Quasi-likelihood models and optimal inference

Wolfgang Wefelmeyer
University of Siegen

Abstract

Consider an ergodic Markov chain on the real line, with parametric models for the conditional mean and variance of the transition distribution. Such a setting is an instance of a quasi-likelihood model. The customary estimator for the parameter is the maximum quasi-likelihood estimator. It is not efficient, but as good as the best estimator that ignores the parametric model for the conditional variance. We construct two efficient estimators. One is a convex combination of solutions of two estimating equations, the other a weighted nonlinear one-step least squares estimator, with weights involving predictors for the third and fourth centered conditional moments of the transition distribution. Additional restrictions on the model can lead to further improvement. We illustrate this with an autoregressive model whose error variance is related to the autoregression parameter.

1 Introduction

According to Wedderburn (1974), a quasi-likelihood model is defined by a relation between mean and variance of the observations. A simple example are i.i.d. observations with known coefficient of variation, but otherwise unknown distribution; efficient estimators for the mean are constructed in Bickel et al. (1993, p. 68). A related regression model is considered by Amemiya (1973). A rich class of quasi-likelihood models is given by generalized linear models with a restriction on the variance of the response. The basic reference is McCullagh and Nelder (1989). Some surveys may be found in Hinkley et al. (1991).

For discrete-time stochastic processes, quasi-likelihood models are defined by specifying parametric models for the conditional mean and variance processes given the past. Examples are the Markov regression models of Zeger and Qaqish (1988), see also Huhtala (1992). For continuous time, a quasi-likelihood model is described by parametric models for the compensator and the predictable quadratic variation of a semimartingale. There

[1] *AMS* 1991 *subject classifications.* Primary 62G20; secondary 62M05.

[2] *Key words and phrases.* Efficient estimator, quasi-likelihood, Markov chain, weighted nonlinear least squares, autoregression.

is a considerable literature on quasi-likelihood models for stochastic processes. Several surveys are collected in Godambe (1991).

We are interested in efficient estimation of the parameter. To keep the model simple and the assumptions specific, we restrict attention to Markov chains and to one-dimensional parameters. A version of our approach for general semimartingales is outlined in Wefelmeyer (1993). In the Introduction we describe some results on estimating functions in quasi-likelihood models for Markov chains. They are essentially known in other settings and easy to derive. Hence we do not prove them. The results will motivate our construction of an efficient estimator.

Let $X_0, \dots, X_n$ be observations from an ergodic real-valued Markov chain with transition distribution $Q(x, dy)$ and invariant distribution $\pi(dy)$. Suppose that we have parametric models for the conditional mean, or autoregression function, and the conditional variance,

$$\int y Q(x, dy) = m_\vartheta(x), \tag{1.1}$$

$$\int \left(y - m_\vartheta(x)\right)^2 Q(x, dy) = v_\vartheta(x), \tag{1.2}$$

but that the transition distribution is unspecified otherwise.

A large class of estimators for ϑ is obtained as solutions of estimating equations of the form

$$\sum_{i=1}^n w_\vartheta(X_{i-1})\left(X_i - m_\vartheta(X_{i-1})\right) = 0. \tag{1.3}$$

Under appropriate conditions, the corresponding estimator is asymptotically normal with variance

$$\pi(w_\vartheta^2 v_\vartheta) \Big/ \left(\pi(w_\vartheta m'_\vartheta)\right)^2.$$

Here $\pi(f)$ is short for the expectation $\int f(x)\pi(dx)$, and prime denotes differentiation with respect to ϑ. Consistency and asymptotic normality may be proved along the lines of Klimko and Nelson (1987).

By the Schwarz inequality, the variance is minimized for $w_\vartheta = m'_\vartheta / v_\vartheta$. The minimal variance is

$$1 \Big/ \pi(m'^2_\vartheta / v_\vartheta). \tag{1.4}$$

A version of this result for general discrete-time processes is in Godambe (1985). For continuous time see Thavaneswaran and Thompson (1986), Hutton and Nelson (1986) and Godambe and Heyde (1987). The denominator $\pi(m'^2_\vartheta / v_\vartheta)$ in (1.4) is called the *quasi-Fisher information.* The optimal estimator is the *maximum quasi-likelihood estimator.* It solves

$$\sum_{i=1}^n v_\vartheta(X_{i-1})^{-1} m'_\vartheta(X_{i-1})\left(X_i - m_\vartheta(X_{i-1})\right) = 0.$$

A different, stronger, optimality property of the maximum quasi-likelihood estimator is obtained in Wefelmeyer (1994c): The estimator attains the asymptotic variance bound for regular estimators which *ignore* the parametric model (1.2) for the conditional variance. This implies that the maximum quasi-likelihood estimator does not use the information about ϑ in (1.2), even though its definition requires (1.2). Crowder (1987) gives two examples in which there is much more information in (1.2) than in (1.1). Amemiya (1973), Firth (1987) and Hill and Tsai (1988) consider the loss in efficiency under the assumption that the underlying model is a specific *parametric* model. Then an efficient estimator is given by the maximum likelihood estimator.

If the transition distribution is unspecified except for (1.1) and (1.2), how can we find a better estimator than the maximum quasi-likelihood estimator? Note first that the estimators obtained from (1.3) are consistent because $X_i - m_\vartheta(X_{i-1})$ are martingale increments by condition (1.1). From condition (1.2) we obtain martingale increments

$$\Big(X_i - m_\vartheta(X_{i-1})\Big)^2 - v_\vartheta(X_{i-1}).$$

These lead to further consistent estimating equations besides (1.3),

$$\sum_{i-1}^{n} w_\vartheta(X_{i-1})\Big(\Big(X_i - m_\vartheta(X_{i-1})\Big)^2 - v_\vartheta(X_{i-1})\Big) = 0. \tag{1.5}$$

It suggests itself to combine estimating equations (1.3) and (1.5),

$$\begin{aligned}\sum_{i=1}^{n}\Big(& w_m(X_{i-1})\Big(X_i - m_\vartheta(X_{i-1})\Big)\\ &+ w_v(X_{i-1})\Big(\Big(X_i - m_\vartheta(X_{i-1})\Big)^2 - v_\vartheta(X_{i-1})\Big)\Big) = 0.\end{aligned} \tag{1.6}$$

Under appropriate conditions, the corresponding estimator is asymptotically normal with variance

$$\pi\Big(w_m^2 v_\vartheta + 2w_m w_v \mu_3 + w_v^2(\mu_4 - v_\vartheta^2)\Big)\Big/\Big(\pi(w_m m'_\vartheta + w_v v'_\vartheta)\Big)^2. \tag{1.7}$$

