

# Sample splitting with Markov chains

ANTON SCHICK

*Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA. E-mail: anton@math.binghamton.edu*

Sample splitting techniques play an important role in constructing estimates with prescribed influence functions in semi-parametric and nonparametric models when the observations are independent and identically distributed. This paper shows how a contiguity result can be used to modify these techniques for the case when the observations come from a stationary and ergodic Markov chain. As a consequence, sufficient conditions for the construction of efficient estimates in semi-parametric Markov chain models are obtained. The applicability of the resulting theory is demonstrated by constructing an estimate of the innovation variance in a nonparametric autoregression model, by constructing a weighted least-squares estimate with estimated weights in an autoregressive model with martingale innovations, and by constructing an efficient and adaptive estimate of the autoregression parameter in a heteroscedastic autoregressive model with symmetric errors.

*Keywords:* contiguity; efficient estimation; ergodicity; heteroscedastic autoregressive model; nonparametric autoregressive model; semi-parametric models; stationary Markov chains;  $V$ -uniform ergodicity; weighted least-squares estimation.

## 1. Introduction

Let  $X_0, X_1, X_2, \dots$  be a stationary Markov chain taking values in the state space  $S$  with countably generated  $\sigma$ -field  $\mathcal{S}$  defined on a probability space  $(\Omega, \mathcal{U}, P)$ . We denote the stationary distribution by  $F$  and the transition probability measure by  $M$ . Let  $h$  be a measurable function from  $S^2$  into  $\mathbb{R}^m$  such that

$$E(\|h(X_0, X_1)\|^2) = \iint F(dx)M(x, dy)\|h(x, y)\|^2 < \infty.$$

Our goal is to construct  $m$ -dimensional random vectors  $Z_n = z_n(X_0, \dots, X_n)$  such that

$$n^{1/2} \left( Z_n - \frac{1}{n} \sum_{j=1}^n h(X_{j-1}, X_j) \right) \xrightarrow{P} 0. \quad (1.1)$$

This problem is of particular importance when the pair  $(F, M)$  is unknown and  $h$  depends on this parameter. It arises naturally in the construction of efficient estimates in semi-parametric Markov chain models. See Wefelmeyer (1999) for an overview.

One is tempted to use  $Z_n$  of the form

$$Z_n = \frac{1}{n} \sum_{j=1}^n H_n(X_{j-1}, X_j; X_0, \dots, X_n),$$

where  $H_n$  is a measurable function from  $S^2 \times S^{n+1}$  into  $\mathbb{R}^m$ . However, there are technical difficulties with this approach even in the case of independent and identically distributed observations since the estimator and the arguments used in averaging are not independent. To overcome this difficulty sample splitting techniques were introduced. Drawing on earlier work of Hájek (1962) and van Eeden (1970), Bickel (1982) proposed to use a small initial part of the sample to estimate the unknown influence function and the remaining observations to evaluate this estimate, and then exploited the independence of the evaluations and the estimate. As only a small part of the sample is used to construct the influence function, his construction is more of theoretical interest than of practical value. His construction might lack the desired accuracy in moderate sample sizes. Schick (1986) improved this construction by using two estimates of the influence function each based on half of the sample, and pairing the estimate based on the second half with arguments from the first half and the estimate based on the first half with arguments from the second half. His proof also exploits the independence of the two parts of the sample. Klaassen (1987) treated the general case of asymptotically linear (not necessarily efficient) estimation and showed that Schick's (1986) sufficient conditions are also necessary. Under additional assumptions, Schick (1987) showed how sample splitting can be avoided altogether. The constructions in Bickel (1982) and Schick (1986; 1987) rely heavily on LeCam's (1960) discretization technique, while Klaassen avoids discretization by further sample splitting.

Using a martingale argument, Bickel's (1982) construction can be generalized to Markov chains. This was done by Maercker (1997) for a special case. She used this approach to construct adaptive estimates of the autoregression parameter in a (conditionally) heteroscedastic autoregressive model of order 1. Unfortunately, Schick's (1986) construction does not have such an obvious generalization. The martingale approach does not work here when the estimate of  $h$  is based on the second part of the sample. His proof in essence relies much more on the independence of the two parts of the sample. We should point out that his construction, however, has been successfully implemented in some nonlinear time series models by Drost *et al.* (1997) and Koul and Schick (1997). These authors exploit the special structure of those models. More precisely, they reconstruct the innovations using discrete auxiliary estimates. This allows them to treat the estimated innovations as independent and identically distributed observations in the proofs by utilizing contiguity arguments and the discrete structure of the auxiliary estimates. Thus their arguments essentially reduce to the situation considered by Schick (1986). Koul and Schick (1997) give conditions under which sample splitting can be avoided in these models. But they demonstrate their usefulness only in the case of ARMA models. It had already been shown by Kreiss (1987a; 1987b) that for such models sample splitting can be avoided. The sample splitting ideas used by Drost *et al.* (1997) and Koul and Schick (1997) work only for the special class of nonlinear time series models considered by these authors. They do not carry over to nonlinear autoregression models with semi-parametric components or to autoregression models with heteroscedasticity of unknown form as considered by Maercker (1997) and revisited in Section 8 below. It was this lack of a theory for dealing with such models that prompted the present research.

The main result of this paper is the extension of the sample splitting technique from independent observations to Markov chain models. We shall not restrict ourselves to the

case of efficient estimation as in Schick (1986) but shall treat more generally the case of asymptotically linear estimation as in Klaassen (1987).

In the Markov chain setting, the two halves of the sample are no longer independent. However, under an appropriate ergodicity assumption, we shall show that  $(X_0, \dots, X_m)$  and  $(X_{n-m}, \dots, X_n)$  are essentially independent if  $m/n$  is close to  $1/2$  and  $n - 2m$  is large. This will allow us to carry out a modified sample splitting technique in the spirit of Schick (1986) for the present case by ignoring the middle observations  $X_{m+1}, \dots, X_{n-m-1}$  and letting  $X_0, \dots, X_m$  now form the first part of the sample and  $X_{n-m}, \dots, X_n$  the second part. The trick of separating blocks of data by a small but increasing number of observations is well known in the Markov chain literature, or more generally in the context of mixing sequences.

The asymptotic independence result is established in Section 2. There we show that, for an appropriately chosen sequence  $\langle m_n \rangle$ , the sequences of distributions

$$\langle \mathcal{L}(X_0, \dots, X_{m_n}, X_{n-m_n}, \dots, X_n | P) \rangle \quad \text{and} \quad \langle \mathcal{L}(X_0, \dots, X_{m_n} | P) \times \mathcal{L}(X_{n-m_n}, \dots, X_n | P) \rangle$$

are mutually contiguous. This result is used in Section 3 to implement sample splitting for Markov chains. The applicability of the results of Section 3 is then demonstrated by means of two examples. Estimation of the innovation variance in a nonparametric autoregression model is discussed in Section 4. Section 5 addresses weighted least-squares estimation with random weights in an autoregressive model with martingale difference innovations. In Section 6 splitting into more than two parts is discussed. Section 7 provides specifics for efficient estimation in semi-parametric models and gives sufficient conditions for constructing efficient estimates in such models. These conditions are generalizations of those given by Schick (1986) and, in view of the results of Klaassen (1987), should be necessary. The results of Section 7 are applied in Section 8 to construct an efficient and adaptive estimate of the parameter of interest in a heteroscedastic autoregressive model of order 1 with symmetric errors. We shall do so under assumptions weaker than those of Maercker (1997).

Our sample splitting technique has already been successfully implemented in several papers. Schick and Wefelmeyer (1999) used it to construct improvements over empirical estimates of linear functionals of the marginal law  $F$  under semi-parametric restrictions on the transition  $Q$ . Kessler *et al.* (1999) applied this technique to construct efficient estimates of the finite-dimensional parameter in a Markov chain model with parametrically specified marginal laws, but an otherwise unspecified transition law. Schick (1999) relied on it to construct efficient estimates in a semi-parametric additive autoregression model. The author is currently working on applying it to other semi-parametric autoregression models.

## 2. A contiguity result

In this section the aforementioned contiguity result will be presented. We assume from now on that the underlying probability space  $(\Omega, \mathcal{U}, P)$  is large enough that it also carries  $S$ -valued random elements  $\tilde{X}_0, \tilde{X}_1, \dots$  which form an independent copy of the chain  $\{X_j : j = 0, 1, 2, \dots\}$ . For a positive integer  $n$ , we let  $M^n$  denote the  $n$ -step transition probability measure. These Markov kernels are defined iteratively by

$$M^n(x, B) = \int M(x, dy)M^{n-1}(y, B), \quad x \in S, B \in \mathcal{F},$$

starting with  $M^1 = M$ .

We shall establish contiguity under the following condition:

$$(B0) \quad \lim_{n \rightarrow \infty} \int F(dx) \sup_{B \in \mathcal{F}} |M^n(x, B) - F(B)| = 0.$$

This condition is slightly weaker than

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{F}} |M^n(x, B) - F(B)| = 0 \quad \text{for all } x \in S.$$

The latter condition is known to hold for aperiodic positive recurrent Harris chains; see Theorem 13.0.1 in Meyn and Tweedie (1993). Other sufficient conditions for (B0) are discussed in the remark after the following lemma. The measurability of  $x \mapsto \sup_{B \in \mathcal{F}} |M^n(x, B) - F(B)|$  was proved in Roberts and Rosenthal (1997).

**Lemma 2.1.** *Suppose (B0) holds. Let  $p_n$ ,  $q_n$  and  $r_n$  be positive integers such that  $p_n < q_n < r_n$  and  $q_n - p_n \rightarrow \infty$ . Then the sequences  $\langle Q_n \rangle$  and  $\langle \tilde{Q}_n \rangle$  of distributions defined by*

$$Q_n = \mathcal{L}(X_0, \dots, X_{p_n}, X_{q_n}, \dots, X_{r_n} | P) \quad \text{and} \quad \tilde{Q}_n = \mathcal{L}(X_0, \dots, X_{p_n}, \tilde{X}_{q_n}, \dots, \tilde{X}_{r_n} | P)$$

*are mutually contiguous.*

**Proof.** Let  $f$  be a measurable function from  $S^{2+p_n+s_n}$  into  $[-1, 1]$ , where  $s_n = r_n - q_n$ . Then there exists a function  $\bar{f}$  from  $S^2$  to  $[-1, 1]$  such that  $\bar{f}(X_{p_n}, X_{q_n})$  is a version of the conditional expectation of  $f(X_0, \dots, X_{p_n}, X_{q_n}, \dots, X_{r_n})$  given  $(X_{p_n}, X_{q_n})$  and  $\bar{f}(X_{p_n}, \tilde{X}_{q_n})$  is a version of the conditional expectation of  $f(X_0, \dots, X_{p_n}, \tilde{X}_{q_n}, \dots, \tilde{X}_{r_n})$  given  $(X_{p_n}, \tilde{X}_{q_n})$ . This shows that

$$\int f d\tilde{Q}_n = E\bar{f}(X_{p_n}, \tilde{X}_{q_n}) = \iint F(dx)F(dy)\bar{f}(x, y)$$

and

$$\int f dQ_n = E\bar{f}(X_{p_n}, X_{q_n}) = \iint F(dx)M^{q_n-p_n}(x, dy)\bar{f}(x, y).$$

