Communications of the Korean Statistical Society
2011, Vol. 18, No. 1, 71–77

DOI: 10.5351/CKSS.2011.18.1.071

Asymptotics in Transformed ARMA Models

In-Kwon Yeo[1,a]

[a]Department of Statistics, Sookmyung Women's University

Abstract

In this paper, asymptotic results are investigated when a parametric transformation is applied to ARMA models. The conditions are determined to ensure the strong consistency and the asymptotic normality of maximum likelihood estimators and the correct coverage probability of the forecast interval obtained by the transformation and backtransformation approach.

Keywords: Coverage probability, equicontinuous, uniform convergence.

1. Introduction

In time series analysis based on the autoregressive moving average(ARMA) models, it is usually assumed that error terms are normally distributed. However, as shown by Fama (1965), Li (1999), McCulloch (1996), and McDonald (1996), many financial time series tend to be heavy-tailed or highly skewed, especially to the left. This implies that the possibility of observing outliers is higher than that under normal assumption. Statistical inferences under the normal assumption are generally based on the sample mean and sample variance. It is well known that these quantities are not robust against outliers. In the forecast of financial time series, lower or upper quantiles play an important role to measure the value at risk for the risk management. These quantiles also are mainly influenced by the shape of the underlying distribution. Therefore, alternative models are requested to overcome the violation of the normality.

Generally, two approaches are used to solve the non-normality problem in real analysis based on ARMA models. The first approach is that alternative distributions are assumed for error terms. For instance *t*-distribution or normal mixture distribution. This approach is often applied to data that are symmetric but have heavy tails on both sides. The generalized error distribution of Nelson (1991) or skewed Student *t*-distribution can be employed for modeling asymmetric data; however, it is not easy to perform statistical inferences under these distributions. The other method is the classical normal techniques applied to transformed data which are approximately normally distributed. This approach is relatively easy to apply, especially in skewed cases.

In this paper, we investigate asymptotic results when a parametric transformation is applied to the ARMA model. Conditions are specified that ensure the strong consistency and the asymptotic normality of maximum likelihood estimators(MLE) and the correct coverage probability of prediction interval by Cho *et al.* (2007). The asymptotics derived in this paper are the extension of results proved by Cho *et al.* (2001a) and Cho *et al.* (2001b) in the Box-Cox transformed linear models.

This research was supported by the Sookmyung Women's University Research Grants 2009.
[1] Associate Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
Email: inkwon@sm.ac.kr

2. Estimation in Transformed ARMA Models

Let $h(x, \lambda)$ be a general class of transformations which are indexed by the transformation parameter λ. A typical example of $h(x, \lambda)$ is Box and Cox (1964) transformation that takes the form as follows, for $x > 0$

$$h(x, \lambda) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(x), & \lambda = 0. \end{cases}$$

Other examples are John and Draper (1980), Burbidge *et al.* (1988), and Yeo and Johnson (2000). Since the transformation-and-backtransformation approach is frequently applied to predict a future value in the original scale, it is required that $h(x, \lambda)$ is monotone in x and so inverting the transformation is available. Throughout this paper, we assume that the second derivative of $h(x, \lambda)$ with respect to x is continuous in λ. This assumption plays a key role to derive asymptotic results based on the maximum likelihood inference.

Suppose that time series $\{X_t\}_{t=1}^n$ follow a stationary ARMA process with the mean μ. Then, the model equation is written as

$$\begin{aligned} \Phi(B)(X_t - \mu) &= \Theta(B)\varepsilon_t, \quad t = 1, 2, \ldots, n, \\ \Phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p, \\ \Theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q, \end{aligned}$$

where B stands for the back-shift operator and ε_t is a white noise. It is usually assumed that the white noise ε_t is normally distributed with a constant variance. However, it is unlike to see that a set of data on the original scale satisfies the normality assumption. A common characteristic of real data, especially financial time series is that their distribution is heavy-tailed or skewed. This implies that the possibility of observing outliers is higher than we expect. It is well-known that the maximum likelihood estimates are not robust against outliers.

