Journal of the Korean Data & Information Science Society
2015, **26**(2), 487–493

http://dx.doi.org/10.7465/jkdi.2015.26.2.487
한국데이터정보과학회지

Robust Bayesian analysis for autoregressive models

Hyunnam Ryu[1] · Dal Ho Kim[2]

[12]Department of Statistics, Kyungpook National University

Received 27 December 2014, revised 10 January 2015, accepted 27 January 2015

Abstract

Time series data sometimes show violation of normal assumptions. For cases where the assumption of normality is untenable, more flexible models can be adopted to accommodate heavy tails. The exponential power distribution (EPD) is considered as possible candidate for errors of time series model that may show violation of normal assumption. Besides, the use of flexible models for errors like EPD might be able to conduct the robust analysis. In this paper, we especially consider EPD as the flexible distribution for errors of autoregressive models. Also, we represent this distribution as scale mixture of uniform and this form enables efficient Bayesian estimation via Markov chain Monte Carlo (MCMC) methods.

Keywords: Autoregressive model, exponential power distribution, Gibbs sampler, robustness.

1. Introduction

Real data often show violation of normal assumptions. Heavy-tailed distributions are frequently encountered in empirical studies. For cases where the assumption of normality is untenable, more flexible models can be adopted to accommodate heavy tails. The exponential power distribution (EPD) is considered as possible candidate for errors of time series model that may show violation of normal assumption. Besides, the use of flexible models for errors like EPD might be able to conduct the robust analysis. An exponential power distribution had been studied by Box and Tiao (1992) in the context of robustness studies. The exponential power density is given by (1.1)

$$f(x|\mu,\sigma,p) = \frac{1}{2p^{\frac{1}{p}}\Gamma\left(1+\frac{1}{p}\right)\sigma}\exp\left(-\frac{1}{p}\left|\frac{x-\mu}{\sigma}\right|^{p}\right), \tag{1.1}$$

where $\mu \in \boldsymbol{R}, \quad \sigma > 0, \quad p \geq 1$. The EPD has three parameters: μ is a location parameter, σ is a scale parameter and p determines the kurtosis, which is given by $\kappa = \Gamma(1/p)\Gamma(5/p)/(\Gamma(3/p))^2$. If parameter p is fixed, the distribution is a location-scale family of distributions. Hence the parameter p is related to the thickness of the tails. Specifically, the EPD is heavy-tailed distributions, if $1 < p < 2$ and light-tailed distributions, if $p > 2$. Also, the special cases of the EPD are the double exponential distribution ($p = 1$), the normal distribution ($p = 2$) and the uniform distribution ($p \to \infty$) as seen in Figure 1.1.

1 Ph.D student, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
2 Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr

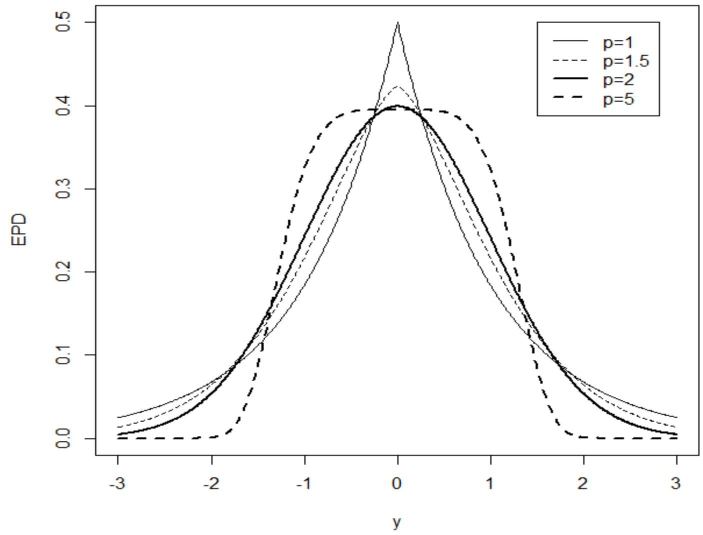

Figure 1.1 Various cases of the EPD

Recently, many studies that consider the skewness as well as the kurtosis in view of robustness have been actively discussed (Zhu and Zinde-Walsh, 2009; DiCiccio and Monti, 2004; Salazar *et al.*, 2012).

