i\|^2\right) = O(n^b) \quad a.s. \tag{2.10}$$

and u_t *is* F_t*-measurable. If the regression lag* h_n *in* (2.1) *is chosen as* $h_n = O(\{\log n\}^\alpha)$, $(\alpha > 1)$ *and* $\log n = o(h_n)$, *then as* $n \to \infty$,

$$\sum_{t=1}^{n} \|\hat{w}_t(n) - w_t\|^2 = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right) \quad a.s. \tag{2.11}$$

and

$$\sum_{t=1}^{n} \|\hat{\varepsilon}_t(n) - w_t\|^2 = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right) \quad a.s., \tag{2.12}$$

where $\{\hat{w}_t(n)\}$ *and* $\{\hat{\varepsilon}_t(n)\}$ *are defined by* (2.2) *and* (2.5), *respectively.*

The proof of this theorem is given in Section 3. We remark that u_t is F_t-measurable implies that the input u_t is a feedback signal (i.e., u_t is a function of the observations $\{y_i, u_{i-1}, i \le t\}$), while (2.10) means that under this feedback controller, the closed-loop system is not explosive. We also mention that in the noncontrolled case ($u_t = 0$), this condition may be applicable to ARMA models with unstable zeros of $\det A(z)$ lying on the unit circle, but it fails for explosive models.

From (2.11)–(2.12), we see that if the sample path behavior of the noise process $\{w_t\}$ is not "too bad" (e.g., $\{w_t\}$ is bounded a.s.), or is Gaussian and white ($\|w_t\| = O\{\log t\}^{1/2}$) or has a growth rate of $O(\{\log t\}^{1-\varepsilon})$, $(\varepsilon > 0)$, then the second term $o(\{\varphi(n)\log\log n\}^2)$ may be negligible. We now give an example to show that in such cases the results of Theorem 2.1 are the best possible.

EXAMPLE 2.1. For an ARMAX model (1.1) with $\det A(z) \neq 0$, $|z| \le 1$, assume that $\{w_t\}$ is a zero mean Gaussian white noise (i.i.d.) sequence independent of $\{u_t\}$ and that $\{u_t\}$ is a Gaussian stationary ARMA process whose spectral density matrix is uniformly positive definite on $[-\pi, \pi]$. If h_n is chosen as in Theorem 2.1, then the honest estimate $\{\hat{w}_t(n)\}$ satisfies

$$(2.13)\quad \lim_{n\to\infty} \frac{\sum_{t=1}^{n}[\hat{w}_t(n) - w_t][\hat{w}_t(n) - w_t]^{\tau}}{h_n \log n} = \sum \ (\triangleq Ew_t w_t^{\tau}) \quad \text{a.s.}$$

The proof is also given in Section 3. This example not only shows that the result (2.11) in Theorem 2.1 is sharp in some sense, but also has its own significance in choosing the lag $\{h_n\}$ for stationary ARMAX model identification.

In some cases, however, it is possible that the sample path growth rate of the noise process is rather fast, as when the noise sequence has only finite second moment. In such cases, the results of Theorem 2.1 may be rough. To deal with this situation, we present the Theorem 2.2.

THEOREM 2.2. *For* ARMAX *model* (1.1), *assume that* $\{w_t, F_t\}$ *is a martingale difference sequence satisfying*

$$(2.14)\qquad E\Big\{\sup_t E\Big[\|w_{t+1}\|^2 | F_t\Big]\Big\} < \infty$$

and that (2.10) *holds. Let the regression* lag $h_n = [c(\log n)^{\alpha}]$ *be the integer part of* $c(\log n)^{\alpha}$ *for some* $\alpha > 1$ *and* $c > 0$. *Then as* $n \to \infty$,

$$(2.15)\qquad \sum_{t=1}^{n} \|\hat{w}_t(n) - w_t\|^2 = O\big([h_n]^2 \log n \{\log\log n\}^{2+\delta}\big) \quad a.s.$$

and

$$(2.16)\qquad \sum_{t=1}^{n} \|\hat{\varepsilon}_t(n) - w_t\|^2 = O\big([h_n]^2 \log n \{\log\log n\}^{2+\delta}\big) \quad a.s.,$$

for any $\delta > 0$, *where* $\{\hat{w}_t(n)\}$ *and* $\{\hat{\varepsilon}_t(n)\}$ *are defined by* (2.2) *and* (2.5), *respectively.*

The proof of this theorem is given in Section 4. We point out that the proofs of both Theorems 2.1 and 2.2 depend essentially on two key techniques: (i) A standard recursion for stochastic Lyapunov functions which has been previously used in, e.g., Moore (1978), Solo (1979), Chen (1982), Lai and Wei (1982a) and Chen and Guo (1986) for the usual fixed lag case, and in Guo, Huang and Hannan (1990) for increasing lag cases; (ii) The martingale limit theory. The proof of Theorem 2.1 hinges on supermartingale exponential inequality, and some techniques of truncations and subsequences; while the proof of Theorem 2.2 relies on the nonnegative supermartingale convergence theory which has been previously used in engineering literature [see, e.g., Ljung (1976); Moore (1978); Solo (1979); Goodwin and Sin (1984)].

We are now in a position to consider the estimation problems of both orders and coefficients.

2.2. *Estimation of the orders and parameters.* Let us introduce the following notations:

$$\phi_t(p,q,r) = \left[y_t^\tau, y_{t-1}^\tau, \dots, y_{t-p+1}^\tau, u_t^\tau, u_{t-1}^\tau, \dots, u_{t-q+1}^\tau, \hat{w}_t^\tau(n), \hat{w}_{t-1}^\tau(n), \dots, \hat{w}_{t-r+1}^\tau(n)\right]^\tau, \tag{2.17}$$

$$X_n(p,q,r) = \left[\phi_0(p,q,r), \phi_1(p,q,r), \dots, \phi_{n-1}(p,q,r)\right]^\tau, \tag{2.18}$$

$$Z_n(p,q,r) = Y_n - Z_n(p,q,r)\left[X_n^\tau(p,q,r)X_n(p,q,r)\right]^{-1}X_n^\tau(p,q,r)Y_n, \tag{2.19}$$

$$\hat{\sigma}_n(p,q,r) = \operatorname{tr}\{Z_n^\tau(p,q,r)Z_n(p,q,r)\}, \tag{2.20}$$

where $\{\hat{w}_t(n)\}$ and Y_n are defined by (2.2) and (2.6), respectively.

We now consider the following information criterion (CIC):

$$\mathrm{CIC}(p,q,r)_n = \hat{\sigma}_n(p,q,r) + (p+q+r)a_n \tag{2.21}$$

[see, Guo, Chen and Zhang (1987)], where the first C stresses that the criterion is designed for *control* systems and where $\{a_n\}$ is a nondecreasing of positive numbers specified later on, in (2.27)–(2.28).

1. Order estimation procedure. For any $n \ge 1$, this procedure consists of two steps:

   Step 1.

$$\text{Take } \hat{m}(n) \text{ to minimize } \mathrm{CIC}(k,k,k)_n,\ 0 \le k \le [\log n]. \tag{2.22}$$

   Step 2.

$$\text{Take } \hat{p}(n) \text{ to minimize } \mathrm{CIC}(p,\hat{m}(n),\hat{m}(n))_n,\ 0 \le p \le \hat{m}(n). \tag{2.23}$$

$$\text{Take } \hat{q}(n) \text{ to minimize } \mathrm{CIC}(\hat{p}(n),q,\hat{m}(n))_n,\ 0 \le q \le \hat{m}(n). \tag{2.24}$$

$$\text{Take } \hat{r}(n) \text{ to minimize } \mathrm{CIC}(\hat{p}(n),\hat{q}(n),r)_n,\ 0 \le r \le \hat{m}(n). \tag{2.25}$$

2. Parameter estimates. For any $n \ge 1$, the estimate $\hat{\theta}(n)$ for the unknown parameter $\theta^* \triangleq [-A_1, \dots, -A_{p_0}, B_1, \dots, B_{q_0}, C_1, \dots, C_{r_0}]^\tau$ is defined by

$$\hat{\theta}(n) = \hat{\theta}_n(\hat{p}(n), \hat{q}(n), \hat{r}(n)),$$

   where $(\hat{p}(n), \hat{q}(n), \hat{r}(n))$ is defined by (2.23)–(2.25) and

$$\hat{\theta}_n(p,q,r) = \left[X_n^\tau(p,q,r)X_n(p,q,r)\right]^{-1}X_n^\tau(p,q,r)Y_n, \tag{2.26}$$

   with $X_n(p,q,r)$ and Y_n the same as those in (2.19).

It is worth noting that in the above order estimation procedure, the first step (2.22) corresponds to estimating the value of $m_0 \triangleq \max\{p_0, q_0, r_0\}$. In the second step, the true orders p_0, q_0 and r_0 are searched between at most $3\hat{m}(n)$ points at each time instant n, rather than $[\hat{m}(n)]^3$ points as in [e.g., Guo, Chen, and Zhang (1989)]. These ideas were previously used in Huang (1989) for estimation of ARMA orders.

