

# Discussion of “High-dimensional autocovariance matrices and optimal linear prediction”\*

Rob J. Hyndman

*Department of Econometrics & Business Statistics*

*Monash University*

*Clayton VIC 3800*

*Australia*

*e-mail: [Rob.Hyndman@monash.edu](mailto:Rob.Hyndman@monash.edu)*

**Abstract:** I propose new ACF and PACF plots based on the autocovariance estimators of McMurry and Politis. I also show that the forecasting methods they propose perform poorly compared to some relatively simple autoregression algorithms already available.

**Keywords and phrases:** Autocorrelation, autoregression, data visualization, forecasting, serial correlation, time series graphics.

Received October 2014.

It is always interesting to see new approaches to old problems, and McMurry and Politis have provided some very stimulating ideas in their approach to autocovariance estimation and forecasting. I am particularly interested in the usefulness of their results for analysing and forecasting real time series.

There are three potential uses of their methods for data analysis that I would like to explore. First, the standard ACF plot could be replaced by a plot based on a tapered estimate of the autocovariance matrix. Second, a corresponding PACF plot could be obtained using the Durbin-Levinson recursions applied to a tapered estimate of the autocovariance matrix. Third, the proposed FSO or PSO predictors could be used for forecasting real time series (as in fact the authors do in Section 5.4). I will reflect on each of these in turn.

## 1. A better ACF plot

The standard ACF plot is notoriously unreliable for large lags, and so any improvement in these estimates would be welcomed by time series analysts. In addition, most software produces ACF plots based on poor graphical choices making them more difficult than necessary to interpret. Consequently, I will propose a better ACF plot based on the tapered and banded autocovariance estimator of McMurry and Politis, and with improved graphical presentation.

---

\*Main article [10.1214/15-EJS1000](https://doi.org/10.1214/15-EJS1000).

We shall define the estimated ACF using a data-based choice of the banding parameter, and with eigenvalue thresholding to ensure positive definiteness. For the time series  $\{X_1, \dots, X_n\}$ , let  $\hat{\rho}_s = \kappa(|s|/\ell)\check{\gamma}_s/\check{\gamma}_0$  where  $s = 0, 1, 2, \dots, n-1$ ,

$$\check{\gamma}_s = n^{-1} \sum_{t=1}^{n-|s|} X_t X_{t+|s|}, \quad \kappa(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ 2 - |x| & \text{if } 1 < |x| \leq 2 \\ 0 & \text{otherwise;} \end{cases}$$

and  $\ell$  is the smallest positive integer such that  $|\check{\gamma}_{\ell+k}/\check{\gamma}_0| < c(\log_{10} n/n)^{1/2}$  for  $k = 1, \dots, K$ . The corresponding autocorrelation matrix is  $\hat{R}$  with  $(i, j)$ th element  $\hat{\rho}_{|i-j|}$ . We then use eigenvalue thresholding to define a new autocorrelation matrix  $\tilde{R} = VDV'$  where  $\hat{R} = V\Lambda V'$  is the eigendecomposition of  $\hat{R}$  and  $D$  is a diagonal matrix with  $(i, i)$ th element equal to the maximum of  $\Lambda_{i,i}$  and  $\epsilon/n^\beta$ . Finally, we define

$$\tilde{\rho}_s = \frac{1}{(n-s)\bar{d}} \sum_{k=1}^{n-s} \tilde{R}_{k,k+s}$$

where  $\bar{d}$  is the mean of the diagonals of  $D$ . Following the suggestions of McMurry and Politis, I use  $c = 2$ ,  $K = 5$ ,  $\epsilon = 20$  and  $\beta = 1$ .

To demonstrate the new estimator, I have applied it to the seasonally differenced US monthly housing sales data (Makridakis et al. 1998, Chapter 3) in Figure 1 from 1973–1995. The plot on the left uses the standard estimator  $\hat{\rho}_s = \check{\gamma}_s/\check{\gamma}_0$  and was obtained using the default settings for the `acf()` function in R (R Core Team 2014), except that I selected 100 lags. The plot on the right uses the estimator  $\tilde{\rho}_s$  defined above.

