Journal of Forecasting, Vol. 6. 239–248 (1987)

Forecasts of Power-transformed Series

ALAN PANKRATZ
UNDERWOOD DUDLEY
DePauw University, U.S.A.

ABSTRACT

Consider a time series transformed by an instantaneous power function of the Box–Cox type. For a wide range of fractional powers, this paper gives the relative bias in original metric forecasts due to use of the simple inverse retransformation when minimum mean squared error (conditional mean) forecasts are optimal. This bias varies widely according to the characteristics of the data. A fast algorithm is given to find this bias, or to find minimum mean squared error forecasts in the original metric. The results depend on the assumption that the forecast errors in the transformed metric are Gaussian. An example using real data is given.

KEY WORDS Box–Cox transformation Minimum mean absolute error forecasts Minimum mean squared error forecasts Time series Tukey's power transformation

Let a time series Y_t to be forecast be transformed by a monotonic, instantaneous power function $T(Y_t)$ into another series X_t to induce homogeneous and normally distributed model errors. For example, if the variance of Y_t is proportional to its level, then $X_t = Y_t^{0.5}$ will have a constant variance. A model for X_t is developed and, given the data, forecasts of X_t are generated. Typically, forecasts of Y_t are desired; for example, managers want forecasts of sales, not the square root of sales.

One way to find forecasts of Y_t is to apply the inverse of the transformation $T(\cdot)$ to forecasts of X_t. This is called the 'naïve' retransformation. Practising forecasters often use this procedure, partly because it seems natural and partly because their software may not provide retransformation options. But to find the optimum forecast of Y_t, from origin n with lead time h, generally we must know the decision-maker's forecast error loss function (or an approximation) and the conditional probability density function (pdf) of Y_{n+h}, given the available data (Granger, 1969). For example, for absolute error loss (that is, the loss function is linear and symmetrical around zero) the naïve retransformation is optimal for any conditional pdf of Y_{n+h}; it gives the minimum mean absolute error (MMAE) forecast, equal to the median of the conditional pdf. In this paper we assume that the error loss function is quadratic; that is, the cost of error is proportional to the squared error. Then the optimal forecast in the original metric, for any conditional pdf of Y_{n+h}, is the minimum mean squared error (MMSE) forecast, equal to the mean of the conditional pdf.

0277–6693/87/040239–10$05.00

Received July 1986
Revised April 1987

While a quadratic forecast error loss function is not always appropriate, it may provide a useful approximation in many cases when the exact loss function is unknown, especially when extreme errors are likely to bring sharply increased costs. For example, consider a manufacturer that produces for sale to retail dealers. If the forecast falls short of orders by a small to moderate amount, insufficient resource acquisition will lead to lost profits; this cost may be approximated by a linear function equal to profit margin times unit sales lost. But if the forecast falls far enough below orders, unhappy dealers missing large portions of their orders may seek other suppliers in the future. Then the (present value) cost of a large forecast shortfall will exceed the value of profit margin times current unit sales lost; in the extreme case, the survival of the firm may be threatened. If the forecast exceeds orders by small to moderate amounts, the producer will incur a cost which may be well approximated by a linear function equal to unit carrying cost (e.g., interest and warehouse handling) times unsold units. But if the forecast is far enough above orders, the producer may acquire unnecessary and less convenient warehouse space, face higher interest charges and additional financial transactions costs due to increased leverage (Weston and Brigham, 1978, Ch. 19), undertake unnecessary yet costly labour search and hiring programmes, and so forth. Then the cost of a large negative error will exceed unit carrying cost times the number of unsold units and, in the extreme case, the viability of the firm may be threatened. In situations like this a quadratic error loss function may be a reasonable approximation even if the true loss function is linear and asymmetric for small to moderate errors.