The variance depends on the centered conditional moments

$$\mu_j(x) = \int \Big(y - m_\vartheta(x)\Big)^j Q(x, dy), \qquad j = 3, 4. \tag{1.8}$$

We will not express the dependence of μ_j and similar terms on Q. The variance (1.7) is minimized for

$$w_m = C_\vartheta^{-1} A_\vartheta, \qquad w_v = C_\vartheta^{-1} B_\vartheta\,,$$

with

$$A_\vartheta = m'_\vartheta(\mu_4 - v_\vartheta^2) - v'_\vartheta \mu_3\,, \tag{1.9}$$

$$B_\vartheta = v'_\vartheta v_\vartheta - m'_\vartheta \mu_3\,, \tag{1.10}$$

$$C_\vartheta = (\mu_4 - v_\vartheta^2) v_\vartheta - \mu_3^2. \tag{1.11}$$

By the Schwarz inequality, $C_\vartheta(x)$ is positive unless $Q(x,\cdot)$ is degenerate. The minimum variance is

$$1\Big/\pi\Big(C_\vartheta^{-1}(A_\vartheta m'_\vartheta + B_\vartheta v'_\vartheta)\Big). \tag{1.12}$$

By the Schwarz inequality, this is strictly smaller than the asymptotic variance (1.4) of the maximum quasi-likelihood estimator unless $B_\vartheta = 0$. Hence the maximum quasi-likelihood estimator is inefficient except when $B_\vartheta = 0$.

The optimal estimator solves

$$\begin{aligned}\sum_{i=1}^n C_\vartheta(X_{i-1})^{-1}\Big(&A_\vartheta(X_{i-1})\big(X_i - m_\vartheta(X_{i-1})\big) \\ &+B_\vartheta(X_{i-1})\Big(\big(X_i - m_\vartheta(X_{i-1})\big)^2 - v_\vartheta(X_{i-1})\Big)\Big) = 0.\end{aligned} \tag{1.13}$$

The weights depend, through μ_3 and μ_4, on the unknown transition distribution Q. Hence the estimator is, in general, not useful. Suppose, for the moment, that besides (1.1) and (1.2) we have parametric models for the third and fourth centered conditional moments,

$$\mu_3(x) = \mu_{3\vartheta}(x), \qquad \mu_4(x) = \mu_{4\vartheta}(x).$$

Such a model is called an *extended quasi-likelihood model.* Consider estimating equations (1.6), with weights w_m and w_v possibly depending on ϑ. Under appropriate conditions, the corresponding estimator is again asymptotically normal. Its asymptotic variance equals again (1.7), of course now with $\mu_3 = \mu_{3\vartheta}$ and $\mu_4 = \mu_{4\vartheta}$. The variance is again minimized for the estimator obtained from (1.13). Now the weights depend on Q through ϑ only. The optimal estimator is the *extended maximum quasi-likelihood estimator.* The optimal weights are determined by Crowder (1986, 1987) for independent observations, and by Godambe (1987) and Godambe and Thompson (1989) for discrete-time stochastic processes. These authors restrict attention to the special case with (1.3) and (1.5) orthogonal, i.e. $\mu_{3\vartheta} = 0$. The general case, also for continuous time, is treated in Heyde (1987).

We return to the ordinary quasi-likelihood model. Then (1.12) is still a variance bound for estimators obtained from an equation of the form (1.6), with weights w_m and w_v possibly depending on ϑ. Two questions arise. Is (1.12) also a variance bound for the much larger class of *regular* estimators? This will be shown in Theorem 1. Can we find an estimator which attains the bound for all Q? We describe such an estimator in Theorem 2. The basic idea is the following. For fixed Q, a regular estimator attaining the bound was obtained above as solution of (1.13). The estimating function, and hence the estimator, depends on Q through μ_3 and μ_4. We want an adaptive version of the estimator. There are several options. The most direct one consists in replacing A_ϑ, B_ϑ, C_ϑ in the estimating equation (1.13) by estimators. They may still depend on ϑ. The resulting estimating equation may be difficult to solve. A second possibility is a random convex combination of two estimators which solve equations of the form (1.3) and (1.5), with appropriate weights. A third option is a weighted nonlinear one-step least squares estimator; such an estimator can be written in closed form.

In particular, the efficient estimator has the following property. Whatever the parametric models for μ_3 and μ_4 in an extended quasi-likelihood model, our estimator is asymptotically as good as the extended maximum quasi-likelihood estimator when $\mu_{3\vartheta}$ and $\mu_{4\vartheta}$ are correctly specified, and strictly better when they are not.

Additional restrictions on the model can lead to further improvement. In Section 3 we assume that $Q(x,dy) = p(y - \vartheta x)dy$, with p a mean zero density. Then the observations come from an autoregressive process with error density p. We specify the error variance as a function of ϑ. The maximum quasi-likelihood estimator is the least squares estimator. We obtain an efficient estimator as a random convex combination of two estimators. One is an estimator for ϑ, the other is a function of an estimator for the error variance. Both these estimators are efficient in the usual autoregressive model, without restriction on the error variance. The first is due to Kreiss (1987), the second to Wefelmeyer (1994a). A simpler, but inefficient, convex combination is described in Wefelmeyer (1994b).

2 Main results

Let $X_0, \ldots, X_n$ be observations from a real-valued Markov chain, with unknown transition distribution $Q(x,dy)$ fulfilling

$$\int yQ(x,dy) = m_\vartheta(x), \tag{2.1}$$

$$\int \Big(y - m_\vartheta(x)\Big)^2 Q(x,dy) = v_\vartheta(x). \tag{2.2}$$

The model can be written as a semiparametric model, with nuisance parameter given by transition distributions with conditional mean 0 and conditional variance 1. Then (2.1) and (2.2) can be generated by conditional location and scale transformations. However, we found it more convenient to treat (2.1) and (2.2) as side conditions on the nonparametric model described by all transition distributions.

Fix ϑ in some open subset of the real line, and a transition distribution Q fulfilling (2.1) and (2.2).

Assumptions The Markov chain is stationary and ergodic, with nondegenerate invariant distribution π. The function C_ϑ defined in (1.11) is bounded away from zero π-almost surely. For τ in a neighborhood of ϑ, and for all x, the functions $m_\tau(x)$ are twice differentiable in τ, with first derivatives at ϑ bounded in x. The second derivatives fulfill Lipschitz conditions at $\tau = \vartheta$,

$$|m''_\tau(x) - m''_\vartheta(x)| \le |\tau - \vartheta| c_m(x), \quad |v''_\tau(x) - v''_\vartheta(x)| \le |\tau - \vartheta| c_v(x).$$

The functions m'_ϑ, m''_ϑ, c_m, v'_ϑ, v''_ϑ, c_v have finite eighth moments. The fourth conditional moment of Q has finite fourth moment,

$$\int \Big(\int y^4 Q(x, dy) \Big)^4 \pi(dx) < \infty.$$

The derivatives m'_ϑ and v'_ϑ are not both π-almost surely equal to zero.

