The Annals of Statistics
1995, Vol. 23, No. 3, 707–731

MARTINGALE EXPANSIONS AND SECOND ORDER INFERENCE[1]

BY PER ASLAK MYKLAND

University of Chicago

The paper develops a one-step triangular array Edgeworth expansion for multivariate martingales that are, essentially, asymptotically ergodic. Both discrete and continuous time are covered. The expansion is in a test function topology. We investigate when the expansion has the usual Edgeworth form, looking in particular at likelihood inference, including Cox regression, and at inference for stationary time series. The triangular array nature of the results make them useful for bootstrapping, and a result pointing in that direction is shown for Cox regression.

1. Introduction. Edgeworth expansions are an important part of statistics. A computerized search of the *Current Index to Statistics* for the years 1980–1989 reveals 489 entries containing the word "expansion" in the title or as a keyword, and most of these entries concern Edgeworth series in some form. New expansions are being developed, or they are being used for bootstrapping, Cornish–Fisher inversion, conditional inference, optimality calculations and so on. References tackling broad expansion issues are, for example, Wallace (1958), Beran (1987), McCullagh (1987) and Hall (1988). Major recent breakthroughs in expansion theory include Bhattacharya and Ranga Rao (1976), Bhattacharya and Ghosh (1978) and Singh (1981).

A feature of this research area is that the applications are far ahead of the theory. Some very important statistical results, such as those in McCullagh (1984) and Beran (1987), can be derived under the assumption that a relevant Edgeworth series exists and has the usual form, without reference to the underlying data structure, which may range from i.i.d. observations via regression to time- and space-dependent variables. Such results can be much more comprehensive than available results guaranteeing the existence of an expansion.

The discrepancy between expansion results and their applications are the most striking for the case of dependent observations. Results here are quite new, and the theory is still in development. Important references here are

Received August 1990; revised October 1994.

[1] Research supported in part by NSF Grants DMS-89-02667 and DMS-93-05601. This manuscript was prepared using computer facilities supported in part by NSF Grants DMS-89-05292, DMS-87-03942 and DMS-86-01732 awarded to the Department of Statistics at the University of Chicago, and by the University of Chicago Block Fund.

AMS 1991 *subject classifications*. Primary 60F99; secondary 60G42, 60G44, 60G55, 60J05, 62E20, 62F12, 62M09, 62M10.

Key words and phrases. Bootstrapping, Edgeworth expansions, likelihood inference, Markov processes, martingales, Cox regression, time series.

Goetze and Hipp (1983), Malinovskii (1986) and Jensen (1989), working with mixing and Markov conditions. Earlier work includes Nagaev (1957) and Statulevicius (1969, 1970).

This paper also concerns dependent variables, in the form of martingales. We shall investigate the conditions for a multivariate martingale to have a one-step asymptotic expansion, thus generalizing the results in Mykland (1992, 1993). Specifically, we shall consider the behavior of a triangular array $(\ell_t^{N,i})_{0 \le t \le T_N}$, $N = 1, 2, \dots,$ of p-dimensional martingales ($i = 1, \dots, p$). The expansion is in a test function topology.

Martingales are particularly important in inference [see, e.g., Klimko and Nelson (1978), Hall and Heyde (1980), Chapter 6, Andersen and Gill (1982), Andersen and Borgan (1985), Tjøstheim (1986) and Wong (1986)]. For example, they come up in likelihood inference and, more generally, when using martingale estimating equations. The martingale is the first order term in the Taylor expansion relating the estimator to the derivatives of the likelihood or the estimating equation. A one-step expansion for the martingale term can then usually be turned into one for $\hat{\theta}_N - \theta_0$ appropriately normalized (or for other common approximate pivots, such as the signed likelihood ratio statistic). However, we shall not investigate in detail the conditions for this to hold. Such conditions have been studied by Skovgaard (1981a, b) for the case of pointwise convergence, and for our test function topology the conditions typically will be considerably weaker. Some results of this type (studentizing the martingale) are included in Mykland (1992, 1993).

A major question, and one on which many results depend, is when is it true that the expansion has the usual form

$$P\left(N^{-1/2}\ell_{T_N}^{N,i} \le x^i\right) \simeq \Phi(x;\kappa) + N^{1/2}\nu(x), \tag{1.1}$$

where, using the summation convention,

$$d\nu(x) = \phi(x;\kappa)\left\{\tfrac{1}{2}v^{i,j}h_{ij}(x) + \tfrac{1}{6}\kappa^{i,j,k}h_{ijk}(x)\right\}dx. \tag{1.2}$$

We have here used the notation of McCullagh (1987); Φ and ϕ are, respectively, the p-variate normal c.d.f. and density with mean 0 and covariance matrix $\kappa = \kappa^{i,j}$. The matrix $\kappa_{i,j}$ is the inverse of $\kappa^{i,j}$, and the Hermite polynomials are given by

$$\begin{aligned} h_i &= \kappa_{i,j}x^j, \\ h_{ij} &= h_ih_j - \kappa_{i,j}, \\ h_{ijk} &= h_ih_jh_k - h_i\kappa_{j,k}[3], \end{aligned} \tag{1.3}$$

where [3] denotes summation over the three different expressions for $h_i\kappa_{j,k}$ which arise by permuting i, j and k. We shall use similar notation in the rest of the paper.

It turns out that, apart from "lattice problems" (the convergence type is discussed in Section 2.2), (1.1)–(1.2) depends on a niceness condition on the asymptotic behavior of the quadratic variation matrix of the martingale (to be defined in Section 2.1). We shall see that in parametric and partial likelihood problems, the condition can be reexpressed as one on the observed information, and that it is automatically satisfied in regular problems (Section 3). Outside the parametric framework, the conditions can still be verified for broad classes of nonlinear time series (Section 5). As an extended example, we do a case study of Cox regression (Section 4).

It should be emphasized that there are a number of examples where (1.1)–(1.2) does not hold. The example in Section 2.2 provides one illustration of this. Earlier examples of nonstandard expansions can be found in, for example, Woodroofe and Keener (1987) and Mykland [(1992), Section 2.3].

2. The martingale expansion.

2.1. *When does the expansion have its usual form?* The asymptotic behavior of martingales is largely controlled by their quadratic variations. If $(\ell_t^{N,i})_{0 \le t \le T_N}$, $N = 1, 2, \dots,$ is a triangular array of p-dimensional martingales, and if the time axis is discrete ($t = 0, t_1, t_2, \dots, T_N$), one can define two $p \times p$-matrices of quadratic variations in the following way: *optional quadratic variation*,

$$[\ell^{N,i}, \ell^{N,j}]_{T_N} = \sum_u \Delta\ell_{t_u}^{N,i}\, \Delta\ell_{t_u}^{N,j}; \tag{2.1}$$

predictable quadratic variation,

$$\langle \ell^{N,i}, \ell^{N,j}\rangle_{T_N} = \sum_u E\Big(\Delta\ell_{t_u}^{N,i}\, \Delta\ell_{t_u}^{N,j} \mid \mathscr{F}_{t_u}^{(N)}\Big). \tag{2.2}$$

Here, $\Delta\ell_{t_u}^{N,i} = \ell_{t_{u+1}}^{N,i} - \ell_{t_u}^{N,i}$, and $(\mathscr{F}_t^{(N)})$ is a filtration or history with respect to which $(\ell_t^{N,i})$ is a martingale. If the time axis is not discrete, the quadratic variations are obtained by taking limits in the above, subject to $(\ell_t^{N,i})$ having sample paths which are right continuous with left limits ("cadlag"). This can always be arranged [cf. Liptser and Shiryaev (1977), Chapter 2]. For a short and self-contained rigorous development defining (2.1)–(2.2), see Jacod and Shiryaev [(1987), Chapter I]; for an example of quadratic variations for a martingale which is not time-discrete, see the discussion of Cox regression in Section 4.

A law of large numbers for either one of the two quadratic variations is the main condition for a central limit theorem for $(\ell_t^{N,i})$ [cf. Hall and Heyde (1980), Chapters 3 and 6].

For expansion purposes, it turns out to be convenient to consider a combination of the two matrices above, namely, the *mixed quadratic variation*,

$$(\ell^{N,i}, \ell^{N,j})_{T_N} = \tfrac{1}{3}[\ell^{N,i}, \ell^{N,j}]_{T_N} + \tfrac{2}{3}\langle \ell^{N,i}, \ell^{N,j}\rangle_{T_N}. \tag{2.3}$$

The structure of the expansion is determined by a certain asymptotic projection of the $(\ell^{N,i}, \ell^{N,j})_{T_N}$-matrix on the $\ell_{T_N}^{N,i}$-vector. We assume that there exists a Borel-measurable function $\rho^{ij}(z^1, z^2, \dots, z^p)$, symmetric in i and j, so that

$$E_{\mathrm{as}}\left\{\left[N^{1/2}\left(\frac{(\ell^{N,i}, \ell^{N,j})_{T_N}}{N} - \kappa^{i,j}\right) - \rho^{ij}\left(\frac{\ell^{N;}_{T_N}}{\sqrt{N}}\right)\right]\frac{\partial^2 g\left(\ell^{N,\cdot}_{T_N}/\sqrt{N}\right)}{\partial x_i\,\partial x_j}\right\} = 0, \tag{2.4}$$

for all bounded measurable Hessians $\partial^2 g(x)/\partial x_i\ \partial x_j$. The subscript "as" means asymptotically—(2.4) should be taken to mean that $\rho^{i,j}$ is some function so that any subsequence has a subsequence for which

$$N^{-1/2}\ell^{N,k}_{T_N} \to_{\mathscr{L}} Z^k \quad \text{and} \quad N^{1/2}\left(\frac{(\ell^{N,i}, \ell^{N,j})_{T_N}}{N} - \kappa^{i,j}\right) \to_{\mathscr{L}} \xi^{ij} \quad \text{jointly;}$$
$$E\left[\xi^{ij} - \rho^{ij}(Z)\right]\frac{\partial^2 g(Z)}{\partial x_i\,\partial x_j} = 0. \tag{2.5}$$

It is clear from (2.4) that ρ^{ij} only needs to be defined up to the equivalence relation $\sim$, where $\rho^{ij} \sim \tilde{\rho}^{ij}$ if they are both symmetric in i and j, and satisfy

$$E\left[\rho^{ij}(Z) - \tilde{\rho}^{ij}(Z)\right]\frac{\partial^2 g(Z)}{\partial x_i\,\partial x_j} = 0$$

for all bounded Hessians $\partial^2 g/\partial x_i\ \partial x_j$. One can take

$$\rho^{ij} = E_{\mathrm{as}}\left(N^{1/2}\left(\frac{(\ell^{N,i}, \ell^{N,j})_{T_N}}{N} - \kappa^{i,j}\right)\middle|\frac{\ell^{N,\cdot}_{T_N}}{\sqrt{N}}\right),$$

provided the right-hand side is well defined.

