

Determination of Relevant Frequencies and Modeling Varying Amplitudes of Harmonic Processes

Winfried Theis and Claus Weihs

University of Dortmund*
Department of Statistics
44221 Dortmund, Germany

Abstract. When a process is dominated by few important frequencies the observations of this process can be modelled by a harmonic process (Bloomfield (2000)). If the amplitudes of these dominating frequencies vary over time their dominance may not be apparent during the whole process.

To discriminate between frequencies relevant for such a process we determine the distribution of the periodogram ordinates, and use this distribution to derive a procedure to assess the relevance of the frequencies. This procedure uses the standardized median (Gather and Schultze (1999)) to determine the variance of the error process. In a simulation study we show that this procedure is very efficient even under difficult conditions such as a low signal-to-noise ratio or AR(1) disturbances. Furthermore, we show that the necessary transformation to estimate the amplitudes from periodogram ordinates leads to a good normality approximation which makes it especially easy to model the development of the amplitudes from these estimates.

1 Introduction

Many processes dominated by few frequencies with varying amplitudes are well-known, e.g. music, resonance, etc.. When such a non-stationary process is observed in a noisy environment or the oscillating part of the process is obscured by an inherent stochastic process it becomes of interest to determine the really relevant frequencies. We encountered such a difficulty when investigating the BTA deep-hole drilling process and one process disturbance – called chatter – observed in this process. It turned out that chatter can be described by specific eigen-frequencies of the drilling tool bar and the development of the amplitudes of these frequencies (Weinert et al. (2002)). As long as the process stays stable the harmonic process is obscured by the noise in the process which led to the question how to determine the relevant frequencies from such data and how to model the time development of the amplitudes on these frequencies.

* This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

In this paper we first determine the distribution of periodogram ordinates of a harmonic process with only a few relevant frequencies, show how to incorporate this distribution to find the relevant frequencies and that there exists a normality approximation which readily facilitated constructing a model for time varying amplitudes. Finally we demonstrate the practical value of the procedure by results from an extensive simulation study.

2 Determination of the Distribution of Periodogram Ordinates

Gallant et al. (1974) consider analysis of variance (ANOVA) models on periodograms. Their argument – based on a Taylor series extension of the distribution function – is to transform the observed ordinates with $g(x) = x^{\frac{1}{4}}$ to increase the convergence of the χ_2^2 -distributed measurements to a normal distribution and thereby make a common ANOVA sensible in this situation.

The periodogram ordinate at frequency f equals n times the squared absolute value of the Fourier-transform F of the time series y_t at frequency f , that is

$$I[y_t](f) = n|F[y_t](f)|^2,$$

where n is the number of observations in the series.

If y_t is a Gaussian process with distribution $\mathcal{N}(0, \sigma^2)$, $F[y_t](f)$ as a linear transformation of y_t has again a normal distribution. $|F[y_t](f)|^2 = (\operatorname{Re}(F[y_t](f)))^2 + (\operatorname{Im}(F[y_t](f)))^2$ is therefore χ^2 distributed with 2 degrees of freedom, which equals an exponential distribution (cf. e.g. Fisz (1970)) with $E(|F[y_t](f)|^2) = 2\sigma_y^2$. On the basis of this argument and using the fact that the Fourier-transform is a linear operator it follows that periodogram ordinates of $\operatorname{AR}(p)$ processes are χ_{2p}^2 -distributed.

When the amplitudes at the relevant frequencies f_k , $k = 1, \dots, K$, of a harmonic process are influenced by some input variables \mathbf{x} and possibly time t , it is of interest to investigate the form of this influence. So the following model is considered:

$$H_t(\mathbf{x}) = \sum_{k=1}^K h_k(\mathbf{x}, t) \cos 2\pi(f_k t + \varphi) + \varepsilon_t, \quad (1)$$

for $t \in \{0, \dots, n-1\}$ and $K \ll n$. The functions of the amplitudes of the relevant frequencies are possibly time-dependent. Since only discrete time is considered, they are defined by $h_k : \mathbb{R}^d \times \mathbb{N} \rightarrow [0, \infty)$. For h_k only the existence of a Fourier-transform is assumed.

When all h_k are time constant it is clear that the expected value of the periodogram ordinates at the relevant frequencies is

$$E(I_{H_t(\mathbf{x})}(f)) = n(|e^{i\pi\varphi}| h_k(\mathbf{x})^2 + 2\sigma_\varepsilon^2) \text{ for } f = f_k, k = 1, \dots, K. \quad (2)$$

Note that the phase is of no interest in this model because it contributes only a constant factor in the complex Fourier transform equal to $e^{i\pi\varphi}$ of which the absolute value is 1. So the phase does not contribute to the estimates described above.

