# INFORMATION PROPERTIES IN SPECTRAL ANALYSIS OF STATIONARY TIME SERIES

Wei Jiang[*], Shaochuan Cheng[†], Youmin Xi[†] and Shouyang Wang[††]

[*]*Hong Kong University of Science and Technology,* [†]*Xi'an Jiaotong University and* [††]*Chinese Academy of Sciences*

*Abstract:* In this paper, we investigate some information properties of parameter estimation in spectral analysis of stationary time series based on a geometrical framework. Stationary ARMA models are studied as a submanifold in the exponential family and the so-called Whittle estimator is analyzed in association with the embedded curvatures. Asymptotic behaviors such as information loss and bias of the estimator are shown to be dependent on the curvatures of this manifold. Simulation studies are performed to compare the estimation error in AR(1) models with the corresponding results in the time domain.

*Key words and phrases:* ARMA model, differential geometry, Fisher information, information loss, Whittle estimator.

## 1. Introduction

Let $\{x_t\}$ be a causal, stationary autoregressive moving-average process $[ARMA(p,q)]$ satisfying the difference equation

$$U(B)x_t = V(B)e_t, \tag{1.1}$$

where $B$ is the backshift operator, $U(B) = 1 - u_1 B - \cdots - u_p B^p$ and $V(B) = 1 - \nu_1 B - \cdots - \nu_q B^q$ are two polynomials in $B$ and $\{e_t\}$ is a sequence of i.i.d. random variables, referred to as innovations. Anderson (1977) proposed a general structure for Maximum Likelihood Estimation (MLE) in both time and frequency domains. The MLE in the frequency domain is actually equivalent to the so-called Whittle estimator in spectral analysis, and has been extensively studied in the case of Gaussian innovations (e.g. Whittle (1953), Brockwell and Davis (1991), Section 10.8). It is known that Whittle's estimator is asymptotically equivalent to the MLE in the time domain, but is computationally simpler and more efficient in the sense of $L_2$ even for non-Gaussian situations (Mikosch, Gadrich, Kluppelberg and Adler (1995)). The aim of the present paper is to study the information properties of this estimator, including information loss and asymptotic bias, in a differential geometric framework.

Information that statistics convey, or sufficiency in statistical inference, is an important criterion for constructing appropriate statistics. Information loss of the underlying statistics may heavily affect the accuracy of the parameters' attainability and the asymptotic inference. In the framework of statistical differential geometry, or the so-called information geometry, the effects of information loss on parameter estimation and inference have been studied. Rao (1945) first proposed Riemannian geometry on a family of distributions, in which the probability measure space is regarded as a manifold. Further, a metric tensor, which is basically the Fisher information matrix, is introduced on this manifold. It was structurally formalized by Amari (1982) as the expected geometry. Barndroff-Nielsen (1988) proposed the observed geometry for recovery of information loss and developed string theory for invariant asymptotic expansions. Taking a nonlinear regression model as a curve in the space of all regression models, Bates and Watts (1980) investigated the geometric structure of the nonlinear model, and successfully explained the fundamental role of intrinsic and parameter-effect curvatures in nonlinear problems.

For correlated observations, a few studies have been done in the geometric framework. In the time domain, given a finite sample of ARMA model (1.1), Ravishanker, Melnick and Tsai (1990) characterized ARMA models as members of the curved exponential family, developed the geometric calculus per observation up to order $O(1)$, and provided the estimation bias and revised hypothesis testing based on the geometric structure. By identifying the nonlinearity characteristics of the transformation function $W(z) = U^1(z)V(z)$, Ravishanker (1994) studied the effects of nonlinearity for various ARMA models. In the frequency domain, the key part of spectra is also a nonlinear function of parameters $W(z)$, which has exactly the same form as that in the time domain. As periodograms and spectra are intrinsic parts of the $L_2$ structure of stationary stochastic processes, this nonlinearity inevitably affects the accuracy of the estimates, e.g. the quadratic estimates of spectral density (Anderson (1994), Section 9.2). Therefore it is necessary to evaluate the effect of nonlinearity, especially in small samples. Note that unlike time domain analysis, where complicated and delicate numerical procedures may be required to evaluate the density function, frequency domain analysis using Whittle's estimator is simpler and has received considerable attention in recent years. The objective of the paper is to study the first- and second-order properties of Whittle's estimator and to investigate the relationship between its asymptotic performance and model nonlinearity via Edgeworth expansions (Taniguchi (1991)).

The paper is organized as follows. In Section 2, we introduce a few basic concepts in statistical geometry. In Section 3, we define a separable curved exponential family as a special case of the curved exponential family, and develop

its curvatures and their calculations. In Section 4, stationary time series observations are modeled as a curved exponential manifold in the frequency domain, and the stochastic expansion of MLE associated with the curvatures is given. Section 5 discusses the main results concerning the information loss of MLE and estimation bias, and simulation is performed to compare the estimation bias with that in the time domain.

