The Algebraic Structure of Transformed Time Series

Tucker McElroy and Osbert Pang

Abstract Invertible transformations are often applied to time series data to generate a distribution closer to the Gaussian, which naturally has an additive group structure. Estimates of forecasts and signals are then typically transformed back to the original scale. It is demonstrated that this transformation must be a group homomorphism (i.e., a transformation that preserves certain arithmetical properties) in order to obtain coherence between estimates of quantities of interest in the original scale, and that this homomorphic structure is ensured by defining an induced group structure on the original space. This has consequences for the understanding of forecast errors, growth rates, and the relation of signal and noise to the data. The effect of the distortion to the additive algebra is illustrated numerically for several key examples.

Key words: forecasting, growth rates, homomorphism, signal extraction

1 Introduction

The analysis of time series data is often focused on producing estimates of signals, forecasts, and/or growth rates, all of which are typically estimated by methodologies that assume an additive group structure of the data. For example, many signal extraction estimates assume that the sum of signal and noise equals the original data process; forecasts have their performance evaluated by taking their difference with the future value (this defines the forecast error). However, it is not uncommon for data to be initially transformed by an invertible function so as to make a Gaussian distribution more plausible. Any signal estimates, forecasts, or growth rates would

---

T. McElroy
U.S. Census Bureau, e-mail: tucker.s.mcelroy@census.gov

O. Pang
U.S. Census Bureau, e-mail: osbert.c.pang@census.gov

then be transformed back into the original scale by inverting the transformation. This mapping necessarily distorts the additive group structure.

For example, many monthly retail series exhibit dramatic seasonal behavior and hence are candidates for seasonal adjustment [3][9]. Due to the underlying linkage of retail to inflation exponential growth is not uncommon, and typically a logarithmic transformation is suitable for producing a more symmetric, light-tailed marginal distribution. Seasonal adjustments, which are an application of signal extraction techniques, can then be produced using an additive group structure. Inverting the initial transformation by exponentiating maps the addition operator to the multiplication operator. That is, in the original data scale the seasonal and nonseasonal estimates no longer sum to the data process, but instead their product equals the data process.

This is the only transformation with an intuitive induced algebra on the original space. All transformations induce a group structure on the original space, which can be used to understand how the data process is decomposed into signal and noise, or how growth rates are to be properly understood; however, multiplication is the only familiar induced group structure[1]. The main result of this paper is to explicitly derive the induced group structure, and study its impact on several examples, such as the hyperbolic sine and logistic transformations.

Section 2 gives background concepts, with a brief discussion of the statistical motivation for our results, which arise from time series data that have been affected by the use of transformations. Section 3 contains our main results, and develops the algebra of the parent space, which is induced by the additive group structure of the transformed space. Section 4 continues the main examples and provides plots of level curves for the new group operations. Section 5 gives two empirical examples that compare the new group operator to addition in the parent space for the square root and logistic transformations. Section 6 provides our conclusions and discusses the implications for interpreting signal extractions, forecasts, and growth rates.

2 Statistical Background

Let us label the original domain of the data as the "parent space," and all variables will be written in bold. The "transformation space" arises from application of a one-to-one mapping φ, which is chosen so as to reduce heteroscedasticity, skewness, and kurtosis in an effort to produce data that is closer to having a Gaussian structure. For Gaussian time series variables, the additive group structure is extremely natural: optimal mean square error estimates of quantities of interest (such as future values, missing values, unknown signals, etc.) are linear in the data, and hence are intimately linked to the addition operator. Errors in estimation are assessed by comparing estimator and target via subtraction – this applies to signal extraction,

[1] The original use of the logarithm, as invented by John Napier, was to assist in the computation of multiplications of numbers [7].

forecasting, and any other Gaussian prediction problem. Therefore the additive operator is quite natural for relating quantities in the transformation space.

It is for the above reasons (the linearity of estimators when the data is Gaussian) that the sum of signal and noise estimates equals the data process; no other algebraic operation is natural for relating Gaussian signal and noise. Given an observed time series $\{\mathbf{X}_t\}$ in the parent space, say for $1 \leq t \leq n$, the analyst would select φ via exploratory analysis such that $X_t = \varphi(\mathbf{X}_t)$ is representable as a sample from a Gaussian process. Most of the classical results on signal extraction [2][8] and projection [5] are interpretable in terms of a Gaussian distribution. More precisely, the estimates commonly used in time series applications minimize the mean squared prediction error among all linear estimators, and are also conditional expectations when the process is Gaussian. If φ does not produce a Gaussian distribution, at a minimum it should reduce skewness and kurtosis in the marginal distributions.

Also, it is necessary that φ be invertible, and it will be convenient for it to be a continuously differentiable function. Denoting the joint probability distribution function (pdf) of the transformed data by $p_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$, the joint pdf of the original data is then

$$p_{\mathbf{X}_1,\cdots,\mathbf{X}_n}(\mathbf{x}_1,\cdots,\mathbf{x}_n) = p_{X_1,\cdots,X_n}(x_1,\cdots,x_n) \cdot \Pi_{t=1}^n \frac{\partial \varphi(\mathbf{x}_t)}{\partial x}. \tag{1}$$

Of course, here $x_t = \varphi(\mathbf{x}_t)$. If we select a parametric family to model $p_{X_1,\cdots,X_n}$, e.g., a multivariate Gaussian pdf, then (1) can be viewed as a function of model parameters rather than of observed data, and we obtain the likelihood. It is apparent that the Jacobian factor does not depend on the parameters, and hence is irrelevant for model fitting purposes. That is, the model parameter estimates are unchanged by working with the likelihood in the parent space.

