Journal of the Korean Statistical Society 43 (2014) 475–482

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

# Non-stationary quasi-likelihood and asymptotic optimality

S.Y. Hwang [a,*], Tae Yoon Kim [b]

[a] *Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea*
[b] *Department of Statistics, Keimyung University, Daegu, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 10 October 2013
Accepted 15 February 2014
Available online 5 March 2014

*AMS 2000 subject classifications:*
primary 62M10
secondary 62P10

*Keywords:*
Asymptotic optimality
Non-stationary quasi-likelihood
Non-stationary ARCH
Branching Markov processes
Non-stationary random-coefficient AR

ABSTRACT

This article is concerned with non-stationary time series which does not require the full knowledge of the likelihood function. Consequently, a quasi-likelihood is employed for estimating parameters instead of the maximum (exact) likelihood. For stationary cases, Wefelmeyer (1996) and Hwang and Basawa (2011a,b), among others, discussed the issue of asymptotic optimality of the quasi-likelihood within a restricted class of estimators. For non-stationary cases, however, the asymptotic optimality property of the quasi-likelihood has not yet been adequately addressed in the literature. This article presents the asymptotic optimal property of the non-stationary quasi-likelihood within certain estimating functions. We use a random norm instead of a constant norm to get limit distributions of estimates. To illustrate main results, the non-stationary ARCH model, branching Markov process, and non-stationary random-coefficient AR process are discussed.


## 1. Introduction

When the likelihood of the data from a stochastic process is known to us, the likelihood-based methods are readily available for estimating parameters. A unified approach for asymptotic optimal inference based on the likelihood (from a stationary process) is to use the general framework of local asymptotic normality (LAN). Refer to, among others, Hall and Mathiason (1990) and Wefelmeyer (1996) for an excellent review of the LAN. The LAN is extended to local asymptotic mixed normality (LAMN) to suit non-stationary processes. Asymptotic optimality properties of the maximum (exact) likelihood estimates under various criteria are well documented in the literature via the LAN and the LAMN for the stationary case and non-stationary case, respectively. See, for instance, Basawa and Scott (1983), Hwang and Basawa (2011a) and Hwang, Basawa, Choi, and Lee (2013) for the various asymptotic optimal properties of the maximum (exact) likelihood estimates via the LAN and LAMN approaches.

In this article, we suppose that the likelihood function is *unknown*. For instance, in a time series, the error distribution may not be known and/or the time series is specified only through a first few conditional moments. A quasi-likelihood (QL, for short) method in the context of estimating functions is suited to cases of unknown likelihood (cf. Heyde, 1997). Introducing the Godambe information criterion, Godambe (1985) established finite sample optimality of the QL within a certain class of estimating functions. To establish asymptotic optimality properties of the QL, one needs to impose constraints on the class of estimators among which the quasi-likelihood performs "best". Readers refer to, e.g., Wefelmeyer (1996) and Hwang and Basawa (2011b, 2014) for the stationary QL case. However, for the non-stationary QL case, the issue of asymptotic optimality has not been adequately pursued in the literature. Our main goal is to establish certain asymptotic optimality of

* Corresponding author. Tel.: +82 1091809078.
*E-mail address:* shwang@sm.ac.kr (S.Y. Hwang).

http://dx.doi.org/10.1016/j.jkss.2014.02.004
1226-3192/

the non-stationary QL within a restricted class of estimating functions (see Section 2 for details). To illustrate main results, the non-stationary ARCH model, branching Markov process and non-stationary random coefficient AR model are discussed.

## 2. Non-stationary quasi-likelihood: main results

We are concerned with the non-stationary stochastic process $\{X_t,\ t = 0, 1, 2, \ldots\}$ with the starting value $X_0 = x_0$. The probability measure associated with $\{X_t\}$ involves a $(k \times 1)$ vector parameter $\theta$ which belongs to an open subset $\Theta$ of $R^k$, the k-dimensional Euclidean space. It is noted that $\theta$ denotes a parameter of interest, allowing nuisance parameter $\eta$ in addition to $\theta$. Semiparametric cases can also be included by treating the error distribution as $\eta$. This will be illustrated via examples in Section 3. Let $\{X_1, X_2, \ldots, X_n\}$ denote the sample. Suppose that the likelihood of the data is *unknown* and we are interested in estimating $\theta$ based on the sample. We are then led to the theory of estimating functions instead of the likelihood-based methods. Throughout the paper, the estimating function is shortened as EF. We regard quasi-likelihood (QL) as a member of certain class of EFs within which the asymptotic optimality property of the QL can be derived. To proceed, one needs to restrict the class of EFs under consideration in order to discuss the optimality of the non-stationary QL.