## 2. Preliminaries

We give a brief introduction to a few basic concepts of statistical geometry in this section. More detailed mathematical treatments can be found in Amari (1985).

Consider a probability space as a set of probability distribution functions indexed by $\theta : \prod = \{f(x,\theta)|\theta \in \Theta\}$, given the observation $x = (x_1, \ldots, x_n)$ of a random variable $X$. Let $\Theta \subseteq R^p$ denote the parameter space of $\prod$ and $l(x,\theta) = \log f(x,\theta)$ denote the corresponding log-likelihood function. $\prod$ is defined as a *p-dimensional manifold* since there exists a map $\theta$ from $\prod$ to $\Theta$, and $\theta$ is also called a set of coordinates of $\prod$, which corresponds to a parameterization of the model.

A manifold $\prod$ becomes a *Riemannian manifold* if we define a *Riemannian metric tensor* $g(\theta)$ which measures the distance between two points. Rao (1945) and later Amari (1982) used the Fisher information matrix $(g(\theta))_{ij} = g_{ij}(\theta) = E_{f(x,\theta)}[\partial l(x,\theta)/\partial \theta_i \cdot \partial l(x,\theta)/\partial \theta_j]$ as such a measure. In a Riemannian manifold, a path $\gamma$ is defined as a smooth map from $[a,b] \subset R$ to $\prod$, i.e., $\gamma(t) : t \rightarrow (\gamma^1(t), \ldots, \gamma^p(t))$, and its length is calculated as

$$L(a,b) = \int_a^b \sqrt{\sum_{i,j} g_{ij}(\gamma(t)) \frac{d}{dt}\gamma^i(t) \frac{d}{dt}\gamma^j(t)} dt.$$

This length is invariant to a change of parameterization on $\prod$ due to the tensorial nature of the metric. A geodesic between two points in a Riemannian manifold is defined as a curve joining them with minimum path length, and the length is called their *geodesic distance*.

The geometric structure extended by Amari (1982) is defined by a pair of tensors $g_{ij}(\theta)$ and $T_{ijk}(\theta)$, where $T_{ijk} = E_{f(x;\theta)}[\partial l(x,\theta)/\partial \theta_i \cdot \partial l(x,\theta)/\partial \theta_j \cdot \partial l(x,\theta)/\partial \theta_k]$, and the *α-connections* are defined by $\Gamma_{ijk}^{\alpha} = \Gamma_{ijk}^0 - \alpha T_{ijk}/2$, where $\Gamma_{ijk}^0 = [\partial g_{jk}/\partial \theta_i + \partial g_{ik}/\partial \theta_j - \partial g_{ij}/\partial \theta_k]/2$. In the above structure, each α-connection defines a set of geodesics. In general, a geodesic is a solution of a set of differential equations determined by a connection. Intuitively, geodesics are the "straight lines" of geometry.

Based on the above definitions, we can define $\alpha$-*flatness*. A manifold $M$ is called $\alpha$-*flat* if a coordinate system $\theta$ exists such that, for all $\theta_1$ and $\theta_2$, the $\alpha$-geodesic joining them is of the form $\gamma(t) = (1-t)\theta_1 + t\theta_2$. Amari (1982) proved that any full exponential family is both +1-flat and -1-flat. Furthermore the point which is closest to a fixed point in $M$ in the sense of Kullback-Leibler divergence on a submanifold of an exponential family (*i.e.*, a curved exponential family), can be found by dropping a -1-geodesic which cuts the submanifold orthogonally.

## 3. Separable Exponential Geometry

Amari (1982) provided a substantial discussion of the exponential geometry. In the log-likelihood function of an exponential family

$$l(T, \theta) = \theta'T - \Psi(\theta), \tag{3.1}$$

where $T = (T_1, \ldots, T_n)'$ is a random vector and $\theta = (\theta_1, \ldots, \theta_n)'$ is a set of natural coordinates, the characteristic function $\Psi(\theta)$ plays an important role in evaluating all moments, in particular, $E(T_i) = \partial\Psi/\partial\theta_i$ and $\mathrm{Cov}\,(T_i, T_j) = \partial^2\Psi/\partial\theta_i\partial\theta_j$. When $\Psi(\theta)$ is a separable function of the components of $\theta$, i.e. $\Psi(\theta) = \sum \varphi_k(\theta_k)$, it is easily seen that the variables $\{T_i\}$ are uncorrelated with the covariance matrix

$$A^2 = \begin{bmatrix} \ddot{\varphi}_1(\theta_1) & & \\ & \ddots & \\ & & \ddot{\varphi}_n(\theta_n) \end{bmatrix}.$$

In this case, the geometry induced by $\Psi(\theta)$ is referred to as a *separable exponential geometry*, and the corresponding space of probability distributions is a *separable manifold*.

