Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

QUASI-LIKELIHOOD REGRESSION MODELS FOR MARKOV CHAINS

Wolfgang Wefelmeyer
University of Siegen

Abstract

We consider regression models in which covariates and responses jointly form a higher order Markov chain. A quasi-likelihood model specifies parametric models for the conditional means and variances of the responses given the past observations. A simple estimator for the parameter is the maximum quasi-likelihood estimator. We show that it does not use the information in the model for the conditional variances, and construct an efficient estimating function which involves estimators for the third and fourth centered conditional moments of the responses. In many applications one assumes that the innovations are not arbitrary martingale increments but independently and identically distributed. We determine how much additional information about the parameter such an assumption contains. To make the exposition more readable, we first treat the case in which only the conditional mean is specified.

AMS 1991 subject classifications. Primary 62G20; secondary 62M05.

Key words and Phrases. Efficient estimator, quasi-likelihood, Markov chain, weighted nonlinear least squares, autoregression.

1 Introduction

Suppose we observe covariates X_i and responses Y_i which jointly form a homogeneous p-order Markov chain $Z_i = (X_i, Y_i)$. We write $Q(Z_{i-1}, \ldots, Z_{i-p}, dz)$ for its transition distribution, and for the conditional mean and variance of the response we write

$$\begin{aligned} m(z_{p-1}, \ldots, z_0) &= \int\int Q(z_{p-1}, \ldots, z_0, dx, dy)y, \\ v(z_{p-1}, \ldots, z_0) &= \int\int Q(z_{p-1}, \ldots, z_0, dx, dy)(y - m(z_{p-1}, \ldots, z_0))^2. \end{aligned}$$

If we have a parametric model $m = m_\vartheta$ for the conditional mean of the response, we can introduce a large class of martingale estimating functions

$$M_{\vartheta n} = \sum_{i=p}^{n} w_\vartheta(Z_{i-1}, \ldots, Z_{i-p})(Y_i - m_\vartheta(Z_{i-1}, \ldots, Z_{i-p})), \tag{1.1}$$

with $w_\vartheta(z_{p-1}, \ldots, z_0)$ an arbitrary *weight function.* The corresponding estimators for ϑ are defined as solutions of $M_{\vartheta n} = 0$. We indicate in Section 3 that an efficient estimator is obtained with the choice

$$w_\vartheta(Z_{i-1}, \ldots, Z_{i-p}) = \hat{v}_{i-1}(Z_{i-1}, \ldots, Z_{i-p})^{-1} \dot{m}_\vartheta(Z_{i-1}, \ldots, Z_{i-p}), \tag{1.2}$$

with $\hat{v}_{i-1}$ an estimator for v based on the observations up to time $i-1$. By efficiency we mean asymptotic optimality among all regular estimators in the sense of an appropriate version of Hájek's (1970) convolution theorem, not just optimality within some class of estimating functions. For the case of no covariates, a first-order chain and a one-dimensional parameter, a rigorous proof is given in Wefelmeyer (1996a). The estimating function is an adaptive version of the *quasi-score function.*

The model is described by all transition distributions Q which fulfill $m = m_\vartheta$ for some ϑ. It could be interpreted as a *semiparametric* model by writing

$$Q(z_{p-1}, \ldots, z_0, dx, dy) = M(z_{p-1}, \ldots, z_0, dx, dy - m_\vartheta(z_{p-1}, \ldots, z_0))$$

with $\int\int M(z_{p-1}, \ldots, z_0, dx, dy)y = 0$, and considering M as *nuisance parameter.*

In many applications one uses more specific models,

$$Y_i = m_\vartheta(Z_{i-1}, \ldots, Z_{i-p}) + \varepsilon_i,$$

where the ε_i are i.i.d. with mean zero and known or unknown distribution, rather than arbitrary martingale increments. We call such models *regression-autoregression models.* Again, $m = m_\vartheta$. We show that the specific structure contains additional information about ϑ, except when the ε_i are normal. The efficient estimators that have been constructed for specific such models are, however, not based on estimating functions.

If we have, in addition to $m = m_\vartheta$, a parametric model $v = v_\vartheta$ for the conditional variance of the response, with the same parameter ϑ, then the model is called a *quasi-likelihood model.* The best weight in (3.1) is then $w_\vartheta = v_\vartheta^{-1} \dot{m}_\vartheta$, giving the *maximum quasi-likelihood estimator.* It is as good as the estimator corresponding to the estimated weights (1.2). This implies that the maximum quasi-likelihood estimator does not use any of the information in the model assumption $v = v_\vartheta$. We also note that if the model $v = v_\vartheta$ is misspecified, then the weights (1.2) lead to a strictly better estimator.

In a quasi-likelihood model one can introduce further martingale estimating functions

$$\sum_{i=p}^{n} w_\vartheta(Z_{i-1}, \dots, Z_{i-p}) \Big((Y_i - m_\vartheta(Z_{i-1}, \dots, Z_{i-p}))^2 - v_\vartheta(Z_{i-1}, \dots, Z_{i-p}) \Big) .$$

We show in Section 4 that an appropriate combination with (3.1), with weights involving estimators for the conditional centered third and fourth moments of the response, gives an efficient estimator. For the case of no covariates, a first-order chain and a one-dimensional parameter, a rigorous proof is given in Wefelmeyer (1996b). The estimating function is an adaptive version of the *extended quasi-score function.* Recent reviews of quasi-likelihood methods are McCullagh (1991) and Firth (1993).

Again, in many applications one uses more specific models

$$Y_i = m_\vartheta(Z_{i-1}, \dots, Z_{i-p}) + v_\vartheta(Z_{i-1}, \dots, Z_{i-p})^{1/2} \varepsilon_i,$$

where the ε_i are i.i.d. with mean zero and variance one. We call such models *heteroscedastic regression-autoregression models* and show again that the specific structure contains additional information about ϑ.

We do not give precise regularity conditions for our results. They can be obtained by fairly straightforward, if tedious, modifications of Wefelmeyer (1996a, 1996b).

2 Notation

We observe k-dimensional covariates X_i and real-valued responses Y_i. We suppose that $Z_i = (X_i, Y_i)$ form a homogeneous and ergodic p-order Markov chain. For the p values of the chain preceding Z_i we write $\boldsymbol{Z}_{i-1} = (Z_{i-1}, \dots, Z_{i-p})'$. The chain starts with an initial value $\mathbf{Z}_{p-1} = (Z_{p-1}, \dots, Z_0)'$. For the transition distribution of Z_i given $\boldsymbol{Z}_{i-1} = \mathbf{t}$ we write $Q(\mathbf{t}, dz)$. Here and in the following, we will always write $z = (x, y)$ for the variables corresponding to the random variables $Z_i = (X_i, Y_i)$, and $t = (r, s)$ corresponding to $Z_{i-1} = (X_{i-1}, Y_{i-1})$. Boldface letters denote corresponding p-dimensional vectors, with components numbered backwards.

The conditional distribution of the response Y_i given the past depends only on the value $\boldsymbol{Z}_{i-1} = \mathbf{t}$ and is given by the marginal of the transition distribution of Z_i,

$$Q_r(\mathbf{t}, dy) = Q(\mathbf{t}, \mathbf{R}^k \times dy).$$

For the conditional mean and variance of the response given the past observations we write

$$m(\mathbf{t}) = \int Q_r(\mathbf{t}, dy)y,$$
$$v(\mathbf{t}) = \int Q_r(\mathbf{t}, dy)(y - m(\mathbf{t}))^2.$$

Let $\pi(d\mathbf{z})$ denote the stationary law of $\mathbf{Z_i}$. For the expectation of a function $f(\mathbf{Z_i})$ under π we write

$$\pi f = \int \pi(d\mathbf{z})f(\mathbf{z}).$$

Similarly, for the expectation of a function $f(\boldsymbol{Z}_{i-1}, Y_i)$ under the stationary law $\pi \otimes Q_r$ we write

$$\pi \otimes Q_r f = \int \int \pi(d\mathbf{t})Q_r(\mathbf{t}, dy)f(\mathbf{t}, y).$$

3 Modeling the conditional mean of the response

3.1. Estimating functions. Suppose we have a parametric model $m = m_\vartheta$ for the conditional mean of the response, where ϑ is a q-dimensional parameter. Recall that a large class of martingale estimating functions can be constructed as follows. Note that $Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})$ are martingale increments with respect to the filtration generated by the $\boldsymbol{Z}_i$. Choose a q-dimensional vector $w_\vartheta(\mathbf{t})$ of *weight functions.* Then $w_\vartheta(\boldsymbol{Z}_{i-1})$ is predictable, and the components of the vector $w_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))$ are again martingale increments, so that the *estimating functions*

$$M_{\vartheta n} = \sum_{i=p}^{n} w_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \tag{3.1}$$

form a martingale. An estimator is obtained as solution $\vartheta = \hat{\vartheta}_n$ of the *estimating equation* $M_{\vartheta n} = 0$. We do not give conditions for existence and uniqueness here.