There may be estimates of interest in the transformation space, which are some functions of the transformed data. Typically we have some quantity of interest Z that we estimate via $\widehat{Z}$ in the transformed space, perhaps computed as a linear function of the transformed data (though the linearity of the statistic is not required for this discussion). If we have a measure of the uncertainty in $\widehat{Z}$, we can compute probabilities such as $\mathbb{P}[a \leq \widehat{Z} \leq b]$ and $\mathbb{P}[a \leq \widehat{Z} - Z \leq b]$. Since φ is invertible, the former probability can be immediately converted into a confidence interval for the parent space, via

$$\mathbb{P}\left[\varphi^{-1}(a) \leq \varphi^{-1}(\widehat{Z}) \leq \varphi^{-1}(b)\right].$$

This assumes that φ is increasing (else the inequalities will be flipped around). Then our estimate of $\mathbf{Z} = \varphi^{-1}(Z)$ would be $\varphi^{-1}(\widehat{Z})$, with uncertainty interval given by the above equation; a knowledge of the probability in the transformed space immediately provides the probability in the parent space. However, when uncertainty about an estimate is assessed in terms of its relation to a target quantity Z, which may be stochastic, it is less obvious how to proceed. This is typically the situation in time series analysis, where Z is often a signal or a future value of the data process, and so is a stochastic quantity. If we apply the inverse transformation to $\mathbb{P}[a \leq \widehat{Z} - Z \leq b]$ we obtain

$$\mathbb{P}\left[\varphi^{-1}(a) \leq \varphi^{-1}(\widehat{Z}-Z) \leq \varphi^{-1}(b)\right], \tag{2}$$

which tells us nothing of the relationship of $\mathbf{Z}$ to its estimate $\varphi^{-1}(\widehat{Z})$. That is, there should be some algebraic relation between $\mathbf{Z}$ and $\varphi^{-1}(\widehat{Z})$ such that a suitable notion of their discrepancy can be assessed probabilistically, and (2) can become interpretable in terms of natural quantities in the parent domain. Supposing that some operator $\oplus$ were defined such that $\varphi^{-1}(\widehat{Z}-Z) = \varphi^{-1}(\widehat{Z}) \oplus \mathbf{Z}^{-1}$, for an appropriate notion of the inverse of $\mathbf{Z}$, then we could substitute into (2) and obtain a confidence interval for the statistical error. The next section develops the unique operator $\oplus$ possessing the requisite properties.

3 Algebraic Structure of the Parent Space

Given an additive operation in the transformed space, e.g., $x_t + x_{t-1}$, it is crucial to define a corresponding composition rule $\oplus$ in the parent domain such that φ is a group homomorphism. A group is a set together with an associative composition law, such that an identity element exists and every element has an inverse [1]. A homomorphism is a transformation of groups such that the laws of composition are respected. The groups under consideration are $\mathscr{R} = (\mathbb{R}, +)$ for the transformed space, and $\mathscr{G} = (\varphi^{-1}(\mathbb{R}), \oplus)$ for the parent space. Consider the situation of latent components in the transformed space, where $X_t = S_t + N_t$ is a generic signal-noise decomposition. Then the components in the parent space are $\varphi^{-1}(S_t) = \mathbf{S}_t$ and $\varphi^{-1}(N_t) = \mathbf{N}_t$, which can be quantities of interest in their own right. How do we define an algebraic structure that allows us to combine $\mathbf{S}_t$ and $\mathbf{N}_t$, such that the result is always $\mathbf{X}_t$? What is needed is a group operator $\oplus$ such that

$$\mathbf{S}_t \oplus \mathbf{N}_t = \mathbf{X}_t = \varphi^{-1}(S_t + N_t) = \varphi^{-1}(\varphi(\mathbf{S}_t) + \varphi(\mathbf{N}_t)).$$

This equation actually suggests the definition of $\oplus$: any two elements $\mathbf{a}, \mathbf{b}$ in the parent group $\mathscr{G}$ are summed via the rule

$$\mathbf{a} \oplus \mathbf{b} = \varphi^{-1}(\varphi(\mathbf{a}) + \varphi(\mathbf{b})). \tag{3}$$

This definition "lifts" the additive group structure of $\mathscr{R}$ to $\mathscr{G}$ such that: (1) $\varphi^{-1}(0) = 1_G$ is the unique identity element of $\mathscr{G}$; (2) $\mathscr{G}$ has the associative property; (3) the unique inverse of any $\mathbf{a} \in \mathscr{G}$ is given by $\mathbf{a}^{-1} = \varphi^{-1}(-\varphi(\mathbf{a}))$. These properties are verified below, and establish that $\mathscr{G}$ is indeed a group. Moreover, the group is Abelian and φ is a group isomorphism.

First, $\mathbf{a} \oplus \varphi^{-1}(0) = \varphi^{-1}(\varphi(\mathbf{a}) + 0) = \mathbf{a}$, which together with the reverse calculation shows that $\varphi^{-1}(0)$ is an identity; uniqueness similarly follows. Associativity is a book-keeping exercise. For the inverse, note that $\mathbf{a} \oplus \varphi^{-1}(-\varphi(\mathbf{a})) = \varphi^{-1}(\varphi(\mathbf{a}) - \varphi(\mathbf{a})) = \varphi^{-1}(0) = 1_{\mathscr{G}}$. This shows that $\mathscr{G}$ is a group, and commutativity follows from (3) and the commutativity of addition; hence $\mathscr{G}$ is an Abelian group. Finally, φ is a bijection as well as a homomorphism, i.e., it is an isomorphism.

