*The Annals of Statistics*
1983, Vol. 11, No. 1, 78–85

# ON THE ESTIMATION OF THE PARAMETERS OF MARKOV PROBABILITY MODELS USING MACRO DATA

BY ADRIAAN P. VAN DER PLAS

*University of Amsterdam*

In this paper we consider the problem of estimating the parameters of a Markov model using so-called *macro data*. It will be shown that the stochastic process of the macro data is a Markov chain, which uniquely determines the probability structure of the underlying Markov model. A *conditional least squares estimator* exists under very weak conditions and this estimator is strongly consistent as time tends to infinity. Moreover this estimator is shown to be asymptotically normal under some additional assumptions.

**1. Introduction.** For the description of e.g. social mobility, voting change and changes in the sizes of firms, so-called *Markov models* are used frequently. These models consist of a fixed number $K$ of independently distributed Markov chains with a Markov matrix $P$ depending on a vector of parameters $\theta$.

In this paper, we consider the problem of making statistical inferences about $\theta$ in the case where very limited information is available. To be more specific, the samples observed are the so-called *macro data*, i.e. the number of Markov chains in each state for consecutive instants.

In Section 2 we show that the discrete-time stochastic process of these macro data is a Markov chain, with which it is possible to identify the probability structure of the process consisting of the underlying Markov chains. However the likelihood function based on these macro data will have an unattractive form if $K > 1$. Therefore we propose to use the method of least squares.

Miller (1952) proposed the method of least squares in the case where the chains take values in a finite space and where $\theta = P$. Then, for a fixed and sufficiently large sample $T$ of the macro data, Madansky (1959) proved that this estimator for the transition probabilities is consistent for the number of chains $K \to \infty$ under the assumption that the initial distribution is not invariant. The reader is referred to Lee et al (1970) for these types of "linear" estimators. We should note that the asymptotic behavior concerning the consistency of these estimators for $T \to \infty$, $K$ fixed, was considered numerically only.

We consider the case where the state space $S$ may be infinite and where the Markov matrix $P = P(\theta)$ depends on a finite dimensional parameter $\theta$. Then under very weak conditions we show that there exists a conditional least squares estimator for $\theta$, which is a.s. consistent (Theorem 3.2) and under additional, more technical assumptions this estimator turns out to be asymptotically normal (Theorem 3.5) for $T \to \infty$, $K$ fixed.

We are indebted to Jennrich (1969), who gave the first rigorous account of the method of non-linear least squares estimation in a regression model. In another context, Klimko and Nelson (1978) considered a non-linear least squares estimator for more general discrete-time stochastic processes. However, they assume rather strong conditions in order to prove the existence and a.s. consistency of the estimator.

**2. The model.** Suppose that a stochastic system may be described by $K$ independent, irreducible and ergodic Markov chains $X_k = \{X_k(t);\, t = 0, 1, \cdots\}$, each with a countable, possibly finite, state space $S$, each with Markov matrix $P = (p_{rs})_{r,s\in S}$ and each starting

Received November 1981; revised July 1982.

*AMS* 1970 *subject classifications.* 60J10, 62F10, 62M05

*Key words and phrases.* Markov models, macro data, grouped chain, least squares estimator, a.s. convergence, asymptotic normality, a.s. uniform convergence.

with the invariant distribution $\pi = (\pi_r)_{r \in S}$, i.e. the unique probability measure satisfying $\pi = \pi P$ (see, e.g., Feller, 1968).

If we set $X(t) = (X_1(t), X_2(t), \cdots, X_K(t))$ for $t = 0, 1, \cdots$, the system is described by a discrete time stochastic process $X = \{X(t); t = 0, 1, \cdots\}$ on the state space $S^K$. This process $X$ is a Markov chain with Markov matrix $\bar{P} = (\bar{p}_{xy})_{x,y \in S^K}$, where $\bar{p}_{xy} = \prod_{k=1}^{K} p_{x_k y_k}$ and with initial distribution $\bar{\pi} = (\bar{\pi}_x)_{x \in S^K}$, where $\bar{\pi}_x = \prod_{k=1}^{K} \pi_{x_k}$ for $x = (x_1, x_2, \cdots, x_K)$ and $y = (y_1, y_2, \cdots, y_K)$. From the assumptions on the Markov chains $X_k$ it follows that the chain $X$ is also irreducible and ergodic. Hence we have the following result.

