DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 3 | Issue 2 Article 9

11-1-2004

Variance Stabilizing Power Transformation for Time Series

Victor M. Guerrero
Instituto Tecnológico Autónomo de México, guerrero@itam.mx

Rafael Perera
Instituto Tecnológico Autónomo de México, rafael.perera@public-health.oxford.ac.uk

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Guerrero, Victor M. and Perera, Rafael (2004) "Variance Stabilizing Power Transformation for Time Series," *Journal of Modern Applied Statistical Methods*: Vol. 3 : Iss. 2 , Article 9.
DOI: 10.22237/jmasm/1099267740
Available at: http://digitalcommons.wayne.edu/jmasm/vol3/iss2/9

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Journal of Modern Applied Statistical Methods
November, 2004, Vol. 3, No. 2, 357-369
Copyright © 2004 JMASM, Inc.
1538 – 9472/04/$95.00

Variance Stabilizing Power Transformation for Time Series

Victor M. Guerrero Rafael Perera
Department of Statistics
Instituto Tecnológico Autónomo de México

A confidence interval was derived for the index of a power transformation that stabilizes the variance of a time-series. The process starts from a model-independent procedure that minimizes a coefficient of variation to yield a point estimate of the transformation index. The confidence coefficient of the interval is calibrated through a simulation.

Key words: Autocorrelation, coefficient of variation, confidence interval, model assumptions

Introduction

Applied model-based statistical analysis usually requires some assumptions to be satisfied by the data under study. When working with time-series, covariance-stationarity is often required to begin the modeling process. Therefore it is reasonable to look for a variance stabilizing transformation that will make the data get closer to fulfilling this assumption. Within the forecasting area, recall de Bruin and Franses' (1999) conclusion that data transformations should be considered prior to forecasting.

There are two approaches to search for the transformation. (i) Select the transformation before actually building a statistical model for the time series, or (ii) decide which transformation to use during the model building process. In the latter approach both model form and parameter estimation interact with the search for the transformation.

Victor M. Guerrero is Professor of Statistics at Instituto Tecnológico Autónomo de México. His research interests are in the analysis and forecasting of economic time series. He gratefully acknowledges financial support from Asociacion Mexicana de Cultura, A. C., in carrying out this work. Email him at: guerrero@itam.mx. Rafael Perera graduated from the University of Oxford and is employed at the Department of Primary Health Care at Oxford University. Email him at: rafael.perera@public-health.oxford.ac.uk.

In the former, the scale where the analysis should be carried out is fixed before attempting to build a statistical model. This approach allows the analyst to select a transformation without conditioning on or interfering with a given model. Therefore it is called model-independent.

The focus in this article is on a model-independent method that is useful to select a power transformation that best stabilizes the variance of a time series variable $Z_t > 0$, for t=1,…,N. Such a method was proposed by Guerrero (1993) as a tool to be employed when the analyst wants to use the power transformation family: $T(Z_t)= Z_t^{\lambda}$ if $\lambda \neq 0$ and $T(Z_t)= \log(Z_t)$ if $\lambda=0$ or when using its Box-Cox version: $Z_t^{(\lambda)}=(Z_t^{\lambda}-1)/\lambda$ if $\lambda \neq 0$ and $Z_t^{(\lambda)}=\log(Z_t)$ if $\lambda=0$.

One of the most important works that proposed the second approach for choosing a transformation is the textbook by Box and Jenkins (1976). They suggested using the Box-Cox transformation in order to validate not only the constant variance assumption, but all the underlying assumptions of an Auto-Regressive Integrated Moving Average (ARIMA) model by estimating the transformation index (λ) together with the model parameters. Chen and Lee (1997) proposed a Bayesian method to choose the value of λ for a given model structure. Those works are supported by sound statistical theory, although in practice they present the problem that the model form may depend on the transformation selected. In fact, Gourieroux and Jasiak (2002) have shown that the autocorrelations (hence the ARIMA model structure) change as a function of the nonlinear transformation employed. Therefore, fixing the model form before selecting the transformation index could be inappropriate in some cases.

An advantage of the second approach for choosing the transformation is that a measure of variability, as well as a reference distribution, can be obtained for the estimated transformation index. Thus, it is possible to discriminate among different alternative values of λ based on à priori considerations. For instance, deciding whether the data should be analyzed in the original scale (λ = 1) or in logarithms (λ = 0), can be performed on the basis of the data at hand. This does not happen with the first approach because no model form and no reference distribution exist that will support the decision on an empirical basis. This fact can be considered a drawback of this approach. In this article, we consider this problem and work out a feasible solution by means of a confidence interval for the true λ value.

In the following section a summary of Guerrero's (1993) method is presented that produces a point estimate of the index λ by minimizing a coefficient of variation. Then, a confidence interval is derived for the true value of λ. Approximate expressions for some sample moments involved in the calculations are provided, and a reference distribution for the true coefficient of variation employed by the method is suggested. Some small sample simulations are used to calibrate the confidence coefficient of the interval and to get an insight into the performance of the procedure. Nominal confidence levels are related to realized levels and, useful empirical results are obtained. A section is devoted to illustrate the use of the method through some empirical applications. These examples help to understand how the method works in practice.

Selection of the Transformation

Guerrero (1993) proposed two methods for selecting the power transformation index λ. Underlying these methods is the theoretical result that states that the choice of the transformation index should be done in such a way that $[\mathrm{var}(Z_t)]^{1/2}/[E(Z_t)]^{1-\lambda} = c$ holds valid for all t and some constant $c > 0$. To use this result, it is necessary to estimate both the mean and the variance involved. In applied time series analysis there is usually only one observation at each time t, therefore $\mathrm{var}(Z_t)$ cannot be estimated and that result cannot be applied directly. In order to operationalize the result, work with the observations grouped into H≥2 subseries. This enables the calculation of pairs of sample means and standard deviations, for example, $(\bar{Z}_h, S_h)$ for h=1,...,H, and then search for the λ value that produces

$$S_h / \bar{Z}_h^{1-\lambda} = c \text{ for h=1,…,H} \qquad (1)$$

for some constant c>0. The elements in this equation are given by $\bar{Z}_h = \sum_{r=1}^{R} Z_{h,r} / R$ and $S_h^2 = \sum_{r=1}^{R} (Z_{h,r} - \bar{Z}_h)^2 / (R-1)$, where $Z_{h,r}$ denotes the *r*th observation of subseries h. The subseries $\{Z_{h,1}, ..., Z_{h,r}, ..., Z_{h,R}\}$, for h=1,...,H, are formed by grouping R consecutive observations of the original series $\{Z_t : t = 1, ..., N\}$, trying to keep homogeneity between the subseries. For this to happen they must be equal-sized. Therefore, some number (n) of observations, with 0≤n<R, will have to be left out of the calculations, leaving R=(N-n)/H. The subseries size must be chosen appropriately, and be equal to the length of the seasonality, if such an effect is present in the series.

