Journal of Forecasting, Vol. 12, 37–48 (1993)

Time-series analysis supported by Power Transformations

VICTOR M. GUERRERO
Instituto Tecnológico Autónomo de México, México

ABSTRACT

This paper presents some procedures aimed at helping an applied time-series analyst in the use of power transformations. Two methods are proposed for selecting a variance-stabilizing transformation and another for bias-reduction of the forecast in the original scale. Since these methods are essentially model-independent, they can be employed with practically any type of time-series model. Some comparisons are made with other methods currently available and it is shown that those proposed here are either easier to apply or are more general, with a performance similar to or better than other competing procedures.

KEY WORDS ARIMA models Bias reduction Forecasting Taylor series approximation Time-series models Variance-stabilizing

INTRODUCTION

The applicability of statistical models is often enhanced through the use of power transformations, and time-series models are no exception. However, it is important to select an adequate transformation as was suggested by several discussants of the paper by Chatfield and Prothero (1973), and we should bear in mind that the use of transformations does not necessarily improve forecasts (we refer the reader to the discussion of this topic in Granger and Newbold, 1986, p. 119). Once a model has been constructed in a transformed scale, the forecasts obtained in that metric may need to be retransformed. The problem is that this retransformation procedure induces bias in the forecasts.

The plan of this article is as follows. The following section reviews some procedures related to the use of transformations in time-series analysis. Then, we suggest two procedures for selecting a variance-stabilizing power transformation. Approximate correction of the bias is then considered and a debiasing factor is obtained in closed form, to be readily usable in applied work. Finally, we present an empirical illustration of the methods suggested here and present some conclusions.

0277–6693/93/010037–12$11.00

Received January 1991
Accepted March 1992

SUMMARY OF BASIC PROCEDURES

Procedures for selecting a power transformation

Box and Jenkins (1970) suggest using a power transformation to obtain an adequate autoregressive integrated moving average (ARIMA) model for the time series at hand. In particular, they advocate the use of the Box and Cox (1964) method, which selects a transformation from the family

$$\begin{aligned} Z_t^{(\lambda)} &= (Z_t^{\lambda} - 1)/\lambda, \qquad && \lambda \neq 0 \\ &= \log(Z_t), && \lambda = 0 \end{aligned} \tag{1}$$

where $\{Z_t\}$ is an observed time series with $Z_t > 0$, λ is the index of the transformation and $\log(\cdot)$ denotes natural logarithm. This family of transformations is employed rather than the usual one, given by

$$\begin{aligned} T(Z_t) &= Z_t^{\lambda}, \qquad && \lambda \neq 0 \\ &= \log(Z_t), && \lambda = 0 \end{aligned} \tag{2}$$

to avoid the discontinuity at $\lambda = 0$ and to preserve the original order in the data.

The Box–Cox method maximizes a likelihood function, which depends on λ and the unknown parameters of the model. This method validates the model as a whole and, in particular, helps to stabilize the variance of the series. In practice, one either searches for the power which maximizes the likelihood function over a grid of λ values or uses the algorithm of Ansley *et al.* (1977) for direct implementation of the method. This algorithm has not gained wide acceptance in applied work, mainly because the autocorrelation structure of the series depends on the current value of λ (see Granger and Newbold, 1986, p. 118). Thus the ARIMA model form may change during the search. Besides, a method that is relatively cheap and easy to implement is still required to obtain an initial value of λ for initializing the algorithm. Therefore if an 'easy-to-use' method were available to produce an approximate value of λ, such a value might be deemed good enough for practical purposes, depending on the sensitivity of the data to the choice of transformation.

Within the context of unobserved component models of trend, seasonality and irregularity, an unstable seasonal component is commonly associated with nonconstant variance of the series. This idea comes from the fact that in a two-way table with one observation per cell (in the present case a month by year table) it is impossible to distinguish between nonhomogeneous variance and interaction from the observed data alone. However, some other modeling possibilities for taking account of increasing seasonality, besides data transformation, were suggested by Bowerman *et al.* (1990). Durbin and Murphy (1975) proposed a method for discriminating among the additive, multiplicative or mixed additive–multiplicative models. That method is related to the problem of selecting a variance-stabilizing power transformation by linking the additive model with $\lambda = 1$, the multiplicative model with $\lambda = 0$, and the mixed model with $\lambda \neq 0, 1$.

The 'amplitude-trend diagram' proposed by Durbin and Kenny (1978) is a graphical procedure similar in nature to the 'range-mean plot' mentioned below. On the other hand, Cleveland *et al.* (1978) search for a power transformation that yields an additive decomposition of the series and requires a preliminary estimate of both the trend and seasonal components. Similarly, Harvey and Fernandes (1989) considered the problem of building structural time-series models for count or qualitative data. They use a specific variance-stabilizing transformation for a given distribution, as is the case of the square-root transformation for Poisson data.

