STATISTICS &
PROBABILITY
LETTERS

www.elsevier.com/locate/stapro
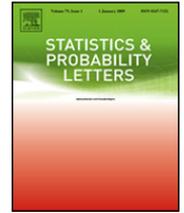
The perils of inferring serial dependence from sample autocorrelations of moving average series☆

Tucker McElroy

Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, United States

A B S T R A C T

We demonstrate that oscillatory patterns in the higher lags of sample autocorrelations can arise whenever the true process is a finite order MA, and that this phenomenon exists even when the true autocorrelations are zero. Therefore the visually apparent structure is a statistical artifact, and the analyst should not attempt to model it directly. Instead one should utilize Box–Jenkins methodology, whereby appropriate significance levels for testing zero correlation can be obtained by fitting successively higher order MA models.

Published by Elsevier B.V.

1. Introduction

This paper is concerned with the phenomena of spurious structure apparent in sample autocorrelation plots of time series. Consider the two sample autocorrelation plots in Fig. 1. One is the sample autocorrelation plot of a real quarterly time series of length 50, and the other is the sample autocorrelation of a simulated MA(3) process of length 50 (full details on the series are provided in Section 3). The plots are standard output from R (R Development Core Team, 2011), with confidence bands (the dashed blue lines) automatically generated; these are useful for testing whether a particular sample autocorrelation is nonzero *under the null hypothesis that the actual process is white noise*. There is little to visually distinguish the two plots qualitatively.

From classical time series methodology (Box and Jenkins, 1976), we know the MA(3) process has true autocorrelations equal to zero at lags exceeding three, and thus the sample autocorrelations converge in probability to zero at these higher lags. The asymptotic properties are described through the classical Bartlett formula, summarized in Theorems 7.2.1 and 7.2.2 of Brockwell and Davis (1991). Although the real series may not obey asymptotic theory, the Gaussian simulation should; yet both series in Fig. 1 appear to manifest large sample autocorrelations at higher lags. In fact, these sample autocorrelations appear to be significantly different from zero when compared to $\pm 1.96/\sqrt{50}$, a paradox for the MA(3) process.

The paradox in the last sentence is resolved by noting that this significance is with respect to a null hypothesis that the true process is white noise. If instead we utilize the null hypothesis that the true process is an MA(3), and we test whether higher lag sample autocorrelations are nonzero, the confidence bands will be widened in accordance with the Bartlett formula (equation 7.2.5 of Brockwell and Davis, 1991; also see Example 7.2.2). In fact, for the simulated process where the true MA coefficients are known, application of the Bartlett formula results in the variance of these estimates being expanded by the factor 11/4. As a result, all apparent structure in the autocorrelation plot (of the simulation) at lags four and higher is completely spurious. (Note that the Bartlett formula is only appropriate for linear series, as emphasized

---
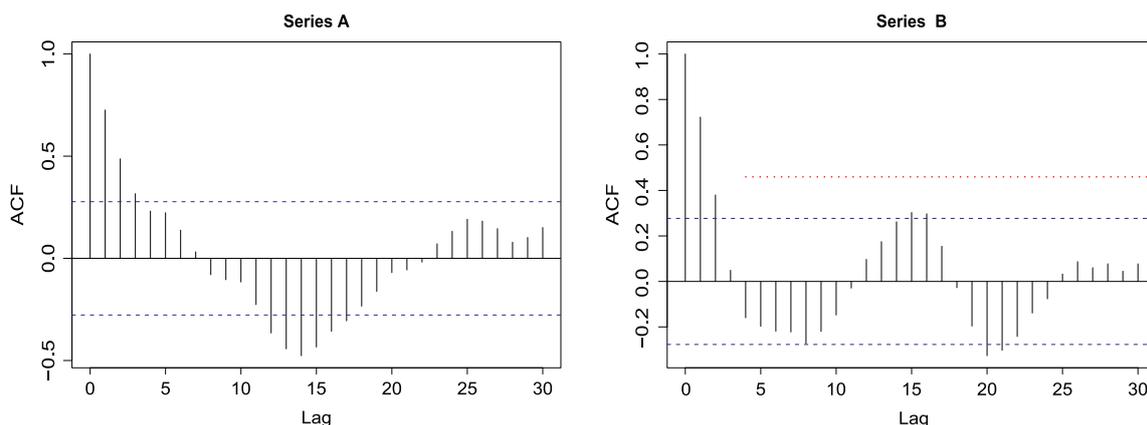
Fig. 1. Sample autocorrelations for Series A (a seasonally differenced quarterly series of outstanding loans) and Series B (a simulation from an MA(3) process), with Barlett confidence bands (dashed) computed under a white noise hypothesis. For Series B, the Bartlett bands (dotted) are also provided for the MA(3) hypothesis —appropriately displayed for lags 4 and higher, just for the positive y values for clarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in Romano and Thombs (1996); hence it is important to know whether the innovations in the Wold decomposition are independent or merely uncorrelated.)

