A Non-linear Analysis of Seasonal Variation

Michael D. Godfrey, Princeton University

Reset in TEX, September, 2004. Michael D. Godfrey

ISL, Stanford University

Note: The first reference in the paper is given as a Research Memorandum. This was subsequently published as:

(1966) Godfrey, M. D. and H. Karreman, "A Spectrum Analysis of Seasonal Adjustment," in *Essays in Honor of Oskar Morgenstern*, ed. M. Shubik, Princeton University Press, pp. 367-421.

21 September 2004

Reprinted from 1964 Proc. ASA, Business and Economic Statistics Sect.

A Non-linear Analysis of Seasonal Variation

Michael D. Godfrey, Princeton University

1. Introduction

It has been customary to accept as the basis for a seasonal adjustment procedure a model for the seasonal component of a time series which has the form:

$$S_t = \sum_{j=1}^{n} \alpha_j e^{i\omega_j t}, \tag{1.1}$$

where

S_t = seasonal component at time t

α_j = complex-valued coefficients

ω_j = seasonal frequencies such that $\omega_j = j\omega_1$

$\omega_1 = 1$ cycle per year.

n = a constant, such that there are $2n$ equally spaced observations per year.*

This is the usual finite Fourier series decomposition of an arbitrary shaped periodic function of period

$$T = \frac{1}{\omega_1}.$$

The model for the observed time series is then given as:

$$X_t = S_t + Z_t, \tag{1.2}$$

where

X_t = observed time series

Z_t = a periodic stochastic time series.

It frequently appears to be appropriate to consider the logarithms of the original variables when postulating an additive model of this form. However, we will not discuss this issue here.

The estimation problem for this process is that of estimating S_t from a realization of X_t. The estimation of S_t permits the computation of Z_t, which is generally supposed to reveal certain information which is not evident in X_t. For the case where Z_t is a wide sense stationary process and the $\alpha_j's$ of equation (1.1) are constants independent of time, the estimation problem seems to be completely solved (Hannan (1963), Jorgenson (1964)).

However, statisticians (Wald (1936), Shiskin (1960)) have long recognized that observed "seasonal" patterns in economic time series do not conform to the fixed, strictly periodic, form given by equation (1.1). This observation has led to attempts to modify this basic model. These attempts have followed two principal lines. Both lines involve making the $\alpha_j's$ of equation (1.1) functions which change with time.

The first line of development has resulted in the postulate that each α_j is a function of time only. Estimation requirements obviously demand that each α_j be a slowly varying function time when compared to $cos(\omega_1)$. The estimation procedures for this model simply involve estimating the $\alpha_j's$ as constants over short successive time intervals and then constructing the time series $\alpha_j(t)$ from these estimates. The pioneering practical work of Shiskin (1960) and his co-workers is well known in this area as is the analytic work of Hannan (1964).

An interesting point to observe about this method is that it nearly preserves the basic linearity of the process which is implied in equation (1.1). In the case where the $\alpha_j's$ are allowed to vary slowly with time, but independently of each other, we are assuming that the seasonal is still generated by a linear, but mildly non-stationary process.

The other line of development has involved making the $\alpha_j's$ both functions of time and of each other. Only the average values of the $\alpha_j's$ over the realization are estimated using equation (1.1). In the procedure as originally developed by A. Wald (1938) the average values were constants, however (Godfrey and Karreman (1964)) the average values may also be estimated as independent functions of time in order to deal with certain kinds of non-stationarity. After the average values of the $\alpha_j's$ have been estimated it is assumed that over relatively short time-intervals the $\alpha_j's$ vary proportionately. Thus the ratios of the $\alpha_j's$ remain nearly constant, but the $\alpha_j's$ vary to take account of the changing amplitude of the seasonal variation.

This second line of development takes account of a certain kind of non-linearity by removing the assumption of independence when estimating the variation of the $\alpha_j's$. However, the method still starts with the basic linear model in order to estimate the mean values of the $\alpha_j's$, and then it imposes a rather special character on the variation of the $\alpha_j's$.

The seasonal adjustment method to be proposed here

* $2n$ is used because the highest frequency which can be discriminated given an equispaced record at time points $t = 1, 2, \ldots, m$ is $\omega_n = \pi$.

Reprinted from 1964 Proc. ASA, Business and Economic Statistics Sect.

is developed on the basis of a rather more general model concerning the process by which the seasonal component is generated. The model will be shown to include the two models indicated above as special cases.