Call an estimator T_n for ϑ *asymptotically linear* with *influence function* $f(\mathbf{t}, \mathbf{y})$ if

$$n^{1/2}(T_n - \vartheta) = n^{-1/2} \sum_{i=p}^{n} f(\boldsymbol{Z}_{i-1}, Y_i) + o_P(1)$$

and $\int Q_r(\mathbf{t}, dy)f(\mathbf{t}, y) = 0$ for all $\mathbf{t}$. Then the components of the vector $f(\boldsymbol{Z}_{i-1}, Y_i)$ are martingale increments. If the components of f are $\pi \otimes Q_r$-square integrable, a martingale central limit theorem holds, and T_n is asymptotically normal with covariance matrix $\pi \otimes Q_r ff'$. See, e.g., Billingsley (1968, p. 206).

Let us recall how one shows that the solution $\vartheta = \hat{\vartheta}_n$ of the estimating equation $M_{\vartheta n} = 0$ is asymptotically linear. We use a dot on top of a vector of functions to denote the matrix of partial derivatives with respect to ϑ. A Taylor expansion gives

$$0 = M_{\hat{\vartheta}_n n} = M_{\vartheta n} + \dot{M}_{\vartheta n}(\hat{\vartheta}_n - \vartheta) + \cdots$$

with matrix of partial derivatives

$$\dot{M}_{\vartheta n} = \sum_{i=p}^{n} \dot{w}_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) - \sum_{i=p}^{n} w_\vartheta(\boldsymbol{Z}_{i-1})\dot{m}_\vartheta(\boldsymbol{Z}_{i-1}).$$

Note that $\dot{m}_\vartheta$ is a *row* vector. Since the entries of the matrix $\sum_{i=p}^{n} \dot{w}_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))$ are mean zero martingales, $\frac{1}{n}\sum_{i=p}^{n} \dot{w}_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))$ is negligible if the entries of the matrix $\dot{w}_\vartheta(\mathbf{t})(y - m_\vartheta(\mathbf{t}))$ are $\pi \otimes Q_r$-square integrable. Furthermore,

$$\frac{1}{n}\sum_{i=p}^{n} w_\vartheta(\boldsymbol{Z}_{i-1})\dot{m}_\vartheta(\boldsymbol{Z}_{i-1}) \to \pi w_\vartheta \dot{m}_\vartheta.$$

The above arguments show that $\hat{\vartheta}_n$ has influence function

$$f(\mathbf{t}, y) = (\pi w_\vartheta \dot{m}_\vartheta)^{-1} w_\vartheta(\mathbf{t})(y - m_\vartheta(\mathbf{t})) \tag{3.2}$$

and asymptotic covariance matrix

$$(\pi w_\vartheta \dot{m}_\vartheta)^{-1} \pi v w_\vartheta w'_\vartheta \, (\pi \dot{m}'_\vartheta w'_\vartheta)^{-1}. \tag{3.3}$$

For dependent observations, weak conditions for asymptotic linearity of estimators may be found in Hosoya (1989), Andrews and Pollard (1994) and Andrews (1994).

Remark 1. We have restricted attention to weights $w_\vartheta(\boldsymbol{Z}_{i-1})$ which depend only on the p previous observations $Z_{i-1}, \dots, Z_{i-p}$ of the p-order Markov chain. Let us show that there is no point in using weights w_ϑ which depend on observations preceding Z_{i-p}. Note first that with such weights we would also get a covariance matrix of the form (3.3), with π now denoting the stationary law of more than p successive observations. Let $\overline{w}_\vartheta$ denote the weight function in which the additional arguments appearing in w_ϑ have been integrated out. Then $\pi \overline{w}_\vartheta \dot{m}_\vartheta$ equals $\pi w_\vartheta \dot{m}_\vartheta$. By Jensen's ineqality, $\pi v w_\vartheta w'_\vartheta - \pi v \overline{w}_\vartheta \overline{w}'_\vartheta$ is positive semi-definite. Hence

$$(\pi w_\vartheta \dot{m}_\vartheta)^{-1} \pi v w_\vartheta w'_\vartheta \, (\pi \dot{m}'_\vartheta w'_\vartheta)^{-1} - (\pi \overline{w}_\vartheta \dot{m}_\vartheta)^{-1} \pi v \overline{w}_\vartheta \overline{w}'_\vartheta \, (\pi \dot{m}'_\vartheta \overline{w}'_\vartheta)^{-1}$$

is positive semi-definite. □

3.2. Known conditional variance. Suppose, for the moment, that the conditional variance v of the response is *known*. Then we can determine a weight function which is *optimal* in the sense that it minimizes the asymptotic covariance matrix (3.3). Recall (e.g., Horn and Johnson, 1985, p. 472) that if a block matrix $\begin{pmatrix} A & B \\ B' & C \end{pmatrix}$ is symmetric and positive definite, so is $C - B'A^{-1}B$ and hence also $B'^{-1}CB^{-1} - A^{-1}$. Applying this result to the covariance matrix of $(v^{-1/2}\dot{m}'_\vartheta, v^{1/2}w_\vartheta)'$ under π, we see that

$$(\pi w_\vartheta \dot{m}_\vartheta)^{-1}\, \pi v w_\vartheta w'_\vartheta\, (\pi \dot{m}'_\vartheta w'_\vartheta)^{-1} - (\pi v^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1} \quad \text{is positive semi-definite.}$$

This means that the covariance matrix (3.3) is minimized for

$$w_\vartheta = v^{-1}\dot{m}'_\vartheta, \tag{3.4}$$

and that the minimal covariance matrix is

$$(\pi v^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1}. \tag{3.5}$$

This result is well known in the context of quasi-likelihood models; see Subsection 4.1. The influence function (3.2) of the estimator corresponding to the optimal weight $w_\vartheta = v^{-1}\dot{m}'_\vartheta$ is

$$f(\mathbf{t}, y) = (\pi v^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1}\, v(\mathbf{t})^{-1}\dot{m}_\vartheta(\mathbf{t})'(y - m_\vartheta(\mathbf{t})). \tag{3.6}$$

3.3. An adaptive estimating function. If v is not known, we can construct an estimating function which is *adaptive* in the sense that for each v it is asymptotically as good as the best estimating function (3.1) for *known* v, with weight $w_\vartheta = v^{-1}\dot{m}'_\vartheta$. It suffices to replace the conditional variance v by an estimator; compare Wefelmeyer (1996a). Specifically, let $\hat{v}_{i-1}(\mathbf{t})$ be estimators for $v(\mathbf{t})$ based only on the observations $Z_0, \ldots, Z_{i-1}$ up to time $i-1$. For the construction of such estimators see, e.g., Collomb (1984) and Truong and Stone (1992). We obtain the *adaptive estimating function*

$$\sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\dot{m}_\vartheta(\boldsymbol{Z}_{i-1})'(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})). \tag{3.7}$$

This estimating function is an adaptive version of the *quasi-score function* discussed in Subsection 4.1. Since the weight is predictable, the estimating function is again a martingale. If the $\hat{v}_{i-1}$ are strongly consistent, a Taylor expansion as in Subsection 3.1 shows that the corresponding estimator is asymptotically normal, its asymptotic covariance matrix is again (3.5), and its influence function is again (3.6). For the case of a one-dimensional parameter, and when there are no covariates, a rigorous proof is given in Wefelmeyer (1996a).