To keep the proofs short, we do not strive for minimal assumptions. Perhaps one can avoid second derivatives of m_ϑ and v_ϑ and prove (2.12) below by an appropriate version of the stochastic equicontinuity argument for M-estimators introduced by Huber (1967) in the i.i.d. case and by Bickel (1975) for the linear model. A recent reference is Welsh (1989). In the more specific setting of Section 3 the assumptions will be close to minimal.

To begin we show local asymptotic normality. A local model is introduced as follows. Let H denote the set of bounded functions $h(x, y)$ such that for all x,

$$\int h(x, y) Q(x, dy) = 0, \tag{2.3}$$

$$\int y h(x, y) Q(x, dy) = m'_\vartheta(x), \tag{2.4}$$

$$\int \big(y - m_\vartheta(x) \big)^2 h(x, y) Q(x, dy) = v'_\vartheta(x). \tag{2.5}$$

For $h \in H$ and $u \in \mathbb{R}$ we must construct a transition distribution Q^{nuh} such that (2.1) and (2.2) hold for $Q = Q^{nuh}$ and $\vartheta = \vartheta + n^{-1/2}u$. Consider first

$$Q_0^{nuh}(x, dy) = \big(1 + n^{-1/2} u h(x, y) \big) Q(x, dy).$$

Straightforward calculation shows that for $Q = Q_0^{nuh}$ and $\vartheta = \vartheta + n^{-1/2}u$, relations (2.1) and (2.2) hold up to terms of order n^{-1}. These terms cancel if we add to h an appropriate correction r_n of order $n^{-1/2}$ and set

$$Q^{nuh}(x, dy) = \big(1 + n^{-1/2} u (h(x, y) + r_n(x, y)) \big) Q(x, dy).$$

A possible choice of r_n is the following. Set

$$\begin{aligned} \bar{p}(y) &= y I(|y| \le n^{1/4}), \\ \bar{q}(x, y) &= (y - m_\vartheta(x))^2 I(|y - m_\vartheta(x)| \le n^{1/4}). \end{aligned}$$

Center these two functions for conditional expectation 0,

$$\begin{aligned} p(x, y) &= \bar{p}(y) - \int \bar{p}(y) Q(x, dy), \\ q(x, y) &= \bar{q}(x, y) - \int \bar{q}(x, y) Q(x, dy). \end{aligned}$$

Set

$$r_n(x, y) = a(x) p(x, y) + b(x) q(x, y).$$

Choose $\vartheta_{nu}(x)$ and $\tau_{nu}(x)$ between ϑ and $\vartheta + n^{-1/2}u$ such that

$$\begin{aligned} m_{\vartheta+n^{-1/2}u}(x) &= m_\vartheta(x) + n^{-1/2}um'_{\vartheta_{nu}(x)}(x), \\ v_{\vartheta+n^{-1/2}u}(x) &= v_\vartheta(x) + n^{-1/2}uv'_{\tau_{nu}(x)}(x). \end{aligned}$$

Define truncated centered moments

$$\begin{aligned} \bar{v}(x) &= \int \bar{q}(x,y)Q(x,dy), \\ \bar{\mu}_3(x) &= \int \Big(y - m_\vartheta(x)\Big)^2 p(x,y)Q(x,dy), \\ \tilde{\mu}_3(x) &= \int \Big(y - m_\vartheta(x)\Big)\bar{q}(x,y)Q(x,dy), \\ \bar{\mu}_4(x) &= \int \Big(y - m_\vartheta(x)\Big)^2 \bar{q}(x,y)Q(x,dy). \end{aligned}$$

Elementary computations show that r_n has the desired properties if

$$\begin{aligned} a\bar{v} + b\tilde{\mu}_3 &= m'_{\vartheta_{nu}} - m'_\vartheta \equiv s, \\ a\bar{\mu}_3 + b(\bar{\mu}_4 - v_\vartheta\bar{v}) &= v'_{\tau_{nu}} - v'_\vartheta + n^{-1/2}um'^2_{\vartheta_{nu}} \equiv t. \end{aligned}$$

Since s and t are of order $n^{-1/2}$, so are a and b. Here and in the following, we often suppress the dependence on n, and also on ϑ. We must set

$$\begin{aligned} a &= D^{-1}\Big((\bar{\mu}_4 - v_\vartheta\bar{v})s - \tilde{\mu}_3 t\Big), \\ b &= D^{-1}(\bar{v}t - \bar{\mu}_3 s), \end{aligned}$$

with $D = (\bar{\mu}_4 - v_\vartheta\bar{v})\bar{v} - \tilde{\mu}_3\bar{\mu}_3$ a determinant. Since C_ϑ is bounded away from zero π-almost surely, so is D. Hence D^{-1} is bounded. This ends the construction of the local model.

Write P_n for the joint distribution of $X_0, \ldots, X_n$ if Q is true, and P_n^{nuh} if Q^{nuh} is true. The family P_n^{nuh}, $h \in H$, $u \in \mathbb{R}$, is the *local model* at Q. Since it lies in the given model, it does not exclude reasonable estimators from competing. On the other hand, it is large enough to give a variance bound which is globally attainable, e.g. by the estimator in Theorem 2 below. Write $\pi \otimes Q$ for the invariant joint distribution $\pi(dx)Q(x,dy)$ of two successive observations, and

$$\pi \otimes Q(f) = \int\int f(x,y)Q(x,dy)\pi(dx).$$

We have *local asymptotic normality*,

$$\log dP_n^{nuh}\Big/dP_n = un^{-1/2}\sum_{i=1}^n h(X_{i-1}, X_i) - \frac{1}{2}u^2\pi \otimes Q(h^2) + o_{P_n}(1)$$

and

$$n^{-1/2}\sum_{i=1}^n h(X_{i-1}, X_i) \Rightarrow N_h \qquad \text{under } P_n,$$

where N_h is normal with mean zero and variance $\pi \otimes Q(h^2)$. The functions h will be called *score functions.* Local asymptotic normality for Markov chains is basically due to Roussas (1965). For nonparametric versions see Penev (1991), Greenwood and Wefelmeyer (1992), and Bickel (1993). Under our conditions a proof may be obtained directly, or by modifying the argument of Höpfner (1993), who treats Markov step processes. We need only check an appropriate version of Hellinger differentiability for Q^{nuh}, condition $H1''$ in Höpfner *et al.* (1990). Here it reads

$$\int \Big(\big(1 + n^{-1/2} h(x,y)\big)^{1/2} - 1 - n^{-1/2} \frac{1}{2} h(x,y) \Big)^2 Q(x, dy) \le n^{-1} r_n(x)$$

with r_n decreasing to zero pointwise and π-integrable for large n. This is true because h is bounded and hence $\pi \otimes Q$-square integrable. The only of the Assumptions we have used for local asymptotic normality is ergodicity.

