BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR–91/04

CONVERGENCE OF FINITE MULTISTEP PREDICTORS FROM INCORRECT MODELS AND ITS ROLE IN MODEL SELECTION

by

David F. Findley
Institute of Statistical Science, Academia Sinica
Taipei 11529, Taiwan

and

U.S. Bureau of the Census
Statistical Research Division
Washington, DC 20233

Version:12 March 1991

Convergence of Finite Multistep Predictors from Incorrect Models and Its Role in Model Selection

David F. Findley

Institute of Statistical Science, Academia Sinica

and U.S. Bureau of the Census

Dedicated to Professor Gottfried Köthe *in memoriam*

Keywords: Time Series Models, Baxter's Inequality

1. Introduction

In a recent paper [5], we described a generalization to univariate time series models of the hypothesis testing procedure of Vuong [12] for comparing incorrect statistical models for independent data. The focus of [5] was on model selection criteria related to one–step–ahead forecasting performance. We suggested there that when p–step–ahead prediction is the goal, with $p > 1$, then different test statistics should be used for each choice of p. To identify appropriate test statistics and determine their asymptotic distribution, information is needed about the rate at which predictors based on n observations converge as $n \rightarrow \infty$. This note provides some of the needed convergence and distributional results, also for the case of r–dimensional vector time series. Our approach rests on generalizations of the finite–section inequality and related convergence results of Baxter [1],[2] for one–step–ahead predictors of scalar time series. To make our results accessible to a larger circle of readers, we will not formulate them in the Banach algebra framework utilized by Hirschman [8], but it will be clear to the mathematical reader that this level of generality is attainable and natural.

2. Baxter's Inequality (Matrix Form)

For any complex–valued matrix C, let C^T denote its transpose and $C^* \equiv \overline{C}^T$ its complex conjugate, or Hermitian, transpose. If $C^* = C$, then C is said to be Hermitian symmetric. Let $f(\theta)$ denote a continuous, positive definite Hermitian symmetric, r $\times$ r matrix function on $[-\pi, \pi]$ satisfying $f(-\theta) = f(\theta)^T$. It is well known, see [6,p.160] or [7], that such an $f(\theta)$ has factorizations of the form

$$f(\theta) = A(e^{i\theta})A^*(e^{i\theta}) \tag{2.1}$$

and

$$f(\theta) = B^*(e^{i\theta})B(e^{i\theta}), \tag{2.2}$$

where A(z) and B(z) are non–singular–matrix–valued analytic functions on $\{|z| < 1\}$,

$$A(z) = \sum_{j=0}^{\infty} a_j z^j$$

$$B(z) = \sum_{j=0}^{\infty} b_j z^j$$

whose coefficient matrices, a_j, b_j, have real entrices. We shall impose magnitude restrictions on these coefficients with the aid of an *increasing* sequence of weights $\nu(j) \geq 1$, j=0,1,$\cdots$, such that $\nu(j) \leq \nu(k)\nu(|j-k|)$ for any j,k ≥ 0. This last conditon insures that the norm defined for matrix functions $C(\theta) = \sum_{j=-\infty}^{\infty} c_j e^{ij\theta}$ by means of

$$||C(\theta)|| = \sum_{j=-\infty}^{\infty} \nu(|j|)|c_j|,$$

with $|c_j|$ equal to the square root of the largest eigenvalue of $c_j^* c_j$, see [10,pp.265–6], has the property that $||C(\theta)D(\theta)|| \leq ||C(\theta)||\,||D(\theta)||$. Let $\mathcal{C}_\nu$ denote the set of all continuous r × r–matrix–valued functions $C(\theta)$ for which $||C(\theta)|| < \infty$, and let $\mathcal{C}_\nu^+$ (respectively $\mathcal{C}_\nu^-$) denote the subset whose j–th Fourier coefficient c_j is 0 for all j < 0 (respectively, j > 0). It follows from the preceding norm inequality that $f(\theta) \in \mathcal{C}_\nu$ if $A(e^{i\theta})$ in (2.1) belongs to $\mathcal{C}_\nu^+$ (which implies that $A^*(e^{i\theta}) \in \mathcal{C}_\nu^-$). Since A(z) is nonsingular for all $|z| \leq 1$, it follows from an argument like that given in [3,p.78] that $A^{-1}(e^{i\theta})$ belongs to $\mathcal{C}_\nu^+$ if $A(e^{i\theta})$ does, and then $A^*(e^{i\theta})^{-1} \in \mathcal{C}_\nu^-$. For us, the important choices of $\nu(j)$ are $\nu(j) \equiv 1$, $\nu(j) \equiv 2^\alpha + j^\alpha \quad (\alpha \geq 0)$ and $\nu(j) \equiv \rho^{-j} \quad (0 < \rho < 1)$.