Generally, most of distributions can be expressed by the scale mixtures. Also, we can represent EPD as the scale mixture of uniform distributions (Walker and Gutiérrez-Peña, 1999). This form enables efficient Bayesian estimation via Markov chain Monte Carlo (MCMC) methods (Fernandez and Steel, 1998).

$$X|U=u \sim \text{Uniform}\left(\mu - 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u^{\frac{1}{p}}, \mu + 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u^{\frac{1}{p}}\right)$$
$$U \sim \text{Gamma}\left(1 + \frac{1}{p}, 2^{-\frac{p}{2}}\right) \tag{1.2}$$

where Gamma(α, β) denotes a gamma distribution with parameters α and β.

In this paper, we especially consider EPD as the flexible distribution for errors of autoregressive (AR) models. Also, we compare the exponential power errors with Student's t errors, that is one of well-known heavy-tailed distributions, for the autoregressive models.

2. Bayesian analysis for autoregressive model

2.1. Autoregressive model with exponential power errors

Consider the autoregressive model with order 1 in which the t th observation y_t satisfies

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, ..., n, \tag{2.1}$$

where e_t is the error such that $e_1, ..., e_n$ are independent and identically distributed according to the EPD with location parameter zero, scale parameter σ and shape parameter p. Also

α is an intercept coefficient and ρ is an autoregressive coefficient. Equivalently, this model can be expressed by the scale mixture of uniform distributions as (2.2).

$$\begin{aligned} y_t|u_t &\sim \text{Uniform}\left(\alpha + \rho y_{t-1} - 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u_t^{\frac{1}{p}}, \alpha + \rho y_{t-1} + 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u_t^{\frac{1}{p}}\right) \\ u_t &\sim \text{Gamma}\left(1 + \frac{1}{p}, 2^{-\frac{p}{2}}\right), \quad t = 1, \ldots, n. \end{aligned} \tag{2.2}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$ be n observations and $\boldsymbol{u} = (u_1, \ldots, u_n)'$ be the mixing parameters. Thus the likelihood function is given by

$$L(\alpha, \rho, \sigma|\boldsymbol{u}, \boldsymbol{y}) \propto \sigma^{-n} \exp\left(-2^{-\frac{p}{2}} \sum_{t=1}^{n} u_t\right) \prod_{t=1}^{n} I(u_t > \delta_t), \tag{2.3}$$

where $\delta_t = 2^{\frac{p}{2}} \frac{1}{p} \frac{|y_t - \alpha - \rho y_{t-1}|^p}{\sigma^p}$.

Since there is no sufficient prior information, we assign diffused uniform priors for the intercept coefficient and the autoregressive coefficient, and a diffused gamma prior on the inverse of the variance components. And we consider the uniform prior for autoregressive coefficient from -1 to 1, to acquire the stationarity. Prior distributions are assumed to be mutually independent. We have the following priors : $\alpha \sim \text{Uniform}(-100, 100)$, $\rho \sim \text{Uniform}(-1, 1)$ and $(\sigma^2)^{-1} \sim \text{Gamma}(a, b)$. Here $X \sim \text{Gamma}(a, b)$ denotes a gamma distribution with shape parameter a and rate parameter b having the expression $f(x) \propto x^{a-1} \exp(-bx)$, $x \geq 0$. The full posterior of the parameters given the data is as follows :

$$\pi(\alpha, \rho, \sigma|\boldsymbol{u}, \boldsymbol{y}) \propto (\sigma^2)^{-(a + \frac{n}{2} + 1)} \exp\left(-\frac{b}{\sigma^2}\right) \exp\left(-2^{-\frac{p}{2}} \sum_{t=1}^{n} u_t\right) \prod_{t=1}^{n} I(u_t > \delta_t). \tag{2.4}$$

