Ren and Ren, International Journal of Applied Economics, 16(1), March 2019, 1-6

Mean or Median for Log-normally Distributed Time Series in Economics and Finance

Louie Ren

School of Business Administration, University of Houston-Victoria, Victoria, TX, USA

Peter Ren

College of Business, University of Houston-Downtown, Houston, TX, USA

Abstract

We examine the forecast precision on measuring un-biasness and forecast accuracy on measuring variations from overall mean and overall median for i.i.d. time series following a skewed lognormal probability distribution. In this manuscript, we introduce a new weighted forecast measure. It is a linearly weighted combination of forecast precision and forecast accuracy. Under the new forecast measure, the mean supersedes the median when the skewness is less or equal to 2. When the skewness is greater than 2, the median outperforms the mean by a very narrowed range of weights ω for the weighted forecast measure. We recommend the use of the mean as the predictor, not the median even though the time series following an extremely skewed lognormal distribution in economics and finance. The proposed weight forecast measure Lω can also apply to analyze any skewed time series in economics and finance.

Keywords: Mean, Median, Lognormal distribution, Weighted linear combination of forecast accuracy and forecast precision, Sensitivity analysis

JEL Classification: C52, C53, E47

---

1. Introduction

In forecasting, "precision" measures the un-biasness of the forecast estimator and is defined as the expectation of the difference of actual observations from a forecast estimator. "Accuracy" measures the dispersion of the forecast estimator and is defined as the variance (or the standard deviation) of the difference (Baldwin and Kain, 2006, Chen, Twycross, and Garibaldi, 2017, Flores and Wichern, 2005, Mazais, Miķelsons, and Salenieks, 2007, and Morlidge, 2016).

Two popular forecast estimators for time series without trend, seasonality, and irregularity, or in the case of independently and identically distributed (i.i.d.) time series are the overall average, $\overline{X_n}$, and the overall median, Md(n), based on past n-period historical observations. The overall average, $\overline{X_n}$, is an unbiased forecast estimator for any time series from a skewed or symmetric probability distribution. The overall median, Md(n), is also an unbiased forecast estimator when the time series is from a symmetric probability distribution but a biased forecast estimator when the time series is from a skewed probability distribution. However, the overall median generates a smaller variance than that from the overall average, making the overall median a more accurate forecast estimator in comparison to the overall mean.

In this manuscript, we introduce a new weighted forecast measure, L_ω. It is a linearly weighted combination of forecast precision and forecast accuracy, that is, $L_\omega = (1-\omega)(\mu_e) + \omega(\sigma_e)$, where $0 \le \omega \le 1$, μ_e, and σ_e are the mean and the s.d. of error term, respectively. The weight, ω, serves to reflect the forecaster's emphasis on precision relative to accuracy. When $\omega = 0$, $L_{\omega=0}$ is the measure of the forecast precision from μ_e. When $\omega = 1$, $L_{\omega=1}$ is the measure of the forecast accuracy from σ_e. When $\omega=0.2$, the forecaster is more concerned (80%) about the forecast precision from μ_e rather than the forecast precision from σ_e. When $\omega=0.8$, the forecaster is more concerned (80%) about the forecast accuracy from σ_e rather than the forecast precision from μ_e. In this study, we will compare forecast precisions and accuracies from the overall mean, $\overline{X_n}$, and overall median, Md(n), on i.i.d. time series following lognormal distributions since lognormal distributions have a vast applications in economics and finance (Antoniou, Ivanov, Ivanov, and Zrelov, 2004).

2. Data Analysis

A positive random variable X has a lognormal distribution with parameters (μ, σ) when its logarithm has a normally distribution with parameters (μ, σ). From Johnson, Kotz, and Balakrishnan (1997), we know that X has a probability density function

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \ x > 0.$$

$E(X) = e^{(\mu+\frac{\sigma^2}{2})}$, $Median(X) = e^{\mu}$, $Var(X) = (e^{\sigma^2}-1)e^{(2\mu+\sigma^2)}$, and

$$\text{the Skewness} = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \quad (1)$$

When using the overall average, $\overline{X_n}$, as the forecast, the mean and the s.d. of errors (e's) when a time series X_t follows a lognormal distribution can be found as:

$$\mu_{e,\bar{x}} = E(X_{n+1} - \overline{X_n}) = E(X_{n+1}) - E(\overline{X_n}) = \mu - \mu = 0 \quad (2)$$

$$\sigma_{e,\bar{x}} = \sqrt{Var(X_{n+1}) + Var(\overline{X_n})} = \sqrt{\sigma_x^2 + \frac{\sigma_x^2}{n}} = \sqrt{(1+\frac{1}{n})\sigma_x^2} = \sqrt{(1+\frac{1}{n})(e^{\sigma^2} - 1)e^{(2\mu+\sigma^2)}} \quad (3)$$