We present now our matrix–function version of the inequality of [2]. Our proof is an adaptation of Baxter's, see also [8]. For the reader's convenience, the complete proof will be given.

Proposition 2.1. Assume that the factors $A(e^{i\theta})$ and $B(e^{i\theta})$ of $f(\theta)$ in (2.1) and (2.2) belong to $\mathcal{C}_\nu^+$. Then there exist a positive integer n_0 and a constant M, depending only on these functions, with the following property: if $n \geq n_0$, then for any given $r \times r$ matrices g_j, $0 \leq j \leq n-1$, the matrix polynominal $h(\theta) = \sum_{k=0}^{n-1} h_k e^{ik\theta}$ which satisfies

$$\int_{-\pi}^{\pi} e^{-ij\theta} h(\theta) f(\theta) d\theta = g_j \tag{2.3}$$

for $j = 0, \cdots, n-1$ will also satisfy the ν–norm inequality

$$||h(\theta)|| \leq M||g(\theta)||, \tag{2.4}$$

where $g(\theta) = \sum_{j=0}^{n-1} g_j e^{ij\theta}$.

Proof: Some additional notation will be helpful. If $C \in \mathcal{C}_\nu$ is such that $C(\theta)$ is nonsingular for all θ, we will sometimes use $\hat{C}$ to denote the function $C(\theta)^{-1}$. Also, for any positive integer m, we define two useful additive components of C:

$$C_{(m)} \equiv \sum_{j=m}^{\infty} c_j e^{ij\theta}, \quad C_{(-m)} \equiv \sum_{j=-\infty}^{-m} c_j e^{ij\theta}.$$

Observe that $||C_{(\pm m)}|| \leq ||C||$.

Using this notation, set $G = (hf)_{(-1)}$ and $H = (hf)_{(n)}$. Then from (2.3),

$$hf = G + g + H. \tag{2.5}$$

From (2.5), (2.1) and (2.2) we obtain

$$hA = G\hat{A}^* + g\hat{A}^* + H\hat{A}^* \tag{2.6}$$

and

$$hB^* = G\hat{B} + g\hat{B} + H\hat{B}. \tag{2.7}$$

The essence of the proof of (2.4) is the verification of

$$||G\hat{A}^*|| \leq Const.||g|| \tag{2.8}$$

and

$$||H\hat{B}|| \leq Const.||g||, \tag{2.9}$$

with constants independent of g, because, from (2.5),

$$\begin{aligned} ||h|| &\leq ||Gf^{-1}|| + ||gf^{-1}|| + ||Hf^{-1}|| \\ &\leq ||G\hat{A}^*||\;||\hat{A}|| + ||g||\;||f^{-1}|| + ||H\hat{B}||\;||\hat{B}^*||. \end{aligned}$$

We start with (2.8): Since $hA \in C_\nu^+$, it follows from (2.5) and $(G\hat{A}^*)_{(-1)} = G\hat{A}^*$ that

$$\begin{aligned} G\hat{A}^* &= -(g\hat{A}^*)_{(-1)} - (H\hat{A}^*)_{(-1)} \\ &= -(g\hat{A}^*)_{(-1)} - (H\hat{A}^*_{(-n-1)})_{(-1)}. \end{aligned}$$

Hence

$$\begin{aligned}||G\hat{A}^*|| &\leq ||g\hat{A}^*|| + ||H\hat{A}^*_{(-n-1)}|| \\ &\leq ||g||\ ||\hat{A}^*|| + ||H\hat{B}||\ ||B||\ ||\hat{A}^*_{(-n-1)}||.\end{aligned}$$