What goes wrong if we use another composition rule to define $\mathscr{G}$? We would lose the group structure, and more importantly we no longer have the important property that $\varphi(\mathbf{X}_t) = \varphi(\mathbf{S}_t) + \varphi(\mathbf{N}_t)$. For example, suppose that $\varphi(x) = \mathrm{sign}(x)\sqrt{|x|}$, and for illustration suppose that $\mathbf{X}_t, \mathbf{S}_t, \mathbf{N}_t$ are all positive. But if an additive structure is assigned to the parent space, then we would have $\mathbf{X}_t = \mathbf{S}_t + \mathbf{N}_t$, and as a consequence $\sqrt{\mathbf{X}_t} = \sqrt{\mathbf{S}_t + \mathbf{N}_t} \neq \sqrt{\mathbf{S}_t} + \sqrt{\mathbf{N}_t}$. Instead, $\oplus$ should be defined via (for positive inputs $\mathbf{a}$ and $\mathbf{b}$) the following: $\mathbf{a} \oplus \mathbf{b} = \mathbf{a} + \mathbf{b} + 2\sqrt{\mathbf{ab}}$. Now this example results in an unfamiliar operator for $\oplus$, but when φ is the logarithm, we obtain multiplication. Although some conceptual realignment is required, the requisite algebraic structure is uniquely determined by φ and cannot be wished away.

Example 1: Logarithm

Suppose that $\varphi(x) = \log x$ and the domain is all positive real numbers. Then $\mathbf{a} \oplus \mathbf{b} = \exp\{\log \mathbf{a} + \log \mathbf{b}\} = \mathbf{a} \cdot \mathbf{b}$, i.e., the group operator is multiplication. The identity element of $\mathscr{G}$ is unity, and inverses of elements are their reciprocals. This is a familiar case, and it works out nicely; the homomorphic property of the logarithm is well-known. In application, seasonal noise is viewed in the parent domain as a "seasonal factor" that divides the data, with the residual being the seasonally adjusted data.

Example 2: Box-Cox

Suppose that $\varphi(x) = \mathrm{sign}(x)|x|^{\lambda}$, which is essentially a Box-Cox transform (see [4]) when $\lambda \in (0,1]$. The case $\lambda = 1$ is trivial, and $\lambda \to 0$ essentially encompasses the case of logarithmic transformation. Typically the transform is utilized on positive data, but we include the sign operator to ensure the homomorphic property, as well as invertibility of φ. The composition law in $\mathscr{G}$ is then

$$\mathbf{a} \oplus \mathbf{b} = \mathrm{sign}\left(\mathrm{sign}(\mathbf{a})|\mathbf{a}|^{\lambda} + \mathrm{sign}(\mathbf{b})|\mathbf{b}|^{\lambda}\right) \cdot \left|\mathrm{sign}(\mathbf{a})|\mathbf{a}|^{\lambda} + \mathrm{sign}(\mathbf{b})|\mathbf{b}|^{\lambda}\right|^{1/\lambda}.$$

The identity is also zero, and $\mathbf{a}^{-1} = \mathrm{sign}(-\mathbf{a})|\mathbf{a}|$. When we restrict the spaces to $\mathbb{R}^{+}$, the rule simplifies to $\mathbf{a} \oplus \mathbf{b} = (\mathbf{a}^{\lambda} + \mathbf{b}^{\lambda})^{1/\lambda}$ (but then additive inverses are not well-defined, and $\mathscr{R}$ becomes a semi-group).

Example 3: Logistic

Suppose that $\varphi(x) = \log(x) - \log(1-x)$ defined on $(0,1)$, with inverse $e^{x}/(1+e^{x})$. This transform is sometimes used for bounded data that represents a percentage or rate. The composition law is

$$\mathbf{a} \oplus \mathbf{b} = \frac{\mathbf{ab}}{1 - \mathbf{a} - \mathbf{b} + 2\mathbf{ab}}$$

with identity element $1/2$ and inverses $\mathbf{a}^{-1} = 1 - \mathbf{a}$. This rule tells one way that percentages may be composed so as to ensure the result is again a percentage.

Example 4: Hyperbolic Sine

The function $\varphi(x) = (e^x - e^{-x})/2$ is the hyperbolic sine transformation, which maps $\mathbb{R}$ to $\mathbb{R}$, with inverse $\varphi^{-1}(y) = \log(y + \sqrt{y^2+1})$. Then the composition law is

$$\mathbf{a} \oplus \mathbf{b} = \varphi^{-1}\left((e^{\mathbf{b}} - e^{-\mathbf{a}})(1 + e^{\mathbf{a}-\mathbf{b}})/2\right).$$

The identity element is zero, and inverses are the same as in $\mathscr{R}$, i.e., $\mathbf{a}^{-1} = -\mathbf{a}$.

Example 5: Distributional Transforms

Any random variable with continuous invertible cumulative distribution function (cdf) F can be transformed to a standard Gaussian variable via $\varphi = \Xi \circ F$, where Ξ is the quantile function of the standard normal. Letting Φ denote the Gaussian cdf and $Q = F^{-1}$ the given variable's quantile function, clearly $\varphi^{-1} = Q \circ \Phi$. This transform takes a random variable with cdf F in the parent domain to a Gaussian variable, and the corresponding composition rule is

$$\mathbf{a} \oplus \mathbf{b} = Q\left\{\Phi\left(\Xi[F(\mathbf{a})] + \Xi[F(\mathbf{b})]\right)\right\}.$$

For example, F might correspond to a χ^2, student t, uniform, or Weibull distribution. A χ^2 variable on 2 degrees of freedom (i.e., an exponential variable) has $F(x) = 1 - e^{-x}$, with $Q(u) = -\log(1-u)$. Then

$$\mathbf{a} \oplus \mathbf{b} = -\log\left\{1 - \Phi\left(\Xi[1 - e^{-\mathbf{a}}] + \Xi[1 - e^{-\mathbf{b}}]\right)\right\}$$

defines the composition law.

