Time domain parameters can be estimated with less statistical error than frequency domain parameters in the analysis of heart rate variability

Oliver Kuss* (Dr. sc.hum.), Barbara Schumann (MPH), Alexander Kluttig (MPH),

Karin Halina Greiser (Dr. med., MPH), Johannes Haerting (Prof. Dr.)

Institute of Medical Epidemiology, Biostatistics, and Informatics

Medical Faculty, University of Halle-Wittenberg

Magdeburger Str. 8

06097 Halle (Saale), Germany

Tel.: +49-345-5573582, Fax: +49-345-5573580

e-mail: Oliver.Kuss@medizin.uni-halle.de

* Corresponding author (responsible for reviewing page proof and reprint requests)

Short title: Time Series Parameters in HRV measurement

Acknowledgement of financial support:

This study was funded in part by a grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Collaborative Research Center 598 “Heart failure in the elderly – cellular mechanisms and therapy” at the Medical Faculty of the Martin-Luther-University Halle-Wittenberg and by the Wilhelm Roux program of the Medical Faculty of the Martin-Luther-University Halle-Wittenberg.

Abstract

Introduction

Measures of heart rate variability (HRV) can be divided in time- and frequency domain parameters. It is frequently ignored that estimation of frequency domain parameters is a two step procedure where statistical error from the first step (spectral estimation) is neglected in subsequent analyses.

Methods

We performed a simulation study to quantify the statistical error by using frequency domain instead of time domain parameters. We generated tachograms from a stationary AR(1) process for a wide range of parameters and compared the resulting estimation error (in terms of precision and variability) for the SDNN and LF, HF, and LF/HF power.

Results

Estimation of frequency domain parameters is associated with (up to 10-fold) increased variability, as compared to the SDNN. Moreover, the SDNN has higher precision.

Conclusion

Frequency domain parameters should be applied in HRV analysis only if important physiological reasons suggest their use. If used, frequency domain parameters should be interpreted with caution, taking into account the statistical weaknesses of spectral estimation.

Key Words

Heart Rate, Electrocardiography, Autonomic Nervous System, Fourier Analysis, Computer Simulation [MESH], Heart Rate Variability, Frequency domain, Time domain

Introduction

Cardiovascular disease (CVD) is still the leading cause of death and morbidity in industrialized nations, accounting, for example, for about 42 % of all deaths in Germany [1]. With incidence and prevalence of CVD increasing in the elderly, public health relevance and economic impact of CVD is expected to grow in the ageing populations in the western world.

A large number of risk factors contribute to CVD, among those, autonomic dysfunction as indicated by reduced heart rate variability (HRV) is associated with an increased risk of CVD incidence and mortality [2-5].

Heart rate variability is defined as the variation in time between consecutive heartbeats, and several statistics have been proposed for its measurement. Most of them are based on the tachogram, the time series of beat-to-beat time differences (RR intervals). HRV analysis thus has to rely largely on the statistical methods of time series analysis, and, corresponding to the dichotomy of methods in time series analysis, the proposed measures can be divided in time domain and frequency domain methods.

Whereas the estimation of time domain parameters is straightforward, estimation of frequency domain parameters is complicated by the fact that the estimation process consists of two steps: In the first step, the spectrum of the tachogram is estimated, and in the second step the frequency domain HRV parameters are derived from this estimated spectrum. In general, the statistical error that is induced by the estimation of the spectrum is ignored in further analysis, thus systematically overestimating the reliability of frequency domain HRV parameters. Additionally, as the common spectral estimates have rather insufficient statistical properties (e.g., the periodogram is an inconsistent estimator of the true spectrum) we expect this statistical error to be of relevant size. To our knowledge and up to now there has been no systematic investigation on the size of this error and its potential impact on interpretation of HRV parameters.

In the following we give the results of a simulation study to compare statistical precision and variability of the SDNN, LF power, HF power, and the LF/HF ratio as the most prominent time domain and frequency domain parameters, respectively.