The above-mentioned methods require previous identification of a model for selecting the transformation parameter λ. That is not the case with the procedure associated with the 'range-mean plots' (see Jenkins, 1979) in which the mean and range of some subsets of the series are plotted to observe if the range is independent of the mean ($\lambda = 1$), if they are linked through a linear relationship ($\lambda = 0$) or if some other pattern is present. The drawbacks of this procedure are its inaccuracy and the subjectivity involved in the visual inspection of the plots. Nevertheless, as one reviewer pointed out, most experienced analysts tend to rely on the visual inspection of data plots. That is why one of the most frequently employed procedures consists of applying a power transformation and looking at the results for different λ values. If those values are arbitrarily chosen, this approach becomes very time consuming and a data-based procedure is preferable.

Procedures for reducing transformation bias

It is well known (e.g. Granger and Newbold, 1976) that optimal forecasts obtained for a power-transformed series do not retain their optimal properties when brought back to the original scale. This happens because an estimated mean of a symmetric distribution in the transformed scale becomes an estimated median after an application of the inverse transformation. Some papers that propose methods for correcting the bias when trying to estimate an expected value in the original scale, are those of Neyman and Scott (1960), Miller (1984), and Taylor (1986). Although those authors did not consider explicitly time-series models, their solutions can be easily adapted to these models.

The solution of Neyman and Scott is based on an expansion method which becomes too complex to be applied in practice. Miller's solution is much simpler, but applies only when λ is a fractional power, and Taylor's is also a simple solution, but lacks accuracy when λ is near zero. On the other hand, Granger and Newbold (1976) as well as Pankratz and Dudley (1987), addressed specifically the time-series case and devised procedures for obtaining debiasing factors, which are either too complicated or do not admit a closed-form expression except when λ is a fractional positive power.

SELECTION OF A VARIANCE-STABILIZING POWER TRANSFORMATION

Let $X > 0$ be a random variable with mean $E(X) < \infty$ and variance $\mathrm{Var}(X)$ which depends on the value of the mean. To find a variance-stabilizing power transformation for X, we expand $T(X)$ about $E(X)$, so that a linear approximation yields

$$\mathrm{Var}[T(X)] = \{T'[E(X)]\}^2 \, \mathrm{Var}(X) \tag{3}$$

where T$'$ denotes the derivative of T. Then, we deduce that the power λ associated with a variance-stabilizing transformation must satisfy

$$[\mathrm{Var}(X)]^{1/2} / [E(X)]^{1-\lambda} = a \tag{4}$$

for some constant $a > 0$. Evidently, in practical applications we require either knowing the theoretical values of $E(X)$ and $\mathrm{Var}(X)$ or being able to estimate these parameters from observations on the variable X.

Let us now consider an observed time series $\{Z_t\}$, probably with nonconstant variance. In order to stabilize its variance, one should select the power λ which keeps constant the relationship (4) for every observation of the series $t = 1, \dots, N$. A problem with time-series data when only one observation is made at each time t is that we are unable to estimate dispersion. To solve this problem, we can group the N observations of the series into H subseries, so that a local estimate of mean and variance within each subseries can be obtained. The idea is to have several subseries in order to obtain pairs of values $(\bar{Z}_h, S_h)$ which are comparable for $h = 1, \dots, H$. Thus we should try to keep homogeneity between those subseries by: (1) eliminating n observations $(0 \leqslant n < R)$, either from the beginning or the end of the original series, in such a way that all the subseries result with the same number, $R = (N - n)/H$, of observations and (2) choosing appropriately the size of the subseries, preferably of the same length as the seasonality.

Let $Z_{h,r}$ be the rth observation of subseries h, We then have

$$\bar{Z}_h = \sum_{r=1}^{R} Z_{h,r}/R, \qquad S_h = \left[\sum_{r=1}^{R} (Z_{h,r} - \bar{Z}_h)^2/(R-1) \right]^{1/2} \tag{5}$$

and λ should be chosen in such a way that

$$S_h/\bar{Z}_h^{1-\lambda} = a, \qquad h = 1, \dots, H \tag{6}$$

holds for some constant $a > 0$.

Comment

When the series under study shows seasonal variation, selecting the subseries of equal length as the seasonality is equivalent to working with a two-way table and the underlying structure of the series can be related to the classical decomposition model. In that case, we could replace $\bar{Z}_h$ by an estimate of mean trend and S_h by another measure of dispersion expressible in the same units as $\bar{Z}_h$. Then what follows would also apply. Nevertheless, when no seasonality is present it would be advisable to make $R = 2$, so that S_h can still be estimated and the loss of information by grouping is kept to a minimum.