Ignoring "lattice problems" and assuming that $\kappa^{i,j}$ is positive definite, the condition for (1.1)–(1.2) is that ρ^{ij} can be taken to be linear:

THE LINEARITY CONDITION:

$$\rho^{ij}(Z) \sim v^{i,j}_k Z^k + v^{i,j}. \tag{2.6}$$

Exact statements to this effect are contained in Theorem 1 and Proposition 2.

EXAMPLE. Consider estimation of θ_i in the model

$$Y_t = \sum_{i=1}^{p} X^i_t \theta_i + \varepsilon_t,$$

where $X^{\cdot}_t$ forms a stationary process and where the ε's are i.i.d., with ε_t independent of $X^{\cdot}_n$, $n \le t$.

If $\hat{\theta}$ is the least squares estimator, then $\hat{\theta}_i - \theta_i = S_{N,ij}\ell^j_N$, where $S_{N,ij}$ is the inverse of $S^{ij}_N = \Sigma^N_{n=1}X^i_n X^j_n$, and where ℓ^j_N is the martingale given by $\ell^j_t = \Sigma^t_{n=1}X^j_n \varepsilon_n$. The quadratic variations of ℓ^j are $[\ell^i, \ell^j]_N = \Sigma^N_{n=1}X^i_n X^j_n \varepsilon^2_n$ and $\langle \ell^i, \ell^j \rangle_N = \mathrm{Var}(\varepsilon)S^{ij}_N$. Weak moment and mixing assumptions [see, e.g., Hall and Heyde (1980), Chapter 5, Chan (1990) and Tjøstheim (1990)] now yield that ℓ^i_N and its quadratic variations are asymptotically jointly normal. Hence p^{ij} can be defined by the asymptotic conditional expectation, and the linearity condition is satisfied.

Conditions for the expansion to hold (cf. Theorem 1 below) can also be verified under mixing assumptions; see Proposition 9 in Section 5.2 for an example of this.

If (2.6) is satisfied, then the quantity appearing in (1.2) in the place of the third cumulant is given by

$$\kappa^{i,j,k} = v^{ij}_{\alpha}\kappa^{\alpha,k}[3]. \tag{2.7}$$

Under conditions which are somewhat stronger than those necessary for the expansion to hold (see Proposition 7), one can show that

$$\kappa^{i,j,k} = \lim_{N\to\infty} \frac{1}{N} E\ell^{N,i}_{T_N}\ell^{N,j}_{T_N}\ell^{N,k}_{T_N}. \tag{2.8}$$

The linearity condition (2.6) is usually satisfied in inference problems (cf. Sections 3–5), but there is no all-embracing guarantee. If it does not hold, then (1.2) is replaced by

$$\begin{aligned} d\nu(x) = \frac{1}{2}\phi(x;\kappa)\bigg\{ & \frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\beta}\rho^{\alpha\beta}(x) - \kappa_{\alpha,\beta}\rho^{\alpha\beta}(x) \\ & + \rho^{\alpha\beta}(x)\big(x^j\kappa_{\alpha,j}\big)\big(x^i\kappa_{i,\beta}\big) \\ & - 2\frac{\partial}{\partial x_\alpha}\rho^{\alpha\beta}(x)\big(x^i\kappa_{i,\beta}\big)\bigg\}\, dx, \end{aligned} \tag{2.9}$$

subject to (some version of) ρ being twice differentiable and not growing more than exponentially as $|x| \to \infty$.

These conditions need not hold either, in which case the signed measure $d\nu(x)$ must be abandoned altogether. The expansion result, however, continues to hold, in a way. We shall now discuss its general form.

2.2. *A more general asymptotic expansion.* The expansion which holds under the fewest assumptions has the form

$$Eg\Big(c_N^{-1/2}\ell^{N,\cdot}_{T_N}\Big) = Eg\big(N(0,\kappa^{i,j})\big) + r_N J(g) + o(r_N), \tag{2.10}$$

where $J(g)$ is a functional of g replacing the signed measured $d\nu(x)$, and c_N and r_N are convergence rates replacing N and $N^{-1/2}$. Instead of (2.4), we let ρ^{ij} be symmetric and satisfy

$$E_{\mathrm{as}}\left\{\left[r_N^{-1}\left(\frac{(\ell^{N,i},\ell^{N,j})_{T_N}}{c_N}-\kappa^{i,j}\right)-\rho^{ij}\left(\frac{\ell^{N,\cdot}_{T_N}}{\sqrt{N}}\right)\right]\frac{\partial^2}{\partial x_i\,\partial x_j}g\left(\frac{\ell^{N,\cdot}_{T_N}}{\sqrt{N}}\right)\right\}=0; \tag{2.11}$$

$J(g)$ is then given by

$$J(g)=\frac{1}{2}E\rho^{ij}(Z)\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}g(Z). \tag{2.12}$$

The expressions (1.2) and (2.9) now follow via integration by parts, subject, in the latter case, to some niceness on the part of ρ^{ij}. If ρ^{ij} is not that nice, the expansion (2.11) is not defined except for twice-differentiable g's. In general, it also only holds for such g's.

We shall denote this kind of test function convergence by $o_2(\cdot)$, and write

$$P\left(c_N^{-1/2}\ell^{N,\cdot}_{T_N}\le x\right)=\Phi(x;\kappa)+r_NJ(\cdot)+o_2(r_N). \tag{2.13}$$

This should be taken to mean that (2.10) holds uniformly in classes $\mathscr{C}$ of functions g that satisfy the following: (i) $\int|\hat{g}(x)|\,dx<\infty$, uniformly in C, and $\{\sum_u x_u^2\hat{g}(x),\ g\in C\}$ is uniformly integrable (here, $\hat{g}$ is the Fourier transform of g, which must exist for each $g\in C$); or (ii) $g(x)=f(z^ix_i)$, with $\sum_i z^iz^i$, f and f'' bounded, uniformly in $\mathscr{C}$, and with $\{f'':\ g\in\mathscr{C}\}$ equicontinuous almost everywhere (under Lebesgue measure).

Case (ii) is important in that it covers one-dimensional martingales and multidimensional characteristic functions.

EXAMPLE. As a case where the signed measure ν does not exist and where one really needs a very weak convergence type, consider the martingale array from Example 5 in Hall and Heyde [(1980), Chapter 3.6, pages 82–83]. Martingales are given by (with some change of notation)

$$\ell_t^N=\sum_{n=1}^{t}X_n^N,$$

where

$$X_n^N=\begin{cases}N^{-1/2}Y_n, & \text{if } 1\le n\le m_N,\\ n^{-1/2}2^{1/2}Y_n\mathbf{I}\left(\ell^N_{m_N}>0\right), & \text{if } m_N<n\le N,\end{cases}$$

the Y_n's being independent $N(0, 1)$, and $m_N = \lfloor N(1 - \delta_N) \rfloor$, where $1 \geq \delta_N \downarrow 0$. In this case,

$$\langle \ell^N, \ell^N \rangle_N = N^{-1} m_N + 2N^{-1}(N - m_N) I\big(\ell^N_{m_N} > 0\big),$$

whence

$$\langle \ell^N, \ell^N \rangle_N - 1 = \delta_N \big(2 I_{(\ell^N > 0)} - 1\big) + o_p(\delta_N).$$

In view of Theorem 1, an expansion [in the sense of (2.13)] exists when r_N^{-1} is $O(\delta_N^{-1})$ and $O(N^{1/2})$, or smaller. In the case where $r_N = \delta_N$ [with $\delta_N = O(N^{-1/2})$ or smaller], Proposition 3 yields that

$$\rho = E_{\text{as}}\big(\delta_N(\langle \ell^N, \ell^N \rangle_N - 1) \mid \ell^N_N\big),$$

that is,

$$\rho(z) = 2\mathbf{I}(z > 0) - 1.$$

Hence the linearity condition (2.6) is not satisfied and ν does not even exist. In the case where $r_N^{-1} = o(\delta_N^{-1})$, ρ is identically zero, so the expansion term is also zero. Even then, however, a pointwise expansion does not exist when $r_N = O(\delta_N^{1/2})$ or smaller [see the discussion in Hall and Heyde (1980)].

Before continuing our discussion of the expansion result, we digress by defining the general kth *order variations at time* t. The optional and predictable versions are, respectively,

$$[\ell^{N,i_1}, \ell^{N,i_2}, \dots, \ell^{N,i_k}]_t = \sum_{t_{u+1} < t} \Delta \ell^{N,i_1}_{t_u} \Delta \ell^{N,i_2}_{t_u} \cdots \Delta \ell^{N,i_k}_{t_u} \tag{2.14}$$

and

$$\langle \ell^{N,i_1}, \ell^{N,i_2}, \dots, \ell^{N,i_k} \rangle_t = \sum_{t_{u+1} < t} E\Big[\Delta \ell^{N,i_1}_{t_u} \Delta \ell^{N,i_2}_{t_u} \cdots \Delta \ell^{N,i_k}_{t_u} \mid \mathscr{F}^{(N)}_{t_u}\Big], \tag{2.15}$$

for discrete time martingales. Equation (2.14) extends to general cadlag martingales by taking the appropriate limit, while $\langle \ell^{N,i_1}, \dots, \ell^{N,i_k} \rangle$ can be defined as the compensator of $[\ell^{N,i_1}, \dots, \ell^{N,i_k}]$ [see, e.g., Jacod and Shiryaev (1987), Chapter I]. Equation (2.14) is clearly also defined for cadlag semimartingales.

We now resume the discussion of when (2.13) holds. It remains to state some integrability conditions.