We recall a well-known characterization of regular and efficient estimators. A convenient reference is Greenwood and Wefelmeyer (1990). As indicated at the beginning of this section, the model can be viewed as semiparametric. For such models, and for the i.i.d. case, versions of the concepts mentioned here are dicussed in the monograph of Bickel et al. (1993): See p. 46 there for regular estimators, p. 63 for the convolution theorem, the information (bound), and efficient estimators, p. 70 for efficient score functions, p. 19 for asymptotically linear estimators and influence functions, and p. 64 for the characterization of regular and efficient estimators.

Let $\overline{H}$ denote the closure of H in $L_2(\pi \otimes Q)$. The *efficient score function* $s \in \overline{H}$ at Q minimizes $\pi \otimes Q(h^2)$ over $h \in \overline{H}$. Hence it is characterized by

$$\pi \otimes Q(s^2) = \pi \otimes Q(sh) \qquad \text{for } h \in H. \tag{2.6}$$

The *information bound* at Q is the squared length of the efficient score function,

$$I = \pi \otimes Q(s^2). \tag{2.7}$$

An estimator $\hat{\vartheta}_n$ is *regular* for ϑ at Q with limit L if, for all $h \in H$ and $u \in \mathbb{R}$,

$$n^{1/2}(\hat{\vartheta}_n - \vartheta - n^{-1/2} u) \Rightarrow L \qquad \text{under } P_n^{nuh}.$$

By the convolution theorem,

$$L = M + N \qquad \text{in distribution,}$$

where M is independent of N, and N is normal with mean zero and variance I^{-1}. This justifies calling an estimator *efficient* for ϑ at Q if its limit under P_n is N. We call I^{-1} a *variance bound* for regular estimators. To state the characterization of regular and efficient estimators, we introduce the following definition. An estimator $\hat{\vartheta}_n$ is *asymptotically linear* for ϑ at Q with *influence function* $f(x,y)$ if

$$n^{1/2}(\hat{\vartheta}_n - \vartheta) = n^{-1/2} \sum_{i=1}^{n} f(X_{i-1}, X_i) + o_{P_n}(1).$$

The characterization reads as follows.

An estimator is regular and efficient for ϑ at Q if and only if it is asymptotically linear with influence function $f(x,y) = I^{-1}s(x,y)$.

To construct an efficient estimator, we need an explicit description of the efficient score function s and the information bound I. There are different ways of guessing the efficient score function. One guess relies on a formal analogy with the i.i.d. case. We expect that for fixed x the efficient score function is a linear combination of $y - m_\vartheta(x)$ and $\big(y - m_\vartheta(x)\big)^2 - v_\vartheta(x)$. Then (2.3) holds. The coefficients in the linear combination must be chosen such that (2.4) and (2.5) hold. A different guess is that the asymptotic variance bound (1.12) for estimators based on equations of the form (1.6) equals the variance bound for the larger class of *regular* estimators. Then the efficient score function is obtained from the estimating equation (1.13). The following theorem shows that both guesses are right.

Theorem 1 *The efficient score function at Q is*

$$s(x,y) = C_\vartheta(x)^{-1}\Big(A_\vartheta(x)\big(y - m_\vartheta(x)\big) + B_\vartheta(x)\Big(\big(y - m_\vartheta(x)\big)^2 - v_\vartheta(x)\Big)\Big).$$

The information bound at Q is positive and equals

$$I = \pi\Big(C_\vartheta^{-1}(A_\vartheta m_\vartheta' + B_\vartheta v_\vartheta')\Big).$$

Here A_ϑ, B_ϑ, C_ϑ are defined in (1.9) *to* (1.11).

Of the Assumptions we only use the nondegeneracy and moment conditions which ensure that s is $\pi \otimes Q$-square integrable and I is well defined and positive.

Proof of Theorem 1. It suffices to check that the function s is in $\overline{H}$ and fulfills (2.6). Then the explicit form of the information bound I is determined from (2.7). To show that $s \in \overline{H}$, we must check (2.3) to (2.5) and $s \in L_2(\pi \otimes Q)$. The calculations leading to (2.3) to (2.6) are straightforward, but tedious, and we omit them. It remains to prove that $s \in L_2(\pi \otimes Q)$, and that I is well defined and positive.

(i) To prove that $s \in L_2(\pi \otimes Q)$, introduce the conditional moments

$$\nu_j(x) = \int y^j Q(x, dy).$$

The following two integrals are finite:

$$\begin{aligned} \int\int A_\vartheta(x)^2 y^2 Q(x,dy)\pi(dx) &= \pi(A_\vartheta^2 \nu_2), \\ \int\int B_\vartheta(x)^2 y^4 Q(x,dy)\pi(dx) &= \pi(B_\vartheta^2 \nu_4). \end{aligned}$$

Since C_ϑ is bounded away from zero, we easily obtain that $s \in L_2(\pi \otimes Q)$.

(ii) Since C_ϑ is bounded away from zero, $Q(x,\cdot)$ is nondegenerate, by the Schwarz inequality. Write

$$\begin{aligned}&A_\vartheta(x)m'_\vartheta(x) + B_\vartheta(x)v'_\vartheta(x)\\ &= \int \Big(v'_\vartheta(x)\big(y - m_\vartheta(x)\big) - m'_\vartheta(x)\Big(\big(y - m_\vartheta(x)\big)^2 - v_\vartheta(x)\Big)\Big)^2 Q(x,dy).\end{aligned}$$

This is positive with positive π-probability since by assumption the derivatives v'_ϑ and m'_ϑ are not both equal to zero π-almost surely. Hence I is well defined and positive.