Godambe (1985) introduced a "linear" class $G$ of EFs $G_n(\theta) : (k \times 1)$ defined by

$$G = \left\{ G_n(\theta) = \sum_{t=1}^{n} w_t(\theta) g_t(\theta) \right\} \tag{2.1}$$

where $\{g_t(\theta)\}$ is a sequence of martingale differences with respect to $F_t$ (here and in what follows, $F_t$ denotes the $\sigma$-field generated by $X_t, X_{t-1}, \ldots, X_1$ for each $t \geq 1$). In (2.1), $g_t(\theta)$ is a pre-specified martingale difference which is referred to as an innovation at time $t$, that is, $E(g_t(\theta)|F_{t-1}) = 0$. And $w_t(\theta)$ is an $F_{t-1}$ measurable weight vector of order $(k \times 1)$. The class $G$ is generated by varying the "coefficient" $w_t(\theta)$ while the given innovation $g_t(\theta)$ being fixed. We use the notation $E_{t-1}$ for denoting the conditional expectation given $F_{t-1}$, viz., $E_{t-1}(\cdot) = E(\cdot|F_{t-1})$ so that $E_{t-1} g_t(\theta) = 0$. It is assumed that $E_{t-1}[g_t^2(\theta)] < \infty$ and $E_{t-1}[\partial g_t(\theta)/\partial\theta] \neq 0$. As a special member of $G$, the quasi-likelihood (QL) $G_n^*(\theta)$ is defined by

$$G_n^*(\theta) = \sum_{t=1}^{n} E_{t-1}[\partial g_t(\theta)/\partial\theta](E_{t-1} g_t^2(\theta))^{-1} g_t(\theta). \tag{2.2}$$

Although it would be more precise to use the term QL-EF for $G_n^*(\theta)$, we shall refer to $G_n^*(\theta)$ as QL for simplicity. In this paper, the superscript $*$ is reserved for denoting terms related to the quasi-likelihood $G_n^*(\theta)$. Due to Godambe (1985), the QL $G_n^*(\theta)$ enjoys maximum of the so-called Godambe's information matrix among the class $G$ of EFs $G_n(\theta)$. Refer also to, e.g., Hwang and Basawa (2011b) and Thavaneswaran, Liang, and Frank (2012) for the expression of Godambe's information matrix. As is noted in, e.g., Hwang and Basawa (2011b), the optimality of QL $G_n^*(\theta)$ established by Godambe (1985) is for the estimating function itself and does not lead to any finite or asymptotic optimality of the estimator derived from the QL equation $G_n^*(\theta) = 0$. Dealing with the estimators directly (rather than EFs themselves), let $\hat{\theta}_{QL}$ denote a consistent solution of the QL equation $G_n^*(\theta) = 0$. For stationary cases, Chandra and Taniguchi (2001), Hwang and Basawa (2011b, 2014) and Wefelmeyer (1996) showed that the asymptotic variance of $\sqrt{n}(\hat{\theta}_{QL} - \theta)$ achieves the "minimum" among all the estimators derived from EFs in the class $G$, under appropriate conditions.

To complement the literature, this article covers non-stationary cases. To do this, we use a random norm for $\hat{\theta}_{QL}$ instead of a constant norm $\sqrt{n}$. Consider $G_n(\theta) = \sum_{t=1}^{n} w_t(\theta) g_t(\theta) \in G$ for which the sum of conditional covariance matrices $V_n(\theta)$ is defined by

$$V_n(\theta) = \sum_{t=1}^{n} w_t(\theta) w_t^T(\theta) E_{t-1} g_t^2(\theta) : (k \times k) \tag{2.3}$$

where the superscript $T$ denotes "transpose". We shall regard two EFs in $G$ as being identical if one is a constant multiple of the other. It will be assumed that $\|V_n(\theta)\| \to \infty$ almost surely as the sample size $n$ tends to infinity. Here, $\|\cdot\|$ is used to denote the matrix (or vector) norm, e.g., $\|A\|^2 = \text{trace}(A^T A)$. The half matrix of a symmetric positive definite matrix $A$ is denoted by $A^{1/2}$ obtained from the spectral decomposition of $A$. The inverse of $A^{1/2}$ is denoted by $A^{-1/2}$. Denote by $I_k$ the identity matrix of order $k$. Fix $\theta \in \Theta$ and define the random local neighborhood $N_\delta(\theta)$ about $\theta$.

$$N_\delta(\theta) = \{\bar{\theta};\ \|V_n^{1/2}(\theta)(\bar{\theta} - \theta)\| < \delta\}, \quad \text{for some } \delta > 0 \tag{2.4}$$

Note that $N_\delta(\theta)$ reduces almost surely to $\theta$ as $n \to \infty$. From now on, we confine ourselves to "Regular EFs" in the class $G$.

**Definition.** Any EF $G_n(\theta) \in G$ is called regular if it satisfies the following three conditions.
(C1) For each fixed $\theta \in \Theta$,

$$\sup \|V_n^{-1/2}(\theta)[\partial G_n(\bar{\theta})/\partial\theta^T - \partial G_n(\theta)/\partial\theta^T] V_n^{-1/2}(\theta)\| = o_p(1)$$

where the "sup" is taken over $\bar{\theta} \in N_\delta(\theta)$ and $o_p(1)$ stands for a term converging to zero in probability.