When using the median, Md(n), as the forecast, the mean and the s.d. of errors (e's) when time series X_t's follow a lognormal distribution can be found as:

$$\mu_{e,Md} = E(X_{n+1} - Md(n)) = E(X_{n+1}) - E(Md(n)) = e^{\mu+\frac{\sigma^2}{2}} - e^{\mu} = e^{\mu}(e^{\frac{\sigma^2}{2}} - 1) \quad (4)$$

$$\sigma_{e,Md} = \sqrt{Var(X_{n+1}) + Var(Md(n))} = \sqrt{\sigma_x^2 + \sigma_{Md(n)}^2} = \sqrt{(e^{\sigma^2} - 1)e^{(2\mu+\sigma^2)} + \frac{1}{4nf(m)^2}} \quad (5)$$

, where $m = e^{\mu}$, and $f(m) = \frac{1}{m\sigma\sqrt{2\pi}} e^{-\frac{-(\ln m-\mu)^2}{2\sigma^2}}$.

For non-standard normal distributions, we can standardize it to a standard normal distribution with a mean $\mu = 0$. Therefore, in this study, we generate probability density functions (pdf) using the following parameters: n = 3, 5, 10, 30, 50, (μ=0, σ=0.15), (μ=0, σ=0.30), (μ=0, σ=0.45), (μ=0, σ=0.55), (μ=0, σ=0.70), (μ=0, σ=1.0), (μ =0, σ=1.5), and (μ=0, σ =2). The skewness for each of them is about 0.5, 1.0, 1.5, 2.0, 3.0, 6.0, 30, and 400, respectively. We plot the pdf of the lognormal distribution for each case in Figure 1, and present the skewness, $\mu_{e,\bar{x}}$, $\sigma_{e,\bar{x}}$, $\mu_{e,Md}$, and $\sigma_{e,Md}$, as shown in Equations (1)-(5) for each case in Table 1.

As an illustration, consider the series with parameters n = 50, μ = 0, and σ = 2, with the most skewness of about 400 highlighted in bold. From Equations (1)-(5), its skewness = $(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$ =

Ren and Ren, International Journal of Applied Economics, 16(1), March 2019, 1-6

$(e^{2^2}+2)\sqrt{e^{2^2}-1}$ =414 ≈ 400, $\mu_{e,\bar{x}}$ = 0, $\sigma_{e,\bar{x}} = \sqrt{(1+\frac{1}{50})(e^{2^2}-1)e^{(2(0)+2^2)}}$ = 54.63, $\mu_{e,Md}$ = $e^0(e^{\frac{2^2}{2}}-1)$ = 6.39, and $\sigma_{e,Md} = \sqrt{(e^{2^2}-1)e^{(2(0)+2^2)}+\frac{1}{4(50)f(m)^2}}$ =54.10 where m = e^0=1, and f(m) = $\frac{1}{(1)(2)\sqrt{2\pi}}e^{-\frac{-(\ln e-0)^2}{2(2^2)}}$ = 0.20. We can generate the following sensitivity graphs for ω, where ω satisfies the equation:

(0)(1-ω)+(54.63)ω = (6.39)(1- ω)+(54.10)ω, i.e.,
6.93ω = 6.39, and ω = 0.92.

We find that the overall average, $\overline{X_n}$, outperforms the overall median, Md(n), for almost all weights ω, 0 ≤ ω ≤0.92. The overall median, Md(n), will only outperform the overall average, $\overline{X_n}$, when ω ≈ 1 (ω = 0.92 for the above case), i.e. when considering only forecast accuracy and not precision. From Table 1, we can see that when skewness is less than or equal to 2.0, the overall average, $\overline{X_n}$, dominates in performance against the overall median, Md(n). For extremely skewed time series with skewness of 6.0, 30, and 400, the overall average, $\overline{X_n}$, dominates in performance against the overall median, Md(n)for almost all weights ω, 0 ≤ ω ≤1. That is, the overall average, $\overline{X_n}$, not the overall median, Md(n), should be used as a predictor when time series follows a lognormal distribution. The proposed weighted forecast measure Lω can also apply to analyze any skewed time series in economics and finance.