(A) For each i, there are $\underline{k}_i$, $\overline{k}_i$, $\underline{k}_i < \kappa^{i,i} < \overline{k}_i$, so that

$$r_N^{-1}\left(\frac{(\ell^{N,i}, \ell^{N,i})_{T_N}}{c_N} - \kappa^{i,i}\right)\mathbf{I}\big(\underline{k}_i \leq c_N^{-1}(\ell^{N,i}, \ell^{N,i})_{T_N} \leq \overline{k}_i\big)$$

is uniformly integrable.

(B) For the same $\underline{k}_i$, $\overline{k}_i$,

$$P\left(\underline{k}_i \le c_N^{-1}(\ell^{N,i}, \ell^{N,i})_{T_N} \le \overline{k}_i\right) = 1 - o(r_N)$$

for each i.

(C) For each i,

$$E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} = O\left(c_N^2 r_N^2\right).$$

The expansion result is now as follows.

THEOREM 1. *Let* $(\ell_t^{N,i})_{0 \le t \le T_N}$ *be a triangular array of zero mean cadlag martingales. Assume conditions* (A)–(C) *and that* ρ^{ij} *is well defined. Then* (2.13) *holds.* □

This result is proved in Section 6, along with the following proposition.

PROPOSITION 2. *Assume the conditions of Theorem* 1. *The expansion term has the usual form* (1.2) *if and only if* $\kappa^{i,j}$ *is nonsingular and the linearity condition* (2.6) *is satisfied. In this case, if* $\kappa^{i,j,k}$ *is required to be symmetric,* (2.7) *holds.*

A number of issues related to Theorem 1 are discussed in the context of the less general expansions in Mykland (1989, 1992, 1993). This includes studentization, confidence intervals and a putting into context of the test function convergence.

The conditions for the expansion can be formulated in terms of the optional or predictable quadratic variations. If $\langle \ell^{N,i}, \ell^{N,j}, \ell^{N,k} \rangle_{T_N}/c_N^{3/2} r_N \to_P \eta^{i,j,k}$ (a constant) for all combinations of indices, then, typically,

$$\frac{[\ell^{N,i}, \ell^{N,j}]_{T_N} - \langle \ell^{N,i}, \ell^{N,j} \rangle_{T_N}}{c_N r_N} = \eta^{i,j,k} \kappa_{k,\alpha} c_N^{-1/2} \ell_{T_N}^{N,\alpha} + U_N, \tag{2.16}$$

where $U_N = O_p(1)$ and, asymptotically, has zero expectation given $c_N^{-1/2} \ell_{T_N}^{N,\cdot}$. Hence, we get the following formulas which relate to the linearity condition (2.6):

$$\frac{(\ell^{N,i}, \ell^{N,j})_{T_N} - [\ell^{N,i}, \ell^{N,j}]_{T_N}}{c_N r_N} = -\frac{2}{3} \eta^{i,j,k} \kappa_{k,\alpha} c_N^{-1/2} \ell_{T_N}^{N,\alpha} - \frac{2}{3} U_N \tag{2.17}$$

and

$$\frac{(\ell^{N,i}, \ell^{N,j})_{T_N} - \langle \ell^{N,i}, \ell^{N,j} \rangle_{T_N}}{c_N r_N} = \frac{1}{3} \eta^{i,j,k} \kappa_{k,\alpha} c_N^{-1/2} \ell_{T_N}^{N,\alpha} + \frac{1}{3} U_N. \tag{2.18}$$

Rigorous conditions under which (2.16)–(2.18) hold can be deduced from Hall and Heyde [(1980), Theorem 3.1 and Corollary 3.2]. Alternatively one can use the following.

PROPOSITION 3. *Suppose that condition* (C) *of Theorem* 1 *holds. Then conditions* (A) *and* (B) *on* $(\ell^{N,i}, \ell^{N,i})_{T_N}$ *are equivalent to the same conditions imposed on either* $[\ell^{N,i}, \ell^{N,i}]_{T_N}$ *or* $\langle \ell^{N,i}, \ell^{N,i} \rangle_{T_N}$. *Suppose furthermore that* $\langle \ell^{N,i}, \ell^{N,j}, \ell^{N,k} \rangle_{uT_N}/c_N^{3/2} r_N$ *converges in probability to a constant for each* $u \in [0,1]$, *with* $\langle \ell^{N,i}, \ell^{N,j}, \ell^{N,k} \rangle_{T_N}/c_N^{3/2} r_N \to_P \eta^{i,j,k}$. *Then, if* $\kappa^{i,j}$ *is nonsingular*, (2.16)–(2.18) *is valid.*

In the above result, if uT_N is not in the time axis for $\langle \ell^{N,i}, \ell^{N,j}, \ell^{N,k} \rangle$, take the closest value to (or the limit on) the right.

2.3. *Average coverage probabilities.* The o_2 topology is necessary for the expansion to hold in general. However, it is not always a necessary evil, as it can be used in connection with average coverage probabilities [cf. Woodroofe (1989)]. Unlike this paper, however, the expansions with which we are dealing here appear to lend themselves to average coverage probabilities in contiguity neighborhoods. (The paper cited discusses average coverage in a global sense.)

To see what happens, we shall focus on a multivariate (2, 1) exponential family, that is,

$$\frac{dP_{\theta+\delta}}{dP_\theta} = \exp\left(\delta' \ell_t - \frac{1}{2} \delta' \langle \ell, \ell \rangle_t \delta \right), \tag{2.19}$$

where ℓ_t is a martingale under P_θ. (In this section, we shall mostly use matrix notation, as we shall be taking square roots of matrices—always the symmetric square root; κ will denote the $\kappa^{i,j}$ matrix.) This is, for example, the model in Woodroofe (1989), and it is also the model for inference for the drift in stochastic differential equations [see, e.g., Lipster and Shiryaev (1977, 1978), Basawa and Prakasa Rao (1980), Chapter 9.5, and Kutoyants (1984)]. The extension to more general models is fairly straightforward.

In this setup, suppose one is interested in average coverage in a contiguity neighborhood of θ, specifically,

$$\bar{\gamma}_c = \int P_{\theta + \delta t^{-1/2}} \big(\langle \ell, \ell \rangle_t^{1/2} (\hat{\theta} - \theta - \delta t^{-1/2}) \in \mathscr{C} \big) \zeta(\delta) \, d\delta, \tag{2.20}$$

where ζ is a probability density. Note that

$$\begin{aligned} &\det\left(\frac{\langle \ell, \ell \rangle_t}{t} \right)^{-1/2} - \det(\kappa)^{-1/2} \\ &\qquad = -\frac{1}{2} \frac{1}{\sqrt{t}} (\det \kappa)^{-1/2} \operatorname{trace}(\kappa^{-1} \xi) + o_p\left(\frac{1}{\sqrt{t}} \right) \end{aligned} \tag{2.21}$$

[see, e.g., McCullagh (1987), Example 1.16, page 21] and

$$\frac{\langle \ell, \ell \rangle_t^{1/2}}{\sqrt{t}} - \kappa^{1/2} = \frac{1}{2} t^{-1/2} \xi \kappa^{-1/2} + o_p(t^{-1/2}). \tag{2.22}$$

Hence, under reasonable conditions on ζ, Taylor expansion yields

$$\bar{\gamma}_c = E_\theta g_c(B_t^{-1}\ell_t) + t^{-1/2}R_c + o(t^{-1/2}),$$

where

$$\begin{aligned} g_c(x) &= \int I(x - \kappa^{1/2}\lambda \in \mathscr{C})\exp(\lambda'\kappa^{1/2}x - \tfrac{1}{2}\lambda'\kappa\lambda)\zeta(\lambda)\,d\lambda \\ &= \int_{\mathscr{C}} \exp(\tfrac{1}{2}(x'x - y'y))\zeta(\kappa^{-1/2}(x-y))\det(\kappa)^{-1/2}\,dy \end{aligned} \tag{2.23}$$

and

$$\begin{aligned} R_c &= -\tfrac{1}{2}E\int I(\kappa^{-1/2}Z - \lambda^{-1/2}\lambda \in \mathscr{C})\exp(\lambda'Z - \tfrac{1}{2}\lambda'\kappa\lambda) \\ &\qquad \times\{\zeta(\lambda)\operatorname{trace}(\kappa^{-1}\xi) + \nabla\zeta(\lambda)\xi\kappa^{-1}\lambda\}\,d\lambda \\ &= -\tfrac{1}{2}E\int I(V \in \mathscr{C})\{\zeta(\lambda)\operatorname{trace}(\kappa^{-1}\rho(\kappa^{1/2}V + \kappa\lambda)) \\ &\qquad\qquad + \nabla\zeta(\lambda)\kappa^{-1}\rho(\kappa^{1/2}V + \kappa\lambda)\lambda\}\,d\lambda, \end{aligned} \tag{2.24}$$

where $V \sim \mathrm{N}(0, I)$. In many instances g_c will now satisfy the regularity conditions imposed on the test functions. For example, this is the case if $\mathscr{C}$ is compact and ζ is compactly supported. To get the final result, one Taylor-expands $E_\theta g_c(\langle \ell, \ell\rangle_t^{-1/2}\ell_t)$ around $E_\theta g_c(t^{-1/2}\kappa^{-1/2}\ell_t)$,

$$\begin{aligned} &Eg_c(\langle \ell, \ell\rangle_t^{-1/2}\ell_t) - Eg_c(t^{-1/2}\kappa^{-1/2}\ell_t) \\ &\qquad = -t^{-1/2}\tfrac{1}{2}E\,\nabla g_c(V)\kappa^{-1}\rho(\kappa^{1/2}V)V + o(t^{-1/2}), \end{aligned} \tag{2.25}$$

and then uses Theorem 1 on $Eg_c(t^{-1/2}\kappa^{-1/2}\ell_t)$. Note, incidentally, that conditions (A) and (B) above assure the uniform integrabilities needed.