Minimizing relative variation

An empirical interpretation of equation (6) leads us to look for a λ value such that the ratios $S_h/\bar{Z}_h^{1-\lambda}$ show minimum variation. However, since for each different power transformation the measurement units of these ratios change, we should select that power by looking for the minimum coefficient of variation (CV) of $S_h/\bar{Z}_h^{1-\lambda}$ as a function of λ.

Estimating a simple linear regression in logarithms

Another empirical interpretation of condition (6) can be put in the form of the following simple linear regression model:

$$\log(S_h) = \log(a) + (1 - \lambda)\log(\bar{Z}_h) + \varepsilon_h, \qquad h = 1, \dots, H \tag{7}$$

with the ε_h's being a random sample of errors uncorrelated with $\log(\bar{Z}_h)$, whose mean is zero and variance σ_ε^2. This formulation was suggested by Taylor within an ecological context (see Perry, 1981). Then it is clear that λ can be estimated by least squares, and the usual regression analysis of time-series data applies.

APPROXIMATE CORRECTION OF BIAS

Let us assume that a time-series model has been built for a power-transformed series with $T(\cdot)$ as in equation (2). Then we might be interested in obtaining forecasts of the original series

$\{Z_t\}$, while the model provides forecasts for $\{T(Z_t)\}$. Typically, the optimal forecasts are estimated expected values in the transformed scale, where we usually assume normality. Thus, by applying the inverse tansformation $T^{-1}(\cdot)$, which is a monotonic function, we obtain estimated median values in the original scale. This retransformation to the median produces a so-called 'naive' forecast, which may be deemed optimal if the cost function associated with the forecast errors is linear and symmetric about zero (see Granger and Newbold, 1986, for details about cost functions). Nevertheless, a correction for bias is required if we really want an estimate of the expected value in the original metric. It should be mentioned that sometimes the effect of bias is negligible when the data are relatively insensitive to transformation.

An approximate debiasing factor

Now let $\mu_j = E_t[T(Z_{t+j})]$ be the optimal forecast of the transformed series, where the subscript t in the expectation operator serves to indicate conditionality on the information known up to time t. Also let $e_{t+j} = T(Z_{t+j}) - \mu_j$ be the forecast error and define $\sigma_j^2 = \mathrm{Var}_t(e_{t+j})$. Then, if $E_t(Z_{t+j})$, $j = 1, 2, \ldots$, are the forecasts we are looking for, we show in the Appendix that

$$E_t(Z_{t+j}) \doteq T^{-1}(\mu_j) \cdot C_\lambda(j) \tag{8}$$

with the approximate debiasing factor for the usual power transformation (2) given by

$$\begin{aligned} C_\lambda(j) &= \{0.5 + 0.5[1 + 2(\lambda^{-1} - 1)\sigma_j^2/\mu_j^2]^{1/2}\}^{1/\lambda}, & \lambda \neq 0 \\ &= \exp(\sigma_j^2/2), & \lambda = 0 \end{aligned} \tag{9}$$

Similarly, the approximate debiasing factor for the Box–Cox transformation becomes

$$\begin{aligned} C_\lambda(j) &= \{0.5 + 0.5[1 + 2(\lambda^{-1} - 1)\sigma_j^2/(\lambda^{-1} + \mu_j)^2]^{1/2}\}^{1/\lambda}, & \lambda \neq 0 \\ &= \exp(\sigma_j^2/2), & \lambda = 0 \end{aligned} \tag{10}$$

We should also correct for bias the following confidence interval, which corresponds to the naive forecast:

$$T^{-1}\{\hat{\mu}_j \pm z_{\alpha/2}[\hat{\mathrm{Var}}_t(e_{t+j})]^{1/2}\} \tag{11}$$

The correction consists simply in multiplying the end points of equation (11) by the debiasing factors (9) or (10), respectively. It is also important to appreciate that in the previous derivations, λ was assumed to be a known fixed value. When λ is estimated from the observed data, the variance of $\hat{\lambda}$ may affect the previous results, although this effect is generally small (see Taylor, 1986).

