Finite Lag Estimation of Non-Markovian Processes[1]

A. Ronald Gallant[2] Halbert White[3]

First draft: August 15, 1997
This draft: October 12, 2023

Paper: www.aronaldg.org/papers/qmle.pdf

[1]Address correspondence to A. Ronald Gallant, P.O. Box 659, Chapel Hill NC 27514, USA, phone 919-428-1130; email aronldg@gmail.com.

[2]Emeritus Professor of Economics, University of North Carolina, Chapel Hill and Emeritus Professor of Economics, Penn State University

[3]Deceased Professor of Economics, University of California, San Diego

Abstract

We consider the quasi maximum likelihood estimator obtained by replacing each transition density in the correct likelihood for a non-Markovian, stationary process by a transition density with a fixed number of lags. This estimator is of interest because it is asymptotically equivalent to the efficient method of moments estimator as typically implemented in dynamic macro and finance applications. We show that the standard regularity conditions of quasi maximum likelihood imply that a score vector defined over the infinite past exists. We verify that the existence of a score on the infinite past implies that the asymptotic variance of the finite lag quasi maximum likelihood estimator tends to the asymptotic variance of the maximum likelihood estimator as the number of lags tends to infinity.

Key words: Maximum likelihood, non-Markovian, quasi maximum likelihood, finite lag approximation, efficient method of moments.

JEL codes: C14, C15, C32, C58

1 Introduction

We consider a stationary process

$$\{y_t \in \mathbb{R}^M : t = 0, \pm 1, \pm 2, ...\}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$ whose finite dimensional density functions are in the family

$$\mathcal{P}_\rho = \{p(y_s, \ldots, y_t, \rho) : s \leq t = 0, \pm 1, \pm 2, \ldots;\ \rho \in \mathcal{R} \subset \mathbb{R}^{p_\rho}\}$$

for some $\rho^o \in \mathcal{R}$. We assume that the process $\{y_t\}$ is not Markovian, in the sense that for some Borel set $A \in \mathbb{R}^M$,

$$P\left[\mathcal{E}\big(I_A(y_t)|\mathcal{F}_{-\infty}^{t-1}\big) = \mathcal{E}\big(I_A(y_t)|\mathcal{F}_{t-L}^{t-1}\big)\right] < 1$$

for all finite L, where $\mathcal{F}_s^t$ denotes the smallest sub-σ-algebra of $\mathcal{F}$ such that the random variables $\{y_s, \ldots, y_t\}$ are measurable.

Such processes can arise in a variety of ways, e.g., a linear system with moving average errors, but we are primarily interested in parameterized processes that are well suited to estimation by efficient method of moments (EMM) implemented by means of a seminonparametric (SNP) score generator. A leading example is a process obtained by discretely sampling a subset of the state variables of a continuous time process that evolves according to a system of nonlinear stochastic differential equations (Gallant and Long, 1997). Other examples are in Gallant and Tauchen (1996).

Although other simulation estimators are applicable in these situations (Ingram and Lee, 1991; Smith, 1993; Gourieroux, Monfort, and Renault, 1993; Duffie and Singleton, 1993), the distinguishing characteristic of the EMM/SNP estimator is that, as shown by Gallant and Long (1997), it is asymptotically equivalent to the quasi maximum likelihood estimator $\hat{\rho}_n$ that is obtained by replacing each transition density in the correct likelihood by a transition density on L lags. Specifically, the objective function

$$Q_n(\rho) = p(y_0, \ldots y_{L-1}, \rho) \prod_{t=L}^{n} p(y_t|y_{t-L}, \ldots, y_{t-1}, \rho)$$

replaces the standard likelihood

$$L_n(\rho) = p(y_0, \rho) \prod_{t=1}^{n} p(y_t|y_0, \ldots, y_{t-1}, \rho),$$

and the estimator is

$$\hat{\rho}_n = \underset{\rho \in R}{\operatorname{argmax}}\, Q_n(\rho).$$

Note that $p(y_t|y_{t-L}, \ldots, y_{t-1}, \rho)$ is the correct density for y_t given $y_{t-L}, \ldots, y_{t-1}$; the error in the approximation of L_n by Q_n is due to truncation, not to misspecified functional form. A salient effect of this truncation is that for finite L, the truncated scores $(\partial/\partial\rho) \log p(y_t|y_{t-L}, \ldots, y_{t-1}, \rho)$ do not necessarily form a martingale difference sequence.

Because the EMM estimator implemented by means of SNP is asymptotically equivalent to $\hat{\rho}_n$, a high level assumption that the quasi maximum likelihood estimator is asymptotically equivalent to the maximum likelihood estimator as L tends to infinity implies that EMM is as efficient as maximum likelihood in the limit. Gallant and Long (1997) obtained their efficiency result by imposing this assumption.

While a high level assumption that the quasi maximum likelihood estimator is asymptotically equivalent to the maximum likelihood estimator is plausible, one would prefer a result that was deduced from more standard and more primitive assumptions. That is our goal here. We show that the Gallant-Long assumption is implied by standard assumptions for maximum likelihood estimation of the parameters of a non-Markovian, stationary system.

Our proof strategy is to construct the score vector for the case when data extend to the infinite past. The construction of the score on the infinite past and its properties are of some interest in their own right. Further, from these properties, one can deduce that these three estimators are asymptotically equivalent: quasi maximum likelihood with an objective function formed from n transition densities that condition on the infinite past, quasi maximum likelihood with an objective function formed from n transition densities that condition on L lags (in the limit as L tends to infinity), and maximum likelihood with an objective function formed from n transition densities that condition back to the first observation. This equivalence implies our main result.

The 1997 version of this paper contained errors that Hal and I were not able to fix.[1] Being now retired, with an abundance of boring leisure, returning to this paper seemed to be a suitable time sink. Regrettably, I was able to fix the errors in finite time.

