Biometrika (1978), 65, 1, *pp.* 135–40
With 3 text-figures
Printed in Great Britain

Interpreting partial autocorrelation functions of seasonal time series models

By DAVID C. HAMILTON and DONALD G. WATTS
Department of Mathematics, Queen's University, Kingston, Ontario

Summary

A simple approach to understanding the behaviour of the partial autocorrelation function of seasonal time series is presented, based on a partial autocorrelation pattern. This pattern, which acts as a signature of the regular component of the model, is a simple composite of the autocorrelation and partial autocorrelation functions of the regular component.

Some key words: Identification; Partial correlation pattern; Seasonal effect.

1. Introduction

A class of linear parametric models for regular and seasonal time series has been developed (Box & Jenkins, 1970). The recommended procedure to fit a series is to identify tentatively an appropriate model, to estimate the parameters in that model, and then to examine the residuals for nonrandom behaviour, in order to improve upon the tentative identification.

The general multiplicative autoregressive integrated moving average model of order $(p,d,q)(P,D,Q)^S$ for a stochastic process Z_t is of the form

$$\Phi(B^S)\,\phi(B)\,\nabla_S^D\,\nabla^d Z_t = \Theta(B^S)\,\theta(B)\,A_t, \tag{1·1}$$

where A_t is a purely random, or white noise, process, the backward shift operator, B, has the effect $BZ_t = Z_{t-1}$ and the difference operators $\nabla = 1-B$ and $\nabla_S = (1-B^S)$ allow for homogeneous nonstationarity (Box & Jenkins, 1970, Chapter 4); Φ, ϕ, θ and Θ are polynomial shift operators.

In the identification stage, the model of the form (1·1) for which the theoretical autocorrelation function ρ_k is most similar to the sample autocorrelation function r_k is tentatively chosen to represent the series. Identification can be a perplexing task because the correlation function is susceptible to confusing distortion in the case of seasonal time series, and because the other important aid to identification, the partial autocorrelation function, has not been well understood for seasonal models. The behaviour of the correlation function of low order regular models of the form (1·1), $P = Q = 0$, is quite well known and is outlined by Box & Jenkins (1970, Chapter 4) and by Stralkowski, Wu & De Vor (1970, 1974).

A complementary tool in the identification of time series models is the partial autocorrelation function whose value at lag k, π_k, is defined to be the last coefficient in an autoregressive process of order k fitted to the autocorrelations. It can also be computed as the ratio of the determinants of two $k\times k$ matrices, $\pi_k = N_k/D_k$, where D_k is the autocorrelation matrix out to lag $k-1$ and N_k is the same as D_k with the last column replaced by $(\rho_1,\ldots,\rho_k)'$. The behaviour of the partial correlation function for low order regular models, $P = Q = 0$, is well known and is discussed by Box & Jenkins (1970, Chapter 4).

2. SEASONAL PARTIAL AUTOCORRELATION FUNCTIONS

2·1 *General remarks*

In this section we present expressions and plots of the partial autocorrelation function of low order multiplicative seasonal models, and show that it can be usefully thought of as a repetition of a combination of the auto and partial correlation functions of the regular component about the seasonal partial values. This combination we call the partial autocorrelation pattern or, simply, the partial pattern. Recognition that the complete partial correlation function consists of repeated partial patterns is then shown to be useful in identifying multiplicative models.

2·2 *The autoregressive order one partial autocorrelation pattern*

The partial correlation function of $(1,0,0)(P,0,Q)^S$ models for lags $k = 1, \ldots, S+1$ may be derived as follows. Let R_{kS} be the kSth autocorrelation of the seasonal component, and

$$G = (R_S + \phi^S R_{2S} + \phi^{2S} R_{3S} + \ldots)/(1 + \phi^S R_S + \phi^{2S} R_{2S} + \ldots); \qquad (2\cdot1)$$

then

$$\rho_j = (\phi^j + \phi^{S-j} G)/(1 + \phi^S G) \quad (j = 1, \ldots, S).$$

Substituting for ρ_j in N_k and D_k and calculating the determinants, we have

$$\pi_1 = \phi(1 + G\phi^{S-2})/(1 + G\phi^S),$$
$$\pi_k = \phi\pi_{k+1}(1 - G\phi^{2(S-k)})/(1 - G\phi^{2(S-k+1)}) \quad (k = 2, \ldots, S-1) \qquad (2\cdot2)$$
$$\pi_S = G(1-\phi^2)/(1-\phi^2 G^2), \quad \pi_{S+1} = -\phi G\{R_S - G + G\phi^2(1 - R_S G)\}/\{G\phi^2(1-G^2)\}.$$

The proof of these results is a straightforward extension of the method used by D. C. Hamilton in his unpublished Master's Thesis, Queen's University, for the $(1,0,0)(1,0,0)^S$ and $(1,0,0)(0,0,1)^S$ models, which involves calculation of the two determinants N_k and D_k, exploiting the patterned structure of the matrices.