Assuming a quadratic forecast error loss function, so that MMSE forecasts are optimal, two questions arise: (i) How much do MMAE forecasts of Y_t deviate from MMSE forecasts? (ii) How can MMSE forecasts be found? Several authors have investigated these questions for various forms of the transformation $T(\cdot)$. Granger and Newbold (1976) present quite general theoretical results on forecasting Y_t when X_t is Gaussian for a wide class of transformations. They attack the problem using Hermite polynomial expansions. Unfortunately the required expansions become so complicated for many fractional power transformations that this approach is often not useful in practice. Granger and Newbold give results for two tractable cases, $X_t = Y_t^{0.5}$ and $X_t = \ln Y_t$. Their results suggest that use of the naïve retransformation may give forecasts in the original metric that deviate substantially from MMSE forecasts.

Nelson and Granger (1979) approach the problem empirically, applying the transformation of Box and Cox (1964) to 21 macroeconomic series. They find a non-biasing procedure (to obtain MMSE forecasts) to be only moderately worthwhile, giving better forecasts in the original metric about 60 per cent of the time for various forecast lead times. They report that the non-biasing procedure gave little improvement in forecast accuracy in simulation experiments.

The present paper gives the theoretical relative bias resulting from use of the naïve retransformation for a wide range of fractional powers under the Box–Cox transformation, assuming that MMSE forecasts are desired and that the forecast errors for X_t are Gaussian. It also gives a computationally fast algorithm for obtaining MMSE forecasts of Y_t from forecasts of X_t. The results are not as general as those of Granger and Newbold (1976), but they cover a wider range of fractional power transformations and are much more easily obtained. We find that the bias due to use of the naïve retransformation ranges from negligible to severe; this suggests that Nelson and Granger (1979) may have found the non-biasing procedure to be only slightly worthwhile largely because of the characteristics of their data.

The next section sets out the Box–Cox transformation. Section 2 gives the theoretical relative bias in MMAE forecasts of Y_t under this transformation, assuming that MMSE forecasts are optimal. Section 3 gives a fast algorithm for evaluating expressions derived in Section 2, along with numerical results. An example using real data is given in Section 4, and Section 5 contains a summary and concluding comments.

1. THE BOX–COX TRANSFORMATION

Much attention has been given to the transformation introduced by Box and Cox (1964):

$$X_t = (Y_t^c - 1)/c, \tag{1}$$

where $Y_t > 0$ and c is real. If Y_t can be negative, a constant H is added to it such that $\Pr[(Y_t + H) \leqslant 0]$ is negligible. Examples of the use of (1), where c is estimated, appear in Wilson (1973), Chatfield and Prothero (1973b), Box and Jenkins (1976), Ansley *et al.* (1977), Nelson and Granger (1979) and Hopwood *et al.* (1981, 1984). Transformation (1) is a variant of the simple transformation $X_t = Y_t^c$ discussed by Tukey (1957). An advantage of (1) is that it is continuous in the limit as c goes to zero (by L'Hospital's rule) with $X_t = \ln Y_t$ in that case.

In the next section we obtain expressions for MMAE and MMSE (conditional median and mean) forecasts of Y_t based on forecasts of X_t.

2. CONDITIONAL MEDIAN AND MEAN FORECASTS OF Y_t

Suppose information set I_n is used to forecast the transformed series X_t from forecast origin time n. I_n may include exogenous variables as well as values of X_t through period n. Given I_n, the optimal (MMSE) h-step ahead forecast of X_{n+h} is $E(X_{n+h} | I_n)$, denoted here as f. Denote the h-step ahead forecast error as u, so that $X_{n+h} = f + u$, where u is truncated Gaussian with mean zero and h-step ahead variance σ_h^2. For an ARIMA model, for instance, this variance depends on the model coefficients as well as the forecast horizon h.

Truncation of u follows from the restriction $Y_{n+h} > 0$, which implies a lower or upper bound for X_{n+h} and thus for u. To ensure $E(u) = 0$, let u be doubly truncated at $\pm u^*$. The results in this paper rest on the usual assumption that u^* is large enough to render the effect of this truncation negligible.

