Robustness of Bayes Forecast to Non-normality

Ashok K. Bansal*

ABSTRACT

Bayesian procedures are in vogue to revise the parameter estimates of the forecasting model in the light of actual time series data. In this paper, we study the Bayes forecast for demand and the risk when (a) 'noise' and (b) mean demand rate in a constant process model have moderately non-normal probability distributions.

1. Introduction

Suppose we believe the time series can be adequately described by the constant model, $x_t=\theta+\varepsilon_t$, when θ is the unknown mean demand and ε_t is the random component(noise) assumed to be normally distributed with known mean θ and precision γ. In Bayesian formulation, we further assume that at the starting of the forecasting process (time zero), the true mean θ is a priori normally distributed with known mean μ and precision τ. 'Extensive form of analysis' with quadratic error loss gives the Bayes estimator, $\theta^*(\gamma)$, for θ which has minimum risk ρ^* associated with the actual sample information. For a constant process, the Bayes forecast for period $T+\delta$ is then $\hat{x}_{(T+\delta)}=\theta^*(T)$ with forecast variance ρ^*.

Earlier workers were mainly concerned with the sensitivity of statistical procedure to the inherent non-normality of the parent distribution. Finucan (1971) studied the effect of non-normality on the posterior variance but his

*Department of Mathematical Statistics, University of Delhi, India. The author is grateful to the Computer Centre, University of Delhi for providing the necessary facilities on IBM 360/44 system to undertake this investigation.

results did not convey anything regarding the effect of individual observations. Such studies can be categorised as Standard Inference Robustness (SIR) in contrast to 'Inference Robustness with respect to Prior' (IRP) when an investigator is interested in examining the sensitivity of the Bayes statistical procedure to non-normality of the assumed prior distribution.

In this paper, we shall compare SIR with IRP of the Bayes forecast and associated forecast Bayes risk (forecast variance) for the actual time series data under 'moderate' amount of non-normality. Unlike Box-Tiao (1973) symmetric non-normal family, we consider a family of non-normal populations represented by the first four terms of the Edgeworth series.

$$f(y)=\phi(y)-\frac{1}{6}\lambda_3\phi^{(3)}(y)+\frac{1}{24}\lambda_4\phi^{(4)}(y)+\frac{1}{72}\lambda_3^2\phi^{(6)}(y), \tag{1}$$

where $y=(z-\alpha)\sqrt{\beta}$, $\phi(y)=(2\pi e^{y^2})^{-\frac{1}{2}}$, $\phi^{(k)}(y)=(d/dy)^k\phi(y)$ with mean α, precision β and non-normality parameter $\lambda=(\lambda_3,\lambda_4)$ measuring skewness $\lambda_3^2(=\beta_1)$ and kurtosis $\lambda_4(=\beta_2-3)$.

When λ is within Barton-Dennis (1952) region (BDR) the theoretical specification of the population by Edgeworth distribution covers a variety of moderately non-normal populations. An advantage with the Edgeworth series approach is that it obtains results as a sum of the normal theory function and certain corrective terms due to skewness and kurtosis. It has also enabled a fairly accurate estimation of the extent of the error involved in use of normal theory procedure for moderately non-normal variates.

2. Bayes Forecast and Forecast Variance

Non-normal Demand (SIR). Let $x=(x_1,x_2\cdots,x_T)$ be the T observations of the time series following constant process. Suppose that this random sample comes from Edgeworth distribution (1) with unknown mean θ, known precision γ and fixed non-normality parameter λ lying in BDR. Let us further suppose that the true mean θ is a random variable having a-priori known normal p.d.f. with mean μ and precision τ. Bansal (1977 b) derived the posterior distribution of θ as

$$\xi_s(\theta|\boldsymbol{x})=(2\pi\tau')^{-\frac{1}{2}}G_s^{-1}[1+\frac{1}{6}\lambda_3(S_3-3S_1)+\frac{1}{24}\lambda_4(S_4-6S_2+3T)+\frac{1}{72}\lambda_3^2(S_3^2-9S_4-6S_3S_1+36S_2+9S_1^2-15T)]\exp[-\frac{1}{2\tau'}(\theta-\hat{x}_N(T))]^2, \quad (2)$$