The plan of the paper is a follows. In Section 2, we specialize the standard assumptions of

[1] The mathematics in Gallant-Long are correct to my knowledge.

nonlinear dynamic modeling to the case of maximum likelihood estimation for non-Markovian data. In Section 3, we construct a notion of a score on the infinite past. In Section 4, we deduce some properties of this score and use them to obtain our main result. In Section 5 we verify our results by a simulation and comment on how one might go about choosing L for a process that truly depends on the infinite past. Although our results are motivated by the EMM/SNP application, they are more general in that they apply to any non-Markovian, stationary process that satisfies standard regularity conditions.

2 Maximum Likelihood Estimation

We begin by formalizing the conventions of Section 1.

DEFINITION 1 Let $\{y_t\}_{t=-\infty}^{\infty}$ with $y_t : \Omega \to \mathbb{R}^M$ be a stationary process defined on a complete probability space $(\Omega, \mathcal{F}, P)$. For contiguous subsequences $(y_s, \dots, y_t)$ from $\{y_t\}$, $s \le t = 0, \pm 1, \pm 2, \dots$, define

$$\mathcal{F}_s^t = \sigma(y_s, \dots, y_t),$$

where $\sigma(y_s, \dots, y_t)$ denotes the smallest, complete sub-σ-algebra of $\mathcal{F}$ such that the random variables $(y_s, \dots, y_t)$ are measurable.

DEFINITION 2 Norms are distinguished as follows:

$$\begin{aligned} \|X\|_r &= \left(\sum_{i=1}^{k} \mathcal{E}|X_i|^r \right)^{1/r} \\ \|X\| &= |X| = \left(\sum_{i=1}^{k} |X_i|^2 \right)^{1/2} \end{aligned}$$

REMARK 1 A consequence of Hölder's inequality, $\mathcal{E}|XY| \le (\mathcal{E}|X|^{p'})^{1/p'}(\mathcal{E}|X|^{q'})^{1/q'}$ for $1/p' + 1/q' = 1$ and scalar X and Y, is that if $1 \le p \le q$, then convergence in L_q implies convergence in L_p. In particular $\|X_n - X\|_2 \to 0$ impales $\|X_n - X\|_1 \to 0$. This because if $Y = 1$, $p' = q/p$, $q' = q/(q-p)$, then $\mathcal{E}|X|^p = \mathcal{E}[|X|^{q/p'} \cdot 1] \le (\mathcal{E}|X|^q)^{1/p'} = (\mathcal{E}|X|^q)^{p/q}$. □

DEFINITION 3 For each ρ in a parameter space $\mathcal{R} \subset \mathbb{R}^{p_\rho}$ and for each $L = 1, 2, \dots,$ let $p_L(x, y, \rho) : \mathbb{R}^{ML} \times \mathbb{R}^M \times \mathcal{R} \to \mathbb{R}^+$ be a continuous probability density function with respect to Lebesgue measure on $\mathbb{R}^{ML} \times \mathbb{R}^{M}$. For $x \in \mathbb{R}^{ML}$ and $y \in \mathbb{R}^{M}$, define

$$\begin{aligned} p_L(x, \rho) &= \int_{\mathbb{R}^M} p_L(x, y, \rho)\, dy \\ p_L(y|x, \rho) &= \frac{p_L(x, y, \rho)}{p_L(x, \rho)} I_{[p_L(x,\rho)>0]}. \end{aligned}$$

Further, $p_L(x, y, \rho)$ satisfies the consistency condition

$$p_L(x, y, \rho) = \int_{\mathbb{R}^M} p_{L+1}[(u, x), y, \rho]\, du.$$

Throughout,

$$x_{t-1} = (y_{t-L}, \ldots, y_{t-1}). \tag{1}$$

We shall drop the subscript L when all arguments are given explicitly; e.g., $p(y_t|y_0, \ldots, y_{t-1}, \rho)$ or $p(y_s, \ldots, y_t, \rho)$. With these conventions,

$$\begin{aligned} \mathcal{P}_\rho &= \{p(y_s, \ldots, y_t, \rho) : s \le t = 0, \pm 1, \pm 2, \ldots;\ \rho \in \mathcal{R}\} \\ &= \{p_L(x_{t-1}, \rho) : t = 0, \pm 1, \pm 2, \ldots;\ L = 1, 2, \ldots;\ \rho \in \mathcal{R}\}. \end{aligned}$$

DEFINITION 4 The quasi maximum likelihood estimator (qmle) is

$$\hat{\rho}_n = \operatorname*{argmax}_{\rho \in \mathcal{R}} Q_n(\rho),$$

where

$$Q_n(\rho) = p_L(x_{L-1}, \rho) \prod_{t=L}^{n} p_L(y_t|x_{t-1}, \rho).$$

The maximum likelihood estimator (mle) is

$$\tilde{\rho}_n = \operatorname*{argmax}_{\rho \in \mathcal{R}} L_n(\rho),$$

where

$$L_n(\rho) = p(y_0, \rho) \prod_{t=1}^{n} p(y_t|y_0, \ldots, y_{t-1}, \rho).$$

The following quantities are needed later to define the asymptotic variances of the maximum likelihood and quasi maximum likelihood estimators.

DEFINITION 5 Let ρ^o denote the true value of the parameter ρ. Define

$$
\begin{aligned}
S_{t,L} &= \frac{\partial}{\partial \rho} \log p_L(y_t | x_{t-1}, \rho^o) \\
S_{t,t} &= S_{t,L} \Big|_{L=t} = \frac{\partial}{\partial \rho} \log p(y_t | y_0, \ldots, y_{t-1}, \rho^o) \\
\mathcal{V}^o_{L,\tau} &= \mathcal{E}\big(S_{t,L}\, S'_{t-\tau,L}\big) && (2) \\
\mathcal{V}^o_{L} &= \mathcal{V}^o_{L,0} + \sum_{\tau=1}^{\infty} \mathcal{V}^o_{L,\tau} + \left(\sum_{\tau=1}^{\infty} \mathcal{V}^o_{L,\tau} \right)' && (3) \\
\mathcal{V}^o &= \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}\big(S_{t,t}\, S'_{t,t}\big) && (4)
\end{aligned}
$$

Regularity conditions for estimation of the parameters of a dynamic model by quasi maximum likelihood and by maximum likelihood are stated in Gallant (1987), Gallant and White (1988), Davidson (1994), and Pötscher and Prucha (1996) and do not differ substantively among authors. For both qmle and mle, the following mixing condition is standard in this literature:

ASSUMPTION 1 The process $\{y_t\}_{t=-\infty}^{\infty}$ is strong mixing of size $-4r/(r-4)$ for some $r > 4$ with respect to the filtration $\{\mathcal{F}_{-\infty}^{t}\}_{t=-\infty}^{\infty}$.