This is perplexing, because the eye is instantly drawn to the pleasant wave-like pattern, and the statistician's instinct is to model that structure. Of course, it has long been known that sample autocovariances have high variability (see Box and Jenkins, 1976), but the point here is the remarkable structure over increasing lag. In particular, one might seek to utilize a damped cosine autocovariance function as a model, perhaps arising from an ARMA or Long Memory model. Such an approach is gravely misplaced for the MA(3) process, because the apparent structure at higher lags is actually completely spurious, i.e., it is a statistical artifact (see below). Given that it is difficult for us to distinguish the two plots on a qualitative basis, the instinct to model the apparent wave structure might also be a mistake for real data exhibiting such an autocorrelation pattern.

Box and Jenkins (1976) pointed out that the sample autocorrelation function tends to be itself autocorrelated over lags. This phenomenon is not due to small sample size; similar results have been encountered by the author for samples of length 100 to 200, for both real and simulated time series. However, the amplitude of the waves does tend to diminish as sample size increases, being squeezed by the Bartlett error formula at $\sqrt{n}$ rate. Neither is the phenomenon due to a quirky simulation in the MA(3) process—changing the random seed results in a different wave pattern, to be sure (but more or less the same structure at the first three lags), which is nevertheless qualitatively the same. In fact, such behavior *always* manifests in MA processes, as is demonstrated in Section 2; it is a fundamental property of the sample autocovariance function, viewed as a sequence in the lag.

Since the autocorrelation function for an MA process truncates to zero eventually, we might guess that we can identify the MA order by finding a truncation point (i.e., where the correlations are no longer significantly different from zero) in the sample autocorrelation plot. (Note that the identification problem also requires study of the sample partial autocorrelations.) In the scenario depicted in Fig. 1, the cutoff point is unclear due to the wave-pattern of the sample autocorrelations. As recommended in Box and Jenkins (1976), one can fit successively higher order MA models to the data, testing for significance using the Bartlett formula for the previous MA model. Proceeding in this manner, one can obtain the true MA order; then, modulo Type I errors, one avoids finding spurious time series structure in the wave-pattern of the sample autocorrelations.

The first point of this paper is that caution is needed in the interpretation of sample autocorrelation plots, in order to avoid an over-fitting pit-fall (e.g., fitting a long memory model to an MA(3) process). Although the human eye finds structure in the higher lags in Fig. 1, these patterns may be spurious. Employing the Box–Jenkins procedure of fitting successively higher-order MA models will be successful in avoiding misspecification for the MA(3) example, whereas attempts to directly model the wave structure in the sample autocorrelations may lead to incorrect results. This point is not new, given its formulation in Box and Jenkins (1976), but given our novel explanation of the phenomenon of such wave-structure, it seems worth reiterating.

The second objective of the paper is to give a mathematical explanation for the appearance of these oscillatory waves in the sample autocorrelations of MA processes. Although quite simple to derive, we have not seen these results before in the time series literature, and we believe that our representation – that the sample autocovariance function can be viewed as the output of a symmetric filter acting on the sample autocovariances of white noise – sheds light on the situation.