So, for n sufficiently large, we will have

$$||G\hat{A}^*|| \leq ||\hat{A}^*||\ ||g|| + \frac{1}{2}||H\hat{B}||. \tag{2.10}$$

With a similar calculation based on the fact that the j–th Fourier coefficient of hB^* (respectively, $H\hat{B}$) is 0 if j $\geq$ n (respectively, j $<$ n), one sees that when n is large enough that $||A^*||||\hat{B}_{(n+1)}|| \leq \frac{1}{2}$, then

$$||H\hat{B}|| \leq ||\hat{B}||\ ||g|| + \frac{1}{2}||G\hat{A}^*||. \tag{2.11}$$

From the inequality obtained by adding (2.10) and (2.11), one obtains (2.8) and (2.9) with the constant equal to $2(||\hat{A}^*|| + ||\hat{B}||)$. This completes the derivation of (2.4).

Remark 2.1. In the univariate case (r=1), the assumption that $A(e^{i\theta})$ belongs to C^+_ν is equivalent to the assumption that $f(\theta) \in C_\nu$, see the proof of Theorem 3.8.4 of [3]. No multivariate generalization of this result appears to be known. A partial result for the special case in which $f(\theta) = F(e^{i\theta})$, with F(z) analytic in $\{\rho < |z| < \rho^{-1}\}$ for some $0 < \rho < 1$, can be obtained from Theorems 3.1 and 3.2 of [11], which imply that for such an $f(\theta)$, the functions A(z) and B(z) are analytic in $\{|z| < \rho^{-1/2}\}$.

Remark 2.2. In Hannan and Deistler's monograph [7, p.270], a vector generalization of Baxter's result is stated with only the assumption that $A(e^{i\theta}) \in C^+_\nu$ and with no mention of the factorization (2.2). E.J. Hannan (personal communication) agrees that the condition $B(e^{i\theta}) \in C^+_\nu$ is also needed. It is possible that this property follows from the assumed property of $A(e^{i\lambda})$, but this seems difficult to estiablish.

Remark 2.3. The derivation of (2.4) does not require our assumption that $f(\theta) = f(\theta)^T$. This was used for convenience of reference in later sections where we wish to maintain the familiar context wherein the Fourier coefficients of $f(\theta)$ are real matrices.

3. Convergence of Predictor Coefficients

Suppose that z_t is a mean zero, weakly stationary, r–dimensional vector time series with spectral density matrix $f(\theta)$,

$$E z_t z_{t-j}^T = 2\pi f_j = \int_{-\pi}^{\pi} e^{-ij\theta} f(\theta) d\theta \quad (j = 0, \pm 1, \cdots),$$

with $f(\theta)$ satisfying the assumptions of section 2. (E denotes expectation). For any integers p, n $\geq$ 1 or n=∞, the optimal linear predictor of z_{t+p} from $z_t, \cdots, z_{t-(n-1)}$ is given by

$$z_{t+p|t}^{(n)} \equiv \sum_{k=0}^{n-1} \pi_k^{(n)}[p] y_{t-k},$$

where the coefficient matrices are determined by the property that the error process $e_{t+p|t}^{(n)} \equiv z_{t+p} - z_{t+p|t}^{(n)}$ is uncorrelated with $z_t, \cdots, z_{t-(n-1)}$,

$$\int_{-\pi}^{\pi} e^{-ij\theta}(e^{-ip\theta} - \sum_{k=0}^{n-1} \pi_k^{(n)}[p] e^{ik\theta}) f(\theta) d\theta = 0, \quad (j = 0, \cdots, n-1). \tag{3.1}$$