In Section 5 we will prove the following theorem:

THEOREM 2.3. (i) *Under the conditions of Theorem 2.1, if* $\lambda_{\min}^0(n)$ *satisfies*

$$h_n \log n + [\varphi(n)\log\log n]^2 = o(\lambda_{\min}^0(n)) \quad a.s.$$

and if in the criterion (2.21), *the sequence* $\{a_n\}$ *is chosen to satisfy*

$$\{h_n \log n + [\varphi(n)\log\log n]^2\}/a_n \to 0 \quad a.s. \text{ as } n \to \infty, \tag{2.27}$$

and

$$a_n/\lambda_{\min}^s 0(n) \to 0 \quad a.s. \text{ as } n \to \infty, \tag{2.28}$$

where $\lambda_{\min}^0(n)$ *is defined as*

$$\lambda_{\min}^0(n) = \lambda_{min}\left\{\sum_{t=0}^{n-1} \phi_t^0 \phi_t^{0\tau}\right\}, \tag{2.29}$$

with $\phi_t^0 = [y_t^\tau, \ldots, y_{t-m_0+1}^\tau, u_t^\tau, \ldots, u_{t-m_0+1}^\tau, w_t^\tau, \ldots, w_{t-m_0+1}^\tau]^\tau$, $m_0 \triangleq \max\{p_0, q_0, r_0\}$. *Then for the estimation algorithm defined by* (2.17)–(2.26), *as* $n \to \infty$,

$$\hat{m}(n) \to m_0 \quad a.s., \tag{2.30}$$

$$(\hat{p}(n), \hat{q}(n), \hat{r}(n)) \to (p_0, q_0, r_0) \quad a.s. \tag{2.31}$$

and

$$\|\hat{\theta}(n) - \theta^*\|^2 = O\left(\frac{h_n \log n}{\lambda_{\min}^0(n)}\right) + o\left(\frac{\{\varphi(n)\log\log n\}^2}{\lambda_{\min}^0(n)}\right) \quad a.s. \tag{2.32}$$

(ii) *Under the conditions of Theorem 2.2, if* $\lambda_{\min}^0(n)$ *satisfies*

$$[h_n]^2 \log n (\log\log n)^{2+\delta} = o(\lambda_{\min}^0(n)) \quad a.s.,$$

and if $\{a_n\}$ *is chosen to satisfy* (2.28) *and*

$$\{[h_n]^2 \log n(\log\log n)^{2+\delta}\}/a_n \to 0 \quad a.s. \text{ for some } \delta > 0, \tag{2.33}$$

then (2.30) *and* (2.31) *also hold and*

$$\|\hat{\theta}(n) - \theta^*\|^2 = O\left(\frac{[h_n]^2 \log n\{\log\log n\}^{2+\delta}}{\lambda_{\min}^0(n)}\right) \quad a.s. \tag{2.34}$$

As one would have probably noted, a major feature of the above theorem is that the criterion CIC (2.21) depends on $\{a_n\}$, which in turn depends on the growth rate of $\lambda_{\min}^0(n)$ as exposed in (2.28). This is naturally expected because there are no specific constraints on the input sequence $\{u_t\}$ except those in (2.10) and the $\{u_t\}$ determines completely the excitation extent—the growth rate of $\lambda_{\min}^0(n)$.

Let us now consider the continuously excited controller used in the area of adaptive control. To be specific, let $\{v_t\}$ be a sequence of l-dimensional i.i.d.

random vectors independent of $\{w_t\}$ with properties:

$$Ev_n = 0, \qquad Ev_n v_n^\tau = \mu I, \qquad E\|v_n\|^3 < \infty, \qquad \mu > 0. \tag{2.35}$$

Assume that u_t^0 is any l-dimensional and $\sigma\{w_i, v_{i-1}, i \le t\}$-measurable random vector (any feedback controller is of this kind). The continuously excited controller u_n is defined as [see, e.g., Caines and Lafortune (1984); Chen and Guo (1986)]

$$u_n = u_n^0 + v_n. \tag{2.36}$$

Corollary 2.1. *Assume that for system* (1.1), *the identifiability condition* (1.5) *holds and that*

$$\liminf_{n\to\infty} \lambda_{\min}\left\{\frac{1}{n}\sum_{i=0}^{n-1} w_i w_i^\tau\right\} > 0 \quad a.s. \tag{2.37}$$

If the control law (2.36) *is applied to the system* (1.1) *and*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}\left(\|y_i\|^2 + \|u_i\|^2\right) < \infty \quad a.s., \tag{2.38}$$

then Theorem 2.3 *still holds with* $\lambda_{\min}^0(n)$ *replaced by* n *in* (2.28), (2.32) *and* (2.34).

Proof. We need only to note that in this case, the system (1.1) is persistently excited, i.e., $\liminf n \to \infty\ \lambda_{\min}^0(n)/n > 0$, a.s. This fact is a specialization of those proved in Guo, Chen and Zhang (1989), [see also Chen and Guo (1986) for related results]. □

We remark that the property (2.38) is usually regarded as a closed-loop stability criterion in the area of stochastic adaptive control. Apparently, a system of the form (1.1) satisfying (2.38) is not necessarily open-loop stable.

To conclude this section, we mention that the efficiency of the estimates given above is still a concern for the present feedback control cases. Of course, with some further restrictions on the input sequence $\{u_t\}$, it is possible to use the similar ideas as those in the third step of the Hannan–Rissanen method to investigate this problem. Such a discussion will be presented elsewhere.

3. Double array martingale limit theory and the proof of Theorem 2.1. Although various martingale limit theorems have been studied extensively in the literature, there are only few results on limit behaviors of double array martingales [e.g., Stout (1974); Lai and Wei, (1982b)]. These results, due to various restrictions, can hardly be applied to the present situation. For our later use, we now present the following results on double array martingales, which are improvements over those studied in Guo, Huang and Hannan (1990).

LEMMA 3.1. *Let $\{w_t, F_t\}$ be an m-dimensional martingale difference sequence satisfying*

$$\|w_t\| = o(\varphi(t)) \quad a.s., \tag{3.1}$$

where $\varphi(x)$ is described as in Theorem 2.1. Assume that $f_t(k)$, $t, k = 1, 2, \dots,$ is an F_t-measurable, $p \times m$-dimensional random matrix satisfying

$$\|f_t(k)\| \le A < \infty \quad a.s. \text{ for all } t, k \text{ and some deterministic constant } A. \tag{3.2}$$

Then for $h_n = O([\log n]^\alpha)$, the following properties hold as $n \to \infty$,

$$\text{(i)} \max_{1\le k\le h_n} \max_{1\le i\le n} \left\| \sum_{j=1}^{i} f(k) w_{j+1} \right\| = O\left(\max_{1\le k\le h_n} \sum_{j=1}^{n} \|f_j(k)\|^2 \right) + o(\varphi(n)\log\log n) \quad a.s. \tag{3.3}$$

provided that

$$\sup_j E\left(\|w_{j+1}\|^2 | F_j\right) < \infty \quad a.s. \tag{3.4}$$

$$\text{(ii)} \max_{1\le k\le h_n} \max_{1\le i\le n} \left\| \sum_{j=1}^{i} f_j(k) w_{j+1} \right\| = O\left(\max_{1\le k\le h_n} \sum_{j=1}^{n} \|f_j(k)\| \right) + o(\varphi(n)\log\log n) \quad a.s. \tag{3.5}$$

provided that

$$\sup_j E\left(\|w_{j+1}\| \, | F_j\right) < \infty \quad a.s. \tag{3.6}$$

PROOF. (i) We need only to consider the case of scalar variables and $A = 1$. For any $\varepsilon > 0$, let us set

$$\tilde{w}_j = w_j I\{|w_j| \le \varepsilon\varphi(j)\}, \qquad \tilde{\tilde{w}}_j = \tilde{w}_j - E(\tilde{w}_j | F_{j-1}).$$

Then

$$\left| \sum_{j=1}^{i} f_j(k) w_{j+1} \right| \le \left| \sum_{j=1}^{i} w_{j+1} I\{|w_{j+1}| > \varepsilon\varphi(j+1)\} f_j(k) \right| + \left| \sum_{j=1}^{i} E(\tilde{w}_{j+1} | F_j) f_j(k) \right| + \left| \sum_{j=1}^{i} f_j(k) \tilde{\tilde{w}}_{j+1} \right|. \tag{3.7}$$

We have from (3.1) and (3.2) that

$$\begin{aligned} &\max_{1\le k\le h_n} \max_{1\le i\le n} \sum_{j=1}^{i} f_j(k) w_{j+1} I\{|w_{j+1} > \varepsilon\varphi(j+1)\} \\ &\le \sum_{j=1}^{n} |w_{j+1}| I\{|w_{j+1}| > \varepsilon\varphi(j+1)\} \\ &\le o(\varphi(n+1)) \sum_{j=1}^{\infty} I\{|w_{j+1}| > \varepsilon\varphi(j+1)\} = o(\varphi(n+1)) \quad \text{a.s.} \end{aligned} \tag{3.8}$$

Also, letting $a_n = \max_{1 \le k \le h_n} \{\sum_{j=1}^n |f_j(k)|^2\}^{1/2}$, we have under (3.4),

$$\max_{1 \le k \le h_n} \max_{1 \le i \le n} \left| \sum_{j=1}^{i} E(\tilde{w}_{j+1}|F_j) f_j(k) \right|$$

$$\le \left\{ \sum_{j=1}^{n} \left[E\left(w_{j+1} I\left(|w_{j+1}| \le \varepsilon\varphi(j+1)|F_j\right)\right)\right]^2 \right\}^{1/2} a_n$$

$$(3.9) \qquad = \left\{ \sum_{j=1}^{n} \left[E\left(w_{j+1} I\left(|w_{j+1}| > \varepsilon\varphi(j+1)|F_j\right)\right)\right]^2 \right\}^{1/2} a_n$$

$$\le \left\{ \sup_j E\left(|w_{j+1}|^2|F_j\right) \sum_{j=1}^{n} P\left(|w_{j+1}| > \varepsilon\varphi(j+1)|F_j\right) \right\}^{1/2} a_n$$

$$= O(a_n) \quad \text{a.s.},$$

where the last relation is deduced by using (3.1) and the conditional Borel–Cantelli lemma [see, e.g., Stout (1974), page 55].