Methods

Simulated tachograms with a length of N = 512 observations were generated from a autoregressive process of first order (AR(1)) process $Y_t = \phi Y_{t-1} + \varepsilon$, $\varepsilon \sim N(0,\sigma^2)$, $0 < \phi < 1$. An AR(1) process constitutes one of the simplest models for a tachogram, and can be interpreted like a ordinary linear regression equation with Y_t as the response, ϕ as the regression parameter for the single covariate Y_{t-1}, and a normally distributed random error ε. The current value of the process, Y_t, (or in terms of HRV, the current beat-to-beat time difference) is the sum of the previous value Y_{t-1} (multiplied by ϕ) and the random error. The association between Y_t and Y_{t-1} is controlled by the AR(1) parameter ϕ: (keeping σ^2 fixed) the larger ϕ, the higher is the correlation between Y_t and Y_{t-1}. The influence of the random error is controlled by the variance term σ^2: (keeping ϕ fixed) the larger σ^2, the larger is the random error. The idea of regressing the current value on its own predecessor explains the term “autoregressive” for the process. Autoregressive processes of higher orders (AR(p) processes) are straightforward extensions of the AR(1) process by including the p previous values of the process in the model equation ($Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon$, $\varepsilon \sim N(0,\sigma^2)$). For an AR(1) process it can be shown [7, p. 53] that the true SDNN of a tachogram from this process equals

$$\sqrt{\frac{\sigma^2}{1-\phi^2}},$$

that is, the SDNN can be calculated simply from the two parameters ϕ and σ^2 of the underlying process.

The true spectrum of an AR(1) process is [7, p. 154]

$$f(\omega) = \frac{\sigma^2}{2\pi} \times \frac{1}{(1+\phi^2 - 2\phi\cos\omega)}.$$

Based on this equation, true values for LF, HF and the LF/HF ratio can be calculated from the integral [8]

$$\int_a^b f(\omega) = \frac{\sigma^2}{\pi(\phi^2-1)} \times \arctan \frac{(\phi-1)^2(\phi^2-1)(\tan\frac{b}{2} - \tan\frac{a}{2})}{(\phi-1)^4(\phi^2-1)^2 \times \tan\frac{b}{2} \times \tan\frac{a}{2}}$$

by setting b = 0.4 and a = 0.15 for the HF, and b = 0.15 and a = 0.04 for the LF power.

In the simulation study, we varied parameter values for the AR(1) parameter ϕ and the error variance σ^2 and generated 10.000 tachograms for each combination of ϕ and σ^2. Realistic values for ϕ and σ^2 were derived from a random sample of 41 subjects of the CARLA Study [9] with regular tachograms. In this sample, the median ϕ was estimated to be 0.74, and the median σ^2 to be 237.5, corresponding to a median SDNN of 27.2 ms. In figure 1 we give the averaged autocorrelation and the averaged partial autocorrelation function for the 41 subjects together with the true theoretical autocorrelation function of a AR(1) process with $\phi = 0.74$. It can be seen that the approximation of a simple AR(1) process to the estimated autocorrelation function is not perfect, in fact there is some evidence for longer memory in the data (as indicated by the spikes in the partial autocorrelation function at lags 3 and 4). However, we felt that, at least for the sake of simulation, correspondence is sufficient. Especially, there is no sign of a periodical behaviour in the sample. We finally varied ϕ from 0.50 to 0.95 in steps of 0.05 and σ^2 between 150, 250, and 350 and simulated tachograms for each combination of these. Note that an AR(1) process is stationary for values $\phi < |1|$.

From each tachogram we calculated the estimated SDNN as the standard deviation of the RR intervals, and LF, HF, and LF/HF ratio as areas under the curve from the raw periodogram. Additionally we also estimated all frequency domain parameters from a smoothed periodogram where smoothing was performed with a Tukey-Hanning kernel with a fixed bandwidth of 5 periodogram ordinates.

In order to compare SDNN, LF, HF, and LF/HF ratio, we were interested in two different phenomena. First, we sought to describe how precisely the respective parameters estimate the corresponding true values (statistical precision or statistical bias); second, we intended to describe how strong the estimated values vary around the estimated values (statistical variability).

To describe statistical precision we calculated the relative percentage bias (RPB) [10]

$$\mathrm{RPB} = \left(\frac{\bar{\hat{\theta}} - \theta}{\theta} \right) \times 100 ,$$

where θ stands for the respective true parameter value (true SDNN, LF, HF, or LF/HF ratio as derived from the equations above) and $\bar{\hat{\theta}}$ for the corresponding mean estimated value

$$\bar{\hat{\theta}} = \frac{1}{10.000} \sum_{i=1}^{10.000} \hat{\theta}_i ,$$

$\hat{\theta}_i$ being the estimated value from the i-th simulation run.