Using this we find that

$$\begin{aligned} \left| \int f d\tilde{Q}_n - \int f dQ_n \right| &\leq \int F(dx) \sup_{|g| \leq 1} \left| \int M^{q_n-p_n}(x, dy)g(y) - \int F(dy)g(y) \right| \\ &= 2 \int F(dx) \sup_{B \in \mathcal{F}} |M^{q_n-p_n}(x, B) - F(B)|. \end{aligned}$$

This bound and (B0) show that the variational distance between  $Q_n$  and  $\tilde{Q}_n$  converges to 0. This implies the desired contiguity.  $\square$

**Remark.** Let  $V$  be a measurable function from  $S$  into  $[1, \infty)$ . Then (B0) is implied by

$$(B1) \quad \int F(dx) \sup_{|h| \leq V} \left| \int M^n(x, dy) h(y) - \int F(dy) h(y) \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Of course, (B1) is implied by  $V$ -uniform ergodicity:

$$(B2) \quad \sup_{x \in S} \sup_{|h| \leq V} \frac{\left| \int M^n(x, dy) h(y) - \int F(dy) h(y) \right|}{V(x)} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

An even stronger condition is geometric  $V$ -uniform ergodicity:

(B3) For some  $D < \infty$  and for some  $\alpha < 1$ ,

$$\sup_{x \in S} \sup_{|h| \leq V} \frac{\left| \int M^n(x, dy) h(y) - \int F(dy) h(y) \right|}{V(x)} \leq D\alpha^n \quad \text{for all } n.$$

For irreducible and aperiodic Markov chains, Meyn and Tweedie (1994) have shown that (B2) and (B3) are equivalent and that either condition is equivalent to the following condition.

(DC) For some small set  $C$  there exist a measurable function  $V_C \geq 1$  and positive constants  $\lambda_C < 1$ ,  $b_C < \infty$  such that the drift condition

$$\int M(x, dy) V_C(y) \leq \lambda_C V_C(x) + b_C 1_C(x), \quad x \in S,$$

holds, and  $V_C$  is equivalent to  $V$  in the sense that  $V/c \leq V_C \leq cV$  for the same constant  $C \geq 1$ .

An event  $C$  in  $\mathcal{S}$  is called a small set if there exist some probability measure  $\nu$ , some constant  $\delta > 0$  and some positive integer  $j_0$  such that

$$M^{j_0}(x, B) \geq \delta \nu(B), \quad x \in C, B \in \mathcal{S}. \quad (2.1)$$

For the definitions of the other concepts we refer the reader to Meyn and Tweedie (1993).

Under slightly stronger assumptions than (DC), Meyn and Tweedie (1994) have shown that the constants  $D$  and  $\alpha$  appearing in (B3) are computable. More precisely, from their Theorem 2.3 one obtains the following.

**Theorem 2.2.** *Let  $\mathcal{Q}$  be a family of transition probabilities each generating an irreducible and aperiodic Markov chain. Let  $V$  be a measurable function from  $S$  to  $[1, \infty)$ . Suppose there are constants  $\lambda < 1$ ,  $b < \infty$  and  $\delta > 0$ , an event  $C$  in  $\mathcal{S}$  and a probability measure  $\nu$  on  $\mathcal{S}$  with  $\nu(C) = 1$  such that*

$$\sup_{Q \in \mathcal{Q}} \int Q(x, dy) V(y) \leq \lambda V(x) + b 1_C(x), \quad x \in S, \quad (2.2)$$

$$\inf_{Q \in \mathcal{Q}} Q(x, B) \geq \delta v(B), \quad x \in C, B \in \mathcal{S} \quad (2.3)$$

and

$$\sup_{x \in C} V(x) < \infty \quad (2.4)$$

Then there are constants  $D < \infty$  and  $\alpha < 1$  and probability measures  $\{\pi_Q : Q \in \mathcal{Q}\}$  such that

$$\pi_Q(B) = \int \pi_Q(dx) Q(x, B), \quad B \in \mathcal{S}, \quad (2.5)$$

for each  $Q \in \mathcal{Q}$ , and

$$\sup_{Q \in \mathcal{Q}} \sup_{x \in S} \sup_{|h| \leq V} \frac{\left| \int Q^n(x, dy) h(y) - \int \pi_Q(dy) h(y) \right|}{V(x)} \leq D \alpha^n \quad \text{for all } n. \quad (2.6)$$

**Remark.** A signed Markov kernel is a map from  $S \times \mathcal{S}$  into  $\mathbb{R}$  such that  $x \mapsto K(x, B)$  is measurable for each  $B \in \mathcal{S}$  and  $B \mapsto K(x, B)$  is a signed measure for each  $x \in S$ . Let  $v$  be a signed measure,  $K$  and  $L$  be signed Markov kernels and  $V$  be a positive measurable function on  $S$ . Then we set

$$\|v\|_V = \sup_{|h| \leq V} \left| \int v(dy) h(y) \right| \quad \text{and} \quad \|K\|_V = \sup_{|h| \leq V} \left| \int K(\cdot, dy) h(y) \right|;$$

we let  $vK$  denote the signed measure defined by

$$vK(B) = \int v(dx) K(x, B), \quad B \in \mathcal{S};$$

and we let  $KL$  denote the signed Markov kernel defined by

$$KL(x, B) = \int K(x, dy) L(y, B), \quad x \in S, B \in \mathcal{S}.$$

Now let  $\mathcal{Q}$ ,  $V$  and  $\{\pi_Q : Q \in \mathcal{Q}\}$  be as in Theorem 2.2 and define Markov kernels  $\{\Pi_Q : Q \in \mathcal{Q}\}$  by

$$\Pi_Q(x, B) = \pi_Q(B), \quad x \in S, B \in \mathcal{S}.$$

Then we can write (2.5) as  $\pi_Q = \pi_Q Q$  and (2.6) as

$$\sup_{Q \in \mathcal{Q}} \|\mathcal{Q}^n - \Pi_Q\|_V \leq DV \alpha^n, \quad \text{for all } n. \quad (2.7)$$

Also, by (2.2),  $\int V d\pi_Q = \int QV d\pi_Q \leq \lambda \int V d\pi_Q + b$ , so that

$$\sup_{Q \in \mathcal{Q}} \int V d\pi_Q \leq \frac{b}{1-\lambda}. \quad (2.8)$$

Now let  $Q$  and  $R$  be elements in  $\mathcal{Q}$ . Then we have  $\pi_Q - \pi_R = \pi_Q(Q - R) + (\pi_Q - \pi_R)R$ . Iterating this, we arrive at

$$\pi_Q - \pi_R = \pi_Q(Q - R) + \sum_{j=1}^n \pi_Q(Q - R)R^j + (\pi_Q - \pi_R)R^{n+1}, \quad n = 1, 2, \dots$$

Since  $(\nu - \mu)\Pi_R = 0$  for two probability measures  $\nu$  and  $\mu$ , this can also be written as

$$\pi_Q - \pi_R = \pi_Q(Q - R) + \sum_{j=1}^n \pi_Q(Q - R)(R^j - \Pi_R) + (\pi_Q - \pi_R)(R^{n+1} - \Pi_R), \quad n = 1, 2, \dots \quad (2.9)$$

Thus, in view of (2.7), we have

$$\|\pi_Q - \pi_R\|_V \leq \frac{1+D}{1-\alpha} \|\pi_Q(Q - R)\|_V \leq \frac{1+D}{1-\alpha} \int \|Q - R\|_V d\pi_Q. \quad (2.10)$$

This result can be used to derive continuity properties of the map  $Q \mapsto \pi_Q$ . It will be exploited in Section 8 to establish continuity of the second moments of the stationary distribution for heteroscedastic autoregression models of order 1.

### 3. The 2-split

In this section we implement the sample splitting technique. Our first result gives the basic construction. It shows how to get a sequence satisfying (1.1) based on the actual observations from one that also uses an independent copy thereof. Subsequent results then deal with the problem of finding solutions to (1.1) that are allowed to use an independent copy of the observations.

Throughout this section we let

$$\mathbf{X}_n = (X_0, \dots, X_n) \quad \text{and} \quad \tilde{\mathbf{X}}_n = (\tilde{X}_0, \dots, \tilde{X}_n),$$

we let  $\langle m_n \rangle$  denote a sequence of positive integers such that

$$n - 2m_n \rightarrow \infty \quad \text{and} \quad \frac{n - 2m_n}{\sqrt{n}} \rightarrow 0, \quad (3.1)$$

and put

$$\mathbf{X}_{n,1} = (X_0, \dots, X_{m_n}) \quad \text{and} \quad \mathbf{X}_{n,2} = (X_{n-m_n}, \dots, X_n).$$

We are now ready to state the basic construction lemma.

**Lemma 3.1.** *Suppose that (B0) holds. Set*

$$Z_n = \frac{1}{2}z_{m_n}(\mathbf{X}_{n,1}, \mathbf{X}_{n,2}) + \frac{1}{2}z_{m_n}(\mathbf{X}_{n,2}, \mathbf{X}_{n,1}), \quad (3.2)$$

where  $z_N$  is a measurable function from  $S^{2N+2}$  to  $\mathbb{R}^m$  for each positive integer  $N$ . Suppose that

$$n^{1/2} \left( z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) - \frac{1}{n} \sum_{j=1}^n h(X_{j-1}, X_j) \right) \xrightarrow{P} 0. \quad (3.3)$$

Then  $\langle Z_n \rangle$  satisfies (1.1).