This assumption is interpreted as follows. The measure of dependence between two σ-algebras used in strong mixing is

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup_{F \in \mathcal{F}, G \in \mathcal{G}} |P(F \cap G) - P(F)P(G)|.$$

The mixing coefficient of $\{y_t\}_{t=-\infty}^{\infty}$ is

$$\alpha_\tau = \sup_{-\infty < t < \infty} \alpha(\mathcal{F}_{t+\tau}^{\infty}, \mathcal{F}_{-\infty}^{t}).$$

The process $\{y_t\}_{t=-\infty}^{\infty}$ is said to be strong mixing of size $-q$ with respect to the filtration $\{\mathcal{F}_{-\infty}^{t}\}_{t=-\infty}^{\infty}$ if there is a $\delta > 0$ such that $\alpha_\tau = \mathcal{O}(\tau^{-q-\delta})$.

The regularity conditions for qmle may be summarized as follows:

ASSUMPTION 2 The finite dimensional densities of $\{y_t\}_{t=-\infty}^{\infty}$ are in the family $\mathcal{P}_\rho$ for some ρ^o in $\mathcal{R}$. For each L, the parameter space $\mathcal{R}$ contains the closure $\bar{\mathcal{R}}_L^o$ of an open ball $\mathcal{R}_L^o$ containing ρ^o, and the Kullback-Leibler discrepancy

$$\bar{q}_L(\rho) = \int\!\!\int \log p_L(y|x,\rho)\, p_L(x,y,\rho^o)\, dy dx$$

has an isolated minimum over $\bar{\mathcal{R}}_L^o$ at ρ^o. The matrix $\mathcal{V}_L^o$ given by (3) is nonsingular. Let $g_t(y_{t-L}, \ldots, y_t, \rho)$ represent either the function $\log p_L(y|x,\rho)$, its first partial derivatives with respect to the elements of ρ, the cross products of its first partial derivatives, or its second partial derivatives. The family $\{g_t\}_{t=0}^{\infty}$ is Near Epoch Dependent on $\{y_t\}$ of size $-2(r-2)/(r-4)$ for each $\rho \in \bar{\mathcal{R}}^o$, where r is that of Assumption 1, is generalized Lipschitz, and is dominated by a sequence of random variables $\{d_t\}_{t=0}^{\infty}$ with bounded r-th moment in the sense that $\sup_{\rho\in\bar{\mathcal{R}}^o} |g_t(y_{t-L}, \ldots, y_t, \rho)| \leq d_t$ and $\|d_t\|_r \leq \Delta < \infty$ for all t.

For definitions of Near Epoch Dependent and generalized Lipschitz (aka A-smooth) conditions, see[2] Gallant (1987, pp. 496, 515) Gallant and White (1988, pp. 21, 27), Davidson (1994, pp. 261, 339), or Pötscher and Prucha (1996, pp. 38, 50).

The expression for the asymptotic variance of the maximum likelihood estimator simplifies through elimination of quantities involving second derivatives when integration and differentiation interchange, and it is customary to impose this interchange condition when deriving the asymptotics of the maximum likelihood estimator. For the same reason, it is also convenient for the quasi maximum likelihood estimator.

ASSUMPTION 3

$$\begin{aligned} \int \frac{\partial}{\partial\rho}\, p_L(x,\rho)\, dx\Big|_{\rho=\rho^o} &= \frac{\partial}{\partial\rho}\int p_L(x,\rho)\, dx\Big|_{\rho=\rho^o} \\ \int \frac{\partial^2}{\partial\rho\partial\rho'}\, p_L(x,\rho)\, dx\Big|_{\rho=\rho^o} &= \frac{\partial}{\partial\rho}\int \frac{\partial}{\partial\rho'}\, p_L(x,\rho)\, dx\Big|_{\rho=\rho^o} \end{aligned}$$

Similarly for $p(y_0, \ldots, y_t, \rho)$.

The asymptotic distribution of the qmle is

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\big[0, (\mathcal{V}_{L,0}^o)^{-1}(\mathcal{V}_L^o)(\mathcal{V}_{L,0}^o)^{-1}\big]$$

under Assumptions 1, 2, and 3.

The regularity conditions for mle may be may be summarized as follows:

[2]A preprint of the relevant chapter of Gallant (1987) is https://repository.lib.ncsu.edu/bitstream/handle/1840.4/8431/ISMS_1985_1667.pdf.

ASSUMPTION 4 The finite dimensional densities of $\{y_t\}_{t=-\infty}^{\infty}$ are in the family $\mathcal{P}_\rho$ for some ρ^o in $\mathcal{R}$. The parameter space $\mathcal{R}$ contains the closure $\bar{\mathcal{R}}^o$ of an open ball $\mathcal{R}^o$ centered at ρ^o and

$$\bar{l}(\rho) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{E} \log p(y_t | y_0, \ldots, y_{t-1}, \rho),$$

has an isolated minimum over $\bar{\mathcal{R}}^o$ at ρ^o. The matrix $\mathcal{V}^o$ given by (4) is finite and nonsingular. Let $g_t(y_0, \ldots, y_t, \rho)$ represent either the function $\log p(y_t | y_0, \ldots, y_{t-1}, \rho)$, its first partial derivatives with respect to the elements of ρ, the cross products of its first partial derivatives, or its second partial derivatives. The family $\{g_t\}_{t=0}^{\infty}$ is Near Epoch Dependent on $\{y_t\}$ of size $-2(r-2)/(r-4)$ for each $\rho \in \bar{\mathcal{R}}^o$, where r is that of Assumption 1, is generalized Lipschitz, and is dominated by a sequence of random variables $\{d_t\}_{t=0}^{\infty}$ with bounded r-th moment in the sense that $\sup_{\rho\in\bar{\mathcal{R}}^o} |g_t(y_0, \ldots, y_t, \rho)| \leq d_t$ and $\|d_t\|_r \leq \Delta < \infty$ for all t.