2. The sample autocorrelations of MA processes

Consider the available sample written as a column vector $X = (X_1, X_2, \ldots, X_n)'$, and note that the sample autocovariance function at lag h (if we do not center by the sample mean) is

$$R(h) = n^{-1}X'L_n^h X,$$

where L_n is the $n \times n$ lag matrix, defined as being nonzero only on the first sub-diagonal, which has all ones. See Pollock (1999). Suppose that the time series sample is drawn from an MA(q) process $X_t = \theta(B)Z_t$ with $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, where B is the backshift operator. Also $F = B^{-1}$ denotes the forward shift operator. The matrix polynomial $\theta(L_{n+q})$ is then defined as in Pollock (1999) by taking appropriate powers of the lag matrix. That is, $\theta(L_{n+q}) = 1_{n+q} + \sum_{k=1}^{q} \theta_k L_{n+q}^k$, where 1_{n+q} is the $n + q$-dimensional identity matrix and L_{n+q}^k denotes the kth power of the lag matrix–this matrix has all zero entries except for the kth subdiagonal, which has unit entries. Then writing $\theta_0 = 1$ and $\theta_k = 0$ for $k < 0$ or $k > q$, we see that the jkth entry of $\theta(L_{n+q})$ is θ_{j-k}. Then define the $n \times n + q$-dimensional matrix Θ by $\Theta = [0_{n \times q} \; 1_n]\theta(L_{n+q})$, so that the jkth entry of Θ is $\Theta_{jk} = \theta_{j-k+q}$. Then

$$X = \Theta Z \qquad \Theta = [0_{n \times q} \; 1_n]\theta(L_{n+q})$$

defines a matrix relation between the available sample X and the unknown innovations $Z = (Z_{1-q}, Z_{2-q}, \ldots, Z_n)'$ that appear in the various observed data. As a result

$$R(h) = n^{-1}Z'\left[\Theta'L_n^h\Theta\right]Z.$$

We show in Theorem 1 that when $h > q$, the matrix in square brackets is approximately Toeplitz with corresponding spectral function $|\theta(z)|^2 z^h$, where $z = e^{-i\lambda}$. Define the inverse Fourier Transform (FT) of an arbitrary function g with domain $[-\pi, \pi]$ to be $\gamma_h(g) = (2\pi)^{-1}\int_{-\pi}^{\pi} z^{-h}g(\lambda)\,d\lambda$. We can then define a square matrix $\Sigma(g)$ by assigning $\gamma_{j-k}(g)$ to be its jkth entry. Because $\Sigma(g)$ has entries only depending on the difference of the row and column index, it is Toeplitz by definition (Pollock, 1999). Then we can show the following result.

Theorem 1. *If $\{X_t\}$ is the output of a moving average filter of length q – say $X_t = \theta(B)Z_t$ for a degree q polynomial $\theta(B)$ – acting on a stationary time series, then when $h > q$ its sample autocovariance satisfies*

$$R(h) = n^{-1}Z'\Sigma\left(|\theta(z)|^2 z^h\right)Z + O_P(n^{-1})$$

as $n \to \infty$.

Proof of Theorem 1. First observe that

$$\left[\Theta'L_n^h\Theta\right]_{jk} = \sum_{\ell,m}\Theta'_{j\ell}\left[L_n^h\right]_{\ell m}\Theta_{mk} = \sum_{\ell,m}\theta_{\ell-j+q}1_{\{\ell=m+h\}}\theta_{m-k+q} = \sum_{\ell=h+1}^{n}\theta_{\ell-j+q}\theta_{\ell-h-k+q}.$$

On the other hand,

$$\gamma_{j-k}\left(|\theta(z)|^2 z^h\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} z^{-j+k+h}|\theta(z)|^2\,d\lambda$$

$$= \sum_{\ell,m}\theta_\ell\theta_m\frac{1}{2\pi}\int_{-\pi}^{\pi} z^{\ell-m-j+k+h}\,d\lambda$$

$$= \sum_{\ell}\theta_\ell\theta_{\ell-j+k+h}$$

$$= \sum_{\ell=-\infty}^{\infty}\theta_{\ell-j+q}\theta_{\ell-h-k+q}.$$

Because $\theta(B)$ has degree q and $h > q$, this summation is really over $\ell = 1, \ldots, n + h$. Then the difference between the jkth entries of $\Sigma(|\theta(z)|^2 z^h)$ and $\Theta'L_n^h\Theta$ is

$$\sum_{1\le\ell\le h,\; n+1\le\ell\le n+h}\theta_{\ell-j-q}\theta_{\ell-h-k+q}.$$