A *separable $(n,m)$-curved exponential manifold* is defined as a submanifold embedded in a separable manifold with a mapping from $\beta$ to $\theta$, $\theta = \theta(\beta) = (\theta_1(\beta), \ldots, \theta_n(\beta))'$, in which $\beta \in B \subset R^m$. Let $V(\beta) = (\partial\theta/\partial\beta)$, $W(\beta) = (\partial^2\theta/\partial\beta^2)$, where $V$ is a matrix of order $n \times m$, and $W$ is an array of order $n \times m \times m$ with elements $W_{iab} = \partial^2\theta_i/\partial\beta_a\partial\beta_b$ $(i = 1, \ldots, n;\ a, b = 1, \ldots, m)$. At a point $\beta_0$ of interest, consider a line $l_h$ in the direction $h$ that passes through $\beta_0$, *i.e.*, $\beta(b) = \beta_0 + bh$ where $b$ is a real value. There is a corresponding curve $C_h$ in the space of $\theta : \theta = \theta_h(b) = \theta(\beta_0 + bh)$, called a *lifted line*. The directional derivatives of $\eta$ in the direction $h$ at $\beta_0$ are defined by $\dot{\theta}_h = \partial\theta/\partial b = Vh$, and $\ddot{\theta}_h = \partial^2\theta/\partial b^2 = h'Wh$. As pointed out by Amari (1985), $\ddot{\theta}_h$ is a curvature vector that can be decomposed into tangent and normal components: $\ddot{\theta}_h = \ddot{\theta}_h^T + \ddot{\theta}_h^N$, and the intrinsic curvature and parameter-effects curvature can be defined as follows.

**Definition 3.1.** $K_h^N = \|\ddot{\theta}_h^N\|/\|\dot{\theta}_h\|^2 = \|(h'Wh)^N\|/(h'V'Vh)$ and $K_h^T = \|\ddot{\theta}_h^T\|/\|\dot{\theta}_h\|^2 = \|(h'Wh)^T\|/(h'V'Vh)$ are called the *intrinsic curvature* and *parameter-effects curvature* at $\beta_0$ in the direction of $h$, respectively.

For a separable exponential manifold, standardize T as

$$T = E(T) + \Lambda\varepsilon, \tag{3.2}$$

where $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = I_n$, $I_n$ is the unit matrix of order $n$. Introducing a linear transformation in $\theta : \eta = \Lambda(\theta_0)(\theta - \theta_0)$ at $\theta_0 = \theta(\beta_0)$, we get a new set of $\eta$-coordinates for this curved exponential manifold, where $l(T, \eta) = \theta(\eta)T - \Psi(\theta(\eta))$. The lifted line becomes $\eta = \eta_h(b) = \eta(\theta(\beta_0 + bh))$ in the new coordinates with the directional derivatives of $\eta$ in the direction of $h$ changed to $\dot{\eta}_h = \partial\eta/\partial b = \overline{V}h$ and $\ddot{\eta}_h = \partial^2\eta/\partial b^2 = h'\overline{W}h$, respectively, where $\overline{V} = \Lambda V$ and $\overline{W} = [\Lambda][W]$. Thus, with $\eta$-coordinates, the intrinsic curvature and the parameter-effects curvature at $\beta_0$ in the direction of $h$ change to

$$\overline{K}_h^N = \|(h'\overline{W}h)^N\|/(h'\overline{V}'\overline{V}h) \tag{3.3}$$

and

$$\overline{K}_h^T = \|(h'\overline{W}h)^T\|/(h'\overline{V}'\overline{V}h), \tag{3.4}$$

respectively. Since the transformation is linear which does not influence the assessment of curvatures, analogous to the discussion in Amari (1985), it is easy to prove that $\overline{K}_h^N$ and $\overline{K}_h^T$ are the same as $K_h^K$ and $K_h^T$, respectively.

To compute $\overline{K}_h^N$ and $\overline{K}_h^T$ at point $\beta_0$, we perform a $QR$ decomposition on $\overline{V}$:

$$\overline{V} = (\overline{Q}, \overline{H}) \begin{bmatrix} \overline{R} \\ 0 \end{bmatrix} = \overline{Q}\,\overline{R}, \tag{3.5}$$

and a transformation of $\beta : \vartheta = \overline{R}(\beta - \beta_0)$, or $\beta = \beta_0 + \overline{L}\vartheta$, where $\overline{L} = \overline{R}^{-1}$. This results in a new expression of $\eta$ with respect to $\vartheta : \eta(\theta) = \eta(\theta(\beta_0 + \overline{L}\vartheta))$, and $\beta_0$ becomes $\vartheta = 0$ with the derivatives changed to

$$\frac{\partial\eta}{\partial\vartheta} = \frac{\partial\eta}{\partial\theta} \cdot \frac{\partial\theta}{\partial\beta} \cdot \frac{\partial\beta}{\partial\vartheta} = \Lambda V \overline{L} = \overline{Q}, \qquad \frac{\partial^2\eta}{\partial\vartheta^2} = \overline{L}'\overline{W}\overline{L} := \overline{U}.$$