If the amplitudes are slowly time-varying (i.e. slower than the smallest estimated frequency), the corresponding model in frequency domain is in terms of the complex Fourier-transform:

$$F[H_t(\mathbf{x})](f) = \begin{cases} F[\varepsilon](f) + B_f & \text{for } f \neq f_k, \\ F[h_k(\mathbf{x}, t) \cos 2\pi(f_k t + \varphi)](f) + F[\varepsilon](f) & \text{for } f = f_k \end{cases}, \quad (3)$$

where $k = 1, \dots, K$, and $B_f \neq 0$ is only true for frequencies near to one of f_k , $k = 1, \dots, K$ and possible harmonics.

Again a result on the distribution of the periodogram ordinates is readily gained by the same arguments as above: they are χ^2 -distributed. It is only close to the relevant frequencies that you get non-central χ^2 -distribution with non-centrality parameter $\nu = n(h_k(\mathbf{x}))^2 + 2\sigma_\varepsilon^2$.

A more general determination of the distribution of periodogram ordinates can be found in Wittwer (1986). In her paper G. Wittwer determines the moment generating function and the general properties of the distribution of the periodogram ordinates for stationary sequences.

3 Regression Models on Periodogram Ordinates

3.1 Modelling varying Amplitudes

The periodogram is only able to estimate the amplitudes of Fourier frequencies, so it is of interest to know what happens when the relevant frequencies are Fourier frequencies. When the amplitudes are varying over time, we want to estimate the form of this variation. This is done by dividing the time series into sections of equal length and calculating the periodogram on these sections. Then the estimates of the amplitudes on each relevant frequency are used as objective in a – linear or nonlinear – regression to fit a proposed functional form. It can be easily proved that a linear trend in the amplitudes is transformed into a linear trend in the periodogram ordinates. When calculating the fourier transformations it turns out that using the periodogram to estimate a function of the amplitudes over time possibly underestimates the values of the function (cf. Theis (2004)).

When f_k is a non-Fourier frequency the finite Fourier transform introduces additional non-zero terms to the periodogram because it only considers Fourier frequencies. This comes from the fact that $e^{i2\pi(f_k - f)t}$ is not only non-zero at the nearest Fourier frequencies but also in a neighbourhood. This has to be taken into account when deciding how many significant appearances of a frequency in an experiment are necessary to make that frequency a relevant frequency.

3.2 Estimating the Variance of ε (σ_ε^2)

The Fourier-transform of a harmonic process with a small number of relevant frequencies K compared to the number of observations n can be viewed as a sample from a χ^2 -distribution contaminated by some non-central χ^2 -distributed observations, where the distributions have the same degrees of freedom. As remarked before the expected value of the majority of observations is $2\sigma_\varepsilon^2$, i.e. proportional to the variance of the disturbance process. A robust estimator of the expected value of this distribution is thus proportional to an estimator for the variance of the disturbance process with known proportionality factor.

Since it is assumed that in a regression situation the error processes are independent between experiments and identically distributed over all experiments, the following procedure looks promising:

1. Estimate the periodogram $I[H_t(\mathbf{x}_l)]$ for all input values \mathbf{x}_l , $l \in \{1, \dots, L\}$
2. Merge all $I[H_t(\mathbf{x}_l)](f)$ into one sample
3. Calculate a robust estimator for the expected value of $I[H_t(\mathbf{x}_l)](f)$, e.g. the standardized median $med_{st.}(X) = \frac{1}{\log(2)} med(X)$ on the merged sample

Step 2 enlarges the database for the robust estimate, because it is assumed, that the observations with different input values are independent and the realisations of $I_{H_t(\mathbf{x}_l)}(f)$ for different Fourier frequencies are independent due to the orthogonality relations of the Fourier transform. If $K \ll n$ and $\frac{K}{n}$ is lower than breakdown point of the robust estimator, which equals $\frac{1}{2}$ for the standardized median (Gather and Schulze (1999)), one gets an estimator – in the case of Gaussian white noise – for $2\sigma_\varepsilon^2$.

Transforming Periodogram Ordinates

Given that the goal of the regression on periodogram ordinates is to estimate the influences on the amplitudes, the observations have to be transformed in the following way to get an estimator for $h_k(\mathbf{x})$ (cf. (2)):

$$\hat{h}_k(\mathbf{x}) = \sqrt{\frac{I[H_t(\mathbf{x})](f_k) - 2n\sigma_\varepsilon^2}{n}},$$

Johnson et al. (1994) state that this square-root of the non-central χ^2 -distributed variable is a normal approximation. It depends on the value of the non-centrality parameter, which in return depends here on the value of the functions h_k , $k = 1, \dots, K$, and the number of observations. The impact of this approximation is tested in a simulation study.