4 Numerical Illustrations

In order to assess the degree of distortion that $\oplus$ generates in quantities, in comparison to the $+$ operator, one can examine the level curves $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$ for various values of $\mathbf{c}$, i.e.,

$$L_{\mathbf{c}} = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \oplus \mathbf{b} = \mathbf{c}\} = \{(\mathbf{a}, \mathbf{c} \oplus \mathbf{a}^{-1}) : \mathbf{a}, \mathbf{c} \in \varphi^{-1}(\mathbb{R})\}.$$

When φ is the identity mapping, the level curves are just the lines of slope -1, with y intercepts given by various values of $\mathbf{c}$. By plotting the various level curves $L_{\mathbf{c}}$ in comparison to the straight lines for the operator $+$, we can form a notion of the extent of distortion involved to the group structure of $\mathscr{R}$.

To compute the level curves, we must calculate $\mathbf{c} \oplus \mathbf{a}^{-1}$ in each case, which we write as $f_{\mathbf{c}}(\mathbf{a})$ for short; then the level curve is the graph of $f_{\mathbf{c}}$. For the logarithmic transform, $f_{\mathbf{c}}(\mathbf{a}) = \mathbf{c}/\mathbf{a}$. For the Box-Cox, the general formula is cumbersome. For example, when $\mathbf{a}, \mathbf{c} > 0$ and $\lambda = 1/2$ we obtain $f_{\mathbf{c}}(\mathbf{a}) = \text{sign}(\sqrt{\mathbf{c}} - \sqrt{\mathbf{a}}) \cdot |\sqrt{\mathbf{c}} - \sqrt{\mathbf{a}}|^2$. For the logistic, we have

$$f_{\mathbf{c}}(\mathbf{a}) = \frac{\mathbf{c}(1-\mathbf{a})}{\mathbf{a} - \mathbf{c} + 2\mathbf{c}(1-\mathbf{a})}.$$

For hyperbolic sine, we have $f_{\mathbf{c}}(\mathbf{a}) = \varphi^{-1}((e^{-\mathbf{a}} - e^{-\mathbf{c}})(1 + e^{\mathbf{c}+\mathbf{a}})/2)$, which does not simplify neatly. For distributional transforms,

$$f_{\mathbf{c}}(\mathbf{a}) = Q\left\{\Phi\left(\Xi[F(\mathbf{c})] - \Xi[F(\mathbf{a})]\right)\right\}.$$

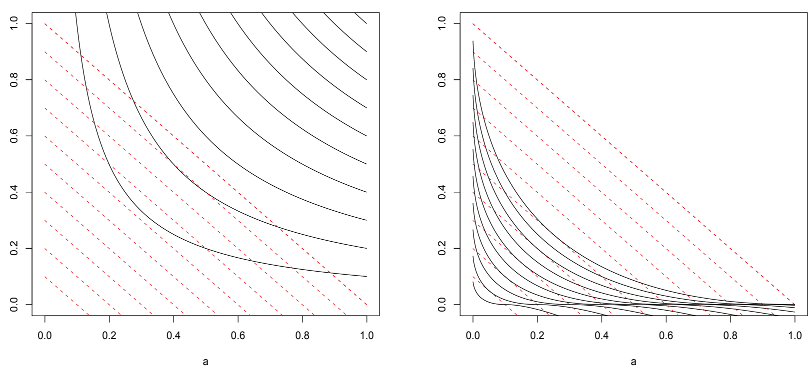

Fig. 1 Level curves L_c for the logarithmic (left panel) and square root Box-Cox (right panel) transformations. The red (dotted) lines are level curves for $a+b$, while the black (solid) lines are level curves for $a \oplus b$, where $c = .1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0$.

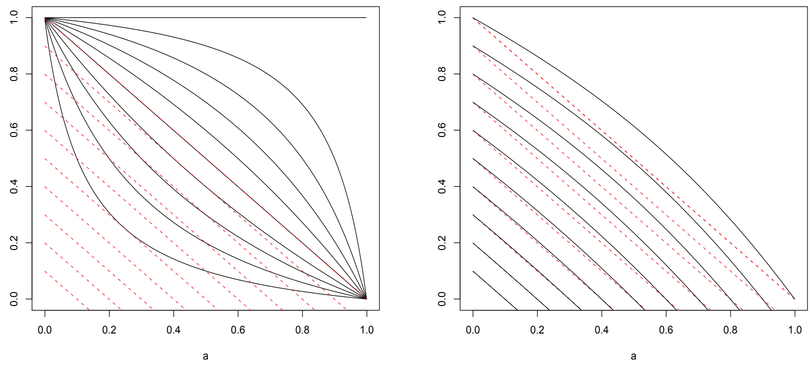

Fig. 2 Level curves L_c for the logistic (left panel) and hyperbolic sine (right panel) transformations. The red (dotted) lines are level curves for $a+b$, while the black (solid) lines are level curves for $a \oplus b$, where $c = .1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0$.

Various level curves are plotted in Figures 1 - 4. We focus on values of $\mathbf{c} = i/10$ for $1 \leq i \leq 10$, and all values of $\mathbf{a} \in [0,1]$. We consider Examples 1 and 2 in Figure 1, Examples 3 and 4 in Figure 2, and Example 5 in Figures 3 and 4, where the distributional transforms include student t with 2 degrees of freedom, χ^2 with 1 degree of freedom, the uniform (on $[0,1]$) distribution, and the Weibull with shape parameter 1.5 and unit scale.