**Proof.** Let  $\tilde{\mathbf{X}}_{n,2} = (\tilde{X}_{n-m_n}, \dots, \tilde{X}_n)$ . It follows from (3.1) and (3.3) that

$$n^{1/2} \left( z_{m_n}(\mathbf{X}_{n,1}, \tilde{\mathbf{X}}_{n,2}) - \frac{1}{m_n} \sum_{j=1}^{m_n} h(X_{j-1}, X_j) \right) \xrightarrow{P} 0.$$

In view of the mutual contiguity of  $\langle \mathcal{L}(\mathbf{X}_{n,1}, \tilde{\mathbf{X}}_{n,2}|P) \rangle$  and  $\langle \mathcal{L}(\mathbf{X}_{n,1}, \mathbf{X}_{n,2}|P) \rangle$  this yields

$$n^{1/2} \left( z_{m_n}(\mathbf{X}_{n,1}, \mathbf{X}_{n,2}) - \frac{1}{m_n} \sum_{j=1}^{m_n} h(X_{j-1}, X_j) \right) \xrightarrow{P} 0.$$

Similarly, one can verify that

$$n^{1/2} \left( z_{m_n}(\mathbf{X}_{n,2}, \mathbf{X}_{n,1}) - \frac{1}{m_n} \sum_{j=n-m_n+1}^n h(X_{j-1}, X_j) \right) \xrightarrow{P} 0.$$

Since  $n^{3/2}(1/(2m_n) - 1/n) \rightarrow 0$  by (3.1), we are left to show that

$$R_n = \frac{1}{\sqrt{n}} \sum_{j=m_n+1}^{n-m_n} h(X_{j-1}, X_j) \xrightarrow{P} 0.$$

But this follows as

$$E(R_n^2) \leq \frac{(n-2m_n)^2}{n} E(h^2(X_0, X_1)) \rightarrow 0$$

by the properties of  $\langle m_n \rangle$ . □

**Remark.** The above construction does not utilize the middle  $n - 2m_n$  observations. Thus we should let  $n - 2m_n$  tend to infinity very slowly; we recommend  $n - 2m_n \sim \log n$  for the geometrically ergodic case. Let us now point out a construction that utilizes all observations. To describe it, let  $k_n$  denote the integer part of  $n/2$ ,  $\langle d_n \rangle$  be a sequence of positive integers such that  $d_n \rightarrow \infty$  and  $d_n n^{-1} \rightarrow 0$  (we recommend  $d_n \sim \log n$ ), and  $z_{k,l}$  be a measurable function from  $S^{k+l+2}$  to  $\mathbb{R}^m$  for non-negative integers  $k$  and  $l$ . The construction we have in mind is

$$\begin{aligned} Z_n &= \frac{1}{2} z_{k_n, n-k_n-d_n}(X_0, \dots, X_{k_n}, X_{k_n+d_n}, \dots, X_n) \\ &\quad + \frac{1}{2} z_{n-k_n, k_n-d_n}(X_{k_n}, \dots, X_n, X_0, \dots, X_{k_n-d_n}). \end{aligned}$$

This estimator does not waste the middle observations. It satisfies (1.1) if



$$n^{1/2} \left( z_{n,l_n}(X_0, \dots, X_n, \tilde{X}_0, \dots, \tilde{X}_{l_n}) - \frac{1}{n} \sum_{j=1}^n h(X_{j-1}, X_j) \right) \xrightarrow{P} 0$$

whenever  $l_n/n \rightarrow 1$ . We have chosen to work with the construction given in the above lemma as it simplifies notation. Similar remarks apply to the constructions given in the later parts of this paper.

Let us now give sufficient conditions for (3.3). We begin with the case where

$$z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = \frac{1}{n} \sum_{j=1}^n h_n(X_{j-1}, X_j; \tilde{\mathbf{X}}_n) \quad (3.4)$$

for some measurable function  $h_n$  from  $S^2 \times S^{n+1}$  to  $\mathbb{R}^m$ .

Conditions (3.5) and (3.6) appearing in the next theorem are reminiscent of (2.2) and (2.3) in Schick (1986). The difference is that we now work with conditional expectations.

**Theorem 3.2.** *Let (B0) hold. Suppose that*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \int M(X_{j-1}, dy) (h_n(X_{j-1}, y; \tilde{\mathbf{X}}_n) - h(X_{j-1}, y)) \xrightarrow{P} 0 \quad (3.5)$$

and

$$\frac{1}{n} \sum_{j=1}^n \int M(X_{j-1}, dy) \|h_n(X_{j-1}, y; \tilde{\mathbf{X}}_n) - h(X_{j-1}, y)\|^2 \xrightarrow{P} 0. \quad (3.6)$$

Then (3.3) holds with  $z_n$  defined in (3.4), i.e.

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (h_n(X_{j-1}, X_j; \tilde{\mathbf{X}}_n) - h(X_{j-1}, X_j)) \xrightarrow{P} 0. \quad (3.7)$$

Consequently, the estimate  $\langle Z_n \rangle$  defined by

$$Z_n = \frac{1}{2m_n} \sum_{j=1}^{m_n} h_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,2}) + \frac{1}{2m_n} \sum_{j=n-m_n+1}^n h_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,1}).$$

satisfies (1.1).

**Proof.** Let  $H_{n,j}$  denote the summand in (3.7) so that

$$H_{n,j} = h_n(\mathbf{X}_{j-1}, X_j; \tilde{\mathbf{X}}_n) - h(X_{j-1}, X_j)$$

and  $\bar{H}_{n,j}$  denote the summand in (3.5) so that

$$\bar{H}_{n,j} = \int M(X_{j-1}, dy) (h_n(X_{j-1}, y; \tilde{\mathbf{X}}_n) - h(X_{j-1}, y)).$$

In view of (3.5), it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (H_{n,j} - \bar{H}_{n,j}) \xrightarrow{P} 0. \quad (3.8)$$

Let  $\mathcal{A}_{n,j}$  denote the  $\sigma$ -field generated by  $X_0, \dots, X_j, \tilde{X}_0, \dots, \tilde{X}_n$ , for  $j = 0, 1, \dots, n$ . Set  $Z_{n,j} = n^{-1/2}(H_{n,j} - \bar{H}_{n,j})$ ,  $j = 1, \dots, n$ . Note that  $\bar{H}_{n,j} = E(H_{n,j} | \mathcal{A}_{n,j-1})$  almost surely. Thus, for each unit vector  $a \in \mathbb{R}^m$ ,  $a^\top Z_{n,1}, \dots, a^\top Z_{n,n}$  is a martingale difference array for the filtration  $\{\mathcal{A}_{n,j} : j = 0, \dots, n\}$  with

$$\sum_{j=1}^n E((a^\top Z_{n,j})^2 | \mathcal{A}_{n,j-1}) \leq \frac{1}{n} \sum_{j=1}^n E((a^\top H_{n,j})^2 | \mathcal{A}_{n,j-1}) \xrightarrow{P} 0$$

in view of (3.6). This shows that  $\sum_{j=1}^n a^\top Z_{n,j} \xrightarrow{P} 0$  for every unit vector  $a$  and proves (3.8).  $\square$

**Remark.** A sufficient condition for (3.6) is

$$\iint F(dx) M(x, dy) \|h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y)\|^2 \xrightarrow{P} 0. \quad (3.9)$$

A sufficient condition for (3.5) is

$$\int M(x, dy) (h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y)) = 0, \quad x \in S. \quad (3.10)$$

Another sufficient condition for (3.5) is given in the next lemma.

**Lemma 3.3.** *Suppose that  $M$  is geometrically  $V$ -uniformly ergodic for some  $F$ -square-integrable function  $V$  from  $S$  to  $[1, \infty)$ , that*

$$\sup_{x \in S} \left\| \frac{\int M(x, dy) (h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y))}{V(x)} \right\| = O_p(1), \quad (3.11)$$

and that

$$\int F(dx) \left\| \int M(x, dy) (h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y)) \right\|^2 = o_p(1). \quad (3.12)$$

Then (3.5) is equivalent to

$$n^{1/2} \iint F(dx) M(x, dy) (h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y)) = o_p(1). \quad (3.13)$$

**Proof.** For simplicity, assume that  $m = 1$ . Let

$$\delta_n(x) = \int M(x, dy) (h_n(x, y; \tilde{\mathbf{X}}_n) - h(x, y)), \quad x \in S.$$

Let  $L_V$  be the set of all measurable functions  $l$  from  $S$  to  $\mathbb{R}$  such that

$$|l|_V := \sup_{x \in S} \frac{|l(x)|}{V(x)} < \infty.$$

It is easy to check that  $(L_V, |\cdot|_V)$  is a Banach space. Let  $U$  denote the bounded linear operator on  $L_V$  which assigns to  $l \in L_V$  the element  $Ul$  in  $L_V$  defined by

$$Ul(x) = \sum_{j=0}^{\infty} \left( \int M^j(x, dy) l(y) - \int F(dy) l(y) \right), \quad x \in S.$$

That  $U$  is well defined and bounded follows from the geometric  $V$ -uniform ergodicity of  $M$ . Indeed, since there are constants  $D < \infty$  and  $\alpha < 1$  such that

$$\left| \int M^j(x, dy) l(y) - \int F(dy) l(y) \right| \leq |l|_V V(x) D \alpha^j, \quad x \in S, j = 0, 1, \dots,$$

we have

$$|Ul|_V \leq \frac{D}{1-\alpha} |l|_V, \quad l \in L_V.$$

Since  $V$  is  $F$ -square-integrable and

$$\int F(dx) \left| \int M^j(x, dy) l(y) - \int F(dy) l(y) \right|^2 \leq \int F(dx) l^2(x), \quad j = 0, 1, \dots,$$

it is now easy to check that  $\int (Ul_n)^2 dF \rightarrow 0$  whenever  $|l_n|_V = O(1)$  and  $\int l_n^2 dF \rightarrow 0$ . This shows that

$$\int F(dy) (U \delta_n(y))^2 = o_p(1). \quad (3.14)$$

For  $l \in L_V$ , let  $Al$  denote the map on  $S^2$  defined by

$$Al(x, y) = Ul(y) - \int M(x, dz) Ul(z), \quad x, y \in S.$$

The martingale representation of Gordin (1969) yields

$$\frac{1}{n} \sum_{j=1}^n \delta_n(X_{j-1}) = \int F(dx) \delta_n(x) + \frac{1}{n} \sum_{j=1}^n A \delta_n(X_{j-1}, X_j) - \frac{1}{n} (U \delta_n(X_n) - U \delta_n(X_0))$$

on the event  $\{|\delta_n|_V < \infty\}$  whose probability tends to 1 by (3.11). Since  $\int M(x, dy) A \delta_n(x, y) = 0$  and  $\iint F(dx) M(x, dy) (A \delta_n(x, y))^2 \leq \int F(dy) (U \delta_n(y))^2 = o_p(1)$ , we obtain from Theorem 3.2 that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n A \delta_n(X_{j-1}, X_j) = o_p(1).$$

This and (3.14) yield

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \delta_n(X_{j-1}) = \sqrt{n} \int F(dx) \delta_n(x) + o_p(1),$$

which is the desired result.  $\square$

**Remark.** Note that (3.9) implies (3.12). Thus if  $M$  is geometrically  $V$ -uniformly ergodic for some  $F$ -square-integrable  $V$ , then (3.9), (3.11) and (3.13) imply (3.6) and (3.5). Note also that we can replace  $\tilde{\mathbf{X}}_n$  by  $\mathbf{X}_n$  in (3.9) and (3.11)–(3.13). In the next theorem we replace (3.11) by a slightly stronger condition. This allows us to weaken (3.9).

**Theorem 3.4.** *Suppose that  $M$  is geometrically  $V$ -uniformly ergodic for some  $F$ -square-integrable function  $V$  from  $S$  to  $[1, \infty)$  and that*

$$\sup_{x, y \in S} \frac{\|h_n(x, y; \mathbf{X}_n) - h(x, y)\|}{\sum_{i=1}^r V^{\alpha_i}(x) V^{1-\alpha_i}(y)} = O_p(1) \quad (3.15)$$

for numbers  $\alpha_1, \dots, \alpha_r$  in  $[0, 1]$ . Assume also that

$$\iint F(dx) M(x, dy) \|h_n(x, y, \mathbf{X}_n) - h(x, y)\| = o_p(1) \quad (3.16)$$

and

$$n^{1/2} \iint F(dx) M(x, dy) (h_n(x, y, \mathbf{X}_n) - h(x, y)) = o_p(1). \quad (3.17)$$

Then the estimate  $\langle Z_n \rangle$  defined by

$$Z_n = \frac{1}{2m_n} \sum_{j=1}^{m_n} h_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,2}) + \frac{1}{2m_n} \sum_{j=n-m_n+1}^n h_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,1})$$

satisfies (1.1).