When the normality assumption is seriously violated, one of the common approaches is that we take a transformation so that the transformed data are approximately normal and then apply the classical normal techniques to the transformed data. This approach is relatively easy to handle, especially in skewed cases. In this paper, we assume that the transformed time series $\{h(X_t, \lambda)\}_{t=1}^n$ follow a stationary Gaussian ARMA(p, q) process with the mean ν. Then,

$$\Phi(B)\{h(X_t, \lambda) - \nu\} = \Theta(B)\epsilon_t, \tag{2.1}$$

where error terms ϵ_t's are independent and distributed as $N(0, \sigma^2)$. Rewriting (2.1) as

$$\epsilon_t = \{1 - \Theta(B)\}\epsilon_t + \Phi(B)\{h(X_t, \lambda) - \nu\}, \tag{2.2}$$

the log-likelihood function of the parameters $\boldsymbol{\xi} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \nu, \sigma^2, \lambda)$ is

$$L(\boldsymbol{\xi}; \boldsymbol{x}) = -\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\epsilon_t^2 + \sum_{t=1}^{n}\log\{J(x_t, \lambda)\},$$

where $J(x_t, \lambda) = |\partial h(x_t, \lambda)/\partial x_t|$ denotes the Jacobian term. In this paper, we estimate the parameters via the conditional log-likelihood function which is written as

$$L_c(\boldsymbol{\xi}; \boldsymbol{x}) = -\frac{n-p}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{n}\epsilon_t^2 + \sum_{t=p+1}^{n}\log\{J(x_t, \lambda)\}. \tag{2.3}$$

Holding λ fixed, we initially obtain the estimates $(\hat{\boldsymbol{\phi}}(\lambda), \hat{\boldsymbol{\theta}}(\lambda), \hat{\nu}(\lambda))$ maximizing $L_c(\boldsymbol{\xi}; \boldsymbol{x})$ with respect to $(\boldsymbol{\phi}, \boldsymbol{\theta}, \nu)$ and then compute

$$\hat{\sigma}^2(\lambda) = \frac{1}{n-p} \sum_{t=p+1}^{n} \epsilon_t^2 \left(\hat{\boldsymbol{\phi}}(\lambda), \hat{\boldsymbol{\theta}}(\lambda), \hat{\nu}(\lambda)\right)$$

where $\epsilon_t(\hat{\boldsymbol{\phi}}(\lambda), \hat{\boldsymbol{\theta}}(\lambda), \hat{\nu}(\lambda))$ is defined as (2.2) of which the parameters are replaced by the estimates $(\hat{\boldsymbol{\phi}}(\lambda), \hat{\boldsymbol{\theta}}(\lambda), \hat{\nu}(\lambda))$. The MLE $\hat{\lambda}$ is obtained by maximizing the conditional profile likelihood function obtained by substituting $(\hat{\boldsymbol{\phi}}(\lambda), \hat{\boldsymbol{\theta}}(\lambda), \hat{\nu}(\lambda), \hat{\sigma}^2(\lambda))$ into (2.3). Consequently, the MLE of $\boldsymbol{\xi}$ is $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\phi}}(\hat{\lambda}), \hat{\boldsymbol{\theta}}(\hat{\lambda}), \hat{\nu}(\hat{\lambda}), \hat{\sigma}^2(\hat{\lambda}), \hat{\lambda})$.

Lemma 1. *Let $Y_1, Y_2, \ldots,$ be a sequence of stationary processes on R. Let Ξ be a compact topological space, and let f be a complex-valued function on $\Xi \times R$ and measurable in z for each $\xi \in \Xi$. Assume that*

(i) there is a function φ such that $|f(\xi, y)| < \varphi(y)$ for all $\xi \in \Xi$ and $E\{\varphi(Y)\} < \infty$,

(ii) there exists a sequence S_M of measurable sets such that $P[Y \in (R - \cup_{M=1}^{\infty} S_M)] = 0$,

(iii) for each M, $f(\xi, y)$ is equicontinuous in ξ for $y \in S_M$.

Then $n^{-1} \sum_{i=1}^{n} f(\xi, Y_i) \xrightarrow{a.s.} E[f(\xi, Y)]$ uniformly in $\xi \in \Xi$ and $E[f(\xi, Y)]$ is continuous.