Under Assumptions 1, 3, and 4,

$$\sqrt{n}(\tilde{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\left[0, (\mathcal{V}^o)^{-1}\right].$$

3 A Score Vector on the Infinite Past

In this section we construct a score vector on the infinite past for the family $\mathcal{P}_{\rho^o}$.

LEMMA 1 Assumption 3 implies

$$\mathcal{E}\left[\frac{\partial}{\partial\rho} \log p(y_\sigma, \ldots, y_\tau, \rho^o) \,\Big|\, \mathcal{F}_s^t\right] = \frac{\partial}{\partial\rho} \log p(y_s, \ldots, y_t, \rho^o) \tag{5}$$

for every $\sigma \leq s \leq t \leq \tau$.

Assumptions 3 and 4 imply

$$\left\|\mathcal{E}\left[\frac{\partial}{\partial\rho} \log p(y_\sigma, \ldots, y_t, \rho^o) \,\Big|\, \mathcal{F}_s^\tau\right] - \frac{\partial}{\partial\rho} \log p(y_s, \ldots, y_t, \rho^o)\right\|_2 = a_{t-s} \tag{6}$$

for every $\sigma \leq s \leq t \leq \tau$, where[3] $a_{t-s} = \mathcal{O}\left(|t-s|^{-2(r-2)/(r-4)-\delta}\right)$ for some $\delta > 0$.

[3]The rate of convergence of a_{t-s} follows naturally from Assumptions 1, 2, and 4 but that all is relevant in this section is that $a_{t-s} = o(1)$, i.e., $a_{t-s} \to 0$ as $t-s \to \infty$.

Proof Let $u = (y_\sigma, \ldots, y_{s-1})$, $v = (y_s, \ldots, y_t)$, and $w = (y_{t+1}, \ldots, y_\tau)$.

$$
\begin{aligned}
&\mathcal{E}\left[\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_\tau, \rho^o) \,\middle|\, \mathcal{F}_s^t\right] \\
&\qquad = \int\int \frac{\partial}{\partial\rho}\log p(u,v,w,\rho)\,\frac{p(u,v,w,\rho^o)}{p(v,\rho^o)}\,dudw\Big|_{\rho=\rho^o} \\
&\qquad = \int\int \frac{\partial}{\partial\rho}p(u,v,w,\rho)\,\frac{p(u,v,w,\rho^o)}{p(u,v,w,\rho^o)p(v,\rho^o)}\,dudw\Big|_{\rho=\rho^o} \\
&\qquad = \frac{\partial}{\partial\rho}\int\int \frac{p(u,v,w,\rho)}{p(v,\rho^o)}\,dudw\Big|_{\rho=\rho^o} \\
&\qquad = \frac{(\partial/\partial\rho)p(v,\rho)}{p(v,\rho^o)}\Big|_{\rho=\rho^o} \\
&\qquad = \frac{\partial}{\partial\rho}\log p(v,\rho^o).
\end{aligned}
$$

In the definition of Near Epoch Dependence (Gallant, 1987, pp. 496–7), replace t in that definition by ℓ here, put ℓ midway between s and t and m such that $[\ell - m, \ell + m] = [s, t]$, put $k_\ell = \tau - \sigma$, $W_\ell(\ldots, y_\sigma, \ldots, y_\tau, \ldots) = (y_\sigma, \ldots, y_\tau) \subset \mathbb{R}^{k_\ell}$, and, for any λ, put $g(W_\ell) = \mathcal{E}\left[\lambda' \frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^\tau\right]$. Then, because $\mathcal{F}_s^t \subset \mathcal{F}_s^\tau$ implies

$$
\mathcal{E}\left\{\mathcal{E}\left[\lambda'\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^\tau\right] \,\middle|\, \mathcal{F}_s^t\right\} = \mathcal{E}\left[\lambda'\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^t\right],
$$

Near Epoch Dependence implies that

$$
\left\|\mathcal{E}\left[\lambda'\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^\tau\right] - \mathcal{E}\left[\lambda'\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^t\right]\right\|_2 = a_{t-s}.
$$

Apply Equation 5 to $\mathcal{E}\left[\lambda'\frac{\partial}{\partial\rho}\log p(y_\sigma, \ldots, y_t, \rho^o) \,\middle|\, \mathcal{F}_s^t\right]$ to get Equation 6. □

An implication of Equation 5 of Lemma 1 is that the process

$$
M_t = \frac{\partial}{\partial\rho}\log p(y_0, ..., y_t, \rho^o) \quad t = 1, 2, ...
$$

is a square integrable martingale with respect to the filtration $\mathcal{F}_0^t$, which is a well known fact (Hall and Heyde, 1980, p. 157). A less well known implication, if known elsewhere at all, is that if time is reversed by considering lags to be the index set, then process that results

$$
M_L = \frac{\partial}{\partial\rho}\log p(y_{t-L}, ..., y_t, \rho^o) \quad L = 1, 2, ...
$$

is a square integrable martingale with respect to the filtration $\mathcal{F}_L = \mathcal{F}_{t-L}^t$.