It follows that the difference between $z_{t+p|t}^{(n)}$ and the p–step–ahead predictor based on the infinite past, $z_{t+p|t}^{\infty} \equiv \sum_{k=0}^{\infty} \pi_k^{(\infty)}[p] z_{t-k}$, that is,

$$z_{t+p|t}^{(n)} - z_{t+p|t}^{(\infty)} = e_{t+p|t}^{(\infty)} - e_{t+p|t}^{(n)},$$

is uncorrelated with $z_t, \cdots, z_{t-(n-1)}$. Therefore,

$$\int_{-\pi}^{\pi} e^{-ij\theta}\Big(\sum_{k=0}^{n-1}\{\pi_k^{(n)}[p] - \pi_k^{(\infty)}[p]\}e^{ik\theta}\Big)f(\theta)d\theta = g_j \tag{3.2}$$

with

$$\begin{aligned} g_j &= \int_{-\pi}^{\pi} e^{-ij\theta}\Big(\sum_{k=n}^{\infty}\pi_k^{(\infty)}[p]e^{ik\theta}\Big)f(\theta)d\theta \\ &= 2\pi\sum_{k=n}^{\infty}\pi_k^{(\infty)}[p]f_{j-k} \end{aligned} \tag{3.3}$$

for j=0, $\cdots$,n-1. We are assuming that $f(\theta)$ and its factors $A(e^{i\theta})$, $B(e^{i\theta})$ belong to C_ν for some weighting sequence $\nu(j)$ of the sort considered. Since $\nu(j) \le \nu(k)\nu(k-j)$ when $0 \le j \le n-1$ and $k \ge n$, it is a consequence of (3.3) that

$$\sum_{j=0}^{n-1}\nu(j)|g_j| \le \Big(2\pi\sum_{m=0}^{\infty}\nu(m)|f_{-m}|\Big)\sum_{k=n}^{\infty}\nu(k)|\pi_k^{(\infty)}[p]|. \tag{3.4}$$

The first factor on the right is finite because $f \in C_\nu$. Theorem 7.3 of [13] shows that the prediction error transfer function $e^{(\infty)}[p](\theta) \equiv e^{-ip\theta} - \sum_{k=0}^{\infty}\pi_k^{(\infty)}[p]e^{ik\theta}$ has the formula

$$e^{(\infty)}[p](\theta) = \Big(\sum_{j=0}^{p-1}\psi_j e^{i(j-p)}\Big)\psi^{-1}(\theta), \tag{3.5}$$

where $\psi(\theta) = A(e^{i\theta})A(0)^{-1}$, from which it follows that $e^{(\infty)}[p](\theta) \in C_\nu$. Thus the second factor on the right in (3.4) is also finite and, applying (2.4), we arrive at the following generalization of the filter coefficient inequality (11) of [1],

$$\sum_{k=0}^{n-1}\nu(k)|\pi_k^{(n)}[p] - \pi_k^{(\infty)}[p]| \le M_0\sum_{k=n}^{\infty}\nu(k)|\pi_k^{(\infty)}[p]| < \infty. \tag{3.6}$$

Remark. The inequality given in [1] for the case r=1 is for weighted versions of the coefficients $\pi_k^{(n)}[1]$ and $\pi_k^{(\infty)}[1]$, and is not as convenient for our application as (3.6).

Next, we observe that since $1 \le \nu(0) \le \nu(1) \le \cdots$, we have

$$\sum_{k=n}^{\infty}|\pi_k^{(\infty)}[p]| \le \nu(n)^{-1}\sum_{k=n}^{\infty}\nu(k)|\pi_k^{(\infty)}[p]| = o(1/\nu(n)). \tag{3.7}$$

The result we are after follows from (3.7) and the version of (3.6) associated with $\nu(k) \equiv 1,\ 0 \leq k \leq \infty$.

Proposition 3.1. The p–step ahead predictor coefficient matrices associated via (3.1) with a spectral density matrix $f(\theta)$ which satisfies the conditions of section 2 will have the property that

$$\sum_{k=0}^{n-1} |\pi_k^{(n)}[p] - \pi_k^{(\infty)}[p]| + \sum_{k=n}^{\infty} |\pi_k^{(\infty)}[p]| = o(1/\nu(n)), \tag{3.8}$$

for any $p \geq 1$.