3. The case of likelihood inference. The structure provided by likelihood estimation makes it possible to relate when our conditions hold to more standard quantities and criteria. This is true both for parametric inference and for partial likelihood in the sense of Andersen and Gill (1982) and Wong (1986). First of all, the requirements imposed on the quadratic variation can be rephrased as ones on the observed information (Section 3.1). Also, contiguity arguments show that the linearity condition (2.6) is satisfied in regular problems and that, typically, the convergence rates are as for i.i.d. observations (Section 3.2). Thus, in the $o_2(N^{-1/2})$ sense, the usual Edgeworth expansion (1.1)–(1.2) continues to hold in fairly great generality. It follows that the same is true for the standard expansions for $\sigma_N^{ij}(\hat{\theta}_{Ni} - \theta_{0,i})$ [where $\sigma_N^{ij} = \sqrt{N}\,\delta^{ij}$ (Kronecker δ) or is the square root of the expected or observed information] and for the signed likelihood ratio statistic. In Section 3.3 we discuss whether this extends to conditional properties, and there it turns out that the picture is more mixed.

The purpose of this section is to see what can be shown without imposing conditions on the dependence structure such as stationarity, mixing or geometric ergodicity. The implications of such assumptions are discussed in Section 5.

Apart from in theorems, we work with fairly regular situations in the sense that $r_N = c_N^{-1/2} = N^{-1/2}$, the parameter θ_0 does not change with N, $\hat{\theta}_N - \theta_0 = O_p(N^{-1/2})$ and the log-likelihood $L_t^N(\theta)$ has as many derivatives as required, denoted by $L_t^{N,i}(\theta) = \partial L_t^N(\theta)/\partial\theta^i$, $L_t^{N,ij}(\theta) = \partial^2 L_t^N(\theta)/\partial\theta^i\,\partial\theta^j$ and so on.

3.1. *Square variation and observed information.* The connection between the quadratic variation and the observed information is given by the Bartlett identities for moments [see, e.g., McCullagh (1987), page 202]. In their conditional version, the second and third are

$$E\left(\Delta L_t^{N,ij} + \Delta L_t^{N,i}\,\Delta L_t^{N,j} \mid \mathscr{F}_t^{(N)}\right) = 0 \tag{3.1}$$

and

$$E\left(\Delta L_t^{N,ijk} + \Delta L_t^{N,i}\,\Delta L_t^{N,jk}[3] + \Delta L_t^{N,i}\,\Delta L_t^{N,j}\,\Delta L_t^{N,k} \mid \mathscr{F}_t^{(N)}\right) = 0, \tag{3.2}$$

where Δ refers to any time increment from t on. All quantities are evaluated at θ_0. From (3.1) it now follows that

$$m_t^{N,ij} = L_t^{N,ij} + (L^{N,i}, L^{N,j})_t \tag{3.3}$$

is a martingale, and (3.2) states that

$$\langle m^{N,ij}, L^{N,k}\rangle_t[3] + L_t^{N,ijk} \tag{3.4}$$

is a martingale. Hence, typically, if

$$\frac{L_{T_N}^{N,ijk}}{N} \to_P \lambda^{ijk} \quad \text{(a constant)} \tag{3.5}$$

and if there is a symmetric $\bar{\rho}^{ij}$ satisfying

$$E_{\mathrm{as}}\left\{\left[N^{1/2}\left(\frac{L_{T_N}^{N,ij}}{N} + \kappa^{i,j}\right) - \bar{\rho}^{ij}\left(\frac{L_{T_N}^{N,\cdot}}{\sqrt{N}}\right)\right]\frac{\partial^2}{\partial x_i\,\partial x_j} g\left(\frac{L_{T_N}^{N,\cdot}}{\sqrt{N}}\right)\right\} = 0, \tag{3.6}$$

then ρ^{ij} is defined, and one can take

$$\rho^{ij}(Z) = -\bar{\rho}^{ij}(Z) - \tfrac{1}{3}\lambda^{ijk} h_k(Z). \tag{3.7}$$

A diversity of regularity conditions can assure this. The following is an example.

PROPOSITION 4. *Assume the following: that the first four Bartlett identities hold; that κ^{ij} is nonsingular; that $L_{uT_N}^{N,ijk}/N$ converges in probability to a constant for all $u \in [0, 1]$ and for each i, j, k; and that, for each i, $EL_{T_N}^{N,ii} = O(N)$, $E[L^{N,iii}, L^{N,iii}]_{T_N} = O(N)$ and $EL_{T_N}^{N,iiii} = O(N)$. Then $N^{1/2}(L_{T_N}^{N,ij}/N + \kappa^{i,j})$ can replace $N^{1/2}((L^{N,i}, L^{N,j})_{T_N}/N - \kappa^{i,j})$ in the conditions of Theorem* 1, *and the expansion is given by* (3.7) *and* (2.12) [*or* (1.1) *and* (2.6)–(2.7) *or* (2.9), *as the case may be*].

As in Proposition 3, if uT_N is not in the time axis for $L^{N,iii}$, take the closest value to (or the limit on) the right.

3.2. *The linearity condition in regular problems.* Condition (2.6) can be verified with the Hájek–Le Cam convolution theorem. Set $Q_N^{ij}(\theta) = (L^{N,i}(\theta), L^{N,j}(\theta))_{T_N}^{\theta}/N - \kappa^{i,j}(\theta)$, where the θ's on $\kappa^{i,j}$ and the quadratic variation indicate that the relevant expectations are taken with respect to P_θ. Under conditions set out in Le Cam (1972), Roussas (1972) or Millar (1985), the convolution theorem now yields that $N^{1/2}Q_N^{ij}(\hat{\theta}_N)$ is asymptotically independent of $N^{1/2}(\hat{\theta}_N - \theta_0)$. If $Q_N^{ijk}(\theta_N^*)/N \to_P w^{ijk}$ (a constant) as $N \to \infty$ whenever $(\theta_N^* - \theta_0) = O_p(N^{-1/2})$, then a Taylor expansion gives that

$$(3.8)\quad N^{1/2}Q_N^{ij}(\theta_0) = N^{1/2}Q_N^{ij}(\hat{\theta}_N) - N^{1/2}(\hat{\theta}_{N,k} - \theta_{0,k})w^{ijk} + o_p(1).$$

The linearity condition is, therefore, satisfied.

Another approach is to take the previous section as a point of departure and to study the linearity condition for the observed information. As an illustration, we here state a formal result. The subscript "as" is as defined after equation (2.4); "$\theta_{N,\mathrm{as}}$" means that the relevant limits in law are under the sequence P_{θ_N}.

PROPOSITION 5. *Suppose that $\kappa^{i,j}(\theta)$ is differentiable and nonsingular at θ_0 and that $L_{T_N}^N(\theta)$ is thrice continuously differentiable in a neighborhood of θ_0. Also assume that, whenever $\theta_N - \theta_0 = O(c_N^{-1/2})$, it holds that $L^{N,ijk}(\theta_N)/c_N = \lambda^{ijk} + o_p(1)$, $L^{N,ij}(\theta_N)/c_N = -\kappa^{i,j}(\theta_N) + O_p(r_N)$ and*

$$E_{\theta_N,\mathrm{as}}\left[r_N^{-1}\left(\frac{L^{N,ij}(\theta_N)}{c_N} + \kappa^{i,j}(\theta_N)\right)\right] = E_{\theta,\mathrm{as}}\left[r_N^{-1}\left(\frac{L^{N,ij}(\theta_0)}{c_N} + \kappa^{i,j}(\theta_0)\right)\right].$$

Then $\bar{\rho}^{ij}$ from (3.6), provided it exists, satisfies the linearity condition (2.6). Also, unless $0 < \limsup_{N\to\infty}|r_N c_N^{1/2}| < \infty$, $\bar{\rho}^{ij} \equiv 0$ (a.s.).

It follows that, if necessary by rescaling, we can take $r_N = N^{-1/2}$, $c_N = N$, whence the expansion gets the form (1.1)–(1.2). Also note that, even when the conditions of the proposition do not hold pointwise (in θ_0), they typically hold almost everywhere under Lebesgue measure [cf. Le Cam (1986), Chapter 8].

Additional conditions and results of this sort are stated in Mykland (1992) for continuous (2, 1) exponential families, and these results generalize under appropriate assumptions.

3.3. *The question of conditional inference.* We shall here concentrate on the one-parameter case. Further pursuing (3.4) yields that, under weak regularity conditions, the signed square root of the likelihood ratio

$$W = \mathrm{sgn}(\hat{\theta}_N - \theta_0)\left\{2\left(\Lambda_{T_N}^N(\hat{\theta}_N) - \Lambda_{T_N}^N(\theta_0)\right)\right\}^{1/2}$$

has the same one-step expansion as $L_{T_N}^{N,1}(\theta_0)/(L^{N,1}(\theta_0), L^{N,1}(\theta_0))_{T_N}^{1/2}$, thus validating its usual unconditional convergence behavior up to $o_2(N^{-1/2})$.

As a major point about W is its conditional behavior, it would be nice if we could verify that, too. This poses some serious problems, however.

McCullagh (1984) shows that W is second order independent of all second order ancillaries, provided a joint expansion exists for W and the candidates for second order ancillaries. In particular, the most natural local ancillaries, such as those discussed in Efron and Hinkley (1978) and Cox (1980), require a one-step expansion for the observed information (jointly with $L^{N,1}$) to show their second order ancillarity. At the time of writing, it looks as if this cannot be done purely with martingales, however. One possibility is that one would need to use, for example mixing, Markov or ARMA process assumptions to get such an expansion. We are still investigating this question.

This raises the question of whether one can show the existence of any second order ancillary under the weaker conditions with which we are dealing. The answer is, typically, in the affirmative, but the resulting local ancillaries have less statistical meaning than those proposed in Efron and Hinkley (1978) and Cox (1980). Such ancillaries can be obtained by considering statistics of the form $\nu^N_{T_N}(\hat{\theta}_N)$, where $(\nu^N_t(\theta))_{0 \le t \le T_N}$ is a martingale under P_θ for each θ. These can be the martingales falling out of the Bartlett identities [i.e., (3.3) etc.] or they can be created via a Girsanov-type construction [see, e.g., Jacod and Shiryaev (1987), Theorem I-3.11, page 155]; $\nu^N_{T_N}(\hat{\theta}_N)$ can be approximated by

$$\nu^N_{T_N}(\theta_0) + \big(\hat{\theta}_N - \theta_0\big)\Big(\nu^{N,1}_{T_N}(\theta_0) - \big(\nu^N(\theta_0), L^{N,1}(\theta_0)\big)_{T_N}\Big)$$
$$+\big(\hat{\theta}_N - \theta_0\big)\big(\nu^N(\theta_0), L^{N,1}(\theta_0)\big)_{T_N} + \big(\hat{\theta}_N - \theta_0\big)^2 \frac{\nu^{N,11}_{T_N}(\theta_0)}{2},$$

and thus fitted into the martingale expansion framework. Second order ancillaries can then be found by the method of McCullagh (1984).

