Advances in Science, Technology and Environmentology
Special Issue on the "Financial & Pension Mathematical Science"
Vol. B12 (2015.3), 63–70

VARIANCE STABILIZING PROPERTIES OF BOX-COX TRANSFORMATION FOR DEPENDENT OBSERVATIONS

By Yan LIU

Waseda University

Box-Cox transformation is one of the most famous transformations to stabilize the variance of estimators. In this note, we focus on the dependent random variables with the multivariate Tweedie distributions. Under a new condition between dispersion parameters, we derive the formula for power parameter in the Box-Cox transformation for variance stabilization. The result shows that even in the dependent case, the same formula as that for identically and independent distributed random variables holds.

1. Introduction. Use of transformation in statistics has been considered for a long time. The main purposes for the use of transformation are summarized as: to make data have (a) the invariant variance under the changes of the mean after transformation; (b) the normally distributed variate after transformation; (c) the efficiency of the arithmetic average for any particular group of measurements after transformation; (d) linear and additive real effects after transformation, in the primitive study by Bartlett (1947). With these motivations, Box and Cox (1964) proposed a parametric family of transformations from the raw data X to the transformed data $X^{(\lambda)}$ by

$$X^{(\lambda)} = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log X, & \text{if } \lambda = 0. \end{cases} \tag{1.1}$$

The transformation is well known as the Box-Cox transformation. It is used in various fields such as medical science, epidemiology and environmentology because of its simplicity and efficiency.

There are two ways to apply the transformation. One is to apply the transformation to the parameter and the other is to apply to the raw data. The first approach began with Fisher's z-transformation for population correlation coefficient ρ in Fisher (1921). The phenomenon is well explained by the normal approximation in Konishi (1978). The general procedure for finding normalizing transformations for parameter is given in Konishi (1981). Taniguchi and Puri (1995) developed the higher order asymptotic theory for transformations of the maximum likelihood estimators in general statistical models.

The other way is directly applying the transformation to the raw data. The analyses can be further divided into two categories. One method to analyze the use of transformation is to assume the normality after transformation. The idea that the transformed observations are distributed as an appropriate linear model with normality and the parameter λ is estimated by maximum likelihood estimator originated from Box and Cox (1964). The assumption of

AMS 2000 subject classifications: Primary 62E20; Secondary 62F12
Keywords and phrases: variance stabilization, multivariate Tweedie distribution, Box-Cox transformation

normality was removed in Bickel and Doksum (1981). Recently, Hosoya and Terasaka (2009) assumed the transformed observations are distributed as a linear stationary process and they provided the whole picture of the inference by the Whittle estimator. This idea, however, is often associated with the issue that the transformed observations are not distributed as they are assumed although the estimator of the parameters in the model is assumed to be maximum likelihood estimator.

The second method to consider the problem is to assume the observations come from the Tweedie class in the exponential dispersion models. Tweedie (1947) introduced a new class with the relationship between the variance and the mean of the observations. This class matches the original idea for data transformation in Bartlett (1947). Jørgensen (1987) introduced exponential dispersion models and indicated that the Tweedie class in the exponential dispersion models has the property of power variance functions. Nishii (1991) examined the Tweedie class and the power transformation by the asymptotic expansion. Especially, he provided the exact power of the Box-Cox transformation corresponding to the Tweedie class. The approach, however, only deals with the transformation of the raw data in the i.i.d. case. In this note, we consider the dependent case for multivariate Tweedie distributions, which were introduced in Furman and Landsman (2010). In this note, we focus on the dependent random variables. We adopt the multivariate Tweedie class to derive the exact value of the power parameter for transformation. Surprisingly, not only when the samples are i.i.d., even in the case of dependent random variables, the formula given in Nishii (1991) also holds.

This note is organized as follows. In Section 2, we review the concept of variance stabilization and give the sufficient condition of variance stabilization of the power parameter in the Box-Cox transformation. In Section 3, we provide the definition of the multivariate dispersion models and their fundamental properties. We will present our main result on the optimal power parameter of variance stabilization in the Box-Cox transformation for dependent random variables in Section 4.

The notations and symbols are listed in the following: $\mathbb{R}$ denotes the set of all real numbers; ∂_i denotes the derivative respective to the ith element of θ, i.e., $\partial/\partial\theta_i$; $\mathrm{cum}_{(m)}(X)$ denotes the mth cumulant of X; $\to_p$ and $\to_d$ denote the convergence in probability and the convergence in law, respectively.