The proposed methods stemmed from two empirical interpretations of equation (1). The first one led to minimizing the coefficient of variation of $S_h / \bar{Z}_h^{1-\lambda}$ as a function of λ. This method is not linked to a formal statistical model and therefore no assumptions need to be validated to be applied correctly in practice. The second empirical interpretation led to a method based on a simple linear regression in logarithms. The assumption of zero error autocorrelation that underlies this method needs careful attention as it is seldom valid when working with time series. Thus, the main method, because of its robustness against violation of assumptions, is the one that minimizes relative variation. We shall concentrate on that method.

A Confidence Interval for λ

To be able to make inferences about λ, estimated as the minimizer of the coefficient of variation, we require a reference statistical distribution. To get such a distribution we start by assuming that the random variables $W_h(\lambda) = S_h / \bar{Z}_h^{1-\lambda}$ for h=1,…,H, can be represented by a Moving Average model of order 1. That is $W_1(\lambda) - \mu = a_1$ and $W_h(\lambda) - \mu = a_h - \theta a_{h-1}$ for h=2,…,H, with $\{a_h\}$ a zero-mean white noise Gaussian process, $\mu>0$ and θ a constant parameter such that $|\theta|<1$ to ensure it is invertible. Thus, $E[W_h(\lambda)] = \mu$, $\mathrm{var}[W_h(\lambda)] = \sigma^2$ and $\mathrm{corr}[W_h(\lambda), W_{h'}(\lambda)] = \rho$ if $h' = h \pm 1$, and zero otherwise, with $\rho = -\theta/(1+\theta^2) \in (-0.5, 0.5)$.

Such a model makes sense because λ is obtained in such a way that $W_1(\lambda),\ldots,W_H(\lambda)$ are approximately constant, but a slight autocorrelation structure is expected in the process $\{W_h(\lambda)\}$ given that $\bar{Z}_h$ and S_h are calculated from time series observations. This assumption was validated by the simulations reported below as the expected behavior was observed. For the sake of simplicity, do not write $\mu(\lambda)$, $\sigma^2(\lambda)$ and $\rho(\lambda)$ even though these parameters are functions of λ.

The sample counterparts of μ and σ^2 will be denoted as $m = \sum_{h=1}^{H} W_h(\lambda)/H$ and $se^2 = \sum_{h=1}^{H}[W_h(\lambda) - m]^2/(H-1)$ so that $CV(\lambda) = se/m$ is the sample coefficient of variation. In what follows we shall derive an approximate distribution for CV(λ), from which a confidence interval for the true λ value can be obtained. Several proposals may be found in the literature to obtain the distribution, hence confidence intervals, for a Normal coefficient of variation (see Vangel, 1996, and the references therein), but none of them allows for autocorrelation in the observations.

We first apply the Theorem in Appendix 1 (known as the Delta Method) to the bivariate case, with $X_1 = se$, $X_2 = m$ and $g(X_1, X_2) = X_1/X_2$, to get $E[CV(\lambda)] \approx E(se)/E(m)$ and

$$\mathrm{var}\left[CV(\lambda)\right] \approx \frac{E^2(se)}{E^2(m)}\left[\frac{\mathrm{var}(se)}{E^2(se)} + \frac{\mathrm{var}(m)}{E^2(m)} - \frac{2\,\mathrm{cov}(se,m)}{E(se)E(m)}\right].$$

Then, evaluate each term in this expression as indicated in Appendix 2, so that $E(m) = \mu$,

$$\mathrm{var}(m) = \sigma^2[1 + 2\rho(H-1)/H]/H,$$
$$E(se) \approx \sigma\{1 - 2\rho/H - 1/[2(H-1)]\}^{1/2},$$

$\mathrm{var}(se) \approx \sigma^2/[2(H-1)]$ and $\mathrm{cov}(se, m) \approx 0$. Hence,

$$E[CV(\lambda)] \approx \frac{\sigma}{\mu}\left[1 - \frac{2\rho}{H} - \frac{1}{2(H-1)}\right]^{1/2}$$

and

$$\mathrm{var}\left[CV(\lambda)\right] \approx \frac{\sigma^2}{2(H-1)\mu^2}\left\{1 + \frac{2\sigma^2}{\mu^2}\left(\frac{H-1}{H}\right)\left[1 - \frac{2\rho}{H} - \frac{1}{2(H-1)}\right]\left[1 + \frac{2\rho(H-1)}{H}\right]\right\}$$

$$\approx \sigma^2/[2(H-1)\mu^2]$$

where the last approximation follows from the fact that σ/μ must be close to zero, since λ is chosen to accomplish that goal. It is clear that $E[CV(\lambda)] \to \sigma/\mu$ as $H \to \infty$ and that it is a decreasing function of ρ. In fact, when ρ≥0 we observe that $E[CV(\lambda)] < \sigma/\mu$ for all H, and the opposite occurs when ρ<0. Similarly, it is easy to see that $\mathrm{var}[CV(\lambda)] \to 0$ as $H \to \infty$.