Then, to prove (3.3) we need only to consider the last term on the R.H.S. of (3.7). Set

$$S_i(k) = \sum_{j=1}^{i} \overline{w}_{j+1} f_j(k), \qquad 1 \le i \le n,$$

$$S_0(k) = 0, \qquad d(x) = 2\varepsilon\varphi(x+1), \qquad \lambda(x) = d(x)^{-1},$$

$$T_i(k,t) = \exp\left\{ \lambda(e^t) S_i(k) - \frac{3\lambda^2(e^t)}{4} \sum_{j=1}^{i} E\left(\overline{w}_{j+1}^2|F_j\right)\left[f_j(k)\right]^2 \right\}, \quad 1 \le i \le e^t,$$

$$T_0(k,t) = 0.$$

We know that for any fixed k and t, $\{T_i(k,t),\ 0 \le i \le e^t\}$ is a supermartingale [see Lemma 5.4.1, Stout (1974)]. Further, from the properties of $\varphi(x)$ we have

$$\lambda(i) = [2\varepsilon\varphi(i+1)]^{-1} \ge [2\varepsilon\varphi(e^t+1)]^{-1} = \lambda(e^t), \qquad i \le e^t.$$

Then

$$S_i(k) - \frac{3}{4}\lambda(i) \sum_{j=1}^{i} E\left(\overline{w}_{j+1}^2|F_j\right) f_j^2(k)$$

$$\le S_i(k) - \frac{3\lambda(e^t)}{4} \sum_{j=1}^{i} E\left(\overline{w}_{j+1}^2|F_j\right) f_j^2(k), \qquad i < e^t.$$

So

$$P\left\{\max_{1\le k\le ct^\alpha}\max_{1\le i\le e^t}\left[S_i(k)-\frac{3\lambda(i)}{4}\sum_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]\right.$$

$$\left.\ge(2+\alpha)d(e^t)\log\log e^t\right\}$$

$$\le\sum_{k=1}^{[ct^\alpha]}P\left\{\max_{1\le i\le e^t}\left[S_i(k)-\frac{3\lambda(e^t)}{4}\sum_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]\right.$$

$$\left.\ge(2+\alpha)d(e^t)\log\log e^t\right\}$$

$$\le\sum_{k=1}^{[ct^\alpha]}P\left\{\max_{1\le i\le e^t}T_i(k,t)\ge\exp[(2+\alpha)\log t]\right\}$$

$$\le\sum_{k=1}^{[ct^\alpha]}t^{-(2+\alpha)}=O(t^{-2})\quad\text{[by Corollary 5.4.1 in Stout (1974)].}$$

Then, according to the Borel–Cantelli Lemma, we have

$$\limsup_{t\to\infty}\max_{1\le k\le ct^\alpha}\max_{1\le i\le e^t}\frac{\left[S_i(k)-(3\lambda(i))/4\Sigma_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]}{\varphi(e^t+1)\log t}$$

$$\le 2(2+\alpha)\varepsilon\quad\text{a.s.}$$

Now let $h_n\le c(\log n)^\alpha$, then $h_n\le ct^\alpha$ for $n\le e^t$, we have from (2.9) and the above inequality that for $n\in[e^{t-1},e^t]$,

$$\max_{1\le k\le h_n}\max_{1\le i\le n}\frac{\left[S_i(k)-(3\lambda(i))/4\Sigma_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]}{\varphi(n)\log\log n}$$

$$\le\frac{\varphi(e^{t+1})}{\phi(e^t)}\frac{\varphi(e^t)}{\varphi(e^{t-1})}\frac{\log t}{\log(t-1)}$$

$$\times\frac{\max_{1\le k\le Ct^\alpha}\max_{1\le i\le e^t}\left[S_i(k)-(3\lambda(i))/4\Sigma_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]}{\varphi(e^t+1)\log t}$$

$$=O(\varepsilon)\quad\text{a.s. as } t\to\infty.$$

Thus,

$$\max_{1\le k\le h_n}\max_{1\le i\le n}\left[S_i(k)-\frac{3\lambda(i)}{4}\sum_{j=1}^{i}E\left(\overline{w}_{j+1}^2|F_j\right)f_j^2(k)\right]$$

$$=O(\varepsilon\varphi(n)\log\log n)\quad\text{a.s.}$$

and

$$\max_{1\le k\le h_n}\max_{1\le i\le n} S_i(k) \le \max_{1\le k\le h_n}\frac{3}{4}\lambda(n)\sum_{j=1}^{n} E\big(\overline{w}_{j+1}^2|F_j\big) f_j^2(k) + O(\varepsilon\varphi(n)\log\log n)$$

$$(3.10)\quad \le \max_{1\le k\le h_n}\frac{3}{4}\lambda(n)\sum_{j=1}^{n} E\Big[|\overline{w}_{j+1}|\lambda(j)^{-1}\Big|F_j\Big] f_j^2(k) + O(\varepsilon\varphi(n)\log\log n)$$

$$\le \sup_j E\big(|w_{j+1}|\,|F_j\big)\max_{1\le k\le h_n}\sum_{j=1}^{n} f_j^2(k) + O(\varepsilon\varphi(n)\log\log n).$$

The similar result holds also for $\{-S_i(k)\}$. Then the desired result (3.3) follows from (3.7)–(3.10) and the arbitrariness of ε.

(ii) Note that the condition (3.4) is crucial only in the proof of (3.9). When (3.4) is relaxed to (3.6), (3.9) can be replaced by

$$(3.9')\quad \max_{1\le k\le h_n}\max_{1\le i\le n}\left|\sum_{j=1}^{i} E\big(\tilde{w}_{j+1}|F_j\big) f_j(k)\right| \le \sup_j E\big(|w_{j+1}|\,|F_j\big)\max_{1\le k\le h_n}\sum_{j=1}^{n}|f_j(k)|.$$

Note also that

$$\max_{1\le k\le h_n}\sum_{j=1}^{n}|f_j(k)|^2 \le \max_{1\le k\le h_n}\sum_{j=1}^{n}|f_j(k)|,$$

then (3.5) follows for (3.7), (3.8), (3.9′) and (3.10). □

We remark that if h_n is only assumed to satisfy $h_n = O(n^\alpha)$, $\alpha > 0$, then the results (3.3) and (3.5) still hold, provided that $\log\log n$ in them is replaced by $\log n$.

LEMMA 3.2. *Under the conditions of Lemma* 3.1 *except* (3.2); *if* (3.4) *holds, then*

$$(3.11)\quad \max_{1\le k\le h_n}\max_{1\le i\le n}\left\|\sum_{j=1}^{i} f_j(k)w_{j+1}\right\| = O(a_n\log a_n) + o(a_n\varphi(n)\log\log n)\quad a.s.,$$

where

$$a_i = \max_{1\le k\le h_n} g_i(k),\qquad g_i(k) = \left[\sum_{j=1}^{i}\|f_j(k)\|^2 + 1\right]^{1/2},\qquad g_0(k) = 1.$$

PROOF. Let $x_j(k) = f_j(k)/g_j(k)$, $1 \le j \le n$. Then $\|x_j(k)\| \le 1$. So we have from Lemma 3.1 that

$$\max_{1\le k\le h_n}\max_{1\le i\le n}\left\|\sum_{j=1}^{i} x_j(k)w_{j+1}\right\| = O\left(\max_{1\le k\le h_n}\sum_{j=1}^{n}\|x_j(k)\|^2\right) + o(\varphi(n)\log\log n) \quad \text{a.s.} \tag{3.12}$$

Note that (omitting the dependence on k)

$$\sum_{j=1}^{n}\|x_j\|^2 = \sum_{j=1}^{n}\frac{g_j^2 - g_{j-1}^2}{g_j^2} \le \sum_{j=1}^{n}\int_{g_{j-1}^2}^{g_j^2}\frac{dx}{x} = 2\log g_n. \tag{3.13}$$

Also,

$$\begin{aligned}\sum_{j=1}^{i} f_j w_{j+1} = \sum_{j=1}^{i} g_j x_j w_{j+1} &= -\sum_{j=1}^{i}\sum_{t=j}^{i-1}(g_{t+1}-g_t)x_j w_{j+1} + g_i\sum_{j=1}^{i}x_j w_{j+1}\\ &= -\sum_{t=1}^{i-1}(g_{t+1}-g_t)\sum_{j=1}^{t}x_j w_{j+1} + g_i\sum_{j=1}^{i}x_j w_{j+1}.\end{aligned}$$

So

$$\begin{aligned}&\max_{1\le k\le h_n}\max_{1\le i\le n}\left\|\sum_{j=1}^{i} f_j(k)w_{j+1}\right\|\\ &\quad\le \max_{1\le k\le h_n}\max_{1\le t\le n}\left\|\sum_{j=1}^{t} x_j(k)w_{j+1}\right\|\\ &\qquad\times \max_{1\le k\le h_n}\max_{1\le i\le n}\left\{\sum_{t=1}^{i-1}[g_{t+1}(k)-g_t(k)] + g_i(k)\right\}\\ &\quad\le 2a_n\max_{1\le k\le h_n}\max_{1\le i\le n}\left\|\sum_{j=1}^{i} x_j(k)w_{j+1}\right\|.\end{aligned} \tag{3.14}$$

Thus, (3.11) follows from (3.12)–(3.14). □

LEMMA 3.3. *Suppose $\{w_j, F_j\}$ is an m-dimensional martingale difference sequence satisfying*

$$\sup_j E\left(\|w_{j+1}\|^2 \middle| F_j\right) < \infty, \qquad \|w_t\| = o(\varphi(t)) \quad a.s. \tag{3.15}$$

Let $\psi_j(k) = [x_{j1}^\tau, x_{j2}^\tau, \dots, x_{jk}^\tau]^\tau$, where x_{jt} is F_j-measurable, $t = 1, 2, \dots,$ and *assume that* h_n *and* $\varphi(x)$ *are the same as those in Lemma* 3.1. *Set*

$$M_i(k) = \sum_{j=1}^{i} \psi_j(k)\psi_j(k)^\tau + \beta I, \qquad \beta > 0.$$

$$S_i(k) = \sum_{j=1}^{i} \psi_j(k) w_{j+1}^\tau, \qquad S_0(k) = 0.$$

$$V_i(k) = \left\| [M_i(k)]^{-1/2} S_i(k) \right\|^2,$$

$$U_i(k) = \sum_{j=1}^{i} \left\| \psi_j^\tau(k) [M_j(k)]^{-1} S_j(k) \right\|^2, \qquad 1 \le i \le n, 1 \le k \le h_n.$$