Replacing direction $h$ by $d = \overline{R}h$ and using $P^T = \overline{Q}\,\overline{Q}'$ and $P^N = \overline{H}\,\overline{H}'$ as the projection matrices in (3.3)-(3.4), the curvatures can be calculated as $K_d^N = \overline{K}_d^N = \|d'[\overline{H}'][\overline{U}]d\|$ and $K_d^T = \overline{K}_d^T = \|d'[\overline{Q}'][\overline{U}]d\|$. It is found that the following two curvature arrays are simpler and more important for further derivations.

**Definition 3.2.** $\overline{I} = [\overline{H}'][\overline{U}]$ and $\overline{P} = [\overline{Q}'][\overline{U}]$ are called the *intrinsic curvature array* and *parameter-effects curvature array*, respectively, where $\overline{I}$ is of order $(n-m) \times m \times m$, and $\overline{P}$ is of order $m \times m \times m$.

By denoting $\overline{\overline{H}} = \Lambda\overline{H}$ and $\overline{\overline{Q}} = \Lambda\overline{Q}$, we have $\overline{I} = \overline{L}'[\overline{\overline{H}}][W]\overline{L}$ and $\overline{P} = \overline{L}'[\overline{\overline{Q}}][W]\overline{L}$. With an analysis analogous to the one given in Bates and Watts

(1980) and Efron (1975), it can easily be proved that $\overline{I}$ and $\overline{P}$ are important terms which reflect the nonlinear feature of the curved exponential family (3.1). It is important to note that $\overline{I}$ is an intrinsic part of this model and does not depend on the parameter that we choose, while $\overline{P}$ does. Actually, in nonlinear regression models, $\Lambda$ is the unit matrix, and $\overline{I}(\overline{P})$ is the same as $I(P)$ defined in Bates and Watts (1980). Thus, a nonlinear regression model with Gaussian innovations is a special case of the separable exponential geometry we discussed above. In the next section, we will apply this framework to ARMA models with Gaussian innovations.

## 4. Geometric Structure of ARMA Models and Stochastic Expansions

In ARMA model (1.1), the circular model is defined by $x_k = x_{M-k}$, $k = 0, \ldots, p-1$, and $e_l = e_{M-l}$, $l = 0, \ldots, q-1$, where $X = (x_1, \ldots, x_M)'$ is a set of $M$ observations. It is assumed that with respect to (1.1), $U(B) = 0$ and $V(B) = 0$ have no common roots and all roots lie outside the unit circle. The spectral density function is given by $f(\lambda) = \sigma^2 |V(e^{i\lambda})|^2/(2\pi|U(e^{i\lambda})|^2)$ and its periodogram is

$$I(\lambda_t) = \frac{1}{2\pi M} \sum_{s,r=1}^{M} e^{i\lambda_t(s-r)} x_s x_r,$$

where $\lambda_t = 2\pi t/M$, $t = 0, \ldots, M-1$. Based on the above assumptions, the log-likelihood function can be written as

$$l(X, \beta) = -M \log 2\pi - \frac{1}{2} \sum_{t=0}^{M-1} \left\{ \log f(\lambda_t, \beta) + \frac{I(\lambda_t)}{f(\lambda_t, \beta)} \right\},$$

where $\beta = (u_1, \ldots, u_p; \nu_1, \ldots, \nu_q; \sigma^2)'$ is a vector of the unknown parameters. With fixed $M$, the manifold induced by this distribution is $(n, m)$-curved exponential with $n = M$, $m = p + q + 1$, where $\theta = \theta(\beta) = (\theta_1(\beta), \ldots, \theta_n(\beta))'$ with $\theta_k(\beta) = (2f(\lambda_k, \beta))^{-1}$, and $T = (T_1, \ldots, T_n)'$ with $T_k = -I(\lambda_k)$. Since the characteristic function of this manifold is $\Psi(\theta) = \sum \varphi(\theta_k)$ where $\varphi(\theta) = -\log \theta + const$, the diagonal elements of $\Lambda$ are $\theta_i^{-1}$, i.e., the manifold is separable. It is not hard to prove that the $T_k$ are independent variables and $2I(\lambda_k)/f(\lambda_k)$ follows a $\chi^2(2)$ distribution (see Anderson (1977) for a detailed discussion).