In some applications  $h$  will be of the form

$$h(x, y) = W\phi(x, y), \quad x, y \in S, \quad (3.18)$$

for some  $m \times l$  matrix  $W$  and some measurable function  $\phi$  from  $S \times S$  into  $\mathbb{R}^l$  such that

$$E(\|\phi(X_0, X_1)\|^2) = \iint F(dx) M(x, dy) \|\phi(x, y)\|^2 < \infty.$$

In this case a possible choice of  $z_n$  is given by

$$z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) \frac{1}{n} \sum_{j=1}^n \phi_n(X_{j-1}, X_j; \tilde{\mathbf{X}}_n), \quad (3.19)$$

where  $W_n$  is a measurable function from  $S^{2n+2}$  into the set of all  $m \times l$  matrices and  $\phi_n$  is a measurable function from  $S^2 \times S^{n+1}$  into  $\mathbb{R}^l$ . We can prove the following result.

**Theorem 3.5.** *Suppose (B0) holds and  $h$  is given by (3.18). Suppose that*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \int M(X_{j-1}, dy) (W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) \phi_n(X_{j-1}, y; \tilde{\mathbf{X}}_n) - W\phi(X_{j-1}, y)) \xrightarrow{P} 0, \quad (3.20)$$

$$W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) \xrightarrow{P} W \quad (3.21)$$

and

$$\frac{1}{n} \sum_{j=1}^n \int M(X_{j-1}, dy) \|\phi_n(X_{j-1}, y; \tilde{\mathbf{X}}_n) - \phi(X_{j-1}, y)\|^2 \xrightarrow{P} 0. \quad (3.22)$$

Then (3.3) holds with  $z_n$  defined in (3.19), i.e.

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) \phi_n(X_{j-1}, X_j; \tilde{\mathbf{X}}_n) - W\phi(X_{j-1}, X_j)) \xrightarrow{P} 0. \quad (3.23)$$

Consequently, the estimate  $\langle Z_n \rangle$  defined by

$$\begin{aligned} Z_n &= \frac{1}{2m_n} \sum_{j=1}^{m_n} W_{m_n}(\mathbf{X}_{n,1}, \mathbf{X}_{n,2}) \phi_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,2}) \\ &\quad + \frac{1}{2m_n} \sum_{j=n-m_n+1}^n W_{m_n}(\mathbf{X}_{n,2}, \mathbf{X}_{n,1}) \phi_{m_n}(X_{j-1}, X_j; \mathbf{X}_{n,1}) \end{aligned}$$

satisfies (1.1).

**Proof.** Set  $\hat{W}_n = W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n)$ ,

$$\Phi_j = \phi(X_{j-1}, X_j) - \int M(X_{j-1}, dy) \phi(X_{j-1}, y)$$

and

$$\Phi_{n,j} = \phi_n(X_{j-1}, X_j; \tilde{\mathbf{X}}_n) - \int M(X_{j-1}, dy) \phi_n(X_{j-1}, y; \tilde{\mathbf{X}}_n).$$

In view of (3.20), it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\hat{W}_n \Phi_{n,j} - W\Phi_j) = \hat{W}_n \frac{1}{\sqrt{n}} \sum_{j=1}^n (\Phi_{n,j} - \Phi_j) + (\hat{W}_n - W) \frac{1}{\sqrt{n}} \sum_{j=1}^n \Phi_j \xrightarrow{P} 0. \quad (3.24)$$

It follows from Theorem 3.2 that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\Phi_{n,j} - \Phi_j) \xrightarrow{P} 0.$$

One can also verify easily that  $1/\sqrt{n} \sum_{j=1}^n \Phi_j$  is bounded in probability. Combining the above with (3.21) yields the desired result (3.24).  $\square$

#### 4. Estimating the innovation variance in nonparametric autoregression

In the following example we take the state space to be the real line and the Markov kernel  $M$  to be defined by

$$M(x, dy) = g(y - r(x)) dy,$$

where  $r$  is a Lipschitz continuous function on  $\mathbb{R}$  and  $g$  is a bounded positive Lebesgue density with zero mean and finite fourth moment. We also assume the limits

$$\alpha = \lim_{x \rightarrow -\infty} \frac{r(x)}{x} \quad \text{and} \quad \beta = \lim_{x \rightarrow \infty} \frac{r(x)}{x}$$

exist and satisfy

$$\alpha < 1, \quad \beta < 1 \quad \text{and} \quad \alpha\beta < 1.$$

Then the chain is Lebesgue-irreducible, aperiodic and satisfies (DC) with a function  $V$  of the form

$$V(x) = \begin{cases} 1 + ax, & x \geq 0, \\ 1 - bx, & x < 0, \end{cases}$$

for properly chosen positive constants  $a$  and  $b$ ; see Bhattacharya and Lee (1995). Thus (B0) holds. Note also that the stationary density  $f$  is bounded.

A consistent and asymptotically normal estimate of the variance of the innovation density  $g$  is given by

$$\frac{1}{n} \sum_{j=1}^n (X_j - r(X_{j-1}))^2$$

if the regression function  $r$  is known. Let us now show that, if  $r$  is unknown, one can still construct an estimate  $\langle Z_n \rangle$  that is asymptotically equivalent to the above estimate in the sense that

$$n^{1/2} \left( Z_n - \frac{1}{n} \sum_{j=1}^n (X_j - r(X_{j-1}))^2 \right) \xrightarrow{P} 0.$$

For this we describe the function  $z_n$  used in the construction given in Lemma 3.1:

$$z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = \frac{\sum_{j=1}^n I[\tilde{f}_n(X_{j-1}) \geq c_n] (X_j - \tilde{r}_n(X_{j-1}))^2}{\sum_{j=1}^n I[\tilde{f}(X_{j-1}) \geq c_n]},$$

where  $c_n$  is a small positive number,  $\tilde{r}_n$  is the kernel regression estimate

$$\tilde{r}_n(x) = \frac{\sum_{j=1}^n \tilde{X}_j w((x - \tilde{X}_{j-1})/d_n)}{\sum_{j=1}^n w((x - \tilde{X}_{j-1})/d_n)}, \quad x \in \mathbb{R},$$

and  $\tilde{f}_n$  is the kernel estimate of the stationary density  $f$ ,

$$\tilde{f}_n(x) = \frac{1}{nd_n} \sum_{j=1}^n w\left(\frac{x - \tilde{X}_{j-1}}{d_n}\right), \quad x \in \mathbb{R}, \quad (4.1)$$

both based on the same symmetric kernel  $w$  with support  $[-1, 1]$  and the same window length  $d_n$ .

We shall conclude the desired (3.3) from an application of Theorem 3.5 with  $W = 1$ ,  $\phi(x, y) = (y - r(x))^2$ ,  $W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = n/\sum_{j=1}^n I[\tilde{f}_n(X_{j-1}) \geq c_n]$  and  $\phi_n(x, y; \tilde{\mathbf{X}}_n) = (y - \tilde{r}_n(x))^2 I[\tilde{f}_n(x) \geq c_n]$ . It is easy to check that the assumptions of this theorem are implied by the following three conditions:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\tilde{r}_n(X_{j-1}) - r(X_{j-1}))^2 I[\tilde{f}_n(X_{j-1}) \geq c_n] \xrightarrow{P} 0, \quad (4.2)$$

$$\frac{1}{n} \sum_{j=1}^n I[\tilde{f}_n(X_{j-1}) \geq c_n] \xrightarrow{P} 1, \quad (4.3)$$

$$\frac{1}{n} \sum_{j=1}^n (\tilde{r}_n(X_{j-1}) - r(X_{j-1}))^4 I[\tilde{f}_n(X_{j-1}) \geq c_n] \xrightarrow{P} 0. \quad (4.4)$$

To verify these conditions we assume that

$$c_n \rightarrow 0, \quad nd_n^4 \rightarrow 0 \quad \text{and} \quad nd_n^2 c_n^4 \rightarrow \infty.$$

Using this, the Lipschitz continuity of  $r$ , the boundedness of the stationary density  $f$  and the finiteness of the fourth moment of the innovation density  $g$ , we find by standard arguments that

$$E((\tilde{r}_n(X_0) - r(X_0))^2 I[\tilde{f}_n(X_0) \geq c_n]) = O(d_n^2) + O(n^{-1} d_n^{-1} c_n^{-2}) = o(n^{-1/2})$$

and

$$E((\tilde{r}_n(X_0) - r(X_0))^4 I[\tilde{f}_n(X_0) \geq c_n]) = O(d_n^4) + O(n^{-2} d_n^{-4} c_n^{-4}) = o(1).$$

From these we can immediately conclude (4.2) and (4.4). Finally, we can verify that

$$E(|\tilde{f}_n(X_0) - f(X_0)|) \leq \sup_{t \in \mathbb{R}} f(t) \int E(|\tilde{f}_n(x) - f(x)|) dx \rightarrow 0.$$

This yields (4.3).

## 5. Weighted least-squares estimation

In this example we treat an autoregressive model with martingale difference innovations. The state space is the real line  $\mathbb{R}$  endowed with its Borel  $\sigma$ -field  $\mathcal{B}$ . We assume that the Markov kernel  $M$  satisfies the following conditions. Of course, (M2) below means that  $E(X_j|X_{j-1}) = \rho X_{j-1}$ .

(M1) There exists a positive, bounded, continuous function  $q$  such that

$$M(x, B) = \int_B q(x, y) dy, \quad x \in \mathbb{R}, B \in \mathcal{B}.$$

(M2) For some  $\rho$  in  $(-1, 1)$ ,

$$\int M(x, dy)y = \rho x, \quad x \in \mathbb{R}.$$

(M3) The conditional central fourth moment  $\chi$  satisfies the growth conditions

$$\sup_{x \in \mathbb{R}} \frac{\chi(x)}{1 + x^4} < \infty \quad \text{and} \quad \lim_{|x| \rightarrow \infty} \frac{\chi(x)}{1 + x^4} = 0.$$

(M4) The conditional variance  $v$  is bounded away from zero and satisfies

$$|v(y) - v(x)| \leq L(1 + |x|)|y - x|^\gamma, \quad x, y \in \mathbb{R}, |x - y| < 1,$$

for constants  $L < \infty$  and  $\gamma > 0$ .