To describe our efficient one-step estimator for ϑ, we need an initial $n^{1/2}$-consistent estimator ϑ_n for ϑ, and *strongly consistent* predictors μ_{ji} for the centered conditional moments $\mu_j(X_i)$ of the transition distribution,

$$\mu_{ji} - \mu_j(X_i) \to 0 \qquad \text{almost surely for } j = 2,3,4.$$

Recall that μ_j is defined in (1.8). For ϑ_n one may choose a $n^{1/2}$-consistent solution of

$$n^{-1/2}\sum_{i=1}^n \big(X_i - m_\vartheta(X_{i-1})\big) = o_{P_n}(1).$$

We will not discuss conditions for the existence of such a solution here. For μ_{ji} one may take $\hat\mu_{ji}(X_i)$, where $\hat\mu_{ji}$ is a Nadaraya-Watson type kernel estimator for the function μ_j. For stochastic processes, such kernel estimators are discussed, e.g., by Collomb (1984) and Truong and Stone (1992). We do not repeat their assumptions here.

With these estimators, we obtain predictors for $A_\vartheta(X_i), B_\vartheta(X_i), C_\vartheta(X_i)$,

$$\begin{aligned}A_{\vartheta i} &= m'_\vartheta(X_i)(\mu_{4i} - \mu_{2i}^2) - v'_\vartheta(X_i)\mu_{3i},\\ B_{\vartheta i} &= v'_\vartheta(X_i)\mu_{2i} - m'_\vartheta(X_i)\mu_{3i},\\ C_i &= (\mu_{4i} - \mu_{2i}^2)\mu_{2i} - \mu_{3i}.\end{aligned}$$

Later we replace ϑ by the initial estimator ϑ_n. In particular, we obtain an estimator for the information bound I,

$$I_n = n^{-1}\sum_{i=1}^n C_{i-1}^{-1}\big(A_{\vartheta_n,i-1}m'_{\vartheta_n}(X_{i-1}) + B_{\vartheta_n,i-1}v'_{\vartheta_n}(X_{i-1})\big).$$

Theorem 2. *The estimator*

$$\begin{aligned}\hat\vartheta_n = \vartheta_n + I_n^{-1}n^{-1}\sum_{i=1}^n C_{i-1}^{-1}\Big(&A_{\vartheta_n,i-1}\big(X_i - m_{\vartheta_n}(X_{i-1})\big)\\ &+B_{\vartheta_n,i-1}\Big(\big(X_i - m_{\vartheta_n}(X_{i-1})\big)^2 - v_{\vartheta_n}(X_{i-1})\Big)\Big)\end{aligned}$$

is regular and efficient for ϑ at Q.

The estimator $\hat{\vartheta}_n$ involves predictors $\mu_{j,i-1}$ based on $X_0, \dots, X_{i-1}$ rather than estimators $\hat{\mu}_{jn}(X_{i-1})$ making full use of the observations $X_0, \dots, X_n$. We have chosen predictors because with them the processes

$$\sum_{i=1}^n C_{i-1}^{-1} A_{\vartheta,i-1}\big(X_i - m_\vartheta(X_{i-1})\big)$$

and

$$\sum_{i=1}^n C_{i-1}^{-1} B_{\vartheta,i-1}\Big(\big(X_i - m_\vartheta(X_{i-1})\big)^2 - v_\vartheta(X_{i-1})\Big)$$

are martingales. This will be used in the proof of Theorem 2. For a similar approach see Wefelmeyer (1994c).

In applications it will often be more convenient to use a weighted average of two estimators which solve equations of the form (1.3) and (1.5). Specifically, let A_i, B_i, C_i be predictors or estimators for $A_\vartheta(X_i), B_\vartheta(X_i), C_\vartheta(X_i)$. As above, these estimators may depend on ϑ. Let $\vartheta = \vartheta_n^m$ be a $n^{1/2}$-consistent solution of

$$n^{-1/2}\sum_{i=1}^n C_{i-1}^{-1} A_{i-1}\big(X_i - m_\vartheta(X_{i-1})\big) = o_{P_n}(1),$$

and let $\vartheta = \vartheta_n^v$ be a $n^{1/2}$-consistent solution of

$$\sum_{i=1}^n C_{i-1}^{-1} B_{i-1}\Big(\big(X_i - m_\vartheta(X_{i-1})\big)^2 - v_\vartheta(X_{i-1})\Big).$$

Write

$$I^m = \pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta), \qquad I^v = \pi(C_\vartheta^{-1} B_\vartheta v'_\vartheta).$$

Let a_n be a consistent estimator for $1\big/(1 + I^v/I^m)$. Then the convex combination $a_n\vartheta_n^m + (1-a_n)\vartheta_n^v$ is efficient for ϑ.

Proof of Theorem 2. By the characterization of regular and efficient estimators we must prove that $\hat{\vartheta}_n$ is asymptotically linear with influence function $I^{-1}s$, where s is the efficient score function and I the information bound determined in Theorem 1. We will prove the following two expansions, all sums extending over i from 1 to n:

$$\begin{aligned}
(2.8)\qquad & n^{-1/2}\sum C_{i-1}^{-1} A_{\vartheta_n,i-1}\big(X_i - m_{\vartheta_n}(X_{i-1})\big)\\
&= n^{-1/2}\sum C_\vartheta(X_{i-1})^{-1} A_\vartheta(X_{i-1})\big(X_i - m_\vartheta(X_{i-1})\big)\\
&\quad - n^{1/2}(\vartheta_n - \vartheta)\pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta) + o_{P_n}(1),\\
(2.9)\qquad & n^{-1/2}\sum C_{i-1}^{-1} B_{\vartheta_n,i-1}\Big(\big(X_i - m_{\vartheta_n}(X_{i-1})\big)^2 - v_{\vartheta_n}(X_{i-1})\Big)\\
&= n^{-1/2}\sum C_\vartheta(X_{i-1})^{-1} B_\vartheta(X_{i-1})\Big(\big(X_i - m_\vartheta(X_{i-1})\big)^2 - v_\vartheta(X_{i-1})\Big)\\
&\quad - n^{1/2}(\vartheta_n - \vartheta)\pi(C_\vartheta^{-1} B_\vartheta v'_\vartheta) + o_{P_n}(1).
\end{aligned}$$

The assertion follows from these two expansions if we prove that I_n is a consistent estimator of

$$I = \pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta) + \pi(C_\vartheta^{-1} B_\vartheta v'_\vartheta).$$

To show that I_n is consistent, we split I_n in the same way as I, and prove

$$n^{-1} \sum C_{i-1}^{-1} A_{\vartheta_n,i-1} m'_{\vartheta_n}(X_{i-1}) = \pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta) + o_{P_n}(1), \tag{2.10}$$

$$n^{-1} \sum C_{i-1}^{-1} B_{\vartheta_n,i-1} v'_{\vartheta_n}(X_{i-1}) = \pi\Big(C_\vartheta^{-1} B_\vartheta v'_\vartheta\Big) + o_{P_n}(1). \tag{2.11}$$

From now on we restrict attention to (2.8) and (2.10). Relations (2.9) and (2.11) are proved analogously.