Conditional median forecasts

Given an optimal forecast for X_{n+h}, suppose that a forecast of Y_{n+h} is desired. (For simplicity let $H = 0$. If $H > 0$, H must be subtracted from the forecast of Y_{n+h} to return the forecast to the original level of the Y_t series.) One procedure to forecast Y_{n+h} is simply to apply the inverse of the original transformation to f, giving a 'naïve' forecast of Y_{n+h} as $m = T^{-1}(f)$. Letting $1/c = p$, this naïve forecast under transformation (1) is $m = (cf + 1)^p$.

This inversion generally does not give the MMSE forecast in the original metric; it gives the MMAE forecast equal to the median of the conditional pdf of Y_{n+h}. This follows because the pdf of u is symmetrical so that f is the median as well as the mean of X_{n+h}. And since $Y_t = T^{-1}(X_t)$ is an increasing function of X_t for (1), the 50th percentile of Y_{n+h} is the inverse transformation of the 50th percentile of X_{n+h}. For $c \neq 1$, the conditional pdf of Y_{n+h} is skewed so that m is not its mean.

Conditional mean forecasts

For the MMSE forecast of Y_{n+h} we require the expected value $M = E(Y_{n+h} | I_n)$. We have $Y_{n+h} = T^{-1}(X_{n+h}) = T^{-1}(f + u)$, and M may be found directly from the conditional pdf of X_{n+h} as

$$M = \int_{-u^*}^{u^*} T^{-1}(f + u)\phi(u)\,\mathrm{d}u, \tag{2}$$

where ϕ is the (negligibly truncated) Gaussian pdf,

$$\left[(1/\sigma_h\sqrt{[2\pi]})\exp\{-(u/\sigma_h)^2/2\}\bigg/\int_{-u^*}^{u^*}\phi(u)\,du\right],$$

in which the value of the integral is close enough to 1.0 that it may be ignored in practice. Define $w=u/\sigma_h$, so $u=\sigma_h w$, $w^*=u^*/\sigma_h$, $du=\sigma_h\,dw$, and

$$M=\int_{-w^*}^{w^*}(cf+c\sigma_h w+1)^p g(w)\,dw, \tag{3}$$

where g is the (negligibly truncated) standard Gaussian pdf and the integrand is $T^{-1}(f+u)$ under transformation (1), with c (and hence p), f and σ_h treated as constants. Nelson and Granger (1979) give (3) for the Box–Cox case, though their expression (3.4) has a typographical error: -1 in the second factor of the integrand should be $+1$.

Factoring $m=T^{-1}(f)=(cf+1)^p$ out of (3) gives the alternative expression

$$M=m\int_{-w^*}^{w^*}(1+rw)^p g(w)\,dw=mG, \tag{4}$$

where $r=\sigma_h/(f+p)$. (Equation (4) also holds for the simple power transformation $X_t=Y_t^c$, with $r=\sigma_h/f$.) Thus the MMSE forecast of Y_{n+h} may be found as the product of the simple inverse transformation (conditional median) m and a factor G which is the value of the integral in (4).

There are restrictions on r implied by the assumptions that $Y_t>0$ and that truncation of the pdf $g(w)$ is negligible. This is seen by noting that, under transformation (1), the condition $Y_{n+h}>0$ requires $Y_{n+h}^c=cX_{n+h}+1=c(f+u)+1=c(f+\sigma_h w)+1>0$. Thus, if $c>0$, we must have $w>-(f+p)/\sigma_h$, or $w>-1/r$. This also ensures $(1+rw)>0$ in (4) when $c>0$ (preventing negative numbers to fractional powers). If $c<0$, we must have $w<-(f+p)/\sigma_h$, or $w<-1/r$. Thus the standardized truncation points are $w^*=\pm|1/r|$. If truncation effects for the standard Gaussian pdf are considered negligible for, say, $|w^*|\geqslant 4$, then we must have $|r|\leqslant 0.25$.

