Tomáš Cipra

Note on inappropriate trend and seasonal elimination

Commentationes Mathematicae Universitatis Carolinae, Vol. 27 (1986), No. 2, 277--283

Persistent URL: http://dml.cz/dmlcz/106450

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE
27,2 (1986)

NOTE ON INAPPROPRIATE TREND AND SEASONAL ELIMINATION

Tomáš CIPRA

Abatract: Behaviour of time series is investigated in two simple cases when one has used an inappropriate method for elimination of trend or seasonality: (i) differencing in a simple polynomial model and (ii) regression in a simple seasonal model of Box and Jenkins. It can imply distorting consequences for the further analysis of the time series.

Key words: Differencing, regression, seasonal elimination, seasonal model, time series, trend elimination.

Classification: 62M10, 62J05

---

1. Introduction. Elimination of trend or seasonality is one of the usual initial steps in time series analysis. Its aim is (if it is possible) to transfer the analysed series to a stationary form for which a corresponding stationary model can be built. In this brief note we demonstrate for two simple situations that one must be very cautious when choosing a suitable elimination method since an inappropriate procedure can lead to serious distortions from the point of view of the further analysis.

In Section 2 the situation is considered when the differences of the order d are applied to the usual polynomial model of the order d in order to detrend it. We remind in this section the well-known result (which can be dated to the times of Slutzky) that the mentioned differencing creates spurious autocorrelations to the order d in the detrended series and that this series tends to be dominated by high frequency periodicities (see e.g. [1] for theoretical results or [3] for a more heuristic approach when d = = 1).

Section 2 demonstrates the case when differencing is used inappropriately instead of regression. In Section 3 the opposite case is considered: regression is shown to be the inappropriate elimination technique for seasonal models of Box and Jenkins [2] instead of seasonal differencing since the behaviour of the residuals remains strongly periodical. This conclusion can be compared with the one obtained in [3] and [4] for the case when the linear regression is applied for a random walk (i.e. for the simplest Box-Jenkins model of the type ARIMA).

2. Differencing in simple polynomial model. Let us consider the following polynomial model of the order d

$$x_t = \beta_0 + \beta_1 t + \ldots + \beta_d t^d + \varepsilon_t, \tag{2.1}$$

where $\beta_0, \beta_1, \ldots, \beta_d$ are parameters and ε_t's are uncorrelated random variables with zero mean values and variances σ^2. If the differences of a time series $\{y_t\}$ are defined recursively as

$$\Delta^k y_t = \Delta(\Delta^{k-1} y_t), \quad k \geq 2, \tag{2.2}$$

where

$$\Delta y_t = \Delta^1 y_t = y_t - y_{t-1} \tag{2.3}$$

then it holds for (2.1)

$$\Delta^d x_t = d!\, \beta_d + \sum_{j=0}^{d} (-1)^j \binom{d}{j} \varepsilon_{t-j}. \tag{2.4}$$

Obviously the detrended series $\{\Delta^d x_t\}$ is (weakly) stationary with the constant mean value $d!\, \beta_d$. The explicit formula for its autocorrelation function $\rho_k = \mathrm{corr}(\Delta^d x_t, \Delta^d x_{t-k})$ is given in the following

Theorem 1. Under the previous assumptions the autocorrelation function ρ_k of $\{\Delta^d x_t\}$ has the form

$$(2.5)\quad \rho_k = (-1)^k \binom{2d}{d-k} \Big/ \binom{2d}{d}, \quad 0 \leq k \leq d,$$
$$= 0 \qquad\qquad , \quad k > d.$$

Proof. See [1, Section 3.4.3].

In other words, the use of the differences of the order d to eliminate the polynomial trend in (2.1) creates spurious autocorrelations for the lags $k = 1,\dots,d$. It is even

$$(2.6)\quad \rho_1 = - \frac{d}{d+1}$$

so that there is a distinctive negative first autocorrelation for larger d. Especially, it is $\rho_1 = -0.5$ for $d = 1$ (see also [3]).

Theorem 2. Under the previous assumptions the spectral density $f(\lambda)$ of $\{\Delta^d x_t\}$ has the form

$$(2.7)\quad f(\lambda) = \frac{\sigma^2}{2\pi}\, 2^d (1 - \cos\lambda)^d .$$

Proof. See [1, Section 7.5.5].