Because the variance of CV(λ) is proportional to the square of its mean, the logarithm becomes an adequate variance-stabilizing power transformation (see Guerrero, 1993, eq. 4). In turn, assume that (roughly) $\log[CV(\lambda)] \sim N(\eta, \delta^2)$. From the Lognormal distribution, $E[CV(\lambda)] = \exp(\eta + \delta^2/2)$ and $\mathrm{var}[CV(\lambda)] = [\exp(\delta^2) - 1]\exp(\delta^2 + 2\eta)$. Thus, solve for η and δ², to get

$$\eta = \log\{E[CV(\lambda)]\} - \delta^2/2$$
$$\approx \log(\sigma/\mu) + \log\{1 - 2\rho/H - 1/[2(H-1)]\}/2 - \delta^2/2$$

and

$$\delta^2 = \log\left\{\frac{\mathrm{var}[\mathrm{CV}(\lambda)]}{\mathrm{E}^2[\mathrm{CV}(\lambda)]}+1\right\}$$

$$\approx \log(1-2\rho/\mathrm{H})-\log\{1-2\rho/\mathrm{H}-1/[2(\mathrm{H}-1)]\}$$

It is known that

$$1-\alpha \approx \Pr\left\{\frac{\log \mathrm{CV}(\lambda)-\eta}{\delta} \geq -z_\alpha\right\}$$

$$= \Pr\left\{\mathrm{CV}(\lambda) \geq \frac{\sigma}{\mu}\frac{1-2\rho/\mathrm{H}-1/[2(\mathrm{H}-1)]}{(1-2\rho/\mathrm{H})^{1/2}}\exp(-\delta z_\alpha)\right\}$$

with z_α the 100α upper percentile of the unit Normal distribution. The previous assertion leads us to an approximate 100(1-α)% confidence interval for the true coefficient of variation. Because, there is a one-to-one correspondence between coefficient of variation and λ value, it follows that an approximate 100(1-α)% confidence interval for λ is given by

$$CI_{1-\alpha}(\lambda) = \left\{\lambda : \frac{\sigma}{\mu} \leq \frac{CV(\lambda)(1-2\rho/H)^{1/2}}{1-2\rho/H-1/[2(H-1)]}\exp(\delta z_\alpha)\right\}. \tag{2}$$

In order for this confidence interval to be useful in practice, estimate $\mathrm{CV}(\lambda)$ as the minimum sample coefficient of variation, denoted as $\mathrm{CV}(\hat{\lambda})$. Similarly, use the estimated first-order autocorrelation coefficient,

$$\hat{\rho} = \frac{\sum_{h=1}^{H-1}\left[W_h(\hat{\lambda})-m\right]\left[\mathrm{W}_{h+1}(\hat{\lambda})-m\right]}{\sum_{h=1}^{H}\left[W_h(\hat{\lambda})-m\right]^2}$$

and an estimate of δ, say $\hat{\delta}$, can be obtained by using $\hat{\rho}$ in place of ρ. Keep in mind that the interval (2) was derived from several approximations, in such a way that the actual confidence level may differ from the nominal level and calibration is required.

To appreciate numerically the effect that α, H and ρ have on the length of the confidence interval, some calculations are presented in Table 1 for selected values of those constants. This table shows values of the function

$$f(\alpha,\mathrm{H},\rho) \equiv \exp(\delta z_\alpha)(1-2\rho/\mathrm{H})^{1/2}/\{1-2\rho/\mathrm{H}-1/[2(\mathrm{H}-1)]\}$$

which is the expanding factor of $\mathrm{CV}(\lambda)$ that defines the length of $\mathrm{CI}_{1-\alpha}(\lambda)$. It is clear that $f(\alpha,\mathrm{H},\rho)$ gets smaller as: (i) α gets larger, (ii) H gets larger (in fact, $f(\alpha,\mathrm{H},\rho)\to 1$ as $\mathrm{H}\to\infty$) and/or (iii) ρ moves from positive to negative values. The first two of these conclusions have a clear interpretation in terms of confidence and sample size. The third has no clear explanation, but it should be borne in mind when trying to understand why two similar situations, differing only in the sign of ρ, will yield different results (especially when α and H are small). In practical applications, typically H≥6, so that ρ should not be expected to be the decisive factor in defining the size of the confidence interval, but we should be aware of its potential relevance.

In order to better understand how the method works, in Figure 1 the graph is presented of CV(λ) against λ for the Sales Data that will be considered as an illustrative example below. Observe that the confidence interval is obtained by slicing the curve produced by the coefficient of variation of the variable $\mathrm{S}_h/\overline{\mathrm{Z}}_h^{1-\lambda}$, for h=1,...,H, as a function of λ. The minimum of this curve yields $\mathrm{CV}(\hat{\lambda})$ and the required confidence interval is built by projecting on the horizontal axis the points where the curve reaches $\mathrm{CV}(\hat{\lambda})f(\alpha,\mathrm{H},\hat{\rho})$, for a given α value.

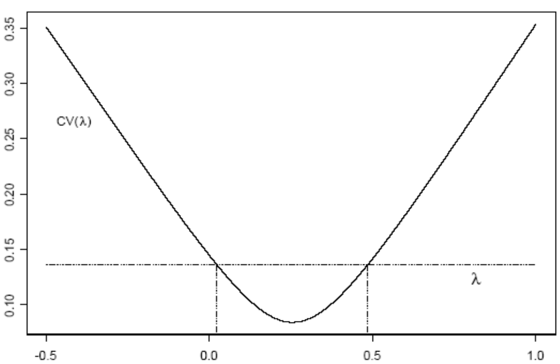
Table 1: Expanding factor of CV(λ) as a function of α, H and ρ.

α	H\ρ	-0.45	-0.25	-0.05	0	0.05	0.25	0.45
0.01	2	5.75	7.86	12.1	13.9	16.2	39.7	544
	6	2.06	2.18	2.33	2.36	2.40	2.58	2.80
	50	1.26	1.27	1.28	1.28	1.28	1.29	1.29
0.05	2	3.69	4.83	6.99	7.87	8.98	19.4	189
	6	1.68	1.77	1.87	1.90	1.92	2.05	2.20
	50	1.18	1.19	1.19	1.19	1.19	1.20	1.21
0.1	2	2.92	3.73	5.22	5.82	6.56	13.3	108
	6	1.50	1.58	1.66	1.68	1.71	1.81	1.93
	50	1.14	1.14	1.15	1.15	1.15	1.16	1.16

Methodology

The confidence interval for λ was derived from several approximations that may cause the actual confidence level to differ from its nominal level. In order to calibrate the confidence coefficient, a small simulation study based on the following two model specifications was conducted.

Figure 1. 95% Confidence interval for λ built from CV($\hat{\lambda}$) for the Sales Data.