To describe statistical variability we calculated the relative percentage standard deviation (RPSD) [10]

$$\text{RPSD} = \left(\frac{\text{SD}(\hat{\theta})}{\bar{\hat{\theta}}} \right) \times 100$$

with $\text{SD}(\hat{\theta}) = \sqrt{\frac{1}{10.000 - 1} \sum_{i=1}^{10.000} (\hat{\theta}_i - \bar{\hat{\theta}})^2}$ as the empirical standard deviation over all simulations for the respective parameter. It should be noted that there is no reference to the true parameter in the formula for RPSD, that is, it measures only the variability around the estimated average parameters, thus keeping precision (as measured by RPB) and variability (as measured by RPSD) strictly separated.

To make this rather technical description of our study design more comprehensible, we give an example for a fixed parameter constellation. Suppose we fix ϕ = 0.75 and σ^2 = 250. A theoretical tachogram from an AR(1) process with these values will have a true SDNN of 23.9 ms, a true LF power of 70.3 ms^2, a true HF power of 84.0 ms^2 and a true LF/HF ratio of 0.84. In our simulation, we now generated 10.000 tachograms from this process and estimate the SDNN, LF, HF, and LF/HF ratio from each of the tachograms. As we actually know the true values which generated the tachograms, we can easily compare the estimated values to the true ones and calculate precision (RPB) and variability (RPSD) as described.

The simulation study and all analyses were programmed in SAS®, 9.1 (SAS Institute, Cary, NC, USA).

Results

In figures 2 and 3 we give the results of our simulation study in terms of the defined outcomes, precision (RPB) and variability (RPSD). As results were essentially unchanged across the different values of σ^2, we only report results for σ^2 = 250. We also omit confidence intervals in the graphs, first to enhance their readability, and

second, because confidence intervals are very small in our setting with 10.000 observations for each estimate. Moreover, length of confidence intervals could be further reduced by enlarging the number of simulation runs. To give an example, the 95% confidence interval for $RPB_{LF,\ periodogram}$ (the estimate being - 10.5%) is [-10.0%, -11.1%].

In terms of precision (figure 2), there is only a small bias of the SDNN and the LF/HF ratio across the whole range of the AR(1) parameter ϕ. HF consistently underestimates the true value by 4-6 %, and the underestimation in terms of the LF is even larger (10-12 %). At first sight it might look counterintuitive that the LF/HF ratio is essentially unbiased. This can be explained in part by the fact that both LF and HF are negatively biased and these biases cancel out in their ratio LF/HF. However, in calculating the mean bias of the LF/HF ratio we do not build a ratio of mean biases of LF and HF, but rather calculate the mean bias of a ratio, which is not the same.

We might further be surprised that the SDNN becomes more biased the closer ϕ approaches 1, because in principle we expect the SDNN to be unbiased. However, unbiasedness is an asymptotic property, that is, unbiasedness is guaranteed only in infinite samples. Of course, our sample is finite, and the relevant sample size in this case is the length of the original tachogram, which is 512 in our case. We ran some additional simulations with longer tachograms (data not shown), and the bias of SDNN was indeed observed to diminish.

Biases from the smoothed periodograms are more erratic than those from the raw periodograms and become more serious with larger values of ϕ.

In terms of variability (figure 3), the SDNN is observed to have by far the smallest variability, being smaller than 10 % across most of the range of ϕ. Opposed to this, the frequency domain parameters show larger variability, which is 20-25 % for the HF, 30-35 % for the LF, and 40-45 % for the LF/HF ratio. Again, smoothed spectra do not solve the problems of the raw periodograms. In summary, variability for the frequency domain parameters is 1.4 to 10-fold larger than the variability for the SDNN, the “best” case occurring with $\phi = 0.95$ ($RPSD_{SDNN}$ = 13.6, $RPSD_{HF,smoothed}$ = 19.1), and the “worst” case with $\phi = 0.50$ ($RPSD_{SDNN}$ = 4.0, $RPSD_{LF/HF,\ periodogram}$ = 40.2).

Discussion

Our simulation study yielded a clear result: In HRV analysis, time domain parameters can be estimated with smaller bias and considerably smaller variability as compared to frequency domain parameters.