Thus, the conditional density of α, ρ, σ, $\boldsymbol{u}$ are given by

$$[\alpha|\rho, \sigma^2, \boldsymbol{u}, \boldsymbol{y}] \sim \text{Uniform}\Big(\max(A_t - B_t), \min(A_t + B_t)\Big). \tag{2.5}$$

$$A_t = y_t - \rho y_{t-1}, \quad B_t = 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u_t^{\frac{1}{p}},$$

$$[\rho|\alpha, \sigma^2, \boldsymbol{u}, \boldsymbol{y}] \sim \text{Uniform}\left(\max\left(\frac{1}{y_{t-1}}(C_t - D_t)\right), \min\left(\frac{1}{y_{t-1}}(C_t + D_t)\right)\right), \tag{2.6}$$

$$C_t = y_t - \alpha, \quad D_t = \text{sign}(y_{t-1}) 2^{-\frac{1}{2}} p^{\frac{1}{p}} \sigma u_t^{\frac{1}{p}},$$

$$[\sigma^2|\alpha, \rho, \boldsymbol{u}, \boldsymbol{y}] \sim \text{IG}\left(a + \frac{n}{2}, b\right) \cdot I\left(\sigma^2 > \max\left(2(u_t p)^{-\frac{2}{p}} |y_t - \alpha - \rho y_{t-1}|^2\right)\right), \tag{2.7}$$

$$[u_t|\boldsymbol{u}_{-t}, \alpha, \rho, \sigma^2, \boldsymbol{y}] \sim \text{Exponential}\left(2^{-\frac{p}{2}}\right) \cdot I\left(u_t > 2^{\frac{p}{2}} \frac{1}{p} \frac{|y_t - \alpha - \rho y_{t-1}|^p}{(\sigma^2)^{\frac{p}{2}}}\right), \ t = 1, \ldots, n. \tag{2.8}$$

where $\text{IG}(\alpha, \beta)$ denotes a inverse gamma distribution with parameters α and β and Exponential(λ) denotes a exponential distribution with parameter λ.

Each conditional density has standard distribution. So we can infer the parameters using Gibbs sampler to sample from the full conditional relevant parameters.

2.2. Autoregressive model with Student's t errors

We also think about the autoregressive models with Student's t errors that is one of the heavy-tailed models to compare to results of EPD. Similarly, consider the autoregressive model with order 1 in which t th observation y_t satisfies

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, ..., n, \tag{2.9}$$

where e_t is the error such that $e_1, ..., e_n$ are independent and identically distributed according to the Student's t distribution with location parameter zero and scale parameter σ. And α is intercept coefficient and ρ is autoregressive coefficient. As we know, this model can be expressed by the scale mixture of normals. So, we can express this model as follows.

$$\begin{aligned} y_t|u_t &\sim N\left(\alpha + \rho y_{t-1}, \frac{\sigma^2}{u_t}\right) \\ u_t &\sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \quad t = 1, \ldots, n. \end{aligned} \tag{2.10}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_n)'$ be n observations and $\boldsymbol{u} = (u_1, \ldots, u_n)'$ be the mixing parameters. Thus the likelihood function is given by

$$L(\alpha, \rho, \sigma|\boldsymbol{u}, \boldsymbol{y}) \propto (\sigma^2)^{-\frac{n}{2}} \prod_{t=1}^{n} u_t^{\frac{\nu}{2}-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t(y_t - \alpha - \rho y_{t-1})^2 - \frac{\nu}{2}\sum_{t=1}^{n} u_t\right). \tag{2.11}$$

Since there is no sufficient prior information, we assign diffused uniform priors for the intercept coefficient and the autoregressive coefficient, and a diffused gamma prior on the inverse of the variance components on the same priors as the EPD case. And we consider the uniform prior for autoregressive coefficient from -1 to 1, to acquire the stationarity. The prior distributions are assumed to be mutually independent. We have the following priors : $\alpha \sim \text{uniform}(-100, 100)$, $\rho \sim \text{uniform}(-1, 1)$ and $(\sigma^2)^{-1} \sim \text{Gamma}(a, b)$. Here $X \sim \text{Gamma}(a, b)$ denotes a gamma distribution with shape parameter a and rate parameter b having the expression $f(x) \propto x^{a-1}\exp(-bx)$, $x \geq 0$. The full posterior of α, ρ, σ given the data $\pi(\alpha, \rho, \sigma|\boldsymbol{u}, \boldsymbol{y})$ is as follows.

$$\begin{aligned} \pi(\alpha, \rho, \sigma|\boldsymbol{u}, \boldsymbol{y}) \propto & (\sigma^2)^{-(a+\frac{n}{2}+1)} \exp\left(-\frac{b}{\sigma^2}\right) \prod_{t=1}^{n} u_t^{\frac{\nu}{2}-\frac{1}{2}} \\ & \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t(y_t - \alpha - \rho y_{t-1})^2 - \frac{\nu}{2}\sum_{t=1}^{n} u_t\right). \end{aligned} \tag{2.12}$$