3.4. Efficiency of the adaptive estimating function. Does the adaptive estimating function (3.7) lead to an *efficient* estimator? In other words, is this estimator optimal not only among estimators based on estimating functions of the form (3.1), but also in the larger class of *regular* estimators? To answer this question, we must indicate that the model given by $m = m_\vartheta$ is *locally asymptotically normal* in an appropriate sense, and determine a bound for the asymptotic covariance matrices of regular estimators of ϑ in the sense of a convolution theorem. The basic reference for this theory in the i.i.d. case is Bickel et al. (1993). For the case of a one-dimensional parameter, and when there are no covariates, a rigorous proof of the efficiency of the adaptive estimating function is in Wefelmeyer (1996a). To accomodate covariates, we recall that by Cox (1972) the likelihood factors into two terms. The first is the *partial likelihood* and depends only on the conditional law $Q_r(\mathbf{t}, dy)$ of the responses. The second depends only on the conditional law of the covariates given the past observations *and* the present responses. Our model $m = m_\vartheta$ is a condition on Q_r only. Hence the second factor of the likelihood varies independently of ϑ. This means that the bound for the asymptotic covariance matrices can be determined from the *partial* likelihood.

Fix $Q_r(\mathbf{t}, dy)$. The model is described by a parametric family of side conditions $m = m_\vartheta$. To introduce a local model, we perturb $Q_r(\mathbf{t}, dy)$ such that the perturbed transition distribution is still in the model. This means that a perturbed condition $m = m_\vartheta$, with ϑ replaced by $\vartheta + n^{-1/2}u$, say, holds. Such perturbations are conveniently described as follows. Consider the affine space of q-dimensional vectors $h(\mathbf{t}, y)$ of functions with

$$\int Q_r(\mathbf{t}, dy)h(\mathbf{t}, y) = 0, \tag{3.8}$$

$$\int Q_r(\mathbf{t}, dy)yh(\mathbf{t}, y) = \dot{m}_\vartheta(\mathbf{t})'. \tag{3.9}$$

These vectors will play the role of *score functions*. Set

$$Q_r^{nhu}(\mathbf{t}, dy) = Q_r(\mathbf{t}, dy)(1 + n^{-1/2}h(\mathbf{t}, y)'u). \tag{3.10}$$

Then

$$\begin{aligned}\int Q_r^{nhu}(\mathbf{t}, dy)y &= m_\vartheta(\mathbf{t}) + n^{-1/2}\int Q_r(\mathbf{t}, dy)yh(\mathbf{t}, y)'u \\ &= m_\vartheta(\mathbf{t}) + n^{-1/2}\dot{m}_\vartheta(\mathbf{t})u \\ &= m_{\vartheta+n^{-1/2}u}(\mathbf{t}) + o(n^{-1/2}).\end{aligned}$$

This shows that Q_r^{nhu} is indeed (approximately) in the model. The *partial likelihood ratio* is

$$\overline{Z}^{nhu} = \prod_{i=p}^{n} \frac{dQ_r^{nhu}(\boldsymbol{Z}_{i-1}, \cdot)}{dQ_r(\boldsymbol{Z}_{i-1}, \cdot)}(Y_i).$$

By a Taylor expansion, the partial likelihood ratio is shown to be *locally asymptotically normal*,

$$\log \overline{Z}^{nhu} = n^{-1/2} \sum_{i=p}^{n} h(\boldsymbol{Z}_{i-1}, Y_i)'u - \frac{1}{2} u'\pi \otimes Q_r hh'u + o_P(1),$$

with $n^{-1/2} \sum_{i=p}^{n} h(\boldsymbol{Z}_{i-1}, Y_i)$ asymptotically normal with mean zero and covariance matrix $\pi \otimes Q_r hh'$. By the convolution theorem, an estimator is regular and efficient if and only if it is asymptotically linear with influence function $\Sigma^{-1}s$, where $\Sigma = \pi \otimes Q_r ss'$ and s is the *efficient score function*, minimizing $\pi \otimes Q_r hh'$ over the affine space of vectors h fulfilling (3.8) and (3.9). It is characterized by $\pi \otimes Q_r sh' = \Sigma$ for all h. It is straightforward to check that the solution is

$$s(\mathbf{t}, y) = v(\mathbf{t})^{-1} \dot{m}_\vartheta(\mathbf{t})(y - m_\vartheta(\mathbf{t})). \tag{3.11}$$

Hence $\Sigma = \pi v^{-1} \dot{m}_\vartheta' \dot{m}_\vartheta$, so that the efficient influence function is (3.6). In particular, the minimal asymptotic covariance matrix for regular estimators of ϑ is (3.5). The estimator based on the adaptive estimating function (3.7) also has influence function (3.6) and is therefore efficient.

3.5. Regression-autoregression models. Suppose that the responses have an autoregressive structure,

$$Y_i = m_\vartheta(\boldsymbol{Z}_{i-1}) + \varepsilon_i,$$

where the ε_i are i.i.d. with known or unknown mean zero density $g(y)$. Then the conditional distribution of the response Y_i given $\boldsymbol{Z}_{i-1} = \mathbf{t}$ has the form

$$Q_r(\mathbf{t}, dy) = g(y - m_\vartheta(\mathbf{t}))dy, \tag{3.12}$$

with conditional mean $m_\vartheta(\mathbf{t})$. We call it a *regression-autoregression model.* It is a submodel of the model given by $m = m_\vartheta$. Conditions for (geometric) ergodicity are given in Bhattacharya and Lee (1995). The question arises whether in this submodel there are even better estimators than the one based on the adaptive estimating function (3.7).

We show that the minimal asymptotic covariance matrix of regular estimators of ϑ is, in general, strictly smaller than (3.5). The regression-autoregression model (3.12) is a semiparametric model, with nuisance parameter g. The local model can be obtained by perturbing ϑ and g. Consider the linear space of functions $k(y)$ with

$$\mathrm{E}\, k(\varepsilon) = 0, \tag{3.13}$$

$$\mathrm{E}\, \varepsilon k(\varepsilon) = 0. \tag{3.14}$$

Then $g^{nk}(y) = g(y)(1 + n^{-1/2}k(y))$ is again a mean zero probability density. Set

$$Q_r^{nku}(\mathbf{t}, dy) = g^{nk}(y - m_{\vartheta + n^{-1/2}u}(\mathbf{t}))(dy).$$

Write ℓ' for the logarithmic derivative g'/g of g. By a Taylor expansion,

$$\begin{aligned} &Q_r^{nku}(\mathbf{t}, dy) \\ &= Q_r(\mathbf{t}, dy)\Big(1 + n^{-1/2}(k(y - m_\vartheta(\mathbf{t})) - \dot{m}_\vartheta(\mathbf{t})u\, \ell'(y - m_\vartheta(\mathbf{t})))\Big) \\ &\qquad + o(n^{-1/2}). \end{aligned}$$

The perturbation is seen to be (approximately) of the form (3.10), with $h(\mathbf{t}, y)'u$ replaced by $k(y - m_\vartheta(\mathbf{t})) - \dot{m}_\vartheta(\mathbf{t})u\, \ell'(y - m_\vartheta(\mathbf{t}))$. Hence the corresponding partial likelihood ratio is locally asymptotically normal with variance

$$\begin{aligned} &\int\!\!\int \pi(d\mathbf{t})g(y - m_\vartheta(\mathbf{t}))dy(k(y - m_\vartheta(\mathbf{t})) - \dot{m}_\vartheta(\mathbf{t})u\, \ell'(y - m_\vartheta(\mathbf{t})))^2 \\ &= \int\!\!\int \pi(d\mathbf{t})g(y)dy\, (k(y) - \dot{m}_\vartheta(\mathbf{t})u\, \ell'(y))^2 . \end{aligned} \tag{3.15}$$

For the parametric case, g known and hence $k = 0$, see Hwang and Basawa (1993, 1994). For the semiparametric case considered here, see Koul and Schick (1996). These references do not consider covariates.