1) $Z_t = T_t + S_t + I_t$,

where $T_t = t$, $S_t = \sum_{q=1}^{12} \delta_q D_{t,q}$ with $D_{t,q} = 1$ if $t = 12i + q$ for i = 0,1,…,H-1 and $D_{t,q} = 0$ otherwise. $\{I_t\}$ ~ independent $N(0, \sigma_t^2)$ with $\sigma_t^2 = E(Z_t)^{2(1-\lambda)}$ and $E(Z_t) = T_t + S_t$.

2) $Z_t = (1+\phi)Z_{t-1} - \phi Z_{t-2} + a_t - \theta a_{t-1}$, where $\{a_t\}$ ~ independent $N(0, \sigma_t^2)$ with $\sigma_t^2 = E(Z_t)^{2(1-\lambda)}$ and $E(Z_t) = [(2-\phi) - \phi^t]/(1-\phi)$.

The first one is a seasonal model with seasonality length R=12. The parameter values for the seasonal effects were chosen as $\delta_1 = 2$, $\delta_2 = 4$, $\delta_3 = 5$, $\delta_4 = 0$, $\delta_5 = -1$, $\delta_6 = -2$, $\delta_7 = -3$, $\delta_8 = -3$, $\delta_9 = -2$, $\delta_{10} = 0$, $\delta_{11} = 1$, $\delta_{12} = -1$ so that $\sum_{q=1}^{12} \delta_q = 0$. The sample sizes were of the form N=12H, with H=6, 12, 20, 30. The second is an ARIMA(1,1,1) model with initial values $Z_0 = 1$, and $Z_1 = 2$ with parameter values $\phi = 0.7$ and $\theta = 0.3$. In this case, the subseries size was taken as R = 4 and the sample sizes were N=24, 48, 80, 120, so that the values of H became again 6, 12, 20 and 30. Another exercise was carried out with the latter model and R=3, and sample sizes N=18, 36, 60, 90 to get the same values for H as before. For both models, λ=0,0.5,1 was employed; thus, when λ≠1 there is nonconstant variance, because it depends on the mean of the series.

Jennings' (1987) suggestion about the way that simulation studies should be reported was followed in order to provide information not only on coverage rates but also on bias. In Table 2, some results are presented from the simulations for the seasonal model. Similarly, Tables 3 and 4 show the corresponding results for the nonseasonal model, with R=4 and R=3, respectively.

Table 2: Observations below\above of a 100(1-α)% nominal confidence interval for λ and actual α, with Model 1 and R=12 (1000 samples).

H	α	z_{nom}	λ: 0	λ: 0.5	λ: 1	Actual α average	z_{act}
6	0.005	2.575	5\5	9\13	4\10	0.015	2.170
	0.045	1.695	16\29	25\26	24\26	0.049	1.651
	0.050	1.641	18\32	26\26	26\26	0.051	1.635
	0.125	1.150	44\55	52\50	45\45	0.097	1.299
	0.130	1.128	44\58	56\53	49\47	0.102	1.270
12	0.040	1.750	3\6	11\1	3\4	0.009	2.365
	0.045	1.695	4\6	11\2	3\4	0.010	2.326
	0.195	0.860	19\25	35\23	28\19	0.050	1.645
	0.200	0.841	21\27	36\25	29\19	0.052	1.626
	0.295	0.539	40\48	61\50	59\40	0.099	1.289
	0.300	0.521	42\48	63\53	60\41	0.102	1.270
20	0.125	1.151	5\2	9\4	3\4	0.009	2.365
	0.130	1.128	5\3	9\4	5\4	0.01	2.326
	0.295	0.539	30\20	28\16	29\24	0.049	1.651
	0.300	0.521	32\20	31\16	29\26	0.051	1.635
	0.365	0.341	40\39	56\45	51\54	0.095	1.310
	0.370	0.330	43\42	57\47	57\59	0.102	1.270
30	0.190	0.879	5\1	12\2	8\2	0.010	2.326
	0.195	0.860	6\1	14\2	8\2	0.011	2.290
	0.335	0.421	18\17	37\15	30\24	0.047	1.679
	0.340	0.411	18\18	40\16	33\26	0.050	1.645
	0.395	0.251	37\35	76\25	60\58	0.097	1.300
	0.400	0.251	38\36	81\28	61\61	0.102	1.270

Table 3: Observations below\above of a 100(1-α)% nominal confidence interval for λ and actual α, with Model 2 and $R = 4$ (1000 samples).

H	α	z_{nom}	λ: 0	λ: 0.5	λ: 1	Actual α average	z_{act}
6	0.005	2.575	8\9	7\12	9\13	0.019	2.075
	0.040	1.752	19\23	14\38	20\31	0.048	1.665
	0.045	1.695	21\23	14\46	23\36	0.054	1.607
	0.115	1.202	39\45	24\82	42\60	0.097	1.299
	0.120	1.175	39\46	24\87	43\66	0.102	1.270
12	0.035	1.812	1\3	0\10	7\6	0.009	2.365
	0.40	1.751	1\4	0\12	7\7	0.010	2.326
	0.165	0.974	10\19	11\61	20\28	0.050	1.645
	0.170	0.954	10\19	12\63	20\28	0.051	1.635
	0.245	0.693	19\41	20\113	37\65	0.098	1.353
	0.250	0.675	22\48	21\119	38\68	0.105	1.254
20	0.075	1.434	0\0	1\9	4\11	0.008	2.410
	0.080	1.405	1\0	1\10	4\13	0.010	2.326
	0.220	0.772	6\4	10\56	20\54	0.050	1.645
	0.225	0.755	6\5	10\58	20\57	0.052	1.626
	0.305	0.510	13\21	19\111	34\97	0.098	1.353
	0.310	0.496	14\21	20\112	35\101	0.101	1.275
30	0.105	1.254	0\0	2\17	1\9	0.010	2.326
	0.110	1.227	0\0	2\17	3\10	0.011	2.290
	0.270	0.611	2\3	11\66	16\53	0.050	1.645
	0.275	0.599	2\3	11\68	18\57	0.053	1.619
	0.340	0.411	7\10	24\117	37\94	0.096	1.308
	0.345	0.400	9\11	27\124	39\95	0.102	1.270

In these tables, the nominal confidence levels of the intervals were selected by trial and error. That is, we increased the confidence level by an amount of 0.005 units and looked for the levels that yield actual coverage rates of 99%, 95% and 90%, which are the most commonly used in practice. The *actual* α values were obtained by averaging over the different coverage rates obtained for λ=0,0.5,1. The group size R=12 was used for the monthly seasonal series because this is the usual practice. There is no commonly accepted value for nonseasonal time series. For instance, Guerrero's (1993) advice was to employ R=2 in order to minimize the loss of information by grouping. However, with this choice the estimation of variability required is very poor and perhaps a value R>2 could perform better. By looking at Table 2 it is reasonably clear that H=6 serves to obtain actual confidence levels similar to the nominal ones.