2. The Model of Seasonal Variation

The basic assumption of the procedure to be described below is that the seasonal variation is produced by a few unobservable factors. These factors may be thought of as time series which determine the time paths of the time-dependent $\alpha_j{}'s$ of equation (1.1). Thus the estimation problem is that of estimating these unobservable factors. The estimation of the factors may follow a time series generalization of the classical principal factor method of factor analysis. The number of (orthogonal) factors required will depend on the explanatory power of each additional factor. Clearly, the power of the method will depend on how many factors are required to explain the seasonal variation. If only one factor is required, then the method will form a a simple generalization of Wald's method (Wald (1936)). If n factors are required, then the method will be very similar to Hannan's method (Hannan (1963)) as the $\alpha_j{}'s$ will be independently estimated.

We may now proceed to the formal description of the model and the technique of its estimation. We assume the usual additive model for the observed time series

$$X_t = S_t + Z_t \tag{2.1}$$

where

$X_t =$ observed time series

$S_t =$ seasonal variation

$Z_t =$ time series with continuous spectrum whose density is smooth in the neighborhood of season frequencies.

Thinking in terms of the frequency decomposition of S_t, we nay represent S_t by:

$$S_t = \sum_{i=1}^{n} x_{i,t} e^{i\omega_i t} \tag{2.2}$$

where the complex-valued time series $x_{i,t}$ each have spectral power only in the neighborhood of zero frequency. We now postulate that the $x_{i,t}$ are not independent series, but rather they are generated by a smaller set of p $(p < n)$ series. Therefore, we have:

$$x_{i,t} = \sum_{k=1}^{p} \sum_{j=0}^{m} A_{k,i}(j) y_{k,t-j} \tag{2.3}$$

where

$y_{k,t-j} =$ form, for $k = 1, \ldots, p$ a set of time series,

$m =$ maximum number of lags used,

$A_{k,i}(j) =$ matrix of coefficients.

Each series $x_{i,t}$ is generated by a linear combination of past values of some or all of the $y_{k,t}$.

Instead of estimating the $x_{i,t}$ as if they were independent, it would appear reasonable to attempt to estimate the smaller set of variables $y_{k,t}$. We may therefore write equation (2.3) as

$$x_i = \sum_{k=1}^{p} \sum_{j=0}^{m} A_{k,i} z^j y_k \tag{2.4}$$

using the z-transform operator defined by:

$$x = \sum_{t=0}^{\infty} x_t z^t.$$

We then represent the quantities $\sum_{j=0}^{m} A_{k,i} z^j$ by $B_{k,i}$ and solve the equation (2.4) for the $y_k{}'s$. Thus we have

$$y_k = \sum_{i=1}^{p} C_{k,i} x_i \tag{2.5}$$

where

$$C_{k,i} = B_{k,i}^{-1}$$

We may then consider the estimation of y_k by first computing $\hat{y}_{k,t}$ by:

$$\hat{y}_{k,t} = \sum_{i=1}^{p} \sum_{j=0}^{m} \hat{C}_{k,i}(j) \hat{x}_{i,t-j} + \epsilon_{i,t}(k) \tag{2.6}$$

where: $\epsilon_{i,t}(k) =$ independent random numbers.

We then estimate $y_{i,t}$ by

$$\tilde{y}_{i,t} = D(\hat{y}_{i,t}) \tag{2.7}$$

where: $D(\ )$ is a low pass filter.

While equation (2.3) states that the n series of seasonal coefficients $x_{i,t}$ are determined by a set of p time series $y_{i,t}$, we will want to estimate the $y_{i,t}$ from equations (2.6 and 2.7). We will then use an equation of the form of equation (2.3) to get estimates of the $x_{i,t}$. This equation may be stated as:

$$\tilde{x}_{i,t} = \sum_{k=1}^{p} \sum_{j=0}^{m} \hat{A}_{k,i}(j) \tilde{y}_{k,t-j} \tag{2.8}$$

We will now describe the estimation procedure in somewhat greater detail.

Reprinted from 1964 Proc. ASA, Business and Economic Statistics Sect.

3. Estimation of the Model

The estimation procedure to be presented here is one which seams natural, but it is only a first attempt at estimation and by no means represents a complete solution of the estimation problem. Further work, both theoretical and practical, is required in this general area.

The first step in the estimation procedure is to compute the time series $\hat{x}_{i,t}$ by complex demodulation. The $\hat{x}_{i,t}$ are thus given by:

$$\hat{x}_{i,t} = L(x_t e^{i\omega_i t}) \tag{3.1}$$

where

$L(\ )$ - denotes a low-pass filter with bandwidth of about ω_1.

The $\hat{x}_{i,t}$ form estimates of the complex-valued Fourier coefficients at frequency ω_i. In effect, each term in the series $\hat{x}_{i,t}$ is an estimate over the time interval $t - m_1$ to $t - m_2$ where $m = m_1 + m_2$, and m is the number of terms in the low-pass filter.