Comparisons with other methods

To appreciate how the suggested debiasing factor compares with the factors derived by Taylor (1986) and Pankratz and Dudley (1987) when using the Box–Cox transformation, we now present these factors adapted to our notation. Taylor's factor becomes

$$\begin{aligned} C_{(\lambda)}^{T}(j) &= 1 + 0.5\lambda^{-1}(\lambda^{-1} - 1)\sigma_j^2/(\lambda^{-1} + \mu_j)^2, & \lambda \neq 0 \\ &= 1 + \sigma_j^2/2, & \lambda = 0 \end{aligned} \tag{12}$$

and the Pankratz–Dudley factor is

$$\begin{aligned} C_{(\lambda)}^{PD}(j) &= 1 + \sum_{k=1}^{\infty} \lambda^{-1}(\lambda^{-1} - 1)\ldots(\lambda^{-1} - 2k + 2)(\lambda^{-1} - 2k + 1) \\ &\quad [0.5\sigma_j^2/(\lambda^{-1} + \mu_j)^2]^k/k! & \lambda \neq 0 \\ &= \exp(\sigma_j^2/2) & \lambda = 0 \end{aligned} \tag{13}$$

Table I. Percentage bias from use of the naive forecast as measured by the correction factors $C^{T}_{(\lambda)}$, $C^{PD}_{(\lambda)}$, and $C_{(\lambda)}$

Power	$\sigma_j/\lvert\lambda^{-1}+\mu_j\rvert$								
λ	0.05			0.15			0.25		
	$C^{T}_{(\lambda)}$	$C^{PD}_{(\lambda)}$	$C_{(\lambda)}$	$C^{T}_{(\lambda)}$	$C^{PD}_{(\lambda)}$	$C_{(\lambda)}$	$C^{T}_{(\lambda)}$	$C^{PD}_{(\lambda)}$	$C_{(\lambda)}$
2	0.0	0.0	0.0	0.3	0.3	0.3	0.8	0.8	0.8
3/2	0.0	0.0	0.0	0.3	0.3	0.3	0.7	0.7	0.7
1	0	0	0	0	0	0	0	0	0
1/2	−0.2	−0.2	−0.2	−2.2	−2.2	−2.2	−5.9	−5.9	−5.8
1/4	−1.5	−1.5	−1.5	−11.9	−12.0	−12.1	−27.3	−27.9	−28.2
−1/4	−2.4	−2.5	−2.5	−18.4	−23.1	−21.9	−38.5	[a]	−57.8
−1/2	−0.7	−0.8	−0.8	−6.3	−7.1	−6.9	−15.8	−23.8	−19.8
−1	−0.2	−0.3	−0.3	−2.2	−2.4	−2.3	−5.9	−7.7	−6.7

[a] Result declared unreliable by Pankratz and Dudley.

Here it is noticeable in particular that Taylor's factor has the simplest expression, but for $\lambda = 0$ does not provide an adequate solution. While the factor of Pankratz and Dudley is more complicated than equation (10) and it admits a closed form expression only when λ is a fractional positive power.

We observe that expressions (10), (12) and (13) depend on the value of $\sigma_j^2/(\lambda^{-1}+\mu_j)^2$, when $\lambda \neq 0$. That fact is now used to provide a numerical evaluation of the bias which results from using the median (naive forecast) when we are in fact interested in the expected value. Thus, we selected the range $-1 \leqslant \lambda \leqslant 2$ and three different values of $\sigma_j/\lvert\lambda^{-1}+\mu_j\rvert$ to measure the percentage relative bias according to the different factors. The figures in Table I were obtained by defining the percent relative bias as

$$100[T^{-1}(\mu_j) - E_t(Z_{t+j})]/E_t(Z_{t+j}) = 100(C^{-1} - 1) \tag{14}$$

where C denotes the corresponding debiasing factor. This table allows us to see that all three factors give similar results, except when λ is close to zero and the variance σ_j^2 is large relative to $(\lambda^{-1}+\mu_j)^2$.

EMPIRICAL ILLUSTRATIONS

In order to illustrate the use of the suggested methods empirically, let us consider the Sales Data studied by Chatfield and Prothero (1973). This series consists of 77 monthly observations on sales of an engineering firm. The data cover the period from January 1965 up to May 1971 and were analyzed originally in the logarithmic scale (in Figure 1 we show plots of the original and log-transformed data). Use of this transformation was criticized by Wilson (1973) who found by maximum likelihood that a more appropriate power transformation was $\hat{\lambda} = 0.34$. Similarly, Box and Jenkins (1973) used range-mean plots to suggest that $\hat{\lambda} = 0.25$ could also be an adequate power. Abraham and Ledolter (1986) also employed the Sales Data to illustrate the use of double seasonal differencing. They found results similar to those obtained with adequate data transformation.