To simplify the calculations, we will now assume that ϑ and g are locally orthogonal in the sense that the mixed term in the variance (3.15) vanishes, or equivalently,

$$\mathrm{E}\, k(\varepsilon)\ell'(\varepsilon) = 0 \quad \text{for all } k, \quad \text{or} \quad \pi\dot{m}_\vartheta = 0. \tag{3.16}$$

The first condition is fulfilled if the density of ε is assumed symmetric. Then ℓ' is odd, and since both g and g^{nk} are symmetric, k must be even. The second holds in many applications; see also Examples 1 and 2 below. If (3.16) holds, then we can estimate ϑ asymptotically as well not knowing g as knowing g. We say that the model is *adaptive* with respect to g. We refer to Drost et al. (1994) and Drost and Klaassen (1995) for a discussion of adaptivity for general semiparametric GARCH models. Under (3.16), the variance (3.15) reduces to

$$\mathrm{E}\, k(\varepsilon)^2 + \mathrm{E}\, \ell'(\varepsilon)^2\, u'\pi\dot{m}_\vartheta'\dot{m}_\vartheta u,$$

and the efficient score function, as defined at the end of Subsection 3.4, is

$$s(\mathbf{t}, y) = -\dot{m}_\vartheta(\mathbf{t})'\ell'(y - m_\vartheta(\mathbf{t})).$$

Hence $\Sigma = \mathrm{E}\,\ell'(\varepsilon)^2\,\pi\dot{m}_\vartheta'\dot{m}_\vartheta$, the efficient influence function is

$$f(\mathbf{t}, y) = -(\mathrm{E}\,\ell'(\varepsilon)^2)^{-1}\,(\pi\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1}\,\dot{m}_\vartheta(\mathbf{t})'\ell'(y - m_\vartheta(\mathbf{t})),$$

and the minimal asymptotic covariance matrix of regular estimators of ϑ is

$$(\mathrm{E}\,\ell'(\varepsilon)^2)^{-1}\,(\pi\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1}.$$

Of course, this covariance matrix cannot be larger than the minimal asymptotic covariance matrix (3.5) for the larger model $m = m_\vartheta$. To check this, note first that in the regression-autoregression model we have

$$v(\mathbf{t}) = \int g(y - m_\vartheta(\mathbf{t}))dy\,(y - m_\vartheta(\mathbf{t}))^2 = \mathrm{E}\,\varepsilon^2.$$

Hence (3.5) is $\mathrm{E}\,\varepsilon^2\,(\pi\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1}$. To prove the desired inequality, it suffices to recall that $\mathrm{E}\,\ell'(\varepsilon)^2$ is the Fisher information for location, and that its inverse is not larger than $\mathrm{E}\,\varepsilon^2$. Hence

$$\mathrm{E}\,\varepsilon^2\,(\pi\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1} - (\mathrm{E}\,\ell'(\varepsilon)^2)^{-1}\,(\pi\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1} \quad \text{is positive semi-definite.}$$

We note that the difference between the two matrices is proportional to the difference between the asymptotic variance $\mathrm{E}\,\varepsilon^2$ of the empirical estimator for the mean of g and the asymptotic variance $(\mathrm{E}\,\ell'(\varepsilon)^2)^{-1}$ of the maximum likelihood estimator for the mean in the location model generated by g . The inequality is strict unless $\ell'(y)$ is proportional to y. In particular, for normal ε_i, the adaptive estimating function (3.7) gives an efficient estimator in the regression-autoregression model.

To summarize: The regression-autoregression model is a quasi-likelihood model with the additional restriction that the conditional law of the response does not depend on the past except through the mean. The additional restriction can be exploited to construct an estimator with asymptotic covariance matrix reduced by the factor $(\mathrm{E}\,\ell'(\varepsilon)^2\mathrm{E}\,\varepsilon^2)^{-1}$ as compared to the adaptive estimating function. The reduction can be considerable if the density g is far from normal. On the negative side, the construction requires estimating the logarithmic derivative ℓ' of g, see Koul and Schick (1996) when there are no covariates, and the estimator is inconsistent if in reality the additional restriction does not hold.

Example 1. Set $m_\vartheta(\boldsymbol{Z}_{i-1}) = \vartheta'\boldsymbol{Y}_{i-1}$. Then the conditional mean of the response does not depend on the covariates, but the conditional variance may still depend on them. An efficient estimating function is (3.7); here it has the form

$$\sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\boldsymbol{Y}_{i-1}(Y_i - \vartheta'\boldsymbol{Y}_{i-1}).$$

It gives the weighted least squares estimator

$$\hat{\vartheta}_n = \left(\sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}\right)^{-1} \sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}Y_i\boldsymbol{Y}_{i-1}.$$

The corresponding regression-autoregression model is the *p-order autoregression model*

$$Y_i = \vartheta'\boldsymbol{Y}_{i-1} + \varepsilon_i,$$

where the ε_i are i.i.d. with mean zero density g. Here $v(\boldsymbol{Z}_{i-1}) = \mathrm{E}\,\varepsilon^2$ does not depend on the observations, and the weighted least squares estimator reduces to the ordinary least squares estimator

$$\hat{\vartheta}_n = \left(\sum_{i=p}^{n} \boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}\right)^{-1} \sum_{i=p}^{n} Y_i\boldsymbol{Y}_{i-1}.$$

It is not efficient in the autoregression model unless the ε_i are normal.

Huang (1986) proves local asymptotic normality of the autoregression model. An efficient estimator is constructed by Kreiss (1987a) for symmetric g, and by Kreiss (1987b) for arbitrary mean zero g. □

Example 2. Set $m_\vartheta(\boldsymbol{Z}_{i-1}) = \alpha'\boldsymbol{X}_{i-1} + \beta'\boldsymbol{Y}_{i-1}$. Then $\vartheta = (\alpha, \beta)'$ is of dimension $q = k + p$. Write $\boldsymbol{S}_{i-1} = (\boldsymbol{X}_{i-1}, \boldsymbol{Y}_{i-1})$. An efficient estimating function is (3.7); here it has the form

$$\sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\boldsymbol{S}_{i-1}(Y_i - \vartheta'\boldsymbol{S}_{i-1}).$$

It gives the weighted least squares estimator

$$\hat{\vartheta}_n = \left(\sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\boldsymbol{S}_{i-1}\boldsymbol{S}'_{i-1}\right)^{-1} \sum_{i=p}^{n} \hat{v}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}Y_i\boldsymbol{S}_{i-1}.$$

The corresponding regression-autoregression model is the *p-order autoregression model* with k-dimensional *linear regression trend*

$$Y_i = \alpha' X_{i-1} + \beta'\boldsymbol{Y}_{i-1} + \varepsilon_i = \vartheta'\boldsymbol{S}_{i-1} + \varepsilon_i,$$

where the ε_i are i.i.d. with mean zero density g. As in Example 1, $v(\boldsymbol{Z}_{i-1}) = \mathrm{E}\,\varepsilon^2$ does not depend on the observations, and the weighted least squares estimator reduces to the ordinary least squares estimator

$$\hat{\vartheta}_n = \left(\sum_{i=p}^{n} \boldsymbol{S}_{i-1}\boldsymbol{S}_{i-1}'\right)^{-1} \sum_{i=p}^{n} Y_i \boldsymbol{S}_{i-1}.$$

Swensen (1985) proves local asymptotic normality for the case of nonrandom X_i. See Garel and Hallin (1995) for a recent more general version and references. □

4 Quasi-likelihood models

4.1. The quasi-score function. A *quasi-likelihood model* is given by parametric models $m = m_\vartheta$ and $v = v_\vartheta$ for the conditional mean and variance of the response, with ϑ a common q-dimensional parameter. Consider again the estimating functions (3.1),

$$M_{\vartheta n} = \sum_{i=p}^{n} w_\vartheta(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})).$$

Exactly as in the case of a *known* conditional variance v, Subsection 3.2, the best weight is determined as $w_\vartheta = v_\vartheta^{-1}\dot{m}_\vartheta'$. It gives the *quasi-score function*

$$\sum_{i=p}^{n} v_\vartheta(\boldsymbol{Z}_{i-1})^{-1}\dot{m}_\vartheta(\boldsymbol{Z}_{i-1})'(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})). \tag{4.1}$$

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 corresponding estimator is the *maximum quasi-likelihood estimator.* Its asymptotic covariance matrix is (3.5) with $v = v_\vartheta$,

$$(\pi v_\vartheta^{-1}\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1},$$

the inverse of the *quasi-Fisher information matrix.* Its influence function is (3.6) with $v = v_\vartheta$,

$$f(\mathbf{t}, y) = (\pi v_\vartheta^{-1}\dot{m}_\vartheta'\dot{m}_\vartheta)^{-1}\, v_\vartheta(\mathbf{t})^{-1}\dot{m}_\vartheta(\mathbf{t})'(y - m_\vartheta(\mathbf{t})).$$

The quasi-score function is asymptotically as good as the adaptive estimating function (3.7). This implies that it does not use any of the information in the model assumption $v = v_\vartheta$.