the Bayes forecast

$$\hat{x}_S(T)=\hat{x}_N(T)-\frac{1}{12}\tau'\gamma G_S^{-1}[6\gamma^{-\frac{1}{2}}\lambda_3\{1+\gamma(\tau'+A_2)\}+2\lambda_4\{\gamma A_3+3A_1(1+\gamma\tau')\}+\lambda_3^2[\gamma^2\{A_2A_3+(TA_3+9A_1A_2)\tau'+15TA_1\tau'^2\}-\gamma\{3A_1A_2+(T+6)A_3+6(2T+3)\tau'A_1\}+3(T+4)A_1]], \quad (3)$$

and the associated forecast variance (Bayes risk)

$$\rho_s^*=\tau'G_s^{-1}[1+\frac{1}{6}\lambda_3\sqrt{\gamma}\{\gamma A_3+3A_1(3\gamma\tau'-1)\}+\frac{1}{24}\lambda_4\{\gamma^2A_4+18\tau'(\gamma A_2-1)+15\gamma^2\tau'^2-6\gamma A_2+3T\}+\frac{1}{72}\lambda_3^2[\gamma^3\{A_3^2+9\tau'(2A_1A_3+3A_2^2)+45\tau'^2(3A_1^2+2TA_2)+210\tau'^2\}-\gamma^2\{9A_4+6A_3A_1+54\tau'(A_1^2+(T+3)A_2)+315\tau'^2\}+9\gamma(4A_2+A_1^2+18\tau')-15T]]-[\hat{x}_S(T)-\hat{x}_N(T)] \quad (4)$$

where

$$G_s=1+\frac{1}{6}\sqrt{\gamma}\lambda_3\{\gamma A_3-3A_1(1-\gamma\tau')\}+\frac{1}{24}\lambda_4\{\gamma^2A_4-6\gamma A_2(1-\gamma\tau')+(1-\gamma\tau')^2+3(T-1)\}+\frac{1}{72}\lambda_3^2[\gamma^3\{A_3^2+3\tau'(2A_1A_3+3A_2^2)+9\tau'^2(2TA_2+3A_1^2)+30\tau'^3\}-3\gamma^2\tau'\{2A_1A_3+6A_1^2+6(T+3)A_2+21\tau'^2\}+9\gamma(3A_2+A_1^2)+9\gamma\tau'-15T], \quad (5)$$

$A_k=\sum_{i=1}^{T}(x_i-\hat{x}_N(T))^k$, $S_k=\sum_{i=1}^{T}\{\sqrt{\gamma}(x_i-\theta)\}^k$ for $k=1, 2, \cdots, 6$. $\tau'=(\tau+T\gamma)^{-1}$ and $\hat{x}_N(T)=(\tau\mu+T\gamma\bar{x})\tau'$ are normal theory forecast variance and forecast for period $(T+1)$, respectively.

Non- normal Noise (IRP). Here we assume that the observed time series, $(x_1, x_2,\cdots,x_T)=\boldsymbol{x}$, generated by constant process form a random sample from $N(0,\gamma^{-1})$ with known precision γ. In the absence of 'true' prior p.d.f., earlier workers assumed convenient $N(\mu, \tau^{-1})$ conjugate prior distribution with known parameters μ, τ. Bansal (1977 a) relaxed this conventional assumption and assumed prior p.d.f. as Edgeworth with known μ, τ and fixed non-normality parameter λ lying in BDR to study the Robustness of Bayes estimator for θ against non-normality with increasing sample information.

The posterior p.d.f. of θ with respect to prior (1) is

$$\xi_p(\theta|\boldsymbol{x})=(2\pi\tau')^{-\frac{1}{2}}G_p^{-1}\exp\left[-\frac{1}{2\tau'}(\theta-\hat{x}_N(T))^2\right]\left[1+\frac{1}{6}\lambda_3\sqrt{\tau}\{\tau(\theta-\mu)^2-3\}\right.$$
$$\times(\theta-\mu)+\frac{1}{24}\lambda_4\{\tau^2(\theta-\mu)^4-6\tau(\theta-\mu)^2+3\}+\frac{1}{72}\lambda_3^2\{\tau^3(\theta-\mu)^6-15\tau^2$$
$$\left.\times(\theta-\mu)^4+45\tau(\theta-\mu)^2-15\}\right], \quad (6)$$

the Bayes forecast

$$\hat{x}_p(T)=\hat{x}_N(T)+\left[\frac{1}{2}\lambda_3\tau'\sqrt{\tau}\{p^2\tau(x-\mu)^2+\tau\tau'-p(\bar{x}-\mu)\}+\frac{1}{6}\lambda_4p^2\tau\tau'\right.$$
$$\{(\bar{x}-\mu)^3\tau\tau'-3(\bar{x}-\mu)\}+\frac{1}{72}\lambda_3^2p^3\tau\tau'\{\tau^2p^2(\bar{x}-\mu)^5-10\tau\tau'(\bar{x}-\mu)^3$$
$$\left.+15(\bar{x}-\mu)\}\right]G_p^{-1}, \quad (7)$$