Then, as $n \to \infty$

$$\max_{1 \le k \le h_n} V_n(k) = O(\delta_n) + o\big([\varphi(n)\log\log n]^2\big) \quad a.s. \tag{3.16}$$

$$\max_{1 \le k \le h_n} U_n(k) = O(\delta_n) + o\big([\varphi(n)\log\log n]^2\big) \quad a.s., \tag{3.17}$$

where $\delta_n = h_n \log_n + \lambda_{\max}(M_n(h_n))$

PROOF. Denoting $c_i(k) = [1 + \psi_i^\tau(k) M_{i-1}(k)^{-1} \psi_i(k)]^{-1}$, by the matrix inverse formula we know that

$$[M_i(k)]^{-1} = [M_{i-1}(k)]^{-1} - c_i(k)[M_{i-1}(k)]^{-1}\psi_i(k)\psi_i^\tau(k)[M_{i-1}(k)]^{-1}.$$

Then by a standard treatment [see, e.g., Moore (1978); Solo (1979); Chen (1982); Lai and Wei (1982a); Chen and Guo (1986); Guo, Huang and Hannan (1990)] we have the following relationship:

$$\begin{aligned} \operatorname{tr}\{S_i(k)^\tau [M_i(k)]^{-1} S_i(k)\} &= \operatorname{tr}\{S_{i-1}(k)^\tau [M_{i-1}(k)]^{-1} S_{i-1}(k)\} \\ &\quad + 2c_i(k)\psi_i(k)^\tau M_{i-1}(k)^{-1} S_{i-1} w_{i+1} \\ &\quad - c_i(k) \left\| \psi_i(k)^\tau [M_{i-1}(k)]^{-1} S_{i-1}(k) \right\|^2 \\ &\quad + c_i(k)\psi_i(k)^\tau [M_{i-1}(k)]^{-1} \psi_i(k) \|w_{i+1}\|^2 \\ &\le \operatorname{tr}\{S_{i-1}(k)^\tau [M_{i-1}(k)]^{-1} S_{i-1}(k)\} \\ &\quad + 2\psi_i(k)^\tau [M_i(k)]^{-1} S_{i-1}(k) w_{i+1} \\ &\quad - \left\| \psi_i(k)^\tau M_i(k)^{-1} S_{i-1}(k) \right\|^2 \\ &\quad + \psi_i(k)^\tau M_i(k)^{-1} \psi_i(k) \|w_{i+1}\|^2. \end{aligned} \tag{3.18}$$

For any fixed k, summing up from $i = 1$ to n we have

$$\begin{aligned} V_n(k) &+ \sum_{i=1}^{n} \left\| \psi_i(k)^\tau [M_i(k)]^{-1} S_{i-1}(k) \right\|^2 \\ &\le \sum_{i=1}^{n} \psi_i(k)^\tau M_i(k)^{-1} \psi_i(k) \|w_{i+1}\|^2 \\ &\quad + 2\sum_{i=1}^{n} \psi_i(k)^\tau M_i(k)^{-1} S_{i-1}(k) w_{i+1}. \end{aligned} \tag{3.19}$$

From Lai and Wei [(1982a), Lemma 2] we know that for any fixed k,

$$\psi_i(k)^\tau M_i(k)^{-1}\psi_i(k) = \frac{\det(M_i(k)) - \det(M_{i-1}(k))}{\det(M_i(k))} \le 1,$$

and hence

$$\max_{1\le k\le h_n}\sum_{i=1}^{n}\psi_i(k)^\tau M_i(k)^{-1}\psi_i(k) \le \max_{1\le k\le h_n}\int_{\det(M_0(k))}^{\det(M_n(k))} x^{-1}\,dx = O(\delta_n) \quad \text{a.s} \tag{3.20}$$

Thus, it follows from (3.5), (3.15) and (3.20) that

$$\begin{aligned}
&\max_{1\le k\le h_n}\sum_{i=1}^{n}\psi_i(k)^\tau M_i(k)^{-1}\psi_i(k)\|w_{i+1}\|^2 \\
&= \max_{1\le k\le h_n}\left\{\sum_{i=1}^{n}\psi_1^\tau(k)M_1^{-l}(k)\psi_i(k)\left[\|w_{i+1}\|^2 - E\left(\|w_{i+1}\|^2|F_i\right)\right]\right. \\
&\qquad \left. + \sum_{i=1}^{n}\psi_1^\tau(k)M_1^{-l}(k)\psi_i(k)E\left(\|w_{i+1}\|^2|F_i\right)\right\} \\
&= o\left(\varphi(n)^2\log\log n\right) + O\left(\max_{1\le k\le h_n}\sum_{i=1}^{n}\psi_i(k)^\tau M_i(k)^{-1}\psi_i(k)\right) \\
&= o\left(\varphi(n)^2\log\log n\right) + O(\delta_n) \quad \text{a.s}
\end{aligned} \tag{3.21}$$

Let

$$a_n = \left\{\max_{1\le k\le h_n}\sum_{i=1}^{n}\left\|\psi_i(k)^\tau M_i(k)^{-1}S_{i-1}(k)\right\|^2\right\}^{1/2},$$

then it follows from (3.19), Lemma 3.2 and (3.21) that

$$\begin{aligned}
\max_{1\le k\le h_n} V_n(k) + [a_n]^2 &\le o\left(\varphi(n)^2\log\log n\right) + O(\delta_n) \\
&\qquad + o\left(a_n\varphi(n)\log\log n\right) + O(a_n\log a_n) \\
&\le O(\delta_n) + o\left([a_n]^2\right) + o\left([\varphi(n)\log\log n]^2\right) \quad \text{a.s.}
\end{aligned}$$

From this it is easy to conclude that (3.16) holds and that

$$[a_n]^2 = O(\delta_n) + o\left([\varphi(n)\log\log n]^2\right) \quad \text{a.s.} \tag{3.22}$$

Finally, (3.17) follows from (3.21), (3.22) and the inequality

$$\max_{1\le k\le h_n} U_n(k) \le 2[a_n]^2 + 2\max_{1\le k\le h_n}\sum_{i=1}^{n}\psi_i(k)^\tau M_i(k)^{-1}\psi_i(k)\|w_{i+1}\|^2. \qquad \square$$

Proof of Theorem 2.1. Set

$$Z_n = \left[\hat\varepsilon_1(n), \hat\varepsilon_2(n), \dots, \hat\varepsilon_n(n)\right]^\tau, \qquad W_n = [w_1, w_2, \dots, w_n]^\tau, \tag{3.23}$$

$$\begin{aligned}
E_n(k) &= \left[e_1(k), e_2(k), \dots, e_n(k)\right]^\tau, \\
e_t(k) &= \sum_{j=k+1}^{\infty}\left[H_j u_{t-j} - G_j y_{t-j}\right],
\end{aligned} \tag{3.24}$$

where we stipulate that $u_t = 0$, $t < -q_0$ and $y_t = 0$, $t < -p_0$, and where $\{H_j, G_j\}$ are defined from the expansion

$$C(z)^{-1}A(z) = I + \sum_{j=1}^{\infty} G_j z^j, \qquad C(z)^{-1}B(z) = \sum_{j=1}^{\infty} H_j z^j$$

From (1.1), (2.1), (2.6), (2.7) and (3.23)–(3.24), we know

$$Y_n = \Phi_n \alpha(h_n) + E_n(h_n) + W_n,$$

where $\alpha(k) = [-G_1, -G_2, \dots, -G_k; H_1, H_2, \dots, H_k]^\tau$. Substituting the above identity into (2.5), we have

$$Z_n = E_n(h_n) + W_n - \Phi_n(\Phi_n^\tau \Phi_n)^{-1}\Phi_n^\tau[E_n(h_n) + W_n].$$

So

$$\begin{aligned}\|Z_n - W_n\|^2 &= \left\|\left[I - \Phi_n(\Phi_n^\tau\Phi_n)^{-1}\Phi_n^\tau\right]E_n(h_n) - \Phi_n(\Phi_n^\tau\Phi_n)^{-1}\Phi_n^\tau W_n\right\|^2 \\ &\le \left\|\left[I - \Phi_n(\Phi_n^\tau\Phi_n)^{-1}\Phi_n^\tau\right]E_n(h_n)\right\|^2 + \left\|\Phi_n(\Phi_n^\tau\Phi_n)^{-1}\Phi_n^\tau W_n\right\|^2 \\ &\le \left\|E_n(h_n)\right\|^2 + \left\|(\Phi_n^\tau\Phi_n)^{-1/2}\Phi_n^\tau W_n\right\|^2\end{aligned}$$

Since for some $\rho \in (0,1)$, $\|H_j\| = O(\rho^j)$ and $\|G_j\| = O(\rho^j)$, $\forall j \ge 0$, we have from (2.10) and (2.34) that $\|E_n(h_n)\| \to 0$ a.s. as $n \to \infty$. Then (2.12) follows from (3.16) and the above inequality immediately.