We now apply the methods proposed here to the Sales Data. First, to determine the power which minimizes the coefficient of variation we built Table II with the ratios given by

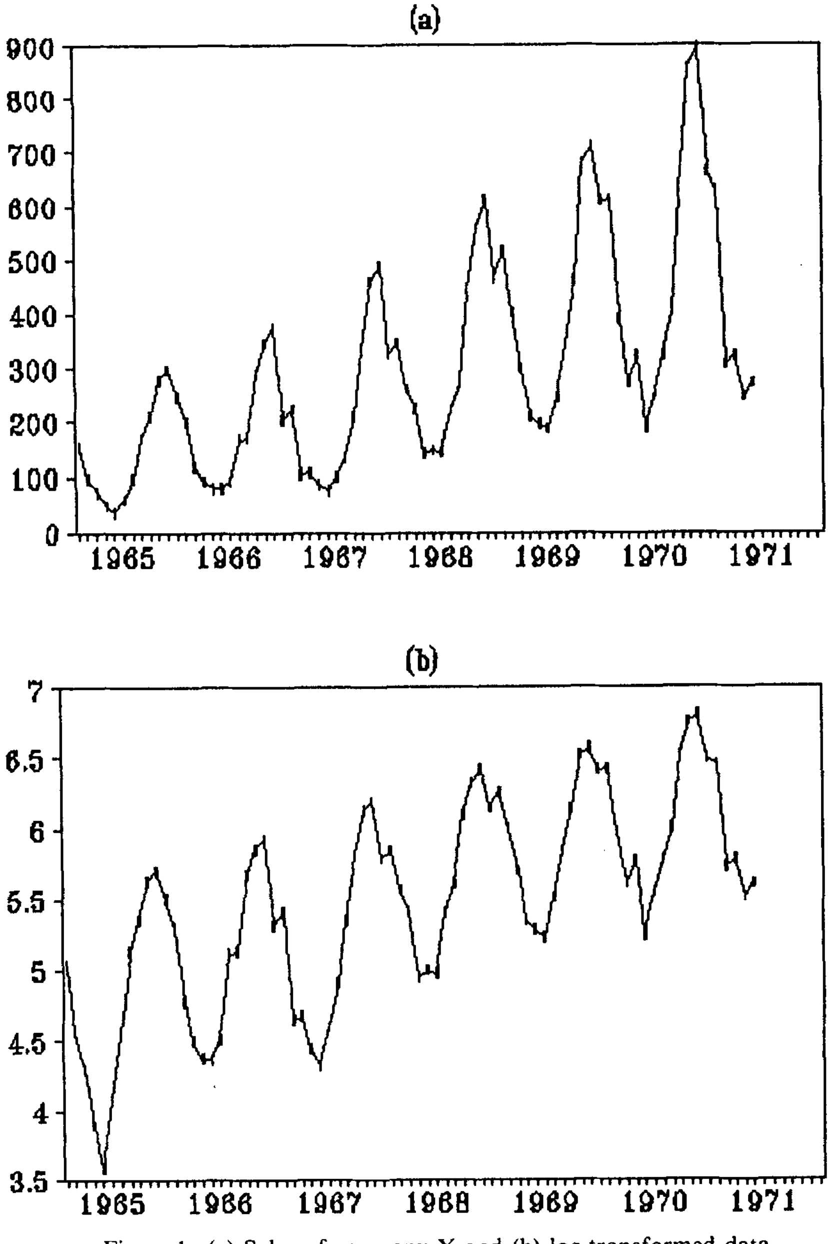

Figure 1. (a) Sales of company X and (b) log-transformed data

expression (6), in which the subseries were formed by the 12 observations within each complete year.

In Table II we can see that the minimum coefficient of variation $C\hat{V} = 0.0838$ is attained at $\hat{\lambda} = 0.254$ and Figure 2 shows a plot of the corresponding transformed data. In that plot, we can visually appreciate that the seasonal pattern becomes more stable with this choice of power than with $\lambda = 0$ or $\lambda = 1$ (see also Figure 1).

Table II. Selection of the power transformation which minimizes the coefficient of variation

	λ				
Year	0	0.25	0.254	0.34	1
1965	0.6313	2.1976	2.2419	3.4432	92.6958
1966	0.5652	2.0860	2.1300	3.3352	104.1449
1967	0.6712	2.5870	2.6434	4.2046	148.1126
1968	0.5133	2.1717	2.2224	3.6501	164.4364
1969	0.4682	2.1015	2.1526	3.6080	189.9833
1970	0.4950	2.3279	2.3862	4.0644	242.0687
m	0.5576	2.2453	2.2961	3.7176	156.9070
s.e.	0.0804	0.1883	0.1925	0.3451	55.4753
CV	0.1441	0.0839	0.0838	0.0928	0.3536