3. Conclusion

Based on sensitivity analysis, we should always use the overall average, $\overline{X_n}$, as our forecaster for a lognormal distributed time series in economics and finance when forecast accuracy and forecast precision are considered. We should only use the overall median, Md(n), when emphasizing on forecast accuracies for dispersions over forecast precisions for unbiasness.

References

Antoniou, I., Vi. Ivanov, Va., Ivanov, and P. Zrelov. 2004. "On the log-normal distribution of stock market data," *Physica A*, 331, 617-638.

Baldwin, M. E. and J. S. Kain. 2006. "Sensitivity of several performance measures to displacement error, bias, and event frequency," *Weather and Forecasting*, 21(4), 636-648.

Chen, C., J. Twycross, and J. M. Garibaldi. 2017. "A new accuracy measure based on bounded relative error for time series forecasting," *Plos ONE*, 12(3), 1-23.

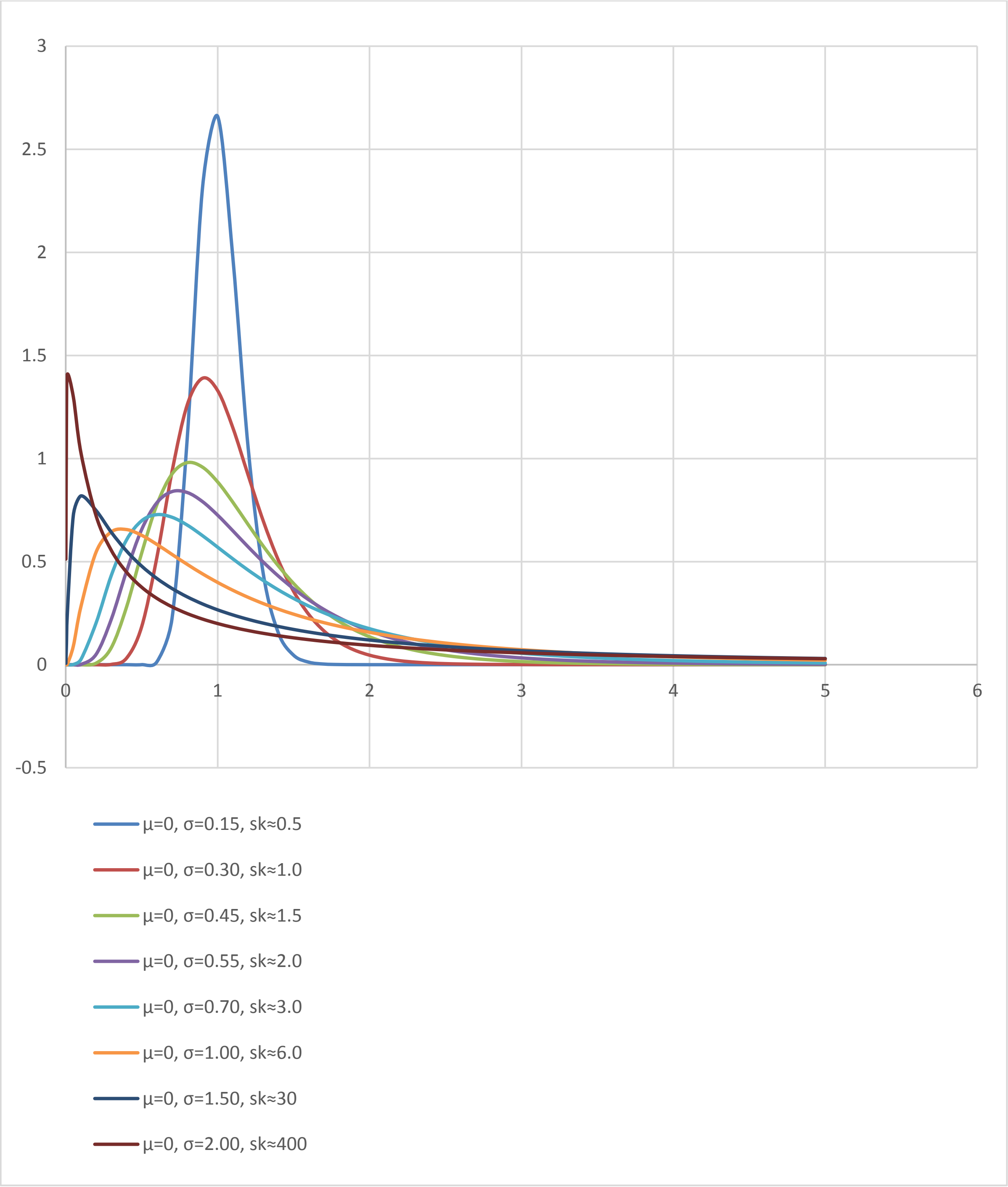
Flores, B. E. and E. W. Wichern. 2005. "Evaluating forecasts: a look at aggregate bias and accuracy measures," *Journal of Forecasting*, 24(6), 433-451.