THEOREM 2.1. *The Markov chain $X$ is irreducible and ergodic; the initial distribution $\bar{\pi}$ is invariant.* □

Next, we define the *macro data* $N(t) = (N_s(t))_{s \in S}$ by $N_s(t) = \#\{k \mid X_k(t) = s\}$ for $s \in S$, $t = 0, 1, 2, \cdots$; $\#$ denotes the number of elements in the set. Let $R = \{n \mid n = (n_s)_{s \in S}, n_s \in \{0, 1, \cdots, K\}, \sum_{s \in S} n_s = K\}$ and note that $N(t)$ takes values in $R$ for each $t$. Now we are in a position to formulate the following theorem regarding the stochastic process of the macro data $N = \{N(t); t = 0, 1, 2, \cdots\}$ and its relation to the process $X$.

THEOREM 2.2.
(i) *The stochastic process $N$ is an irreducible and ergodic Markov chain with invariant initial distribution $\rho = (\rho_n)_{n \in R}$, where*

$$\rho_n = K! \prod_{s \in S} \Pi_s^{n_s}/n_s! \quad \text{for } n = (n_s) \in R. \tag{2.1}$$

(ii) *Let $Q$ be the Markov matrix of the chain $N$. Then the mapping $P \to Q$ is one-to-one.*

PROOF. The proof of the first part of the theorem is similar to the proof in Kemeny and Snell (1960, page 125 and page 130) for Markov chains with finite state space. Here the Markovian character of the stochastic process $N$ follows from the exchangeability of the random variables $(X_1(t), X_1(t+1)), (X_2(t), X_2(t+1)), \cdots, (X_K(t), X_K(t+1))$ for $t = 0, 1, 2, \cdots$.

To prove the second part of the theorem consider $m, n \in R$ for which $m_r = K$ and $n_s = K$ for some $r, s \in S$. The transition $m \to n$ of the chain $N$ occurs if and only if the transition $r \to s$ occurs in each of the $K$ chains $X_k$, $k = 1, 2, \cdots, K$. Hence $q_{mn} = (p_{rs})^K$. □

REMARKS 2.3. The transition probabilities of the Markov matrix $Q = (q_{mn})_{m,n \in R}$ are given by

$$q_{mn} = \sum_{(m_{rs})} \prod_{r \in S} m_r! \prod_{s \in S} \{p_{rs}\}^{m_{rs}}/m_{rs}!, \tag{2.2}$$

where the sum is taken over all $(m_{rs})_{r,s \in S}$, $m_{rs} \in \{0, 1, \cdots, K\}$ such that $\sum_{r \in S} m_{rs} = n_s$ and $\sum_{s \in S} m_{rs} = m_r$. To see this, define the chain $M(t) = (M_{rs}(t))_{r,s \in S}$, where

$$M_{rs}(t) = \#\{k \mid X_k(t-1) = r \wedge X_k(t) = s\} \quad \text{for} \quad t = 1, 2, \cdots,$$

and observe that

$$P\{M(t) = m(t) \mid N(t-1) = m\} = \prod_{r \in S} m_r! \prod_{s \in S} \{p_{rs}\}^{m_{rs}}/m_{rs}!,$$

where $m(t) = (m_{rs})_{r,s \in S}$ for $t = 1, 2, \cdots$.

**3. On the least squares estimator.** In the following, we assume that $\Theta$ is a compact subspace of a $p$-dimensional Euclidean space and that $\theta = (\theta_1, \theta_2, \cdots, \theta_p) \in \Theta$; $\theta^0$ indicates the true parameter value.

Suppose that the Markov matrix $P$ of the model described in Section 2 depends on the parameter $\theta$. Thus $P = P(\theta) = (p_{rs}(\theta))_{r,s \in S}$. We want to make inferences about $\theta$, but the only information with respect to $\theta$ is a sample of so-called macro data $N_T = (N(0), N(1), \cdots, N(T))$ of the chain $N$ with Markov matrix $Q = Q(\theta)$.

Now Theorem 2.2 (ii) shows that it is possible to estimate $\theta$ by observing $N$, provided that the mapping $\theta \to P(\theta)$ is injective. However the formula (2.2) for the transition probabilities of the Markov matrix $Q$ of $N$ as function of the Markov matrix $P$ of $X$ indicates that the likelihood function for the sample $N_T$ of $N$ will be unattractive if $K > 1$.