Proof: The result is derived by applying the ergodic theorem to the uniform convergence of Rubin (1956). □

Lemma 2. *Let $\{G_n(\cdot)\}$ be a sequence of random functions defined on a probability space and depend on $\boldsymbol{\xi}$ in compact set $\boldsymbol{\Xi}$. Suppose that*

(i) there exist a continuous function $G(\boldsymbol{\xi})$ defined on $\boldsymbol{\Xi}$ such that $G_n(\boldsymbol{\xi}) \xrightarrow{a.s.} G(\boldsymbol{\xi})$ uniformly in $\boldsymbol{\xi} \in \boldsymbol{\Xi}$,

(ii) $G(\boldsymbol{\xi})$ has a unique minimum at $\boldsymbol{\xi}_0 \in \boldsymbol{\Xi}$.

Then, $\hat{\boldsymbol{\xi}}_n = \arg\min G_n(\boldsymbol{\xi})$ is a strongly consistent estimator of $\boldsymbol{\xi}_0$.

Since Lemma 2 is a standard result, we omit the proof.

Theorem 1. *Suppose the parameter space*

$$\boldsymbol{\Xi} = \{(\boldsymbol{\phi}, \boldsymbol{\theta}, \nu, \sigma^2, \lambda) \,|\, \boldsymbol{\phi} \in \Xi_\phi, \boldsymbol{\theta} \in \Xi_\theta, \nu \in \Xi_\nu, \sigma^2 \in \Xi_\sigma, \lambda \in \Xi_\lambda\}$$

and the log-likelihood function $L_c(\boldsymbol{\xi}; \boldsymbol{X})$ satisfy the following conditions;

(i) the parameter space $\boldsymbol{\Xi}$ is compact,

(ii) $E[h(X, \lambda)^2]$ and $E[\frac{\partial}{\partial X} h(X, \lambda)]$ are finite for all $\lambda \in \Xi_\lambda$,

(iii) $\Pi(\boldsymbol{\xi}) = E[L_c(\boldsymbol{\xi}; \boldsymbol{X})]$ has a unique global maximum at $\boldsymbol{\xi}_0 \in \boldsymbol{\Xi}$.

Then, for finite p,

(A) $\lim_{n\to\infty}\{\sup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} (n-p)^{-1} L_c(\boldsymbol{\xi}; \boldsymbol{X})\} = \sup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \Pi(\boldsymbol{\xi})$ with probability one,

(B) the MLE $\hat{\boldsymbol{\xi}} \in \Xi$ *is a strongly consistent estimator of* $\boldsymbol{\xi}_0$.

Proof: For simple notation, let

$$L_1(\boldsymbol{\xi}; x_t) = -\frac{1}{2}\log\left(\sigma^2\right) - \frac{\epsilon_t^2(\boldsymbol{\phi}(\lambda), \boldsymbol{\theta}(\lambda), \lambda)}{2\sigma^2} + \log\{J(x_t, \lambda)\}.$$

Boundary conditions of assumptions (i) and (ii) guarantee that $E\left[L_1(\boldsymbol{\xi}; X_t)\right]$ is finite. Since $L_1(\boldsymbol{\xi}; x)$ is continuous in $(\boldsymbol{\xi}, x)$ over the compact set $\Xi \times S_M$ where $S_M = [-M, M]$, $L_1(\boldsymbol{\xi}; x)$ is equicontinuous in $\boldsymbol{\xi}$ for $x \in S_M$, see Kosmala (1995), and, by Lemma 1, we conclude that $(n-p)^{-1}L_c(\boldsymbol{\xi}; \boldsymbol{X}) - \Pi(\boldsymbol{\xi}) \xrightarrow{a.s.} 0$ uniformly in $\boldsymbol{\xi} \in \Xi$. Equivalently,

$$\lim_{n\to\infty}\left\{\sup_{\boldsymbol{\xi}\in\Xi}\left|(n-p)^{-1}L_c(\boldsymbol{\xi}; \boldsymbol{X}) - \Pi(\boldsymbol{\xi})\right|\right\} = 0$$

with probability one. The result (A) follows directly since

$$\left|\sup_{\boldsymbol{\xi}\in\Xi}(n-p)^{-1}L_c(\boldsymbol{\xi}; \boldsymbol{X}) - \sup_{\boldsymbol{\xi}\in\Xi}\Pi(\boldsymbol{\xi})\right| \le \sup_{\boldsymbol{\xi}\in\Xi}\left|(n-p)^{-1}L_c(\boldsymbol{\xi}; \boldsymbol{X}) - \Pi(\boldsymbol{\xi})\right|.$$

From the continuity of $\Pi(\boldsymbol{\xi})$ by Lemma 1, the result (A), and the assumption (iii), the proof of the result (B) is straightforward. □

Theorem 2. *Suppose the conditions of Theorem 1 hold. Furthermore, if*

(iv) $\boldsymbol{\xi}_0$ *is an interior point of* Ξ.