Using the general notation of separable exponential geometry in the last section, the Maximum Likelihood Estimate (MLE) of $\beta$ can be obtained by solving the system of equations

$$\frac{\partial l(X, \beta)}{\partial \beta} \Big|_{\hat{\beta}} = V'(\hat{\beta})(T - \nabla\Psi(\theta(\hat{\beta}))) = 0,$$

where $\nabla$ denotes the gradient operator. This is equivalent to Whittle's estimator, or *Quasi-Maximum Likelihood Estimator* (*q*-MLE) developed by some authors

(Taniguchi (1991)). To derive the stochastic expansion of this estimator, we make the following assumptions. First we assume $(C_1) : \lim_{n\to\infty} n^{-1}\overline{V}'(\beta)\overline{V}(\beta) = \overline{\Omega}(\beta)$, where $\overline{\Omega}(\beta)$ is positive definite. This can be rewritten as $\overline{V}'(\beta)\overline{V}(\beta) = O(n)$, or $\overline{V}(\beta) = O(n^{1/2})$. Second, the Fisher information matrix related to the coordinates is assumed to be consistent with $g_{ab}$, where $g_{ab}$ is defined by Amari (1985), i.e., $(C_2) : \overline{\lim}_{n\to\infty} n^{-1}\sum_{i=1}^n (W_{iab})^2 = E_{ab}(\beta)$.

From the $QR$ decomposition of $\overline{V}(\beta)$, a vector $\hat{\gamma}$ exists such that

$$T - \nabla\Psi(\theta(\beta)) = \Lambda(\beta)\overline{H}(\beta)\hat{\gamma} = \overline{\overline{H}}(\beta)\hat{\gamma}.$$

By expanding $\overline{\overline{H}}(\beta)$ at $\beta_0$ to $O_p(n^{-1})$, we get $\overline{\overline{H}}(\hat{\beta}) = \overline{\overline{H}}(\beta_0) + (\overline{\overline{F}}(\beta_0) \cdot \Delta\beta)' + O_p(n^{-1})$ where $\overline{\overline{F}}(\beta) = (\overline{\overline{F}}_{\lambda ia}) = (\partial\overline{\overline{H}}_{i\lambda}/\partial\beta_a)$. Lemma 4.1, without proof, is for further investigation of the Edgeworth expansion of relevant statistics. The main result of the Edgeworth expansion is given in Theorem 4.1.

**Lemma 4.1.** $V'\overline{\overline{F}} + \lfloor \overline{\overline{H}}' \rfloor [W] = 0.$

**Theorem 4.1.** *If conditions $C_1$, $C_2$ are satisfied, the first order expansions of $\Delta\beta$ and $\hat{\gamma}$ are*

$$\Delta\beta = \overline{L}\mu + O_p(n^{-1}), \qquad \hat{\gamma} = \nu + O_p(n^{-1/2}),$$

*where $\mu = \overline{Q}'\varepsilon$, $\nu = \overline{H}'\varepsilon$. The second order expansion of $\Delta\beta$ is*

$$\Delta\beta = \overline{L}\mu + \overline{L}\{[\nu'][\overline{I}]u - \frac{1}{2}\mu'\overline{P}\mu + \mu'\overline{S}\mu\} + O_p(n^{-3/2}),$$

*where $\overline{P} = \overline{L}'[\overline{\overline{Q}}'][W]\overline{L}$, $\overline{S} = [\overline{Q}'][\overline{\overline{\Sigma}}^{(2)}]$, $\overline{\overline{\Sigma}}_i^{(2)} = \overline{Q}_i'\overline{Q}_i$, $\overline{\overline{\Sigma}}_i^{(2)}$ is the ith face of $\overline{\overline{\Sigma}}^{(2)}$. Furthermore, each component of $\hat{\gamma}$ is asymptotically normal $N(0,1)$, i.e., $\hat{\gamma}$ is an asymptotically ancillary statistic for recovering information from $\hat{\beta}$.*

**Proof.** Following the decomposition (3.2), we have

$$
\begin{aligned}
\Lambda(\beta_0)\varepsilon &= T - \nabla\Psi(\theta(\beta_0)) \\
&= [T - \nabla\Psi(\theta(\hat{\beta}))] + [\nabla\Psi(\theta(\hat{\beta})) - \nabla\Psi(\theta(\beta_0))] \\
&= \overline{\overline{H}}(\hat{\beta})\hat{\gamma} + [\nabla\Psi(\theta(\hat{\beta})) - \nabla\Psi(\theta(\beta_0))] \\
&= \overline{\overline{H}}\hat{\gamma} + [\hat{\gamma}'][\overline{\overline{F}}\Delta\beta] + \Lambda^2 V\Delta\beta + O_p(n^{-1}) \\
&= \Lambda\overline{H}\hat{\gamma} + \Lambda\overline{V}\Delta\beta + [\hat{\gamma}'][\overline{\overline{F}}\Delta\beta] + O_p(n^{-1}).
\end{aligned}
$$