4. Cox regression: a case study. It is worthwhile to look in particular depth at the Cox model for censored survival data [see, e.g., Cox (1972), Andersen and Gill (1982) and Andersen and Borgan (1985)]. This is both because of its wide use and because it presents a somewhat nonstandard application of the martingale expansions. Features of the latter are that, typically, $T_N = 1$ (or, more generally, a constant independent of N) and that the optional and predictable variations do not have the discrete forms (2.1)–(2.2) and (2.14)–(2.15).

We work with the form of the model studied in Andersen and Gill (1982). The point processes $N_{1,t}, \dots, N_{N,t}$ are defined on [0, 1]; $N_{i,t}$ has intensity

$$\lambda_{i,t} = Y^i_t \lambda_{0,t} \exp\big\{\theta_{0,j} Z^{i,j}_t\big\}, \tag{4.1}$$

where (Y^i_t) is predictable and is 0 or 1 according to whether the ith process is under observation, $Z^{i,j}_t$ is a predictable covariate process, $\theta_{0,j}$ are parameters to be estimated and $\lambda_{0,t}$ is a baseline intensity. The $N_{i,t}$'s are assumed orthogonal in the sense that two of them do not jump at the same time. Associated with N_i is the martingale

$$M_{i,t} = N_{i,t} - \int_0^t \lambda_{i,s}\, ds. \tag{4.2}$$

In the notation of Section 3, the derivative of the partial log-likelihood at the true parameter θ_0 is

$$L_t^k = \int_0^t Z_s^{i,k}\, dM_{i,s} - \int_0^t Z_s^{i,k} \lambda_{i,s} \bar{\lambda}_s^+\, d\bar{M}_s, \tag{4.3}$$

where $x^+ = 0$ if $x = 0$ and x^{-1} otherwise, and where $\bar{x} = \sum_{i=1}^N x_i$.

All of the above is in Andersen and Gill (1982), and a more detailed description can be found there.

An important consequence of the orthogonality of the N_i's is that

$$\left[M_{i_1}, \dots, M_{i_k}\right]_t = \delta_{i_1 i_2 \cdots i_k} N_{i_1,t} \tag{4.4}$$

and

$$\langle M_{i_1}, \dots, M_{i_k}\rangle_t = \delta_{i_1 i_2 \cdots i_k} \int_0^t \lambda_{i_1,s}\, ds, \tag{4.5}$$

$\delta_{i_1 \cdots i_k}$ being a Kronecker δ. From this one can calculate the "variations" of L_t^k. The first couple of them are as follows:

$$\begin{aligned} [L^i, L^j]_t = {} & \int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j}\, dN_{\alpha,s} + \int_0^t \left(Z_s^{\alpha,i}\lambda_{\alpha,s}\bar{\lambda}_s^+\right)\left(Z_s^{\beta,j}\lambda_{\beta,s}\bar{\lambda}_s^+\right) d\bar{N}_s \\ & -[2]\int_0^t Z_s^{\alpha,i}\left(Z_x^{\beta,j}\lambda_{\beta,s}\lambda_s^+\right) dN_{\alpha,s}, \end{aligned} \tag{4.6}$$

$$\begin{aligned} \langle L^i, L^j\rangle_t = {} & \int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j}\lambda_{\alpha,s}\, ds \\ & - \int_0^t \left(Z_s^{\alpha,i}\lambda_{\alpha,s}\bar{\lambda}_s^+\right)\left(Z_s^{\beta,j}\lambda_{\beta,s}\bar{\lambda}_s^+\right)\bar{\lambda}_s\, ds; \end{aligned} \tag{4.7}$$

$$\begin{aligned} [L^i, L^j, L^k]_t = {} & \int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j} Z_s^{\alpha,k}\, dN_{\alpha,s} \\ & - \int_0^t \left(Z_s^{\alpha,i}\lambda_{\alpha,s}\bar{\lambda}_s^+\right)\left(Z_s^{\beta,j}\lambda_{\beta,s}\bar{\lambda}_s^+\right)\left(Z_s^{\gamma,k}\lambda_{\gamma,s}\bar{\lambda}_s^+\right) d\bar{N}_s \\ & +[3]\int_0^t Z_s^{\alpha,i}\left(Z_s^{\beta,j}\lambda_{\beta,s}\bar{\lambda}_s^+\right)\left(Z_s^{\gamma,k}\lambda_{\gamma,s}\bar{\lambda}_s^+\right) dN_{\alpha,s} \\ & -[3]\int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j}\left(Z_s^{\beta,k}\lambda_{\beta,s}\bar{\lambda}_s^+\right) dN_{\alpha,s}, \end{aligned} \tag{4.8}$$

$$\langle L^i, L^j, L^k \rangle_t = \int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j} Z_s^{\alpha,k} \lambda_{\alpha,s}\, ds$$

$$(4.9) \qquad - [3] \int_0^t Z_s^{\alpha,i} Z_s^{\alpha,j} \lambda_{\alpha,s} \left(Z_s^{\beta,k} \lambda_{\beta,s} \overline{\lambda}_s^{+} \right) ds$$

$$+ 2 \int_0^t \left(Z^{\alpha,i} \lambda_{\alpha,s} \overline{\lambda}_s^{+} \right) \left(Z^{\beta,j} \lambda_{\beta,s} \overline{\lambda}_s^{+} \right) \left(Z^{\gamma,k} \lambda_{\gamma,s} \overline{\lambda}_s^{+} \right) \overline{\lambda}_s\, ds.$$

Also,

$$(4.10) \quad E[L^i, L^i, L^i, L^i]_1 \le 16 E \int_0^1 \left[\left(Z_s^{\alpha,i} \right)^4 \lambda_{\alpha,s} + \left(Z_s^{\alpha,i} \lambda_{\alpha,s} \overline{\lambda}_s^{+} \right)^{4-} \overline{\lambda}_s \right] ds.$$

These quantities can then be fed into the expansion theorem under various scenarios. As an example, we shall show a result in a particularly simple situation: each N_i only jumps once, the covariates are bounded and the patients are i.i.d. The result is on a triangular array form, making it suitable for use also in bootstrapping.

PROPOSITION 6. *Suppose that* $(N_{i,t}, Y_t^i, Z_t^{i,j}, j = 1, \dots, p)_{0 \le t \le 1}$ *are i.i.d.* P^N *for* $i = 1, \dots, N$, *with* $(Z_t^{i,j})$ *bounded (uniformly in* N*) and* Y_t^i *becoming* 0 *no later than when* $N_{i,t}$ *jumps. Suppose further that there are* $\kappa^{i,j}$ *and* $v^{i,j}$ *so that*

$$(4.11) \quad \begin{aligned} &\int_0^1 E^N Z_s^{1,i} Z_s^{1,j} \lambda_{1,s}\, ds - \int_0^1 E^N \left(Z_s^{1,i} \lambda_{1,s} \right) E^N \left(Z_s^{1,j} \lambda_{1,s} \right) \left(E \lambda_{1,s} \right)^{+} ds \\ &\quad = \kappa^{i,j} + N^{-1/2} v^{i,j} + o(N^{-1/2}) \end{aligned}$$

and that there are $\kappa^{i,j,k}$ *so that*

$$(4.12) \qquad E^N A^i A^j A^k = \kappa^{i,j,k} + o(1),$$

where

$$(4.13) \qquad A^i = \int_0^1 \left[Z_s^{1,i} - E^N \left(Z_s^{1,i} \lambda_{1,s} \right) \left(E^N \lambda_{1,s} \right)^{+} \right] dM_{1,s}.$$

Then the conditions of Theorem 1 *are satisfied, as is the linearity condition* (2.6); $\kappa^{i,j}$, $v^{i,j}$ *and* $\kappa^{i,j,k}$ *are as given by* (4.11)–(4.12).

5. General martingale inference. Outside the likelihood case, estimators can still often be approximated by martingales, and finding a one-step expansion for $\hat{\theta}_N - \theta_0$ (suitably normalized) then reduces by a Taylor argument to finding an expansion for a martingale. This is generally true when using martingale estimating equations, estimating θ_0 by a $\hat{\theta}_N$ satisfying $L_N^i(\hat{\theta}_N) = 0$, $i = 1, \dots, p$, where the $L_N^i(\theta_0)$'s are martingales. A major non-likelihood example of this is the conditional least squares procedure of

Klimko and Nelson (1978) [see also Hall and Heyde (1980), Tjøstheim (1986) and Tong (1990)]; $\hat{\theta}_N$ is here the value minimizing

$$L_N(\theta) = \sum_{s=0}^{N-1} (X_{s+1} - E_\theta(X_{s+1} \mid \mathscr{F}_s))^2. \tag{5.1}$$

In the following, we shall deal only with $L_{T_N}^{N,i}(\theta)$ at the true value θ_0, and for this reason we refer to the martingale by the generic $\ell_{T_N}^{N,i}$.

5.1. *The form of the cumulant.* In Section 2 we introduced the quantity $\kappa^{i,j,k}$ as deriving from a regression. We here tie it to a true cumulant.

PROPOSITION 7. *Let $(\ell_{T_N}^{N,i})$ be a p-dimensional triangular array of zero-mean martingales. Assume that, for each i, $E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} = O(N)$ and $E\{\sqrt{N}[(\ell^{N,i}, \ell^{N,i})_{T_N}/N - \kappa^{i,i}]\}^{1.5+\delta} = O(1)$, for some $\delta > 0$, and that the linearity condition (2.6) is satisfied. Then*

$$\kappa^{i,j,k} = \lim_{N\to\infty} \frac{1}{N} E\ell_{T_N}^{N,i}\ell_{T_N}^{N,j}\ell_{T_N}^{N,k}. \tag{5.2}$$

The result remains valid if $O(N)$, $\sqrt{N}$, $/N$ and $1/N$ are replaced by, respectively, $O(c_N^2 r_N^2)$, r_N^{-1}, c_N^{-1} and $(c_N^{3/2} r_N)^{-1}$.

Under stationarity, we can reexpress (5.2) as follows.