(v) $E[(\frac{\partial}{\partial\lambda}h(X,\lambda))^2]$ *and* $E[\frac{\partial^2}{\partial\lambda\partial X}h(X,\lambda)]$ *are finite for all* $\lambda \in \Xi_\lambda$

(vi) $E[\nabla L_c(\boldsymbol{\xi}_0; \boldsymbol{X})] = \boldsymbol{0}$, *where the column vector* $\nabla L_c(\boldsymbol{\xi}_0; \boldsymbol{X})$ *is the gradient of the log-likelihood function evaluated at* $\boldsymbol{\xi}_0$,

(vii) $E[\nabla^2 L_c(\boldsymbol{\xi}_0; \boldsymbol{X})]$ *is non-singular, where* $\nabla^2 L_c(\boldsymbol{\xi}_0; \boldsymbol{X})$ *is the Hessian of the log-likelihood function evaluated at* $\boldsymbol{\xi}_0$. *Then*

$$(C)\quad \sqrt{n}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right) \xrightarrow{\mathcal{L}} N\left(\boldsymbol{0}, \mathbf{W}^{-1}(\boldsymbol{\xi}_0)\right) \tag{2.4}$$

where $\mathbf{W}(\boldsymbol{\xi}_0) = -E[\nabla^2 L_1(\boldsymbol{\xi}_0; X_t)]$.

Proof: Expanding $(n-p)^{-1/2}\nabla L_c(\hat{\boldsymbol{\xi}}; \boldsymbol{X})$ about $\boldsymbol{\xi}_0$, we obtain that

$$\frac{1}{\sqrt{n-p}}\nabla L_c\left(\hat{\boldsymbol{\xi}}; \boldsymbol{X}\right) = \frac{1}{\sqrt{n-p}}\nabla L_c\left(\boldsymbol{\xi}_0; \boldsymbol{X}\right) + \frac{1}{\sqrt{n(n-p)}}\nabla^2 L_c\left(\boldsymbol{\xi}_*; \boldsymbol{X}\right)\sqrt{n}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right), \tag{2.5}$$

where $\boldsymbol{\xi}_* = c_n\hat{\boldsymbol{\xi}} + (1-c_n)\boldsymbol{\xi}_0$, $c_n \in (0,1)$ for $n \ge 1$. Since $\nabla L_c(\hat{\boldsymbol{\xi}}; \boldsymbol{X}) = 0$ at the maximum, the left hand size of (2.5) converages to zero in probability. By the central limit theorem in Gaussian stationary time series, see Dunsmuir (1983), $(n-p)^{-1/2}\nabla L_c(\boldsymbol{\xi}_0; \boldsymbol{X})$ is asymptotically normal with mean 0 and variance $\mathbf{W}(\boldsymbol{\xi}_0)$. Further, another application of Lemma 1 and the consistency of $\hat{\boldsymbol{\xi}}$ ensure that

$$-\frac{1}{\sqrt{n(n-p)}}\nabla^2 L_c(\boldsymbol{\xi}_*; \boldsymbol{X}) \xrightarrow{p} -E\left[\nabla^2 L_1(\boldsymbol{\xi}_0; X)\right] = \mathbf{W}(\boldsymbol{\xi}_0).$$

Consequently, $\sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \mathbf{W}^{-1}(\boldsymbol{\xi}_0))$ □

3. Forecast Interval in Transformed ARMA Models

One of main goals in time series analysis is to forecast future values. The mean square error(MSE) is a common criterion to choose an optimum forecast. Let $\{Z_t\}_{t=1}^n$ follow a stationary ARMA(p, q) process with mean ν and let $Z_n(l)$ denote the l-step ahead forecast at time n. The MSE of the l-step ahead forecast at time n is $Et[(Z_{n+l} - Z_n(l))^2]$ and the minimum MSE forecast is represented by the conditional expectation $Z_n(l) = E[Z_{n+l} \,|\, Z_n, Z_{n-1}, \ldots]$. Under the stationary process assumption, ARMA(p, q) process is written as the following moving average representation,