This result was found despite the fact that the simulation study was set up in the best possible way to give valid results for the frequency domain parameters: We generated tachograms from a stationary process without ectopic beats or other arrhythmic events, and chose tachogram length as a power of 2 ($512 = 2^9$). These optimized conditions will not be given with real life data, instead observed tachograms will be non-stationary (thus invalidating standard spectral estimation), especially when larger recording periods are being used. Moreover, a number of pre-processing steps (trend removal, interpolation of ectopic beats, zero padding, tapering, resampling) will be necessary for valid spectral estimation [6]. As different studies use different protocols for pre-processing, we feel that results of frequency domain parameters are hardly comparable. It would be of great help to have an extension of the Task Force statement [6] explicitly fixing the pre-processing steps for frequency domain analysis.

To give an impression of the actual relevance of our results we offer the following explanation in terms of statistical power. Assume we plan a study to assess the mean difference in HRV between two groups (e.g., male and female) and use the SDNN and the HF power as HRV measures. We consider a 10 % difference in the respective measure as clinically relevant, corresponding, for example, to a gender difference of 40 versus 44 ms SDNN, or 100 versus 110 ms^2 HF power. The 2-group t-test for independent data would be an appropriate statistical test to compare means. The necessary sample size can be computed by

$$n = \frac{\sigma_m^2 + \sigma_f^2}{(\mu_m - \mu_f)^2}\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 .$$

Keeping size (α), power ($1-\beta$), and the relative difference in means constant, the sample size rises *quadratically* with the standard deviation. That is, a *three-fold* larger standard deviation for the HF power results in a *nine-fold* larger sample size to detect a 10 % difference in HF values, as compared to SDNN values. Note that in the

simulation the HF power was shown to have an up to *six-fold* larger standard deviation, corresponding to a *thirty-six-fold* increase in the necessary sample size.

It is a limitation of our simulation study that it uses a rather simplistic model for tachogram generation, and we showed (figure 1) that real data do not necessarily follow such a simple model. We decided to use an AR(1) process mainly because computation of true values of the frequency domain parameter requires complicated integrals, and using AR processes with higher orders would make these integrals essentially intractable. However, the estimation procedure for time and frequency domain parameters is identical for AR processes of first order and for AR processes with higher orders. That is, statistical weaknesses of spectral estimation will be exactly the same with AR(1) vs AR(p) processes. Moreover, with higher orders of the underlying processes, true spectra will have more complicated forms, e.g., with multiple modes, and spectral estimation is expected to perform even worse. Therefore, there is no reason to believe that spectral estimation works better in these situations.

Another limitation is that our simulation only compared time domain and frequency domain parameters. It has recently been shown [11] that nonlinear measures of HRV offer additional and independent prognostic information for predicting cardiovascular events. However, as some of these nonlinear measures are also calculated from the periodogram (e.g., the 1/f slope) we expect similar problems with precision and variability for these nonlinear HRV measures.

We emphasize that our simulation only allows conclusions concerning statistical issues. The possibility of identifying periodic oscillations in the HRV signal and correlating these oscillations to the autonomic nervous system has promoted the application of frequency domain parameters in HRV analysis [12]. For example, under controlled conditions the HF power directly reflects vagally mediated respiratory effects, tested in clinical studies with pharmacologically determined cardiac vagal tone [12,13]. In contrast, interpretation of LF power is more complex as it is sensitive to various other factors, thus complicating a straightforward physiological interpretation [14]. However, LF and HF power together allow a characterisation of sympatho-vagal balance. That is, if we are mainly interested in the

balance of sympathetic and parasympathetic activity we probably still have to rely on frequency domain parameters, in spite of the described statistical weaknesses.

We finally conclude that frequency domain parameters should be applied in HRV analysis only if important physiological reasons suggest their use. If used, frequency domain parameters should be interpreted with caution, taking the statistical weaknesses of spectral estimation into account. Because of its higher robustness, the calculation of time domain parameters should be added to the analysis of associations of HRV with morbidity and mortality. Although the use of time domain parameters has been recommended by the task force for long-term ECGs only, several studies have demonstrated good predictive ability of SDNN also from short-term ECGs [15,16].

References

1. Statistisches Bundesamt (Federal Statistical Office): Statistisches Jahrbuch 2006 für die Bundesrepublik Deutschland - Statistical Yearbook 2006 for the Federal Republic of Germany. Statistisches Bundesamt (Federal Statistical Office), Wiesbaden, 2006

2. Tsuji H, Venditti FJJ, Manders ES, Evans JC, Larson MG, Feldman CL, Levy D: Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study. Circulation 1994;90:878-83.