In Tables 3 and 4, the value of R was sought that makes the method work well also for H=6, when the series is nonseasonal. It was found that R=4 is preferable to R=3 in terms of having less bias and more comparable results for the different λ values. However, in Tables 2, 3 and 4, the value of the estimated autocorrelation coefficient was not considered, because it was not under our control. The simulations were carried out with the statistical package S-Plus 2000 (MathSoft, Inc.).

On the basis of these simulations, it was concluded that the nominal confidence level depends on the following factors: (i) the actual confidence level, (ii) the value of H, and (iii) the value of R. Thus, in order to calibrate the confidence intervals we estimated the following linear regression model (standard errors in parentheses) with $\overline{R}^2 = 0.9503$, $\hat{\sigma} = 0.1265$ and sample size=69

$$z_{nom} = \underset{(0.0932)}{0.8845} - \underset{(0.0038)}{0.0200}R - \underset{(0.0091)}{0.1426}H + \underset{(0.0002)}{0.0028}H^2 + \underset{(0.0361)}{0.9838}z_{act}$$

This result indicates that the Normal approximation derived previously requires a statistically significant numerical correction. With this equation, the appropriate z_{nom} may be calculated, given the values of R and H, as well as the desired z_{act}, corresponding to the actual confidence level. Such a nominal value can then be introduced in expression (2) to obtain an appropriate confidence interval.

Illustrative Applications

The Sales dataset corresponds to the seasonal time series provided by Chatfield and Prothero (1973). The original series has N=77 observations on sales of an engineering firm. A time plot of the series without transformation appears in Figure 2(a) and power-transformed with λ=0.254 in Figure 2(b). This transformation index was obtained as minimizer of the coefficient of variation with H=6 subseries and R=12 observations per subseries (so that n=5 observations were left out of the calculations). In this case the autocorrelation required by the confidence interval was estimated as $\hat{\rho} = 0.2554$.

The following confidence intervals were obtained for the true λ value. 99%: (-0.0594,0.5646); 95%: (0.0216,0.4846); and 90%: (0.0616,0.4456). Figure 1 shows a graph of the coefficient of variation CV(λ) for these data, together with a 95% confidence interval for λ. Thus, with a confidence level of 95%, it can be determined that λ=0 is not supported by the data as the index of a variance stabilizing power transformation. In other words, the logarithm is not a reasonable transformation to stabilize the variance of this time series. However, values such as λ=0.25 or λ=0.34, are reasonably adequate to represent the true value of λ, even with 90% confidence. This result is in agreement with the basic conclusion reached by previous authors (see Guerrero, 1993).

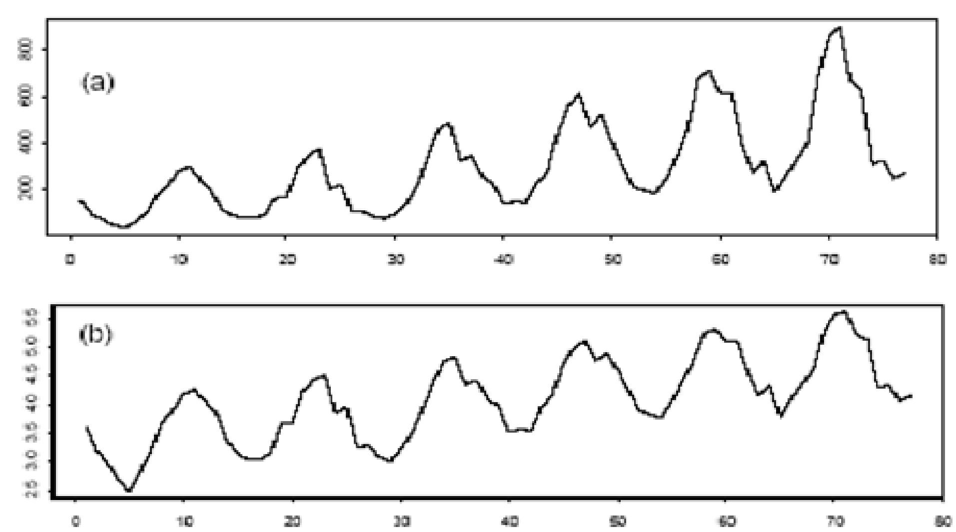
Now, for comparative purposes, assume that no autocorrelation exists in the series $W_h(\hat{\lambda}) = S_h / \overline{Z}_h^{1-\hat{\lambda}}$, for h=1,…,H, in such a way that Vangel´s (1996) proposal can be used. In this situation, the 100(1 - α) % confidence interval is given by

$$\left\{ \lambda : \frac{\sigma}{\mu} \leq CV\left(\hat{\lambda}\right) \left[\frac{\left(\frac{\chi^2_{H-1,1-\alpha} + 2}{H} - 1 \right)}{CV^2\left(\hat{\lambda}\right) + \frac{\chi^2_{H-1,1-\alpha}}{H-1}} \right]^{-1/2} \right\}$$

Table 4: Observations below\above of a 100(1-α)% nominal confidence interval for λ and actual α, with Model 2 and R=3 (1000 samples).