Relative bias

Define the relative bias B from using m to forecast Y_{n+h} as

$$B=(m-M)/M \tag{5}$$

For $c=0$ (the natural log transform), Nelson (1973, Ch. 6) and Granger and Newbold (1976) show that $m=\exp(f)$ and $M=\exp(f+\sigma_h^2/2)$, in which case $G=\exp(\sigma_h^2/2)$ and $B=\exp(-\sigma_h^2/2)-1$. For $X_t=Y_t^{0.5}$, Granger and Newbold also show that $m=f^2$ and $M=f^2+\sigma_h^2$, so $G=1+(\sigma_h/f)^2$ and $B=-(\sigma_h/f)^2/[1+(\sigma_h/f)^2]$.

The next section gives a fast procedure for evaluating G, along with values of B, for a wide range of fractional c values when $|r|\leqslant 0.25$.

3. NUMERICAL EVALUATION ALGORITHM AND RESULTS

The integral G in (4) has no closed form solution for general c and must be evaluated numerically. As shown in the Appendix, assuming $1/c=p$ is a positive integer and expanding the binomial in (4) gives

$$M=m\left\{1+\sum_{k=1}^{\infty}p(p-1)\cdots(p-(2k-2))(p-(2k-1))r^{2k}/2^k k!\right\}, \tag{6}$$

where the factor in brackets is the non-biasing factor G.

The sum in (6) is finite when p is a positive integer. If p is an even, positive integer, all terms after $k=1/2c$ are zero; if p is an odd, positive integer, all terms after $k=(1-c)/2c$ are zero.

We use the expression in (6) to approximate the non-biasing factor when p is not a positive integer. A numerical study of the sum shows that truncation after not more than eight terms gives good results for most $\{c,|r|\}$ values likely to arise in practice.

Table 1 shows values of $100B$ (the per cent bias from use of m) for various $\{c,|r|\}$. These values were found by truncating the sum in (6) after eight terms. Table 1 entries shown as ** are cases where (6) gives results notably different from those obtained by evaluating (4) with Romberg integration. (Romberg integration is a commonly used numerical integration method. See Stark (1970, Ch. 4) for an introductory discussion.) This problem appears to be inherent in the integral G: for those cases, Romberg method results are highly sensitive to the truncation value w^*. It is not

Table 1. Per cent bias from use of conditional median forecast in original metric when conditional mean forecast is optimal under Box–Cox transformation

	$\|r\|$					
c	0.02	0.05	0.10	0.15	0.20	0.25
3.00	0.0	0.0	0.1	0.3	0.5	0.8
2.00	0.0	0.0	0.1	0.3	0.5	0.8
1.00	0	0	0	0	0	0
0.75	−0.0	−0.1	−0.2	−0.5	−0.9	−1.4
0.50	−0.0	−0.2	−1.0	−2.2	−3.8	−5.9
0.25	−0.2	−1.5	−5.7	−12.0	−19.7	−27.9
0.10	−1.8	−10.4	−34.0	−57.8	−75.1	−85.9
−0.10	−2.2	−13.2	−46.8	**	**	**
−0.25	−0.4	−2.5	−10.1	−23.1	−43.8	**
−0.50	−0.1	−0.8	−3.1	−7.1	−13.4	−23.8
−0.75	−0.1	−0.4	−1.6	−3.7	−6.9	−12.1
−1.00	−0.0	−0.3	−1.0	−2.4	−4.4	−7.7
−2.00	−0.0	−0.1	−0.4	−0.9	−1.6	−2.7
−3.00	−0.0	−0.1	−0.2	−0.5	−1.0	−1.6

** = unreliable results.

clear whether (6) or other numerical integration methods will give better results in such extreme cases. Use of the log transformation is recommended in those situations since they arise only when c is near zero (on the negative side). Expression (6) is more efficient computationally than commonly used methods, such as Romberg integration, since the latter typically require repeated exponentiations; after some parameter assignments, evaluating G with (6) requires no exponentiations.