By multiplying both sides of the above equation by $\overline{H}'\Lambda^{-1}$ and noticing that $\overline{H}'\overline{V} = 0$, $\Delta\beta = O_p(n^{-1/2})$, it is easy to see that $\hat{\gamma} = \nu + O_p(n^{-1/2})$. The higher order properties can be studied via evaluating the above equation up to the second order, that is

$$\Lambda\varepsilon = \overline{\overline{H}}\hat{\gamma} + [\hat{\gamma}']\left[\overline{\overline{F}}\Delta\beta\right] + \Lambda^2 V\Delta\beta + \frac{1}{2}(\Delta\beta)'\Sigma(\Delta\beta) + O_p(n^{-1}), \qquad (4.1)$$

where

$$\Sigma_{iab} = \frac{\partial^2}{\partial\beta_a\partial\beta_b}\left(\frac{\partial\Psi}{\partial\theta_i}\right) = \Lambda_i^2\frac{\partial^2\theta_1}{\partial\beta_a\partial\beta_b} + \ddot{\varphi}(\theta_i)\frac{\partial\theta_i}{\partial\beta_a}\frac{\partial\theta_i}{\partial\beta_b}.$$

It turns out that $\Sigma$ is composed of two terms: $\Sigma^{(1)}$ and $\Sigma^{(2)}$, where $\Sigma^{(1)} = [\Lambda^2][W]$ and $\Sigma^{(2)} = -2[\Lambda][\overline{V}_t'\overline{V}_t]$. By multiplying both sides of (4.1) with $\overline{V}'\Lambda^{-1}$, we have $\overline{V}'\varepsilon = (\overline{V}'\overline{V})\Delta\beta + [\hat{\gamma}'][\overline{V}'\overline{\overline{F}}]\Delta\beta + (\Delta\beta)'[\overline{V}'\Lambda^{-1}][\Sigma](\Delta\beta)/2 + O_p(n^{-1/2})$. It follows that the second order expansion of $\Delta\beta$ can be obtained by multiplying both sides with $(\overline{V}'\overline{V})^{-1}$, i.e.,

$$\Delta\beta = (\overline{V}'\overline{V})^{-1}\overline{V}'\varepsilon - (\overline{V}'\overline{V})^{-1}\{\frac{1}{2}(\Delta\beta)'[V'][\Sigma](\Delta\beta) + [\hat{\gamma}'][\overline{V}'\overline{\overline{F}}]\Delta\beta + O_p(n^{-1/2})\}$$

$$= \overline{L}\mu - \overline{L}\,\overline{L}'\{\frac{1}{2}\mu'[V'][\overline{L}'\Sigma\overline{L}]\mu - [\nu'][\overline{R}'\overline{I}\,\overline{R}]\overline{L}\mu\} + O_p(n^{-3/2})$$

$$= \overline{L}\mu + \overline{L}\,\overline{L}'\{-\frac{1}{2}\mu'[V'][\overline{L}'\Sigma\overline{L}]\mu + \overline{R}'[\nu'][\overline{I}]\mu\} + O_p(n^{-3/2}).$$

Now it leaves the computation of the two components in $[V'][\overline{L}'\sum\overline{L}]$:

$$[V'][\overline{L}'\Sigma^{(1)}\overline{L}] = \overline{L}'[V'][[\Lambda^2][W]]\overline{L} = \overline{L}'[\overline{V}'\Lambda][W]\overline{L} = \overline{L}'[\overline{R}'\overline{Q}'\Lambda][W]\overline{L}$$

$$= \overline{L}'[\overline{R}'][\overline{\overline{Q}}'][W]\overline{L} = [\overline{R}'][\overline{P}]$$

and

$$[V'][\overline{L}'\Sigma^{(2)}\overline{L}] = -2\overline{L}'[V'][[\Lambda][\overline{\Sigma}^{(2)}]]\overline{L} = -2[\overline{V}'][\overline{L}'\overline{\Sigma}^{(2)}\overline{L}]$$

$$= -2[\overline{R}'][[\overline{Q}'][\overline{\Sigma}^{(2)}]] = -2[\overline{R}'][\overline{S}].$$

Therefore, we have $\Delta\beta = \overline{L}\mu + \overline{L}\{[\nu'][\overline{I}]\mu - \frac{1}{2}\overline{L}'\mu[\overline{R}'][\overline{P}]\mu + \overline{L}'\mu'[\overline{R}'][\overline{S}]\mu\} + O_p(n^{-3/2}) = \overline{L}\mu + \overline{L}\{[\nu'][\overline{I}]\mu - \frac{1}{2}\overline{L}'[\overline{R}'][\mu'\overline{P}\mu] + \overline{L}'[\overline{R}'][\mu'\overline{S}\mu]\} + O_p(n^{-3/2})$. The proof is complete.