(i) Proof of (2.10). Write the left side of (2.10) as

$$\begin{aligned} &n^{-1} \sum C_\vartheta(X_{i-1})^{-1} A_\vartheta(X_{i-1}) m'_\vartheta(X_{i-1}) \\ &+n^{-1} \sum \Big(C_{i-1}^{-1} A_{\vartheta,i-1} - C_\vartheta(X_{i-1})^{-1} A_\vartheta(X_{i-1})\Big) m'_\vartheta(X_{i-1}) \\ &+n^{-1} \sum C_{i-1}^{-1} \Big(A_{\vartheta_n,i-1} m'_{\vartheta_n}(X_{i-1}) - A_{\vartheta,i-1} m'_\vartheta(X_{i-1})\Big). \end{aligned}$$

By the ergodic theorem, the first of these three terms converges to $\pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta)$. We must show that the second and third terms are of order $o_{P_n}(1)$. For the second term, note that

$$A_{\vartheta i} - A_\vartheta(X_i) = m'_\vartheta(X_i)\Big(\mu_{4i} - \mu_4(X_i) - \mu_{2i}^2 + v_\vartheta(X_i)^2\Big) - v'_\vartheta(X_i)\Big(\mu_{3i} - \mu_3(X_i)\Big).$$

It follows easily from the Assumptions that the second term is bounded by an expression of the form $\varepsilon_n \sum r(X_{i-1})$, with r π-integrable. Hence the second term is of order $o_{P_n}(1)$. For the third term, we note that

$$\begin{aligned} A_{\vartheta_n i} - A_{\vartheta i} &= \Big(m'_{\vartheta_n}(X_i) - m'_\vartheta(X_i)\Big)(\mu_{4i} - \mu_{2i}^2) \\ &\quad -\Big(v'_{\vartheta_n}(X_i) - v'_\vartheta(X_i)\Big)\mu_{3i}. \end{aligned}$$

The assumptions imply as before that the third term is of order $o_{P_n}(1)$.

(ii) Proof of (2.8). Choose ϑ_{ni} between ϑ and ϑ_n such that

$$m_{\vartheta_n}(X_{i-1}) = m_\vartheta(X_{i-1}) + (\vartheta_n - \vartheta) m'_{\vartheta_{ni}}(X_{i-1}).$$

Write the left side of (2.8) as

$$\begin{aligned} &n^{-1/2} \sum C_{i-1}^{-1} A_{\vartheta_n,i-1}\Big(X_i - m_\vartheta(X_{i-1})\Big) \\ &-n^{1/2}(\vartheta_n - \vartheta) n^{-1} \sum C_{i-1}^{-1} A_{\vartheta_n,i-1} m'_{\vartheta_{ni}}(X_{i-1}). \end{aligned}$$

The second of these two terms is dealt with exactly as in part (i) of the proof:

$$n^{-1} \sum C_{i-1}^{-1} A_{\vartheta_n,i-1} m'_{\vartheta_{ni}}(X_{i-1}) = \pi(C_\vartheta^{-1} A_\vartheta m'_\vartheta) + o_{P_n}(1).$$

To prove (2.8), it remains to show that

$$n^{-1/2}\sum\Big(C_{i-1}^{-1}A_{\vartheta_n,i-1}-C_\vartheta(X_{i-1})^{-1}A_\vartheta(X_{i-1})\Big)\Big(X_i-m_\vartheta(X_{i-1})\Big)$$
$$=o_{P_n}(1). \tag{2.12}$$

Choose ϑ_{ni} between ϑ and ϑ_n such that

$$A_{\vartheta_n,i-1}=A_{\vartheta,i-1}+(\vartheta_n-\vartheta)A'_{\vartheta_{ni},i-1}.$$

Write the left side of (2.12) as

$$\begin{aligned}
&n^{-1/2}\sum\Big(C_{i-1}^{-1}A_{\vartheta,i-1}-C_\vartheta(X_{i-1})^{-1}A_\vartheta(X_{i-1})\Big)\Big(X_i-m_\vartheta(X_{i-1})\Big)\\
&\quad+(\vartheta_n-\vartheta)n^{-1/2}\sum C_{i-1}^{-1}A'_{\vartheta,i-1}\Big(X_i-m_\vartheta(X_{i-1})\Big)\\
&+n^{1/2}(\vartheta_n-\vartheta)n^{-1}\sum C_{i-1}^{-1}(A'_{\vartheta_{ni},i-1}-A'_{\vartheta,i-1})\Big(X_i-m_\vartheta(X_{i-1})\Big).
\end{aligned}$$

The first term has predictable quadratic variation

$$n^{-1}\sum\Big(C_{i-1}^{-1}A_{\vartheta,i-1}-C_\vartheta(X_{i-1})^{-1}A_\vartheta(X_{i-1})\Big)^2 v_\vartheta(X_{i-1}).$$

This is shown to be of order $o_{P_n}(1)$ as in part (i) of the proof. Hence by Lenglart's inequality (Jacod and Shiryaev, 1987, p. 35, Lemma 3.30a), the first term is of order $o_{P_n}(1)$. The second term is $\vartheta_n-\vartheta=o_{P_n}(1)$, multiplied by a term with predictable quadratic variation

$$n^{-1}\sum C_{i-1}^{-2}A'^2_{\vartheta,i-1}v_\vartheta(X_{i-1}).$$

This is of order $O_{P_n}(1)$. Hence the second term is also of order $o_{P_n}(1)$. The third term is $n^{1/2}(\vartheta_n-\vartheta)=O_{P_n}(1)$, multiplied by an expression which is again shown to be of order $o_{P_n}(1)$.

3 An autoregressive model

Let $X_0,\dots,X_n$ be observations from an autoregressive process

$$X_i=\vartheta X_{i-1}+\varepsilon_i\,,$$

where ε_i are i.i.d. with unknown density p having mean zero,

$$E\varepsilon=0. \tag{3.1}$$

Suppose that the error variance is related to the regression parameter,

$$E\varepsilon^2=v(\vartheta). \tag{3.2}$$

Then we have parametric models of the form (1.1) and (1.2) for the conditional mean and variance,

$$\int y p(y - \vartheta x) dy = \vartheta x, \tag{3.3}$$

$$\int (y - \vartheta x)^2 p(y - \vartheta x) dy = v(\vartheta). \tag{3.4}$$

Hence the model is a quasi-likelihood model. The results of Section 2 are, however, not directly applicable because of the special structure of the transition distribution,

$$Q(x, dy) = p(y - \vartheta x) dy. \tag{3.5}$$

This is an additional restriction besides (3.1) and (3.2). It also involves ϑ.