As seen in Table 1, the bias on the conditional median forecasts of Y_t increases as c approaches zero or as $|r|$ increases. The small bias for many $\{c,|r|\}$ may help to explain the modest improvement in forecast accuracy reported by Nelson and Granger (1979) when a non-biasing procedure was used with the Box–Cox transformation. In 12 of their 21 cases using real data and maximum likelihood estimates of c, they find $|c|>0.6$, and in 11 of those 12 c is positive. Table 1 shows that the bias for a given $|r|$ is smaller when $c>0$.

It also appears that the values of $|r|$ arising in Nelson and Granger's study may be rather small.

We have not examined all of their data, but three of their series show the following range of $|r$ values for forecast lead times from $h=1$ to $h=10$:

Nelson-Granger series	c	Range for $|r|$
H	-1.52	0.014–0.052
J	0.14	0.003–0.020
P	0.63	0.004–0.023

As seen in Table 1, for these $\{c, |r|\}$ values the bias from using m is typically a small fraction of -1 per cent. If Nelson and Granger's other series have similar $\{c, |r|\}$ values, it is not surprising that their overall empirical results show only a small gain in forecast accuracy when a non-biasing procedure is used.

4. SALES DATA EXAMPLE

In this section the non-biasing procedure is applied to forecasts of a real data series originally studied by Chatfield and Prothero (1973a). (The data are given in their article.)

The series is 77 observations of the monthly sales of an engineering product with a strong seasonal pattern. In their original analysis, Chatfield and Prothero use a log transformation. Use of this transformation was criticized by Box and Jenkins (1973) and Wilson (1973). The former suggest the use of $c=0.25$ based on graphical analysis; the latter finds the optimal c to be 0.34 using a maximum likelihood criterion. We use $c=1/3$, so $1/c=p=3$, with $X_t=Y_t^{1/3}$.

Chatfield and Prothero and their critics consider various ARIMA models for series X_t, but the one receiving the most attention is

$$(1-\phi_1 B)w_t=(1-\theta_{12}B^{12})a_t, \tag{7}$$

where w_t is X_t after both regular and seasonal differencing. The first 65 observations give the estimates $\hat{\phi}_1=-0.53$ and $\hat{\theta}_{12}=0.54$; these were obtained using the maximum likelihood option in the SCA system (Liu *et al.*, 1983). The coefficients of skewness and kurtosis for the model residuals are smaller than their respective standard errors, so the assumption of Gaussian forecast errors seems reasonable.

Forecasts of X_{n+h} for $h=1,\ldots,12$ from time origin $n=65$ were produced using model (7). The corresponding conditional median (m) and conditional mean (M) forecasts of Y_t are shown in Table 2. The m forecasts (column 3) are simply the X_t-metric forecasts raised to the power $1/c=p=3$, while the M forecasts (column 7) are the values of m multiplied by the non-biasing factor G (column 6). Since p is a positive integer the series in (6) is finite; with $p=3$, only one term is needed to find exact values of G:

$$G=1+p(p-1)r^2/2=1+3r^2, \tag{8}$$

where (under the simple power transformation) $r=s_h/f$, f is the h-step ahead forecast in the X_t metric, and s_h is the corresponding estimated forecast standard deviation. We obtained this estimate from the SCA package (Liu *et al.*, 1983) forecast option which uses the psi-weights implied by the ARIMA model in (7) and the variance estimate of the residuals, following Box and Jenkins (1976, Ch. 5).

Column 5 of Table 2 is $s_h/f=r$ for forecast lead times $h=1,\ldots,12$. The largest s_h/f value of 0.123

Table 2. Conditional median (m) and mean (M) forecasts from model (7) for forecast horizons $h=1,\ldots,12$ from forecast origin $n=65$

(1) t	(2) Y	(3) m	(4) $Y-m$	(5) s/f	(6) G	(7) M	(8) $Y-M$
66	257	256	+1	0.052	1.008	258	−1
67	324	304	+20	0.055	1.009	306	+18
68	404	423	−19	0.059	1.011	427	−23
69	677	572	+105	0.059	1.010	577	+100
70	858	768	+90	0.059	1.010	776	+82
71	895	806	+89	0.062	1.012	815	+80
72	664	650	+14	0.071	1.015	660	+4
73	628	665	−37	0.075	1.017	676	−48
74	308	469	−161	0.089	1.024	480	−172
75	324	365	−41	0.101	1.031	376	−52
76	248	335	−87	0.109	1.035	347	−99
77	272	261	+11	0.123	1.045	272	0

corresponds to the most severe truncation of the standard Gaussian density in (4) at about $w^* = \pm 8$, so there is no apparent problem in assuming the truncation effects to be negligible.

