Finite Lag Estimation of Non-Markovian Processes[1]

A. Ronald Gallant
Department of Economics, CB# 3305
University of North Carolina
Chapel Hill NC 27599, USA

Halbert White
Department of Economics, 0508
University of California, San Diego
La Jolla CA 92093, USA

August 1997
Last revised: November 1998

[1]Supported by the National Science Foundation. The corresponding author is A. Ronald Gallant, Department of Economics, University of North Carolina, CB# 3305, 6F Gardner Hall, Chapel Hill NC 27599-3305 USA; Phone 1-919-966-5338; FAX 1-919-966-4986; e-mail ron.gallant@unc.edu. Updated versions of this paper are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/papers as file qmle.ps, which is PostScript.

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. Under standard regularity conditions we derive a necessary and sufficient condition for the existence of a score vector defined over the infinite past. Upon imposing this condition, we show 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.

1 Introduction

We consider a stationary process

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

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

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

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 $A \in \Re^M$,

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

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 this situation (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 $p(y_t | y_{t-L}, \ldots, y_{t-1}, \rho^o)$ 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 $(\partial/\partial\rho) \log p(y_t | y_0, \ldots, y_{t-1}, \rho^o)$ 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 plan of the paper is as follows. In Section 2, we discuss the standard assumptions of nonlinear dynamic modeling in a setting applicable 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. 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.

ASSUMPTION 1 Let $\{y_t\}_{t=-\infty}^{\infty}$ with $y_t : \Omega \to \Re^M$ be a stationary process defined on a complete probability space $(\Omega, \mathcal{F}, P)$.

For contiguous subsequences $(y_s, \ldots, y_t)$ from $\{y_t\}$, $s \le t = 0, \pm 1, \pm 2, \ldots$, define

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

where $\sigma(y_s, \ldots, y_t)$ denotes the smallest complete sub-σ-algebra of $\mathcal{F}$ such that the random variables $(y_s, \ldots, y_t)$ are measurable. Throughout, for $\mathcal{F}$-measurable integrable g, $\mathcal{E}(g) = \int g(\omega)\, dP(\omega)$, $\|g\|_r = (\mathcal{E}|g|^r)^{1/r}$, and $\|g\| = \|g\|_2$.

ASSUMPTION 2 For each ρ in a parameter space $\mathcal{R} \subset \Re^{p_\rho}$ and for each $L = 1, 2, \ldots$, $p_L(\cdot, \rho) : \Re^{M(L+1)} \to \Re^+$ is a probability density function with respect to Lebesque measure on $\Re^{M(L+1)}$. For $x \in \mathcal{R}^{ML}$ and $y \in \mathcal{R}^M$, define

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

We assume that for each $\rho \in \mathcal{R}$, $(x, y) \to p_L(x, y, \rho)$ satisfies the consistency condition

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

Further, for each $(x, y) \in \Re^{M(L+1)}$, $p_L(x, y, \cdot) : \mathcal{R} \to \Re^+$, $p_L(x, \cdot) : \mathcal{R} \to \Re^+$, and $p_L(y|x, \cdot) : \mathcal{R} \to \Re^+$ are continuous on $\mathcal{R}$, $L = 1, 2, \ldots$.

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

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

PROPOSITION 1 **(a)** Given assumptions 1 and 2, there exists a function $\hat{\rho}_n : \Omega \to \mathcal{R}$ measurable – $\mathcal{F}$ such that

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

where

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

We call $\hat{\rho}_n$ the quasi-maximum likelihood estimator (qmle).

(b) Given assumptions 1 and 2, there exists a function $\tilde{\rho}_n : \Omega \to \mathcal{R}$ measurable – $\mathcal{F}$ such that

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

where

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

We call $\tilde{\rho}_n$ the maximum likelihood estimator (mle).

We assume that the model $\{\mathcal{P}_\rho,\ \rho \in \mathcal{R}\}$ is correctly specified in the following sense.