$$Z_t = \nu + \Psi(B)\varepsilon_t = \nu + \sum_{j=0}^{\infty} \psi_j B^j \epsilon_t,$$

where $\Psi(B) = \Theta(B)/\Phi(B)$ and $\psi_0 = 1$. Then,

$$Z_n(l) = E\left(Z_{n+l} \,|\, Z_n, Z_{n-1}, \ldots\right) = \mu + \sum_{j=l}^{\infty} \psi_j \epsilon_{n+l-j}$$

and the forecast error is $e_n(l) = Z_{n+l} - Z_n(l) = \sum_{j=0}^{l-1} \psi_j \epsilon_{n+l-j}$. Suppose ϵ_t is Gaussian white noise with variance σ^2. Then, the conditional distribution of Z_{n+l} given $\{Z_t\}_{t=1}^n$ is $N(Z_n(l), \sigma^2 \sum_{j=0}^{l-1} \psi_j^2)$ and the $100(1-\alpha)\%$ forecast interval for Z_{n+l} is estimated as follows;

$$\hat{Z}_n(l) \pm z_{\frac{\alpha}{2}} \hat{\sigma} \sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2}, \tag{3.1}$$

where $\hat{Z}_n(l)$ is obtained by replacing parameters by estimators in $Z_n(l)$ and z_α denotes the $(1-\alpha)^{th}$ quantile of the standard normal distribution.

Cho *et al.* (2007) investigated the estimation of forecast intervals based on the transformation-and-backtransformation approach in the ARMA model and showed that the coverage probabilities of their forecast intervals are closer to the nominal level than those of the forecast interval without transformation through a simulation study. In this paper, we provide a theoretical validation of the forecast interval estimation the via the transformation and backtransformation approach.

Replacing components of (3.1) by $\hat{\xi}$ and the forecast on the transformed scale, we obtain the $100(1-\alpha)\%$ forecast interval on the transformed scale as follows;

$$\hat{h}\left(X_n, \hat{\lambda}\right)(l) \pm z_{\frac{\alpha}{2}} \hat{\sigma}\left(\hat{\lambda}\right) \sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2\left(\hat{\lambda}\right)}.$$

Let $L(\hat{\xi})$ and $U(\hat{\xi})$ be the lower and the upper bound of estimated forecast interval, respectively, on the transformed scale. Then, forecast interval $(L^*(\hat{\xi}), U^*(\hat{\xi}))$ on the original scale is given as $L^*(\hat{\xi}) = h^{-1}(L(\hat{\xi}), \hat{\lambda})$, $U^*(\hat{\xi}) = h^{-1}(U(\hat{\xi}), \hat{\lambda})$.

In order to find an asymptotic forecast interval for a future value, we consider the following quantity

$$T_n^{(l)}(\lambda) = \frac{h(X_{n+l}, \lambda) - \hat{h}(X_n, \lambda)(l)}{\hat{\sigma}(\lambda) \sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2(\lambda)}},$$

Lemma 3. *Suppose the conditions of Theorem 1 hold. Then, the limiting distribution of $T_n^{(l)}(\lambda)$ is the same as the distribution of*

$$T^{(l)}(\lambda) = \frac{\sum_{j=0}^{l-1} \psi_j(\lambda)\epsilon_{n+l-j}}{\sigma(\lambda)\sqrt{\sum_{j=0}^{l-1} \psi_j^2(\lambda)}}.$$

Proof: Note that $T_n^{(l)}(\lambda)$ is decomposed as follows;

$$T_n^{(l)}(\lambda) = \frac{h(X_{n+l}, \lambda) - h(X_n, \lambda)(l)}{\hat{\sigma}(\lambda)\sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2(\lambda)}} + \frac{h(X_n, \lambda)(l) - \hat{h}(X_n, \lambda)(l)}{\hat{\sigma}(\lambda)\sqrt{\sum_{j=0}^{l-1} \hat{\psi}_j^2(\lambda)}}. \tag{3.2}$$

Since $\hat{h}(X_n, \lambda)(l)$ is a function of $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\nu}, \lambda)$ and the uniform convergency of Theorem 1 for $\hat{\boldsymbol{\xi}}$ can be applied, the second term of right hand side of (3.2) converges to zero and the denominator of the second term converages to $\sigma(\lambda)\sqrt{\sum_{j=0}^{l-1} \psi_j^2(\lambda)}$. □

Note that, for Gaussian white noises with variance $\sigma^2(\lambda)$, the distribution of $T(\lambda)$ is the standard normal distribution.