These $\hat{x}_{i,t}$ may be considered as linear estimates of the seasonal variation. If we set m equal to the number of observations (and $m_1 = m_2 = m/2$) we then have a single (complex-valued) number for each series $\hat{x}_{i,t}$. These are clearly estimates of the Fourier coefficients in the stationary linear model. Letting m be less than n gives us estimates of time varying Fourier coefficients which are similar to what the Census Method [Shiskin (1960) and see also the relevant analysis of Rosenblatt (1963)] and the BLS method produce. As m is decreased, for a given n, the variability of $\hat{x}_{i,t}$ increases. Depending on one's assumptions about the variability of the amplitudes of the seasonal there will be some m which minimizes the expected error of the estimate over the realization of n observations.

In forming estimates of time varying coefficients it has been general practice to make $m_1 = m_2$. This implies an indifference toward the direction of time. For the purpose of forecasting future values of Z_t it would seem more natural to set $m_2 \leq 0$. In this case we are dealing with a linear prediction problem and the "best" operator may be the Wiener optimum predicting filter.

However, for our present purposes the appropriate filter is simply one which removes the frequency terms above $\cos(\omega_1 t)$ which will appear in the $x_{i,t}e^{i\omega_i t}$ series before filtering. For this purpose a low order (say 8 or 10 term) weighted moving average should be adequate.

Given these complex-valued series we now wish to estimate the transformation which will give the p series which explain most of the variation in the n series. A natural way to do this is to estimate the principal factors at a set of modulating frequencies. We may then combine these factors in order to form the series $\hat{y}_{i,t}$ as given in equation (2.6).

The modulating frequencies which should be chosen for this estimation are naturally $\lambda = 0, \pi/kn, 2\pi/kn, \ldots, \pi/n$ where k may equal, say, 2 or 3.

The number of modulating frequencies which we can hope to estimate is severely limited by the amount of data available. The frequency π/kn has a period of $2k$ years. Thus, for $k = 3$ we are attempting to estimate a modulating frequency with a period of 6 years. In order to estimate the principal factors we band-pass filter the series $\hat{x}_{i,t}$ around λ_j to obtain a series $\hat{x}_{i,t}(\lambda_j)$. We then compute the covariance matrix:

$$R(\lambda_j) = \begin{pmatrix} r_{1,1}(\lambda_j) & \ldots & r_{1,n}(\lambda_j) \\ & \ddots & \vdots \\ & & r_{n,n}(\lambda_j) \end{pmatrix}$$

where:

$$r_{i,q}(\lambda_j) = \frac{1}{N}\sum_{t=1}^{N} x_{i,t}(\lambda_j)x^*_{q,t}(\lambda_j) \tag{3.2}$$

The eigenvectors associated with the p largest eigenvalues of $R(\lambda_j)$ define the linear combination of the p series $y_{i,t}$ which best explain the variation of the $\hat{x}_{i,t}$ at the frequency λ_j. We therefore have a $p \times n$ matrix, $F(\lambda_j)$, at each frequency λ_j.

From these matrices we can compute the coefficients $\hat{C}_{k,i}(j)$ of equation (2.6). These coefficients will then give the linear combination of the $\hat{x}'_{i.t}s$ which explain most of the variance of the factors $\hat{y}_{i,t}$.

We can now derive estimates of the variation in the seasonal factors, $\tilde{y}_{i,t}$, on the basis of assumptions concerning the model which generates these factors. The most natural assumption is that the $\tilde{y}_{i,t}$ are generated by a low order autoregressive model. Thus, we estimate the seasonal factors by:

$$\tilde{y}_{i,t} = \sum_{s=1}^{l} b(s)\hat{y}_{i,t-s} \tag{3.3}.$$

The value of l and the values of the autoregressive coefficients $b(s)$ should be determined after inspection of the $\hat{y}_{i,t}$ and, probably, computation of their spectra. The "optimum" values of the $b(s)$ depend on the exact assumption concerning the $\tilde{y}_{i,t}$ and the amount of data available. However, in practice a standard set of $b(s)'s$ will almost certainly give nearly as good results.

We can now compute the $\tilde{x}_{i,t}$ according to:

$$\tilde{x}_{i,t} = \sum_{k=1}^{p}\sum_{j=0}^{m} \hat{A}_{k,i}(j)\tilde{y}_{k,t-j} \tag{3.4}$$

where the $\hat{A}_{k,i}(j)$ is derived from the $\hat{C}_{k,i}(j)$.