3. Liao D, Cai J, Rosamond WD, Barnes RW, Hutchinson RG, Whitsel EA, Rautaharju P, Heiss G: Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study. The ARIC Study. Atherosclerosis Risk in Communities Study. Am J Epidemiol 1997;145:696-706.

4. Dekker JM, Crow RS, Folsom AR, Hannan PJ, Liao D, Swenne CA, Schouten EG: Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: the ARIC Study. Atherosclerosis Risk In Communities. Circulation 2000;102:1239-44.

5. Nolan J, Batin PD, Andrews R, Lindsay SJ, Brooksby P, Mullen M, Baig W, Flapan AD, Cowley A, Prescott RJ, Neilson JM, Fox KA: Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-heart). Circulation 1998, 98:1510-1516.

6. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996;17:354–81.

7. Hamilton JD: Time Series Analysis, 1st ed. Princeton University Press, Princeton, 1994

8. Jeffrey A: Handbook of Mathematical Formulas and Integrals. 1st ed. Academic Press, San Diego, 1995

9. Greiser KH, Kluttig A, Schumann B, Kors JA, Swenne CA, Kuss O, Werdan K, Haerting J: Cardiovascular disease, risk factors and heart rate variability in the elderly general population: design and objectives of the CARdiovascular disease, Living and Ageing in Halle (CARLA) Study. BMC Cardiovasc Disord 2005;5:33.

10. Burton A, Altman DG, Royston P, Holder RL: The design of simulation studies in medical statistics. Stat Med 2006;25:4279-92.

11. Stein PK, Domitrovich PP, Huikuri HV, Kleiger RE: Traditional and nonlinear heart rate variability are each independently associated with mortality after myocardial infarction. J Cardiovasc Electrophysiol 2005;16:13-20.

12. Akselrod S, Gordon D, Ubel FA, Shannon DC, Berger AC, Cohen RJ. Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. Science 1981; 213:220-2.

13. Hayano J, Sakakibara Y, Yamada A, Yamada M, Mukai S, Fujinami T, Yokoyama K, Watanabe Y, Takata K. Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects. Am J Cardiol 1991;67:199-204.

14. Lombardi F: Physiological understanding of HRV components. pp. 40-47. In Malik M, Camm AJ (eds): Dynamic Electrocardiography. 1st ed. Blackwell, Oxford, 2004

15. de Bruyne MC, Kors JA, Hoes AW, Klootwijk P, Dekker JM, Hofman A, van Bemmel JH, Grobbee DE. Both decreased and increased heart rate variability on the standard 10-second electrocardiogram predict cardiac mortality in the elderly: the Rotterdam Study. Am J Epidemiol 1999;150:1282-8.

16. Dekker JM, Schouten EG, Klootwijk P, Pool J, Swenne CA, Kromhout D. Heart rate variability from short electrocardiographic recordings predicts mortality from all

causes in middle-aged and elderly men. The Zutphen Study. Am J Epidemiol 1997;145:899-908.

Figure legends

Figure 1:

Estimated and theoretical autocorrelation functions. The estimated autocorrelation function and its partial counterpart were estimated from 41 CARLA [9] subjects with regular tachograms. The theoretical autocorrelation function is the true autocorrelation function for a AR(1) process with $\phi = 0.74$.

Figure 2:

Relative percent bias (RPB) for the respective HRV parameters.

Figure 3:

Relative percent standard deviation (RPSD) for the respective HRV parameters.