Of course, the conditional variance  $v$  and the conditional fourth central moment  $\chi$  are defined by

$$v(x) = \int M(x, dy)(y - \rho x)^2 \quad \text{and} \quad \chi(x) = \int M(x, dy)(y - \rho x)^4, \quad x \in \mathbb{R}.$$

It follows from (M1) that  $M^n(x, A) > 0$  for every event  $A$  with positive Lebesgue measure and every  $n$ . Thus the Markov chain generated by  $M$  will be Lebesgue-irreducible and aperiodic; see Theorem A1.2 in Tong (1990) for the latter. Moreover, (M1)–(M3) imply conditions (2.2)–(2.4) in Theorem 2.2 with  $\mathcal{Q}$  the singleton  $\{M\}$ ,  $V(x) = 1 + x^4$ ,  $\lambda = (|\rho| + 1)/2$ ,  $C = [-K, K]$  for some sufficiently large  $K$ ,  $b = \sup_{x \in C} \int q(x, y)(1 + y^4) dy$ ,  $v$  the uniform distribution on  $C$  and  $\delta = \inf_{|x| \leq K, |y| \leq K} q(x, y)$ . Thus, in view of this theorem, the Markov chain can be chosen to be stationary. By the properties of  $q$ , the stationary distribution  $F = \pi_M$  has a positive and bounded Lebesgue density  $f$ . Moreover,

$$E(X_0^4) = \int x^4 f(x) dx < \infty.$$

An estimate of  $\rho$  is given by the least-squares estimate



$$\hat{\rho}_n = \frac{\sum_{j=1}^n X_{j-1}X_j}{\sum_{j=1}^n X_{j-1}^2} = \rho + \frac{\sum_{j=1}^n X_{j-1}(X_j - \rho X_{j-1})}{\sum_{j=1}^n X_{j-1}^2}.$$

This estimate satisfies

$$n^{1/2} \left( \hat{\rho}_n - \rho - \frac{1}{n} \sum_{j=1}^n X_{j-1}(X_j - \rho X_{j-1}) / \mathbb{E}(X_0^2) \right) \xrightarrow{P} 0,$$

so that  $\mathcal{L}(n^{1/2}(\hat{\rho}_n - \rho) | P) \Rightarrow \mathcal{N}(0, \mathbb{E}(X_0^2 v(X_0)) / (\mathbb{E}(X_0^2))^2)$ . If  $v$  were known, one could use instead the weighted least-squares estimate

$$\hat{\rho}_n^* = \frac{\sum_{j=1}^n u(X_{j-1})X_j}{\sum_{j=1}^n u(X_{j-1})X_{j-1}} = \rho + \frac{\sum_{j=1}^n u(X_{j-1})(X_j - \rho X_{j-1})}{\sum_{j=1}^n u(X_{j-1})X_{j-1}},$$

where

$$u(x) = \frac{x}{v(x)}, \quad x \in \mathbb{R}.$$

This estimate satisfies

$$n^{1/2} \left( \hat{\rho}_n^* - \rho - \frac{1}{n\tau} \sum_{j=1}^n u(X_{j-1})(X_j - \rho X_{j-1}) \right) \xrightarrow{P} 0,$$

so that  $\mathcal{L}(n^{1/2}(\hat{\rho}_n^* - \rho) | P) \Rightarrow \mathcal{N}(0, 1/\tau)$ , where  $\tau = \mathbb{E}(u(X_0)X_0) = \mathbb{E}(X_0^2/v(X_0))$ . It follows from the Cauchy–Schwarz inequality that

$$(\mathbb{E}(X_0^2))^2 \leq \mathbb{E}(X_0^2 v(X_0)) \mathbb{E}(X_0^2 / v(X_0))$$

with equality if and only if  $v$  is constant. This shows that the asymptotic variance of the least-squares estimate is larger than that of the weighted least-squares estimate unless  $v$  is constant. Wefelmeyer (1997) has shown that an estimate  $\tilde{\rho}_n$ , that is asymptotically equivalent to the weighted least-squares estimate in the sense that

$$n^{1/2}(\tilde{\rho}_n - \hat{\rho}_n^*) \xrightarrow{P} 0,$$

is a least dispersed regular estimate of  $\rho$ . He also sketched a possible construction of such an estimate that does not require knowledge of  $v$ . His construction, however, relies on a uniformly consistent estimate of  $v$  which is unavailable in the present setting as the support of  $f$  is  $\mathbb{R}$ .

Let us now construct an estimate that is asymptotically equivalent to the weighted least-squares estimate but does not require knowledge of  $v$ . We use the method of Lemma 3.1 with

$$z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = \frac{\sum_{j=1}^n \tilde{u}_n(X_{j-1})X_j}{\sum_{j=1}^n \tilde{u}_n(X_{j-1})X_{j-1}} = \rho + \frac{\sum_{j=1}^n \tilde{u}_n(X_{j-1})(X_j - \rho X_{j-1})}{\sum_{j=1}^n \tilde{u}_n(X_{j-1})X_{j-1}},$$

where  $\tilde{u}_n(x) = u_n(x; \tilde{\mathbf{X}}_n)$  is an estimate of  $u(x)$  to be determined later. To verify the desired (3.3) we apply Theorem 3.5 with  $W = [1, 1/E(X_0u(X_0))]$ ,  $W_n(\mathbf{X}_n, \tilde{\mathbf{X}}_n) = [1, n/\sum_{j=1}^n \tilde{u}_n(X_{j-1})X_{j-1}]$ ,

$$\phi(x, y) = \begin{pmatrix} u(x)(y - \rho x) \\ \rho - \rho x \end{pmatrix} \quad \text{and} \quad \phi_n(x, y; \tilde{\mathbf{X}}_n) = \begin{pmatrix} \tilde{u}_n(x)(y - \rho x) \\ \rho - \rho x \end{pmatrix}.$$

It is now easy to see that the assumptions of this theorem are implied by

$$\frac{1}{n} \sum_{j=1}^n (\tilde{u}_n(X_{j-1}) - u(X_{j-1}))^2 v(X_{j-1}) \xrightarrow{P} 0.$$

A sufficient condition for this is

$$\int (\tilde{u}_n(x) - u(x))^2 v(x) f(x) dx \xrightarrow{P} 0. \quad (5.1)$$

Let us now construct  $\tilde{u}_n$  with this property. For this we rely on an estimate  $\tilde{v}_n$  of  $v$ .

Let  $\tilde{\rho}_n$  be the least-squares estimate based on the data  $\tilde{\mathbf{X}}_n$ . Let  $w$  be a twice continuously differentiable symmetric density with support  $[-1, 1]$  and  $d_n$  be a small positive number. We use the kernel estimate

$$\tilde{v}_n(x) = \frac{\sum_{j=1}^n (\tilde{X}_j - \tilde{\rho}_n \tilde{X}_{j-1})^2 w((x - \tilde{X}_{j-1})/d_n)}{\sum_{j=1}^n w((x - \tilde{X}_{j-1})/d_n)}, \quad x \in \mathbb{R}.$$

The estimate  $\tilde{v}_n(x)$  may not be very accurate if  $|x|$  is large or when  $\tilde{f}_n(x)$  is small, where  $\tilde{f}_n$  is the kernel estimate of stationary density  $f$  defined in (4.1). For this reason we take

$$\tilde{u}_n(x) = \frac{x}{\tilde{v}_n(x)} \tilde{I}_n(x), \quad x \in \mathbb{R},$$

where

$$\tilde{I}_n(x) = I[\tilde{f}_n(x) \geq c_n, |x| \leq A_n]. \quad x \in \mathbb{R},$$

$c_n$  is a small positive number and  $A_n$  is a large positive number. More precisely, we assume that

$$d_n \rightarrow 0, \quad c_n \rightarrow 0, \quad A_n \rightarrow \infty, \quad A_n d_n^\gamma \rightarrow 0 \quad \text{and} \quad \frac{A_n^5}{c_n^3 n d_n^{5/2}} \rightarrow 0.$$

Since  $\int u^2(x)v(x)f(x) dx < \infty$  and  $P(\tilde{f}_n(X_0) \geq c_n, |X_0| \leq A_n) \rightarrow 1$ , we find that

$$\int u^2(x)v(x)(1 - \tilde{I}_n(x))f(x) dx \xrightarrow{P} 0. \quad (5.2)$$

We shall show that

$$\sup_{x \in \mathbb{R}} \tilde{I}_n(x) |\tilde{v}_n(x) - v(x)| \xrightarrow{P} 0. \quad (5.3)$$

Since  $\int u^2(x)v(x)f(x) dx < \infty$  and  $v$  is bounded away from zero, this will imply

$$\int \tilde{I}_n(x)(\tilde{u}_n(x) - u(x))^2 v(x)f(x) dx \xrightarrow{P} 0. \quad (5.4)$$

The desired (5.1) follows from (5.2) and (5.4).

To verify (5.3), let

$$\bar{v}_n(x) = \frac{\sum_{j=1}^n (\tilde{X}_j - \rho \tilde{X}_{j-1})^2 w((x - \tilde{X}_{j-1})/d_n)}{\sum_{j=1}^n w((x - \tilde{X}_{j-1})/d_n)}, \quad x \in \mathbb{R}.$$

and

$$v_n(x) = \frac{\sum_{j=1}^n v(\tilde{X}_{j-1}) w((x - \tilde{X}_{j-1})/d_n)}{\sum_{j=1}^n w((x - \tilde{X}_{j-1})/d_n)}, \quad x \in \mathbb{R}.$$

It follows from (M4) that

$$\sup_{x \in \mathbb{R}} \tilde{I}_n(x) |v_n(x) - v(x)| = O(A_n d_n^\nu). \quad (5.5)$$

Since the fourth moment of  $f$  is finite,  $E(X_0^2(X_1 - \rho X_0)^2) = \int x^2 v(x)f(x) dx$  is finite in view of (M3). Using the inequality  $a^2 P(|Y| > a) \leq E(Y^2 I[|Y| > a])$ , we obtain

$$\max_{1 \leq j \leq n} n^{-1/2} |\tilde{X}_{j-1}(\tilde{X}_j - \rho \tilde{X}_{j-1})| \xrightarrow{P} 0 \quad \text{and} \quad \max_{1 \leq j \leq n} n^{1/2} |\tilde{X}_{j-1}| \xrightarrow{P} 0.$$

This and the  $\sqrt{n}$ -consistency of the least-squares estimate show that

$$\sup_{x \in \mathbb{R}} \tilde{I}_n(x) |\tilde{v}_n(x) - \bar{v}_n(x)| \xrightarrow{P} 0. \quad (5.6)$$

Next, verify that

$$\sup_{x \in \mathbb{R}} \tilde{I}_n(x) |\bar{v}_n(x) - v_n(x)| \leq \frac{1}{C_n} \sup_{|x| \leq A_n} |B_n(x)|, \quad (5.7)$$

where

$$B_n(x) = \frac{1}{nd_n} \sum_{j=1}^n ((\tilde{X}_j - \rho \tilde{X}_{j-1})^2 - v(\tilde{X}_{j-1})) w\left(\frac{x - \tilde{X}_{j-1}}{d_n}\right), \quad x \in \mathbb{R}.$$

Note that  $B_n$  is twice continuously differentiable and that  $|B_n''(x)| \leq \|w''\|_\infty S_n/d_n^3$  for all  $x \in \mathbb{R}$ , where

$$S_n = \frac{1}{n} \sum_{j=1}^n |(\tilde{X}_j - \rho \tilde{X}_{j-1})^2 - v(\tilde{X}_{j-1})|.$$

Since  $nE(B_n^2(x)) \leq d_n^{-2}E(\chi(X_0)w^2((x - X_0)/d_n))$  and the maps  $f$  and  $w$  are bounded, we obtain, in view of (M3), that

$$v_0 = \sup_{|x| \leq A_n} E(B_n^2(x)) = O\left(\frac{A_n^4}{nd_n}\right).$$

Similarly, we derive

$$v_1 = \sup_{|x| \leq A_n} E((B_n'(x))^2) = O\left(\frac{A_n^4}{nd_n^3}\right).$$

Let  $R$  be a positive integer and  $x_i = -A_n + (2i - 1)A_n/R$ ,  $i = 1, \dots, R$ . If  $|x - x_i| \leq A_n/R$ , then we find by a Taylor expansion that  $|B_n(x)| \leq |B_n(x_i)| + A_n|B_n'(x_i)|/R + A_n^2\|w''\|_\infty S_n/(R^2 d_n^3)$ . Thus it follows from a standard argument that

$$P\left(\sup_{|x| \leq A_n} |B_n(x)| \geq 3\eta c_n\right) \leq \frac{Rv_0}{c_n^2 \eta^2} + \frac{A_n^2 v_1}{Rc_n^2 \eta^2} + \frac{A_n^2 \|w''\|_\infty E(S_n)}{R^2 d_n^3 c_n \eta}$$

for  $\eta > 0$ . Taking  $R = R_n \sim d_n^{-3/2} c_n^{-1} A_n$  shows that

$$\sup_{|x| \leq A_n} |B_n(x)|/c_n \xrightarrow{P} 0. \quad (5.8)$$

The desired result (5.3) now follows from (5.5)–(5.8).