By the arguments of Subsection 3.1, the quasi-score function can be used even if the model is not true. In this sense it is *robust* against misspecification of the conditional variance of the response. If the true conditional variance is v, then by (3.3) for $w_\vartheta = v_\vartheta^{-1}\dot{m}'_\vartheta$ the maximum quasi-likelihood estimator has asymptotic covariance matrix

$$(\pi v_\vartheta^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1}\ \pi v v_\vartheta^{-2}\dot{m}'_\vartheta \dot{m}_\vartheta\ (\pi v_\vartheta^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1}.$$

However, unless $v = v_\vartheta$, this covariance matrix is strictly larger than the covariance matrix $(\pi v^{-1}\dot{m}'_\vartheta \dot{m}_\vartheta)^{-1}$ which is attained by the estimator based on the adaptive estimating equation.

4.2. Further estimating functions. Note that $(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})$ are martingale increments with respect to the filtration generated by the $\boldsymbol{Z}_i$. We obtain martingale estimating functions

$$M_{\vartheta n} = \sum_{i=p}^{n} w_\vartheta(\boldsymbol{Z}_{i-1})\Big((Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})\Big) \tag{4.2}$$

which we can combine with estimating functions (3.1) to get estimating functions of the form

$$\begin{aligned}&\sum_{i=p}^{n} w_{1\vartheta}(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \\ &+\sum_{i=p}^{n} w_{2\vartheta}(\boldsymbol{Z}_{i-1})\Big((Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})\Big).\end{aligned} \tag{4.3}$$

It will be convenient to introduce the $q\times 2$ matrix of weights $\mathbf{w}_\vartheta = (w_{1\vartheta}, w_{2\vartheta})$ and the two-dimensional vector of martingale increments

$$\mathbf{i}_\vartheta(\mathbf{t}, y) = \Big(y - m_\vartheta(\mathbf{t}), (y - m_\vartheta(\mathbf{t}))^2 - v_\vartheta(\mathbf{t})\Big)',$$

and to rewrite the estimating function (4.3) as

$$\sum_{i=p}^{n} \mathbf{w}_\vartheta(\boldsymbol{Z}_{i-1})\mathbf{i}_\vartheta(\boldsymbol{Z}_{i-1}, Y_i).$$

We also introduce the $2 \times q$ matrix of derivatives $\mathbf{d}_\vartheta = (\dot{m}_\vartheta, \dot{v}_\vartheta)'$. For the conditional centered third and fourth moments of the response we write

$$\mu_j(\mathbf{t}) = \int Q_r(\mathbf{t}, dy)(y - m_\vartheta(\mathbf{t}))^j, \quad j = 3, 4.$$

The conditional covariance matrix of the martingale increments $\mathbf{i}_\vartheta$ is

$$C = \begin{pmatrix} v_\vartheta & \mu_3 \\ \mu_3 & \mu_4 - v_\vartheta^2 \end{pmatrix}. \tag{4.4}$$

As in Subsection 3.1, the estimator corresponding to the estimating equation (4.2) is shown to be asymptotically linear, with influence function

$$f(\mathbf{t},y) = (\pi \mathbf{w}_\vartheta \mathbf{d}_\vartheta)^{-1} \mathbf{w}_\vartheta(\mathbf{t}) \mathbf{i}_\vartheta(\mathbf{t},y)$$

and asymptotic covariance matrix

$$(\pi \mathbf{w}_\vartheta \mathbf{d}_\vartheta)^{-1} \pi \mathbf{w}_\vartheta C \mathbf{w}_\vartheta' (\pi \mathbf{d}_\vartheta' \mathbf{w}_\vartheta')^{-1}. \tag{4.5}$$

4.3. Known conditional centered third and fourth moments. Suppose, for the moment, that we *know* the conditional centered third and fourth moments μ_3 and μ_4 of the response. The weights $w_{1\vartheta}$ and $w_{2\vartheta}$ which minimize the asymptotic covariance matrix (4.5) are

$$\mathbf{w}_\vartheta = \mathbf{d}_\vartheta' C^{-1}, \tag{4.6}$$

and the minimal asymptotic covariance matrix is

$$(\pi \mathbf{d}_\vartheta' C^{-1} \mathbf{d}_\vartheta)^{-1}. \tag{4.7}$$

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 of conditionally orthogonal martingale increments, i.e. $\mu_3 = 0$. The general case, also for continuous time, is treated in Heyde (1987). A different derivation may be found in Kessler (1995). The influence function of the estimator corresponding to the optimal weight is

$$f(\mathbf{t},y) = (\pi \mathbf{d}_\vartheta' C^{-1} \mathbf{d}_\vartheta)^{-1} \mathbf{d}_\vartheta(\mathbf{t})' C(\mathbf{t})^{-1} \mathbf{i}_\vartheta(\mathbf{t},y). \tag{4.8}$$

4.4. An extended adaptive estimating function. If the conditional centered third and fourth moments μ_3 and μ_4 of the response are not known, we can construct an estimating function which is *adaptive* in the sense that for each μ_3 and μ_4 it is asymptotically as good as the best estimating function (4.3) for *known* μ_3 and μ_4, with weight (4.6). Similarly as in Subsection 3.3, replace, in (4.6), the matrix $C(\mathbf{t})$ by an estimator $\hat{C}_{i-1}(\mathbf{t})$, using estimators $\hat{\mu}_{j,i-1}(\mathbf{t})$ for $\mu_j(\mathbf{t})$ based on the observations $Z_0, \ldots, Z_{i-1}$. This gives the *extended adaptive estimating function*

$$\sum_{i=p}^{n} \mathbf{d}_\vartheta(\boldsymbol{Z}_{i-1})' \hat{C}(\boldsymbol{Z}_{i-1})^{-1} \mathbf{i}_\vartheta(\boldsymbol{Z}_{i-1}). \tag{4.9}$$

The estimating function is an adaptive version of the extended quasi-score function discussed in Remark 4 below. It gives an estimator whose influence function is (4.8).

The extended adaptive estimating function can be written more explicitly. Estimate the determinant of $C(\boldsymbol{Z}_{i-1})$ by

$$\hat{D}_{i-1}(\boldsymbol{Z}_{i-1}) = v_\vartheta(\boldsymbol{Z}_{i-1})\left(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - v_\vartheta(\boldsymbol{Z}_{i-1})^2\right) - \hat{\mu}_{3,i-1}(\boldsymbol{Z}_{i-1})^2,$$

and write the estimating function as

$$\begin{aligned}
&\sum_{i=p}^{n} \hat{D}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\left(\left(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - v_\vartheta(\boldsymbol{Z}_{i-1})^2\right)\right. \\
&\dot{m}_\vartheta(\boldsymbol{Z}_{i-1})' - \hat{\mu}_{3,i-1}(\boldsymbol{Z}_{i-1})\dot{v}_\vartheta(\boldsymbol{Z}_{i-1})')\,(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \\
&\sum_{i=p}^{n} \hat{D}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \\
&+\sum_{i=p}^{n} \hat{D}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\left(v_\vartheta(\boldsymbol{Z}_{i-1})\dot{v}_\vartheta(\boldsymbol{Z}_{i-1})' - \hat{\mu}_{3,i-1}(\boldsymbol{Z}_{i-1})\dot{m}_\vartheta(\boldsymbol{Z}_{i-1})'\right) \\
&\qquad\qquad \times \left((Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})\right).
\end{aligned}$$

In the important special case of orthogonal martingale increments, $\mu_3 = 0$, the extended adaptive estimating function can be replaced by the simpler version

$$\begin{aligned}
&\sum_{i=p}^{n} v_\vartheta(\boldsymbol{Z}_{i-1})^{-1}\dot{m}_\vartheta(\boldsymbol{Z}_{i-1})'(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \\
&+\sum_{i=p}^{n}\left(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - v_\vartheta(\boldsymbol{Z}_{i-1})^2\right)^{-1}\dot{v}_\vartheta(\boldsymbol{Z}_{i-1})' \\
&\left((Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})\right).
\end{aligned}$$

