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

Yulia R. Gel[†]

Department of Mathematical Sciences, University of Texas, Dallas, USA
e-mail: ygl@utdallas.edu

and

Wilfredo Palma[‡]

Department of Statistics, Pontificia Universidad Catolica de Chile, Santiago, Chile
and Millennium Institute of Astrophysics, Chile
e-mail: wilfredo@mat.puc.cl

We would like to congratulate the authors for this highly motivating work on prediction of stationary time series. This paper does an excellent job addressing the relationship between the problems of estimating high dimensional covariance matrices and finding best linear predictors, and systematizing the current approaches to regularization in time series. Several regularization procedures are proposed, tackling the issue of calculating optimal predictors in the case that p-order dependence approaches the sample size n. The proposed techniques are validated by extensive Monte Carlo simulations and real-data applications.

It was nice to be asked to discuss such a stimulating paper. We divided our comments into several sections, striving to target different strains of ideas presented in the paper and future research directions motivated by these strains.

Offline vs. online prediction problem The current paper focuses on the utility of regularization procedures for time series in a prediction framework. In contrast to (auto)covariance matrix estimation, one of the primary motivations behind using regularization in forecasting time series is to improve a prediction

performance by incorporating more historical data than it is suggested by the Akaike Information criteria and other classical model selection procedures. We can further segment the prediction problem into two major cases:

1) *Offline Problem.* Let us consider a sample of T observations where T is fixed. Our main goal is then to most effectively utilize the available T observations for different forecasting horizons, i.e. predicting $\hat{y}_{T+1}, \ldots, \hat{y}_{T+h}$ given the previous history $y_1, \ldots, y_T$. However, since T is fixed, we are not substantially concerned about computational costs of model estimation and our focus is primarily on the quality of the delivered forecasts.

2) *Online Problem.* The situation changes drastically if a number of observation T is (potentially indefinitely) increasing. Hence, the collected history of previous time series data is rapidly expanding and model selection, model estimation and forecasting are to be performed in an online setting, or real time mode. Obviously, under such conditions, not only the accuracy of the delivered forecasts is of high importance but how fast and computationally efficient are the methods that are being used to produce these forecasts. (Indeed, even in 1960–70s the reduced computational costs due to recursivity of a Kalman filter were one of the main reasons for its high popularity in signal processing and other related fields where online system identification and online forecasting are widely spread (Ljung, 1999). Nowadays, online forecasting problems range from finance to fMRI processing to astronomy and are rapidly evolving in a Big Data paradigm.) In such an online framework, it is not clear whether it is worthwhile to use a regularized $T \times T$-autocovariance matrix because at the next time point $T+1$ we need to estimate a $(T+1) \times (T+1)$-regularized autocovariance matrix and a model of order $T + 1$ and so on, and to the best of our knowledge there exists no reasonably fast method for scaling a $T \times T$-sample autocovariance matrix up to a $(T+1) \times (T+1)$-sample autocovariance matrix. (The Levinson-Durbin (LB) algorithm provides some paths in this direction but the sample autocovariance matrix is only approximately Teoplitz and it also remains unclear how to combine the LB algorithm with the updates of a regularization parameter.) Gel and Barabanov (2007) and Bickel and Gel (2011) addressed this issue by estimating a $p_T \times p_T$-sample regularized autocovariance matrix so the update of an autocovariance matrix and the respective time series model are performed in batches of observations rather than at each single newly arrived observation.

In this light can the authors comment which of the forecasting problems, i.e. offline or online, they envision to be the best fit for the proposed prediction methodology? If it is an online problem, how the computational costs due to non-recursivity of a sample autocovariance estimation could be addressed? If it is an offline problem, are there any issues and limitations of an increasing forecasting horizon h and what is the effect of different h on a regularization procedure?

Assessing forecasting quality   The set of Monte Carlo simulations for assessing the prediction performance via RMSPE is very illustrative. Table 1 to

Table 3 effectively illustrate the forecasting quality of the the twelve approaches. At the first glance, all these methods exhibit comparable RMSPE values, however the standard errors evidenced by some of the techniques seem to be quite different form others. For example, the FSO-Th-Raw method seems to exhibit a lower relative performance, especially for high values of either the first order autoregressive parameter $|\phi|$ or the first order moving average parameter $|\theta|$, respectively. It must be noticed though that these simulations involve only one-step predictions. It would be interesting to evaluate these procedures in the context of larger prediction horizons.