## 6. The $q$ -split

Up to now we have only considered splitting into two parts. But the above ideas can be easily generalized to splitting into more than two parts. Let  $q$  be a integer,  $q \geq 2$ , denoting the number of splits. Write  $n = qm_n + (q - 1)t_n + r_n$ , where  $m_n$ ,  $t_n$  and  $r_n$  are non-negative integers such that

$$t_n \rightarrow \infty \quad \text{and} \quad \frac{n - qm_n}{\sqrt{n}} \rightarrow 0. \quad (6.1)$$

Define the  $q$  parts by

$$\mathbf{X}_{n,i} = (X_{(i-1)(m_n+t_n)}, \dots, X_{im_n+(i-1)t_n}), \quad i = 1, \dots, q.$$

To describe our results we assume from now on that the underlying space is large enough for there to be  $q - 1$  chains  $\{\tilde{X}_{2,j} : j = 0, 1, 2, \dots\}$ ,  $\dots$ ,  $\{\tilde{X}_{q,j} : j = 0, 1, 2, \dots\}$  which are independent copies of the original chain  $\{X_j : j = 0, 1, 2, \dots\}$  and are independent of it. Set

$$\tilde{\mathbf{X}}_{i,N} = (\tilde{X}_{i,0}, \dots, \tilde{X}_{i,N}), \quad i = 2, \dots, q, N = 1, 2, \dots$$

The next lemma is an immediate consequence of Lemma 2.1.

**Lemma 6.1.** *If (B0) and (6.1) hold, then the sequences of distributions*

$$\langle \mathcal{L}(\mathbf{X}_{n,1}, \dots, \mathbf{X}_{n,q} | P) \rangle \quad \text{and} \quad \langle \mathcal{L}(\mathbf{X}_{n,1}, \tilde{\mathbf{X}}_{2,m_n}, \dots, \tilde{\mathbf{X}}_{q,m_n} | P) \rangle$$

*are mutually contiguous.*

By analogy with (3.2), we now consider the estimate

$$Z_n = \frac{1}{q!} \sum_{i_1, \dots, i_q} z_{m_n}(\mathbf{X}_{n,i_1}, \dots, \mathbf{X}_{n,i_q}), \quad (6.2)$$

where the summation is over all permutations  $(i_1, \dots, i_q)$  of  $(1, \dots, q)$  and  $z_N$  is a measurable function from  $S^{q(N+1)}$  to  $\mathbb{R}^m$  for each positive integer  $N$ . We now obtain the following analogue to Lemma 3.1.

**Lemma 6.2.** *Suppose (B0) and (6.1) hold and*

$$n^{1/2} \left( z_n(\mathbf{X}_n, \tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}) - \frac{1}{n} \sum_{j=1}^n h(X_{j-1}, X_j) \right) \xrightarrow{P} 0. \quad (6.3)$$

*Then  $\langle Z_n \rangle$  defined by (6.2) satisfies (1.1).*

**Remark.** Of course, sufficient conditions for (6.3) can be obtained from those given in Theorems 3.2 and 3.5 by simply replacing  $\tilde{\mathbf{X}}_n$  by  $\tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}$ .

## 7. Specifics for semi-parametric models

In this section we shall formulate a result more suitable for semi-parametric models. A standard approach to these models is the use of discretized  $\sqrt{n}$ -consistent estimates. Such estimates can be treated as non-stochastic sequences in the proofs. When combined with contiguity arguments, this often leads to much simpler proofs. See Bickel (1982) or Schick (1986) for this approach. For the ensuing development it is imperative that we stress the dependence on the finite-dimensional parameter of interest, but we suppress the dependence on the infinite-dimensional nuisance parameter. The formulation will be for the more general  $q$ -split procedure discussed in the previous section.

Let  $\Theta$  be an open subset of  $\mathbb{R}^m$ . For each  $\vartheta \in \Theta$ , let  $P_\vartheta$  be a probability measure for which the chains  $X_0, X_1, \dots, \tilde{X}_{2,0}, \tilde{X}_{2,1}, \dots, \tilde{X}_{q,0}, \tilde{X}_{s,1}, \dots$  are independent stationary

Markov chains with the same transition probability  $M_{\vartheta}$  and the same stationary distribution  $F_{\vartheta}$ . Assume now that  $P = P_{\theta}$  for some fixed point  $\theta$  in  $\Theta$ . We want to construct an estimate  $\langle Z_n \rangle$  such that

$$n^{1/2} \left( Z_n - \theta - \frac{1}{n} \sum_{j=1}^n W(\theta) \phi(X_{j-1}, X_j; \theta) \right) \xrightarrow{P_{\theta}} 0, \quad (7.1)$$

where, for each  $\vartheta \in \Theta$ ,  $W(\vartheta)$  is an  $m \times l$  matrix and  $\phi$  is a map from  $S^2 \times \Theta$  into  $\mathbb{R}^l$  such that

$$\int M_{\vartheta}(x, dy) \phi(x, y; \vartheta) = 0 \quad \text{and} \quad \int \int F_{\vartheta}(dx) M_{\vartheta}(x, dy) \|\phi(x, y; \vartheta)\|^2 < \infty.$$

Let  $\mathbf{X}_{n,1}, \dots, \mathbf{X}_{n,q}$  and  $\tilde{\mathbf{X}}_{2,N}, \dots, \tilde{\mathbf{X}}_{q,N}$  be as defined in the previous section. One is tempted to take

$$Z_n = U_n + \Psi_{m_n}(U_n), \quad (7.2)$$

where  $U_n$  is an estimate of  $\theta$  and

$$\Psi_{m_n}(\vartheta) = \frac{1}{q!} \sum_{i_1, \dots, i_q} W_{m_n}(\vartheta; \mathbf{X}_{n,i_1}, \dots, \mathbf{X}_{n,i_q}) \Phi_{m_n}(\vartheta; \mathbf{X}_{n,i_1}, \dots, \mathbf{X}_{n,i_q}), \quad \vartheta \in \Theta, \quad (7.3)$$

with  $W_{m_n}$  a measurable function from  $\Theta \times S^{q m_n + q}$  into the set of  $m \times l$  matrices and

$$\Phi_{m_n}(\vartheta; \mathbf{X}_{n,1}, \dots, \mathbf{X}_{n,q}) = \frac{1}{m_n} \sum_{j=1}^{m_n} \phi_{m_n}(X_{j-1}, X_j; \vartheta; \mathbf{X}_{n,2}, \dots, \mathbf{X}_{n,q})$$

for some measurable function  $\phi_{m_n}$  from  $S^2 \times \Theta \times S^{(q-1)(m_n+1)}$  into  $\mathbb{R}^l$ .

By a ‘local sequence’ we mean a sequence  $\langle \theta_n \rangle$  in  $\Theta$  such that  $\langle n^{1/2}(\theta_n - \theta) \rangle$  is bounded. To take full advantage of the technical simplification associated with the use of discretized estimates we shall impose the following conditions.

(A0) Condition (B0) holds for the pair  $(M, F) = (M_{\theta}, F_{\theta})$ .

(A1) The estimate  $\langle U_n \rangle$  is  $\sqrt{n}$ -consistent ( $\langle \mathcal{L}(n^{1/2}(U_n - \theta) | P_{\theta}) \rangle$  is tight) and discretized (for every  $\varepsilon > 0$ , there are an integer  $N$  and events  $\langle A_n \rangle$  such that, for each  $n = 1, 2, \dots$ ,  $P_{\theta}(A_n) \geq 1 - \varepsilon$  and the image of  $A_n$  under  $U_n$  has at most  $N$  elements).

(A2) For every local sequence  $\langle \theta_n \rangle$ ,

$$n^{1/2} \left( \theta_n + \frac{1}{n} \sum_{j=1}^n W(\theta_n) \phi(X_{j-1}, X_j; \theta_n) - \theta - \frac{1}{n} \sum_{j=1}^n W(\theta) \phi(X_{j-1}, X_j; \theta) \right) \xrightarrow{P_{\theta}} 0.$$

(A3) For every local sequence  $\langle \theta_n \rangle$ , the sequences  $\langle \mathcal{L}(X_0, \dots, X_n | P_{\theta_n}) \rangle$  and  $\langle \mathcal{L}(X_0, \dots, X_n | P_{\theta}) \rangle$  are mutually contiguous.

**Remark.** A discretized version  $\langle \hat{\theta}_n^* \rangle$  of a  $\sqrt{n}$ -consistent estimate  $\langle \hat{\theta}_n \rangle$  can be obtained by assigning  $\hat{\theta}_n^*$  the value on the grid  $\{n^{1/2}j : j \in \mathbb{Z}^m\}$  closest to  $\hat{\theta}_n$ , where  $\mathbb{Z}$  denotes the set of

all integers. More sophisticated discretization procedures are discussed in Fabian and Hannan (1982).

**Remark.** In the construction of efficient estimates for semi-parametric models, the local asymptotic normality condition in the parameter of interest for fixed nuisance parameter automatically implies (A3), while (A2) follows under mild assumptions on the least favourable submodel; see Koul and Schick (1997) for the general case and Drost *et al.* (1997) and Jeganathan (1995) for the case of adaptive estimation.

Under (A1)–(A3), (7.1) follows if

$$n^{1/2} \left( \Psi_{m_n}(\theta_n) - \frac{1}{n} \sum_{j=1}^n W(\theta_n) \phi(X_{j-1}, X_j; \theta_n) \right) \xrightarrow{P_{\theta_n}} 0 \quad (7.4)$$

for every local sequence  $\langle \theta_n \rangle$ . Indeed, it follows from (A3) that  $P_{\theta_n}$  can be replaced by  $P_\theta$  and then from (A2) that

$$n^{1/2} \left( \theta_n + \Psi_{m_n}(\theta_n) - \theta - \frac{1}{n} \sum_{j=1}^n W(\theta) \phi(X_{j-1}, X_j; \theta) \right) \xrightarrow{P_\theta} 0$$

for all local sequences  $\langle \theta_n \rangle$ . In view of (A1), we can then replace  $\langle \theta_n \rangle$  by any discretized  $\sqrt{n}$ -consistent estimate, and this is the desired result (7.1).