4.5. Efficiency of the extended adaptive estimating function. To show that the extended adaptive estimating function (4.9) leads to an efficient estimator, we must determine the lower bound for the asymptotic covariance matrices of regular estimators of ϑ. We follow the arguments in Subsection 3.4, adding the model assumption $v = v_\vartheta$. For the case of a one-dimensional parameter, and when there are no covariates, a rigorous proof is in Wefelmeyer (1996b). We perturb Q_r^{nhu} as in (3.10), with h fulfilling (3.8) and (3.9) and also

$$\int Q_r(\mathbf{t}, dy)(y - m_\vartheta(\mathbf{t}))^2 h(\mathbf{t}, y) = \dot{v}_\vartheta(\mathbf{t})'. \tag{4.10}$$

Then Q_r^{nhu} fulfills (3.11) and also

$$\int Q_r^{nhu}(\mathbf{t}, dy)(y - m_{\vartheta+n^{-1/2}u}(\mathbf{t}))^2 = v_{\vartheta+n^{-1/2}u}(\mathbf{t}) + o(n^{-1/2}).$$

The efficient score function s again minimizes $\pi Q_r hh'$, now over the smaller affine space of functions h fulfilling (3.8), (3.9) and (4.10). The solution is

$$s(\mathbf{t}, y) = \mathbf{d}_\vartheta(\mathbf{t})' C^{-1}(\mathbf{t}) \mathbf{i}_\vartheta(\mathbf{t}, y). \tag{4.11}$$

To see this, note that s fulfills (3.8), (3.9) and (4.10) since

$$\int Q_r(\mathbf{t}, dy) s(\mathbf{t}, y) i_\vartheta(\mathbf{t}, y)' = \mathbf{d}_\vartheta(\mathbf{t})',$$

and that s fulfills $\pi Q_r sh' = \pi Q_r ss'$ since h fulfills (3.8), (3.9) and (4.10). Hence the efficient influence function is (4.8), and the minimal asymptotic covariance matrix for regular estimators of ϑ is (4.7). The estimator based on the extended adaptive estimating function (4.9) also has influence function (4.8) and is therefore efficient.

Remark 2. We have shown that our adaptive estimating function (3.7) is as good as the best estimating function (3.1) for *known* v, with weight (3.4). This does not mean that the estimator based on (3.7) remains efficient in the class of all regular estimators if v is assumed known. This is only true if the vectors $h(\mathbf{t}, y)$ fulfill, besides (3.8) and (3.9),

$$\int Q_r(\mathbf{t}, dy)(y - m_\vartheta(\mathbf{t}))^2 h(\mathbf{t}, y) = 0.$$

This condition is not fulfilled by the score function (3.11) unless $\mu_3 = 0$, i.e., unless the two estimating functions (3.1) and (4.2) are orthogonal in the sense that $\mu_3 = 0$. □

Remark 3. In some applications the conditional mean m_ϑ of the response does not depend on ϑ. Then $\dot{m}_\vartheta = 0$, and $\pi \mathbf{w}_\vartheta \mathbf{d}_\vartheta$ is not invertible, so that the calculations in Subsection 4.2 are not valid. In this case, the estimating functions (3.1) are useless in the sense that they do not lead to estimators with finite asymptotic variance. In particular, the quasi-score function (4.1) is useless.

One possible alternative is to restrict attention to estimating functions (4.2) and proceed as in Subsections 3.1 to 3.3, with the model $m = m_\vartheta$ replaced by the model $v = v_\vartheta$. As in Subsection 3.1, the estimator corresponding to the estimating function (4.2) is shown to be asymptotically linear with influence function

$$f(\mathbf{t}, y) = (\pi w_\vartheta \dot{v}_\vartheta)^{-1} w_\vartheta(\mathbf{t}) \left((y - m_\vartheta(\mathbf{t}))^2 - v_\vartheta(\mathbf{t}) \right)$$

and asymptotic covariance matrix

$$(\pi w_\vartheta \dot{v}_\vartheta)^{-1} \pi(\mu_4 - v_\vartheta^2) w_\vartheta w_\vartheta' (\pi \dot{v}_\vartheta' w_\vartheta')^{-1}.$$

If μ_4 is *known*, this covariance matrix is minimized for

$$w_\vartheta = (\mu_4 - v_\vartheta^2)^{-1}\dot{v}_\vartheta',$$

and the minimal asymptotic covariance matrix is

$$(\pi(\mu_4 - v_\vartheta^2)^{-1}\dot{v}_\vartheta'\dot{v}_\vartheta)^{-1}.$$

A good estimating function is

$$\sum_{i=p}^{n}(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - v_\vartheta(\boldsymbol{Z}_{i-1})^2)^{-1}\dot{v}_\vartheta(\boldsymbol{Z}_{i-1})'((\boldsymbol{Y}_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})). \quad (4.12)$$

It would be efficient if we had not specified m at all. In general, however, the assumption that m_ϑ does not depend on ϑ contains information about ϑ. Condition (3.9) on h now reads

$$\int Q_r(\mathbf{t}, dy)yh(\mathbf{t}, y) = 0. \quad (4.13)$$

The score function of the above estimator is

$$s(\mathbf{t}, y) = (\mu_4(\mathbf{t}) - v_\vartheta(\mathbf{t})^2)^{-1}\dot{v}_\vartheta(\mathbf{t})'\left((y - m_\vartheta(\mathbf{t}))^2 - v_\vartheta(\mathbf{t})\right).$$

For this score function to be efficient, condition (4.13) must hold for $h = s$. This is not true unless $\mu_3 = 0$. An analogous result with interchanged roles of m and v was noted in Remark 2.

We note that although the estimating functions (3.1) are useless on their own, they can be used in combination with estimating functions (4.2): For $\dot{m}_\vartheta = 0$ the efficient score function (4.11) reduces to

$$\begin{aligned} &D(\mathbf{t})^{-1}\dot{v}_\vartheta(\mathbf{t})'(-\mu_3(\mathbf{t}), v_\vartheta(\mathbf{t}))\mathbf{i}_\vartheta(\mathbf{t}, y) \\ &= D(\mathbf{t})^{-1}\dot{v}_\vartheta(\mathbf{t})'\left(-\mu_3(\mathbf{t})(y - m_\vartheta(\mathbf{t})) + v_\vartheta(\mathbf{t})\left((y - m_\vartheta(\mathbf{t}))^2 - v_\vartheta(\mathbf{t})\right)\right), \end{aligned}$$

where $D = v_\vartheta(\mu_4 - v_\vartheta^2) - \mu_3^2$ is the determinant of C. The corresponding extended adaptive estimating function is

$$\begin{aligned} \sum_{i=p}^{n} \hat{D}_{i-1}(\boldsymbol{Z}_{i-1})^{-1}\dot{v}_\vartheta(\boldsymbol{Z}_{i-1})'\Big(&- \hat{\mu}_{3,i-1}(\boldsymbol{Z}_{i-1})(Y_i - m_\vartheta(\boldsymbol{Z}_{i-1})) \quad (4.14) \\ &+ v_\vartheta(\boldsymbol{Z}_{i-1})\left((Y_i - m_\vartheta(\boldsymbol{Z}_{i-1}))^2 - v_\vartheta(\boldsymbol{Z}_{i-1})\right)\Big). \end{aligned}$$

For $\mu_3 = 0$ this gives again (4.12). □

Remark 4. An *extended quasi-likelihood model* is given by parametric models $m = m_\vartheta$, $v = v_\vartheta$, $\mu_3 = \mu_{3\vartheta}$ and $\mu_4 = \mu_{4\vartheta}$. Similarly as in Subsection

4.1, the best estimating function (4.3) is seen to have weights (4.6), now with $\mu_3 = \mu_{3\vartheta}$ and $\mu_4 = \mu_{4\vartheta}$. This gives the *extended quasi-score function*

$$\sum_{i=p}^{n} \mathbf{d}_\vartheta(\boldsymbol{Z}_{i-1})' C_\vartheta(\boldsymbol{Z}_{i-1})^{-1} \mathbf{i}_\vartheta(\boldsymbol{Z}_{i-1})$$

with

$$C_\vartheta = \begin{pmatrix} v_\vartheta & \mu_{3\vartheta} \\ \mu_{3\vartheta} & \mu_{4\vartheta} - v_\vartheta^2 \end{pmatrix}.$$

It is asymptotically as good as the estimator given by the extended adaptive estimating function (4.9). Hence it does not use the information in the specifications $\mu_3 = \mu_{3\vartheta}$ and $\mu_4 = \mu_{4\vartheta}$. It is robust against misspecification of μ_3 and μ_4, but then the extended adaptive estimating function is strictly better. □

4.6. Heteroscedastic regression-autoregression models. Suppose that the responses have a heteroscedastic autoregressive structure,

$$Y_i = m_\vartheta(\boldsymbol{Z}_{i-1}) + v_\vartheta(\boldsymbol{Z}_{i-1})^{1/2}\varepsilon_i,$$

where the ε_i are i.i.d. with known or unknown mean zero density g, We may and will also assume that the ε_i have variance one. The conditional distribution of the responses given $\boldsymbol{Z}_{i-1} = \mathbf{t}$ has the form

$$Q_r(\mathbf{t}, dy) = v_\vartheta(\mathbf{t})^{-1/2}\, g\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right) dy,$$

with conditional mean m_ϑ and conditional variance v_ϑ. We call it a *heteroscedastic regression-autoregression model.* It is a submodel of the quasi-likelihood model given by $m = m_\vartheta$ and $v = v_\vartheta$.