The spectral density (2.7) increases in the interval $0 \leq \lambda \leq \pi$ (it is $f(0) = 0$ and $f(\pi) = (\sigma^2/2\pi)4^d$). Therefore the high frequency periodicities will prevail in the frequency spectrum of the detrended series after differencing.

3. Regression in simple seasonal model. In this section let us consider the following simple case of Box-Jenkins seasonal models

$$(3.1)\quad x_t - x_{t-L} = \varepsilon_t, \quad t = 1,\dots,n,$$

where ε_t's have the same form as in Section 2 and L is the length of the season. One can also write

$$\Delta_L x_t = \varepsilon_t, \ t = 1,\dots,n,$$

where $\Delta_L = 1-B^L$ is so called seasonal operator (see [2]). In order to determine fully the model let us initialize it prescribing

(3.2) $x_t = \alpha \cos(\omega_L t) + \beta \sin(\omega_L t), \ t = 1-L, 2-L,\dots,0,$

where

(3.3) $$\omega_L = \frac{2\pi}{L}$$

and α, β are parameters.

Remark. The situation is analogous to that with the random walk (which presents the simplest ARIMA model)

$$\Delta x_t = \varepsilon_t, \ t = 1,\dots,n,$$

where one can set

$$x_{-1} = 0$$

(see [3],[4]). Although the conditions (3.2) represent a little bit special choice of the initial values $x_{1-L},\dots,x_0$ they simplify the following derivation and are sufficient for our demonstration (it would be possible to work with L-1 seasonal dummy variables in a more general case).

Let the regression technique be chosen to eliminate seasonality from the observations $x_1,\dots,x_n$ so that the deseasonalized time series has the form of the OLS residuals estimated from the model

(3.4) $x_t = a \cos(\omega_L t) + b \sin(\omega_L t) + e_t, \ t = 1,\dots,n,$

where a, b are parameters and e_t's are uncorrelated random variables with zero mean values and constant variances. If we assume in addition that n is a multiple of L and that $L > 2$ then the mentioned OLS residuals have the form

(3.5) $\hat{e}_t = x_t - \hat{a} \cos(\omega_L t) - \hat{b} \sin(\omega_L t), \ t = 1,\dots,n,$

where

$$(3.6)\quad \hat{a} = \frac{2}{n}\sum_{t=1}^{n} x_t \cos(\omega_L t),\ \hat{b} = \frac{2}{n}\sum_{t=1}^{n} x_t \sin(\omega_L t).$$

The following denotation will be used in the further text: for an arbitrary integer k let $u_L(k)$ and $v_L(k)$ be the integers determined unambiguously by the relation

$$(3.7)\quad k = u_L(k) + v_L(k)\, L,\ 1 \le u_L(k) \le L.$$

Then one can formulate the following

Theorem 3. Under the previous assumptions it holds

$$(3.8)\qquad E\,\hat{e}_t = 0$$

and for $s \ge 0$

$$(3.9)\quad \operatorname{cov}(\hat{e}_t,\hat{e}_{t-s}) = \frac{\sigma^2}{n}\cos(\omega_L s)\Big\{\frac{L}{3n}[2v_L^3(n) + 9v_L^2(n) + 13v_L(n) + 6]$$
$$- 2[v_L(t)+1]\,[v_L(n-t+1) + v_L(t)/2 + 1]$$
$$- 2[v_L(t-s) + 1]\,[v_L(n-t+s+1) + v_L(t-s)/2+1]\Big\}.$$

If t-s is a multiple of L then the additive term $\sigma^2[v_L(t-s)+1]$ must be added to the righthandside of (3.9).

Proof. The proof makes use of the formulas

$$(3.10)\quad \sum_{t=1}^{n} \sin^2(\omega_L t) = \sum_{t=1}^{n} \cos^2(\omega_L t) = \frac{n}{2},$$
$$\sum_{t=1}^{n} \sin(\omega_L t)\cos(\omega_L t) = 0.$$

Due to (3.1) and (3.2) it is

$$(3.11)\quad x_t = \varepsilon_t + \varepsilon_{t-L} + \ldots + \varepsilon_{u_L(t)} + \alpha\cos(\omega_L t) + \beta\sin(\omega_L t).$$