Finally, by using Lemma 3.3, the proof of (2.11) can be carried out along the lines of that for Theorem 1 in Guo and Huang (1990). □

PROOF OF EXAMPLE 2.1. Let $M_i(h_n)$ and $S_i(h_n)$ be defined as in Lemma 3.3, but with $\psi_j(h_n)$ given by (2.1). Then combining these results in Guo, Huang and Hannan [(1990), Theorem 2.3 and Examples 2.1 and 2.2] with Theorem 4 in Huang (1987) we know that

$$\liminf_{n\to\infty} \lambda_{min}[M_n(h_n)]/n > 0 \quad \text{a.s.} \tag{3.25}$$

Consequently, by Lemma 3.6(iii) in Guo, Huang and Hannan (1990) we have

$$S_n(h_n)^\tau[M_n(h_n)]^{-1}S_n(h_n) = O(h_n \log\log n) \quad \text{a.s.} \tag{3.26}$$

Furthermore, by Lemma 1 and Theorem 4 in Huang (1987), it is easy to conclude that

$$\sup_{t \ge (h_n)^\alpha}\left\|\frac{M_t(h_n)}{t} - R(h_n)\right\| = o(1) \quad \text{a.s. for any } \alpha > 2, \tag{3.27}$$

where $R(h_n) = E\psi t(n)^\tau \ge 0$. Note that $R(h_n)$ is in fact uniformly positive definite because by (3.25) and (3.27),

$$\begin{aligned}\liminf_{n\to\infty}\lambda_{min}[R(h_n)] &\ge \liminf_{n\to\infty}\lambda_{min}\left[\frac{M_n(h_n)}{n}\right] - \liminf_{n\to\infty}\left\|\frac{M_n(h_n)}{n} - R(h_n)\right\| \\ &= \liminf_{n\to\infty}\lambda_{min}\left[\frac{M_n(h_n)}{n}\right] > 0 \quad \text{a.s.}\end{aligned}$$

Hence, by (3.27) we know that there exists a constant $\varepsilon_0 > 0$ such that

$$\liminf_{n\to\infty} \inf_{(h_n)^\alpha \le t} \frac{\lambda_{min}(M_t(h_n))}{t} > \varepsilon_0 \quad \text{a.s.} \tag{3.28}$$

Thus, we have from (3.20) that for any $\alpha > 2$,

$$\begin{aligned}
&\sum_{t=1}^{n} \left[\psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n)\right]^2 \\
&\le \sum_{t=1}^{[(h_n)^\alpha]} \left[\psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n)\right] + \sum_{t=[(h_n)^\alpha]+1}^{n} \frac{\|\psi_t(h_n)\|^4}{\lambda_{min}(M_t(h_n))^2} \\
&= O(h_n \log h_n) + O\left(\sum_{t=[(h_n)^\alpha]+1}^{n} \frac{(h_n)^2(\log t)^2}{t^2}\right) \\
&= O(h_n \log\log n) \quad \text{a.s.}
\end{aligned} \tag{3.29}$$

So it follows from (3.3) that

$$\begin{aligned}
&\left\|\sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n)(w_{t+1}w_{t+1}^\tau - \Sigma)\right\| \\
&= O(h_n \log\log n) + o\left(\log n(\log\log n)^2\right) \\
&= o(h_n \log n) \quad \text{a.s.}
\end{aligned} \tag{3.30}$$

From (3.11) and (3.22) we know that

$$\left\|\sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1} S_{t-1}(h_n) w_{t+1}\right\| = o(h_n \log n) \quad \text{a.s.} \tag{3.31}$$

Also, similar to (3.20) it is not difficult to verify that

$$\begin{aligned}
\sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n) &= \sum_{i=1}^{n} \frac{\det(M_i(h_n)) - \det(M_{i-1}(h_n))}{\det(M_i(h_n))} \\
&\le h_n \log n + o(h_n \log n) \quad \text{a.s.}
\end{aligned} \tag{3.32}$$

On the other hand, similar to Lemma 2 of Lai and Wei (1982a), it is easy to see that

$$\psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n) = \frac{\det(M_t(h_n)) - \det(M_{t-1}(h_n))}{\det(M_{t-1}(h_n))},$$

then by invoking (3.28), we obtain

$$\begin{aligned}
&\sum_{t=[(h_n)^\alpha]+1}^{n} \psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n) \\
&= \sum_{t=[(h_n)^\alpha]+1}^{n} \frac{\det(M_t(h_n)) - \det(M_{t-1}(h_n))}{\det(M_{t-1}(h_n))} \\
&\ge h_n \log n + o(h_n \log n).
\end{aligned}$$

Hence similar to (3.29), it is not difficult to show that for any $\alpha > 2$,

$$\begin{aligned}
\sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n) &= \sum_{t=[(h_n)^\alpha]+1}^{n} \frac{\psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n)}{1+\psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n)} \\
&\geq \sum_{t=[(h_n)^\alpha]+1}^{n} \left\{1-\psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n)\right\} \\
&\qquad \times \psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n) \\
&\geq h_n \log h_n + o(h_n \log n) \\
&\quad - \sum_{t=[(h_n)^\alpha]+1}^{n} \left\{\psi_t(h_n)^\tau [M_{t-1}(h_n)]^{-1}\psi_t(h_n)\right\}^2 \\
&= h_n \log n + o(h_n \log n) \quad \text{a.s.}
\end{aligned} \tag{3.33}$$

Combining (3.32) and (3.33) we get

$$\sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n) = h_n \log n + o(h_n \log n) \quad \text{a.s.} \tag{3.34}$$

Again, by invoking (3.28) and using the similar treatment as used in the derivation of (3.18) and (3.19), we have after some manipulations,

$$\begin{aligned}
&S_n^\tau(h_n)[M_n(h_n)]^{-1}S_n(h_n) + \sum_{t=1}^{n} S_t(h_n)^\tau [M_t(h_n)]^{-1}\psi_t(h_n)\psi_t(h_n)^\tau \\
&\qquad\qquad \times [M_t(h_n)]^{-1}S_t(h_n) + o(h_n \log n) \\
&= \sum_{i=1}^{n} w_{i+1}\psi_i(h_n)^\tau M_i(h_n)^{-1}S_{i-1}(h_n) \\
&\quad + \sum_{i=1}^{n} \left[w_{i+1}\psi_i(h_n)^\tau M_i(h_n)^{-1}S_{i-1}(h_n)\right]^\tau \\
&\quad + \sum_{t=1}^{n} \psi_t(h_n)^\tau M_t(h_n)^{-1}\psi_t(h_n)\left[(w_{t+1}w_{t+1}^\tau - \Sigma) + \Sigma\right].
\end{aligned}$$

Hence, by (3.26), (3.30), (3.31) and (3.34) we conclude that

$$\begin{aligned}
&\sum_{t=1}^{n} S_t(h_n)^\tau [M_t(h_n)]^{-1}\psi_t(h_n)\psi_t(h_n)^\tau [M_t(h_n)]^{-1}S_t(h_n) \\
&\qquad = (h_n \log n)\Sigma + o(h_n \log n).
\end{aligned} \tag{3.35}$$

Finally, the desired result (2.13) follows from (3.35), since a similar argument as used in the proof of Theorem 1 in Guo and Huang (1989) shows that

$$\begin{aligned}
&\sum_{t=1}^{n} [\hat{w}_t(h_n) - w_t][\hat{w}_t(h_n) - w_t]^\tau \\
&\quad = \sum_{t=1}^{n} S_t(h_n)^\tau [M_t(h_n)]^{-1}\psi_t(h_n)\psi_t(h_n)^\tau [M_t(h_n)]^{-1}S_t(h_n) \\
&\qquad + o(h_n \log n).
\end{aligned}$$

4. Nonnegative supermartingale convergence theory and the proof of Theorem 2.2.

LEMMA 4.1. *Suppose that X, Y and W are any $n \times l$ and $n \times r$-dimensional matrices, respectively. Let $M = [X, Y]$ and $M^\tau M$ be invertible. Then*

$$W^\tau X(X^\tau X)^{-1} X^\tau W \le W^\tau M(M^\tau M)^{-1} M^\tau W \le W^\tau W. \tag{4.1}$$

PROOF. The lemma follows immediately by noting that $M(M^\tau M)^{-1}M^\tau$ is the projection operator on the subspace spanned by column vectors of M. □

LEMMA 4.2. *Under the notations of Lemma 3.3, if (3.15) is replaced by (2.14), and $h_n = [c(\log n)^\alpha]$, $\operatorname{tr}\{M_n(h_n)\} = O(h_n n^b)$, for some constants $b, c > 0$ and $\alpha > 1$. Then as $n \to \infty$,*

$$\max_{1 \le k \le h_n} V_n(k) = O\left([h_n]^2 \log n \{\log\log n\}^{2+\delta}\right) \quad a.s., \forall \delta > 0, \tag{4.2}$$

$$\sum_{i=0}^{n} \left\| \psi_i^\tau(h_i) M_i^{-1}(h_i) S_i(h_i) \right\|^2 = O\left([h_n]^2 \log n \{\log\log n\}^{2+\delta}\right) \quad a.s., \forall \delta > 0, \tag{4.3}$$

$$\sum_{i=0}^{n} \left\| \psi_i^\tau(h_n) M_i^{-1}(h_n) S_i(h_n) \right\|^2 = O\left([h_n]^2 \log n \{\log\log n\}^{2+\delta}\right) \quad a.s., \forall \delta > 0. \tag{4.4}$$

PROOF. Set

$$\begin{aligned} T_{n+1} &= \operatorname{tr}\{S_n^\tau(h_n) M_n^{-1}(h_n) S_n(h_n)\}, \\ d_n(h_n) &= \operatorname{tr}\{S_{n-1}^\tau(h_n) M_{n-1}^{-1}(h_n) S_{n-1}(h_n)\}, \end{aligned} \tag{4.5}$$

$$\begin{aligned} \alpha_n &= \frac{1}{2} \left\| \psi_n^\tau(h_n) M_n^{-1}(h_n) S_n(h_n) \right\|^2, \\ \beta_n &= 2\psi_n^\tau(h_n) M_n^{-1}(h_n) \psi_n(h_n), \end{aligned} \tag{4.6}$$

$$\gamma_n = \begin{cases} d_n(h_n), & \text{if } n = n_k \text{ for some integer } k, \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

where n_k is defined by $n_k = [\exp(k/c)^{1/\alpha}] + 1$, so that by the definition of h_n,

$$h_n = k, \qquad n_k \le n < n_{k+1}. \tag{4.8}$$

By (2.14) and Lemma 4.1, it can be verified that

$$E \operatorname{tr}\{S_i^\tau(k) M_i^{-1}(k) S_i(k)\} < \infty, \qquad E\left|\psi_i^\tau(k) M_i^{-1}(k) S_{i-1}(k) w_{i+1}\right| < \infty, \quad \text{for all } i \text{ and } k.$$