PROPOSITION 8. *Under the assumptions of Proposition 7, supposing that we are in a discrete time and nontriangular array setting, and that $\Delta\ell_t = \ell_{t+1} - \ell_t$ is a stationary process, then*

$$\kappa^{i,j,k} = E\,\Delta\ell_0^i\,\Delta\ell_0^j\,\Delta\ell_0^k + [3]\sum_{t=1}^{\infty} E\,\Delta\ell_0^k\big(\Delta\ell_t^i\,\Delta_t^j - \kappa^{i,j}\big) \tag{5.3}$$

with

$$\kappa^{i,j} = E\,\Delta\ell_0^i\,\Delta\ell_0^j. \tag{5.4}$$

5.2. *Expansions for stationary processes.* Estimating parameters from a stationary process typically leads the martingale $L_N^i(\theta_0)$ to have stationary increments. In this case, we can use a whole arsenal of theory to show that this martingale has an expansion.

As an illustration, we treat the case where the process is strongly mixing. Let $\mathscr{F}^t$ be the σ-field representing the future of the process from t to $+\infty$, and set

$$\alpha(n) = \sup_{A\in\mathscr{F}_t,\, B\in\mathscr{F}^{t+n}} |P(A\cap B) - P(A)P(B)|. \tag{5.5}$$

If

$$\sum_{n=1}^{\infty} \alpha(n)^{\delta/(1+\delta)} < \infty, \tag{5.6}$$

for some δ, $0 < \delta < \infty$, it is sufficient for the expansion to exist that

$$E|\Delta \ell_0^i|^{4+\delta} < \infty, \tag{5.7}$$

for each i.

This is stated formally in the following result.

PROPOSITION 9. *Suppose that we are in a discrete time and nontriangular array setting, and that $\Delta\ell_t = \ell_{t+1} - \ell_t$ is a stationary process, satisfying conditions* (5.6) *and* (5.7). *Then the conditions of Theorem* 1 *and the linearity condition* (2.6) *are satisfied*; $v^{i,j} = 0$; *and* $\kappa^{i,j,k}$ *and* $\kappa^{i,j}$ *are given by* (5.3)–(5.4). *If* $\kappa^{i,j}$ *is nonsingular, the expansion has the form* (1.1)–(1.2).

6. Proofs. In the proofs of Propositions 4, 6, 8 and 9, $c_N = N$, and when proving Proposition 5, c_N can be anything. Otherwise, we assume, without loss of generality, that $c_N = 1$.

It is convenient to divide the proof of Theorem 1 into two parts. We begin with the following lemma.

LEMMA 1. *Theorem* 1 *holds for martingales with time axis* $t = 0, 1, 2, \dots, T_N$.

PROOF. If alternative (ii) is used in the definition of $o_2(r_N)$-convergence, the result follows from Mykland [(1993), Theorem 1] by considering the martingale $(z_i \ell_t^{N,i})_{0 \le t \le T_N}$.

In particular, the result holds for the characteristic function

$$h_N(x) = E\exp\left\{ix_j \ell_{T_N}^{N,j}\right\}.$$

Hence, if we use alternative (i) in the definition of $o_2(r_N)$-convergence we can use that, if g_N is a convergent sequence in $\mathscr{C}$,

$$Eg_N\left(\ell_{T_N}^{N,j}\right) = \frac{1}{(2\pi)^p}\int \hat{g}_N(x) h_N(x)\, dx,$$

where $\hat{g}_N$ is the Fourier transform of $g_N(-x)$, and we are finished if we show dominated convergence for

$$r_N^{-1}|\hat{g}_N(x)|\,|h_N(x) - \exp\{-\tfrac{1}{2}x_j\kappa^{j,k}x_k\}|.$$

Embed (ℓ_t^N) in a continuous p-dimensional martingale $(\bar{\ell}_t^N)$. By Mykland [(1992), Lemma 5.1],

$$\begin{aligned}|h_N(x) - \exp\{-\tfrac{1}{2}x_j\kappa^{j,k}x_k\}| &\le \tfrac{1}{2}E|x_j x_k\left(\langle \bar{\ell}^{N,j}, \bar{\ell}^{N,k}\rangle_{T_N} - \kappa^{j,k}\right)| \\ &\le \tfrac{1}{2}\sum_j x_j^2 \sum_k E|\langle \bar{\ell}^{N,k}, \bar{\ell}^{N,k}\rangle_{T_N} - \kappa^{k,k}|.\end{aligned}$$

In view of the uniform integrability of $\sum_j x_j^2 \hat{g}_N(x)$, the result now follows if we can show that $r_N^{-1}E|\langle \bar{\ell}^{N,k}, \bar{\ell}^{N,k}\rangle_{T_N} - \kappa^{k,k}|$ is bounded. By Mykland [(1993), Lemma 1], this reduces to showing that $r_N^{-1}E|\langle \ell^{N,k}, \ell^{N,k}\rangle_{T_N} - \kappa^{k,k}|$ is bounded.

This is not necessarily true with our original definitions, but it is true if we assume that $\underline{k}_j \le \langle \ell^{N,j}, \ell^{N,j}\rangle_{T_N} \le \bar{k}_j$. We can make this assumption, however, without loss of generality, by considering the martingales $\tilde{\ell}_t^{N,j}$ given by

$$\tilde{\ell}_t^{N,j} = \begin{cases} \ell_t^{N,j} & \text{for } t \le T_N, \\ \ell_{T_N}^{N,j}, & \text{for } t > T_N \text{ and } \langle \ell^{N,j}, \ell^{N,j}\rangle_{T_N} \ge \underline{k}_j, \\ \ell_{T_N}^{N,j} + \sum_{s=T_N+1}^{t} \varepsilon_s^j, & \text{otherwise,} \end{cases}$$

where the ε_s^j are i.i.d. standard normal and independent of the $(\ell_t^{N,k})$'s. If we set

$$\tau_N = \inf\left\{t: \langle \tilde{\ell}^{N,j}, \tilde{\ell}^{N,j}\rangle_{t+1} - \kappa^{j,j} > \bar{k}_j \mid \text{for some } j\right\} \tag{6.1}$$

[this is a stopping time by Jacod (1979), Chapter X-1a], then the martingale $\tilde{\tilde{\ell}}_t^{N,j} = \tilde{\ell}_{t\wedge\tau_N}^{N,j}$ satisfies that $\underline{k}_j \le \langle \tilde{\tilde{\ell}}^{N,j}, \tilde{\tilde{\ell}}^{N,j}\rangle_{T_N+K} - \kappa^{j,j} \le \bar{k}_j$ if $K > \max(\underline{k}_j)$, and $\tilde{\tilde{\ell}}_{T_N+K}^{N,j}$ can replace $\ell_{T_N}^{N,j}$ in view of Proposition 3 and a Taylor argument. □

PROOF OF THEOREM 1. Assume in the following that, for some $\bar{k}$,

$$[\ell^{N,i}, \ell^{N,i}]_{T_N} \le \bar{k} + |\Delta \ell_{T_N}^{N,i}|^2, \tag{6.2}$$

where $\Delta \ell_{T_N}^{N,i}$ is the jump of $\ell_t^{N,i}$ at time T_N. This can be done without loss of generality, by using the same kind of argument as that surrounding (6.1). A consequence of (6.2) is that

$$[\ell^{N,i}, \ell^{N,i}]_{T_N} \le \bar{k} + 1 + [\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N},$$

whence $[\ell^{N,i}, \ell^{N,i}]_{T_N}$ is integrable.

For some $\varepsilon_N > 0$ to be determined later, create the following stopping times:

$$\tau_0^N = 0;$$

$$\tau_{u+1}^N = \begin{cases} (\tau_u^N + \varepsilon_N) \wedge T_N, & \text{if } |\ell_s^{N,i} - \ell_{\tau_u}^{N,i}| \le 1 \text{ for all } i, s, \\ & \quad \tau_u^N \le s \le (\tau_u^N + \varepsilon_N) \wedge T_N \\ \inf\left\{s: |\ell_s^{N,i} - \ell_{\tau_u^N}^{N,i}| > 1 \text{ for some } i\right\}, & \text{otherwise.} \end{cases}$$

Let the set E_M^N be given as the indices u, $1 \le u \le M$, for which the second option is used to fix τ_u^N. Let $|E_M^N|$ be the number of elements of E_M^N.

Note that

$$
\begin{aligned}
|E_M^N| &\le \sum_i \sum_{u \in E_M^N} \left(\ell^{N,i}_{\tau^N_u} - \ell^{N,i}_{\tau^N_{u-1}} \right)^2 \\
&\le \sum_i \sum_{1 \le u \le M} \left(\ell^{N,i}_{\tau^N_u} - \ell^{N,i}_{\tau^N_{u-1}} \right)^2 \qquad (6.3) \\
&\le \sum_i \sum_u \left(\ell^{N,i}_{\tau^N_u} - \ell^{N,i}_{\tau^N_{u-1}} \right)^2 .
\end{aligned}
$$

By Jacod and Shiryaev [(1987), Theorem I.4.47, page 52], the r.h.s. of (6.3) converges in probability to $\sum_i [\ell^{N,i}, \ell^{N,i}]_{T_N}$ as $\varepsilon_N \to 0$. Fix ε_N such that

$$P\left(\left| \sum_i \sum_u \left(\ell^{N,i}_{\tau^N_u} - \ell^{N,i}_{\tau^N_{u-1}} \right)^2 - \sum_i [\ell^{N,i}, \ell^{N,i}]_{T_N} \right| > 1 \right) = o(r_N).$$

Then, by our assumptions on $[\ell^{N,i}, \ell^{N,i}]_{T_N}$, there is an integer M_0, $M_0 \ge p(\bar{k} + 1)$, such that

$$P\left(|E_M^N| > M_0 \right) = o(r_N).$$

By using a Taylor argument akin to those involving (6.1) and (6.2) we can therefore, without loss of generality, assume that $|E_M^N| \le M_0$ and, in particular, that $|E_\infty^N| \le M_0$. Hence, if

$$M_N = M_0 + 1 + \text{integer part of } T_N / \varepsilon_N,$$

then $\tau_{M_N} = T_N$ a.s.