As seen in column 6 of Table 2, the values for the non-biasing factor G range from 1.008 to 1.045. Thus using m to forecast Y_t will give forecasts that are biased (downward) away from the MMSE forecasts by roughly 1 to 4.5 per cent, depending on the forecast horizon. This is consistent with the results across the first seven forecasts; in five of these cases the non-biasing procedure brings the forecasts closer to the observed values.

The conditional median forecasts happen to perform slightly better based on (rounded) summary statistics for all 12 forecasts in this very limited example: the conditional median forecasts have a mean error of zero and a root-mean-squared error of 73 for the 12 forecasts, compared to 1 and 75, respectively, for the conditional mean forecasts. This appears to be due to an outlier at period 74: when (7) is applied to the full data set, the residual for period 74 is more than 2.8 times the residual standard deviation. The model overforecasts that period, and M overforecasts more than m since G is positive; the resulting large error imposes a heavy penalty on M in the summary statistics.

5. SUMMARY AND CONCLUSIONS

This paper examines methods of forecasting a time series Y_t after it has been transformed to a new series X_t by an instantaneous power function such as the Box–Cox transformation. Given the available data, forecasts of X_t may be used to find forecasts of Y_t. Applying the inverse of the original transformation to forecasts of X_t (the 'naïve' procedure) yields minimum mean absolute error forecasts of Y_t equal to the median of the conditional pdf of Y_t.

Assuming that minimum mean squared error (MMSE) forecasts are optimal, this paper gives the relative bias in the naïve forecasts of Y_t for a wide range of fractional powers under the Box–Cox transformation. (The results are easily adapted to the case of a simple power transformation.) A fast algorithm is given for finding the bias due to the naïve procedure, or for finding MMSE forecasts of Y_t. For the range of cases considered here, the theoretical bias in the naïve forecasts of Y_t varies from a small fraction of ± 1 per cent to more than -85 per cent. A larger bias occurs as the power (c) gets closer to zero, and as the X_t-metric forecast standard deviation (σ_h) grows relative to the sum of the forecast and the inverse of the power ($f+p$) for the Box–Cox transformation.

The results given here must be used with some care. They assume that the forecast error loss function is quadratic, so that minimum mean squared error forecasts are optimal, and that the forecast errors in the X_t metric are Gaussian. As suggested by Nelson and Granger's (1979) study, real data subjected to ARIMA analysis may often not satisfy the latter assumption; this situation might be improved by the addition of intervention components to account for outliers or step shifts (see Box and Tiao, 1975). Furthermore, the properties of the algorithm in Section 3 may be unsatisfactory when c is near zero on the negative side and the forecast standard deviation in the X_t metric is relatively large; however, other numerical integration methods also give unstable results in those cases and the log transform perhaps should be used.

The results should give forecasters a better sense of the bias they may expect from use of the naïve (simple inverse) retransformation when MMSE forecasts are desired. In addition, the algorithm in Section 3 behaves well enough for a sufficiently wide range of powers that it should be useful in obtaining MMSE forecasts in the original metric in a computationally efficient manner when the underlying assumptions are acceptable.