Then, the conditional density of α, ρ, σ, $\boldsymbol{u}$ are given by

$$[\alpha|\rho, \sigma^2, \boldsymbol{u}, \boldsymbol{y}] \sim N\left(\frac{\sum_{t=1}^{n}(y_t - \rho y_{t-1})u_t}{\sum_{t=1}^{n} u_t}, \frac{\sigma^2}{\sum_{t=1}^{n} u_t}\right), \tag{2.13}$$

$$[\rho|\alpha, \sigma^2, \boldsymbol{u}, \boldsymbol{y}] \sim N\left(\frac{\sum_{t=1}^{n}(y_t - \alpha)y_{t-1}u_t}{\sum_{t=1}^{n} u_t y_{t-1}^2}, \frac{\sigma^2}{\sum_{t=1}^{n} u_t y_{t-1}^2}\right), \tag{2.14}$$

$$[\sigma^2|\alpha,\rho,\boldsymbol{u},\boldsymbol{y}] \sim IG\left(a+\frac{n}{2}, \frac{1}{2}\sum_{t=1}^{n}(y_t-\alpha-\rho y_{t-1})^2 u_t + b\right), \tag{2.15}$$

$$[u_t|\boldsymbol{u}_{-t},\alpha,\rho,\sigma^2,\boldsymbol{y}] \sim \text{Gamma}\left(\frac{\nu}{2}+\frac{1}{2}, \frac{1}{2\sigma^2}(y_t-\alpha-\rho y_{t-1})^2+\frac{\nu}{2}\right). \tag{2.16}$$

Thus each conditional densities has standard distribution. So we can infer the parameters using Gibbs sampler to sample from the full conditional relevant parameters and compare to result of the exponential power case.

3. Numerical studies

Consider the AR(1) model

$$y_t = \alpha + \rho y_{t-1} + e_t, \tag{3.1}$$

where innovations are independent and identically distributed, each with a contaminated normal density $0.9 \times N(0,1) + 0.1 \times N(\delta,1)$. The properties of the estimator of ρ when the EPD is used to approximate the distribution of e_t's was studied by simulation in two cases: a mild contamination having $\delta = 1$ and a more serious contamination having $\delta = 8$. The simulation study was performed with series of length $n = 300$, $\alpha = 0$ and $\rho = 0.7$ (Figure 3.1).

We would like to compare Bayes estimates of the AR(1) model with different distributions of errors such as the exponential power errors having different parameter $p = (1.0, 1.5, 2.0)$ and the Student's t errors, as we already mentioned in the previous section. So, we iterated 2,000 times (thin=60) for the each models and discarded 1,000 burn-in samples.