We next establish the existence of a score on the infinite past:

THEOREM 1 Let Assumptions 2, 3, and 4 hold. Then there exists

$$S_{t,\infty} \in L_2(\Omega, \mathcal{F}, P)$$

such that

$$\lim_{L \to \infty} \|S_{t,L} - S_{t,\infty}\| = 0.$$

Proof Let $\lambda \neq 0$ be given[4] and let $J \geq I \geq L$. By Equation 5 of Lemma 1,

$$\begin{aligned}
&\mathcal{E}\left[(\lambda' S_{t,I})\left(\lambda' S_{t,J} - \lambda' S_{t,I}\right)\right] \\
&= \mathcal{E}\left\{(\lambda' S_{t,I})\left[\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, y_t, \rho^o) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, y_t, \rho^o)\right]\right\} \\
&\quad - \mathcal{E}\left\{(\lambda' S_{t,I})\left[\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o)\right]\right\} \\
&= \mathcal{E}\left\{(\lambda' S_{t,I})\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, y_t, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, y_t, \rho^o) \middle| \mathcal{F}_{t-I}^t\right)\right]\right\} \\
&\quad - \mathcal{E}\left\{(\lambda' S_{t,I})\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right)\right]\right\} \\
&= -\mathcal{E}\left\{(\lambda' S_{t,I})\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right)\right]\right\}
\end{aligned}$$

By the above and Equation 6 of Lemma 1,

$$\begin{aligned}
&\left\| \mathcal{E}\left[(\lambda' S_{t,I})\left(\lambda' S_{t,J} - \lambda' S_{t,I}\right)\right] \right\| \qquad (7) \\
&\leq \mathcal{E}\left\{\|\lambda' S_{t,I}\| \left\| \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right)\right\|\right\} \\
&\leq \mathcal{E}\left\{\|\lambda' S_{t,I}\| \left\| \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \right| \right\|\right\} \\
&\quad + \mathcal{E}\left\{\|\lambda' S_{t,I}\| \left\| \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^t\right) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o)\right\|\right\} \\
&= a_I + a_I
\end{aligned}$$

where $a_I = o(1)$.

Note that

$$0 \leq \|\lambda' S_{t,I} - \lambda' S_{t,J}\|_2^2 = \mathcal{E}(\lambda' S_{t,J})^2 - \mathcal{E}(\lambda' S_{t,I})^2 - 2\mathcal{E}\left[(\lambda' S_{t,I})\left(\lambda' S_{t,J} - \lambda' S_{t,I}\right)\right] \qquad (8)$$

[4] Uniformity in λ will not be required because this is an application of the Cramer-Wold device.

By Assumption 2 the sequence $\{\mathcal{E}(\lambda' S_{t,L})^2\}_{L=1}^{\infty}$ is bounded whence the positive, bounded sequence $\{\mathcal{E}(\lambda' S_{t,L})^2\}_{L=1}^{\infty}$ has at least one limit point. If more than one, let $a < b$ denote two of them and choose subsequences $I_i < J_i$ such that $\lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 = b$, and $\lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,J_i})^2 = a$. From Equations 7 and 8 we would then have

$$\begin{aligned}
b &= \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 + 0 \\
&= \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 + \lim_{i\to\infty} 2\mathcal{E}\left[(\lambda' S_{t,I_i})(\lambda' S_{t,J_i} - \lambda' S_{t,I_i})\right] \quad \text{(Equation 7)} \\
&\leq \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 + \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,J_i})^2 - \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 \quad \text{(Equation 8)} \\
&= \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,J_i})^2 \\
&= a
\end{aligned}$$

which is a contradiction. Therefore $\lim_{L\to\infty} \mathcal{E}(\lambda' S_{t,L})^2$ exists and given $\epsilon > 0$ we may choose L large enough that $I, J > L$ implies

$$0 \leq \|\lambda' S_{t,I} - \lambda' S_{t,J}\|_2^2 = \mathcal{E}(\lambda' S_{t,J})^2 - \mathcal{E}(\lambda' S_{t,I})^2 - 2\mathcal{E}\left[(\lambda' S_{t,I})(\lambda' S_{t,J} - \lambda' S_{t,I})\right] < \epsilon. \tag{9}$$

Equation 9 implies $\{\lambda' S_{t,L}\}_{L=1}^{\infty}$ is Cauchy. Therefore $\{\lambda' S_{t,L}\}_{L=1}^{\infty}$ has an $L_2(\Omega, \mathcal{F}, P)$ limit. □

REMARK 2 Let $(\mathbb{R}^\infty, \mathcal{A}, P_\rho)$ denote the probability space induced on $\mathbb{R}^\infty = \times_{t=-\infty}^{\infty}\mathbb{R}$ by the finite dimensional densities $\mathcal{P}_\rho$ via the Daniel-Kolmogorov construction (Tucker, 1967, p. 30). Inspection of the proof of Theorem 1 reveals that $S_{t,\infty}$ can be viewed as a random variable defined on the probability space $(\mathbb{R}^\infty, \mathcal{A}, P_{\rho^o})$ with infinite dimensional argument $(y_t, y_{t-1}, \ldots)$. Because $(\mathbb{R}^\infty, \mathcal{A}, P_{\rho^o})$ is the range space of the random variables $\{y_t(\omega)\}_{t=-\infty}^{\infty}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, $S_{t,\infty}$ is a composite function that is ultimately defined on $(\Omega, \mathcal{F}, P)$. □

4 Asymptotic Efficiency

In this section we deduce some properties of the score and then prove our main result.