Let M denote the difference between the two matrices, such that M_{jk} is the above quantity. Then

$$R(h) - n^{-1}Z'\Sigma\left(|\theta(z)|^2 z^h\right)Z = -n^{-1}Z'MZ = -n^{-1}\sum_{1\le\ell\le h,\; n+1\le\ell\le n+h}\left[\sum_{j,k=1}^{n+q}Z_j Z_k \theta_{\ell-j-q}\theta_{\ell-h-k+q}\right].$$

For each ℓ in the outer summation, only a finite number of Z_ts are present in the double summation in the square brackets, because there are only a finite number of nonzero θ coefficients. Because each of these variables is $O_P(1)$, the discrepancy is $O_P(n^{-1})$. $\square$

Although Theorem 1 is stated somewhat generally, we are interested in this paper with the case that $\{Z_t\}$ is white noise. Let I denote the periodogram of the white noise sample Z of length $n + q$, and observe that the sample autocovariance function can be expressed as $\gamma_h(I)$, because I is by definition the FT of the sample autocovariance sequence. Then

$$n^{-1}Z'\Sigma\left(|\theta(z)|^2 z^h\right)Z = \frac{1}{2\pi}\int_{-\pi}^{\pi}|\theta(z)|^2 I(\lambda)z^h\,d\lambda,$$
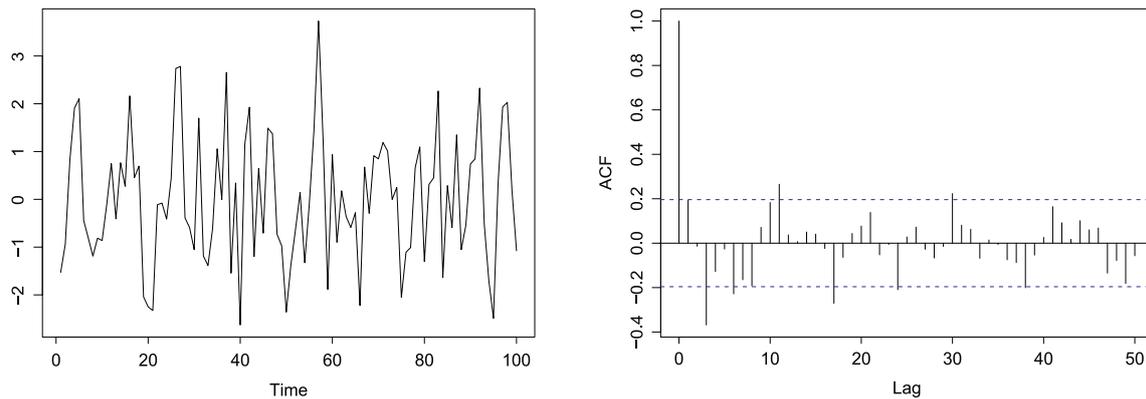
Fig. 2. A simulation from differenced white noise (left panel) of length 200, and the sample autocorrelation plot (right panel) with Barlett confidence bands (dashed) computed under a white noise hypothesis.

which follows from algebra and the fact that $\Sigma_{jk}(|\theta(z)|^2 z^h) = (2\pi)^{-1} \int_{-\pi}^{\pi} |\theta(z)|^2 z^{h-j+k} \, d\lambda$. Next, consider the sequence $\gamma_{-h}(I)$, which equals $(2\pi)^{-1} \int_{-\pi}^{\pi} I(\lambda) z^h \, d\lambda$ (and is the same as $\gamma_h(I)$). Applying $\theta(B)$ and $\theta(F)$ in succession to this sequence – where B and F act upon the index h – yields

$$\theta(B)\theta(F)\gamma_{-h}(I) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\theta(z)|^2 I(\lambda) z^h \, d\lambda.$$

In other words, the sample autocorrelation of the MA process asymptotically equals a double-sided MA filter acting on the sample autocorrelation of the underlying white noise process. Now consider $\{\gamma_h(I)\}$ as a time series in its own right, with h indexing time fixed to be positive ($h > 0$). The expectation is then zero and there is no serial correlation, the variance being n^{-1} times the variance of Z_1. That is, the sample autocovariance of white noise $\{\gamma_h(I)\}_{h \geq 1}$ behaves itself like mean zero white noise (if the original series was Gaussian, then the autocorrelation process is not Gaussian, being a quadratic process). Seen this way, the sample autocorrelation for the MA process at lags $h > q$ is equal to the filter $\theta(B)\theta(F)$ acting on a white noise sequence.