The estimated seasonal series is then given by remodulating and summing the $\tilde{x}_{i,t}$:

$$\tilde{S}_t = \sum_{i=1}^{n} \tilde{x}_{i,t} e^{-i\omega_i t}. \tag{3.5}$$

Finally, the seasonally adjusted series is given by:

$$\tilde{Z}_t = X_t - \tilde{S}_t. \tag{3.6}$$

4. Relation of the Model to Linear processes and to Wald's Model

In order to explore the nature of this model somewhat further it is interesting to consider the expected result if the procedure is applied when in fact the seasonal was generated by a linear process. It will also be interesting to see that special assumptions concerning the form of the matrix $\hat{A}_{k,i}(j)$ or equivalently $F(\lambda_j)$ lead to a procedure which is similar to the Wald method of seasonal adjustment.

4.1 Linear Seasonal Variation

We may consider the effect of estimating by this procedure if the series S_t is composed of n independent series with power at frequencies ω_i for $i = 1, \ldots, n$ such that the Fourier coefficients vary independently with time. In this case it is to be expected that n factors would be required to explain the variation of the n series $\hat{x}_{i,t}$ and that these factors could simply be the series $\hat{x}_{i,t}$. Thus, we are led, in the linear case, to estimation of the time varying Fourier coefficients by n independent autoregressive schemes.

4.2 Relation to Wald's Method

We may also consider the case which, in a sense, is at the other extreme from the linear case. This is the case in which the $\hat{x}_{i,t}$ are given by:

$$\hat{x}_{i,t} = a_i \hat{x}_{1,t} \qquad i = 2, \ldots, n \tag{4.2.1}$$

This restriction on the $\hat{x}_{i,t}$ implies that they are perfectly correlated and vary proportionally without any lag. Wald's assumptions concerning the "pattern" of seasonal variation imply this assumption. (Wald (1936) pg. 78; and Godfrey and Karreman (1964) pg. 15.) If equation (4.2.1) holds then we would expect that we could explain the variation in the $\hat{x}_{i,t}$ series by a single factor $y_{1,t}$ for which the matrix $A_{k,i}(j)$ would be given by:

$$\begin{aligned} A_{k,i}(j) &= [1, a_2, \ldots, a_n] \quad & j = 0,\ k = 1, \\ &= 0, & j \neq 0 \end{aligned} \tag{4.2.2}$$

If we were to estimate $\tilde{y}_{1,t}$ by a 12 month moving average we would be exactly paralleling Wald's method.

5. Conclusion

The method proposed here has been motivated by the observed behavior of a considerable number of economic time series. The most prominent characteristic of many economic time series, as originally observed by Wald (1936), is that the seasonal pattern seems to be stable but amplitude modulated. This results in the amplitude of the fundamental and all the the harmonics varying together. Quite naturally the method presented here is similar to Wald's method and, in fact, it includes his method as a special case. However, as the method is somewhat more general than Wald's method it will provide reasonable estimates in situations such that the modulation of the pattern is not present, but the amplitude of each harmonic varies independently.

This method is far from completely developed or described in this paper. The next natural step is to program the method and study its behavior along the lines of the analysis in Godfrey and Karreman (1964).

Acknowledgment

The research described in this paper was performed at the Econometric Research program, Princeton University, with the support of NFS Grant GS 551.

I take pleasure in acknowledging the value of several conversations concerning the ideas in this paper with David Brillinger. John W. Tukey was kind enough to comment in detail on earlier drafts of the paper. The analysis presented here stems logically from the work initiated at the suggestion of, and done in collaboration with, Herman Karreman. Oskar Morgenstern has provided much helpful advice, particularly with respect to the motivation and development of Wald's method, as well as continuing encouragement.

References

1964 Godfrey, M. D. and Karreman, H., "A Spectrum Analysis of Seasonal Adjustment," Research Memorandum No. 64, *Econometric Research program,* Princeton University.

1963 Hannan, E. J., "The Estimation of the Seasonal Variation in Economic Time Series," *Journal of the American Statistical Association,* Vol. 58, pp. 31-44, March 1963.

1964 Hannan, E. J., "The Estimation of a Changing Seasonal Pattern," Mimeographed Note, 1964.

1964 Jorgenson, D. W., "Minimum Variance, Linear, Unbiased Seasonal Adjustment of Economic Time Series," *Journal of the American Statistical Association,* Vol. 59, pp. 681-724, Sept. 1964.

1963 Rosenblatt, H. M., "Spectral Analysis and Parametric Methods for Seasonal Adjustment of Economic Time Series," paper presented at the meetings of the American Statistical Association, Cleveland, Ohio, Sept. 1963.

1960 Shiskin, J., "Seasonal Adjustment on Electronic Computers," *Proceedings of an International Conference on Seasonal Adjustment Methods,* Paris, November 1960.

1936 Wald, A., *Berechnung und Ausschaltung von Saisonschwankungen,* Beiträge zur Konjunkturforschung no. 9, Verlag von Julius Springer, Vienna, 1936.