REMARK 3 Below we use some miscellaneous facts. From Remark 1, if $X_L \to X$ and $Y_L \to Y$ in $L_2(\Omega, \mathcal{F}, P)$, then $X_L Y_L \to XY$ in $L_1(\Omega, \mathcal{F}, P)$. Because $\int |X_L Y_L - XY|\, dP \to 0$ implies $\int (X_L Y_L - XY)\, dP \to 0$ and $X_L Y_L$ and XY in $L_1(\Omega, \mathcal{F}, P)$ implies $\int X_L Y_L\, dP$ and

$\int XY\,dP$ exist, we have $X_L \to X$ and $Y_L \to Y$ in $L_2(\Omega, \mathcal{F}, P)$ implies $\mathcal{E}X_L Y_L \to \mathcal{E}XY$. By putting $Y_L = Y$, we have $X_L \to X$ in L_2 implies $\mathcal{E}X_L Y \to \mathcal{E}XY$. By putting $Y_L = Y = 1$, we have $X_L \to X$ in L_2 implies $\mathcal{E}X_L \to \mathcal{E}X$. □

LEMMA 2 Let Assumptions 2, 3, and 4 hold. If $s < t$, then

$$\mathcal{E}(S_{t,\infty}) = 0$$

$$\mathcal{E}(S_{s,\infty} S'_{t,\infty}) = 0$$

$$\mathcal{E}(S_{s,\infty} S'_{s,\infty}) = \mathcal{E}(S_{t,\infty} S'_{t,\infty}).$$

Proof A standard result from the theory of maximum likelihood estimation is

$$\mathcal{E}(S_{t,L} | \mathcal{F}_{t-L}^{t-1}) = \int \frac{\partial}{\partial \rho} \log p_L(y|x_{t-1}, \rho)\, p_L(y|x_{t-1}, \rho^o)\, dy \,\Big|_{\rho=\rho^o} = 0, \tag{10}$$

whence $\mathcal{E}(S_{t,L}) = 0$ for every L. Because $S_{t,L} \to S_{t,\infty}$ in L_2, we have that

$$0 = \lim_{L \to \infty} \mathcal{E}(S_{t,L}) = \mathcal{E}(S_{t,\infty}).$$

For $s < t$ and large J,

$$\mathcal{E}(S_{s,L} S'_{t,J} | \mathcal{F}_{t-J}^{t-1}) = S_{s,L} \mathcal{E}(S'_{t,J} | \mathcal{F}_{t-J}^{t-1}) = 0,$$

whence $\mathcal{E}(S_{s,L} S'_{t,J}) = 0$. Because $S_{t,J} \to S_{t,\infty}$ in L_2, we have that

$$0 = \lim_{J \to \infty} \mathcal{E}(S_{s,L} S'_{t,J}) = \mathcal{E}(S_{s,L} S'_{t,\infty}).$$

Because $\mathcal{E}(S_{s,L} S'_{t,\infty}) = 0$ for all L and $S_{t,L} \to S_{t,\infty}$ in L_2, we have that

$$0 = \lim_{L \to \infty} \mathcal{E}(S_{s,L} S'_{t,\infty}) = \mathcal{E}(S_{s,\infty} S'_{t,\infty}).$$

Lastly, $S_{s,L} \to S_{s,\infty}$ and $S_{t,L} \to S_{t,\infty}$ in L_2 together with $\mathcal{E}(S_{s,L} S'_{s,L}) = \mathcal{E}(S_{t,L} S'_{t,L})$ because of stationarity, we have

$$\mathcal{E}(S_{s,\infty} S'_{s,\infty}) = \lim_{L \to \infty} \mathcal{E}(S_{s,L} S'_{s,L}) = \lim_{L \to \infty} \mathcal{E}(S_{t,L} S'_{t,L}) = \mathcal{E}(S_{t,\infty} S'_{t,\infty}).$$

□

Lemma 2 implies

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} S_{t,\infty} \right) = \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}\big(S_{t,\infty} S'_{t,\infty}\big) = \mathcal{E}\big(S_{0,\infty} S'_{0,\infty}\big),$$

which permits the following definition.

DEFINITION 6

$$\mathcal{V}_{\infty}^{o} = \mathcal{E}\big(S_{0,\infty}S_{0,\infty}'\big) = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} S_{t,\infty}\right). \tag{11}$$

LEMMA 3 If Assumptions 1, 2, 3, and 4 hold, then

$$\lim_{L\to\infty}\sum_{\tau=1}^{\infty}\mathcal{E}(S_{t,L}S_{t-\tau,L}') = 0$$

Proof A standard mixing inequality is

$$\left\|\mathcal{E}(\lambda' S_{t,L}|\mathcal{F}_{-\infty}^{t-\tau})\right\| \leq 2(2^{1/2}+1)\left\|\lambda' S_{t,L}\right\|_4 \left[\alpha(\mathcal{F}_{-\infty}^{t-\tau},\mathcal{F}_{t-L}^{\infty})\right]^{1/2-1/4};$$

see, e.g., Gallant (1987, p. 507) or Davidson (1994, p. 211). Assumption 1 implies

$$\sup_{-\infty<t<\infty}\alpha(\mathcal{F}_{-\infty}^{t-\tau},\mathcal{F}_{t-L}^{\infty}) = \mathcal{O}\big[(\tau-L)^{-4r/(r-4)}\big].$$

Assumption 4 bounds $\left\|\lambda' S_{t,L}\right\|_4$ uniformly in t. Applying Cauchy-Schwarz and the above, there is a $\delta>0$ such that

$$\big|\,\mathcal{E}\big[(\lambda' S_{t-\tau,L})\,\mathcal{E}(\lambda' S_{t,L}|\mathcal{F}_{-\infty}^{t-\tau})\big]\big| \leq \left\|\lambda' S_{t-\tau,L}\right\|_2 \left\|\mathcal{E}(\lambda' S_{t,L}|\mathcal{F}_{-\infty}^{t-\tau})\,\right\|_2 \leq \mathcal{O}\big[(\tau-L)^{-1-\delta}\big].$$

Therefore, because $S_{t-\tau,L}$ is measurable $\mathcal{F}_{-\infty}^{t-\tau}$,

$$\sum_{\tau=1}^{\infty}\lambda'\mathcal{E}(S_{t-\tau,L}S_{t,L}')\lambda = \sum_{\tau=1}^{2L}\lambda'\mathcal{E}(S_{t-\tau,L}S_{t,L}')\lambda + \mathcal{O}\left(L^{-\delta}\right).$$