Equations (2·2) show that, for G and ϕ not close to unity and for S greater than about 4, so that $\phi^S \ll 1$,

$$\pi_1 \simeq \phi, \quad \pi_k \simeq \phi\pi_{k+1} \quad (k = 2, \ldots, S-1), \quad \pi_S \simeq G, \quad \pi_{S+1} \simeq -\phi G. \qquad (2\cdot3)$$

The pattern implied by these expressions and confirmed in Fig. 1(b) is an example of what we call a partial autocorrelation pattern. The pattern consists of autocorrelation behaviour to the left of lag zero and inverted partial correlation behaviour to the right. To the accuracy

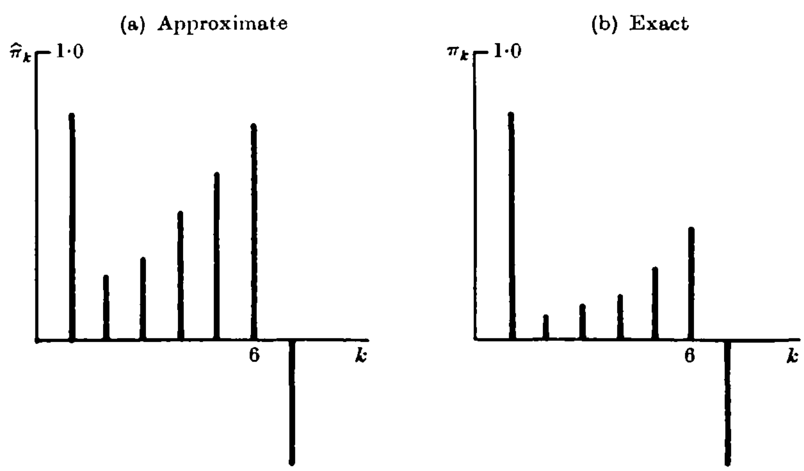

Fig. 1. Approximate and exact partial correlation functions of the $(1,0,0)(1,0,0)^6$ model with $\phi = 0\cdot7$ and $\Phi = 0\cdot6$.

required in identification, it is possible to approximate the partial correlation function by multiplying the partial pattern by the value of the partial correlation function of the seasonal component and adding the result to the partial correlation function of the regular component; see Fig. 1(a). Note that the scaled partial pattern in Fig. 1(b) consists of the values at lags $2, \ldots, 7$, since $S = 6$ in this example.

For lags greater than $S+1$, a detailed numerical study was carried out on models of the form $(1,0,0)(P,0,Q)^S$ with $P \leqslant 2$, $Q \leqslant 2$ to compare the exact partial correlation values around lags $2S$ and $3S$ with those around lag S, since we were unable to obtain analytic expressions for $k > S+1$. These calculations showed that the characteristic partial pattern is scaled and repeated about each nonzero seasonal partial correlation value. In particular, the fractional deviation of subsequent patterns from the first, that is for $j = 2, 3, \ldots$

$$(\pi_{jS,l}/\pi_{jS} - \pi_{S,l}/\pi_S)\,(\pi_{S,l}/\pi_S)^{-1},$$

where $\pi_{jS,l}$ is the lth value in the jth pattern, was seldom greater than 0·15. These results demonstrate that to the degree of accuracy needed for identification purposes, the appearance of each pattern is the same. The partial correlation function for the $(1,0,0)(0,0,1)^6$ model shown in Fig. 2 illustrates this repetition of the pattern.