4. Convergence of Finite Predictors from Incorrect Models

Let $y_1, \cdots, y_n$ denote the observed values of a mean zero, r–dimensional time series y_t. Suppose that a forecast of y_{n+p} is desired and that a time series model specifying a spectral density matrix $f(\theta)$ has been fit to the observations for the purpose of determining a predictor,

$$y_{n+p|n}^{(n)}[p] \equiv \sum_{k=0}^{n-1} \pi_k^{(n)}[p] y_{n-k},$$

whose coefficients satisfy (3.1). If the model is incorrect, as is ordinarily the case, then the prediction error $y_{n+p} - y_{n+p|n}^{(n)}[p]$ will not be uncorrelated with $y_n, \cdots, y_1$ and the forecast error process associated with prediction from the infinite past,

$$e_{t+p|t}^{(\infty)} \equiv y_t - \sum_{k=0}^{\infty} \pi_k^{(\infty)}[p] y_{t-k},$$

will not be a process whose autocorrelations at lags greater than $p-1$ are zero. The inequality (3.8) makes it possible to determine a rate of convergence for the finite predictor in this situation. For a measure of discrepancy, we will use the mean square norm, which is defined for a random vector x by $||x||_E = (Ex^Tx)^{1/2}$. This has the property that if b is a constant matrix, then $||bx||_E \leq |b| \; ||x||_E$ for the matrix norm $|\cdot|$ specified in section 2. The quantity $||e^{(n)}_{n+p|n}||_E$ is a natural measure of forecast standard error. From (3.6), we obtain

$$| \; ||e^{(\infty)}_{n+p|n}||_E - ||e^{(n)}_{n+p|n}||_E \; | \leq ||e^{(\infty)}_{n+p|n} - e^{(n)}_{n+p|n}||_E$$

$$= ||y^{(\infty)}_{n+p|n} - y^{(n)}_{n+p|n}||_E \leq \sum_{k=0}^{n-1} |\pi^{(\infty)}_k[p] - \pi^{(n)}_k[p]| \; ||y_{n-k}||_E + \sum_{k=n}^{\infty} |\pi^{(\infty)}_k[p]| \; ||y_{n-j}||_E.$$

The convergence result needed for the testing procedure described in the next section now follows from (3.8):

Proposition 4.1. Suppose $f(\lambda)$ satisfies the assumptions of section 2 and the time series y_t has bounded second moments, $\sup_{-\infty<t<\infty} ||y_t||_E < \infty$. *Then* $||e^{(\infty)}_{n+p|n}||_E - ||e^{(n)}_{n+p|n}||_E, ||e^{(\infty)}_{n+p|n} - e^{(n)}_{n+p|n}||_E$ *and* $||y^{(\infty)}_{n+p|n} - y^{(n)}_{n+p|n}||_E$ *are all of order $o(1/\nu(n))$.*

Remark 4.1. In the univariate case (r=1), if y_t is stationary and $f(\lambda)$ is the correct spectral density for y_t, then the proofs of Theorems 2.3 and 3.1 of [1] for the case p=1 can be adapted to show that the order of $||e^{(n)}_{n+p|n}||_E - ||e^{(\infty)}_{n+p|n}||_E$ is $o(\nu(n)^{-2})$, for each $p \geq 1$.

Remark 4.2. The approach of Devinatz [4] for obtaining results like those above for one–step–ahead predictors from *correct* models, and its multivariate generalization by Pourahmadi [11], do not seem to lend themselves to obtaining results for incorrect models except under restrictive assumptions, such as $|f_y(\theta)| \leq M_1|f(\theta)|$ for some constant M_1, where $f_y(\theta)$ designates the true spectral density of the series y_t, now assumed covariance stationary.

5. A Prototype Test Statistic for Comparing Models for Prediction

In this section, we assume that y_t is a mean zero, stationary vector process whose m–th order cumulants exist and are absolutely summable, for each m=2,3, $\cdots$ (Assumption 2.6.1 of [3]). Suppose that p–step–ahead forecasts are desired for some $p \geq 1$ and that two competing *incorrect* models for y_t are available, specifying spectral density matrices $f(\theta)$ and $\tilde{f}(\theta)$, both of which satisfy the assumptions of section 2 for the weighting sequence $\nu(j) \equiv 2^{1/2} + j^{1/2}$. Let $e^{(\infty)}_{t+p|t}$ and $\tilde{e}^{(\infty)}_{t+p|t}$ denote the error process of these models arising from predicting y_{t+p} linearly from y_{t-j}, j ≥ 0. These are stationary processes satisfying the same cumulant assumptions as y_t, and the same is true of the difference–of–squared–error process

$$\delta_{t+p} \equiv e^{(\infty)T}_{t+p|t} e^{(\infty)}_{t+p|t} - \tilde{e}^{(\infty)T}_{t+p|t} \tilde{e}^{(\infty)}_{t+p|t}. \tag{5.1}$$