On the other hand, the real data experiments based on 105 times series from the M3 competition are quite illustrative about the performance of the different methods. Given the nonparametric nature of the prediction techniques proposed in the paper, this study illuminates the forecasting performance in a more realistic context. In this sense, Table 5 seems to suggest that all the approaches seem to have a roughly similar prediction quality.

More generally, the comments above relate to a predictive performance metrics for the *same-realization* and *independent-realization* prediction. As noted by Ing and Wei (2005) properties of the classical model selection criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), etc., are different for the *same-realization* and *independent-realization* prediction of an infinite-order autoregressive $AR(\infty)$ process. This phenomenon is due to the fact that for the *same-realization* prediction newly arrived observations are no longer independent of the previous data. It appears that regularization procedures particularly target the case of the *same-realization* prediction. While it is conventional to use a root mean squared error (RMSE) for validating prediction performance, i.e. a classical measure for the *independent-realization* prediction, the current paper also considers standard deviation of RMSE and thus motivates for a search of alternative measures for predictive performance that are more suitable for the *same-realization* prediction.

A path long memory and other open issues The manuscript opens a number of interesting venues for future research. For instance, it would be interesting to further investigate the practical consequences of Theorem 2. The optimal predictor coefficients $\phi_j(n)$ tend to zero as the lag j and n increase. In this sense, it would be expected or desirable that the L_2 norm of the vector of the difference between the estimated coefficients and the true values converges to zero. Another interesting issue is to establish the respective convergence rate for a more general class of processes. In particular, the derived rate r_n applies only to the short memory case (or more generally to weakly dependent stationary processes) but excludes long memory processes.

Along this interesting path, Ing et al. (2015) study the problem of estimating inverse autocovariance matrices in the context of strongly dependent processes. As pointed out by Palma and Pourahmadi (2012), consistent estimation of the autocovariance matrices for stationary long-range dependent processes is a challenging task due to the fact that the off-diagonal elements do not tend to zero

fast enough as compared to the short memory case. Nevertheless, the inverse covariance matrix behaves much better. A mathematical explanation of this phenomenon is as follows. Suppose that $\Gamma_n(f)$ is the variance covariance matrix of the stationary process with spectral density f. Thus, we can write

$$\Gamma_n(f) = \left[\int_{-\pi}^{\pi} f(\lambda) \exp(i\,\lambda\,(\ell - j) d\lambda \right]_{\ell, j = 1, \ldots, n}.$$

The inverse of the covariance matrix can be well approximated by

$$\Gamma_n^{-1}(f) \approx \Gamma_n(\{4\pi^2 f\}^{-1})$$

cf., Lemma 5.2 and Lemma 5.3 of Dahlhaus (1989). Even though the spectral density of a long-memory process has a pole at zero, the inverse spectral density behaves very well at the origin. This approximation allows to circumvent the problem generated by the singularity of the spectrum at zero frequency by considering the well-behaved inverse of the spectral density.

References

BICKEL, P. and GEL, Y. (2011). Banded regularisation of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society*, 73, 711–728. MR2867455

DAHLHAUS, R. (1989). Efficient estimation self-similar processes. *Annals of Statistics*, 17, 1749–1766. MR1026311

GEL, Y. R. and BARABANOV, A. (2007). Strong consistency of the regularized least-squares estimates of infinite autoregressive models. *Journal of Statistical Planning and Inference*, 137, 1260–1277. MR2301478

ING, C. K. and WEI, C. Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics*, 33(5), 2423–2474.

ING, C. K., CHIOU, H. T. and GUO, M. (2015). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli*, to appear.

LJUNG, L. (1999). *System Identification: Theory for the User* (2nd Edition). Prentice Hall.

PALMA, W. and POURAHMADI, M. (2012). Banded regularization and prediction of long-memory time series. *Working Paper*.