Now consider the discrete time martingale $\tilde{\ell}^{N,i}_n$, $n = 0, \dots, M_N$, given by

$$\tilde{\ell}^{N,i}_n = \ell^{N,i}_{\tau_n}.$$

Correspondingly, $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$ and so on are defined. Since $\tilde{\ell}^{N,i}_{M_N} = \ell^{N,i}_{T_N}$, it is enough to prove Theorem 1 if we check the conditions on the "variations" associated with $\tilde{\ell}^{N,i}_n$.

We begin with $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$. Note that

$$
\begin{aligned}
&\left[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i} \right]_{M_N} \\
&\quad = \sum_u \left(\ell^{N,i}_{\tau_u} - \ell^{N,i}_{\tau_{u-1}} \right)^4 \\
&\quad \le \sum_u \left(\ell^{N,i}_{\tau_u} - \ell^{N,i}_{\tau_{u-1}} \right)^2 + 2^4 \sum_{u \in E_\infty^N} \left(\left(\Delta \ell^{N,i}_{\tau_u} \right)^4 + 1 \right) \qquad (6.4) \\
&\quad \le \left[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i} \right]_{M_N} + 2^4 [\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} + 2^4 M_0 .
\end{aligned}
$$

As in Jacod and Shiryaev [(1987), Theorem I.4.47, page 52], $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$ converges in probability to $[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N}$ if we let ε_N be small enough. By (6.4) and (6.6) (below), this convergence is dominated, whence ε_N can be selected to go sufficiently fast to zero to get

$$E\left[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}\right]_{M_N} - E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} = O(r_N^2).$$

Hence the integrability condition on $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$ is satisfied.

Turning to $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$, note that

$$E\left[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}\right]_{M_N} = E[\ell^{N,i}, \ell^{N,i}]_{T_N}. \tag{6.5}$$

Since $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N} \to [\ell^{N,i}, \ell^{N,i}]_{T_N}$ in probability as ε_N becomes small (for fixed N), and since (6.5) yields that this convergence also holds in L^1, we can choose ε_N to go sufficiently fast to zero to get

$$E\left|\left[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}\right]_{M_N} - [\ell^{N,i}, \ell^{N,i}]_{T_N}\right| = o(r_N). \tag{6.6}$$

By Proposition 3, the integrability condition for $[\ell^{N,i}, \ell^{N,i}]_{T_N}$ is satisfied, and by the same reasoning as in Mykland [(1993), Lemma 2], this is also true for $[\tilde{\ell}^{N,i}, \tilde{\ell}^{N,i}]_{M_N}$. Equation (6.6) also yields that ψ_0 exists for $(\tilde{\ell}_n^{N,i})$ and has the same form as for $(\ell_t^{N,i})$. The same type of reasoning gives the same result for ψ_p if one observes that the limit of a sequence of predictable processes is predictable. □

PROOF OF PROPOSITION 2. Assumption (2.6) is obviously sufficient for (1.2). To prove necessity, note that the signed measure ν from (1.2) can be represented (for nice g's)

$$\int g(x)\, d\nu(x) = \frac{1}{2} E\left[v^{i,j} + \frac{\kappa^{i,j,k} h_k(Z)}{3}\right] \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} g(Z),$$

where Z is $N(0, \kappa^{ij})$. Comparing this to (2.12) yields (2.6), and also (2.7) for symmetric definitions of $\kappa^{i,j,k}$ and $v^{i,j}$. □

PROOF OF PROPOSITION 3. From inequality (5.6) in Mykland (1993),

$$E([\ell^{N,i}, \ell^{N,i}]_{T_N} - \langle \ell^{N,i}, \ell^{N,i} \rangle_{T_N})^2 \le E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} \tag{6.7}$$

and

$$E((\ell^{N,i}, \ell^{N,i})_{T_N} - \langle \ell^{N,i}, \ell^{N,i} \rangle_{T_N})^2 \le \tfrac{1}{9} E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} \tag{6.8}$$

for discrete time martingales. By taking limits, the same results hold for general time axis. Hence, one can apply Lemma 2 of that paper to show the first part of Proposition 3.

The second part of the result follows as in Mykland [(1993), Proposition 2], with the appropriate modifications which come from dealing with a martingale with general time axis. □

PROOF OF PROPOSITION 4. The proof is divided into three parts.

Fact 1. $E(m_{T_N}^{N,ii})^2 = O(N)$. In view of (6.8), we can replace $m_{T_N}^{N,ii}$ with $L_{T_N}^{N,ii} + [L^{N,i}, L^{N,i}]_{T_N}$ for this purpose. From the fourth Bartlett identity,

$$E\big[L^{N,ii} + [L^{N,i}, L^{N,i}], L^{N,ii} + [L^{N,i}, L^{N,i}]\big]_{T_N}$$
$$= -\tfrac{4}{3}E[L^{N,i}, L^{N,iii}]_{T_N} - \tfrac{1}{3}EL_{T_N}^{N,iiii} + \tfrac{2}{3}E[L^{N,i}, L^{N,i}, L^{N,i}, L^{N,i}]_{T_N}$$
$$= O(N),$$

in view of our assumptions and since

$$|E[L^{N,i}, L^{N,iii}]_{T_N}| \le E[L^{N,i}, L^{N,i}]_{T_N} + E[L^{N,iii}, L^{N,iii}]_{T_N} = O(N),$$

by our assumptions and by the second Bartlett identity.

Fact 2. For any x_i,

$$E_{\text{as}}\Big(x_i x_j N^{-1/2} m_{T_N}^{N,ij} \mid x_k N^{-1/2} L_{T_N}^{N,k}\Big) = c x_k Z^k, \tag{6.9}$$

where Z^k is as in (2.5) and where

$$c x_i x_j \kappa^{i,j} = -x_i x_j x_k \lambda^{ijk}/3. \tag{6.10}$$

Our assumptions yield that $L_{T_N}^{iii} - \langle L^{N,iii}\rangle_{T_N} = O_p(N^{1/2})$, whence it follows from (3.4) that

$$\langle x_i x_j m^{N,ij}, x_k L^{N,k}\rangle_{T_N}/N \to -x_i x_j x_k \lambda^{ijk}/3. \tag{6.11}$$

By the same reasoning as in Proposition 3 [or by Mykland (1993), Proposition 2], (6.9) follows. Equation (6.10) then follows from (6.9) and (6.11).

Tying it together. The integrability conditions are a consequence of Fact 1. The relationship between ρ^{ij} and $\bar{\rho}^{ij}$ follow from Fact 2 in the following way. Equation (6.10) implies that, if Z is $N(O, \kappa^{i,j})$,

$$E\Big(x_i x_j \lambda^{ijk}\kappa_{k,\alpha} Z^{\alpha}/3 + c x_{\alpha} Z^{\alpha}\Big) x_{\beta} Z^{\beta} = 0,$$

that is,

$$E\Big(x_i x_j \lambda^{ijk} h_k(Z)/3 + c x_{\alpha} Z^{\alpha}\Big) x_k Z^k = 0.$$

The equivalence between orthogonality and independence for Gaussian random variables then yields that

$$E\Big(x_i x_j \lambda^{ijk} h_k(Z)/3 \mid x_k Z^k\Big) = -c x_k Z^k.$$

Thus, (6.9) gives that

$$E_{\text{as}}\Big(x_i x_j N^{-1/2} m_{T_N}^{N,ij} \mid x_k N^{-1/2} L_{T_N}^{N,k}\Big)$$
$$= -E\Big(x_i x_j \lambda^{ijk} h_k(Z)/3 \mid x_k Z^k\Big).$$

The result now follows by, for example, a Fourier argument. □

PROOF OF PROPOSITION 5. By the reasoning used in the proof of Proposition 6.1 of Hall and Heyde [(1980), page 160], we can see that, under P_{θ_0}, $L_{T_N}^{N,i}(\theta_0)/\sqrt{c_N} \to_{\mathscr{L}} Z^i$, where Z is $N(0, \kappa^{i,j}(\theta_0))$. Let us assume that we are dealing with a subsequence so that $r_N^{-1}(L_{T_N}^{N,ij}(\theta_0)/c_N + \kappa^{i,j}) \to_{\mathscr{L}} \bar{\xi}^{ij}$ for all i, j, jointly with $\ell_{T_N}^{N,k}(\theta_0)/\sqrt{c_N}$. Since

$$r_N^{-1}\left(\frac{L^{N,ij}(\theta_N)}{c_N} + \kappa^{i,j}(\theta_N)\right) - r_N^{-1}\left(\frac{L^{N,ij}(\theta_0)}{c_N} + \kappa^{i,j}(\theta_0)\right)$$
$$= r_N^{-1}\big(\kappa^{i,j}(\theta_N) - \kappa^{i,j}(\theta_0)\big) + r_N^{-1}(\theta_{N,k} - \theta_{0,k})L^{N,ijk}(\theta_N^*),$$

it follows that $\liminf_{N\to\infty}|r_N c_N^{1/2}| > 0$ unless $g^{ij,k} = \lambda^{ijk} + \partial\kappa^{i,j}(\theta)/\partial\theta^k|_{\theta=\theta_0} = 0$, and that

$$E_{\theta_0}(\bar{\xi}^{ij}) = E_{\theta_0}\big(\bar{\xi}^{ij} + h\omega_k g^{ij,k}\big)\exp\big\{\omega_k Z^k - \tfrac{1}{2}\omega_i\omega_j\kappa^{i,j}\big\},$$

where $\omega_k = \lim_{v\to\infty} c_N^{1/2}(\theta_{N,k} - \theta_0)$ and $h = \lim_{N\to\infty} r_N^{-1}c_N^{-1/2}$ (if the limits do not exist, take a subsequence). Hence the result follows. □

PROOF OF PROPOSITION 6. Working with (4.7) and (4.10), it follows from (4.11)–(4.12), the i.i.d. assumption, the boundedness of $|Z_s^{\alpha,i}\lambda_{\alpha,s}\bar{\lambda}^+|$ and Proposition 3 that the integrability conditions are satisfied and that $\kappa^{i,j}$ and $v^{i,j}$ in (4.11) are the ones referred to in Theorem 1.