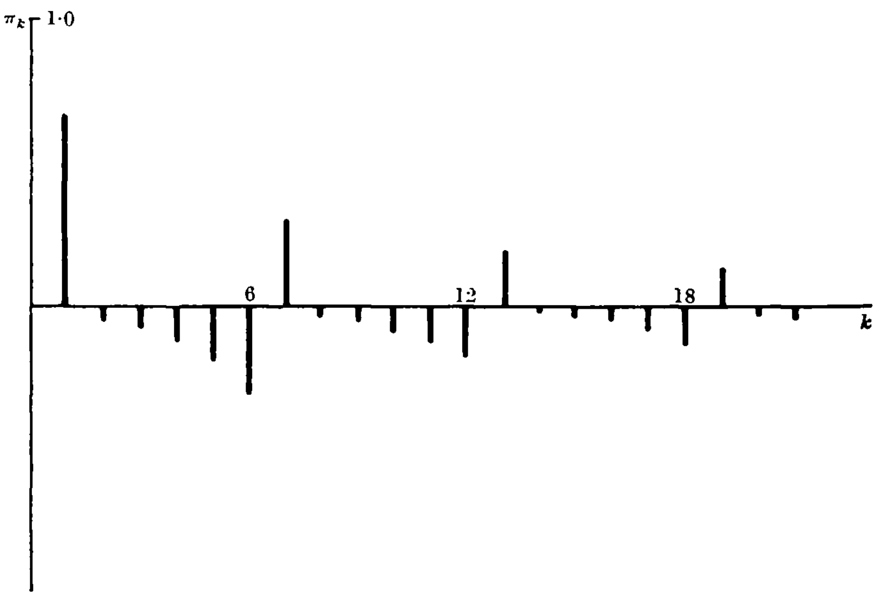

Fig. 2. Partial correlation function of the $(1, 0, 0)$ $(0, 0, 1)^6$ model with $\phi = 0{\cdot}7$ and $\Theta = 0{\cdot}8$.

2·3 *The moving average order one partial autocorrelation pattern*

As a second illustration of a characteristic pattern in the partial correlation function consider a regular moving average component. Equations for the first $2S-2$ values in the partial correlation function of $(0,0,1)(P,0,Q)^S$ models are given below. Let R_S be the autocorrelation at lag S of the seasonal component; then

$$\pi_k = -\theta^k(1-\theta^2)/(1-\theta^{2(k+1)}) \quad (k = 1, \ldots, S-2),$$

$$\pi_{S-1} = -\theta^{S-1}(1-\theta^2)(1-\theta^{2S})^{-1} - \theta(1+\theta^2)^{-1} R_S(1-\theta^{2(S-1)})(1+\theta^2)(1-\theta^{2S})^{-1},$$

$$\pi_S = \frac{R_S\{(1-\theta^2)(1+\theta^{2S}) - R_S\,\theta^S(1-\theta^2)\} - \theta^S(1-\theta^2)}{(1-\theta^{2S+2}) - 2R_S\,\theta^{2S}(1-\theta^2) - R_S^2(\theta^2-\theta^{2S})}, \tag{2·4}$$

$$\pi_{S+k+1} = \theta\pi_{S+k}\{1-(\theta^S - R_S)^2(1-R_S\,\theta^S)^{-2}\theta^{2k+2}\}$$
$$\times\{1-(\theta^S - R_S)^2(1-R_S\,\theta^S)^{-2}\theta^{2k+4}\}^{-1} \quad (k = 0, \ldots, S-3).$$

These results are obtained by inverting the autocorrelation matrix to solve the Yule–Walker equations (Box & Jenkins, 1970, p. 55) using a result of Shaman (1969).

Equations (2·4) show that for lags 1 to $S-2$, π_k is the same as the partial correlation function of the regular component, i.e. there is approximate first-order decay to the right of lag 1, and from lag $S+1$ to $2S-2$ there is also approximate first-order decay. For θ not too large and S not too small, so that $\theta^S \ll 1$, equations (2·4) simplify so that not only is π_k $(k = 1, \ldots, S-2)$ the same as for the regular component but also

$$\pi_{S-1} \simeq -\theta R_S, \quad \pi_S \simeq R_S(1-\theta^2)/(1-R_S^2\theta^2), \quad \pi_{S+k+1} \simeq \theta\pi_{S+k} \quad (k = 0, \ldots, S-3). \qquad (2·5)$$

Once again the analytic expressions reveal a pattern of inverted regular partial correlation behaviour to the right and regular autocorrelation behaviour to the left of the seasonal partial values.

As for the autoregressive process of order 1, computer calculations of partial correlation functions for $(0,0,1)(P,0,Q)^S$ models with P and $Q \leqslant 2$ have shown that to the degree of accuracy necessary for identification purposes, the same pattern is scaled and repeated about each seasonal value.