4 Simulation Study on Time-Varying Amplitudes

4.1 Design Considerations

We chose a full factorial 2^7 design to compare the effects on the Normality assumption, the frequency detection, and the goodness of fits of the following

influences: the signal-to-noise ratio, the number of frequencies, number of observations, Fourier or non-Fourier frequencies, and the distance between the relevant frequencies. Additionally the effect of AR(1)-disturbances was checked.

For the influences the following values for treatment low, high, respectively were chosen: Number of frequencies (1 / 5), Fourier Frequencies (no/yes), Distance of frequencies δ_f (3 / 10 Fourier frequencies), Length of series (2560 / 102400), and Signal-to-Noise ratio (1.1 / 100). The frequencies considered are $\frac{5+\delta_f k-1}{n}$ with, in the case of non-Fourier frequencies, addition of $\frac{1}{\sqrt{2}}$ in the numerator.

The following functions were chosen as ‘true’ models for the variation of the amplitudes:

$$h_{lin}(t) = 2 + 0.001t \text{ or } h_{nonlin}(t) = 2 + \frac{2}{(1 + \exp(\frac{m-t}{d}))} \quad (4)$$

The parameters m, d in equation (4) are changed for each frequency if 5 frequencies are included in the model. This is done by setting $m = 5l$ or $m_i = (2+i)l$ and $d = l$ or $d_i = \frac{l}{i}$ where $i = 1, \dots, 5$. These different values for the parameters in the nonlinear function were chosen to test whether differing functions on the frequencies can be found in the data.

The choice of these functions had the following reasons: the first slow linear trend may be useful as an approximation for a slow nonlinear trend in the amplitude. Generally it can be assumed that amplitudes have an upper bound because oscillating systems break down when the amplitude becomes too large. This is the reason for the chosen logistic function. The inclusion of a mean intercept of 2 in both cases is done to ensure a true harmonic process right from the start of the observations.

For all settings 100 repetitions were evaluated. The function `nls` from **R** (R Core Team (2003)) was used to fit the nonlinear models. Since it is well known that nonlinear regressions tend to fail with some starting values, ten randomly chosen starting values were tested and the first successful set was used for the fit.

4.2 Results

First we checked whether the procedure to find the relevant frequencies was influenced by the time varying amplitudes, or the AR(1) disturbances. Both influences did not show an effect on the performance of the method in the sense that the correct frequencies are always found. This becomes obvious from the histograms of the found relevant frequencies in Figure 1.

The left panel in Figure 1 shows the results on relevant frequencies for experiments with the high level of observations, 10240, and high signal-to-noise ratio of 100 and five non-Fourier frequencies with a distance of ten Fourier frequencies in the simulated model that is, the true values of the frequencies are:

0.00557 0.0251 0.03487 0.04463 0.0544 That there are more frequencies found than just the highest peaks, is due to leakage (Bloomfield (2000)) and does not present a serious problem since it is easily possible to narrow the relevant frequencies by e.g. using higher significance levels or adding a further step where the peak(s) of the amplitudes of the found relevant frequencies is determined.

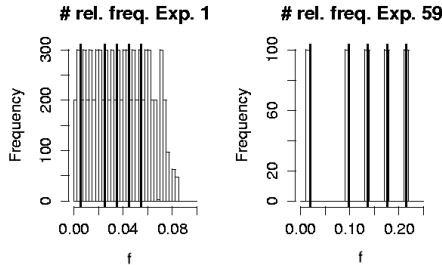


Fig. 1. The histograms of the found relevant frequencies f show clearly that at least the true frequencies (black vertical lines) are found and only a few others besides.

The right panel in Figure 1 gives an impression of the more difficult situation with only 2560 observations and AR(1) disturbances but Fourier frequencies. From this panels it is clear that at least the true frequencies are found by the method for the detection of relevant frequencies. In this case the true frequencies are: 0.0195 0.0977 0.1367 0.1758 0.2148

The proposed normality approximation was checked for appropriateness. First we applied a Shapiro-Wilk test (Shapiro et al.(1968)) to the observations. For each true relevant frequency and each of the ten observations the 100 repetitions were collected and tested for normality on the 5%–level. The test rejected the hypothesis only in 4.86% of the cases for $h_{lin.}$, and in 5.82% of the cases for $h_{nonlin.}$. The number of rejections of normality of the observations for the normal disturbances is slightly higher than with AR(1) disturbances (linear case: 5.21% vs. 4.51%; nonlinear case: 6.04% vs. 5.63%). This was expected by the theoretical model because the goodness of the approximation is influenced by the number of stochastic components, i.e. the order of the disturbance process and the value of the non-centrality parameter. No assignable pattern was found in the rejections.

The distribution of the parameter estimates was also investigated. In the linear case the parameters displayed an even greater degree of normality. The Shapiro-Wilk test rejected only in 3.47% of the situations. In the nonlinear case it cannot be expected to find normality in the parameters. It is hard to define a distribution for the parameters in nonlinear regression, only when a linear approximation approach is chosen as fitting procedure normality is expected (cf. e.g. Ratkowski (1990), p. 20).