So by noting

$$\begin{aligned} \left\| \psi_i^\tau(k) M_i^{-1}(k) S_{i-1}(k) \right\|^2 \ge \frac{1}{2} \left\| \psi_i^\tau(k) M_i^{-1}(k) S_i(k) \right\|^2 \\ - \psi_i^\tau(k) M_i^{-1}(k) \psi_i(k) \|w_{i+1}\|^2, \end{aligned} \tag{4.9}$$

we know from (3.18) that

$$E[T_{n+1}|F_n] \le T_n + \gamma_n - \alpha_n + \beta_n E\left[\|w_{n+1}\|^2|F_n\right]. \tag{4.10}$$

Now, for any $\delta > 0$, let us denote $\lambda_n = h_n(\log h_n)^{1+\delta}\rho_n(\log \rho_n)^{1+\delta}$ with $\pi_n = h_n \log[\operatorname{tr} M_n(h_n)]$. Then it is easy to see that there exists an appropriately large integer N, such that for any $n \le N$, $\lambda_n > 0$, λ_n is F_n-measurable and $\lambda_n \le \lambda_{n+1}$. So it follows from (4.10) that

$$E\left[\frac{T_{n+1}}{\lambda_{n+1}}\middle|F_n\right] \le \frac{T_n}{\lambda_n} - \frac{\alpha_n}{\lambda_n} + \frac{\gamma_n + \beta_n E\left[\|w_{n+1}\|^2|F_n\right]}{\lambda_n}.$$

Hence, if we can prove that

$$\sum_{n=N}^{\infty} \frac{\gamma_n + \beta_n E\left[\|w_{n+1}\|^2|F_n\right]}{\lambda_n} < \infty \quad \text{a.s.}, \tag{4.11}$$

then by the nonnegative supermartingale convergence results in Neveu (1975) [see also Solo, (1979), page 961 and Goodwin and Sin (1984) page 501] we will have

$$\frac{T_n}{\lambda_n} \to T < \infty \quad \text{a.s.} \tag{4.12}$$

$$\sum_{n=N}^{\infty} \frac{\alpha_n}{\lambda_n} < \infty \quad \text{a.s.} \tag{4.13}$$

We proceed as follows. Let k_0 be the positive integer such that $n_{k0} = N$. Then by (4.6) and (4.8) we know that

$$\begin{aligned}
\frac{1}{2}\sum_{n=M}^{\infty}\frac{\beta_n}{\lambda_n} &= \frac{1}{2}\sum_{k=k_0}^{\infty}\sum_{n=n_k}^{n_{k+1}-1}\frac{\beta_n}{\lambda_n}\\
&= \sum_{k=k_0}^{\infty}\sum_{n=n_k}^{n_{k+1}-1}\frac{\det[M_n(k)] - \det[M_{n-1}(k)]}{\det[M_n(k)]\rho_n(\log\rho_n)^{1+\delta}k(\log k)^{1+\delta}}\\
&\le c_0\sum_{k=k_0}^{\infty}\frac{1}{k(\log k)^{1+\delta}} \qquad\qquad (4.14)\\
&\quad\times\sum_{n=n_k}^{n_{k+1}-1}\frac{\det[M_n(k)] - \det[M_{n-1}(k)]}{\det[M_n(k)]\log[\det M_n(k)]\{\log\log[\det M_n(k)]\}^{1+\delta}}\\
&\le c_0\sum_{k=k_0}^{\infty}\frac{1}{k(\log k)^{1+\delta}}\int_{\det[M_{n_k-1}(k)]}^{\det[M_{n_{k+1}-1}(k)]}\frac{dx}{x\log x(\log\log x)^{1+\delta}} < \infty,
\end{aligned}$$

where c_0 is the constant satisfying $c_0 h_n =$ dimension of $M_n(h_n)$.

Next, we consider $\sum_{n=N}^{\infty}\gamma_n/\lambda_n$. For any suitably large n (say $n \ge N$), denote

$$\mu_i(n) = h_n \log[\operatorname{tr} M_i(h_n)]\log^{1+\delta}\{h_n \log[\operatorname{tr} M_i(h_n)]\}, \qquad i \ge 1. \tag{4.15}$$

Then $\mu_i(n)$ is positive, F_i-measurable and $\mu_i(n) \le \mu_{i+1}(n)$. Again, dividing $\mu_i(n)$ on both sides of (3.18), noting (4 9) and taking mathematical expectations, yields

$$E\frac{d_{i+1}(h_n)}{\mu_{i+1}(n)} \le E\frac{d_i(h_n)}{\mu_i(n)} - \frac{1}{2}E\left\{\frac{\left\|\psi_i^\tau(h_n)M_i^{-1}(h_n)S_i(h_n)\right\|^2}{\mu_i(n)}\right\} + 2E\left\{\frac{\psi_i^\tau(h_n)M_i^{-1}(h_n)\psi_i(h_n)E\left[\|w_{i+1}\|^2|F_i\right]}{\mu_i(n)}\right\}. \tag{4.16}$$

Summing up from N to $n-1$, and noting (2 14) and

$$\sum_{i=0}^{n-1}\psi_i^\tau(h_n)M_i^{-1}(h_n)\psi_i(h_n)/\mu_i(n) \le c_1 < \infty, \tag{4.17}$$

for some deterministic constant c_1, we conclude that

$$\sup_{n>N} E\frac{d_n(h_n)}{\mu_n(n)} < \infty. \tag{4.18}$$

Thus, by (4.7) and (4.8),

$$E\sum_{n=N}^{\infty}\frac{\gamma_n}{\lambda_n} = E\sum_{k=k_0}^{\infty}\frac{d_{n_k}(h_{n_k})}{\mu_{n_k}(n_k)h_{n_k}\{\log h_{n_k}\}^{1+\delta}} \le \sup_k\left\{E\frac{d_{n_k}(h_{n_k})}{\mu_{n_k}(n_k)}\right\}\sum_{k=k_0}^{\infty}\frac{1}{k(\log k)^{1+\delta}} < \infty,$$

and so $\sum_{n=N}^{\infty}\gamma_n/\lambda_n < \infty$ a.s This together with (2.14) and (4 14) yields (4.11). Consequently, (4.12) and (4 13) hold.

Now, by the assumption $\operatorname{tr}\{M_n(h_n)\} = O(h_n n^b)$, we know that

$$\lambda_n = O\left([h_n]^2 \log n(\log\log n)^{2+\delta}\right), \forall\, \delta > 0.$$

Therefore assertion (4.3) follows from (4.13) and the Kronecker lemma immediately, while (4.2) follows from (4.12) and Lemma 4.1:

$$\max_{1\le k\le h_n} V_n(k) = V_n(h_n) \le T_{n+1} = O\left([h_n]^2\log n(\log\log n)^{2+\delta}\right) \quad \text{a.s.}$$

To complete the proof of the lemma, we have to verify the last assertion (4.4).

Similar to the derivation of (4.18), summing up both sides of (4 16) from N to n and noting (4.17), we see that

$$\sup_{n>N} E\left[\sum_{i=N}^{n}\left\|\psi_i^\tau(h_n)M_i^{-1}(h_n)S_i(h_n)\right\|^2/\mu_i(n)\right] < \infty.$$

Then by the fact that $\mu_i(n) \le \mu_{i+1}(n)$, $i \le n$, it follows that for $U_n(h_n)$ defined as in Lemma 3 3,

$$\sup_{n>N} E\left[U_n(h_n)/\mu_n(n)\right] < \infty.$$

Consequently, by (4.8),

$$P\left\{U_{n_k-1}(h_{n_k-1}) > \mu_{n_k-1}(n_k - 1)h_{n_k}(\log h_{n_k})^{1+\delta}\right\} = O\left(\frac{1}{k(\log k)^{1+\delta}}\right).$$

Then by $\mu_n(n) = O(h_n \log n\{\log\log n\}^{1+\delta})$ and the Borel–Cantelli Lemma,

$$\limsup_{k\to\infty} \frac{U_{n_k-1}(k-1)}{\left[h_{n_k}\right]^2 \log n_k\left[(\log h_{n_k})(\log\log n_k)\right]^{1+\delta}} < \infty \quad \text{a.s.}$$

Finally, from this, (4.8) and $U_i(h_n) \le U_{i+1}(h_n)$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{U_n(h_n)}{[h_n]^2 \log n[(\log h_n)(\log\log n)]^{1+\delta}} \\
&\quad\le \limsup_{k\to\infty} \sup_{n\in[n_k n_{k+1}-1]} \frac{U_n(h_n)}{[h_n]^2 \log n[(\log h_n)(\log\log n)]^{1+\delta}} \\
&\quad\le \limsup_{k\to\infty} \frac{U_{n_{k+1}-1}(k)}{\sigma_{k+1}} \frac{\sigma_{k+1}}{\sigma_k} < \infty,
\end{aligned}$$

where $\sigma_k \triangleq [h_{n_k}]^2 \log n_k[(\log h_{n_k})(\log\log n_k)]^{1+\delta}$. Hence, (4.4) is true. □

PROOF OF THEOREM 2.2. This proof is again similar to that for Theorem 1 in Guo and Huang (1989), but with results in Lemma 1 of that paper replaced by those in the present Lemma 4.2. □

REMARK 4.1. If (2.14) and the second condition in (2.10) are replaced by, respectively,

(4.19) $$\sup_k E\left[\|w_{k+1}\|^2 \middle| F_k\right] \le \sigma < \infty \quad \text{and} \quad E\left\{\|u_n\|^2 + \|y_n\|^2\right\} = O(n^b),$$

where b and σ are some nonnegative deterministic constants, then the results in Theorem 2.2 can be improved to $O([h_n]^2 \log n[\log\log n]^{1+\delta})$. However, in applications the second condition in (4.19) seems to be less applaudable than that in (2.10).