ASSUMPTION 3 The finite dimensional densities of $\{y_t\}_{t=-\infty}^{\infty}$ are in the family $\mathcal{P}_{\rho^o}$ 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.

Regularity conditions for consistent estimation of the parameters ρ^o of a dynamic model by quasi maximum likelihood and by maximum likelihood are stated in Gallant and White

(1988), Gallant (1987), Davidson (1994), Levine (1982), Pötscher and Prucha (1996) and White (1987, 1994) and do not differ substantively among authors. Under additional standard regularity conditions the qmle and the mle are asymptotically normal. Among other things, these conditions ensure that $\log p(y_t|y_0,\ldots,y_{t-1},\rho)$, is continuously differentiable of order two in ρ over $\bar{\mathcal{R}}_L^0$, and that its first and second partial derivatives with respect ρ, are dominated by a function $d_L(x,y)$ suitably integrable with respect to $p_L(x,y,\rho^o)$. In this literature, mixing conditions such as the following are standard.

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

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 these interchange conditions when deriving the asymptotics of the qmle and the mle. For the qmle these conditions are

ASSUMPTION 5

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

To define the asymptotic variances of the quasi maximum likelihood estimators, we define

$$\begin{aligned}
S_{t,L} &= \frac{\partial}{\partial\rho}\log p_L(y_t|x_{t-1},\rho^o) \\
\mathcal{V}_{L,\tau}^o &= \mathcal{E}\Big(S_{t,L}\,S_{t-\tau,L}'\Big) && (1) \\
\mathcal{V}_L^o &= \mathcal{V}_{L,0}^o + \sum_{\tau=1}^{\infty}\mathcal{V}_{L,\tau}^o + \left(\sum_{\tau=1}^{\infty}\mathcal{V}_{L,\tau}^o\right)'. && (2)
\end{aligned}$$

The matrix $\mathcal{V}_L^o$ given by (2) is assumed to be nonsingular. Under appropriate regularity conditions and with application of the interchange assumption, which eliminates quantities involving second derivatives from the asymptotic variance of the qmle, it can be shown that

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

For the mle the correct specification assumption is stated as follows:

ASSUMPTION 6 The finite dimensional densities of $\{y_t\}_{t=-\infty}^{\infty}$ are in the family $\mathcal{P}_{\rho^o}$ 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.

Standard regularity conditions that deliver asymptotic normality ensure that the log density $\log p(y_t|y_0, \ldots, y_{t-1}; \rho)$ is continuously differentiable over $\bar{\mathcal{R}}^0$ and that its first and second partial derivatives with respect to ρ are dominated by a function $d_t(y_0, \ldots, y_t)$ suitably integrable with respect to $p(y_0, \ldots, y_t)$.

As for the qmle, the asymptotic covariance of the mle simplifies under conditions permitting interchange of integral and derivative. Specifically,

ASSUMPTION 7

$$\begin{aligned}
\int \frac{\partial}{\partial\rho} p(y_0, \ldots, y_t; \rho)\, dy_0 \cdots dy_t\Big|_{\rho=\rho^o} &= \frac{\partial}{\partial\rho} \int p(y_0, \ldots, y_t; \rho)\, dy_0 \cdots dy_t\Big|_{\rho=\rho^o} \\
\int \frac{\partial^2}{\partial\rho\partial\rho'} p(y_0, \ldots, y_t; \rho)\, dy_0 \cdots dy_t\Big|_{\rho=\rho^o} &= \frac{\partial^2}{\partial\rho\partial\rho'} \int p(y_0, \ldots, y_t; \rho)\, dy_0 \cdots dy_t\Big|_{\rho=\rho^o}
\end{aligned}$$

To define the asymptotic covariance matrix for the mle, we define

$$\begin{aligned}
S_{t,t} &= \frac{\partial}{\partial\rho} \log p(y_t | y_0, \ldots, y_{t-1}, \rho^o) \\
\mathcal{V}^o &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}\Big(S_{t,t}\, S'_{t,t}\Big) \qquad (3)
\end{aligned}$$