To calculate $\kappa^{i,j,k}$, write

$$L_t^k = Q_t^k - R_t^k, \tag{6.12}$$

where

$$Q_t^k = \int_0^t \Big[Z_s^{i,k} - E^N\big(Z_s^{1,k}\lambda_{1,s}\big)\big(E^N\lambda_{1,s}\big)^+\Big]\,dM_{i,s} \tag{6.13}$$

and

$$R_t^k = \int_0^t \Big[Z_s^{i,k}\lambda_{i,s}\bar{\lambda}_s^+ - E^N\big(Z_s^{1,k}\lambda_{1,s}\big)\big(E^N\lambda_{1,s}\big)^+\Big]\,d\bar{M}_s. \tag{6.14}$$

Note that, clearly, $\langle R^k, R^k\rangle_1 = O_p(1)$, while $\langle Q^j, R^k\rangle_1 = o_p(N^{1/2})$, since

$$E\Big[Z_s^{i,k} - E^N\big(Z_s^{1,k}\lambda_{1,s}\big)\big(E^N\lambda_{1,s}\big)^+\Big]\lambda_{i,s} = 0$$

except when $E^N\lambda_{1,s} = 0$. Similarly, $\langle Q^i, Q^j, R^k\rangle_1 = o_p(N)$, $\langle Q^i, R^j, R^k\rangle_1 = o_p(N)$ and $\langle R^i, R^j, R^k\rangle_1 = o_p(N)$. Since the integrability conditions of Theorem 1 are satisfied for Q^k and R^k for the same reasons as this is true for L^k, it follows from Proposition 3 that the R-component in (6.12) contributes nothing to the asymptotic regression (2.4)–(2.5). That reduces the problem to calculating $\kappa^{i,j,k}$ for the Q-component in (6.12), which is straightforward. □

PROOF OF PROPOSITION 7. Assume first that the time axis is discrete. Embed $(\ell_t^{N,\cdot})_{0\le t\le T_N}$ in a continuous martingale $(\bar{\ell}_t^{N,\cdot})_{0\le t\le T_N}$. By Itô's lemma [see, e.g., Jacod and Shiryaev (1987), Theorem I.4.57],

$$d\bar{\ell}_t^{N,i}\bar{\ell}_t^{N,j}\bar{\ell}_t^{N,k} = \bar{\ell}_t^{N,i}\bar{\ell}_t^{N,j}\,d\bar{\ell}_t^{N,k}[3] + \bar{\ell}_t^{N,k}\,d\langle\bar{\ell}^{N,i}, \bar{\ell}^{N,j}\rangle_t[3]$$

and

$$d\langle \bar{\ell}^{N,i}, \bar{\ell}^{N,j}\rangle_t \bar{\ell}_t^{N,k} = \langle \bar{\ell}^{N,i}, \bar{\ell}^{N,j}\rangle_t \, d\bar{\ell}_t^{N,k} + \bar{\ell}_t^{N,k} \, d\langle \bar{\ell}^{N,i}, \bar{\ell}^{N,j}\rangle_t.$$

Also, by Mykland [(1993), (5.12)–(5.13)],

$$x_i x_j x_k E\langle \bar{\ell}^{N,i}, \bar{\ell}^{N,j}\rangle_{T_N} \bar{\ell}_{T_N}^{N,k} = x_i x_j x_k E(\ell^{N,i}, \ell^{N,j})_{T_N} \ell_{T_N}^{N,k}.$$

Combining the three equations above yields

$$\begin{aligned} &x_i x_j x_k r_N^{-1} E \ell_{T_N}^{N,i} \ell_{T_N}^{N,j} \ell_{T_N}^{N,k} \\ &\quad = 3 x_i x_j x_k E r_N^{-1} \big((\ell^{N,i}, \ell^{N,j})_{T_N} - \kappa^{i,j} \big) \ell_{T_N}^{N,k}. \end{aligned} \tag{6.15}$$

By taking limits, we can now go over to a general time axis. Since the integrand on the r.h.s. clearly converges to $3x_i x_j x_k \xi^{ij} Z^k$, it remains to show uniform integrability. By Hölder's inequality, it is clearly enough to show that $E|\ell_{T_N}^{N,i}|^p$ and $E(|(\ell^{N,i}, \ell^{N,i})_{T_N} - \kappa^{i,i}|/r_N)^q$ are bounded for p, q so that $p^{-1} + q^{-1} = 1 - \varepsilon$. Assume first that the time axis is discrete. Since, by Burkholder's inequality for continuous martingales [see Jacod (1979)],

$$\begin{aligned} E|\ell_{T_N}^{N,i}|^p &= E|\bar{\ell}_{T_N}^{N,i}|^p \\ &\le c_1 E\langle \bar{\ell}^{N,i}, \bar{\ell}^{N,i}\rangle_{T_N}^{p/2} \\ &\le c_2 \Big\{ E(\ell^{N,i}, \ell^{N,i})_{T_N}^{p/2} + \big(E[\ell^{N,i}, \ell^{N,i}, \ell^{N,i}, \ell^{N,i}]_{T_N} \big)^{p/4} \Big\} \end{aligned}$$

by Liapunov's inequality and inequality (5.11) in Mykland (1993), c_i being a constant, and having assumed that $2 \le p \le 4$. By taking limits we can now eliminate the discrete time assumption, and it is clearly enough that $E(|(\ell^{N,i}, \ell^{N,i})_{T_N} - \kappa^{i,i}|/r_N)^q$ be bounded with $q = p/2$. The result follows. □

Proof of Proposition 8. Follows directly from (6.15) and the Toeplitz lemma [see Hall and Heyde (1980), Section 2.6].

Proof of Proposition 9. It follows from the ergodic theorem [see, e.g., Breiman (1968), Section 6] that $E[\ell^i, \ell^i, \ell^i, \ell^i]_N = O(N)$, for each i, and that

$$[\ell^i, \ell^j, \ell^k]_N / N \to_P \eta^{i,j,k} \tag{6.16}$$

as $N \to \infty$, for each i, j, k. By Hall and Heyde [(1980), Corollary 5.1, page 132],

$$\big(\ell^{N,i}/\sqrt{N}, \sqrt{N}\big([\ell^j, \ell^k]_N / N - \kappa^{j,k} \big); \text{all } i, j, k \big)$$

is asymptotically jointly normal with mean zero, whence the linearity condition (2.6) is satisfied in view of (6.16) and Proposition 3, with $v^{ij} = 0$. The same Corollary 5.1 also yields that $[\sqrt{N}([\ell^i, \ell^i]_N/N - \kappa^{i,i})]^2$ is uniformly integrable, whence, in view of (6.7)–(6.8), $E[\sqrt{N}((\ell^i, \ell^i)_N/N - \kappa^{i,i})]^2 = O(1)$.

Hence the integrability conditions (A) and (B) of Theorem 1 are satisfied, as are the conditions of Proposition 8. □

Acknowledgments. The author would like to thank Kung-Sik Chan, Siu-Kai Chan, Peter McCullagh, Wing Wong, Michael Woodroofe and two referees for very helpful discussions and comments, and Mitzi Nakatsuka for typing the manuscript.

REFERENCES

ANDERSEN, P. K. and BORGAN, Ø. (1985). Counting process models for life history data: a review (with discussion). *Scand. J. Statist.* **12** 97–158.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

BASAWA, I. V. and PRAKASA RAO, B. L. S., (1980). *Statistical Inference for Stochastic Processes.* Academic Press, New York.

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 435–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, MA.

BRÉMAUD, P. (1981). *Point Processes and Queues—Martingale Dynamics.* Springer, New York.

CHAN, K.-S. (1990). Deterministic stability, stochastic stability, and ergodicity. In *Non-linear Time Series. A Dynamical System Approach* (H. Tong, ed.) 448–466. Oxford Science Publications.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information (with discussion). *Biometrika* **65** 457–487.

GOETZE, F. and HIPP, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–953.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic, New York.

JACOD, J. (1979). *Calcul Stochastique et Problèmes de Martingales. Lecture Notes in Math.* **714** Springer, Berlin.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer, New York.

JENSEN, J. L. (1989). Asymptotic expansions for strongly mixing Harris recurrent Markov chains. *Scand. J. Statist.* **16** 47–64.

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

KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

LIPTSER, R. S. and SHIRYAEV, A. N. (1977). *Statistics of Random Processes I. General Theory.* Springer, New York.

LIPTSER, R. S. and SHIRYAEV, A. N. (1978). *Statistics of Random Processes II. Applications.* Springer, New York.

MALINOVSKII, V. K. (1986). Limit theorems for Harris Markov chains. *Theory Probab. Appl.* **31** 269–285.

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.

MILLAR, P. W. (1985). Non-parametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. Verw. Gebiete* **68** 545–556.

MYKLAND, P. A. (1989). Edgeworth and bootstrap methods for dependent variables. Ph.D. dissertation, Univ. California, Berkeley.

MYKLAND, P. A. (1992). Asymptotic expansions and bootstrapping-distributions for dependent variables: a martingale approach. *Ann. Statist.* **20** 643–654.

MYKLAND, P. A. (1993). Asymptotic expansions for martingales. *Ann. Probab.* **21** 800–818.

NAGAEV, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Probab. Appl.* **2** 378–406.

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

ROUSSAS, G. G. (1972). *Contiguous Probability Measures: Some Applications in Statistics*. Cambridge Univ. Press.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

SKOVGAARD, I. M. (1981a). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8** 207–217.

SKOVGAARD, I. M. (1981b). Edgeworth expansions of the distributions of maximum likelihood estimators in the general (non i.i.d.) case. *Scand. J. Statist.* **8** 227–236.

STATULEVICIUS, V. (1969). Limit theorems for sums of random variables that are connected in a Markov chain I, II. *Litovsk. Mat. Sb.* **9** 345–362; 635–672.

STATULEVICIUS, V. (1970). Limit theorems for sums of random variables that are connected in a Markov chain III. *Litovsk. Mat. Sb.* **10** 161–169.

TJØSTHEIM, D. (1986). Estimation in nonlinear time series models. *Stochastic Process. Appl.* **21** 251–273.

TJØSTHEIM, D. (1990). Non-linear time series and Markov chains. *Adv. in Appl. Probab.* **22** 587–611.

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

WALLACE, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14** 88–123.

WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: Linear models. *Ann. Statist.* **17** 1087–1102.

WOODROOFE, M. and KEENER, R. (1987). Asymptotic expansions in boundary crossing problems. *Ann. Probab.* **15** 102–114.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637