The maximum quasi-likelihood estimator solves

$$v(\vartheta)^{-1} \sum_{i=1}^{n} X_{i-1}(X_i - \vartheta X_{i-1}) = 0.$$

Hence it equals the *least squares estimator*

$$\vartheta_n = \sum_{i=1}^{n} X_{i-1} X_i \Big/ \sum_{i=1}^{n} X_{i-1}^2 .$$

It is well known that, in general, the least squares estimator does not even attain the variance bound for the usual autoregression model, without restriction (3.2) on the error variance. This differs from Section 2, where the maximum quasi-likelihood estimator was as good as the best estimator ignoring the corresponding restriction (2.2). The reason is that the least squares estimator fails to use not only the information about ϑ in (3.2), but also the information in (3.5). — For extended maximum quasi-likelihood estimators in an autoregressive model we refer to Heyde (1987).

We turn to the construction of an efficient estimator for ϑ. The arguments are similar to those in Section 2, and we will only sketch them. Fix a density p fulfilling (3.1) and (3.2).

Assumptions. The parameter varies in an open subset of $(-1, 1)$ on which the function v has a continuous and nonvanishing derivative v'. The density p is absolutely continuous with logarithmic derivative ℓ' and finite Fisher information

$$I^* = E\ell'(\varepsilon)^2.$$

The error distribution is nondegenerate and has finite fourth moment.

To prove local asymptotic normality, we introduce a local model as follows. Besides p, fix ϑ. Let K denote the set of all bounded functions $k(y)$ such that

$$\begin{aligned} Ek(\varepsilon) &= 0, \\ E\varepsilon k(\varepsilon) &= 0, \\ E\varepsilon^2 k(\varepsilon) &= v'(\vartheta). \end{aligned}$$

For $k \in K$ and $u \in \mathbb{R}$ define

$$p^{nuk}(y) = \Big(1 + n^{-1/2} u(k(y) + r_n(y))\Big) p(y).$$

As in Section 2 one can choose r_n of order $n^{-1/2}$ such that (3.1) and (3.2) hold for $p = p^{nuk}$ and $\vartheta = \vartheta + n^{-1/2}u$. Write P_n for the joint distribution of $X_0, \dots, X_n$ if p and ϑ are true, and P_n^{nuk} if p^{nuk} and $\vartheta + n^{-1/2}u$ are true. Similarly as in Huang (1986) or Kreiss (1987) one obtains local asymptotic normality,

$$\begin{aligned} \log dP_n^{nuk} \big/ dP_n &= u n^{-1/2} \sum_{i=1}^n \Big(-X_{i-1}\ell'(X_i - \vartheta X_{i-1}) + k(X_i - \vartheta X_{i-1})\Big) \\ &\quad -\frac{1}{2} u^2 \Big((1-\vartheta^2)^{-1} I^* v(\vartheta) + Ek(\varepsilon)^2\Big) + o_{P_n}(1), \end{aligned}$$

and

$$n^{-1/2} \sum_{i=1}^n \Big(-X_{i-1}\ell'(X_i - \vartheta X_{i-1}) + k(X_i - \vartheta X_{i-1})\Big) \Rightarrow N_k \qquad \text{under } P^n,$$

where N_k is normal with mean zero and variance equal to

$$(1-\vartheta^2)^{-1} I^* v(\vartheta) + Ek(\varepsilon)^2. \tag{3.6}$$

Hence the score functions are of the form

$$h(x,y) = -x\ell'(y - \vartheta x) + k(y - \vartheta x). \tag{3.7}$$

Next we determine the efficient score function and the information bound. Let $\overline{K}$ denote the closure of K in $L_2(p)$. The efficient score function minimizes (3.6) over $k \in \overline{K}$. Similarly as in Theorem 1 the minimum is attained if $k(y)$ equals

$$t(y) = C_\vartheta^{-1} v'(\vartheta) \Big(v(\vartheta) \big(y^2 - v(\vartheta)\big) - \mu_3 y \Big)$$

with

$$C_\vartheta = \Big(\mu_4 - v(\vartheta)^2\Big) v(\vartheta) - \mu_3^2 \quad \text{and} \quad \mu_j = E\varepsilon^j.$$

Since the error distribution is nondegenerate, $v(\vartheta)$ is positive. Hence C_ϑ is positive by the Schwarz inequality. The efficient score function is (3.7) for $k = t$,

$$s(x,y) = -x\ell'(y - \vartheta x) + t(y - \vartheta x). \tag{3.8}$$

The information bound is (3.6) for $k = t$,

$$I = (1-\vartheta^2)^{-1} I^* v(\vartheta) + C_\vartheta^{-1} v'(\vartheta)^2 v(\vartheta). \tag{3.9}$$

The information bound (3.9) is strictly larger than the bound $(1-\vartheta^2)^{-1} I^* v(\vartheta)$ in the usual autoregressive model, without restriction (3.2) on the error variance.

By the characterization stated in Section 2, an estimator for ϑ with influence function $I^{-1}s(x,y)$ is regular and efficient. To construct such an estimator, we recall some results for the usual autoregressive model, without restriction (3.2). For this model, Kreiss (1987) has introduced an efficient estimator ϑ_n^* for ϑ, with influence function

$$-(1-\vartheta^2)(I^*\mu_2)^{-1}x\ell'(y-\vartheta x). \tag{3.10}$$

Write $\varepsilon_{in} = X_i - \vartheta_n X_{i-1}$ for the estimated errors of the autoregressive model, with ϑ_n the least squares estimator. The moments μ_j of the error distribution are estimated by the empirical moments

$$\mu_{jn} = n^{-1}\sum_{i=1}^{n} \varepsilon_{in}^j.$$

In particular, μ_{1n} estimates zero. According to Wefelmeyer (1994a), the estimator

$$\mu_{2n}^* = \mu_{2n} - \mu_{2n}^{-1}\mu_{3n}\mu_{1n}$$

is efficient for μ_2 if μ_2 is not restricted by (3.2). The influence function of μ_{2n}^* is

$$(y-\vartheta x)^2 - \mu_2 - \mu_2^{-1}\mu_3(y-\vartheta x).$$

Wefelmeyer (1994a) treats only expectations of bounded functions. The result here follows by the usual truncation argument.