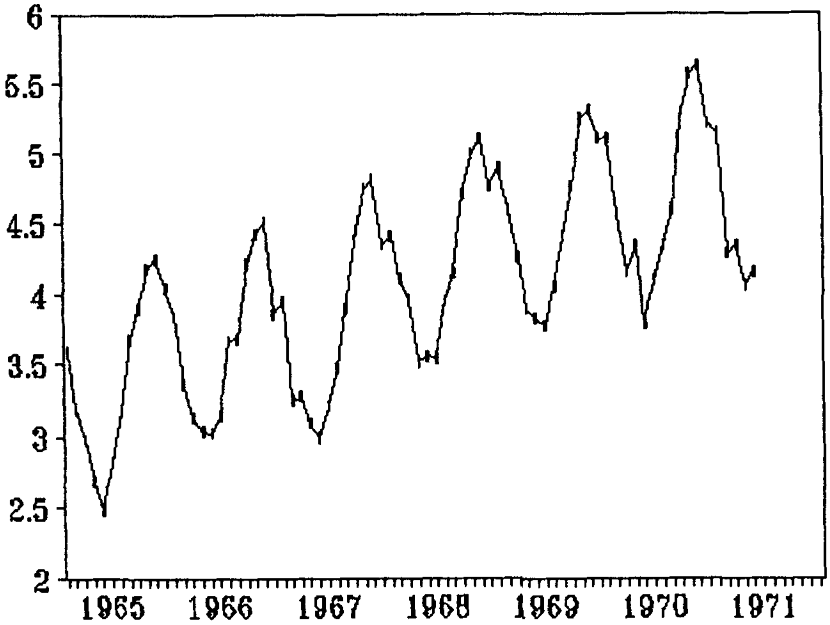

Figure 2. Power-transformed series ($\hat{\lambda} = 0.254$)

By fitting a simple linear regression in logarithms we found the estimated relationship (standard errors in parentheses)

$$\log(S_h) = \underset{(0.482)}{0.798} + \underset{(0.086)}{0.751} \log(\bar{Z}_h), \qquad h = 1, \ldots, 6$$

with $R^2 = 0.95$, $\hat{\sigma}_\varepsilon = 0.0904$, and Durbin–Watson's $d = 2.79$. Therefore, we have $\hat{\lambda} = 0.249$ with a 95% confidence interval for λ given by (0.010, 0.487), so that both $\lambda = 0.25$ and $\lambda = 0.34$ belong in this interval, while $\lambda = 0$ and $\lambda = 1$ do not.

The previous results were obtained by fixing the size of the subseries in $R = 12$ observations

Table III. Selection of λ for different choices of the subseries size

Size of the subseries (R)	Method: Minimum coefficient of variation $\hat{\lambda}$	$CV(\hat{\lambda})$	Linear regression in logarithms $\hat{\lambda}$	s.e.$(\hat{\lambda})$
2	0.209	0.5072	−0.052	0.2028
3	0.125	0.5223	−0.026	0.2292
4	0.514	0.2438	0.567	0.0972
6	0.149	0.0990	0.144	0.0493
8	0.068	0.1637	0.108	0.1045
9	0.157	0.1753	0.180	0.1266
12	0.254	0.0838	0.249	0.0859
18	0.245	0.0334	0.245	0.0427
24	0.205	0.0609	0.205	0.1204
36	0.285	0.0001	–	–

since that is the length of the seasonal period. However, to be able to appreciate the effect of different values of R on the selection of λ, we present in Table III the results obtained with the two suggested methods when R takes on values which are submultiples of 72 (the number of original data points in the 6 complete years of the series). In this table we clearly see that the procedure based on fitting a linear regression in logarithms is more sensitive to the choice of R than minimizing the coefficient of variation. On the other hand, it is important to realize that, with both procedures, the value of $\hat{\lambda}$ depends (sometimes heavily) on the choice of the subseries size. Therefore, it should be clear that R must be selected cautiously and we refer the reader to the comment that appears after equation (6).

Next, to illustrate the application of the bias-reducing factor we consider an ARIMA model for the series transformed with $\lambda = \frac{1}{3}$, as did Pankratz and Dudley (1987). This model is of the same form as model A of Chatfield and Prothero, that is,

$$(1 - \phi B)w_t = (1 - \Theta B^{12})a_t$$

with $w_t = (1 - B)(1 - B^{12})Z_t^{1/3}$, where B is the back shift operator and $\{Z_t\}$ denotes the original series. Pankratz and Dudley obtained by maximum likelihood $\hat{\phi} = -0.53$, $\hat{\Theta} = 0.54$ and then 12 forecasts from time origin $t = 65$ were produced. We also take advantage of the results provided in their Table 2 to obtain the bias-corrected forecasts shown in our Table IV.

In this particular application, Pankratz and Dudley found that the naive forecasts $\hat{\mu}_j^3$ performed slightly better than the corrected ones due to an outlier at time 74. In fact, the naive forecasts have an average forecast error of −1 for the 12 forecasts, compared to −4 for the bias-corrected forecasts with Taylor's factor and −9 with both the Pankratz–Dudley's factor and the factor suggested in this article. However, when considering only the 11 forecasts which exclude the outlying observation, we obtain an average error of 13 for the naive forecasts. With Taylor's factor the corresponding figure is 11, while it is 6 with the Pankratz–Dudley factor and 5 with ours. Thus, we obtain a slight improvement in forecast accuracy with the new factor when the outlier is excluded.