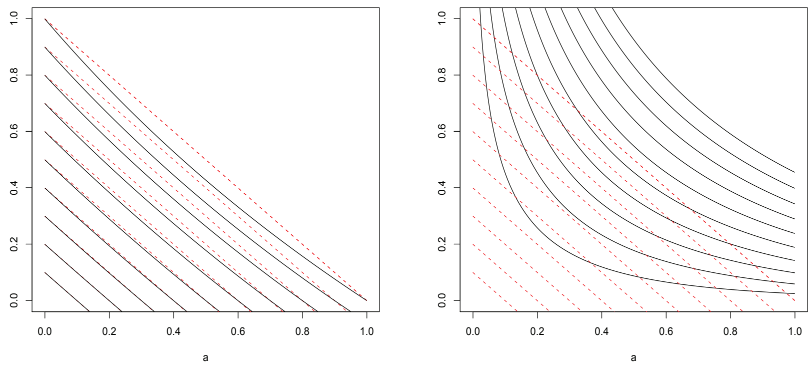

Fig. 3 Level curves L_c for the student t (left panel) and χ^2 (right panel) transformations. The red (dotted) lines are level curves for $a+b$, while the black (solid) lines are level curves for $a \oplus b$, where $c = .1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0$.

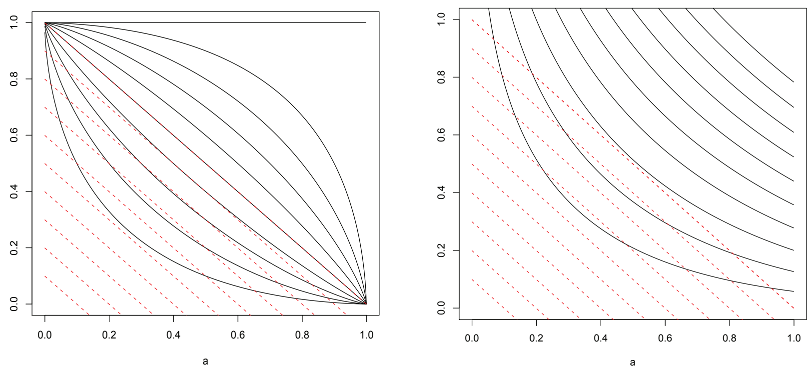

Fig. 4 Level curves L_c for the uniform (left panel) and Weibull (right panel) transformations. The red (dotted) lines are level curves for $a+b$, while the black (solid) lines are level curves for $a \oplus b$, where $c = .1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0$.

We see that the level curves for the logistic and uniform transforms are similar (though not identical, see Figures 2 and 4), which is intuitive since they both map the space $[0,1]$ into $\mathbb{R}$. Also, the logarithmic (Figure 1), χ^2 (Figure 3), and Weibull (Figure 4) are quite similar. The hyperbolic sine (Figure 2) and student t (Figure 3) both offer little distortion, but have opposite curvature.

5 Empirical Illustrations

How much does the $\oplus$ operator differ from addition in practice? To answer this query we provide two data illustrations. In these examples, a data set $\{\mathbf{X}_t\}$ that has undergone some transformation φ will be seasonally adjusted using X-12-ARIMA [6], being applied to $X_t = \varphi(\mathbf{X}_t)$ in the transformed space. We then demonstrate that the additive decomposition into seasonal and nonseasonal components in the transformed space does not reproduce an additive decomposition of these inverted components in the parent space. For this application, if X_t decomposes into a trend-cycle C_t, seasonal S_t, and irregular component I_t, then we have $X_t = C_t + S_t + I_t$. The nonseasonal component will be composed of the trend-cycle and the irregular effect, so let us label the adjusted component N_t as $N_t = C_t + I_t$. Note that while the signal-noise decomposition uses S_t to denote signal and N_t to denote noise, for the additive seasonal decomposition described here, the nonseasonal portion N_t is the signal, and the seasonal component S_t is the noise. What we show is that although $X_t = N_t + S_t$ and $\mathbf{X}_t = \varphi^{-1}(X_t)$ are both true, $\varphi^{-1}(X_t)$ can be quite a bit different from $\varphi^{-1}(N_t) + \varphi^{-1}(S_t) = \mathbf{N}_t + \mathbf{S}_t$ when φ is not a linear transformation.

5.1 Example 1: Square root transform

The first example is the U.S. Census Bureau monthly series of total inventory of nonmetallic mineral products in the U.S., between the years 1992 and 2011. The default square root transform in X-12-ARIMA is $0.25 + 2(\sqrt{\mathbf{X}_t} - 1)$, which is a shifted and scaled version of the basic $\sqrt{\mathbf{X}_t}$ square root transform, and the two adjusted log likelihoods are identical. Using X-12-ARIMA's automdl and transform specs, we compare the square root transform to both a logarithmic transform and to no transform at all. Typically, the model with the smallest AICC would be preferred over other contenders, but since the same SARIMA (0 2 1)(0 1 1) model was found to fit all three transforms of the data, the best transform would equivalently be indicated by the highest log likelihood. Table 1 displays the log likelihoods (adjusted for the transformations) along with the corresponding AICC. We see that the square root transform yields the highest log likelihood in the parent space and also the lowest value for AICC; this leads us to prefer the use of a square root transform for this total inventory series.

Table 1 Comparison of log likelihoods and AICCs for three transforms of total inventory data.