Figure 1

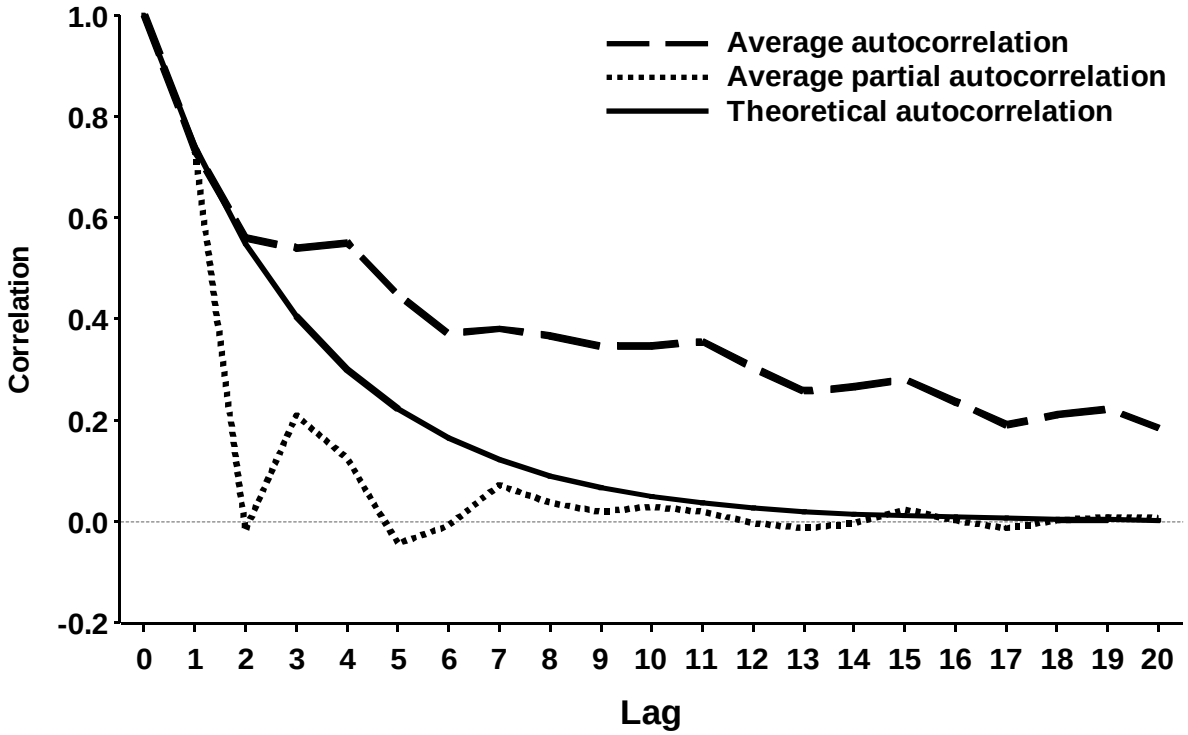

Average autocorrelation
Average partial autocorrelation
Theoretical autocorrelation
Correlation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Lag