## 5. Information Loss in Estimation

Information hidden in data often plays an essential role in a decision making process. Correct decisions such as forecasting and control are highly dependent on valuable hidden information. In statistical inference, it is well known that the Fisher information matrix reflects all the information the relevant data conveys. It can be represented as $J_\beta(T) = \text{Cov}\,(\partial_i l, \partial_j l)$, where $\partial_i l = \partial l(T,\beta)/\partial\beta_i$. When estimating the unknown parameters, the information behind the data transfers to the parameters' estimates, and the dimension of information stored is reduced. However, part of the information may be lost in the estimation process, which results in the second-order inefficiency of the estimate. Amari (1985) proposed

that this information loss can be measured by $\Delta J(\hat{\beta}) = J_\beta(T) - J_\beta(\hat{\beta})$, and computed as

$$\Delta J(\hat{\beta}) = E_\beta \left\{ \text{Var}_\beta \left[ \frac{\partial l(T, \beta)}{\partial \beta} \Big| \hat{\beta} \right] \right\}.$$

Using the main results of Theorems 4.1, the information loss can be evaluated with relation to the intrinsic curvature of the curved manifold.

**Theorem 5.1.** $\Delta J(\hat{\beta}) = \sum_{t=1}^{n-m} \overline{R}'(\overline{I}_t)^2 \overline{R} + O(n^{-1})$ where $\overline{I}_i$ is the ith face of $\overline{I}$.

**Proof.** From the $QR$ decomposition of $\overline{V}(\beta)$ and Lemma 4.1, we have

$$\frac{\partial l}{\partial \beta} = V'(\beta)(T - \nabla\Psi(\theta(\beta))) = V'(\beta)\left\{ \left[ T - \nabla\Psi(\theta(\hat{\beta})) \right] + \left[ \nabla\Psi(\theta(\hat{\beta})) - \nabla\Psi(\theta(\beta)) \right] \right\}$$

$$= \overline{V}'\overline{H}\hat{\gamma} - [\hat{\gamma}'] \left[ [\overline{\overline{H}}'][W] \right] \Delta\beta + O_p(n^{-1}) + \Delta(\beta, \hat{\beta})$$

$$= -[\nu'] \left[ [\overline{\overline{H}}'] \right] [W] \overline{L}\mu + O_p(n^{-1/2}) + \Delta(\beta, \hat{\beta}).$$

It follows that the variance of $\partial l/\partial\beta$ conditioned on $\hat{\beta}$ can be calculated as $\text{Var}_\beta(\partial l/\partial\beta|\hat{\beta}) = \sum(\overline{R}'\overline{I}_t\mu)^2 + O_p(n^{-1})$. From Amari's (1985) definition, the information loss of $\hat{\beta}$ is

$$\Delta J(\hat{\beta}) = E_\beta \left\{ \text{Var}_\beta \left[ \frac{\partial l}{\partial \beta} \Big| \hat{\beta} \right] \right\} = \sum_{t=1}^{n-m} E(\overline{R}'\overline{I}_t\mu)^2 + O_p(n^{-1}) = \sum_{t=1}^{n-m} \overline{R}'(\overline{I}_t)^2 \overline{R} + O_p(n^{-1}).$$

The proof is complete.

As in the case of independent observations, the above theorem indicates that the information loss of MLE is only related to the intrinsic curvature and will not be influenced by parameterization. However, some ancillary statistics may help to recover the lost information. In the time domain, Taniguchi (1991) provides a detailed discussion on the information loss and recovery of MLE. Using Edgeworth expansions, he constructs an asymptotically ancillary statistic, which can recover the information loss in the reduction process from observations $X$ to MLE. In the frequency domain, the Edgeworth expansions of the MLE has been derived in Theorem 4.1. Here, $\hat{\gamma}$ is expected to be used as an ancillary statistic to help recover the information loss. This will be discussed in future work. Furthermore, the asymptotic bias of Whittle's estimator is given in the following theorem.

**Theorem 5.2.** The bias of $\hat{\beta}$ is given by $-\frac{1}{2}\overline{L} \cdot tr[\overline{P} - 2\overline{S}] + O(n^{-3/2})$.