Transform	(Adj.) log likelihood	AICC
None	-1353.0860	2712.2800
Logarithm	-1352.1446	2710.3974
Square root	-1350.8787	2707.8655

We proceed by using X-12-ARIMA to obtain an additive decomposition of the series $\{X_t\}$, where X_t is just the square root of $\mathbf{X}_t$. In checking the difference series $X_t - (N_t + S_t)$, we note that the differences appear to be centered around 0, with a maximum magnitude no greater than 5×10^{-13}; numerical error from rounding and computer precision explains why this difference is not identically 0. Similar results hold for the difference between $\mathbf{X}_t$ and $\mathbf{N}_t \oplus \mathbf{S}_t$, which is just the application of φ^{-1} to $N_t + S_t$. However, there are substantial discrepancies between $\mathbf{X}_t$ and $\mathbf{N}_t + \mathbf{S}_t$, as expected. For $\mathbf{N}_t = \varphi^{-1}(N_t)$ and $\mathbf{S}_t = \varphi^{-1}(S_t)$, Figure 5 shows a plot of the untransformed series $\mathbf{X}_t$ along with $\mathbf{N}_t + \mathbf{S}_t$ on the top panel, and on the bottom panel, we have the difference series obtained by subtracting $\mathbf{N}_t + \mathbf{S}_t$ from $\mathbf{X}_t$. The top panel of Figure 5 confirms that the additive decomposition in transformed space does not translate to an additive decomposition in parent space, and the bottom panel shows that the deviations from 0 in this case are quite pronounced. Furthermore, while the lower panel of Figure 5 indicates that the differences are roughly unbiased (the series is centered around zero), it also displays a highly seasonal pattern evincing some heteroskedasticity. We explain this behavior below.

Noting that the seasonal S_t can be negative, it follows that $\mathbf{S}_t$ can be negative as well; however, if the original data $\mathbf{X}_t$ is always positive, it follows that

$$\mathbf{S}_t \oplus \mathbf{N}_t = \mathbf{S}_t + \mathbf{N}_t + \mathrm{sign}(\mathbf{S}_t \mathbf{N}_t)\,\sqrt{|\mathbf{S}_t|\,|\mathbf{N}_t|}.$$

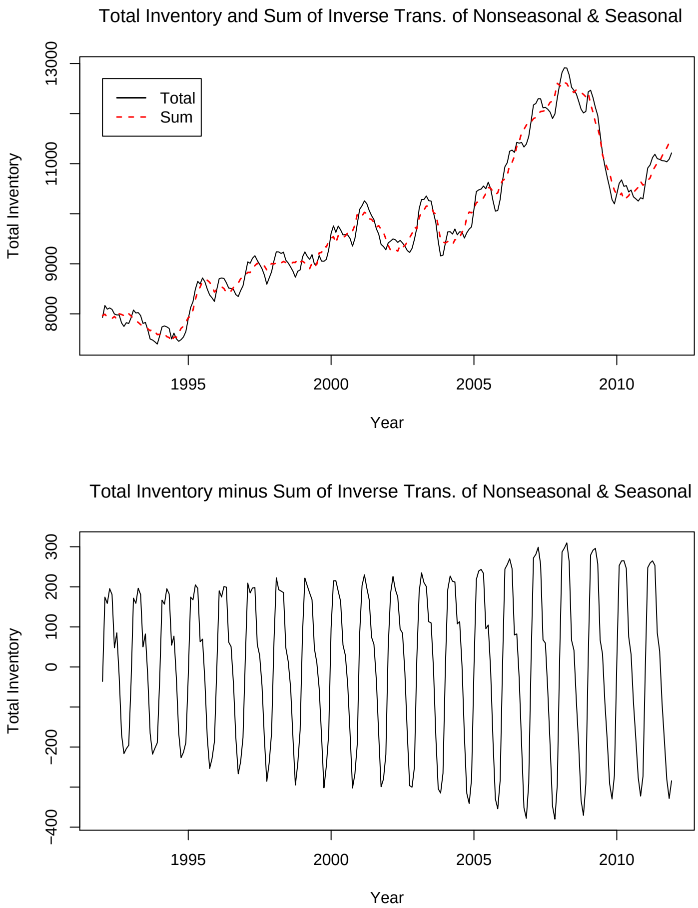

Fig. 5 The top plot shows $\mathbf{X}_t$ and $\mathbf{N}_t + \mathbf{S}_t$ together, while the bottom plot displays $\mathbf{X}_t - (\mathbf{N}_t + \mathbf{S}_t)$, where $\mathbf{X}_t$ is the series for U.S. total inventory of nonmetallic mineral products between 1992 and 2011. $\mathbf{N}_t$ and $\mathbf{S}_t$ are the signed squares of N_t and S_t, the nonseasonal and seasonal components from an additive decomposition of $X_t = \sqrt{\mathbf{X}_t}$.

Typically $\mathbf{N}_t$ is positive as well, so that

$$\mathbf{S}_t \oplus \mathbf{N}_t - (\mathbf{S}_t + \mathbf{N}_t) = \text{sign}(\mathbf{S}_t)\,\sqrt{|\mathbf{S}_t|}\,\sqrt{\mathbf{N}_t}.$$

Thus, the discrepancy between $\oplus$ and the addition operator is equal to the square root of the product of the seasonal and nonseasonal, multiplied by the sign of the seasonal; we can expect this time series to be centered around zero, because S_t is centered around zero. This explains the seasonal behavior of the lower panel in Figure 5.

5.2 Example 2: Logistic transform

The second example is the monthly unemployment rate for 16 to 19 year old individuals of Hispanic origin between the years 1991 and 2011; the data was obtained from the Bureau of Labor Statistics. For rate data, the logistic transform $\varphi(\mathbf{a}) = \log(\mathbf{a}) - \log(1-\mathbf{a})$ is sometimes warranted, as it ensures fits and predictions that are guaranteed to fall between 0 and 1. As in the previous example, we use X-12-ARIMA's automdl and transform specs to help us compare the logistic transform to both a logarithmic transform and to no transform at all. Again, the procedure selects the same SARIMA (0 1 1)(0 1 1) model for all three transforms, so whichever transform has the highest log likelihood in the parent space will also have the lowest AICC. Table 2 displays the log likelihoods (adjusted for the transformations) along with the corresponding AICC, and we see that the logistic transform does indeed result in a better model compared to the other two transformations.

Table 2 Comparison of log likelihoods and AICCs for three transforms of unemployment rate data.