			λ			Actual α	
H	α	z_{nom}	0	0.5	1	average	z_{act}
6	0.005	2.575	1\5	2\12	7\11	0.013	2.229
	0.045	1.695	15\12	7\54	18\39	0.048	1.665
	0.050	1.641	15\13	7\55	21\40	0.050	1.645
	0.125	1.150	29\31	19\103	35\74	0.097	1.299
	0.130	1.128	31\31	21\105	37\78	0.101	1.279
12	0.040	1.750	1\0	0\17	3\8	0.010	2.326
	0.045	1.695	1\0	0\19	4\12	0.012	2.259
	0.150	1.032	5\5	12\61	16\44	0.048	1.669
	0.155	1.011	5\6	15\63	19\45	0.051	1.635
	0.250	0.671	9\21	23\118	42\85	0.099	1.289
	0.255	0.660	10\22	25\120	42\89	0.103	1.268
20	0.080	1.405	0\0	3\16	6\5	0.010	2.326
	0.085	1.370	0\0	3\17	6\7	0.011	2.290
	0.225	0.755	2\2	11\66	20\42	0.048	1.669
	0.2300.	0.740	3\2	11\71	20\405	0.051	1.635
	0.315	0.480	13\10	25\120	39\91	0.099	1.289
	0.320	0.469	15\10	25\125	40\96	0.104	1.252
30	0.080	1.405	0\0	4\15	4\8	0.010	2.326
	0.085	1.370	0\0	4\17	4\11	0.012	2.259
	0.235	0.721	0\0	7\71	23\46	0.049	1.651
	0.240	0.709	0\0	7\74	23\48	0.051	1.635
	0.325	0.451	6\4	18\127	37\105	0.099	1.289
	0.330	0.440	7\5	19\132	38\113	0.105	1.254

Figure 2. Sales data. (a) Original and (b) power-transformed with λ=0.254.



Because $\hat{\lambda} = 0.254$, $CV(\hat{\lambda}) = 0.0838$, H=6 and α=0.05, then $\chi^2_{5,0.95} = 1.15$. The confidence interval gets defined by the λ values satisfying the inequality σ/μ ≤ 0.1708, where it should be recalled that both σ and μ are functions of λ. Hence, (see Figure 1) the 95% confidence interval obtained is: (-0.0797,0.5717). The corresponding interval (2) on the assumption ρ=0, satisfies the inequality σ/μ≤0.1519 and becomes (-0.0204,0.5266). Both intervals obtained on the no-autocorrelation assumption cover the value λ=0, but Vangel's interval is wider than ours. In this exercise, the autocorrelation coefficient changed from being a negative value to zero, leaving everything else constant. This change produced a larger expanding factor of CV(λ), hence a wider interval.

Blowfly Data

Nicholson's blowfly data have been analyzed from several angles. Notably among these is the one that employs a nonlinear model for these data, in place of a power transformation (see Young, 2000). Nevertheless, because we are mainly concerned with the use of power transformations, we emphasize the analysis presented in the paper by Chen and Lee (1997). These authors used 82 observations of the original series (from 218 to 299) for comparison with previous works. They also mentioned that other authors used either a logarithmic or a square root transformation (i.e. λ=0 or λ=0.5). Then, they employed their method, conditioning on an autoregressive AR(1) model form, and made inferences on both λ and the parameters of that model (mean, autoregressive coefficient and error variance).

The point estimate of the transformation index was obtained as the posterior mean of a distribution obtained by Gibbs sampling with a uniform prior on the set {0.30, 0.31, ..., 0.50} The estimated value, $\hat{\lambda} = 0.39$ with standard error 0.001, clearly differs significantly from λ=0 and λ=0.5. However, we believe that Chen and Lee´s method is misleading because it conditions on the model form, while the other methods against which they compared their results are model-independent. Moreover, it should be recalled that the model form may change depending on the value of λ, as indicated by Gourieroux and Jasiak (2002), thus the AR(1) specification might be in doubt.

We applied our procedure to the data employed by Chen and Lee, without conditioning on any given model structure. By so doing, $\hat{\lambda}$=0.3997, with R=4 and H=20; so that n=2 observations were not used. The point estimate of the transformation index took almost the same value as that obtained by Chen and Lee's method. The autocorrelation became in this case $\hat{\rho}$=0.0215 and the confidence intervals were 99%: (-1.0448,1.6272); 95%: (-0.6048,1.2892) and 90%: (-0.3328,1.0682). These intervals are inconclusive, because even with 90% confidence using the data in the original scale, in a square root scale or in logarithms, produces essentially the same results (in terms of variance stabilization). This result would have been expected just by looking at the graphs shown in Figure 3, where no relevant changes are observed in the time series behavior by changing the scale. We calculated again the interval proposed by Vangel (1996) on the assumption that ρ=0 (which may be deemed reasonable since $\hat{\rho}$ is indeed close to zero) with $\hat{\lambda}$=0.3997, CV(λ)=0.54794, H=20 and α=0.05, so that $\chi^2_{19,0.95}$ = 10.12. The corresponding 95% confidence interval was defined by the λ values satisfying the inequality σ/μ≤0.851204, (see the graph of CV(λ) in Figure 4) that is (-1.7678,2.1393). Thus, the previous conclusion holds valid even if the assumption ρ=0 were true.

Similarly, the graph of CV(λ) shown in Figure 4 shows why the intervals are so wide: CV(λ) is extremely flat for the range of usual λ values employed in practice. This is an example where the data are basically insensitive to the choice of a variance stabilizing transformation. To test this idea, we estimated the same AR(1) model for the data with the following choices of the transformation index: λ=1,0.39,0.

Figure 3. Blowfly data. (a) Original, (b) power-transformed with λ=0.3997, and (c) log-transformed.