2·4 *The autoregressive-moving average (p,q) partial autocorrelation pattern*

For mixed and higher order models, useful analytic expressions for the partial correlation function have not been found. However, exact partials have been calculated for a wide range of $(p,0,q)(P,0,Q)^S$ models with p, P, q and $Q \leqslant 2$ and $S \leqslant 12$. As before, numerical comparisons were made between partial autocorrelations about different seasonal lags, and again a strong pattern was revealed.

These calculations and analytic expressions for $p = 1$ and $q = 1$ showed that:

(i) the values at seasonal lags are similar to the values in the partial correlation function of the purely seasonal component,

(ii) to the right of each seasonal value, behaviour like that of the inverted partial correlation function of the regular component appears,

(iii) to the left of each seasonal value, behaviour characteristic of the autocorrelation function of the regular component appears,

(iv) the same pattern is scaled about each value in the partial correlation function of the seasonal component whether the seasonal component is autoregressive, moving average, or mixed.

3. AN EXAMPLE

We illustrate the use of the partial correlation function in identification of a model for seasonal data, by considering a series relating to caffeine levels in instant coffee. The data, given in Table 1, come from a cyclic process with period five units and thus we expect some seasonal effect. A plot of the 178 data points shows a homogeneous series that does not require a variance stabilizing transformation. There is, however, a slight negative drift in the series, warning that differencing may be required.

Sample auto and partial correlation functions are shown in Fig. 3 for the original and differenced series with dashed horizontal lines indicating approximate 95% probability limits for a purely random process. The autocorrelation function of the original data appears to indicate a first-order autoregressive regular component, and seasonality is evidenced by the oscillating nature of the decay. Several different seasonal components could explain this feature, but since the apparent periodicity of the peaks and troughs is not five as we expect, a good identification is not possible. Clear patterns appear in the partial correlation function,

Table 1. *Caffeine levels in instant coffee*

0·429	0·443	0·451	0·455	0·440	0·433	0·423	0·412	0·411	0·426
0·436	0·441	0·446	0·443	0·437	0·426	0·427	0·431	0·436	0·432
0·433	0·424	0·420	0·416	0·405	0·408	0·414	0·417	0·400	0·404
0·396	0·393	0·389	0·409	0·409	0·413	0·403	0·402	0·389	0·392
0·383	0·389	0·386	0·395	0·400	0·412	0·411	0·414	0·412	0·404
0·401	0·406	0·407	0·410	0·410	0·409	0·407	0·407	0·402	0·409
0·403	0·398	0·397	0·396	0·392	0·391	0·395	0·391	0·387	0·389
0·399	0·396	0·392	0·391	0·388	0·395	0·405	0·414	0·425	0·433
0·423	0·417	0·432	0·430	0·418	0·420	0·418	0·404	0·403	0·418
0·419	0·417	0·418	0·420	0·417	0·421	0·422	0·428	0·429	0·428
0·421	0·422	0·412	0·405	0·405	0·411	0·407	0·411	0·413	0·409
0·405	0·408	0·402	0·398	0·396	0·392	0·391	0·399	0·407	0·406
0·393	0·387	0·386	0·383	0·389	0·393	0·396	0·387	0·380	0·359
0·361	0·375	0·399	0·406	0·417	0·421	0·407	0·384	0·393	0·410
0·409	0·413	0·410	0·398	0·382	0·381	0·376	0·389	0·395	0·397
0·401	0·404	0·401	0·393	0·401	0·399	0·400	0·402	0·399	0·397
0·398	0·399	0·388	0·389	0·379	0·384	0·394	0·402	0·395	0·398
0·391	0·375	0·375	0·393	0·394	0·389	0·391	0·385		

however, which better reveal the seasonality and the model. The large values at lags 1 and 2 are repeated at 6 and 7 and again, partially, at 11. This repetition is characteristic of the autoregressive order 2 partial pattern and also indicates the existence of strong seasonality of period 5. Furthermore, by the decay of the values at lags 1, 6 and 11 a firm identification of the seasonal component of $(0, 0, 1)^5$ can be made.