Therefore we propose the method of least squares to estimate $\theta$. More precisely, an estimator $\tilde{\theta}_T$ based on a sample $N_T$, that minimizes the expression

$$\sum_{t=1}^{T} \| N(t) - \mathscr{E}_\theta\{N(t) \mid \mathscr{F}(t-1)\} \|^2 \tag{3.1}$$

is called a *least squares estimator for* $\theta$. Here $\mathscr{E}_\theta(\cdot \mid \cdot)$ denotes the conditional expectation under $P(\theta)$, $\mathscr{F}(t-1)$ denotes the $\sigma$-algebra generated by $N(0), N(1), \cdots, N(t-1)$ and $\| \cdot \|$ denotes the Euclidean norm associated with the inner product $\langle n, m\rangle = \sum_{s\in S} n_s m_s$.

Finally we use the following notations for $\theta \in \Theta$:

$$P_{r\cdot}(\theta) = (p_{rs}(\theta))_{s\in S} \quad \text{for each} \quad r \in S, \tag{3.2}$$

$$L_T(\theta) = T^{-1} \sum_{t=1}^{T} \| N(t) - N(t-1)P(\theta) \|^2 \quad \text{and} \tag{3.3}$$

$$L(\theta) = \mathscr{E} L_1(\theta). \tag{3.4}$$

Now let us return to the formula (3.1). Since the process $N$ is a Markov chain (Theorem 2.2), we find

$$\mathscr{E}_\theta\{N(t) \mid \mathscr{F}(t-1)\} = \mathscr{E}_\theta\{N(t) \mid N(t-1)\} = N(t-1)P(\theta) \quad \text{a.s.}$$

for $t = 1, 2, \cdots, \theta \in \Theta$. Thus, using the notation (3.3), we may rewrite (3.1) as

$$\sum_{t=1}^{T} \| N(t) - N(t-1)P(\theta) \|^2 = TL_T(\theta).$$

ASSUMPTIONS 3.1.
(i) $P(\theta) \neq P(\theta')$ for all $\theta \neq \theta'$ $(\theta, \theta' \in \Theta)$.
(ii) The function $p_{rs}(\cdot)$ is continuous on $\Theta$ for each $r, s \in S$.
(iii) The series $\sum_{s\in S} \{p_{rs}(\theta)\}^2$ is uniformly convergent in $\theta$ for each $r \in S$.

THEOREM 3.2. *Let $N$ be the chain constructed in Section 2 with $P = P(\theta)$, satisfying Assumptions 3.1. Then*

(3.5) *for each $T$ there exists a stochastic vector $\tilde{\theta}_T$ with values in $\Theta$ such that $L_T(\tilde{\theta}_T) = \inf_{\theta\in\Theta} L_T(\theta)$*

*and such that*

$$\lim_{T\to\infty} \tilde{\theta}_T = \theta^0 \quad \text{a.s.} \tag{3.6}$$

PROOF. To prove the first part of the theorem it is sufficient to show that $L_T(\cdot)$ is continuous on $\Theta$, since $\Theta$ is compact. Then we may choose a measurable $\tilde{\theta}_T$, see e.g. in Jennrich (1969).

The mapping $\theta \to P_{r\cdot}(\theta) \in \ell^2(S)$ is continuous on $\Theta$ by Assumptions 3.1(ii)–(iii). For each realization of the chain $X(t)$, $t = 0, 1, 2, \cdots$

(3.7) there exists a finite subset $S(t) \subset S$, such that $\sum_{s\notin S(t)} N_s(t) = 0$ for each $t$.

Hence $N(t-1)P(\theta)$ is a finite sum and the mapping

(3.8) $\theta \to N(t) - N(t-1)P(\theta) \in \ell^2(S)$ is continuous on $\Theta$ for $t = 1, 2, \cdots$.

Thus, $L_T(\cdot)$ is continuous on $\Theta$.