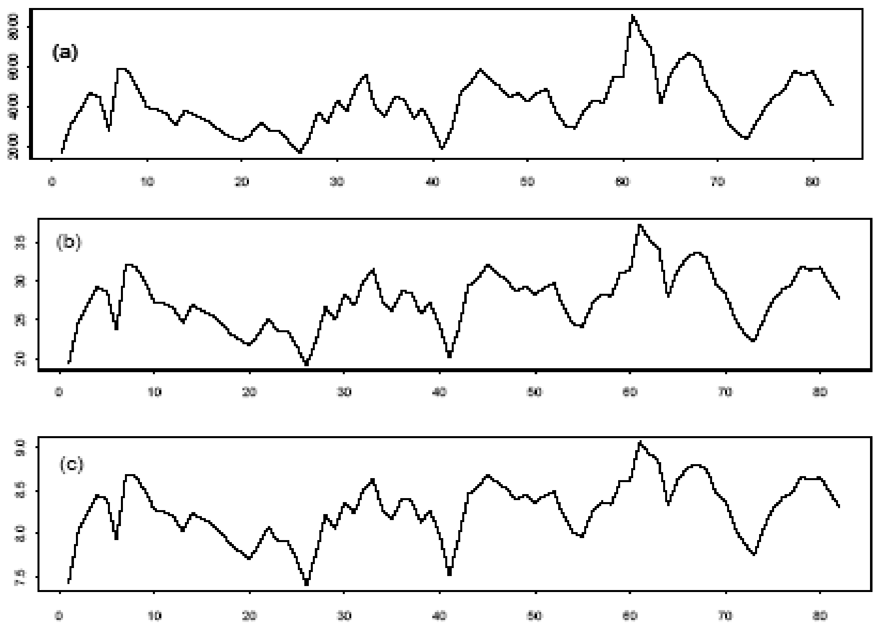
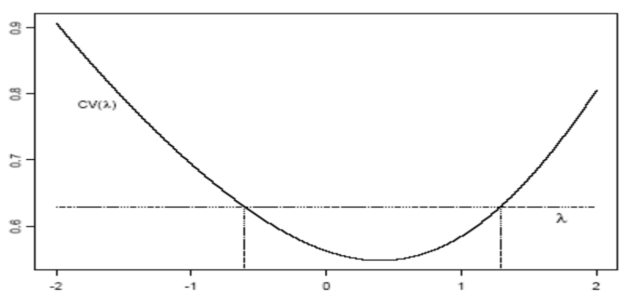


Figure 4. Confidence interval for λ with blowfly data.



The Maximum Likelihood estimation results appear in Table 5, where it may be observed that the estimated AR coefficients $\left(\hat{\phi}_1\right)$ are almost the same in the three different scales. The Ljung-Box statistics Q(24-1), when compared against a Chi-square distribution with 23 degrees of freedom, show no evidence of inadequacy.

The other two estimated parameters (mean $\hat{\phi}_0$ and residual standard error $\hat{\sigma}$) depend heavily on the scale of the analysis and do not allow a direct comparison. The *t*-statistics indicate that the estimated coefficients are significantly different from zero in the three cases and the residual graphs (not shown) are also very similar, showing no evidence of nonconstant variance by visual inspection. Thus, it may be concluded that choosing one particular power transformation, within those indexed by λ=1,0.39,0, depends on some criterion different from variance stabilization. Perhaps, the forecasting ability of the model should be studied in the different scales, as Chen and Lee (1997) finally did, in order to select the λ value, but that task was outside the scope of this article.

Table 5. Estimation results of the AR(1) model for blowfly data with different choices of λ.

λ	1	0.39	0
$\hat{\phi}_0$	4249.83	63.448	8.311
t - stat	11.35	28.27	96.63
$\hat{\phi}_1$	0.735	0.726	0.712
t - stat	10.09	10.07	9.83
$\hat{\sigma}$	890.83	5.500	0.221
Q(24 – 1)	12.08	11.15	11.92

Conclusion

This article presents a procedure to calculate a confidence interval for the true index of a power transformation that best stabilizes the variance of a time series. This is useful as it enables a time series analyst to make statistical inferences about the transformation index, without relying on a model-dependent method. The procedure was derived from a study of the approximate mean and variance of the minimum coefficient of variation employed for choosing the transformation. Then, a small simulation study allowed us to calibrate the confidence coefficient. This calibration was justified because our analytical results were derived from several approximations that may yield inaccurate results in practical applications.

The coverage rates were found to be dependent on the nominal size of the confidence level, the subseries size R and the number H of subseries used. The simulations led to practical conclusions. For instance, the appropriate subseries size, when there is no seasonality in the time series, was found to be R = 4, while the length of the seasonal period is adequate for a seasonal time series (i.e. R = 12 for a monthly time series). A more extensive simulation study would be required to consider negative λ values as well as some other time series models, in order to get more conclusive results.

The empirical illustrations provided evidence on the use the method may have in practical applications. The first example provided an empirical confirmation that our method can be trusted, because we obtained essentially the same results that were established previously by means of Maximum Likelihood. However, our method was applied with less effort, and we did not rely on knowledge of the model structure of the time series, as is required by the Maximum Likelihood method. The second illustration tested the recommendations derived from the simulation study. In fact, it was found that our method led to sensible results and it is relatively easy to apply it.

Finally, it is interesting to note that the confidence interval for the minimum coefficient of variation can also be used to construct confidence intervals for any coefficient of variation. Therefore, the results obtained here may lead to further research in the area of inference for a coefficient of variation in general.

References

Box, G. E. P, Jenkins, G. M. & Reinsel, G. C. (1994). *Time series analysis. Forecasting and control*, 3rd ed. London: Prentice-Hall.

Chatfield C., & Prothero, D. L. (1973). Box-Jenkins seasonal forecasting: problems in a case-study. *Journal of the Royal Statistical Society*, *A-136,* 295-315.

Chen, C. W. S., & Lee, J. C. (1997). On selecting a power transformation in time-series analysis. *Journal of Forecasting*, *16*, 343-354.

de Bruin, P., & Franses, P. H. (1999). Forecasting power-transformed time series data. *Journal of Applied Statistics, 26*, 807-815.

Gourieroux, Ch., & Jasiak, J. (2002). Nonlinear autocorrelograms: An application to inter-trade durations. *Journal of Time Series Analysis*, *23*, 127-154.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations, *Journal of. Forecasting*, *12*, 37-48.

Jennings, D. E. (1987). How do we judge confidence-interval adequacy? *The American Statistician*, *41*, 335-337

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics*, *1*, 5th ed. London: Charles Griffin & Co. Ltd.

Vangel, M. G. (1996). Confidence intervals for a Normal coefficient of variation. *The American Statistician*, *15*, 21-26.

Young, P. C. (2000). Stochastic, dynamic modelling and signal processing: Time variable and state dependent parameter estimation. In *Nonstationary and nonlinear signal processing*, ed. W. J. Fitzgerald, A. Walden, R. Smith, & P. C. Young, pp. 74-114. Cambridge: University Press.