We define $\sigma_p \equiv ||e^{(\infty)}_{t+p|t}||_E$ and $\tilde{\sigma}_p \equiv ||\tilde{e}^{(\infty)}_{t+p|t}||_E$. These quantities measure prediction performance: if $\sigma_p < \tilde{\sigma}_p$, the model specifying $f(\lambda)$ can be regarded as better for p–step prediction than the model specifying $\tilde{f}(\lambda)$. We would like to have a statistical test for deciding from observed prediction errors whether one of σ_p or $\tilde{\sigma}_p$ is smaller than the other.

Let $f_\delta(\theta)$ denote the spectral density function of the process $\delta_{t+p} - E(\delta_{t+p})$, observing that $E(\delta_{t+p}) = \sigma_p^2 - \tilde{\sigma}_p^2$. Theorem 4.4.1 of [3] shows that $N^{1/2}$ times the sample mean from N observations of this process has a limiting normal distribution with mean 0 and variance $2\pi f_\delta(\theta)$. This fact can be expressed as

$$N^{-1/2} \sum_{n=1}^{N} \delta_{n+p} - N^{1/2}(\delta_p^2 - \tilde{\sigma}_p^2) \rightarrow_{dist.} \mathcal{N}(0, 2\pi f_\delta(0)). \tag{5.2}$$

This result cannot be used directly to obtain a test of the hypothesis $\sigma_p = \tilde{\sigma}_p$, because the quantities δ_{n+p} cannot be calculated when only finitely many observations of y_t are available. This is where Proposition 4.1 plays a useful role. It enables us to show that $N^{-1/2}\sum_{n=1}^{N}\delta_{n+p}$ can be approximated with sufficient accuracy by $N^{-1/2}\sum_{n=1}^{N}\delta_{n+p|n}$, where

$$\delta_{n+p|n} \equiv e_{n+p|n}^{(n)T}e_{n+p|n}^{(n)} - \tilde{e}_{n+p|n}^{(n)T}\tilde{e}_{n+p|n}^{(n)}.$$

Using the Kalman filters, these quantities are easily calculated from the available data, when the models have ARMA representations.

Proposition 5.1. If the spectral density matrices $f(\theta)$ *and* $\tilde{f}(\theta)$ *satisfy the assumptions of section 2 with* $\nu(j) \equiv 2^{1/2} + j^{1/2}$, *and if* $\sup_{-\infty<t<\infty}||y_t||_E < \infty$, *then*

$$\lim_{N\to\infty} N^{-1/2}E|\sum_{n=1}^{N}\delta_{n+p} - \sum_{n=1}^{N}\delta_{n+p|n}| = 0. \tag{5.3}$$

Proof : The quantity whose limit is under investigation is bounded above by

$N^{-1/2}\sum_{n=1}^{N}E|\delta_{n+p} - \delta_{n+p|n}|$. This latter quantity is bounded above by $N^{-1/2}$ times the sum of

$$\Delta_n \equiv E|e_{n+p|n}^{(\infty)T}e_{n+p|n}^{(\infty)} - e_{n+p|n}^{(n)T}e_{n+p|n}^{(n)}|, \quad 1 \le n \le N$$

and the analogous quantities associated with $\tilde{f}(\theta)$, to which the argument given below also applies. Toeplitz's Lemma [10,p.250] shows that if

$$\Delta_n = o(n^{-1/2}) \tag{5.4}$$

holds, then $N^{-1/2}\sum_{n=1}^{N}\Delta_n \to 0$. Thus it remains to verify (5.4). By the Cauchy–Schwarz inequality,

$$\Delta_n \le ||e_{n+p|n}^{(\infty)} - e_{n+p|n}^{(n)}||_E \; ||e_{n+p|n}^{(\infty)} + e_{n+p|n}^{(n)}||_E. \tag{5.5}$$

The first factor on the right is $o(n^{-1/2})$ by Proposition 4.1. For the second, we have, since $\sigma_p \equiv ||e^{(\infty)}_{n+p|n}||_E$,

$$||e^{(\infty)}_{n+p|n} - e^{(n)}_{n+p|n}||_E \leq \sigma_p + ||e^{(n)}_{n+p|n}||_E.$$

The quantities $||e^{(n)}_{n+p|n}||_E$ converge to σ_p, by Proposition 4.1 again, so the second factor on the right in (5.5) is bounded, and (5.4) follows. This completes the proof of (5.3).