We return to the autoregressive model with restriction (3.2). Then $\mu_2 = v(\vartheta)$, and we have a new estimator for ϑ, namely $v^{-1}(\mu_{2n}^*)$. Both ϑ_n^* and $v^{-1}(\mu_{2n}^*)$ are not efficient in this smaller model. An efficient estimator is obtained as a random convex combination of the two estimators. The weight involves an estimator for I^*, say the estimator I_n^* of Kreiss (1987), and an estimator for C_ϑ, say

$$C_n = \Big(\mu_{4n} - v(\vartheta_n)^2\Big)v(\vartheta_n) - \mu_{3n}^2.$$

Theorem 3. *The estimator*

$$\hat{\vartheta}_n = a_n\vartheta_n^* + (1-a_n)v^{-1}(\mu_{2n}^*),$$

with

$$a_n = 1\Big/\Big(1 + (I_n^*C_n)^{-1}v'(\vartheta_n)^2\Big),$$

is regular and efficient for ϑ at p.

Proof. We show that $\hat{\vartheta}_n$ has influence function $f(x,y) = I^{-1}s(x,y)$, with s and I defined in (3.8) and (3.9), respectively. From (3.11) we obtain by Taylor expansion that $v^{-1}(\mu_{2n}^*)$ has influence function

$$v'(\vartheta)^{-1}\Big((y-\vartheta x)^2 - v(\vartheta) - v(\vartheta)^{-1}\mu_3(y-\vartheta x)\Big).$$

The influence function of ϑ_n^* is given in (3.10). The estimator a_n is consistent for

$$a = 1\Big/\Big(1 + (I^* C_\vartheta)^{-1} v'(\vartheta)^2\Big).$$

It follows that $\hat{\vartheta}_n$ has influence function

$$\begin{aligned}
&-a(1-\vartheta^2)\Big(I^* v(\vartheta)\Big)^{-1} x\ell'(y-\vartheta x) \\
&+(1-a)v'(\vartheta)^{-1}\Big((y-\vartheta x)^2 - v(\vartheta) - v(\vartheta)^{-1}\mu_3(y-\vartheta x)\Big) \\
&= I^{-1}\Big(-x\ell'(y-\vartheta x) + C_\vartheta^{-1} v'(\vartheta)\Big(v(\vartheta)\Big((y-\vartheta x)^2) - v(\vartheta)\Big) - \mu_3(y-\vartheta x)\Big)\Big) \\
&= I^{-1}\Big(-x\ell'(y-\vartheta x) + t(y-\vartheta x)\Big) \\
&= I^{-1} s(x,y).
\end{aligned}$$

Hence $\hat{\vartheta}_n$ is regular and efficient for ϑ at p.

Acknowledgments I would like to thank the referees and an Associate Editor for their suggestions, which improved an earlier version of the manuscript.

References

Amemiya, T. (1973). Regression analysis when the variance of the dependent variable is proportional to the square of its expectation. *J. Amer. Statist. Assoc.* **68** 928–934.

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

Bickel, P.J. (1993). Estimation in semiparametric models. In *Multivariate Analysis: Future Directions* (C.R. Rao, ed.) 55–73. North-Holland, Amsterdam.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Collomb, G. (1984). Propriétés de convergence presque complète de prédicteur à noyeau. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **66** 441–460.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* **2** 305–330.

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74** 591–597.

Firth, D. (1987). On the efficiency of quasi-likelihood estimation. *Biometrika* **74** 233–245.

Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72** 419–428.

Godambe. V.P. (1987). The foundations of finite sample estimation in stochastic processes – II. In *Proceedings of the 1st World Congress of the Bernoulli Society* (Yu. Prohorov and V.V. Sazonov, eds.) **2** 49–54. VNU Science Press, Utrecht.

Godambe, V.P., ed. (1991). *Estimating Functions.* Oxford University Press.

Godambe, V.P. and Heyde, C.C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation. *J. Statist. Plann. Inference* **22** 137–152.

Greenwood, P.E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* **36** 353–370.

Greenwood, P.E. and Wefelmeyer, W. (1992). Efficiency of empirical estimators for Markov chains. To appear in *Ann. Statist.*

Heyde, C.C. (1987). On combining quasi-likelihood estimating functions. *Stochastic Process. Appl.* **25** 281–287.

Hill, J.R. and Tsai, C.-L. (1988). Calculating the efficiency of maximum quasilikelihood estimation. *Appl. Statist.* **37** 219–230.

Hinkley, D.V., Reid, N. and Snell, E.J., eds. (1991). *Statistical Theory and Modelling. In Honour of Sir David Cox, FRS.* Chapman and Hall, London.

Höpfner, R. (1993). On statistics of Markov step processes: representation of log-likelihood ratio processes in filtered local models. *Probab. Theory Related Fields* **94** 375–398.

Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86** 105–129.

Huang, W.-M. (1986). A characterization of limiting distributions of estimators in an autoregressive process. *Ann. Inst. Statist. Math.* **38** 137–144.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233.

Huhtala, K. (1992). A quasi-likelihood Markov model for the hardenability of steel. In *Proceedings of the Sixth European Conference on Mathematics in Industry* (F. Hodnett, ed.) 191–194. Teubner, Stuttgart.

Hutton, J.E. and Nelson, P.I. (1986). Quasi-likelihood estimation for semimartingales. *Stochastic Process. Appl.* **22** 245–257.

Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

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

Kreiss, J.-P. (1987). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* 2nd ed. Chapman and Hall, London.

Penev, S. (1991). Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* **27** 105–123.

Roussas, G.G. (1965). Asymptotic inference in Markov processes. *Ann. Math. Statist.* **36** 987–992.

Thavaneswaran, A. and Thompson, M.E. (1986). Optimal estimation for semimartingales. *J. Appl. Probab.* **23** 409–417.

Truong, Y.K. and Stone, C.J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61** 439–447.

Wefelmeyer, W. (1993). Estimating functions and efficiency in a filtered model. In *Frontiers in Pure and Applied Probability* **1** (H. Niemi, G. Högnas, A.N. Shiryaev, A.V. Melnikov, eds.) 287–295. VSP, Utrecht.

Wefelmeyer, W. (1994a). An efficient estimator for the expectation of a bounded function under the residual distribution of an autoregressive process. *Ann. Inst. Statist. Math.* **46** 309–315.

Wefelmeyer, W. (1994b). Improving maximum quasi-likelihood estimators. In *Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 467–474. Physica-Verlag, Heidelberg.

Wefelmeyer, W. (1994c). Adaptive estimators for parameters of the autoregression function of a Markov chain. Preprint.

Welsh, A.H. (1989). On M-processes and M-estimation. *Ann. Statist.* **17** 337–361.

Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44** 1019–1031.