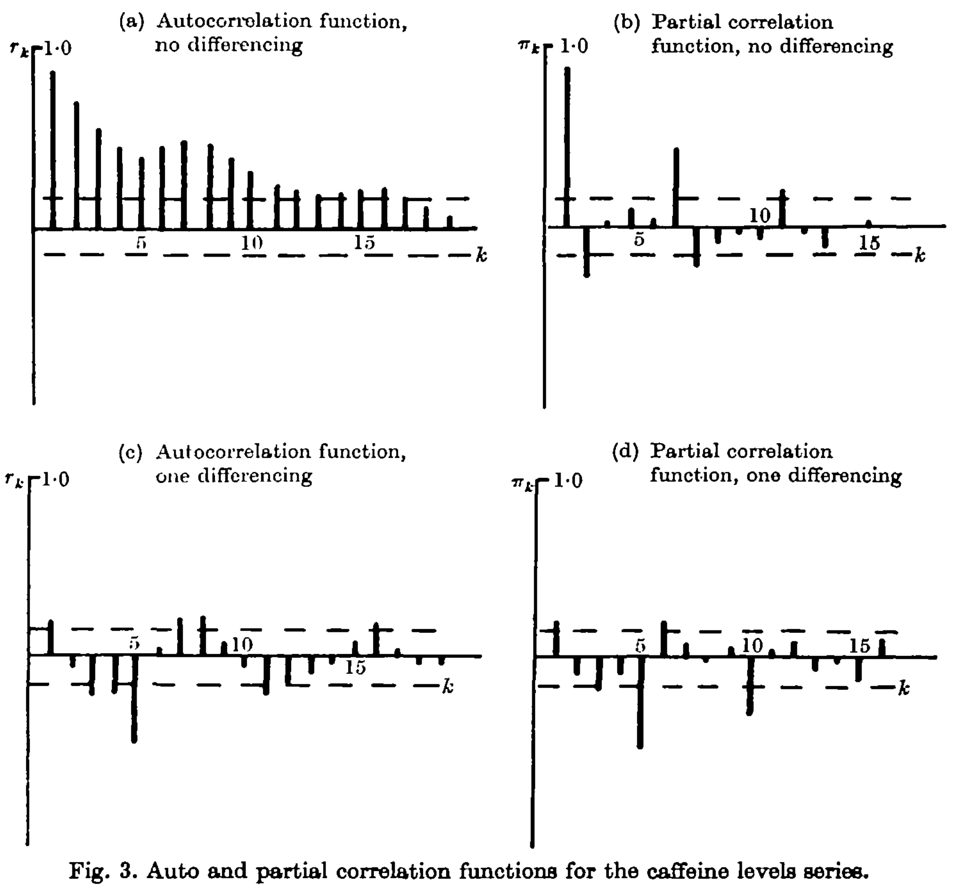

Fig. 3. Auto and partial correlation functions for the caffeine levels series.

After one differencing, identification from the autocorrelation function remains difficult. There is a large value at lag 5, but the true seasonality is obscured by persistent significant values at lags 3, 4, 7, 8, 11 and 16. The regular component could be either a weak first-order autoregressive or a strong autoregressive second-order process which combines with the seasonal component to cause the persistently large values.

In the partial correlation function of the differenced series, the first-order moving average seasonal identification is strongly confirmed by the pattern of values at lags 5, 10 and 15. The regular component is first-order autoregressive as revealed by the large values at lags 1 and 6.

Note that identification via the partial correlations is more consistent through the stages, and correctly reveals the form of the seasonal component. This is in contrast with the autocorrelations in which a strong seasonal component was hidden for the original data and confused for the differenced data.

The $(1,1,0)(0,0,1)^5$ model gave an excellent fit, with estimated residual variance $\hat{\sigma}^2 = 0{\cdot}3275 \times 10^{-4}$ and parameter values $\hat{\theta} = 0{\cdot}31$, $\hat{\Theta} = 0{\cdot}82$ and $\hat{\theta}_0 = -0{\cdot}21 \times 10^{-3}$. The constant term, θ_0, was included to remove deterministic trend and is significant at the 0·05 level. An examination of the residual autocorrelations failed to suggest any improvement to this model.

The authors acknowledge the financial support provided by the National Research Council of Canada. Some of the research was completed when the first author was employed at Statistics Canada. The authors are grateful to Mavis Carroll, General Foods Corporation, Tarrytown, N.Y., for providing data.

REFERENCES

BOX, G. E. P. & JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day.

SHAMAN, P. (1969). On the inverse of the correlation matrix of a first order moving average. *Biometrika* **56**, 595–600.

STRALKOWSKI, C. M., WU, S. M. & DE VOR, R. E. (1970). Charts for the interpretation and estimation of the second order autoregressive model. *Technometrics* **12**, 669–85.

STRALKOWSKI, C. M., WU, S. M. & DE VOR, R. E. (1974). Charts for the interpretation and estimation of the second order moving average and mixed first order autoregressive-moving average models. *Technometrics* **16**, 275–85.

[*Received January* 1977. *Revised October* 1977]