To prove the second part, we first observe that $L(\theta)$ has a unique minimum at $\theta = \theta^0$, since

$$\begin{aligned} L(\theta) &= L(\theta^0) + \mathscr{E} \| N(0)\{P(\theta) - P(\theta^0)\} \|^2 \\ &\geq L(\theta^0) + \textstyle\sum_{r\in S} \| K\{P_{r\cdot}(\theta) - P_{r\cdot}(\theta^0)\} \|^2 P\{N_r(0) = K\} \\ &> L(\theta^0) \quad \text{for all} \quad \theta \neq \theta^0, \end{aligned}$$

where the equality sign follows using $\mathscr{E}\{N(1) - N(0)P(\theta^0) \mid \mathscr{F}(0)\} = 0$ a.s. and the strict inequality is implied by Assumption 3.1(i). Therefore

$$\begin{aligned} 0 < L(\tilde{\theta}_T) - L(\theta^0) &= L(\tilde{\theta}_T) - L_T(\tilde{\theta}_T) + L_T(\tilde{\theta}_T) - L(\theta^0) \\ &\leq L(\tilde{\theta}_T) - L_T(\tilde{\theta}_T) + L_T(\theta^0) - L(\theta^0) \\ &\leq 2 \sup_{\theta\in\Theta} | L(\theta) - L_T(\theta) | \quad \text{a.s.}, \end{aligned}$$

and since the last expression tends to 0 a.s. if $T \to \infty$ by Lemma 4.1, we conclude that $\lim_{T\to\infty} L(\tilde{\theta}_T) = L(\theta^0)$ a.s. Therefore (3.6) is true, since $L(\cdot)$ is continuous on a compact set $\Theta$ and has unique minimum at $\theta = \theta^0$. □

In order to establish the asymptotic normality of the least squares estimator $\tilde{\theta}_T$ for $\theta^0$ we make the following additional assumptions, which are of a local character.

ASSUMPTIONS 3.3. There exists a convex compact subset $\Theta^0$ with a neighborhood $\Theta' \subset \Theta$ such that
(i) $\theta^0$ is an interior point of $\Theta^0$;
(ii) the function $p_{rs}(\cdot)$ is twice continuously differentiable on $\Theta'$ for each $r, s \in S$;
(iii) the series $\sum_{s\in S} | \partial p_{rs}(\theta)/\partial\theta_i |$ is uniformly convergent for $\theta \in \Theta^0$ for each $r \in S$, $i = 1, 2, \cdots, p$;
(iv) the series $\sum_{s\in S} |\partial^2 p_{rs}(\theta)/\partial\theta_i\partial\theta_j|$ is uniformly convergent for $\theta \in \Theta^0$ for each $r \in S$, $i, j = 1, 2, \cdots, p$;
(v) the $p \times p$ matrix $A = A(\theta^0) = (a_{ij}(\theta^0))_{i,j=1}^p$ defined by

$$a_{ij}(\theta^0) = 2\mathscr{E}\left\langle \frac{\partial N(0)P(\theta^0)}{\partial\theta_i}, \frac{\partial N(0)P(\theta^0)}{\partial\theta_j} \right\rangle \quad \text{for} \quad i, j = 1, 2, \cdots, p$$

is non-singular, hence positive definite.

LEMMA 3.4. *Let $N$ be the chain constructed in Section* 2 *with $P = P(\theta)$, satisfying Assumptions* 3.1 *and* 3.3. *Then for $T \to \infty$ we have*

(3.9) $$\frac{\partial^2 L_T(\theta)}{\partial\theta\partial\theta'} \quad \text{a.s. } \textit{converges uniformly for} \quad \theta \in \Theta^0$$

(3.10) $$\frac{\partial^2 L_T(\theta^0)}{\partial\theta\partial\theta'} \to A \quad \text{a.s.}$$

(3.11) $$T^{1/2}\frac{\partial L_T(\theta^0)}{\partial\theta} \to_{\mathscr{L}} \mathscr{N}(0, \Sigma), \textit{ where}$$

$$\Sigma = \Sigma(\theta^0) = \mathscr{E}\left\{\frac{\partial L_1(\theta^0)}{\partial\theta'} \cdot \frac{\partial L_1(\theta^0)}{\partial\theta}\right\} \textit{ is defined by}$$

$$\frac{\partial L_1(\theta^0)}{\partial\theta_i} = 2\left\langle N(1) - N(0)P(\theta^0), \frac{\partial N(0)P(\theta^0)}{\partial\theta_i} \right\rangle \quad \textit{for} \quad i = 1, 2, \cdots, p.$$