APPENDIX: DERIVATION OF NON-BIASING FACTOR G

For simplicity we ignore time subscripts and deal with the simple power transformation $X=Y^c$, and its inverse $Y=X^{1/c}$, where $Y>0$ and c is real. Let $X=f+u$, where $f=E(X)$ and u is Gaussian with mean zero and variance σ^2. Suppose for the moment that $1/c=p$ is a positive integer. We wish to evaluate

$$M=E(Y)=\int_{-\infty}^{\infty}(f+u)^p\phi(u)\,\mathrm{d}u, \tag{A1}$$

where ϕ is the Gaussian pdf, $(1/\sigma\sqrt{[2\pi]})\exp\{-(u/\sigma)^2/2\}$. Let $w=u/\sigma$, so $u=\sigma w, f+u=f+\sigma w=f(1+\sigma w/f)$, $\mathrm{d}u=\sigma\,\mathrm{d}w$, and

$$M=\int_{-\infty}^{\infty}f^p(1+\sigma w/f)^p g(w)\,\mathrm{d}w, \tag{A2}$$

where g is the standard Gaussian pdf, and where f, σ and p are treated as constants. Expanding the binomial,

$$M=(1/\sqrt{[2\pi]})f^p\int_{-\infty}^{\infty}\left\{1+\sum_{k=1}^{\infty}\binom{p}{k}(\sigma w/f)^k\right\}\exp(-w^2/2)\,\mathrm{d}w, \tag{A3}$$

where the sum is finite since $\binom{p}{k}=0$ for $k>p$.

Since $w^k\exp(-w^2/2)$ is an odd function when k is odd,

$$M=(1/\sqrt{[2\pi]})f^p\int_{-\infty}^{\infty}\left\{1+\sum_{k=1}^{\infty}\binom{p}{2k}(\sigma/f)^{2k}w^{2k}\right\}\exp(-w^2/2)\,\mathrm{d}w \tag{A4.1}$$

$$=f^p\left\{1+(1/\sqrt{[2\pi]})\sum_{k=1}^{\infty}\binom{p}{2k}(\sigma/f)^{2k}\int_{-\infty}^{\infty}w^{2k}\exp(-w^2/2)\,\mathrm{d}w\right\} \tag{A4.2}$$

$$=f^p\left\{1+\sum_{k=1}^{\infty}\binom{p}{2k}(\sigma/f)^{2k}(1\cdot 3\cdots(2k-1))\right\} \tag{A4.3}$$

$$=f^p\left\{1+\sum_{k=1}^{\infty}p(p-1)\cdots(p-(2k-2))(p-(2k-1))(\sigma/f)^{2k}/2^k k!\right\}, \tag{A4.4}$$

where the factor in brackets is the non-biasing factor G. When using the Box–Cox transformation, σ/f is replaced by $\sigma/(f+p)$.

Truncating the integral to go from $-w^*$ to w^* introduces negligible error in G even for moderate values of w^*. This may be seen by noting that the resulting error (E) in G when truncating the right-hand tail of the integral is

$$E=(1/\sqrt{[2\pi]})\int_{w^*}^{\infty}(1+\sigma w/f)^p \exp(-w^2/2)\,dw. \tag{A5}$$

Now let $v(w)=(1+\sigma w/f)^p\exp(-w/2)$, so

$$E=(1/\sqrt{[2\pi]})\int_{w^*}^{\infty} v(w)\exp(-(w^2-w)/2)\,dw. \tag{A6}$$

The maximum value of v to the right of its zero at $w=-f/\sigma$ occurs at $w_{max}=2p-f/\sigma$. Then

$$E<(1/\sqrt{[2\pi]})v(w_{max})\int_{w^*}^{\infty}\exp(-(w^2-w)/2)\,dw. \tag{A7}$$

Letting $z=w-1/2$,

$$E<(1/\sqrt{[2\pi]})v(w_{max})\exp(1/8)\int_{w^*-1/2}^{\infty}\exp(-z^2/2)\,dz \tag{A8.1}$$

$$<(1/\sqrt{[2\pi]})v(w_{max})\exp(1/8)\{\exp(-(w^*-1/2)^2/2)/(w^*-1/2)\}. \tag{A8.2}$$

This becomes small very quickly as w^* increases since the factor in brackets is very small and $v(w_{max})$ is not large for values of $\{\sigma, f, p\}$ likely to occur in practice. For example, consider the extreme case of $\sigma/f=0.25$, so $w^*=4$; then for $p=5$, E is less than 1.4×10^{-3}. Similar analysis shows that the error in G resulting from truncating the left-hand tail of the integral is also small.