2. Variance stabilization. In this section, we derive the condition of the power parameter in the Box-Cox transformation for variance stabilization. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a collection of m-dimensional random variables and the parameter θ included in the model satisfy $\theta \in \Theta \subset \mathbb{R}^p$, where Θ is a closed subset. As a general setting, we suppose the parameter also satisfies the equation

$$E\psi_n(\boldsymbol{X}, \theta) = 0, \tag{2.1}$$

and for simplicity, we write $\psi(\theta) \equiv \psi_n(\boldsymbol{X}, \theta)$. Generally, the scalar parameter ρ of interest is a function of θ, i.e.,

$$\rho = g(\theta).$$

Suppose we estimate the parameter θ by solving (2.1), that is,

$$\psi(\hat{\theta}_n) = 0.$$

If the equation ψ satisfies the regularity conditions, then the estimate $g(\hat{\theta}_n)$ is a consistent estimator of ρ and the distribution of $g(\hat{\theta}_n)$ is asymptotically normal. In this note, we consider the transformation T of $g(\hat{\theta}_n)$ and give the sufficient condition for the transformation to be variance stabilizing one.

Definition 2.1. A transformation T is called a variance stabilizing transformation if the asymptotic variance of $T(g(\hat{\theta}_n))$ is independent of θ.

To be more concise, we suppose the estimation function, considerably general enough to contain the non-i.i.d. case, can be approximated by the log-likelihood function as

$$\begin{aligned} \psi_i &= n^{-1/2}\partial_i l_n(\theta), \\ \psi_{ij} &= n^{-1/2}(\partial_i\partial_j l_n(\theta) - E\partial_i\partial_j l_n(\theta)), \\ \psi_{ijk} &= n^{-1/2}(\partial_i\partial_j\partial_k l_n(\theta) - E\partial_i\partial_j\partial_k l_n(\theta)), \end{aligned}$$

where $i,j,k \in \{1,2,\ldots,p\}$, $l_n(\theta) = \log p_n(\boldsymbol{X};\theta)$. Here, we omit the terms of lower order.

Assumption 2.2. Suppose the asymptotic moments of ψ_i, ψ_{ij} and ψ_{ijk} are given as

$$\begin{aligned} E\psi_i\psi_j &= \Sigma_{ij} + O(n^{-1}), \\ E\psi_i\psi_{jk} &= J_{ijk} + O(n^{-1}), \\ E\psi_i\psi_j\psi_k &= \frac{1}{\sqrt{n}}K_{ijk} + O(n^{-3/2}) \end{aligned}$$

and Jth order ($J \geq 3$) cumulants of ψ_i, ψ_{ij} and ψ_{ijk} are all $O(n^{-J/2+1})$.

Under regularity conditions, $\hat{\theta}_n$ is asymptotically normal, that is,

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d \mathcal{N}(\mathbf{0}, \Sigma).$$

As a direct result,

$$\sqrt{n}(g(\hat{\theta}_n) - \rho) \to_d \mathcal{N}(\mathbf{0}, \Sigma^*),$$

where

$$\Sigma^* = \frac{\partial g(\theta)}{\partial \theta^T}\Sigma\frac{\partial g(\theta)}{\partial \theta}.$$

Proposition 2.3 (Taniguchi and Puri (1995)). *If T satisfies the differential equations*

$$\frac{T''}{T'} = (\Sigma^* g_i)^{-1}\{g_j g_k \Sigma^{jj'}\Sigma^{kk'}(K_{j'k'i} + J_{j'k'i} + J_{k'j'i})/2 - g_{ii'}g_{j'}\Sigma^{i'j'}\},$$

then T is the asymptotic variance stabilizing transformation.

This result is direct and straightforward. The proof can be found in Taniguchi and Kakizawa (2000). Among the most popular transformations, we focus on the Box-Cox transformation.

Suppose we apply the Box-Cox transformation to the scalar parameter of interest ρ. That is, from the definition (4.1), the parameter ρ is transformed by

$$\rho^{(\lambda)} = \begin{cases} \frac{\rho^{\lambda}-1}{\rho}, & \text{if } \lambda \neq 0, \\ \log \rho, & \text{if } \lambda = 0. \end{cases} \tag{2.2}$$

As for the asymptotic variance stabilizing transformation, we obtain the following result.