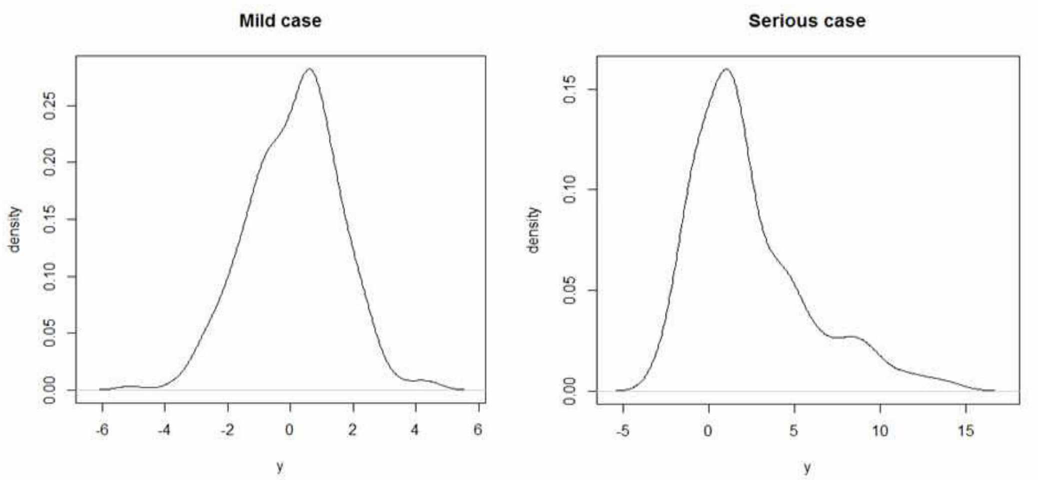

Figure 3.1 Density of the simulated data

In this example, the interest is focused on knowing if the parameter values are recovered, i.e., if the posterior estimations are close to the original parameters. The posterior means, standard deviations and the 95% highest posterior density (HPD) intervals are reported in Table 3.1-3.2. The posterior mean of the AR(1) with exponential power errors ($p = 1.0$) is 0.6596 which is the closest to the true value among the four models in the mild case. And we can also check that the AR(1) with exponential power errors ($p = 1.0$) has the shortest HPD interval (0.5780, 0.7449), and this interval contain the true value $\rho = 0.7$ as well. In the serious case (i.e., heavier case than the mild), the posterior mean of the AR(1) with exponential power errors ($p = 1.0$) is 0.6974, which is not only more accurate than the mild case but also the most similar value to the true value. And 95% HPD intervals of the serious case also have the shortest interval (0.6504, 0.7450) when the exponential power errors have the shape parameter p is 1.0. In addition, we detect that when AR(1) models have the exponential power errors, the shape parameter p is smaller (i.e., tails are heavier), they have more accurate result from heavy-tailed case.

Table 3.1 Bayesian estimates and HPD credible intervals under AR(1): $\delta = 1$

		α				ρ				σ			
		Mean	SD	95% HPD		Mean	SD	95% HPD		Mean	SD	95% HPD	
	$p = 1.0$	0.0188	0.0660	-0.1153	0.1483	0.6596	0.0417	0.5780	0.7449	0.8778	0.0550	0.7724	0.9880
EP	$p = 1.5$	0.0329	0.0628	-0.0901	0.1574	0.6519	0.0425	0.5686	0.7362	0.9973	0.0467	0.9070	1.0920
	$p = 2.0$	0.0415	0.0632	-0.0866	0.1651	0.6517	0.0440	0.5652	0.7395	1.1090	0.0449	1.0221	1.1958
Student t		0.0413	0.0643	-0.0874	0.1671	0.6501	0.0447	0.5606	0.7371	1.1016	0.0444	1.0148	1.1915

Table 3.2 Bayesian estimates and HPD credible intervals under AR(1): $\delta = 8$

		α				ρ				σ			
		Mean	SD	95% HPD		Mean	SD	95% HPD		Mean	SD	95% HPD	
	$p = 1.0$	0.1074	0.0927	-0.0736	0.2944	0.6974	0.0248	0.6504	0.7450	1.4393	0.0851	1.2737	1.6092
EP	$p = 1.5$	0.3253	0.1330	0.0659	0.5901	0.6955	0.0315	0.6322	0.7561	1.9543	0.0930	1.7748	2.1398
	$p = 2.0$	0.7265	0.1719	0.3853	1.0723	0.6924	0.0423	0.6069	0.7758	2.4702	0.1014	2.2785	2.6756
Student t		0.7065	0.1715	0.3723	1.0486	0.6916	0.0411	0.6088	0.7716	2.4450	0.1034	2.2399	2.6481

4. Concluding remarks

In this paper, the EPD has been analyzed from a Bayesian viewpoint. The exponential power family includes the double exponential distribution and the normal distribution as particular cases and provides distributions with either lighter or heavier tails compared to the normal one. Using such a EPD for the error terms in a autoregressive model, we detect that the EPD is able to capture the kurtosis features better than other competing models by adjusting the parameter p when some observational data have heavy-tailed distributions. So, in these situation, the exponential power case can release the normal assumptions. Besides, using the proposed scale mixture representations of distributions, Bayesian inference is provided via an ′easy to implement′ Gibbs sampler. And this involves constructing a Markov chain which has as its stationary distribution the relevant posterior distribution of interest. Thus, this scale mixture have the advantage of simplifying the full conditional distributions and making Bayesian computations more efficient.

References

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian inference in statistical analysis*, Wiley, New York.

DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, **99**, 439-450.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359-571.

Salazar, E., Ferreira, M. R. A. and Migon, H. S. (2012). Objective Bayesian analysis for exponential power regression models. *Aankhyā B*, **74**, 107-125.

Walker, S. G. and Gutiérrez-Peña, E. (1999). Robustifying Bayesian procedures. In *Bayesian Statistics 6*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford University Press, New York, 685-710.

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, **148**, 86-99.