Therefore the serial structure of $R(h)$, viewed as a time series in the index h, is governed by the spectral characteristics of $\theta(z)$. If this polynomial corresponds to a low-pass filter structure (e.g., $\theta(B) = 1 + B + B^2 + B^3$) then $R(h)$ will have slowly oscillatory waves compatible with high positive serial correlation. If instead it corresponds to a high-pass filter (e.g., $\theta(B) = 1 - B$) then $R(h)$ will oscillate rapidly. An example of this high frequency behavior is provided in Fig. 2, where we simulate a Gaussian time series of length 200 with $\theta(B) = 1 - B$, and also plot its sample autocorrelations. (To correct the standard errors on the bands using the Bartlett formula, we should multiply by the square root of 3/2.)

3. Conclusion

We have demonstrated that purely random effects can manifest themselves as an apparent structure in sample autocorrelation plots. In particular, for an MA(q) process the sample autocorrelation at lags $h > q$ behaves statistically like a double-sided moving average filter applied to white noise. Hence if the original process has spectral mass at the lower frequencies, its sample autocorrelation will tend to have a slowly-moving oscillatory pattern much like the original series! That is, whereas the MA process looks like $\theta(B)$ applied to white noise, its sample autocorrelation looks like $\theta(B)\theta(F)$ applied to white noise.

This type of sample autocorrelation pattern is not uncommon in empirical examples, in our experience. In Fig. 1, Series A corresponds to a quarterly time series that was logged and seasonally differenced, the resulting sample autocorrelation plot being computed and presented.[1] Series B corresponds to a simulation from a Gaussian MA process with $\theta(B) = 1 + B + B^2 + B^3$ and unit innovation variance. In this case we can apply the Bartlett formula to the MA(3) process, obtaining that the sum of the autocorrelations in (7.2.5) of Brockwell and Davis (1991) equals 11/4. Whereas the standard error under a white noise hypothesis is 0.277, the standard error under this MA(3) hypothesis is 0.460. Fig. 1 presents the correct band as a dotted red line, appropriate for lags four and higher; now none of the sample autocorrelations are significant.

Note that seasonal differencing entails application of the high-pass filter

$$1 - B^4 = (1 - B)(1 + B + B^2 + B^3).$$

If the raw Series A were not actually seasonal, but rather a random walk, then application of $1 - B^4$ to the data would produce an MA(3) time series, namely $X_t = (1 + B + B^2 + B^3)Z_t$, of which Series B is an actual example. In other words, completely spurious serial correlation patterns are imposed upon the data by inappropriate selection of nonstationary unit root differencing operators!

---

[1] The time series title is "Quarterly amounts outstanding of UK resident monetary financial institutions" in sterling millions; dates are 1997.Q4 through 2011.Q1. The series was obtained from Fida Hussain of the Bank of England.

In summary, we have shown that oscillatory patterns in the higher lags of sample autocorrelations can arise whenever the true process is a finite order MA, and that this phenomenon exists even when the true autocorrelations are zero. The statistical behavior of the wave patterns is that of a forward- and backward-filtered white noise sequence, the filter corresponding exactly to the MA polynomial of the underlying process. Hence the visually apparent structure is a statistical artifact. Because a given time series may exhibit this behavior in its sample autocorrelation plot, great caution is needed in the plot's proper interpretation. The appropriate significance levels for testing zero correlation can be obtained by fitting successively higher order MA models, as outlined in Box and Jenkins (1976).

Acknowledgments

References

Box, G., Jenkins, G., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
Brockwell, P., Davis, R., 1991. Time Series: Theory and Methods. Springer-Verlag, New York.
Pollock, D., 1999. A Handbook of Time-Series Analysis, Signal Processing, and Dynamics. Academic Press, New York.
R Development Core Team, 2011. R foundation for statistical computing.
Romano, J., Thombs, L., 1996. Inference for autocorrelations under weak assumptions. Journal of the American Statistical Association 91, 590–600.