We show that the lower bound for the asymptotic covariance matrices of regular estimators of ϑ is, in general, strictly smaller than the lower bound (4.7) in the quasi-likelihood model. We follow the arguments of Subsection 3.5, now with heteroscedasticity. Since $\mathrm{E}\,\varepsilon^2 = 1$, the functions k fulfill not only (3.13) and (3.14) but also

$$\mathrm{E}\,\varepsilon^2 k(\varepsilon) = 0.$$

With $g^{nk}(y) = g(y)(1 + n^{-1/2}k(y))$ as before we set

$$Q_r^{nku}(\mathbf{t}, dy) = v_{\vartheta+n^{-1/2}u}(\mathbf{t})^{-1/2}\, g^{nk}\left(v_{\vartheta+n^{-1/2}u}(\mathbf{t})^{-1/2}(y - m_{\vartheta+n^{-1/2}u}(\mathbf{t}))\right) dy.$$

By a Taylor expansion,

$$Q_r^{nku}(\mathbf{t}, dy)$$

$$= Q_r(\mathbf{t}, dy)\Big(1 + n^{-1/2}\Big(k\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right)$$
$$\left(-v_\vartheta(\mathbf{t})^{-1/2}\dot{m}_\vartheta(\mathbf{t})u\ \ell'\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right)\right.$$
$$\left(-1/2 v_\vartheta(\mathbf{t})^{-1}\dot{v}_\vartheta(\mathbf{t})u\ (v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\ell'\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right) + 1)\Big)\Big)$$
$$= +o(n^{-1/2}).$$

Hence the corresponding partial likelihood ratio is locally asymptotically normal with variance

$$\int\int \pi(d\mathbf{t})g(y)dy(k(y) - v_\vartheta(\mathbf{t})^{-1/2}\dot{m}_\vartheta(\mathbf{t})u\,\ell'(y) - \frac{1}{2}v_\vartheta(\mathbf{t})^{-1}\dot{v}_\vartheta(\mathbf{t})u(y\ell'(y) + 1))^2. \tag{4.15}$$

The model is adaptive with respect to g if ϑ and g are locally orthogonal in the sense that $k(y)$ is orthogonal to $v_\vartheta(\mathbf{t})^{-1/2}\dot{m}_\vartheta(\mathbf{t})\ell'(y)$ $+\frac{1}{2}v_\vartheta(\mathbf{t})^{-1}\dot{v}_\vartheta(\mathbf{t})\ (y\ell'(y) + 1)$. This condition is rarely fulfilled. For a discussion see Drost et al. (1994) and Drost and Klaassen (1995). To simplify the calculations, we will assume that g is *known*, and calculate the minimal asymptotic covariance matrix for regular estimators in that case. It equals the minimal asymptotic covariance matrix for an adaptive model and is a lower bound for the non-adaptive situation. If g is known, the variance (4.15) reduces to

$$\pi V_\vartheta^{-1} J V_\vartheta^{-1},$$

where

$$J = \begin{pmatrix} \mathrm{E}\,\ell'(\varepsilon)^2 & \frac{1}{2}\mathrm{E}\,\varepsilon\ell'(\varepsilon)^2 \\ \frac{1}{2}\mathrm{E}\,\varepsilon\ell'(\varepsilon)^2 & \frac{1}{4}(\mathrm{E}\,\varepsilon^2\ell'(\varepsilon)^2 - 1) \end{pmatrix}$$

and V_ϑ is the matrix

$$V = \begin{pmatrix} v^{1/2} & 0 \\ 0 & v \end{pmatrix}$$

with $v = v_\vartheta$. Hence the efficient score function is

$$s(\mathbf{t}, y) = \mathbf{d}_\vartheta(\mathbf{t})' V_\vartheta^{-1}\mathbf{e}_\vartheta(\mathbf{t}, y)$$

with

$$\mathbf{e}_\vartheta(\mathbf{t}, y) = \begin{pmatrix} -\ell'\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right) \\ -\frac{1}{2}\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\ell'\left(v_\vartheta(\mathbf{t})^{-1/2}(y - m_\vartheta(\mathbf{t}))\right) + 1\right) \end{pmatrix},$$

and the minimal asymptotic covariance matrix of regular estimators of ϑ in the heteroscedastic regression-autoregression model is

$$(\pi\mathbf{d}_\vartheta' V_\vartheta^{-1} J V_\vartheta^{-1}\mathbf{d}_\vartheta)^{-1}. \tag{4.16}$$

This matrix cannot be larger than the minimal asymptotic covariance matrix (4.7) in the larger model $m = m_\vartheta$ and $v = v_\vartheta$, the quasi-likelihood model. To check this, note first that in the heteroscedastic regression-autoregression model the μ_i are of the form

$$\begin{aligned}\mu_j(\mathbf{t}) &= v_\vartheta^{-1/2} \int g\left(v_\vartheta^{-1/2}(y - m_\vartheta(\mathbf{t}))\right) dy\,(y - m_\vartheta(\mathbf{t}))^j \\ &= v_\vartheta^{j/2} \mathrm{E}\,\varepsilon^j, \quad j = 3, 4.\end{aligned}$$

Hence the matrix (4.4) can be written

$$C(\mathbf{t}) = V_\vartheta(\mathbf{t}) F V_\vartheta(\mathbf{t})$$

with

$$F = \begin{pmatrix} 1 & \mathrm{E}\,\varepsilon^3 \\ \mathrm{E}\,\varepsilon^3 & \mathrm{E}\,\varepsilon^4 - 1 \end{pmatrix},$$

and the minimal asymptotic covariance matrix (4.7) is

$$(\pi \mathbf{d}_\vartheta' C^{-1} \mathbf{d}_\vartheta)^{-1} = (\pi \mathbf{d}_\vartheta' V_\vartheta^{-1} F^{-1} V_\vartheta^{-1} \mathbf{d}_\vartheta)^{-1}.$$

To prove that this matrix is larger than the minimal asymptotic covariance matrix (4.16) in the heteroscedastic regression-autoregression model, it suffices to show that $F - J^{-1}$ is positive semi-definite. This is a well-known result. We recall it briefly. Consider the location-scale model generated by the density g with mean zero and variance one, and the problem of estimating mean and variance based on i.i.d. observations $\varepsilon_1, \ldots, \varepsilon_n$. If the true distribution has mean zero and variance one, the Fisher information matrix is J, and an efficient estimator, say the maximum likelihood estimator, has asymptotic covariance matrix J^{-1}. If we do not know the density g, then the model is completely nonparametric, and an efficient estimator is the empirical estimator for the mean and the variance. If the true distribution has mean zero and variance one, its asymptotic covariance matrix is F. It must be larger than J^{-1}. The inequality is strict unless $\ell'(y)$ is proportional to y. In particular, if the ε_i are normal, then the extended adaptive estimating function (4.9) gives an efficient estimator in the heteroscedastic regression-autoregression model.