**Proof.** With the decomposition (3.2), it is not hard to prove that $E(\varepsilon_i) = 0$, $E(\varepsilon_i\varepsilon_j) = \delta_{ij}$ $(i, j = 1, \ldots, n)$. Thus, $E(\mu) = 0$, $E(\mu_a\mu_b) = \delta_{ab}$, and $E(\mu'P\mu) =$

$\operatorname{tr}[P]$, where $P$ is any array of order $m \times m \times m$ whose faces are all symmetric matrices. From Theorem 4.1, we can obtain the estimation bias

$$Bias(\hat{\beta}) = E(\Delta\beta) = \overline{L}\{E[\nu'][\overline{I}]\mu - \frac{1}{2}E(\mu'\overline{P}\mu) + E(\mu'\overline{S}\mu)\} + O(n^{-3/2})$$
$$= -\frac{1}{2}\overline{L} \cdot tr[\overline{P} - 2\overline{S}] + O(n^{-3/2}).$$

This completes the proof.

The asymptotic bias of MLE in the time domain was addressed in Ravishanker, Melnick and Tsai (1990). However, the derivations are more complicated in the analysis of frequency domain. Here, it is verified again that the estimation bias is only related to the parameter-effects curvature. To have more understanding of the asymptotic performance of Whittle's estimator, the AR(1) model is used to illustrate the difference between MLEs in the time and frequency domains. In the time domain, as Ravishanker, Melnick and Tsai (1990) and Taniguchi (1991) showed, the asymptotic bias is $(-2u/n) + o(n^{-1})$. In the frequency domain, the above asymptotic bias with known variance $\sigma^2$ can be calculated from Theorem 5.2 as $-3u/n + O(n^{-3/2})$, which is the same as Taniguchi's derivation (1991, Chapter 7). Note that the bias is $O(n^{-1})$, the same order as that in the time domain, but a bit larger. We do not intend to interpret the difference here, however, it is worth noting that both biases vanish as $u$ gets close to zero. This coincides with our expectation that the model approaches linearity when $u$ diminishes. Table 1 shows simulation results for $n = 100$, where the comparisons between MLEs in both time and frequency domains are summarized based on 1000 replications.

Table 1. Estimating the parameters of AR(1) processes via Whittle's estimator and MLE in the time domain.

| $\phi$ | Whittle's Estimator | | MLE in Time Domain | |
|---|---|---|---|---|
| | Mean | St. Dev. | Mean | St. Dev. |
| -0.9 | -0.877 | 0.096 | -0.875 | 0.098 |
| -0.6 | -0.577 | 0.063 | -0.582 | 0.065 |
| -0.3 | -0.290 | 0.047 | -0.297 | 0.050 |
| 0.0 | 0.002 | 0.001 | 0.001 | 0.001 |
| 0.3 | 0.285 | 0.051 | 0.289 | 0.053 |
| 0.6 | 0.590 | 0.040 | 0.593 | 0.045 |
| 0.9 | 0.859 | 0.099 | 0.871 | 0.099 |

## References

Amari, S. (1982). Differential geometry of curved exponential families–curvature and information loss. *Ann. Statist.* **10**, 357-387.

Amari, S. (1985). *Differential-geometrical methods in statistics.* Lecture Notes in Statistics, Vol. 28. Springer-Verlag, New York.

Anderson, T. W. (1977). Estimation of autoregressive moving average models in the time and frequency domains. *Ann. Statist.* **5**, 842-865.

Anderson, T. W. (1994). *The Statistical Analysis of Time Series.* (Wiley Classics Library Edition). Wiley, New York.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of a nonlinearity (with discussion). *J. Roy. Statist. Soc. Ser. B* **40**, 1-25.

Barndorff-Nielsen, O. E. (1988). *Parametric statistical models and likelihood.* Lecture Notes in Statistics, Vol. **50**. Springer-Verlag, New York.

Barndorff-Nielsen O. E. and Cox, D. R. (1994). *Inference and asymptotics.* Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* 2nd edition. Springer, Berlin.

Efron, B. (1975). Defining the curvature of a statistical problem (with application to second order efficiency) (with Discussion). *Ann. Statist.* **3**, 1189-1242.

Mikosch, T., Gadrich, T., Kluppelberg, C. and Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.* **23**, 305-326.

Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81-91.

Ravishanker, N., Melnick, E. and Tsai, C. L. (1990). Differential geometry of ARMA models. *J. Time Series Anal.* **11**, 259-274.

Ravishanker, N. (1994). Relative curvature measures of nonlinearity for time series models. *Comm. Statist. Simulation Comput.* **2**, 415-430.

Taniguchi, M. (1991). *Higher order asymptotic theory for time series analysis.* Lecture Notes in Statistics, Vol. 68. Springer-Verlag, New York.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2**, 423-434.

Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Hong Kong.

E-mail: jiangwei@ust.hk

Management School, Xi'an Jiaotong University, Xi'an 710049, China.

E-mail: ymxi@xjtu.edu.cn

Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China.

E-mail: sywang@rose.nsfc.gov.cn