Johnson, N., S. Kotz, and N. Balakrishnan. 1997. *Continuous Univariate Distributions*, Ch. 14, Vol. 1, 2nd ed., New York, N.Y.: John Wiley and Sons.

Mazais, J., J. Miķelsons, and N. Salenieks. 2007. "Process parameter and measurement accuracy and precision," *Transport and Engineering*, 2311-2315.

Morlidge, S. 2016. "Using error analysis to improve forecast performance," *Foresight: The International Journal of Applied Forecasting*, 41, 37-44.

Figure 1. Plots of Probability Density Functions (pdf) for Various Lognormal Distributions



Ren and Ren, International Journal of Applied Economics, 16(1), March 2019, 1-6

Table 1. Skewness, $\mu_{e,\bar{x}}$, $\sigma_{e,\bar{x}}$, $\mu_{e,Md}$, and $\sigma_{e,Md}$ for Various Length of Historical Data, (μ, σ)

Length n	(μ, σ)	Skewness*	$\mu_{e,\bar{x}}$	$\sigma_{e,\bar{x}}$	$\mu_{e,Md}$	$\sigma_{e,Md}$
3	(0, 0.15)	0.5	0	0.18	0.01	0.19
	(0, 0.30)	1.0	0	0.37	0.05	0.39
	(0, 0.45)	1.5	0	0.61	0.11	0.62
	(0, 0.55)	*2.0*	*0*	*0.80*	*0.16*	*0.80*
	(0, 0.70)	3.0	0	1.17	0.28	1.14
	(0, 1.0)	6.0	0	2.50	0.65	2.28
	(0, 1.5)	30	0	10.36	2.08	9.04
	(0, 2.0)	400	0	62.46	6.39	54.12
5	(0, 0.15)	0.5	0	0.17	0.01	0.17
	(0, 0.30)	1.0	0	0.35	0.05	0.36
	(0, 0.45)	1.5	0	0.57	0.11	0.58
	(0, 0.55)	*2.0*	*0*	*0.76*	*0.16*	*0.76*
	(0, 0.70)	3.0	0	1.11	0.28	1.09
	(0, 1.0)	6.0	0	2.37	0.65	2.23
	(0, 1.5)	33	0	9.83	2.08	9.01
	(0, 2.0)	400	0	59.26	6.39	54.11
10	(0, 0.15)	0.5	0	0.16	0.01	0.16
	(0, 0.30)	1.0	0	0.34	0.05	0.34
	(0, 0.45)	1.5	0	0.55	0.11	0.55
	(0, 0.55)	*2.0*	*0*	*0.73*	*0.16*	*0.72*
	(0, 0.70)	3.0	0	1.07	0.28	1.05
	(0, 1.0)	6.0	0	2.27	0.65	2.20
	(0, 1.5)	33	0	9.41	2.08	8.99
	(0, 2.0)	400	0	56.74	6.39	54.10
30	(0, 0.15)	0.5	0	0.16	0.01	0.16
	(0, 0.30)	1.0	0	0.33	0.05	0.33
	(0, 0.45)	1.5	0	0.53	0.11	0.53
	(0, 0.55)	*2.0*	*0*	*0.70*	*0.16*	*0.70*
	(0, 0.70)	3.0	0	1.03	0.28	1.03
	(0, 1.0)	6.0	0	2.20	0.65	2.17
	(0, 1.5)	33	0	9.12	2.08	8.98
	(0, 2.0)	400	0	54.99	6.39	54.10
50	(0, 0.15)	0.5	0	0.15	0.01	0.15
	(0, 0.30)	1.0	0	0.32	0.05	0.33
	(0, 0.45)	1.5	0	0.53	0.11	0.53
	(0, 0.55)	*2.0*	*0*	*0.70*	*0.16*	*0.70*
	(0, 0.70)	3.0	0	1.03	0.28	1.02
	(0, 1.0)	6.0	0	2.18	0.65	2.17
	(0, 1.5)	33	0	9.06	2.08	8.98
	(0, 2.0)	**400**	**0**	**54.63**	**6.39**	**54.10**

* approximation

Ren and Ren, International Journal of Applied Economics, 16(1), March 2019, 1-6

Figure 2. Sensitivity Analysis