5. Proof of Theorem 2.3. We first show that under the conditions of Theorem 2.3,

(5.1) $$\limsup_{n\to\infty} \hat{m}(n) \le m_0 \quad \text{a.s.}$$

Let us set

(5.2) $$\phi_t^0(p,q,r) = \left[y_t^\tau, y_{t-1}^\tau, \dots, y_{t-p+1}^\tau, u_t^\tau, u_{t-1}^\tau, \dots, u_{t-q+1}^\tau, w_t^\tau, w_{t-1}^\tau, \dots, w_{t-r+1}^\tau\right]^\tau,$$

(5.3) $$X_n^0(p,q,r) = \left[\phi_0^0(p,q,r), \phi_1^0(p,q,r), \dots, \phi_{n-1}^0(p,q,r)\right]^\tau,$$

(5.4) $$\theta(p,q,r) = \left[-A_1, \dots, -A_p, B_1, \dots, B_q, C_1, \dots, C_r\right]^\tau,$$

where $A_p = 0$, $p > p_0$, $B_q = 0$, $q > q_0$, $C_r = 0$, $r > r_0$. When $p = q = r = k$, we will simply write (k) for (k, k, k) in ϕ_t, ϕ_t^0, X_n and X_n^0, etc. Then with Y_n, $X_n(p, q, r)$ and W_n defined by (2.6), (2.18) and (3.23), it follows from (1.1) that for any $k \leq m_0$,

$$Y_n = X_n^0(k)\theta(k) + W_n = X_n(k)\theta(k) + \left[X_n^0(k) - X_n(k)\right]\theta(k) + W_n. \tag{5.5}$$

Let

$$R_n = \left[X_n^0(k) - X_n(k)\right]\theta(k). \tag{5.6}$$

It is easy to see that R_n does not depend on k when $k \geq m_0$. Then by (2.19), (5.5) and (5.6) it follows that

$$Z_n(k) = [R_n + W_n] - X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)[R_n + W_n]$$

and

$$\begin{aligned} Z_n^\tau(k)Z_n(k) = [R_n + W_n]^\tau[R_n + W_n] - [R_n + W_n]^\tau X_n(k) \\ \times [X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)[R_n + W_n]. \end{aligned} \tag{5.7}$$

It is evident that for $k \geq m_0$,

$$\begin{aligned} &[R_n + W_n]^\tau X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)[R_n + W_n] \\ &\quad = O\left(\left\| R_n^\tau X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)R_n \right\|\right) \\ &\qquad + O\left(\left\| W_n^\tau X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)W_n \right\|\right). \end{aligned} \tag{5.8}$$

By Lemma 4.1 and Theorem 2.1, it follows that

$$\begin{aligned} &\max_{m_0 \leq k \leq \log n} \left\| R_n^\tau X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)R_n \right\| \\ &\quad \leq \|R_n^\tau R_n\| = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right) \quad \text{a.s.} \end{aligned} \tag{5.9}$$

On the other hand, by Lemmas 3.3 and 4.1,

$$\begin{aligned} &\max_{m_0 \leq k \leq \log n} \left\| W_n^\tau X_n(k)[X_n^\tau(k)X_n(k)]^{-1}X_n^\tau(k)W_n \right\| \\ &\quad = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right). \end{aligned} \tag{5.10}$$

Thus, from (5.7)–(5.10) we have

$$\begin{aligned} &\max_{m_0 \leq k \leq \log n} \left| \operatorname{tr} Z_n^\tau(k)Z_n(k) - \operatorname{tr}[R_n + W_n]^\tau[R_n + W_n] \right| \\ &\quad = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right) \quad \text{a.s.} \end{aligned}$$

Consequently, by (2.21),

$$\begin{aligned} &\max_{m_0 < k \leq \log n} \{\mathrm{CIC}(m_0)_n - \mathrm{CIC}(k)_n\} \\ &\quad = \max_{m_0 < k \leq \log n} \Big\{ \operatorname{tr} Z_n^\tau(m_0)Z_n(m_0) - \operatorname{tr}[R_n + W_n]^\tau[R_n + W_n] \\ &\qquad\qquad - \operatorname{tr} Z_n^\tau(k)Z_n(k) + \operatorname{tr}[R_n + W_n]^\tau[R_n + W_n] - 3(k - m_0)a_n \Big\} \\ &\quad \leq 2 \max_{m_0 \leq k \leq \log n} \left| \operatorname{tr} Z_n^\tau(k)Z_n(k) - \operatorname{tr}[R_n + W_n]^\tau[R_n + W_n] \right| - 3a_n \\ &\quad = O(h_n \log n) + o\left(\{\varphi(n)\log\log n\}^2\right) - 3a_n < 0 \\ &\qquad\qquad \text{a.s. for sufficiently large } n, \end{aligned}$$

because of (2.27) Hence by the definition (2.22) for $\hat{m}(n)$, we see that (5.1) holds.

We now show that

$$\liminf_{n\to\infty} \hat{m}(n) \geq m_0 \quad \text{a.s.} \tag{5.11}$$

Let us write $\hat{\theta}_n(k)$ defined by (2.26) in its component form

$$\hat{\theta}_n(k) = \left[-\hat{A}_1, \dots, -\hat{A}_k, \hat{B}_1, \dots, \hat{B}_k, \hat{C}_1, \dots, \hat{C}_k\right]^\tau \tag{5.12}$$

and set for any $k \leq m_0$,

$$\hat{\theta}_n^0(k) = \left[-\hat{A}_1, \dots, -\hat{A}_{m0}, \hat{B}_1, \dots, \hat{B}_{m0}, \hat{C}_1, \dots, \hat{C}_{m0}\right]^\tau, \tag{5.13}$$

where $\hat{A}_i = 0$, $\hat{B}_i = 0$, $\hat{C}_i = 0$ for $i > k$.

Then it follows from (2.19), (2.26) and (5.5) that for $k \leq m_0$,

$$\begin{aligned} Z_n(k) &= Y_n - X_n(k)\hat{\theta}_n(k) = Y_n - X_n(m_0)\hat{\theta}_n^0(k) \\ &= Y_n - X_n(m_0)\theta(m_0) + X_n(m_0)\left[\theta(m_0) - \hat{\theta}_n^0(k)\right] \\ &= W_n + \tilde{X}_n\theta(m_0) + X_n(m_0)\tilde{\theta}_n(k), \end{aligned}$$

where $\tilde{X}_n = X_n^0(m_0) - X_n(m_0)$, $\tilde{\theta}_n(k) = \theta(m_0) - \hat{\theta}_n^0(k)$.

Hence for $k \leq m_0$,

$$\begin{aligned} \operatorname{tr} Z_n^\tau(k) Z_n(k) &= \operatorname{tr} \tilde{\theta}_n^\tau(k) X_n^\tau(m_0) X_n(m_0)\tilde{\theta}_n(k) \\ &\quad + 2\operatorname{tr} \tilde{\theta}_n^\tau(k) X_n^\tau(m_0)\left[W_n + \tilde{X}_n\theta(m_0)\right] \\ &\quad + \operatorname{tr}\left[W_n + \tilde{X}_n\theta(m_0)\right]^\tau\left[W_n + \tilde{X}_n\theta(m_0)\right] \end{aligned} \tag{5.14}$$

and for $k < m_0$,

$$\|\tilde{\theta}_n(k)\|^2 \geq \min\left\{\|A_{p0}\|^2, \|B_{q0}\|^2, \|C_{r0}\|^2\right\} \triangleq \delta_0 > 0. \tag{5.15}$$

With $\lambda_{\min}^0(n)$ defined by (2.29), it is easy to verify by (2.17) and (2.18) that

$$\lambda_{\min}^0(n) \leq 2\lambda_{\min}\{X_n^\tau(m_0)X_n(m_0)\} + 2\sum_{i=0}^{n-1}\left\|\phi_i(m_0) - \phi_i^0\right\|^2,$$

so by Theorem 2.1 and (2.27) and (2.28), it is evident that

$$\lambda_{\min}\{X_n^\tau(m_0)X_n(m_0)\} \geq \frac{1}{3}\lambda_{\min}^0(n), \quad \text{for sufficiently large } n,$$

then by (5.15), we obtain for any $k < m_0$,

$$\operatorname{tr} \tilde{\theta}_n^\tau(k) X_n^\tau(m_0) X_n(m_0)\tilde{\theta}_n(k) \geq \frac{\delta_0}{3}\lambda_{\min}^0(n), \quad \text{for sufficiently large } n. \tag{5.16}$$

Similar to (5.9) and (5.10), we have

$$\begin{aligned} &\left\|\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1/2}X_n^\tau(m_0)\left[W_n + \tilde{X}_n\theta(m_0)\right]\right\|^2 \\ &\qquad = O(h_n\log n) + o\left(\{\varphi(n)\log\log n\}^2\right). \end{aligned}$$