Transform	(Adj.) log likelihood	AICC
None	506.2443	-1006.3864
Logarithm	508.9064	-1011.7107
Logistic	511.0460	-1015.9900

We proceed by performing a logistic transform on $\mathbf{X}_t$ and then running X-12-ARIMA on the transformed series to obtain an additive seasonal decomposition. Checking the series of differences $X_t - (N_t + S_t)$, we find that the magnitude of the differences is bounded by 6×10^{-15}. These deviations from 0 are entirely explained by numerical error produced from passing the data through X-12-ARIMA. Similar results hold for $\mathbf{X}_t - (\mathbf{N}_t \oplus \mathbf{S}_t)$. But there are notable discrepancies between $\mathbf{X}_t$ and $(\mathbf{N}_t + \mathbf{S}_t)$, as in the previous illustration, as shown in Figure 6. The top panel shows that the additive nature of the decomposition in transformed space is not preserved when mapped back to the parent space, while the bottom panel shows that this discrepancy (in the parent space) is a time series centered around -0.5. Also, the lower panel of discrepancies $\mathbf{X}_t - (\mathbf{N}_t + \mathbf{S}_t)$ exhibits seasonal structure; we explain this phenomenon next.

For the logistic transform, the composition operator $\oplus$ is defined as

$$\mathbf{S}_t \oplus \mathbf{N}_t = \frac{\mathbf{S}_t \cdot \mathbf{N}_t}{1 - \mathbf{S}_t - \mathbf{N}_t + 2\mathbf{S}_t \cdot \mathbf{N}_t},$$

where $\mathbf{S}_t$ and $\mathbf{N}_t$ in the parent space are mapped using $1/(1+e^{-S_t})$ and $1/(1+e^{-N_t})$ from the transformed space. To explain the behavior of the lower panel in Figure 6, we calculate the difference:

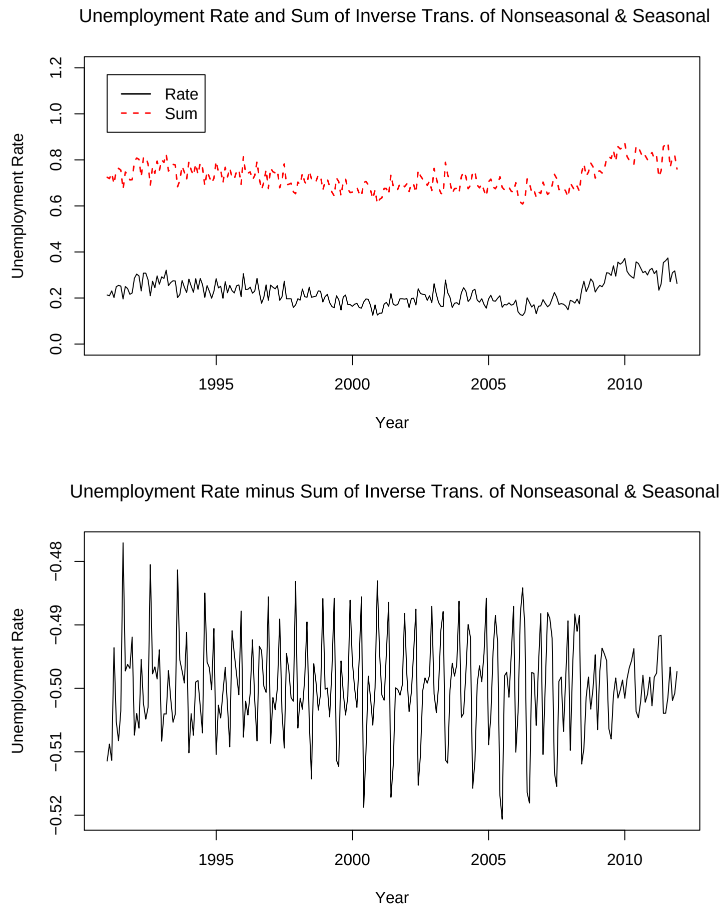

Fig. 6 The top panel shows $\mathbf{X}_t$ and $\mathbf{N}_t + \mathbf{S}_t$ together, while the bottom panel displays $\mathbf{X}_t - (\mathbf{N}_t + \mathbf{S}_t)$, where $\mathbf{X}_t$ is the unemployment rate among 16-19 year old Hispanic individuals between 1991 and 2011. $\mathbf{N}_t$ and $\mathbf{S}_t$ are the inverse transforms of the N_t and S_t from the additive decomposition of $X_t = \log(\mathbf{X}_t) - \log(1 - \mathbf{X}_t)$.

$$\begin{aligned}
\mathbf{S}_t \oplus \mathbf{N}_t - (\mathbf{S}_t + \mathbf{N}_t) &= \frac{\mathbf{S}_t \cdot \mathbf{N}_t}{1 - (\mathbf{S}_t + \mathbf{N}_t) + 2\mathbf{S}_t \cdot \mathbf{N}_t} - (\mathbf{S}_t + \mathbf{N}_t) \\
&= -\frac{1}{2} + \frac{1}{2}(\mathbf{S}_t + \mathbf{N}_t - 1)\left\{ \frac{1}{1 - (\mathbf{S}_t + \mathbf{N}_t) + 2\mathbf{S}_t \cdot \mathbf{N}_t} - 2 \right\}.
\end{aligned}$$

Given that $\mathbf{S}_t$ and $\mathbf{N}_t$ are both restricted between 0 and 1, the second term in the final expression above is a time series that fluctuates about zero (we cannot claim that its expectation is zero). This explains why the discrepancies in parent space were centered around -0.5. The second part of the sum helps account for the variation around the -0.5 center in the discrepancies $\mathbf{S}_t \oplus \mathbf{N}_t - (\mathbf{S}_t + \mathbf{N}_t)$.