The blue lines in the left-hand panel shows the 5% critical values at  $\pm 1.96n^{-1/2}$  under the null hypothesis of white noise. These are often misleading as we are

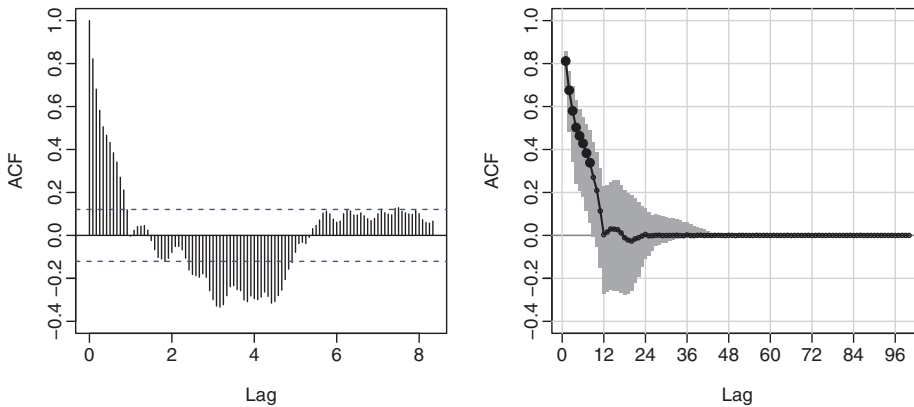


FIG 1. Left: traditional ACF plot with 5% critical values at  $\pm 1.96n^{-1/2}$ . Right: Proposed new ACF plot based on a tapered estimator of the autocorrelation with bootstrapped confidence intervals shown as a gray shaded bars. Values significantly different from zero are shown as large solid points.

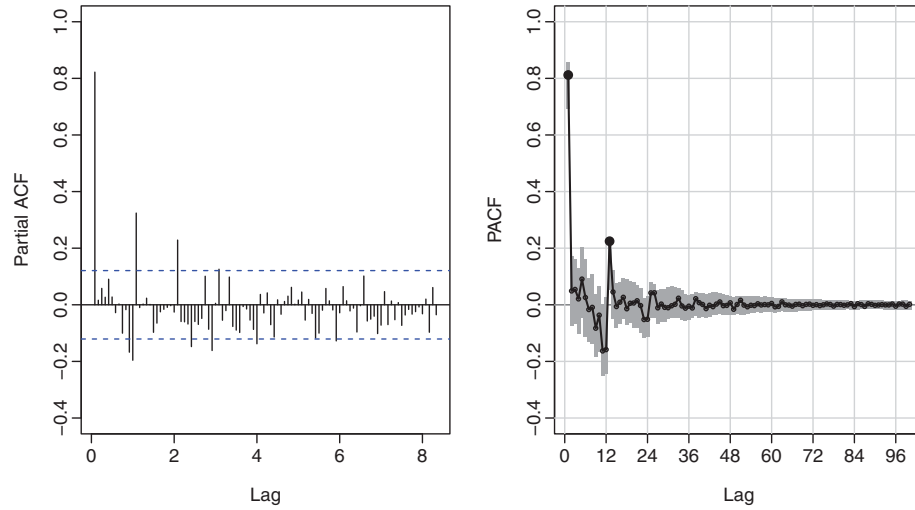


FIG 2. Left: traditional PACF plot with 5% critical values at  $\pm 1.96n^{-1/2}$ . Right: Proposed new PACF plot based on a tapered estimator of the autocorrelation with bootstrapped confidence intervals shown as a gray shaded bars. Values significantly different from zero are shown as large solid points.

usually interested in whether the autocorrelations are significantly different from zero under a null hypothesis that the data are from a stationary process (rather than a white noise process). In particular, the long section of significant negative autocorrelations in the left-hand panel is probably not of particular interest, and I frequently have to tell my confused students to ignore such features. The horizontal axis is labelled “Lag” but the axis is marked in units equal to one year rather than in lag units. It is possible to over-ride these defaults, but good software should have good default settings.

The right-hand panel demonstrates an alternative plot. The shaded bars show 95% bootstrapped confidence intervals based on the linear process bootstrap (McMurry & Politis 2010) obtained using the same autocovariance estimate that is plotted. Autocorrelations that are significantly different from zero are highlighted using large solid circles, while insignificant autocorrelations are shown using small open circles. The x-axis shows the number of lags with tick-marks at multiples of years. Finally, the pointless lag 0 autocorrelation has been omitted. This version of an ACF plot should be much easier for students to read and interpret correctly.