Let $u=(y_{t-\tau-L},\ldots,y_{t-\tau-1})$ and $v=(y_{t-L},\ldots,y_t)$;

$$\mathcal{E}(S_{t,L}S_{t-\tau,L}') = \mathcal{E}[\mathcal{E}(S_{t,L}S_{t-\tau,L}'|\mathcal{F}_{u\cup v})] = \mathcal{E}[S_{t,L}\mathcal{E}(S_{t-\tau,L}'|\mathcal{F}_{u\cup v})].$$

In the definition of Near Epoch Dependence (Gallant, 1987, pp. 496–7), choose ℓ and m such that $[\ell-m,\ell+m]=[t-\tau-L,t-\tau-1]$, put $k_\ell=(t-\tau-1)-(t-\tau-L)$, $W_\ell(\ldots,y_{t-\tau-L},\ldots,y_t,\ldots)=(y_{t-\tau-L},\ldots,y_{t-\tau-1})\in\mathbb{R}^{k_\ell}$ and, for any λ, put $g_\ell(W_\ell)=\mathcal{E}(\lambda' S_{t-\tau,L}'|\mathcal{F}_{u\cup v})$. Because $\mathcal{F}_u\subset\mathcal{F}_{u\cup v}$,

$$\mathcal{E}\left[\mathcal{E}(\lambda' S_{t-\tau,L}'|\mathcal{F}_{u\cup v})\middle|\mathcal{F}_u\right] = \mathcal{E}(\lambda' S_{t-\tau,L}'|\mathcal{F}_u).$$

Near Epoch Dependence implies

$$\left\|\mathcal{E}(\lambda' S_{t-\tau,L}|\mathcal{F}_{u\cup v}) - \mathcal{E}(\lambda' S_{t-\tau,L}|\mathcal{F}_u)\right\|_2 = \left\|\mathcal{E}(\lambda' S_{t-\tau,L}|\mathcal{F}_{u\cup v}) - 0\right\|_2 = \left\|a_L\right\|_2 = \mathcal{O}\left(L^{-2(r-2)/(r-4)-\delta}\right)$$

for some $\delta > 0$. Then, because of the domination condition in Assumption 4

$$\lim_{L\to\infty} \sum_{\tau=1}^{\infty} \mathcal{E}(\lambda' S_{t,L} S'_{t-\tau,L} \lambda) = \lim_{L\to\infty} \sum_{\tau=1}^{2L} \mathcal{E}(\lambda' S_{t,L} a_L) + \lim_{L\to\infty} \mathcal{O}\left(L^{-\delta}\right) = \lim_{L\to\infty} 2L\Delta\, a_L = 0.$$

□

THEOREM 2 Assumptions 1, 2, 3, and 4 imply that

$$\mathcal{V}^o = \lim_{L\to\infty} \mathcal{V}^o_{L,0} = \lim_{L\to\infty} \mathcal{V}^o_L = \mathcal{V}^o_\infty,$$

where, from Equations 4, 2, 3, and 11,

$\mathcal{V}^o = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}\left(S_{t,t} S'_{t,t}\right)$, variance of the mle score

$\mathcal{V}^o_{L,\tau} = \mathcal{E}\left(S_{t,L}\, S'_{t-\tau,L}\right)$, autocovariance function of the qmle score

$\mathcal{V}^o_L = \mathcal{V}^o_{L,0} + \sum_{\tau=1}^{\infty} \mathcal{V}^o_{L,\tau} + \left(\sum_{\tau=1}^{\infty} \mathcal{V}^o_{L,\tau}\right)'$, variance of the qmle score

$\mathcal{V}^o_\infty = \mathcal{E}\left(S_{0,\infty} S'_{0,\infty}\right) = \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} S_{t,\infty}\right)$, variance of the infinite past score.

Proof Lemma 3 implies $\lim_{L\to\infty} \sum_{\tau=1}^{\infty} \mathcal{V}^o_{L,\tau} = 0$. Further, $S_{t,L} \to S_{t,\infty}$ in L_2 implies

$$\lim_{L\to\infty} \mathcal{V}^o_{L,0} = \lim_{L\to\infty} \mathcal{E}\left(S_{t,L}\, S'_{t,L}\right) = \mathcal{E}\left(S_{t,\infty}\, S'_{t,\infty}\right) = \mathcal{V}^o_\infty.$$

Therefore, $\lim_{L\to\infty} \mathcal{V}^o_L = \mathcal{V}^o_\infty$. By stationarity,

$$\lim_{t\to\infty} \mathcal{E}\left(S_{t,t} S'_{t,t}\right) = \lim_{t\to\infty} \mathcal{E}\left(S_{0,t} S'_{0,t}\right) = \lim_{L\to\infty} \mathcal{V}^o_L = \mathcal{V}^o_\infty$$

whence

$$\mathcal{V}^o = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}\left(S_{t,t} S'_{t,t}\right) = \mathcal{V}^o_\infty.$$

□

REMARK 4 Recall that the asymptotic distribution of the qmle is

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\left[0, (\mathcal{V}^o_{L,0})^{-1} (\mathcal{V}^o_L) (\mathcal{V}^o_{L,0})^{-1}\right]$$

and that the asymptotic distribution of the mle is

$$\sqrt{n}(\tilde{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\left[0, (\mathcal{V}^o)^{-1}\right].$$

□

Table 1. Infinite Lag Autoregression

	Estimate			Standard Deviation		
Lag	μ	σ	ϕ	μ	σ	ϕ
1	0.00018	1.00760	0.68999	0.03298	0.01622	0.12768
5	0.00013	1.00051	0.77845	0.02688	0.01586	0.06767
10	0.00015	0.99926	0.79608	0.02410	0.01582	0.06509
20	0.00015	0.99911	0.79257	0.02262	0.01587	0.06436
t-1	0.00014	0.99913	0.79118	0.02221	0.01578	0.06136

For $R = 5000$ repetitions, samples of length $n = 2000$ were generated according to the distribution shown in Equation 12 using $L = 2000$ to approximate an infinite number of lags. The quasi maximum likelihood estimate (L fixed) of $\rho = (\mu, \sigma, \phi)$ and the maximum likelihood estimate ($L = t - 1$) were computed for each repetition. Shown in the table are the means and standard deviations computed from these estimates for L as shown.