Figure 2 gives an insight into the goodness of the fits of the functions on the truly relevant frequencies. In all cases it was obvious that the observations lie below the values of the true functions and therefore the fitted functions underestimate the true values as well.

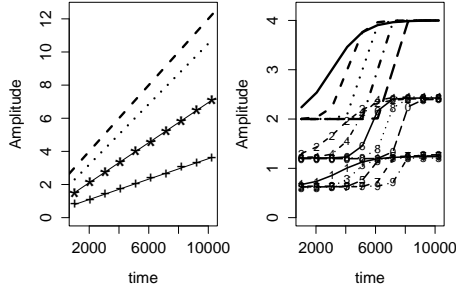


Fig. 2. Left: fitted linear functions (true: dashed line) on varying amplitudes, with AR(1) disturbances on a non-Fourier frequency, therefore two lines on the neighbouring Fourier frequencies. Right: non-linear functions of five non-Fourier frequencies, marked by numbers. True functions are again in the upper half of the graphic.

The left panel of Figure 2 shows a fit for the case of non-Fourier frequency and linear time dependence of the amplitude. This graphic gives the impression that the underestimation may be cured by summing over neighbouring frequencies in an appropriate way. This is emphasized by the dotted line which is the sum of the fitted values.

The right panel in Figure 2 underlines the previous impression as well. Furthermore, it is obvious that the general form of the influences on the amplitudes is found even if they are different for the different frequencies. This is also not influenced by the number of observations or the kind of disturbances. All fits show that the general fit of the regressions is very good which was also found when checking for the goodness of fit over all situations in the simulation study.

Studying the effect of the varying amplitudes on the performance of the proposed variance estimator, a slight overestimation of the true standard deviation σ occurred. The two most important influences on the difference between the true and the estimated σ are the signal-to-noise ratio followed by the number of observations. It turns out that a high signal-to-noise ratio also leads to better estimates of the standard deviation of the disturbance term. Of course a higher number of observations leads to a better estimation since then there are more observations following the distribution of the Fourier transformation of the AR(1) or white noise normally distributed disturbances.

5 Conclusions

We introduced a method for the identification of relevant frequencies of a harmonic process with error processes based on the normal distribution. The crucial idea for this method is to look at the estimates of the periodogram ordinates as a contaminated sample of a χ^2 distribution and use this to get an estimate of the variance of the error process. Additionally we showed that the necessary transformation of the periodogram ordinates to get an estimator for the amplitude leads to a normal approximation. Finally, we established the fact that the linearity of the Fourier transformation makes it possible to evaluate time trends in the amplitude by regression methods.

Our simulation proved all theoretical results to work even in difficult situations, i.e. low signal-to-noise ratio, non-Fourier frequencies and differing influences on the relevant frequencies. The only significant drawback of the method is the underestimation of the true amplitudes which may be tackled by summing over an appropriate neighbourhood of the found relevant frequencies.

References

- BLOOMFIELD, P. (2000) *Fourier Analysis of Time Series*. Wiley, New York, 2nd edition.
- FISZ, M. (1970) *Wahrscheinlichkeitsrechnung und Mathematische Statistik*. VEB Deutscher Verlag der Wissenschaften, 1970.
- GALLANT, A. R., GERIG, T. M. and EVANS, J. W. (1974) Time Series Realizations Obtained According to An Experimental Design. *JASA*, 69: 639–645, 1974.
- GATHER, U. and SCHULTZE V. (1999): Robust estimation of scale of an exponential distribution. *Statistica Neerlandica*, 53(3), 327–341.
- JOHNSON, N. L., KOTZ, S., and BALAKRISHNAN, N. (1994) *Continuous univariate distributions*, volume 1,2. Wiley, 1994.
- R DEVELOPMENT CORE TEAM (2003): R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, URL: <http://www.R-project.org>
- SHAPIRO, S.S., WILK, M.B. and CHEN, M.J. (1968) A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63: 1343–1372, 1968.
- THEIS, W. (2004): *Modelling Varying Amplitudes*, Dissertation, Fachbereich Statistik, Universität Dortmund, <http://eldorado.uni-dortmund.de:8080/FB5/l57/forschung/2004/Theis>
- WEINERT, K., WEBBER, O., HÜSKEN, M., MEHNEN, J., and THEIS, W. (2002): Analysis and Prediction of Dynamic Disturbances of the BTA Deep Hole Drilling Process. In Teti, R., editor, *Proceedings of the 3rd CIRP International Seminar on Intelligent Computation in Manufacturing Engineering*.
- WITTEWER, G. (1986): On the Distribution of the Periodogram for Stationary Random Sequences. *statistics*, 17(2):201–219, 1986.