PROOF. For each $r \in S$ and $i = 1, 2, \cdots, p$ the mapping

$$\theta \to \frac{\partial P_{r\cdot}(\theta)}{\partial\theta_i} \in \ell^1(S) \subset \ell^2(S) \quad \text{is continuous on } \Theta^0$$

by Assumption 3.3(ii), (iii). So, using (3.7), we have for $t = 1, 2, \cdots$, that the mapping

(3.12) $$\theta \to \frac{\partial N(t-1)P(\theta)}{\partial \theta_i} \in \ell^2(S) \quad \text{is continuous on } \Theta^0 \text{ for each } \quad i = 1, 2, \cdots, p.$$

Analogously we have that for $t = 1, 2, \cdots$ the mapping

(3.13) $$\theta \to \frac{\partial^2 N(t-1)P(\theta)}{\partial \theta_i \partial \theta_j} \in \ell^2(S) \quad \text{is continuous on } \Theta^0 \text{ for } i, j = 1, 2, \cdots, p$$

by Assumptions 3.3(ii), (iv).

Combining (3.8), (3.12) and (3.13) it is seen that the following functions are continuous on $\Theta^0$:

$$\frac{\partial L_T(\cdot)}{\partial \theta_i} = -2T^{-1} \sum_{t=1}^T F_i(t, \cdot),$$

where

$$F_i(t, \theta) = \langle N(t) - N(t-1)P(\theta), \partial N(t-1)P(\theta)/\partial \theta_i \rangle$$

for $i = 1, 2, \cdots, p$, $\theta \in \Theta^0$, and

$$\frac{\partial^2 L_T(\cdot)}{\partial \theta_i \partial \theta_j} = 2T^{-1} \sum_{t=1}^T \{(A_{ij} - B_{ij})(t, \cdot)\},$$

where

$$A_{ij}(t, \theta) = \left\langle \frac{\partial N(t-1)P(\theta)}{\partial \theta_i}, \frac{\partial N(t-1)P(\theta)}{\partial \theta_j} \right\rangle,$$

$$B_{ij}(t, \theta) = \left\langle N(t) - N(t-1)P(\theta), \frac{\partial N(t-1)P(\theta)}{\partial \theta_i \partial \theta_j} \right\rangle$$

for $i, j = 1, 2, \cdots, p$, $\theta \in \Theta^0$. Furthermore, note that for $t = 1, 2, \cdots$, the random variables $|F_i(t, \theta)|$, $|A_{ij}(t, \theta)|$ and $|B_{ij}(t, \theta)|$ are bounded for $\theta \in \Theta^0$, $i, j = 1, 2, \cdots, p$. So (3.9) follows by applying Lemma 4.1 from the appendix and since $\mathscr{E}_\theta\{B_{ij}(t, \theta) \mid \mathscr{F}(t-1)\} = 0$ a.s. for $\theta \in \Theta^0$, $i, j = 1, 2, \cdots, p$ $(t = 1, 2, \cdots)$ the formula (3.10) is an immediate consequence of the Ergodic theorem (see, e.g., Doob, 1953, page 465). To prove the asymptotic normality of $T^{1/2}(\partial L_T(\theta^0)/\partial \theta)$, we apply Lemma 4.3. Let $T = 1, 2, \cdots$ and $t = 1, 2, \cdots, T$. Define $D(T, t) = (D_1(T, t), D_2(T, t), \cdots, D_p(T, t))'$ by

$$D_i(T, t) = T^{-1/2} F_i(t, \theta^0) \quad \text{for} \quad i = 1, 2, \cdots, p.$$

Note that

$$\mathscr{E}\{D_i(T, t) \mid \mathscr{F}(t-1)\} = T^{-1/2} \mathscr{E}\{F_i(t, \theta^0) \mid \mathscr{F}(t-1)\} = 0 \quad \text{a.s.}$$