Appendix

1. Approximate variances and covariances of functions of random variables

1) *Theorem*. Let $\mathbf{X} = (X_1,...,X_k)'$ be a k-dimensional random vector, g($\mathbf{X}$) be a real-valued function defined on $\mathbf{R}^k$ and $E(X_i) < \infty$ for $i = 1,\ldots,k$. Assume that the partial derivatives $g'_i(\mathbf{X}) = \partial g(\mathbf{X})/\partial X_i$ all exist and let $g'_i[E(\mathbf{X})]$ denote $g'_i(\mathbf{X})$ evaluated at E($\mathbf{X}$). Then, the first-order Taylor expansion $g(\mathbf{X}) \approx g[E(\mathbf{X})] + \sum_{i=1}^{k} g'_i[E(\mathbf{X})][X_i - E(X_i)]$, so that $E[g(\mathbf{X})] \approx g[E(\mathbf{X})]$ and, if not all the $g'_i[E(\mathbf{X})]$ are zero, $\mathrm{var}[g(\mathbf{X})] \approx \sum_{i=1}^{k}\{g'_i[E(\mathbf{X})]\}^2 \mathrm{var}(X_i) + \sum_{i\neq j}^{k}\sum_{j=1}^{k} g'_i[E(\mathbf{X})]g'_j[E(\mathbf{X})]\mathrm{Cov}(X_i, X_j)$

Similarly, for two functions $g_1(\mathbf{X})$ and $g_2(\mathbf{X})$, $\mathrm{cov}[g_1(\mathbf{X}), g_2(\mathbf{X})] \approx \sum_{i=1}^{k} g'_{1i}[E(\mathbf{X})]g'_{2i}[E(\mathbf{X})]\mathrm{var}(X_i)$ $+ \sum_{i\neq j}^{k}\sum_{j=1}^{k} g'_{1i}[E(\mathbf{X})]g'_{2j}[E(\mathbf{X})]\mathrm{Cov}(X_i, X_j)$.

Proof. This result was established by Stuart and Ord (1987, Ch. 10).

2. Expected values, variances and covariance of m and se.

It is known that $E[W_h(\lambda)] = \mu$, $\mathrm{var}[W_h(\lambda)] = \sigma^2$ and $\mathrm{corr}[W_h(\lambda), W_{\ell'}(\lambda)] = \rho$ if $h' = h \pm 1$ and zero otherwise. Then,

1) $E(m) = \mu$,

2) $$\mathrm{var}(m) = E\left\{H^{-2}\sum_{h=1}^{H}\sum_{\ell'=1}^{H}\left[W_h(\lambda) - \mu\right]\left[W_{h'}(\lambda) - \mu\right]\right\}$$

$$= H^{-2}\left(\sum_{h=1}^{H} E[W_h(\lambda) - \mu]^2 + 2\sum_{h=1}^{H-1} E\{[W_h(\lambda) - \mu][W_{h+1}(\lambda) - \mu]\}\right) = \sigma^2[1 + 2\rho(H-1)/H]/H,$$

3) $$E[(H-1)se^2] = E\left\{\sum_{h=1}^{H} E[W_h(\lambda) - \mu]^2 - H(m-\mu)^2\right\}$$

$$= H\sigma^2 - H\sigma^2[1 + 2\rho[H-1]/H]/H = (H-1)\sigma^2(1 - 2\rho/H).$$

Under Normal theory, with $\rho = 0$, the distribution of $(H-1)se^2/\sigma^2$ is Chi-square with H-1 degrees of freedom. Since ρ cannot be far away from zero, it follows that $(H-1)se^2/\sigma^2$ must have a distribution close to a χ^2_{H-1}. The variance of such a distribution is derived by assuming approximately valid the following relationship that holds for a Chi-square distribution: *Variance* = 2 *Mean*, therefore

$\mathrm{var}\left[(H-1)se^2/\sigma^2\right] \approx 2(H-1)(1-2\rho/H)$. From the Theorem in Appendix 1 with k = 1, $X = se^2$ and $g(X) = X^{1/2}$, $\mathrm{var}(se) \approx \left\{g'\left[E\left(se^2\right)\right]\right\}^2 \mathrm{var}\left(se^2\right) \approx \left[\frac{1}{2}E^{-1/2}(se^2)\right]^2 2\sigma^4\left(1-2\rho/H\right)/(H-1) = \frac{1}{2}\sigma^2/(H-1)$,

hence,

4) $E^2(se) = E\left(se^2\right) - \mathrm{var}(se) \approx \sigma^2\left\{1-2\rho/H - 1/[2(H-1)]\right\}$.

Next, the Theorem in Appendix 1 applied with k=2, $\mathbf{X} = \left(se^2, m\right)$, $g_{11}(\mathbf{X}) = se$, $g_{12}(\mathbf{X}) = 0 = g_{21}(\mathbf{X})$ and $g_{22}(\mathbf{X}) = m$, allows

$\mathrm{cov}(se, m) \approx g'_{11}\left[E\left(se^2, m\right)\right]g'_{22}\left[E\left(se^2, m\right)\right]\mathrm{cov}(se^2, m) = \frac{1}{2}E^{-1/2}(se^2)\,\mathrm{cov}\left(se^2, m\right)$. Now,

$$\mathrm{Cov}(se^2, m) = E\left[\frac{1}{H-1}\sum_{h=1}^{H}\left\{\left[W_h(\lambda)-\mu\right]-(m-\mu)\right\}^2(m-\mu)\right]$$

$$= \frac{1}{H-1}E\left\{\sum_{h=1}^{H}\left[W_h(\lambda)-\mu\right]^3 + \sum_{h\neq h'}^{H}\sum_{h'=1}^{H}\left[W_h(\lambda)-\mu\right]^2\left[W_{h'}(\lambda)-\mu\right]\right\} - \frac{H}{H-1}E(m-\mu)^3$$

$$= \frac{1}{H-1}\left\{\sum_{h=1}^{H}E\left[W_h(\lambda)-\mu\right]^3 - HE(m-\mu)^3\right.$$

$$\left. + \sum_{h=1}^{H}E\left(a_h^2 a_{h+1} - \theta a_{h-1}a_h a_{h+1} + \theta^2 a_{h-1}^2 a_{h+1} - \theta a_h^3 + \theta a_{h-1}^2 a_h^2 - \theta^3 a_{h-1}^2 a_h\right)\right\}$$

then the normality assumption implies that the third central moments of a_h, $W_h(\lambda)$ and m are all zero. It follows that $\mathrm{cov}(se^2, m) = 0$ and **5)** $\mathrm{cov}(se, m) \approx 0$.