Example 3. Set $m = 0$ and $v_\vartheta(\boldsymbol{Z}_{i-1}) = \sigma^2(1 + \beta_1 Y_{i-1}^2 + \cdots + \beta_p Y_{i-p}^2)$. Then $\vartheta = (\sigma^2, \beta_1, \ldots, \beta_p)'$ is of dimension $q = 1 + p$. As noted in Remark 3, a good estimator is obtained from the estimating function (4.12). With $\mathbf{Y}_{i-1}^2 = (Y_{i-1}^2, \ldots, Y_{i-p}^2)'$ it reads

$$\sum_{i=p}^{n} \left(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - \sigma^4(1 + \beta' \mathbf{Y}_{i-1}^2)^2\right)^{-1} \begin{pmatrix} 1 + \beta' \mathbf{Y}_{i-1}^2 \\ \sigma^2 \mathbf{Y}_{i-1}^2 \end{pmatrix} \tag{4.17}$$

$$\left(Y_i^2 - \sigma^2(1 + \beta' \mathbf{Y}_{i-1}^2)\right).$$

An efficient estimating function is the extended adaptive estimating function (4.14). It is obtained from (4.17) by adding to the martingale increment $Y_i^2 - \sigma^2(1+\beta'\mathbf{Y}_{i-1}^2)$ the increment

$$-\hat{\mu}_{3,i-1}(\boldsymbol{Z}_{i-1})\sigma^{-2}(1+\beta'\mathbf{Y}_{i-1}^2)^{-1}Y_i.$$

The corresponding heteroscedastic autoregression model is the *p-order ARCH model* introduced in Engle (1982),

$$Y_i = \sigma(1+\beta'\mathbf{Y}_{i-1}^2)^{1/2}\epsilon_i,$$

where the ε_i are i.i.d. with a density g which has mean and variance one. In this model we have

$$\mu_4(\mathbf{t}) = \sigma^4(1+\beta'\mathbf{s}^2)^2 \mathrm{E}\,\varepsilon^4,$$

and the estimating function (4.17) is, up to an irrelevant factor $\sigma^{-4}(\mathrm{E}\,\varepsilon^4-1)^{-1}$,

$$\sum_{i=p}^{n}(1+\beta'\mathbf{Y}_{i-1}^2)^{-2}\begin{pmatrix} 1+\beta'\mathbf{Y}_{i-1}^2 \\ \sigma^2\mathbf{Y}_{i-1}^2 \end{pmatrix}\left(Y_i^2 - \sigma^2(1+\beta'\mathbf{Y}_{i-1}^2)\right). \qquad (4.18)$$

For normal ε_i this gives the maximum likelihood estimator.

A review of ARCH models is Bollerslev et al. (1992). Efficient estimators in this model are constructed in Engle and González-Rivera (1991), Linton (1993) and Drost et al. (1994) under increasingly weaker assumptions.

Example 4. Set $m = 0$ and

$$v_\vartheta(\boldsymbol{Z}_{i-1}) = \sigma^2(1+\beta_1\left(Y_{i-1}-\alpha'X_{i-1})^2 + \cdots + \beta_p(Y_{i-p}-\alpha'X_{i-p})^2\right).$$

Then $\vartheta = (\sigma^2, \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_p)'$ is of dimension $q = 1+k+p$. As in Example 3, the quasi-score function (4.1) is useless, and a good estimator is obtained from the estimating function (4.12). We write $v_\vartheta(\mathbf{t}) = \sigma^2(1+\beta'(\mathbf{s}-\alpha'\mathbf{r})^2)$ with $\mathbf{s}^2 = (s_{p-1}^2, \ldots, s_0^2)'$ and $\beta'\mathbf{r} = (\beta' r_{p-1}, \ldots, \beta' r_0)'$, and obtain

$$\dot{v}_\vartheta(\mathbf{t}) = \begin{pmatrix} 1+\beta'(\mathbf{s}-\alpha'\mathbf{r})^2 \\ -2\sigma^2\beta'(\mathbf{s}-\alpha'\mathbf{r})\mathbf{r} \\ \sigma^2(\mathbf{s}-\alpha'\mathbf{r})^2 \end{pmatrix}.$$

Hence the estimating function (4.12) is

$$\sum_{i=p}^{n}\left(\hat{\mu}_{4,i-1}(\boldsymbol{Z}_{i-1}) - \sigma^4(1+\beta'(\boldsymbol{Y}_{i-1}-\alpha'\boldsymbol{X}_{i-1})^2)^2\right)^{-1} \qquad (4.19)$$

$$\begin{pmatrix} 1+\beta'(\boldsymbol{Y}_{i-1}-\alpha'\boldsymbol{X}_{i-1})^2 \\ -2\sigma^2\beta'(\boldsymbol{Y}_{i-1}-\alpha'\boldsymbol{X}_{i-1})\boldsymbol{X}_{i-1} \\ \sigma^2(\boldsymbol{Y}_{i-1}-\alpha'\boldsymbol{X}_{i-1})^2 \end{pmatrix}\left(Y_i^2 - \sigma^2(1+\beta'(\boldsymbol{Y}_{i-1}-\alpha'\boldsymbol{X}_{i-1})^2)\right).$$

The corresponding heteroscedastic regression-autoregression model is the *p-order ARCH model* with k-dimensional *linear regression trend* introduced in Engle (1982),

$$Y_i = \sigma(1 + \beta'(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})^2)^{1/2}\varepsilon_i,$$

where the ε_i are i.i.d. with a density g which has mean and variance one. In this model we have

$$\mu_4(\mathbf{t}) = \sigma^4(1 + \beta'(\mathbf{s} - \alpha'\mathbf{r})^2)^2 \mathrm{E}\,\varepsilon^4,$$

and the estimating function (4.19) is, up to the irrelevant factor $\sigma^{-4}(\mathrm{E}\,\varepsilon^4 - 1)^{-1}$,

$$\sum_{i=p}^{n}(1 + \beta'(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})^2)^{-2}$$
$$\begin{pmatrix} 1 + \beta'(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})^2 \\ -2\sigma^2\beta'(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})\boldsymbol{X}_{i-1} \\ \sigma^2(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})^2 \end{pmatrix} \left(Y_i^2 - \sigma^2(1 + \beta'(\boldsymbol{Y}_{i-1} - \alpha'\boldsymbol{X}_{i-1})^2)\right).$$

Acknowledgment

The author thanks Feike Drost for helpful discussions on ARCH models, and the referee for numerous suggestions which made the paper more readable.

References

Andrews, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.

Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.* 62, 119–132.

Bhattacharya, R. and Lee, C. (1995). On geometric ergodicity of nonlinear autoregressive models. *Statist. Probab. Lett.* 22, 311–315.

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.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Bollerslev, T., Chou, R. Y. and Kroner, K. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. *J. Econometrics* 52, 5–59.

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

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

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.

Drost, F. C. and Klaassen, C. A. J. (1995). Efficient estimation in semiparametric GARCH models. To appear in: *J. Econometrics.*

Drost, F. C., Klaassen, C .A. J. and Werker, B. J. M. (1994). Adaptive estimation in time series models. Preprint.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.

Engle, R. F. and González-Rivera (1991). Semiparametric ARCH models. *J. Business Economic Statist.* 9, 345–359.

Firth, D. (1993). Recent developments in quasi-likelihood methods. *Bull. Inst. Internat. Statist.* 55, 2, 341–358.

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.* 47, 551–579.

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. 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.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* 14, 323–330.

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

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press.

Hosoya, Y. (1989). The bracketing condition for limit theorems on stationary linear processes. *Ann. Statist.* 17, 401–418.

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

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

Hwang, S. Y. and Basawa, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series models. *Stochastic Process. Appl.* 46, 91–113.

Hwang, S. Y. and Basawa, I. V. (1994). Large sample inference based on multiple observations from nonlinear autoregressive processes. *Stochastic Process. Appl.* 49, 127–140.

Kessler, M. (1995). Martingale estimating functions for a Markov chain. Preprint.

Koul, H. L. and Schick, A. (1996). Efficient estimation in nonlinear time series. Preprint.

Kreiss, J.-P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* 15, 112–133.

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

Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* 9, 539–569.

McCullagh, P. (1991). Quasi-likelihood and estimating functions. In: *Statistical Theory and Modelling* (D. V. Hinkley, N. Reid and E. J. Snell, eds.), 265–286. Chapman and Hall, London.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* 16, 54–70.

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.

Wefelmeyer, W. (1996a). Adaptive estimators for parameters of the autoregression function of a Markov chain. To appear in: *J. Statist. Plann. Inference.*

Wefelmeyer, W. (1996b). Quasi-likelihood models and optimal inference. *Ann. Statist.* 24, 405–422.