Theorem 3. *Suppose the conditions in Theorem 1 and Theorem 2 hold. Then, $L^*(\hat{\boldsymbol{\xi}}) = h^{-1}(L(\hat{\boldsymbol{\xi}}), \hat{\lambda})$ and $U^*(\hat{\boldsymbol{\xi}}) = h^{-1}(U(\hat{\boldsymbol{\xi}}), \hat{\lambda})$ give an asymptotically correct forecast interval for the l-step ahead future value X_{n+l}.*

Proof: A Taylor expansion gives

$$T_n^{(l)}\left(\hat{\lambda}\right) = T_n^{(l)}(\lambda_0) + \left(\hat{\lambda} - \lambda_0\right) \frac{d}{d\lambda} T_n^{(l)}(\lambda)\bigg|_{\lambda=\lambda_*}, \tag{3.3}$$

where $|\lambda_* - \lambda_0| \le |\hat{\lambda} - \lambda_0|$. The condition (v) of Theorem 2 ensures that $\frac{d}{d\lambda} T_n^{(l)}(\lambda)\big|_{\lambda=\lambda_*}$ is bounded in probability. Since $\hat{\lambda}$ uniformly converges to λ_0 with probability one, the second term of right hand side of (3.3) converges to zero and $T_n^{(l)}(\hat{\lambda})$ has the same limiting distribution as $T^{(l)}(\lambda_0)$, that is $N(0, 1)$. This implies that

$$\lim_{n\to\infty} P\left[-z_{\frac{\alpha}{2}} \le T_n^{(l)}\left(\hat{\lambda}\right) \le z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

and

$$\lim_{n\to\infty} P\left[L\left(\hat{\boldsymbol{\xi}}\right) \le h\left(X_{n+1}, \hat{\lambda}\right) \le U\left(\hat{\boldsymbol{\xi}}\right)\right] = 1 - \alpha.$$

Since transformation $h(x, \lambda)$ is monotone in x, applying the inverse transformation, we obtain

$$\lim_{n\to\infty} P\left[h^{-1}\left(L\left(\hat{\boldsymbol{\xi}}\right), \hat{\lambda}\right) \le X_{n+l} \le h^{-1}\left(U\left(\hat{\boldsymbol{\xi}}\right), \hat{\lambda}\right)\right] = 1 - \alpha.$$

□

References

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

Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable, *Journal of the American Statistical Association.* **83**, 123–127.

Cho, H., Oh, S. and Yeo, I. K. (2007). Prediction interval estimation in transformed ARMA models, *The Korean Journal of Applied Statistics.* **20**, 541–550.

Cho, K. Yeo, I., Johnson, R. A. and Loh, W. Y. (2001a). Asymptotic theory for Box-Cox transformation in linear models, *Statistics and Probability Letters,* **51**, 337–343.

Cho, K. Yeo, I., Johnson, R. A. and Loh, W. Y. (2001b). Prediction interval estimation in transformed linear models , *Statistics and Probability Letters,* **51**, 345–350.

Dunsmuir, W. (1983). A central limit theorem for estimation in Gaussian stationary time series observed at unequally spaced times, *Stochastic Processes and Their Applications,* **14**, 279–295.

Fama, E. F. (1965). The behavior of stock market prices, *Journal of Business,* **38**, 34–105.

John, J. A. and Draper, N. R. (1980). An alternative family of transformations, *Applied Statistics,* **29**, 190–197.

Kosmala, W. A. J. (1995). *Introductory Mathematical Analysis,* William C Brown Publishers.

Li, D. (1999). *Value at Risk Based on the Volatility Skewness and Kurtosis*, RiskMetrics Working Paper.

McCulloch, J. H. (1996). Financial applications of stable distribution, *Handbook of Statistics,* **14**, 393–425.

McDonald, J. B. (1996). Probability distribution for financial models, *Handbook of Statistics,* **14**, 427–461.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrika,* **59**, 347–370.

Rubin, H. (1956). Uniform convergence of random functions with applications to statistics, *Annals of Mathematical Statistics,* **27**, 200–203.

Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika,* **87**, 954–959.

Received August 2010; Accepted November 2010