COROLLARY 2.4. Suppose the Box-Cox transformation of the parameter of interest ρ is defined by (2.2). Then the asymptotic variance-stabilizing transformation is achieved when λ satisfies

$$\lambda = 1 + \rho(\Sigma^* g_i)^{-1}\{g_j g_k \Sigma^{jj'} \Sigma^{kk'} (K_{j'k'i} + J_{j'k'i} + J_{k'j'i})/2 - g_{ii'} g_{j'} \Sigma^{i'j'}\}$$

for $i = 1, \ldots, p$.

This is a direct result of Proposition 2.3. When the power parameter λ of the Box-Cox transformation exists, we call λ the optimal power parameter for variance stabilization.

3. Multivariate Tweedie distributions. To illustrate the variance stabilizing properties of Box-Cox transformation on random variables, the most convenient class is considered as the class of Tweedie distributions. The class is contained in the class of exponential dispersion models. For i.i.d. random variables, the probability measures of exponential dispersion models are defined by

$$dP_\theta = \exp(\theta x - \Delta\kappa(\theta))\nu(dx),$$

where ν is a positive non-degenerate measure on $\mathbb{R}$ and the cumulant function $\kappa(\theta) = \log \int_{\mathbb{R}} \exp(\theta x)\nu(dx)$ with the parameter space

$$\Theta = \{\theta \in \mathbb{R}; \int_{\mathbb{R}} \exp(\theta x)\nu(dx) < \infty\} \neq \emptyset.$$

Denote random variable X in exponential dispersion models by $X \sim \mathrm{ED}(\theta, \Delta)$. The most attractive properties of the class of distributions is that it contains the power variance functions

$$V(\mu) = \mu^p. \tag{3.1}$$

The implicit relation between the mean and the variance of random variables to be transform is considered as early as in Bartlett (1947). Solving the differential equation with the relation (3.1), we have

$$\kappa(\theta) = \begin{cases} e^{\theta}, & p = 1, \\ -\log(-\theta), & p = 2, \\ \frac{\alpha-1}{\alpha}\left(\frac{\theta}{\alpha-1}\right)^{\alpha}, & p \neq 1, 2. \end{cases} \tag{3.2}$$

Exponential dispersion models with $\kappa(\theta)$ defined in (3.2) are univariate Tweedie distributions. Denote random variable X distributed as univariate Tweedie distributions by $X \sim \mathrm{T}(p, \theta, \Delta)$.

As shown in Jørgensen (1987), an exponential dispersion model is characterized by its variance functions among the class of all exponential dispersion models. The property guarantees the uniqueness of univariate Tweedie distributions with different parameters.

The drawback of univariate Tweedie distributions is that they only treat i.i.d. random variables. To consider the transformation of dependent random variables, we consider the multivariate Tweedie distributions. Actually, the generalization of univariate Tweedie distributions to multivariate ones has been considered for a long time. However, the main issue of the generalization is that the new family is not as rich as the univariate Tweedie distributions. These difficult things are depicted very well by the comment by Letac, which is cited in Jørgensen (2013). We recited it again here:

> While the names of distributions in $\mathbb{R}$ are generally unambiguous, at the contrary in the jungle of distributions in $\mathbb{R}^k$ almost nothing is codified outside of the Wishart and Gaussian cases. The scenario is usually as follows: choose a one-dimensional *thingy* type (quite often an exponential dispersion model, namely a natural exponential family and all its real power of convolution) such as the gamma or negative binomial; then any law in $\mathbb{R}^k$ whose margins are of *thingy* type are said to be a multidimensional *thingy*. Although the study of all distributions with given marginals are rather in the non-parametric domain of study, actually each author who isolates some parametric family will declare that he or she has THE multidimensional *thingy* family.

This is why it is better to consider a class of multivariate exponential dispersion models which may contain ARMA models. The multivariate distribution of random vector $\boldsymbol{X}$ can be constructed in the following way, which is proposed in Furman and Landsman (2010).

Theorem 3.1 (Furman and Landsman (2010)). *Suppose the matrix A is defined by*

$$A = \begin{pmatrix} \beta_1 & 1 & 0 & 0 & \cdots & 0 \\ \beta_2 & 0 & 1 & 0 & \cdots & 0 \\ \beta_3 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ \beta_n & 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where $\beta_j = \theta_0/\theta_j$. Let $\boldsymbol{X} = A\boldsymbol{Z}$ and $\boldsymbol{Z}$ be an $(n+1)$-variate random vector with mutually independent additive Tweedie margins $Z_i \sim \mathrm{T}(p, \theta_i, \Delta_i)$, $i = 0, 1, \ldots, n$. Then the jth univariate margin of $\boldsymbol{X}$ is

(i) *in the case of $p = 1$ and $\theta_j \equiv \theta$ for all $j \in \{0, 1, 2, \ldots, n\}$,*

$$X_j \sim \mathrm{T}(1, \theta, \Delta_0 + \Delta_j),$$

(ii) *in the other cases,*

$$X_j \sim \begin{cases} \mathrm{T}(2, \theta_j, \Delta_0 + \Delta_j), & p = 2, \\ \mathrm{T}(p, \theta_j, \Delta_0 \beta_j^{\alpha} + \Delta_j), & p \neq 1, 2, \end{cases}$$

where $\alpha = (p-2)/(p-1)$.