Since the random variables $F_i(t, \theta^0)$ are bounded for $i = 1, 2, \cdots, p$ the conditions (4.5) and (4.6) are certainly satisfied. Condition (4.7) follows from the Ergodic theorem, where

$$\begin{aligned}\lim_{T\to\infty} \sum_{t=1}^T D_i(T, t) D_j(T, t) &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^T F_i(t, \theta^0) F_j(t, \theta^0) \\ &= \mathscr{E}\{F_i(1, \theta^0) F_j(1, \theta^0)\} \\ &= \frac{1}{4} \mathscr{E}\left\{\frac{\partial L_1(\theta^0)}{\partial \theta_i} \cdot \frac{\partial L_1(\theta^0)}{\partial \theta_j}\right\} \quad \text{a.s.,}\end{aligned}$$

so (3.11) is true. □

THEOREM 3.5. *Let $N$ be the chain constructed in Section 2 with $P = P(\theta)$, satisfying Assumptions 3.1 and 3.3. Then, for any sequence of least squares estimators $\{\tilde{\theta}_T; T = 1, 2, \cdots\}$ we have*

(3.14) $$T^{1/2}(\tilde{\theta}_T - \theta^0) \to_{\mathscr{L}} \mathcal{N}(0, \Omega) \quad \textit{for} \quad T \to \infty, \textit{ where}$$

(3.15) $$\Omega = \Omega(\theta^0) = A^{-1}\,\Sigma A^{-1} \textit{ is a positive definite } p \times p \textit{ matrix.}$$

PROOF. Since $\tilde{\theta}_T \to \theta^0$ a.s. for $T \to \infty$ (Theorem 3.2) there exists a.s. a stochastic index $T_0$ such that $\tilde{\theta}_T \in \Theta^0$ for $T \geq T_0$. Thus, using the Taylor expansion around the interior point $\theta^0 \in \Theta^0 \subset \Theta'$ we find

$$0 = T^{1/2}\left\{\frac{\partial L_T(\theta^0)}{\partial \theta}\right\}' + \left\{\frac{\partial^2 L_T(\theta_T^*)}{\partial \theta \partial \theta'}\right\} T^{1/2}(\tilde{\theta}_T - \theta^0), \qquad T \geq T_0,$$

where $\theta_T^*$ is a stochastic $p \times 1$ vector taking values in $\Theta^0$ and satisfying $\| \theta_T^* - \theta^0 \| \leq \| \tilde{\theta}_T - \theta^0 \|$ (see, e.g., Jennrich, 1969, Lemma 3). Note that $\theta_T^*$ and $\tilde{\theta}_T$ are tail equivalent. So, using Lemma 3.4 we have that

$$\lim_{T\to\infty} \frac{\partial^2 L_T(\theta_T^*)}{\partial \theta \partial \theta'} = \lim_{T\to\infty} \frac{\partial^2 L_T(\theta^0)}{\partial \theta \partial \theta'} = A \quad \text{a.s.}$$

Now, since $\theta^0$ is an interior point of $\Theta^0$, the theorem follows immediately from Assumption 3.3(v) and Lemma 3.4. □

As a result of Theorem 3.5 we have

THEOREM 3.6. *Let N be the chain constructed in Section* 2 *with* $P = P(\theta)$, *satisfying Assumptions* 3.1 *and* 3.3. *Any sequence of least squares estimators* $\{\tilde{\theta}_T;\ T = 1, 2, \cdots\}$ *satisfies*

(3.16) $$2T\{L_T(\theta^0) - L_T(\tilde{\theta})\} \to_{\mathscr{L}} \textstyle\sum_{i=1}^{p} \lambda_i \chi^2(i) \quad \textit{for} \quad T \to \infty$$

*where* $\lambda_1, \lambda_2, \cdots, \lambda_p$ *are the positive eigenvalues of* $\Omega A$ *and where* $\chi^2(1), \chi^2(2), \cdots, \chi^2(p)$ *are* i.i.d. $\chi_1^2$*-distributed.*

PROOF. A proof may be patterned after Klimko and Nelson (1978). □

We conclude this section with some remarks.

REMARKS 3.7.

(i) Since we are dealing with asymptotic results, the lemmas and theorems in Section 3 still hold if the underlying irreducible and ergodic Markov chains $X_1, X_2, \cdots, X_K$ start with arbitrary initial distributions $a_1, a_2, \cdots, a_K$.

(ii) In Assumptions 3.1 and 3.3 we may replace $\theta$ in 3.1(i), (iii) and 3.3(iii) by $\theta^0$.