5 Simulation

Data were simulated according to the distribution with density

$$f[y_t|y_{t-L}, ..., y_{t-1}, (\mu, \sigma^2, \phi)] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(y_t - \mu - \frac{1}{2}(1-\phi)\sum_{j=1}^{L} \phi^j y_{t-j}\right)^2} \tag{12}$$

with $L = 2000$ to approximate $L = \infty$ and estimated by quasi maximum likelihood, L fixed, and maximum likelihood, $L = t - 1$, for sample size $n = 2000$ and repetitions $R = 5000$. Results are shown in Table 1. As one expects, estimates stabilize and standard errors decline as L increases.

The traditional methods of lag determination are the likelihood ratio test (Gallant, 1987, p. 591), the Akaike information criterion, AIC, (Akaike, 1969), the Hannan-Quinn criterion, HQ, (Hannan, 1987), and the Bayes information criterion, BIC, (Schwartz, 1978). The correct L according to AIC, HQ, or BIC is the L at which the criterion stops decreasing. Of these criteria, AIC will tend to select a larger L than HQ and HQ a larger L than BIC (Pötscher, 1989).

All of the above methods presume that the number of fitted parameters increases as L increases. However, one would prefer to impose stationarity on fitted likelihoods by imposing

Table 2. Finite Lag Autoregression

	Estimate			Standard Deviation		
Lag	μ	σ	ϕ	μ	σ	ϕ
1	0.00014	1.00488	0.72106	0.02726	0.01601	0.11996
5	0.00011	0.99911	0.78706	0.02234	0.01578	0.08121
10	0.00011	0.99978	0.69941	0.02067	0.01584	0.13126
20	0.00010	0.99988	0.67830	0.02021	0.01591	0.12524
t-1	0.00013	0.99992	0.67895	0.02010	0.01590	0.11917

For $R = 5000$ repetitions, samples of length $n = 2000$ were generated according to the distribution shown in Equation 12 using $L = 5$. The quasi maximum likelihood estimate (L fixed) of $\rho = (\mu, \sigma, \phi)$ and the maximum likelihood estimate ($L = t - 1$) were computed for each repetition. Shown in the table are the means and standard deviations computed from these estimates for L as shown.

either equality restrictions on the coefficients of the lags, as in the example, or inequality restrictions. In the former case this can violate the requirement that the number of parameters increases with L. In the latter case, computational complexity increases dramatically and estimates often lie close enough to the boundary to violate Assumption 2 or 4.

One possibility is to add the term τy_{t-L-1} to the location function and test that $\tau = 0$ using the likelihood ratio test or, equivalently, the t-test. Experimentation with several of the simulation repetitions described above indicates that the t-test will select a value of L around eight. The t-test is a bit erratic as L increases so some judgment may be required. Perhaps better is to estimate successively with increasing L until inspection of estimates suggests stability and further decline in standard errors is of no practical importance. With this approach experimentation indicates an L of about fifteen.

For Table 2 data were simulated according to Equation 12 with $L = 5$ and estimated by quasi maximum likelihood, L fixed, and maximum likelihood, $L = t - 1$, for sample size $n = 2000$ and repetitions $R = 5000$.

To repeat the above lag selection exercise, the t-test fairly definitively selects L equal to five. Inspection also suggests L equal to five. Moreover, standard errors do not decrease appreciably after five as they do in the infinite lag case.

Although lag determination is a detour from the main focus of this paper, it does appear that is possible to distinguish between infinite and finite lags in applications.

6 Conclusion

In this paper, we have constructed a score vector $S_{t,\infty}$ defined over the infinite past for a non-Markovian stationary process $\{y_t\}_{t=-\infty}^{\infty}$. We have shown that its variance does not depend on t and that this variance is the same as the asymptotic variance of the score for the maximum likelihood estimator. It is also the limit, as the number of lags L goes to infinity, of the asymptotic variance of the score for the quasi maximum likelihood estimator on L lags. The regularity conditions used to obtain these results are the standard regularity conditions for the asymptotics of the quasi maximum likelihood estimator and the maximum likelihood estimator in nonlinear dynamic models. The consequence of the above is that the asymptotic distribution of the quasi maximum likelihood estimator on L lags tends toward the asymptotic distribution of the maximum likelihood estimator as the number of lags increases.

7 References

Akaike, H. (1969), "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics* 21, 243–247.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford, U.K.

Duffie, D., and Singleton, K. J. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Gallant, A. R. (1987), *Nonlinear Statistical Models.* Wiley, New York.

Gallant, A. R., and Long, J. R. (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika*, 84, 125–141.

Gallant, A. R., and Tauchen, G. (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

Gallant, A. R., and White, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford, U.K..

Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Hall, P., and C. C. Heyde (1980), *Martingale Limit Theory and Its Applications,* Academic Press, New York.

Hannan, E. J. (1987), "Rational Transfer Function Approximation," *Statistical Science* 2, 1029–1054.

Ingram, B. F., and Lee, B. S. (1991), "Simulation Estimation of Time Series Models," *Journal of Econometrics* 47, 197–250.

Pötscher, B. M. (1989), "Model Selection Under Nonstationarity: Autoregressive Models and Stochastic Linear Regression Models," *Annals of Statistics* 17, 1257–1274.

Pötscher, B. M. and Prucha, I.R. (1996), *Dynamic Nonlinear Econometric Models*, Springer, Berlin.

Smith, A. A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics* 6, 461–464.

Tucker, H. G. (1967). *A Graduate Course in Probability.* Academic Press, New York.