Therefore, for any $k \le m_0$,

$$
\begin{aligned}
&\left|2\,\mathrm{tr}\,\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]\right|\\
&\quad= O\Big(\left\|\tilde{\theta}_n^\tau(k)\left[X_n^\tau(m_0)X_n(m_0)\right]^{1/2}\right\|\\
&\qquad\times\left\|\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1/2}X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]\right\|\Big)\\
&\quad= O\Big(\left\{\mathrm{tr}\left[\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)X_n(m_0)\tilde{\theta}_n(k)\right]\right\}^{1/2}\\
&\qquad\times\left\{O(h_n\log n)+o\left(\left[\varphi(n)\log\log n\right]^2\right)\right\}^{1/2}\Big).
\end{aligned}
\tag{5.17}
$$

Hence, it follows from (5.16) and (5.17) that for any $k < m_0$,

$$
\begin{aligned}
&\mathrm{tr}\,\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)X_n(m_0)\tilde{\theta}_n(k)+2\,\mathrm{tr}\,\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]\\
&\quad\ge\frac{\delta_0}{3}\lambda_{\min}^0(n)\{1+O(1)\}\quad\text{a.s. as } n\to\infty.
\end{aligned}
\tag{5.18}
$$

Note that when $k = m_0$, we have by (2.26) and (5.5),

$$
\begin{aligned}
\tilde{\theta}_n(m_0) &= \theta(m_0)-\hat{\theta}_n(m_0)\\
&= -\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1}X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right],
\end{aligned}
\tag{5.19}
$$

So the first two terms on the R.H.S. of (5.14) can be rewritten as

$$
\begin{aligned}
&\mathrm{tr}\,\tilde{\theta}_n^\tau(m_0)X_n^\tau(m_0)X_n(m_0)\tilde{\theta}_n(m_0)\\
&\quad=\mathrm{tr}\left[W_n+\tilde{X}_n\theta(m_0)\right]^\tau X_n(m_0)\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1}\\
&\qquad\times X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right],
\end{aligned}
\tag{5.20}
$$

$$
\begin{aligned}
&2\,\mathrm{tr}\,\tilde{\theta}_n^\tau(m_0)X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]\\
&\quad=-2\,\mathrm{tr}\left[W_n+\tilde{X}_n\theta(m_0)\right]^\tau X_n(m_0)\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1}\\
&\qquad\times X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right].
\end{aligned}
\tag{5.21}
$$

Now, by (5.8)–(5.10) it is obvious that the quantity on the R.H.S. of (5.20) is bounded by $O(h_n \log n) + o(\{\varphi(n)\log\log n\}^2)$. Hence from (2.21), (5.14) and (5.18)–(5.21) we see that for any $k < m_0$,

$$
\begin{aligned}
\mathrm{CIC}(k)_n-\mathrm{CIC}(m_0)_n &= \mathrm{tr}\,\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)X_n(m_0)\tilde{\theta}_n(k)\\
&\quad+2\,\mathrm{tr}\,\tilde{\theta}_n^\tau(k)X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]\\
&\quad+\mathrm{tr}\left[W_n+\tilde{X}_n\theta(m_0)\right]^\tau X_n(m_0)\left[X_n^\tau(m_0)X_n(m_0)\right]^{-1}\\
&\qquad\times X_n^\tau(m_0)\left[W_n+\tilde{X}_n\theta(m_0)\right]+3(k-m_0)a_n\\
&\ge\frac{\delta_0}{3}\lambda_{min}^0(n)\{1+o(1)\}+O(a_n)\\
&\qquad+O(h_n\log n)+o\left(\{\varphi(n)\log\log n\}^2\right)\\
&=\lambda_{\min}^0(n)\left\{\frac{\delta_0}{3}+o(1)\right\}\quad\text{a.s. as } n\to\infty,
\end{aligned}
$$

where the last inequality holds because of (2.27) and (2.28). Thus it is easy to see that (5.11) is true Hence the first assertion (2.30) has been proved.

With $\hat{m}(n) \to m_0$ in mind, the proof of (2.31) can be carried out by a similar argument as that used above [see also, Guo, Chen and Zhang (1989), for related proofs]. As for the assertion (2.32), we note that (5.19) is also valid with (m_0) replaced by (p_0, q_0, r_0), hence

$$
\begin{aligned}
&\|\hat{\theta}_n(p_0, q_0, r_0) - \theta^*\|^2 \\
&\quad = \left\| [X_n^\tau(p_0, q_0, r_0) X_n(p_0, q_0, r_0)]^{-1} X_n^\tau(p_0, q_0, r_0) [W_n + \tilde{X}_n \theta(m_0)] \right\|^2 \\
&\quad \le \{\lambda_{\min}^0(n)\}^{-1} \left\| [X_n^\tau(p_0, q_0, r_0) X_n(p_0, q_0, r_0)]^{-1/2} \right. \\
&\qquad\qquad \left. \times X_n^\tau(p_0, q_0, r_0) [W_n + \tilde{X}_n \theta(m_0)] \right\|^2 \\
&\quad = \{\lambda_{\min}^0(n)\}^{-1} \{O(h_n \log n) + o([\varphi(n) \log\log n]^2)\},
\end{aligned}
$$

where the last inequality follows from a similar argument as that used in (5.8)–(5.10).

Finally, the results in the second part (ii) can be proved in a similar way □

APPENDIX

By (1.8) we have for all $\lambda \in [0, 2\pi]$,

$$
\begin{aligned}
I &> I - C(e^{i\lambda}) - C^\tau(e^{-i\lambda}) + C(e^{i\lambda})C^\tau(e^{-i\lambda}) \\
&= [I - C(e^{i\lambda})][I - C^\tau(e^{-i\lambda})].
\end{aligned} \tag{A1}
$$

So $\|C(e^{i\lambda}) - I\| < 1$, $\forall\ \lambda \in [0, 2\pi]$ Thus, for any complex vector x, $\|x\| = 1$, $|x^* C(e^{i\lambda})x - 1| < 1$, $\forall\ \lambda \in [0, 2\pi]$. Then by the maximum principle we know that $|x^* C(z)x - 1| < 1$, $\forall\ |z| \le 1$ Consequently, $x^* C(z)x \neq 0$, $\forall\ |z| \le 1$, which is tantamount to (1.6).

Integrating both sides of (A1) from $\lambda = 0$ to 2π, we have $\|[C_1, \dots, C_{r_0}]\| < 1$.

Acknowledgments. We would like to thank Professor E. J Hannan for his encouragement and valuable discussions throughout this work and also the referees for their helpful comments on an earlier version of the paper.

REFERENCES

AN, H Z., CHEN, Z. G. and HANNAN, E. J. (1982) Autocorrelation, autoregression and autoregressive approximation *Ann. Statist* **10** 926–936.

ASTROM, K J. (1968). Lectures on the identification problems—the least squares method. Report 6080, Div Aut. Contr., Lund Inst Tech., Sweden

CAINES, P. E. and LAFORTUNE, S (1984). Adaptive control with recursive identification for stochastic linear systems. *IEEE Trans Automat. Control* **29** 312–321

CHEN, H. F. (1982). Strong consistency and convergence rate of least squares identification *Sci Sinica Ser A* **25** 771–784.

CHEN, H F and GUO, L. (1986) Convergence rate of least squares identification and adaptive control for stochastic systems. *Internat. J Control* **44** 1459–1476.

DURBIN, J (1961) The fitting of time series model. *Internat. Statist Rev* **28** 233–244

GOODWIN, G C. and SIN, K. S. (1984). *Adaptive Filtering, Prediction and Control*. Prentice-Hall, Englewood Cliffs, N. J.

GUO, L., CHEN, H. F and ZHANG, J F. (1989). Consistent order estimation for linear stochastic feedback control systems (ARMAX model). *Automatica—J. IFAC* **25** 147–151

GUO, L and HUANG, D W (1989). Least squares identification for ARMAX models without the positive real condition. *IEEE Trans Automat Control* **34** 1094–1098.

GUO, L., HUANG, D. W and HANNAN, E. J. (1990). On ARX(∞) approximation *J Multivariate Anal* **32** 17–47

HANNAN, E. J and DEISTLER, M. (1988) *The Statistical Theory of Linear Systems* Wiley, New York

HANNAN, E J and KAVALIERIS, L. (1983). The convergence of autocorrelations and autoregressions *Austral J. Statist* **25** 287–297.

HANNAN, E. J. and KAVALIERIS, L (1984) Multivariate linear regression models *Adv. Appl Probab.* **16** 492–561

HANNAN, E. J and RISSANEN, J (1982). Recursive estimation of ARMA order *Biometrika* **69** 81–94.

HUANG, D W. (1987) Convergence rate of sample autocorrelations and autocovariances for stationary time series *Sci Sinica Ser. A* **10** (in Chinese); **31** 406–424 (1988) (in English).

HUANG, D W. (1989). Recursive method for ARMA model estimation. II *Acta Math. Appl. Sinica* **5** 332–354

LAI, T. L. and WEI, C Z. (1982a). Least squares identification 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). A law of the iterated logarithm for double arrays of independent random variables with applications to regression and time series models *Ann Probab* **10** 320–335.

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* **31** 898–906.

LJUNG, L. (1976) Consistency of the least squares identification method. *IEEE Trans. Automat Control* **21** 779–781

LJUNG, L (1977). On positive real function and the convergence of some recursive schemes. *IEEE Trans. Automat Control* **22** 539–551.

LJUNG, L. and SODERSTROM, T. (1983). *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Mass.

MOORE, J. B. (1978) On strong consistency of least squares algorithm *Automatica—J IFAC* **14** 505–509.

NEVEU, J. (1975) *Discrete-Parameter Martingales.* North-Holland, Amsterdam.

SOLO, V. (1979) The convergence of AML *IEEE Trans Automat. Control* **24** 958–962.

STOUT, W F (1974). *Almost Sure Convergence* Academic, New York

INSTITUTE OF SYSTEMS SCIENCE
ACADEMIA SINICA
BEIJING
CHINA

SCHOOL OF MATHEMATICS, QUT
2 GEORGE STREET, GPO BOX 2434
BRISBANE, Q. 4001
AUSTRALIA