**4. Appendix.** In this appendix we give some definitions and two lemmas that are used in the text. The first lemma, concerning a result on a.s. convergence of random functions, is a generalization of Theorem 2 in Jennrich (1969) (Jennrich gives no proof). The second lemma is a result on a central limit theorem for vector martingales: we use a multivariate version of Theorem (2.3) in McLeish (1973) given by Heymans and Magnus (1979).

LEMMA 4.1. *Let* $Y = \{Y(t);\ t = 0, 1, \cdots\}$ *be a stationary and ergodic process with values in a Euclidean space E. Let* $\Theta$ *be a compact subspace of some Euclidean space. Let F be a real valued measurable function on* $E \times \Theta$ *such that* $F(y, \theta)$ *is a continuous function of* $\theta$ *for all* $y \in E$.

*Define* $\phi(y) = \sup_{\theta\in\Theta} |F(y, \theta)|$ *for all y and assume that* $\mathscr{E}\phi(Y(0)) < \infty$, *then*

(4.1) $$\lim_{T\to\infty} T^{-1} \textstyle\sum_{t=1}^{T} F(Y(t), \theta) = \mathscr{E} F(Y(0), \theta)$$

a.s. *uniformly for all* $\theta \in \Theta$.

PROOF. First note that $\sup_{\theta\in\Theta}F(y, \theta)$ and $\inf_{\theta\in\Theta}F(y, \theta)$ are measurable functions of $y$, since $\Theta$ is separable and $F(y, \cdot)$ is continuous. Let $\varepsilon > 0$ and $\theta' \in \Theta$. Define

$$M_k = \{\theta \mid \|\theta - \theta'\| < k^{-1}\} \cap \Theta \qquad \text{for} \quad k = 1, 2, \cdots.$$

Since $F(y, \cdot)$ is continuous for each $y$ and since

$$|\sup_{\theta\in M_k}F(y, \theta) - \inf_{\theta\in M_k}F(y, \theta)| \leq 2 \sup_{\theta\in M_k} |F(y, \theta)| \leq 2\phi(y)$$

we have, using the dominated convergence theorem, that there exists $k_0 = k_0(\varepsilon, \theta')$ such that

$$|\mathscr{E}\{\sup_{\theta\in M_k}F(Y(t), \theta) - \inf_{\theta\in M_k}F(Y(t), \theta)\}| < \varepsilon \tag{4.2}$$

for $k \geq k_0$, $t = 1, 2, \cdots$. Furthermore, from the Ergodic Theorem (see e.g. Doob, 1953, page 465) applied to $\{\sup_{\theta\in M_k}F(Y(t), \theta); t = 0, 1, \cdots\}$, it follows that there exists a.s. a stochastic index $T_0^+ = T_0^+(\varepsilon, \theta')$ such that

$$|T^{-1} \textstyle\sum_{t=1}^T \sup_{\theta\in M_k}F(Y(t), \theta) - \mathscr{E} \sup_{\theta\in M_k}F(Y(0), \theta)| < \varepsilon \tag{4.3}$$

for $T \geq T_0^+$. If we set $U_T(\theta) = T^{-1} \sum_{t=1}^T F(Y(t), \theta) - \mathscr{E} F(Y(0), \theta)$, we have from (4.2) and (4.3) for all $\theta \in M_k$, $k \geq k_0$ that

$$\begin{aligned} U_T(\theta) &\leq T^{-1} \textstyle\sum_{t=1}^T \sup_{\theta\in M_k}F(Y(t), \theta) - \mathscr{E} \inf_{\theta\in M_k}F(Y(0), \theta) \\ &\leq |T^{-1} \textstyle\sum_{t=1}^T \sup_{\theta\in M_k}F(Y(t), \theta) - \mathscr{E} \sup_{\theta\in M_k}F(Y(0), \theta)| \\ &\quad + |\mathscr{E}\{\sup_{\theta\in M_k}F(Y(0), \theta) - \inf_{\theta\in M_k}F(Y(0), \theta)\}| < 2\varepsilon \end{aligned}$$

for $T \geq T_0^+$. Similarly there exists a.s. a stochastic index $T_0^- = T_0^-(\varepsilon, \theta')$ such that for all $\theta \in M_k$, $k \geq k_0$ we have $U_T(\theta) > -2\varepsilon$ for $T \geq T_0^-$. So let $T_0 = \max(T_0^+, T_0^-)$; then for all $\theta \in M_k$, $k \geq k_0$ we have

$$|U_T(\theta)| < 2\varepsilon \quad \text{for} \quad T \geq T_0. \tag{4.4}$$

Let the collection of sets $\{\theta \mid \|\theta - \theta^{(j)}\| < k_j^{-1}\}$ for $\theta^{(j)} \in \Theta$ be an open covering of $\Theta$ such that (2.4) holds on

$$M_{k_j} = \{\theta \mid \|\theta - \theta^{(j)}\| < k_j^{-1}\} \cap \Theta \quad \text{for} \quad T \geq T_j.$$