If substituting (3.11) to (3.6) one can verify easily that

$$E\,\hat{a} = \alpha\ ,\ E\,\hat{b} = \beta$$

so that (3.8) is fulfilled. Further one can write

$$\operatorname{cov}(\hat{e}_t,\hat{e}_{t-s}) = E\{\varepsilon_t + \varepsilon_{t-L} + \ldots + \varepsilon_{u_L(t)} + \alpha\cos(\omega_L t) + \beta\sin(\omega_L t)$$
$$- \frac{2}{n}\cos(\omega_L t)\sum_{i=1}^{n}[\varepsilon_i + \varepsilon_{i-L} + \ldots + \varepsilon_{u_L(i)} + \alpha\cos(\omega_L i) +$$
$$+ \beta\sin(\omega_L i)]\cos(\omega_L i) - \frac{2}{n}\sin(\omega_L t)\sum_{i=1}^{n}[\varepsilon_i + \varepsilon_{i-L} + \ldots$$

$$\ldots + \varepsilon_{u_L(i)} + \alpha \cos(\omega_L i) + \beta \sin(\omega_L i)] \sin(\omega_L i)\}$$

$$\{\varepsilon_{t-s} + \varepsilon_{t-s-L} + \ldots + \varepsilon_{u_L(t-s)} + \alpha \cos[\omega_L(t-s)] + \beta \sin[\omega_L(t-s)]$$

$$- \frac{2}{n} \cos[\omega_L(t-s)] \sum_{i=1}^{n} [\varepsilon_i + \varepsilon_{i-L} + \ldots + \varepsilon_{u_L(i)} + \alpha \cos(\omega_L i) +$$

$$+ \beta \sin(\omega_L i)] \cos(\omega_L i) - \frac{2}{n} \sin[\omega_L(t-s)] \sum_{i=1}^{n} [\varepsilon_i + \varepsilon_{i-L} + \ldots$$

$$\ldots + \varepsilon_{u_L(i)} + \alpha \cos(\omega_L i) + \beta \sin(\omega_L i)] \sin(\omega_L i)\}.$$

After some tedious algebraic manipulations one obtains (3.9) (it is e.g.

$$E(\varepsilon_t + \varepsilon_{t-L} + \ldots + \varepsilon_{u_L(t)})(\varepsilon_{t-s} + \varepsilon_{t-s-L} + \ldots + \varepsilon_{u_L(t-s)}) =$$

$$= 0 \qquad , \qquad \text{if } u_L(t-s) \neq L,$$

$$= \sigma^2 [v_L(t-s) + 1], \qquad \text{if } u_L(t-s) = L;$$

similarly

$$E [\sum_{i=1}^{n} (\varepsilon_i + \varepsilon_{i-L} + \ldots + \varepsilon_{u_L(i)}) \cos(\omega_L i)]^2 =$$

$$= E [\sum_{i=1}^{n} (\varepsilon_i + \varepsilon_{i-L} + \ldots + \varepsilon_{u_L(i)}) \sin(\omega_L i)]^2 =$$

$$= \frac{\sigma^2 L}{12} [2v_L^3(n) + 9v_L^2(n) + 13v_L(n) + 6]).$$

According to (3.9) the autocovariances of the deseasonalized series $\{\hat{e}_t\}$ depend on t so that this series is not stationary. Moreover, for t fixed the autocovariances $\mathrm{cov}(\hat{e}_t, \hat{e}_{t-s})$ can show a distinctive seasonality in s (if the values of t-s are multiples of L then there are even discontinuous jumps in the behaviour of these autocovariances). One can conclude that the behaviour of the residuals $\{\hat{e}_t\}$ remains periodical (e.g. from the point of view of Box-Jenkins identification procedure) so that the regression is the inappropriate technique in this case.

References

[1] ANDERSON, T.W.: The Statistical Analysis of Time Series, Wiley,

New York 1971.

[2] BOX, G.E.P., JENKINS G.M.: Time Series Analysis, Forecasting and Control, Holden Day, San Francisco 1970.

[3] CHAN K.H., HAYYA J.C., ORD J.K.: A note on trend removal methods: the case of polynomial regression versus variate differencing, Econometrica 45(1977), 737-744.

[4] NELSON Ch.R., KANG H.: Spurious periodicity in inappropriately detrended time series, Econometrica 49(1981), 741-751.

Dept. of Statistics, Charles University, Sokolovská 83, 186 00 Praha 8, Czechoslovakia

(Oblatum 9.10. 1985)