6 Discussion

The primary applications of time series analysis are forecasting and signal extraction. In the transformed space, the data process is equal to signal plus noise, but their proper relation is different in the parent space, being given by $\mathbf{S}_t \oplus \mathbf{N}_t = \mathbf{X}_t$. Also, for Gaussian time series the forecast error is defined via $\widehat{X}_{t+1} - X_{t+1}$, which in the parent space becomes $\widehat{\mathbf{X}}_{t+1} \oplus \mathbf{X}_{t+1}^{-1}$. If the transformation is logarithmic, the forecast error in the parent space is the ratio of estimate and future value. Other relations can be worked out for the logistic and distributional transformations.

There is also much applied interest in growth rates, which in the transformed space is given by definition as $X_t - X_{t-1}$ (these might also be computed in terms of a signal of interest, say $S_t - S_{t-1}$). For a logarithmic transform, the growth rate becomes $\mathbf{X}_t / \mathbf{X}_{t-1}$ in the parent space, which might be interpreted as a percent increase over the previous value. But a growth rate for another transformation looks much different, e.g., in the logistic case

$$\mathbf{X}_t \oplus \mathbf{X}_{t-1}^{-1} = \frac{\mathbf{X}_t (1 - \mathbf{X}_{t-1})}{\mathbf{X}_{t-1} - \mathbf{X}_t + 2\mathbf{X}_t (1 - \mathbf{X}_{t-1})}.$$

Likewise, growth rate formulas can be written down for the other transformations, although typically the expressions do not simplify so neatly as in the logarithmic and logistic cases.

These new formulas for growth rates, forecast errors, and relations of signal and noise to data can be counter-intuitive. Only with the logarithmic transformation do we attain a recognizable group operation, namely multiplication. In order for φ to be a homomorphism of groups – which is needed so that quantities in the parent space can be meaningfully combined – one must impose a new group operator on the parent space, and oftentimes this operator $\oplus$ results in unfamiliar operations. However, there seems to be no rigorous escape from the demands of the homomorphism, and familiarity can develop from intimacy.

To illustrate a particular conundrum resolved by our formulation, consider the case alluded to in Section 2, where $\mathbf{Z}$ represents a forecast or signal of interest in the parent domain, and $\varphi^{-1}(\widehat{Z})$ is its estimate. Note that $\varphi^{-1}(Z) = \mathbf{Z}^{-1}$, and the corresponding error process is then $\varphi^{-1}(\widehat{Z}) \oplus \mathbf{Z}^{-1}$. The probability (2) becomes

$$\mathbb{P}\left[\varphi^{-1}(a) \leq \varphi^{-1}(\widehat{Z}) \oplus \mathbf{Z}^{-1} \leq \varphi^{-1}(b)\right].$$

Hence the confidence interval for the statistical error (in the parent domain) is expressed as $[\varphi^{-1}(a), \varphi^{-1}(b)]$, which exactly equals the probability that in the transformed domain $\widehat{Z}-Z$ lies in $[a,b]$. This type of interpretation is not possible unless φ is a homomorphism, which the particular definition of $\oplus$ guarantees.

We can also manipulate $\mathbb{P}[a \leq \widehat{Z}-Z \leq b]$ to obtain an interval for $\mathbf{Z}$:

$$\mathbb{P}\left[a \leq \widehat{Z}-Z \leq b\right] = \mathbb{P}\left[\widehat{Z}-b \leq Z \leq \widehat{Z}-a\right] = \mathbb{P}\left[\varphi^{-1}(\widehat{Z}-b) \leq \mathbf{Z} \leq \varphi^{-1}(\widehat{Z}-a)\right].$$

Although the last expression allows us to easily compute the interval for $\mathbf{Z}$, it is not directly expressed in terms of the parent estimate $\varphi^{-1}(\widehat{Z})$. Using the homomorphism property, the interval can be written as

$$\left[\varphi^{-1}(\widehat{Z}) \oplus \mathbf{b}^{-1}, \varphi^{-1}(\widehat{Z}) \oplus \mathbf{a}^{-1}\right].$$

In summary, the chief applications of time series analysis dictate that quantities in the parent space of a transformation must satisfy certain algebraic relations, and the proper way to ensure this structure is to define a group operator $\oplus$ via (3). As a consequence, the notions of statistical error (for forecasts, imputations, signal extraction estimates, etc.) are altered accordingly, as are the definitions of growth rates and the relations of signal and noise to data. Such relations are already intuitive and well-accepted when the transformation is logarithmic, but for other transforms there remains quite a bit of novelty.

Disclaimer

This article is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

References

1. Artin, M.: Algebra. Prentice Hall, New Jersey (1991)
2. Bell, W.: Signal extraction for nonstationary time series. Ann. Stat. **12**, 646–664 (1984)
3. Bell, W., Hillmer, S.: Issues involved with the seasonal adjustment of economic time series. J. Bus. Econ. Stat. **2**, 291–320 (1984)
4. Box., G., Jenkins, G.: Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco (1976)
5. Brockwell, P., Davis, R.: Time Series: Theory and Methods, 2nd Ed. Spring, New York (1991)
6. Findley, D., Monsell, B., Bell, W., Otto, M., Chen, B.: New capabilities and methods of the X-12-ARIMA seasonal adjustment program. J. Bus. Econ. Stat. **16**, 127–177 (1998)
7. McElroy, T.: A to Z of Mathematicians. Facts on File, New York (2005)
8. McElroy, T.: Matrix formulas for nonstationary ARIMA signal extraction. Econom. Theory **24**, 1–22 (2008)
9. McElroy, T.: An alternative model-based seasonal adjustment that reduces over-adjustment. Taiwan Econ. Forecast & Policy **43**, 35–73 (2012)