The matrix $\mathcal{V}^o$ given by (3) is assumed to be nonsingular. For the mle, appropriate regularity conditions ensure that

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

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 5 implies

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

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

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

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

$\square$

LEMMA 2 Let Assumptions 3, 5, and 6 hold. Then there exists

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

such that

$$\lim_{L\to\infty}\|S_{t,L} - S_{t,\infty}\| = 0 \tag{4}$$

if and only if

$$\lim_{I\to\infty}\sup_{J>I}\,\mathcal{E}\left[\mathcal{E}\left(\lambda'\frac{\partial}{\partial\rho}\log p_J(x_{t-1},\rho^o)\Big|\,\mathcal{F}_{t-I}^t\right) - \lambda'\frac{\partial}{\partial\rho}\log p_I(x_{t-1},\rho^o)\right]^2 = 0 \tag{5}$$

for every $\lambda \ne 0$.

Proof Let $\lambda \ne 0$ be given and for $J \ge I \ge L$ let

$$Z_{t,IJ} = \mathcal{E}\left(\frac{\partial}{\partial\rho}\log p_J(x_{t-1},\rho^o)\Big|\,\mathcal{F}_{t-I}^t\right) - \frac{\partial}{\partial\rho}\log p_I(x_{t-1},\rho^o).$$

Using the law of iterated expectations and Lemma 1,

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

Then

$$
0 \;\le\; \|\lambda' S_{t,I} - \lambda' S_{t,J}\|^2 = \mathcal{E}(\lambda' S_{t,J})^2 - \mathcal{E}(\lambda' S_{t,I})^2 + 2\,\mathcal{E}[(\lambda' S_{t,I})\,(\lambda' Z_{t,IJ})]\,. \tag{6}
$$

Suppose that Equation 5 holds. By Assumption 6 the sequence $\{\mathcal{E}(\lambda' S_{t,L})^2\}_{L=1}^{\infty}$ is bounded. Therefore Equation 5 implies

$$
\lim_{I\to\infty} \sup_{J>I} 2\,\mathcal{E}[(\lambda' S_{t,I})\,(\lambda' Z_{t,IJ})] = 0.
$$

The positive, bounded sequence $\{\mathcal{E}(\lambda' S_{t,L})^2\}_{L=1}^{\infty}$ has at least one limit point. If there are more than one, let $a < b$ denote two of them and choose subsequences $L_i \le I_i < J_i$ such that $\lim_{i\to\infty} L_i = \infty$, $\lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 = b$, and $\lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,J_i})^2 = a$. From Equation 6 we would then have

$$
b = \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,I_i})^2 - \lim_{i\to\infty} 2\,\mathcal{E}[(\lambda' S_{t,I_i})\,(\lambda' Z_{t,I_iJ_i})] \le \lim_{i\to\infty} \mathcal{E}(\lambda' S_{t,J_i})^2 = a
$$

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

$$
0 \;\le\; \|\lambda' S_{t,I} - \lambda' S_{t,J}\|^2 = \mathcal{E}(\lambda' S_{t,J})^2 - \mathcal{E}(\lambda' S_{t,I})^2 + 2\,\mathcal{E}[(\lambda' S_{t,I})\,(\lambda' Z_{t,IJ})] < \epsilon. \tag{7}
$$

Equation 7 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.