Remark 3.2. In Jørgensen and Song (1998), it is shown that a class of stationary infinite-order moving average processes with exponential dispersion model margins can be constructed by means of the thinning operation. That means that the coefficients of the process are random variables. For finite-order ARMA processes, the dependence structures can be considered as the same of multivariate Tweedie distributions for sufficiently large n in the class defined in Theorem 3.1. We consider the Box-Cox transformation of multivariate Tweedie distributions with the dependent care which is broader than the class of stationary processes with exponential dispersion model.

Denote the mean and the variance of the jth univariate margin X_j by μ_j and V_j, respectively. The fundamental properties of univariate margin are

$$\mu_j = \beta_j \Delta_0 \kappa'(\theta_0) + \Delta_j \kappa'(\theta_j), \tag{3.3}$$

$$V_j = \beta_j^2 \Delta_0 \kappa''(\theta_0) + \Delta_j \kappa''(\theta_j), \tag{3.4}$$

and $p(= m_{i_1} + m_{i_2} + \cdots + m_{i_n})$th joint cumulant is

$$\mathrm{cum}_{(m_{i_1}, m_{i_2}, \ldots, m_{i_n})}(X_{i_1}, X_{i_2}, \ldots, X_{i_n}) = \prod_{k=1}^{n} \beta_{i_k}^{m_{i_k}} \Delta_0 \kappa^{(p)}(\theta_0). \tag{3.5}$$

4. Power parameter for variance stabilization. In this section, we shall derive the optimal power parameter λ of the Box-Cox transformation for variance stabilization. To be concise, we consider the Box-Cox transformation applied to the random vector $\boldsymbol{X}$ distributed as multivariate Tweedie distribution. The Box-Cox transformation, i.e.,

$$X_j^{(\lambda)} = \begin{cases} \frac{X_j^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log X_j, & \text{if } \lambda = 0, \end{cases} \tag{4.1}$$

is applied to every marginal random variable X_j for $1 \leq j \leq n$. To deal with the transformation, we usually consider the case that the dispersion parameter Δ diverges. For the multivariate case, we suppose $\Delta_j / \Delta_0 = m_j$ and $\Delta = \min_{j \in \{1,2,\ldots,n\}} \Delta_j$. As $\Delta \to \infty$, it holds from (3.3)–(3.5) that

$$V^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \to_d \mathcal{N}(\boldsymbol{0}, \Sigma), \tag{4.2}$$

where $V^{1/2} = \mathrm{diag}(V_1^{1/2}, V_2^{1/2}, \cdots, V_n^{1/2})$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$. The asymptotic variance matrix Σ is given by

$$\Sigma_{i,j} = \begin{cases} 1, & \text{if } i = j, \\ \frac{\beta_i \beta_j \kappa''(\theta_0)}{\sqrt{\beta_i^2 \kappa''(\theta_0) + m_i \kappa''(\theta_i)} \sqrt{\beta_j^2 \kappa''(\theta_0) + m_j \kappa''(\theta_j)}}, & \text{if } i \neq j. \end{cases}$$

Note that the random vector $\boldsymbol{Z}$ to construct $\boldsymbol{X}$ is the maximum likelihood estimator for its mean under Δ-asymptotics. With the asymptotics above, we can derive the optimal power of the Box-Cox transformation for variance stabilization from the results in Section 2.