Table IV. Bias-corrected forecasts in the original scale for the Sales Data. Time origin is $t = 65$ and leads are $j = 1, \ldots, 12$

$t+j$	Z_{t+j}	$\hat{\mu}_j^3$	$\hat{\sigma}_j/\hat{\mu}_j$	$\hat{\mu}_j^3 C_{1/3}^{\mathrm{T}}(j)$	$\hat{\mu}_j^3 C_{1/3}^{\mathrm{PD}}(j)$	$\hat{\mu}_j^3 C_{1/3}(j)$
66	257	256	0.052	257	258	258
67	324	304	0.055	305	306	307
68	404	423	0.059	424	427	427
69	677	572	0.059	574	577	578
70	858	768	0.059	771	776	776
71	895	806	0.062	809	815	815
72	664	650	0.071	653	660	660
73	628	665	0.075	669	676	676
74	308	469	0.089	473	480	480
75	324	365	0.101	369	376	376
76	248	335	0.109	339	347	347
77	272	261	0.123	265	272	273

CONCLUSIONS

A word of caution related to the presence of outliers in the series is in order, since it is well known (see Atkinson, 1985) that sometimes one or two influential observations may result in a wrongly selected transformation. On the other hand, even though no explicit mention was made to time series consisting of discrete data, it should be clear that transformations are also applicable to that situation. Those transformations can be deduced when the functional relationship between mean and variance is known either theoretically or empirically. Similarly, when dealing with data measured as proportions, we suggest consideration of the odds-ratio as the original measurement and the inclusion of an extra parameter in the model to allow for other than logistic transformations, as in Guerrero and Johnson (1982).

We also note that the procedures suggested here for selecting a power transformation are, in essence, model-independent and therefore can be applied when using ARIMA, structural, or unobservable components models for time series. Similarly; to derive the approximate correction factor we did not rely on the assumption of normality, but only on Taylor's series approximations. Finally, it should be noticed that the procedures suggested here provide results which are comparable or even better than those obtained with other methods currently available, being easier to implement and/or more general.

APPENDIX

Derivation of the debiasing factors

Let $\mu_j = E_t[T(Z_{t+j})]$ and $\sigma_j^2 = \mathrm{Var}[T(Z_{t+j}) - \mu_j]$. Then an expression equivalent to equation (3) is

$$\sigma_j^2 \doteq \{T'[E_t(Z_{t+j})]\}^2 \, \mathrm{Var}_t[Z_{t+j} - E_t(Z_{t+j})] \tag{A1}$$

On the other hand, expanding $T(Z_t)$ about $E_t(Z_{t+j})$ we have, to a second-order approximation,

$$T(Z_{t+j}) \doteq T[E_t(Z_{t+j})] + T'[E_t(Z_{t+j})]\,[Z_{t+j} - E_t(Z_{t+j})] + T''[E_t(Z_{t+j})]\,[Z_{t+j} - E_t(Z_{t+j})]^2/2 \tag{A2}$$

Therefore by taking conditional expectation it follows that

$$\mu_j \doteq T[E_t(Z_{t+j})] + T''[E_t(Z_{t+j})]\,\mathrm{Var}_t[Z_{t+j} - E_t(Z_{t+j})]/2 \tag{A3}$$

Substituting equation (A1) in (A3), we therefore obtain

$$\mu_j \doteq T(E_t(Z_{t+j})] + T''[E_t(Z_{t+j})]\{T'[E_t(Z_{t+j})]\}^{-2}\sigma_j^2/2 \tag{A4}$$

In particular, for the logarithmic transformation this expression leads us to

$$\mu_j \doteq \log[E_t(Z_{t+j})] - \sigma_j^2/2 \tag{A5}$$

from which we get

$$E_t(Z_{t+j}) \doteq \exp(\mu_j + \sigma_j^2/2) \tag{A6}$$

which is an exact result when $\log(Z_{t+j})$ is normally distributed. Similarly, when $T(Z_t) = Z_t^\lambda$ with $\lambda \neq 0$, we obtain from equation (A4)

$$\mu_j \doteq E_t^\lambda(Z_{t+j}) + (1 - \lambda^{-1})E_t^{-\lambda}(Z_{t+j})\sigma_j^2/2 \tag{A7}$$

which yields a second-degree equation in $E_t^\lambda(Z_{t+j})$, with solution

$$E_t(Z_{t+j}) \doteq \mu_j^{1/\lambda}\{0.5 \pm 0.5[1 + 2(\lambda^{-1} - 1)\sigma_j^2/\mu_j^2]^{1/2}\}^{1/\lambda} \tag{A8}$$

Here the plus sign is chosen to ensure that $E_t(Z_{t+j}) > \mu_j^{1/\lambda}$ for $\lambda < 1$ and $E_t(Z_{t+j}) < \mu_j^{1/\lambda}$ for $\lambda > 1$, which is a consequence of the fact that the power transformation is monotonic in λ. Consequently, expressions (8) and (9) follow from (A6) and (A8).