Now suppose that $\{\lambda' S_{t,L}\}_{L=1}^{\infty}$ has an $L_2(\Omega, \mathcal{F}, P)$ limit. Then $\{\lambda' S_{t,L}\}_{L=1}^{\infty}$ is a Cauchy sequence and we have from Jensen's inequality, the law of iterated expectations, and Lemma 1, that

$$
\begin{aligned}
&\mathcal{E}\left[(\lambda' S_{t,J} - \lambda' S_{t,I})\right]^2 \\
&\geq \mathcal{E}\left[\mathcal{E}\left(\lambda' S_{t,J} - \lambda' S_{t,I} \middle| \mathcal{F}_{t-I}^{t}\right)\right]^2 \\
&= \mathcal{E}\Bigg[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, y_t, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, y_t, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) \\
&\qquad - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right)\Bigg]^2 \\
&= \mathcal{E}\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) - \mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right)\right]^2 \\
&= \mathcal{E}\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o)\right]^2
\end{aligned}
$$

□

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

In view of Lemma 2, we consider the following

ASSUMPTION 8

$$
\lim_{I \to \infty} \sup_{J > I} \mathcal{E}\left[\mathcal{E}\left(\lambda' \frac{\partial}{\partial \rho} \log p_J(x_{t-1}, \rho^o) \middle| \mathcal{F}_{t-I}^{t}\right) - \lambda' \frac{\partial}{\partial \rho} \log p_I(x_{t-1}, \rho^o)\right]^2 = 0
$$

for every $\lambda \neq 0$.

4 Asymptotic Efficiency

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

REMARK 2 Throughout we use these facts: 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 $L_1(\Omega, \mathcal{F}, P)$ (Chung, 1974, p. 70). 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 3 Let Assumptions 3, 5, 6 and 8 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

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

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 imply

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

□

Lemma 3 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 1

$$\mathcal{V}_{\infty}^{o} = \mathcal{E}\Big(S_{0,\infty}S_{0,\infty}'\Big) = \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} S_{t,\infty}\right). \tag{8}$$

LEMMA 4 If Assumptions 4, 3, 5, 6 and 8 hold, then

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

for every $\lambda \in \Re^{p_\rho}$.

Proof A standard mixing inequality is

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

see, e.g., Gallant (1987, p. 507) or Davidson (1994, p. 211). Assumption 4 implies

$$\sup_{-\infty<t<\infty}\alpha(\mathcal{F}_{t+\tau}^{\infty},\mathcal{F}_{-\infty}^{t}) = \mathcal{O}\left(\tau^{-4r/(r-4)}\right),$$

which is uniform in t by stationarity.

Assumption 6 bounds $\|\lambda' S_{t+\tau,L}\|_4$ uniformly in t. Applying Cauchy-Schwarz and the above, there is a $\delta > 0$ such that

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

where $\mathcal{O}\left(\tau^{-1-\delta}\right)$ does not depend upon L or t. Therefore

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

where $\mathcal{O}\left(T^{-\delta}\right)$ does not depend on L. By the same arguments as in the proof of Lemma 3 we have $\lim_{L\to\infty}\mathcal{E}(S_{t+\tau,L}S_{t,L}') = \mathcal{E}(S_{t+\tau,\infty}S_{t,\infty}') = 0$. Therefore,

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

where T is arbitrary. □

THEOREM 1 Assumptions 4, 3, 5, 6 and 8 imply

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

where the variance matrices are given by (3), (2), and (8), respectively.

Proof Recall that

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

where $\mathcal{V}^o_{L,\tau} = \mathcal{E}\Big(S_{t,L}\, S'_{t-\tau,L}\Big)$. Lemma 4 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

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

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

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

whence

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

□

5 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 maximum likelihood estimator. It is also the limit, as the number of lags L goes to infinity, of the asymptotic variance of 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 plus and additional condition that we show to be necessary and sufficient.

Of particular interest in these derivations are Lemmas 1, 2, and 3. Lemma 1 gives an explicit expression for the conditional expectation of the derivative of a log density when the conditioning variables are a subset of its arguments. Using this result, Lemma 2 provides a necessary and sufficient contditon for the existence of a score defined on the infinite past, which is the L_2 limit of the finite dimensional score. Lemma 3 derives some useful properties of this score.

6 References

Chung, K. L. (1974), *A Course in Probability Theory, Second Edition*, Academic Press, New York.

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.

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

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.

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