THEOREM 4.1. *The power parameter λ of the Box-Cox transformation for variance stabilization is given by*

$$\lambda = \frac{\alpha}{2(\alpha - 1)},$$

in both cases of (i) $\beta_j = 0$*, i.e., i.i.d. samples and* (ii) $\beta_j \neq 0$*, i.e., dependent observations.*

PROOF. (i) Note that

$$\Delta_j^{-1/2}(X_j - \mu_j) \to_d \mathcal{N}(0, \kappa''(\theta_j)).$$

It is easy to see that $g(\theta) = \kappa'(\theta_j)$, $K = \kappa^{(3)}(\theta_j)$ and $J = 0$. As a result, we have

$$\lambda = 1 - \frac{\kappa'(\theta_j)\kappa^{(3)}(\theta_j)}{(\kappa''(\theta_j))^2}.$$

Substitute (3.2) for $\kappa(\theta_j)$, we have

$$\lambda = \frac{\alpha}{2(\alpha - 1)}.$$

(ii) Suppose $\Delta_j/\Delta_0 = m_j$. From (4.2), we have

$$\Delta_0^{-1/2}(X_j - \mu_j) \to_d \mathcal{N}(0, V_j),$$

where

$$\mu_j = \beta_j\kappa'(\theta_0) + m_j\kappa'(\theta_j), \tag{4.3}$$

$$V_j = \beta_j^2\kappa''(\theta_0) + m_j\kappa''(\theta_j). \tag{4.4}$$

From the asymptotics of $\boldsymbol{Z}$, it holds that

$$\Delta_0^{-1/2}\begin{pmatrix} Z_0 \\ Z_j \end{pmatrix} \to_d \mathcal{N}\left(\begin{pmatrix} \kappa'(\theta_0) \\ \kappa'(\theta_j) \end{pmatrix}, \begin{pmatrix} \kappa''(\theta_0) & 0 \\ 0 & \kappa''(\theta_j) \end{pmatrix}\right).$$

Since $g(\theta_0, \theta_j) = \beta_j\kappa'(\theta_0) + m_j\kappa'(\theta_j)$, $K = \kappa^{(3)}(\theta_j)$ and $J = 0$, we find the power parameter λ have to satisfy

$$\lambda = \frac{\mu_j}{V_j}\Big\{-\frac{\beta_j\kappa^{(3)}(\theta_0)}{2\kappa''(\theta_0)}\Big\}, \tag{4.5}$$

$$\lambda = \frac{\mu_j}{V_j}\Big\{-\frac{m_j\kappa^{(3)}(\theta_j)}{2\kappa''(\theta_j)}\Big\}. \tag{4.6}$$

Note that m_j in (4.5) and (4.6) is a ratio parameter between two dispersion parameters Δ_0 and Δ_j. Two equations (4.5) and (4.6) can hold simultaneously if and only if

$$m_j = \frac{\beta_j\kappa''(\theta_j)\kappa^{(3)}(\theta_0)}{\kappa''(\theta_0)\kappa^{(3)}(\theta_j)}. \tag{4.7}$$

Under (4.7) and substitute (3.2) for $\kappa(\theta_0)$ and $\kappa(\theta_j)$ in (4.3), (4.4) and λ with $\beta_j = \theta_0/\theta_j$, we also have

$$\lambda = \frac{\alpha}{2(\alpha - 1)}.$$

□

Acknowledgements. The author would like to thank Professor Masanobu Taniguchi for his carefully reading and valuable comments on this note. The author also would like to thank Professor A. C. Monti for her explanation of data transformations during her stay at Waseda University. This work was supported by Grant-in-Aid for JSPS Fellows Grant Number 26·7404.

References.

Bartlett, M. S. (1947). The use of transformations. *Biometrics* **3** 39–52.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the american statistical association* **76** 296–311.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* **26** 211–252.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32.

Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance: Mathematics and Economics* **46** 351–361.

Hosoya, Y. and Terasaka, T. (2009). Inference on transformed stationary time series. *Journal of econometrics* **151** 129–139.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)* **49** 127–162.

Jørgensen, B. (2013). Construction of multivariate dispersion models. *Brazilian Journal of Probability and Statistics* **27** 285–309.

Jørgensen, B. and Song, P. X.-K. (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability* **26** 78–92.

Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient. *Biometrika* **65** 654–656.

Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis. *Biometrika* **68** 647–651.

Nishii, R. (1991). Exponential dispersion model and power trantransformations. *RIMS kôkyûroku* **749** 147–161.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series.* New York: Springer-Verlag.

Taniguchi, M. and Puri, M. L. (1995). Higher order asymptotic theory for normalizing transformations of maximum likelihood estimators. *Annals of the Institute of Statistical Mathematics* **47** 581–600.

Tweedie, M. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. In *Mathematical Proceedings of the Cambridge Philosophical Society* **43** 41–49. Cambridge Univ Press.

Yan Liu
Department of Applied Mathematics
School of Fundamental Science and Engineering
Waseda University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555
Japan
E-mail: great-rainbow@ruri.waseda.jp