ACKNOWLEDGEMENTS

The author would like to thank three anonymous reviewers and the Departmental Editor, Professor R. H. Shumway, for their helpful comments and suggestions which significantly improved the presentation of this paper. He also acknowledges the benefit of discussions on the topic of this paper with M. C. Minor and the computational help of L. Rosas.

REFERENCES

Abraham, B. and Ledolter, J., 'Forecast functions implied by autoregressive integrated moving average models and other related procedures', *International Statistical Review*, **54** (1986), 51–66.

Ansley, C. F., Spivey, W. A. and Wrobleski, W. J., 'A class of transformations for Box–Jenkins seasonal models', *Applied Statistics*, **26** (1977), 173–8.

Atkinson, A. C., *Plots. Transformations and Regression*, Oxford: Clarendon Press, 1985.

Bowerman, B. L., Koehler, A. B. and Pack, D. J., 'Forecasting time series with increasing seasonal variation', *Journal of Forecasting*, **9** (1990), 419–36.

Box, G. E. P. and Cox, D. R. 'An analysis of transformations', *Journal of the Royal Statistical Society*, **B-26**, (1964), 211–43.

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

Box, G. E. P. and Jenkins, G. M., 'Some comments on a paper by Chatfield and Prothero and on a review by Kendall', *Journal of the Royal Statistical Society*, **A-136** (1973), 337–52.

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

Cleveland, W. S., Dunn, D. M. and Terpenning, I. J. 'SABL: A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnostics', in A. Zellner (ed.), *Seasonal Analysis of Economic Time Series*, US Bureau of the Census, 1978, 201–41.

Durbin, J. and Kenny, P. B., 'Seasonal adjustment when the seasonal component behaves neither purely multiplicatively nor purely additively', in A. Zellner (ed.), *Seasonal Analysis of Economic Time Series*, US Bureau of the Census, 1978, 173–97.

Durbin, J. and Murphy, M. J., 'Seasonal adjustment based on a mixed additive–multiplicative model', *Journal of the Royal Statistical Society*, **A-138** (1975), 385–410.

Granger, C. W. J. and Newbold, P., 'Forecasting transformed series', *Journal of the Royal Statistical Society*, **B-38** (1976), 189–203.

Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, 2nd edn, New York: Academic Press, 1986.

Guerrero, V. M. and Johnson, R. A., 'Use of the Box–Cox transformation with binary response models', *Biometrika*, **69**, (1982), 309–14.

Harvey, A. C. and Fernandes, C., 'Time series models for count or qualitative observations', *Journal of Business and Economic Statistics*, **7**, (1989), 407–17.

Jenkins, G. M., 'Practical experiences with modelling and forecasting time series', in O. D. Anderson (ed.), *Forecasting*, San Francisco: Holden-Day, 1979, 43–166.

Miller, D. M., 'Reducing transformation bias in curve fitting', *The American Statistician*, **38** (1984), 124–6.

Neyman, J. and Scott, E. L., 'Correction for bias introduced by a transformation of variables', *Annals of Mathematical Statistics*, **31** (1960), 643–55.

Pankratz, A. and Dudley, U., 'Forecasts of power-transformed series', *Journal of Forecasting*, **6** (1987), 239–48.

Perry, J. N., 'Taylor's power law for dependence of variance on mean in animal populations', *Applied Statistics*, **30**, (1981), 254–63.

Taylor, J. M. G., 'The retransformed mean after a fitted power transformation', *Journal of the American Statistical Association*, **81**, (1986), 114–18

Wilson, G. T., 'Discussion of a paper by Dr Chatfield and Dr Prothero', *Journal of the Royal Statistical Society*, **A-136**, (1973), 315–19.

Author's biography:

Victor M. Guerrero received his PhD (1979) in Statistics from the University of Wisconsin-Madison and is currently a Professor at the Instituto Tecnológico Autónomo de México (ITAM). His research interests include time-series analysis, forecasting, and econometrics. He has published papers in the *Journal of Forecasting*, *International Statistical Review*, *Biometrika*, *Insurance: Mathematics and Economics*, *Economics Letters*, *International Journal of Forecasting*, *Letters in Probability and Statistics*, *Estadística* and others in Spanish.

Author's address:

Victor M. Guerrero, Instituto Tecnológico Autónomo de México, Rio Hondo #1, México 01000, DF, México.