Since $\Theta$ is compact, there exists a finite number, say $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n)}$ such that $\Theta = \cup_{j=1}^n M_{k_j}$. Thus we see that there exists a.s. a stochastic index $T_{\max} = \max(T_1, T_2, \cdots, T_n)$ such that

$$\sup_{\theta\in\Theta} |U_T(\theta)| \leq \max_{j=1,2,\cdots n} \sup_{\theta\in M_{k_j}} |U_T(\theta)| < 2\varepsilon \quad \text{for} \quad T \geq T_{\max}. \qquad \square$$

DEFINITIONS 4.2. A sequence of random vectors $\{D(t); t = 0, 1, \cdots\}$ with $\sup_t \mathscr{E}\|D(t)\| < \infty$ is called a *sequence of vector martingale differences* if $\mathscr{E}D(0) = 0$ and $\mathscr{E}\{D(t) \mid \mathscr{F}(t-1)\} = 0$ a.s. for $t = 1, 2, \cdots$, where $\mathscr{F}(t-1) = \sigma\{D(0), D(1), \cdots, D(t-1)\}$ is the $\sigma$-algebra generated by $D(0), D(1), \cdots, D(t-1)$. Let $S(T) = \sum_{t=0}^T D(t)$, then $\{S(T); T = 0, 1, \cdots\}$ is a *vector martingale*.

LEMMA 4.3. *Let $\{D(T, t); t = 0, 1, \cdots, T\}$ be a sequence of vector martingale differences for $T = 0, 1, \cdots$, satisfying*

$$\mathscr{E} \max_{t=1,2,\cdots,T} \|D(T, t)\|^2 \textit{ is uniformly bounded,} \tag{4.5}$$

$$\max_{t=1,2,\cdots,T} \|D(T, t)\| \to_{\mathscr{P}} 0 \textit{ for } T \to \infty, \tag{4.6}$$

$$\textstyle\sum_{t=1}^T D(T, t)D'(T, t) \to_{\mathscr{P}} \Sigma \textit{ for } T \to \infty, \tag{4.7}$$

*where $\Sigma$ is a positive semidefinite matrix. Then*

$$S(T) = \sum_{t=1}^{T} D(T, t) \to_{\mathscr{L}} \mathscr{N}(0, \Sigma) \quad \text{for} \quad T \to \infty. \tag{4.8}$$

PROOF. See McLeish (1973), Heymans and Magnus (1979). □

**Acknowledgment.** I am grateful to A. A. Balkema for critical comments and suggestions.

## REFERENCES

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications,* vol. I. Wiley, New York.

HEYMANS, R. D. H. and MAGNUS, J. R. (1979). Asymptotic properties of maximum likelihood estimators in a non-linear model with unknown parameters in the disturbance covariance matrix. Research paper presented at the European Meeting of the Econometric Society.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

KEMENY, J. G. and SNELL, J. L. (1960). *Finite Markov Chains.* Van Nostrand, Princeton.

KLIMKO, L. A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Math. Statist.* **6** 629–642.

LEE, T. C., JUDGE, G. G. and ZELLNER, A. (1970). *Estimating the parameters of the Markov probability model from aggregate time series data.* North-Holland, Amsterdam.

MADANSKY, A. (1959). Least squares estimation in finite Markov processes. *Psychometrika* **24** 137–144.

MCLEISH, D. L. (1973). Dependent central limit theorems and invariance principles. *Ann. Probability* **2** 620–628.

MILLER, G. A. (1952). Finite Markov processes in psychology. *Psychometrika* **17** 149–167.

INSTITUUT VOOR ACTUARIAAT EN ECONOMETRIE
UNIVERSITEIT VAN AMSTERDAM
JODENBREESTRAAT 23
1011 NH AMSTERDAM, THE NETHERLANDS