Figure 2

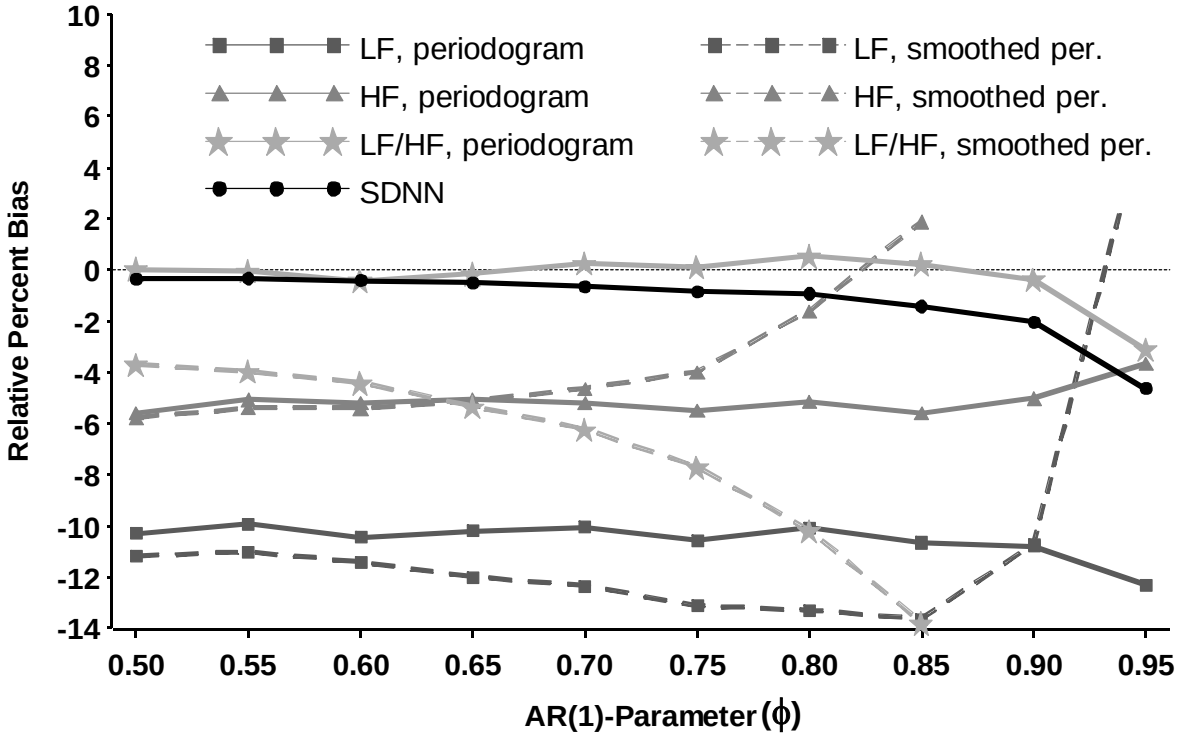
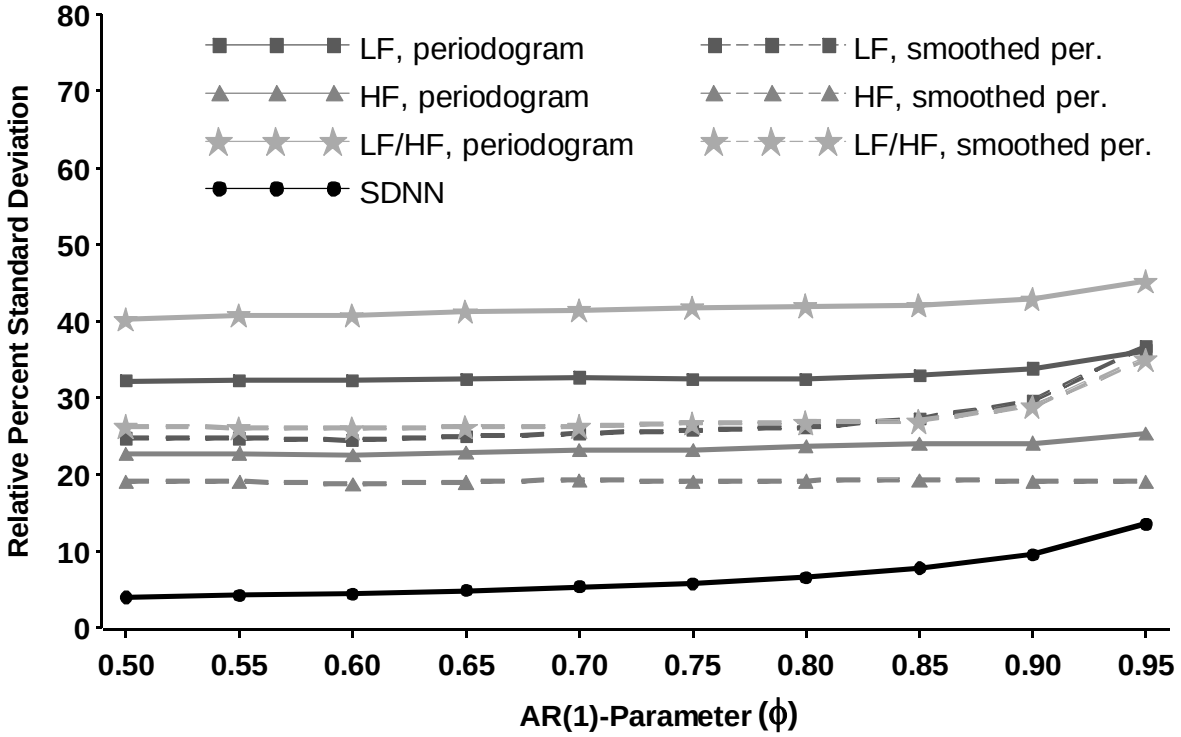

LF, periodogram
LF, smoothed per.
HF, periodogram
HF, smoothed per.
LF/HF, periodogram
LF/HF, smoothed per.
SDNN
Relative Percent Bias
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
AR(1)-Parameter (ϕ)

Figure 3


LF, periodogram
LF, smoothed per.
HF, periodogram
HF, smoothed per.
LF/HF, periodogram
LF/HF, smoothed per.
SDNN
Relative Percent Standard Deviation
0
10
20
30
40
50
60
70
80
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
AR(1)-Parameter (ϕ)