Electronic Journal of Statistics
Vol. 9 (2015) 789–791
ISSN: 1935-7524
DOI: 10.1214/15-EJS1010

Discussion of "High-dimensional autocovariance matrices and optimal linear prediction"*

Wei Biao Wu

Department of Statistics
University of Chicago
Chicago, IL 60637
e-mail: wbwu@galton.uchicago.edu

Abstract: In this note I provide some discussion on the paper by "High-dimensional auto covariance matrices and optimal linear prediction" by T. McMurry and D. Politis.

MSC 2010 subject classifications: Primary 62M10; secondary 62H12.
Keywords and phrases: Linear prediction, covariance matrix estimation, spectral density function.

Received February 2015.

I congratulate Professor McMurry and Professor Politis for their important contribution to the time series prediction problem. If the covariance functions were known, the classical celebrated Wiener-Kolmogorov theory lays a beautiful foundation of the prediction problem. In particular, one can derive various close-form expressions for the best linear predictors. In practice, however, the covariance functions are not known, and they need to be estimated from data. Wu and Pourahmadi (2009) showed that the associated estimated covariance matrix is not operator norm consistent if one simply replaces covariance functions by the estimated ones. In the current paper the authors considered the tapered estimates using partial or full samples. Their main results Theorem 3 and Corollary 2 assert that the linear predictor based on tapered covariance estimates is consistent.

My primary concern is the rate of convergence of the estimated linear predictors. Theorem 3 and Corollary 2 provide the consistency result

$$\hat{X}_{n+1} - \tilde{X}_{n+1} \to 0 \text{ in probability,} \tag{0.1}$$

where $\tilde{X}_{n+1} = \phi(n)^T (X_n, \dots, X_1)^T$ is the population version optimal linear predictor (cf. the authors' Equation (1)), $\phi(n) = \Gamma_n^{-1}\gamma(n)$, and $\hat{X}_{n+1} = \hat{\phi}(n)^T (X_n, \dots, X_1)^T$ is the estimated predictor. I believe that, under suitable conditions, one can derive a rate of convergence. The following is a heuristic argument. Given data $X_1, \dots, X_n$, we predict X_{n+1} by the linear form

*Main article 10.1214/15-EJS1000.

$X^*_{n+1} = a_1 X_n + \cdots + a_n X_1 := X^*_{n+1}(\mathbf{a})$. In comparison with the optimal linear predictor $\tilde{X}_{n+1}$, we have the mean square error

$$\begin{aligned} E|X^*_{n+1}(\mathbf{a}) - \tilde{X}_{n+1}|^2 &= E|\sum_{i=1}^{n}(a_i - \phi_i(n))X_{n+1-i}|^2 \\ &= \sum_{1\le i,j\le n}(a_i - \phi_i(n))(a_j - \phi_j(n))\gamma_{i-j}. \end{aligned} \quad (0.2)$$

Assume that the spectral density $f(\theta) = (2\pi)^{-1}\sum_{j=-\infty}^{\infty}\gamma_j \exp(\sqrt{-1}j\theta)$ satisfies

$$0 < c_1 \le 2\pi f(\theta) \le c_2 < \infty \text{ holds for all } \theta. \quad (0.3)$$

Then

$$c_1|\mathbf{a} - \phi|_2^2 \le E|X^*_{n+1}(\mathbf{a}) - \tilde{X}_{n+1}|^2 \le c_2|\mathbf{a} - \phi|_2^2 \quad (0.4)$$

By Theorem 2 in the paper, under proper conditions, one has

$$|\hat{\phi}_n - \phi_n| = O_p(r_n) \quad (0.5)$$

Using the argument in the paper and the technique for inverse moment bounds for sample autocovariance matrices in Cheng, Ing and Yu (2015), I believe that the following

$$|X^*_{n+1}(\hat{\phi}_n) - \tilde{X}_{n+1}|^2 = O_p(r_n^2). \quad (0.6)$$

or even the stronger statement can be valid

$$E|X^*_{n+1}(\hat{\phi}_n) - \tilde{X}_{n+1}|^2 = O(r_n^2). \quad (0.7)$$

The above bounds provide a refined version of their Theorem 3 and Corollary 2 in that it provides the rate of convergence.

The above consideration also motivates the problem of choosing optimal lag parameter l. Ideally, one should choose l such that it can minimize $E|X^*_{n+1}(\mathbf{a}) - \tilde{X}_{n+1}|^2$, or as an easier version in view of (0.4), one should minimize $|\hat{\phi}_n - \phi_n|_2$. Can the authors show that the data-driven choice of l in Section 2.3 enjoys certain consistency property; for example $\hat{l}/l \to 1$ in probability? In the nonparametric estimation literature such problems are well studied: one can show that, under proper conditions, the selected smoothing parameter is rate-consistent.

Acknowledgements

The research is supported in part by NSF Grants DMS–1405410 and DMS–1106790.

References

CHENG, T.-Z. F., ING, C.-K., AND YU, S.-H. (2015). Inverse moment bounds for sample autocovariance matrices based on detrended time series and their applications. To Appear *Linear Algebra and Its Applications*.

WU, W.-B. AND POURAHMADI, M. (2009). Banding sample covariance matrices of stationary processes. *Statistica Sinica*, 19 1755–1768