Mean absolute convergence as in (5.3) implies convergence in probability. Therefore, from (5.2) and (5.3) we obtain

$$N^{-1/2} \sum_{n=1}^{N} \delta_{n+p|n} - N^{1/2}(\sigma_p^2 - \tilde{\sigma}_p^2) \rightarrow_{dist.} \mathcal{N}(0, 2\pi f_\delta(0)). \tag{5.6}$$

and the following corollary.

Corollary 5.1. If w_N is consistent estimator of $(2\pi f_\delta(0))^{1/2}$ and if $f_\delta(0) \neq 0$, then

$$Z_N[p] \equiv (N^{1/2} w_N)^{-1} \sum_{n=1}^{N} \delta_{n+p|n}$$

is a test statistic which behaves like a $\mathcal{N}(0,1)$ variate for large enough N when $\sigma_p = \tilde{\sigma}_p$, and otherwise behaves like $N^{1/2}(\sigma_p^2 - \tilde{\sigma}_p^2)$, thereby revealing the sign of $\sigma_p - \tilde{\sigma}_p$, and with it the preferred model.

We plan to apply such a statistic, for several choices of p, also taking into account the uncertainties in $f(\theta)$ and $\tilde{f}(\theta)$ due to parameter estimation, to compare the pairs of competing models for the 40 times series considered in [5]. In [5], a statistic which is asympotically equivalent to $Z_N[1]$ was presented and used, as a time series generalization of the test statistic of [12].

Acknowledgements

I feel especially fortunate to have had Professor Gottfried Köthe as my *Doktorvater* .

While a student, I appreciated his great knowledge, the clarity of his thinking and writing, and his humor. As a mature researcher, I also admire the ways in which he encouraged independent thinking and the high example he set of how to treat colleagues and students with integrity and warmth.

This article was written while I was the Visiting Research Fellow of the Institute of Statistical Science of Academia Sinica in Taipei. I am grateful for the excellent research opportunities provided by this Institute and for the hospitality of the research fellows and staff there, especially Ching–Zong Wei and Min–Te Chao.

References

[1] G. Baxter, An asymptotic result for the finite predictor. *Math. Scandinavica* 10(1962), 137–144.

[2] G. Baxter, A norm inequality for a "finite-section" Wiener–Hopf equation. *Illinois J. Math.* 7(1963), 97–103.

[3] D. Brillinger, *Time Series Analysis* . New York: Holt, Rinehart, Winston (1975).

[4] A. Devinatz, Asympotic estimates for the finite predictor. *Math. Scandinavica* 15(1964), 111–120.

[5] D. Findley, Making difficult model comparisons. submitted for publication.

[6] E.J. Hannan, *Multiple Time Series* . New York: Wiley (1970).

[7] E.J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems.* New York: Wiley (1988).

[8] H. Helson and D. Lowdenslager, Prediction theory and Fourier series in several variables. *Acta Math. 99(1958)* , 165–202.

[9] I. Hirschman, Jr., Finite sections of Wiener–Hopf equations and Szegö polynominals. *J. Math. Anal. and Appl.* 11(1965), 290–320.

[10] M. Loève, *Probability Theory, 4th edn. Vol. I* . New York: Springer–Verlag (1977).

[11] B. Noble and J. Daniel, *Applied Linear Algebra, 3rd ed.* Englewood Cliffs: Prentice Hall (1988).

[12] M. Pourahmadi, On the convergence of finite linear predictors of stationary processes. *J. Mult. Anal.* 30(1989), 167–180.

[13] Q. Vuong, (1989) Likelihood ratio tests for model selection and non–nested hypotheses. *Econometrica* 57(1989), 307–333.

[14] Yu. A. Rozanov, *Stationary Random Processes* . San Francisco: Holden–Day (1967).