## 2. A better PACF plot

It is possible to obtain a corresponding estimate of the partial autocorrelation function using the Durbin-Levinson recursions (Morettin 1984) applied to the autocorrelation estimates  $\{\hat{\rho}_0, \hat{\rho}_1, \dots, \hat{\rho}_{n-1}\}$ .

Figure 2 shows the traditional and proposed PACF plots for the same housing sales data as shown in Figure 1. While the same improvements are evident, the new plot obscures some potentially important information. In the left-hand panel, there are significant autocorrelations near lags 12, 24 and 36 indicating some seasonality in the data. Because they decline geometrically, this is suggestive of a seasonal MA(1) process. The tapering and shrinkage applied to the autocovariances has meant the corresponding autocorrelations near lags 24 and 36 in the right-hand plot are insignificant.

It appears that the parameter choices for  $c$ ,  $K$ ,  $\epsilon$  and  $\beta$  may need refinement, especially with seasonal data, to prevent important information being obscured. In other examples (not shown here), the partial autocorrelations increased in size for very large lags (and even became greater than one in absolute value). These problems are due to insufficient shrinkage of the eigenvalues, and provide further indication that better selection of the values of  $\epsilon$  and  $\beta$  is required before these estimators could be routinely used in real data analysis.

### 3. Forecasting performance

One surprising aspect of the results presented by McMurry and Politis is that their proposed forecasting methods do relatively well in the simulations. I had expected that with  $n$  observations, it would be impossible to satisfactorily forecast with an AR( $p$ ) model where  $p = O(n)$ , but they have demonstrated otherwise. This is interesting, and deepens our understanding of the nature of the problem, but it does not help forecasters in practice.

Even for the simulations with low-order stationary AR and MA processes, the proposed methods never give much improvement, and are usually worse than the benchmark AR method. In these ideal circumstances, one would expect the proposed methods to perform at their best.

For the real (M3) data, the reported results in Table 5 show that their Rect-SSBC-Raw method does slightly worse than the simple benchmark AR approach (with an RMSE of 0.8572 compared to 0.8356). The benchmark AR approach uses the `ar()` function in R with default settings (which employs Yule-Walker estimates and uses the AIC to select the order).

By way of comparison, I tried using the `auto.arima()` function from the forecast package (Hyndman & Khandakar 2008) with maximum likelihood estimation and AR order selection using the AICc (Hurvich & Tsai 1997). I restricted the models to purely autoregressive models in order to ensure comparability with the results of McMurry and Politis. The resulting RMSE was 0.7821, substantially and significantly better than any of the results from McMurry and Politis. The corresponding result for the reversed series was 0.7470. So even if we restrict the models to pure autoregressions, it is easy to get much better results than any of those obtained by McMurry and Politis.

It seems that for forecasting purposes, the methods proposed by McMurry and Politis perform poorly compared to the relatively simple algorithms already available. While their results appear to be very useful in estimating high-

dimensional autocovariance matrices and autocorrelation functions, they do not provide “optimal linear prediction” as claimed.

## References

- HURVICH, C. & TSAI, C. (1997), ‘Selection of a multistep linear predictor for short time series’, *Statistica Sinica* **7**, 395–406. [MR1466688](#)
- HYNDMAN, R. J. & KHANDAKAR, Y. (2008), ‘Automatic time series forecasting: The forecast package for R’, *Journal of Statistical Software* **26**(3), 1–22.
- MAKRIDAKIS, S. G., WHEELWRIGHT, S. C. & HYNDMAN, R. J. (1998), *Forecasting: Methods and Applications*, 3rd edition edn, John Wiley and Sons, New York.
- MCMURRY, T. L. & POLITIS, D. N. (2010), ‘Banded and tapered estimates for autocovariance matrices and the linear process bootstrap’, *J. Time Series Analysis* **31**(6), 471–482. [MR2732601](#)
- MORETTIN, P. A. (1984), ‘The Levinson algorithm and its applications in time series analysis’, *International Statistical Review* **52**(1), 83–92. [MR0967203](#)
- R CORE TEAM (2014), *R: A Language and Environment for Statistical Computing*, Vienna, Austria. <http://www.r-project.org>