It follows from (A0) and Lemma 6.1 that  $\langle Q_{n,\theta} \rangle$  and  $\langle \tilde{Q}_{n,\theta} \rangle$  are mutually contiguous, where

$$Q_{n,\vartheta} = \mathcal{L}(\mathbf{X}_{n,1}, \dots, \mathbf{X}_{n,q} | P_\vartheta) \quad \text{and} \quad \tilde{Q}_{n,\vartheta} = \mathcal{L}(\mathbf{X}_{n,1}, \tilde{\mathbf{X}}_{2,m_n}, \dots, \tilde{\mathbf{X}}_{q,m_n} | P_\vartheta), \quad \vartheta \in \Theta.$$

Combined with (A3), this gives that  $\langle Q_{n,\theta_n} \rangle$  and  $\langle \tilde{Q}_{n,\theta_n} \rangle$  are mutually contiguous for every local sequence  $\langle \theta_n \rangle$ . Thus (7.4) is implied by

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\tilde{W}_n(\theta_n) \tilde{\phi}_n(X_{j-1}, X_j; \theta_n) - W(\theta_n) \phi(X_{j-1}, X_j; \theta_n)) \xrightarrow{P_{\theta_n}} 0, \quad (7.5)$$

where

$$\tilde{W}_n(\vartheta) = W_n(\theta_n; \mathbf{X}_n, \tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}), \quad \vartheta \in \Theta,$$

and

$$\tilde{\phi}_n(x, y; \vartheta) = \phi_n(x, y; \vartheta; \tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}), \quad x, y \in S, \vartheta \in \Theta.$$

Sufficient conditions for (7.5) can be obtained by slightly modifying the assumptions of Theorem 3.5. This is done in the next theorem.

**Theorem 7.1.** *Let (A0)–(A3) and (6.1) hold and suppose that, for every local sequence  $\langle \theta_n \rangle$ ,*

$$\limsup_n \|W(\theta_n)\| + \int F_{\theta_n}(dx) M_{\theta_n}(x, dy) \|\phi(x, y; \theta_n)\|^2 < \infty, \quad (7.6)$$

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \int M_{\theta_n}(X_{j-1}, dy) \tilde{W}_n(\theta_n) \tilde{\phi}_n(X_{j-1}, y; \theta_n) \xrightarrow{P_{\theta_n}} 0, \quad (7.7)$$

$$\tilde{W}_n(\theta_n) - W(\theta_n) \xrightarrow{P_{\theta_n}} 0 \quad (7.8)$$

and

$$\frac{1}{n} \sum_{j=1}^n \int M_{\theta_n}(X_{j-1}, dy) \|\tilde{\phi}_n(X_{j-1}, y; \theta_n) - \phi(X_{j-1}, y; \theta_n)\|^2 \xrightarrow{P_{\theta_n}} 0. \quad (7.9)$$

Then (7.1) holds with  $\langle Z_n \rangle$  defined in (7.2) and (7.3).

## 8. Efficient estimation in a heteroscedastic autoregressive model of order 1 with symmetric errors

In this section we consider a heteroscedastic autoregression model of order 1 with a smooth scale function and symmetric innovations. More precisely, we assume that the transition kernel  $M = M_\rho$  is defined by

$$M_\rho(x, dy) = \frac{1}{\sigma(x)} g\left(\frac{y - \rho x}{\sigma(x)}\right) dy,$$

where  $g$  is a positive symmetric Lebesgue density with variance 1,  $\sigma$  is a Lipschitz continuous function from  $\mathbb{R}$  into  $(0, \infty)$  which is bounded away from 0 and  $\rho$  is a real number. To guarantee stationarity and ergodicity we impose the condition

$$\rho^2 + \limsup_{|x| \rightarrow \infty} \frac{\sigma^2(x)}{x^2} < 1. \quad (8.1)$$

Since the density  $g$  is positive, the corresponding chain is Lebesgue-irreducible and aperiodic. It follows from (8.1) that, for a small enough interval  $R$  containing  $\rho$ , the class  $\mathcal{Q} = \{M_r : r \in R\}$  satisfies (2.2)–(2.4) of Theorem 2.2 with  $V(x) = 1 + x^2$  and  $C = [-K, K]$  for some sufficiently large  $K$ . The corresponding invariant measures  $\{F_r = \pi_{M_r} : r \in R\}$  have positive bounded Lebesgue densities  $\{f_r : r \in R\}$  and uniformly bounded second moments (see (2.8)). One can even show that

$$\lim_{r \rightarrow \rho} \|F_r - F_\rho\|_V = \lim_{r \rightarrow \rho} \int (1 + x^2) |f_r(x) - f_\rho(x)| dx \rightarrow 0. \quad (8.2)$$

Indeed, in view of the remark after Theorem 2.2, this follows from

$$\lim_{r \rightarrow \rho} \int \|M_r - M_\rho\|_V dF_\rho \rightarrow 0. \quad (8.3)$$

Since  $\int V(x) f_\rho(x) dx < \infty$  and  $\sup_{r \in R} \|M_r - M_\rho\|_V \leq LV$  for some finite constant  $L$  by Theorem 2.2, the Lebesgue dominated convergence theorem shows that (8.3) is implied by



$$\lim_{r \rightarrow \rho} \|M_r - M_\rho\|_V(x) = 0, \quad x \in \mathbb{R}. \quad (8.4)$$

The substitution  $u = (y - \rho x)/\sigma(x)$  yields  $\|M_r - M_\rho\|_V(x) = \int (1 + (\rho x + \sigma(x)u)^2) |g(u - (r - \rho)x/\sigma(x)) - g(u)| du$ . Thus (8.4) follows if we show that

$$\Delta(t) = \int (1 + y^2) |g(y - t) - g(y)| dy \rightarrow 0 \quad \text{as } t \rightarrow 0. \quad (8.5)$$

Since  $y^2 - (y - t)^2 = 2t(y - t) + t^2$ , we find that

$$\Delta(t) \leq \int |(1 + (y - t)^2)g(y - t) - (1 + y^2)g(y)| dy + \int (t^2 + 2|t||y - t|)g(y - t) dy.$$

Therefore (8.5) follows by the  $L_1$ -continuity for translations (cf. Theorem 9.5 in Rudin 1974) and the fact that  $\int (1 + y^2)g(y) dy < \infty$ .

Assume from now on also that the innovation density  $g$  has finite Fisher information for location, i.e.  $g$  is absolutely continuous and

$$J(g) = \int l^2(x)g(x) dx < \infty, \quad \text{where } l(x) = -\frac{g'(x)}{g(x)}, \quad x \in \mathbb{R}.$$

We do not assume that  $g$  and  $\sigma$  are known. We want to construct an estimate  $\langle Z_n \rangle$  which satisfies

$$n^{1/2} \left( Z_n - \rho - \frac{1}{n} \sum_{j=1}^n \frac{1}{\gamma(\rho)} \frac{X_{j-1}}{\sigma(X_{j-1})} l \left( \frac{X_j - \rho X_{j-1}}{\sigma(X_{j-1})} \right) \right) \xrightarrow{P_\rho} 0 \quad (8.6)$$

with

$$\gamma(\rho) = J(g) E_\rho \left( \frac{X_0^2}{\sigma^2(X_0)} \right).$$

It follows from the results in Maercker (1997) that such an estimate is an efficient and adaptive estimate of  $\rho$  in the presence of the nuisance parameters  $\sigma$  and  $g$ . She also constructs such an estimate by extending the approach of Bickel (1982). This approach uses only a small part of the sample to construct the score function. We shall use essentially all the data by relying on our sample splitting technique. Moreover, we shall relax some of her assumptions such as (iii) of her Assumption 2.1 and (ii), (iii) and (v) of her Assumption 2.2. In particular, we are happy to get rid of her Assumption 2.2(v) as it is not clear how to choose  $\gamma_m$  if the error density  $g$  is unknown. We achieve this mainly by using a different approach to estimating the volatility function. Rather than estimating  $\sigma$  directly, we estimate a multiple of it. This is made precise below. But let us first observe that assumptions (A0)–(A3) from Section 7 hold.

Here, (A0) is met as shown above, and (A1) holds with  $U_n$  a discretized version of the least-squares estimate. It follows from results in Maercker (1997) that (A3) holds if we verify her Assumption B. But this is a consequence of (8.2), and so is the continuity of  $r \mapsto \gamma(r)$  at  $\rho$ . The latter and the results of Maercker (1997) show that (A2) holds. Thus we

can use the construction given in the previous section if we produce estimates that satisfy (7.6)–(7.9).

Before we do so, let us first rewrite (8.6) as

$$n^{1/2} \left( Z_n - \rho - \frac{1}{n} \sum_{j=1}^n \frac{1}{\gamma(\rho) s(X_{j-1})} \frac{X_{j-1}}{s(X_{j-1})} l_0 \left( \frac{X_j - \rho X_{j-1}}{s(X_{j-1})} \right) \right) \xrightarrow{P_\rho} 0, \quad (8.7)$$

where, with  $\mu = \int |y| g(y) dy$ ,

$$s(x) = \mu \sigma(x), \quad g_0(x) = \mu g(\mu x) \quad \text{and} \quad l_0(x) = -\frac{g'_0(x)}{g_0(x)} = \mu l(\mu x).$$

Note also that

$$\gamma(\rho) = J(g_0) E_\rho \left( \frac{X_0^2}{s^2(X_0)} \right).$$

In view of this, we shall construct estimates of  $s$  and  $g_0$  rather than of  $\sigma$  and  $g$ . Note that

$$s(X_0) = E_\rho(|X_1 - \rho X_0| | X_0) \quad (8.8)$$

and that  $g_0$  is a density of  $(X_1 - \rho X_0)/s(X_0)$  and, by symmetry of  $g$ , also a density of  $-(X_1 - \rho X_0)/s(X_0)$ .

For the remainder of this section, we let  $\langle \rho_n \rangle$  be a local sequence for  $\rho$ . We take

$$W(\rho_n) = \frac{1}{\gamma(\rho_n)},$$

$$\phi(x, y; \rho_n) = \frac{x}{s(x)} l_0 \left( \frac{y - \rho_n x}{s(x)} \right), \quad x, y \in \mathbb{R}.$$

We shall apply the construction with  $q \geq 3$  and rely on different estimates of  $s$ . We shall mimic  $(X_j - \rho_n X_{j-1})/s(X_{j-1})$  by using an estimate  $\tilde{s}_{n,1}$  of  $s$  based on all the auxiliary samples  $\tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}$  and mimic  $(\tilde{X}_{i,j} - \rho_n \tilde{X}_{i,j-1})/s(\tilde{X}_{2,j-1})$  by using an estimate  $\tilde{s}_{n,i}$  of  $s$  based on all the auxiliary samples except  $\tilde{\mathbf{X}}_{i,n}$ . Let  $B_1 = \{2, \dots, q\}$  and  $B_i = B_1 - \{i\}$  for  $i = 2, \dots, q$ .