ACKNOWLEDGEMENTS

We gratefully acknowledge constructive comments from two referees and from Johannes Ledolter in his capacity as departmental editor, and helpful discussions with John Morrill, Paul Newbold, Carl Singer, Richard Smock and Mike Zeller. The first author received partial support from DePauw University's Fisher Fund.

REFERENCES

Ansley, C. F., Spivey, W. A. and Wrobleski, W. J., 'A class of transformations for Box–Jenkins seasonal models', *Applied Statistics*, **26** (1977), 173–178.

Box, G. E. P. and Cox, D. R., 'An analysis of transformations', *Journal of the Royal Statistical Society*, **B26** (1964), 211–243.

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*, **A135** (1973), 337–345.

Box, G. E. P. and Cox, D. R., *Time Series Analysis: Forecasting and Control*, rev. edn., San Francisco: Holden-Day, 1976.

Box, G. E. P. and Tiao, G. C., 'Intervention analysis with applications to economic and environmental problems', *Journal of the American Statistical Association*, **70** (1975), 70–79.

Chatfield, C. and Prothero, D. L., 'Box–Jenkins seasonal forecasting: problems in a case study', *Journal of the Royal Statistical Society*, **A136** (1973a), 295–315.

Chatfield, C. and Prothero, D. L., 'A reply', *Journal of the Royal Statistical Society*, **A136** (1973b), 345–352.

Granger, C. W. J., 'Prediction with a generalized cost of error function', *Operational Research Quarterly*, **20** (1969), 199–207.

Granger, C. W. J. and Newbold, P., 'Forecasting transformed series', *Journal of the Royal Statistical Society*, **B38** (1976), 189–203.

Hopwood, W. S., McKeown, J. C. and Newbold, P., 'Power transformations in time-series models of quarterly earnings per share', *The Accounting Review*, **56** (1981), 927–933.

Hopwood, W. S., McKeown, J. C. and Newbold, P., 'Time series forecasting models involving power transformations', *Journal of Forecasting*, **3** (1984), 57–61.

Liu, L.-M., Hudak, G. B., Muller, M. E., Box, G. E. P. and Tiao, G. C., *The SCA System for Univariate–Multivariate Time Series and General Statistical Analysis*, DeKalb, IL: Scientific Computing Associates, 1983.

Nelson, C. R., *Applied Time Series Analysis for Managerial Forecasting*, San Francisco: Holden-Day, 1973.

Nelson, H. L. and Granger, C. W. J., 'Experience with using the Box–Cox transformation when forecasting economic time series', *Journal of Econometrics*, **10** (1979), 57–69.

Stark, P. A., *Introduction to Numerical Methods*, New York: Macmillan, 1970.

Tukey, J. W., 'On the comparative anatomy of transformations', *Annals of Mathematical Statistics*, **28** (1957), 602–632.

Weston, J. F. and Brigham, E. F., *Managerial Finance*, Hinsdale, IL: Dryden, 1978.

Wilson, G. T., 'Discussion of paper by Chatfield and Prothero', *Journal of the Royal Statistical Society*, **A136** (1973), 315–319.

Authors' biographies:

Alan Pankratz teaches in the Department of Economics and Management, DePauw University. He has published articles in the *Journal of Money, Credit, and Banking*, *Southern Economic Journal* and *Journal of Economics and Business*, and is the author of *Forecasting with Univariate Box–Jenkins Models* (Wiley, 1983).

Underwood Dudley teaches in the Department of Mathematics and Computer Science, DePauw University. He is the author of *Number Theory*, 2nd edn. (W. H. Freeman and Co., 1978).

Authors' addresses:

Alan Pankratz, Department of Economics and Management, DePauw University, Greencastle, IN 46135, U.S.A.

Underwood Dudley, Department of Mathematics and Computer Science, DePauw University, Greencastle, IN 46135, U.S.A.