The identity (8.8) suggests the kernel estimates

$$\tilde{s}_{n,i}(x) = \frac{\sum_{v \in B_i} \sum_{j=1}^n |\tilde{X}_{v,j} - \rho_n \tilde{X}_{v,j-1}| w((x - \tilde{X}_{v,j-1})/d_n)}{\sum_{v \in B_i} \sum_{j=1}^n w((x - \tilde{X}_{v,j-1})/d_n)}, \quad x \in \mathbb{R}, i = 1, \dots, q,$$

utilizing a twice continuously differentiable symmetric density  $w$  with support  $[-1, 1]$  and window length  $d_n$ . These estimates can be poor if  $|x|$  is large or if their denominators are too small. To eliminate this problem we shall use the random weights

$$\tilde{I}_{n,i}(x) = I[\tilde{f}_{n,i}(x) \geq c_n, |x| \leq A_n], \quad x \in \mathbb{R}, i = 1, \dots, q,$$

where  $\tilde{f}_{n,i}$  is the kernel density estimate

$$\tilde{f}_{n,i}(x) = \frac{1}{\text{card}(B_i)nd_n} \sum_{v \in B_i} \sum_{j=1}^n w \left( \frac{x - \tilde{X}_{v,j-1}}{d_n} \right), \quad x \in \mathbb{R}, i = 2, \dots, q.$$

Now set

$$e_{n,j} = \frac{X_j - \rho_n X_{j-1}}{\tilde{s}_{n,1}(X_{j-1})} \quad \text{and} \quad \tilde{e}_{i,n,j} = \frac{\tilde{X}_{i,j} - \rho_n \tilde{X}_{i,j-1}}{\tilde{s}_{n,i}(\tilde{X}_{i,j-1})}, \quad j = 1, \dots, n, i = 2, \dots, q.$$

We shall use the variables  $\tilde{e}_{2,n,1}, \dots, \tilde{e}_{2,n,n}, \dots, \tilde{e}_{q,n,1}, \dots, \tilde{e}_{q,n,n}$  to estimate  $g_0$  via a weighted kernel method based on the logistic density  $k$  and window length  $a_n$  as follows:

$$\tilde{g}_n(x) = \frac{1}{2(q-1)na_n} \sum_{i=2}^q \sum_{j=1}^n \tilde{I}_{n,i}(\tilde{X}_{i,j-1}) \left[ k \left( \frac{x - \tilde{e}_{i,n,j}}{a_n} \right) + k \left( \frac{x + \tilde{e}_{i,n,j}}{a_n} \right) \right], \quad x \in \mathbb{R}.$$

Instead of the logistic density one can take any other density  $k$  that is symmetric, bounded, three times differentiable and positive, has finite variance and satisfies  $|k^{(i)}(x)| \leq Ck(x)$  for  $x \in \mathbb{R}$ ,  $i = 1, 2, 3$  and some  $C < \infty$ . We estimate  $l_0$  by

$$\tilde{l}_n(x) = -\frac{\tilde{g}'_n(x)}{b_n + \tilde{g}_n(x)}, \quad x \in \mathbb{R}.$$

Finally, set

$$\tilde{\phi}_n(x, y; \rho_n) = \phi_n(x, y; \rho_n; \tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}) = \frac{x \tilde{I}_{n,1}(x)}{\tilde{s}_{n,1}(x)} \tilde{l}_n \left( \frac{y - \rho_n x}{\tilde{s}_{n,1}(x)} \right), \quad x, y \in \mathbb{R},$$

and

$$\tilde{W}_n(\rho_n) = W_n(\rho_n, \mathbf{X}_n, \tilde{\mathbf{X}}_{2,n}, \dots, \tilde{\mathbf{X}}_{q,n}) = \frac{1}{\tilde{\tau}_n \tilde{J}_n},$$

where

$$\tilde{\tau}_n = \frac{1}{n} \sum_{j=1}^n \frac{X_{j-1}^2 \tilde{I}_{n,1}(X_{j-1})}{\tilde{s}_{n,1}^2(X_{j-1})} \quad \text{and} \quad \tilde{J}_n = \frac{\sum_{j=1}^n \tilde{I}_{n,1}(X_{j-1}) \tilde{l}_n^2(e_{n,j})}{\sum_{j=1}^n \tilde{I}_{n,1}(X_{j-1})}.$$

We shall now verify conditions (7.6)–(7.9) under the following rate requirements on the sequences  $d_n, c_n, A_n, a_n$  and  $b_n$ :

$$d_n \sim n^{1/3}, \quad A_n \sim \log n, \quad c_n^{-1} \sim \log n, \quad na_n^4 b_n^2 \rightarrow \infty \quad \text{and} \quad n^{2/3} a_n^5 b_n c_n^2 \rightarrow \infty.$$

The first condition follows from the continuity of  $r \mapsto \gamma(r)$  at  $\rho$  proved earlier. As in Section 5, we can verify that

$$\sup_{x \in \mathbb{R}} \tilde{I}_{n,i}(x) |\tilde{s}_{n,i}(x) - s(x)| \xrightarrow{P_{\rho_n}} 0, \quad i = 1, \dots, q.$$

Thus it suffices to prove the remaining three conditions with  $\tilde{s}_{n,i} = c \vee \tilde{s}_{n,i}$ , where  $c = \frac{1}{2} \inf_{t \in \mathbb{R}} s(t)$ . Then we verify that

$$\int \mathbb{E}_{\rho_n} \tilde{I}_{n,i}(x) \left( \frac{1}{\tilde{s}_{n,i}(x)} - \frac{1}{s(x)} \right)^2 (1+x^2) f_{\rho_n}(x) dx = O(n^{-1} d_n^{-1} c_n^{-2} + d_n^2) = O(n^{-2/3} c_n^{-2}), \quad (8.9)$$

for  $i = 1, \dots, q$ . This and Lemmas 10.1 and 10.2 in Schick (1993) give

$$\iint F_{\rho_n}(dx) M_{\rho_n}(x, dy) \tilde{I}_{n,1}(x) \left( \tilde{I}_n \left( \frac{y - \rho_n x}{\tilde{s}_{n,1}(x)} \right) - l_0 \left( \frac{y - \rho_n x}{s(x)} \right) \right)^2 \xrightarrow{P_{\rho_n}} 0 \quad (8.10)$$

and

$$\iint F_{\rho_n}(dx) M_{\rho_n}(x, dy) \tilde{I}_{n,1}(x) \left( \frac{x}{\tilde{s}_{n,1}(x)} \tilde{I}_n \left( \frac{y - \rho_n x}{\tilde{s}_{n,1}(x)} \right) - \frac{x}{s(x)} l_0 \left( \frac{y - \rho_n x}{s(x)} \right) \right)^2 \xrightarrow{P_{\rho_n}} 0. \quad (8.11)$$

Here we have used the fact that  $a_n^{-5} b_n^{-1} n^{-2/3} c_n^{-2} \rightarrow 0$ . In view of  $\mathbb{E}_{\rho}(1 - \tilde{I}_{n,1}(X_0)) \rightarrow 0$ , (8.9) and (8.10) yield

$$\tilde{\tau}_n \xrightarrow{P_{\rho_n}} \tau \quad \text{and} \quad \tilde{J}_n \xrightarrow{P_{\rho_n}} J(g_0)$$

and hence (7.8), while (8.11) implies (7.9). Finally, since  $l_0(y) = l_0(-y)$ ,  $\tilde{I}_n(y) = \tilde{I}_n(-y)$  and  $g_0(y) = g_0(-y)$ , we have

$$\int l_0 \left( \frac{y - \rho_n x}{s(x)} \right) g_0 \left( \frac{y - \rho_n x}{s(x)} \right) dy = 0 \quad \text{and} \quad \int \tilde{I}_n \left( \frac{y - \rho_n x}{\tilde{s}_{n,1}(x)} \right) g_0 \left( \frac{y - \rho_n x}{s(x)} \right) dy = 0.$$

This implies (7.7).

## References

- Bickel, P.J. (1982) On adaptive estimation. *Ann. Statist.*, **10**, 647–671.
- Bhattacharya, R.N. and Lee, C. (1995) Ergodicity of nonlinear first order autoregressive models. *J. Theoret. Probab.*, **8**, 207–219.
- Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997) Adaptive estimation in time-series models. *Ann. Statist.*, **25**, 786–817.
- Fabian, V. and Hannan, J. (1982) On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **59**, 459–478.
- Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Math. Dokl.*, **10**, 1174–1176.
- Hájek, J. (1962) Asymptotically most powerful rank-order tests. *Ann. Math. Statist.*, **33**, 1124–1147.
- Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**, 818–887.

- Kessler, M., Schick, A. and Wefelmeyer, W. (1999) The information in the marginal law of a Markov chain. Preprint, Department of Mathematics, University of Binghamton.
- Klaassen, C.A.J. (1987) Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.*, **15**, 1548–1563.
- Koul, H.L. and Schick, A. (1997) Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli*, **3**, 247–277.
- Kreiss, J.-P. (1987a) On adaptive estimation in stationary ARMA processes. *Ann. Statist.*, **15**, 112–133.
- Kreiss, J.-P. (1987b) On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions*, **5**, 59–76.
- LeCam, L. (1960) Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.*, **3**, 37–98.
- Maercker, G. (1997) *Statistical Inference in Conditional Heteroskedastic Autoregressive Models*. Shaker Verlag, Aachen.
- Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*. Berlin: Springer-Verlag.
- Meyn, S.P. and Tweedie, R.L. (1994) Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, **4**, 981–1011.
- Roberts, G.O. and Rosenthal, J.S. (1997) Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, **2**, 13–25.
- Rudin, W. (1974) *Real and Complex Analysis* (2nd edition). New York: McGraw-Hill.
- Schick, A. (1986) On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, **14**, 1139–1151.
- Schick, A. (1987) A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference*, **16**, 89–105. Correction **22** (1989), 269–270.
- Schick, A. (1993) On efficient estimation in regression models. *Ann. Statist.*, **21**, 1486–1521. Correction and addendum **23** (1995), 1862–1863.
- Schick, A. (1999) Efficient estimation in a semiparametric additive autoregressive model. *Stat. Inference Stoch. Process.*, **2**, 69–98.
- Schick, A. and Wefelmeyer, W. (1999) Efficient estimation of invariant distributions of some semiparametric Markov chain models. *Math. Methods Statist.*, **8**, 426–440.
- Tong, H. (1990) *Nonlinear Time Series: A Dynamical Approach*. New York: Oxford University Press.
- van Eeden, C. (1970) Efficiency robust estimation of location. *Ann. Math. Statist.*, **41**, 172–181.
- Wefelmeyer, W. (1997) Adaptive estimators for parameters of the autoregressive function of a Markov chain. *J. Statist. Plann. Inference*, **58**, 389–398.
- Wefelmeyer, W. (1999) Efficient estimation in Markov chain models: an introduction. In S. Ghosh (ed.), *Asymptotics, Nonparametrics, and Time Series*, Statist. Textbooks Monogr. 158, pp. 427–459. New York: Dekker.

Received December 1998 and revised June 2000