and the associated risk is given by

$$\rho_p^*=\tau'G_p^{-1}\left[1+\frac{1}{6}\lambda_3\sqrt{\tau}\{\tau p^3(\bar{x}-\mu)^2+3p(2\tau\tau'-p)\}(\bar{x}-\mu)\right.$$
$$+\frac{1}{24}\lambda_4\{\tau^2p^4(\bar{x}-\mu)^4+6\tau p^2(2\tau\tau'-p)(\bar{x}-\mu)^2-3p(4\tau\tau'-p)\}$$
$$+\frac{1}{72}\lambda_3^2\{\tau^3p^6(\bar{x}-\mu)^6+15\tau^2p^4(2\tau\tau'-p)(\bar{x}-\mu)^4-45\tau p^3$$
$$\left.(4\tau\tau'-p)(\bar{x}-\mu)^2+15p^2(6\tau\tau'-p)\}\right]-[\hat{x}_p(T)-\hat{x}_N(T)]^2, \quad (8)$$

with

$$p=\gamma T\tau' \text{ and } G_p=1+\frac{1}{6}\lambda_3p^2\sqrt{\tau}\{p\tau(\bar{x}-\mu)^2-3\}(\bar{x}-\mu)$$
$$+\frac{1}{24}\lambda_4p^2\{\tau^2p^2(\bar{x}-\mu)^4-6\tau p(\bar{x}-\mu)^2+3\}+\frac{1}{72}\lambda_3^2p^3\{\tau^3p^3(\bar{x}-\mu)^6$$
$$-15\tau^2p^2(\bar{x}-\mu)^4+45\tau p(\bar{x}-\mu)^2-15\}. \quad (9)$$

3. Discussion

It is easily seen that for $\boldsymbol{\lambda}=0$, expressions (1)-(4) and (6)-(8) collapse to corresponding normal theory results. Further, we observe that the sample observations $(x_1,x_2,\cdots,x_T)$ appear as sample mean $\bar{x}$ in the IRP. On the other hand, in SIR x_i's appear in a more meaningful (though complicated) form. This implies that the effect of non-normality in the parent will vary with individual sample observations. Earlier studies by the author, however, indicated that the posterior p.d.f. for θ is less sensitive to non-normality in the prior p.d.f. of θ

than the non-normality in the p.d.f. of 'noise' ε_t.

In order to bring out difference between IRP and SIR, we considered a random sample $\boldsymbol{x}=(86, 94, 97, 95, 106, 107, 103, 92, 98, 104)$ of size 10 from a constant process with error variance 150 and assumed that, the unknown θ had true mean 100 and variance 25. The Bayes forecast and the associated forecast risk for $T=1, 2, 5, 10$ were computed for the normal and three other non-normal distributions with λ lying in BDR.

Table 1 : Forecast with Edgeworth as Noise and Demand (underlined) p.d.f. with Non-normality λ for Observed Constant Process.

T	Actual Demand	Sample mean	Forecast for $(T+1)^{th}$ period			
			Normal	Edgeworth		
			(0.0, 0.0)	(0.3, 0.5)	(0.4, 1.5)	(0.5, 2.4)
1	86.0	86.0	98.00	98.10 <u>99.97</u>	99.92 <u>100.69</u>	104.22 <u>101.40</u>
2	94.0	90.0	97.50	97.68 <u>99.52</u>	99.07 <u>100.33</u>	100.90 <u>101.12</u>
5	106.0	95.6	98.00	97.94 <u>99.18</u>	98.43 <u>99.72</u>	98.78 100.23
10	104.0	98.2	98.88	98.64 <u>99.33</u>	98.82 <u>99.59</u>	98.91 <u>99.79</u>

We observe that the departure from normality assumption, in each case, yields larger forecasts and the error in forecast steadily decreases with increasing sample information.

Further, the Bayes forecast is less sensitive(though higher) to non-normality in the prior p.d.f. of θ than that in 'noise' ε_t.

As reported by Finucan (1971), we observe that the forecast variance is reduced for increasing non-normality in the demand whereas it is always higher than normal theory value. Once again, it suggests that actual observations in the sample play an important role in deciding the effect of non-normality of 'noise' ε_t on the forecast variance. Further, the effect of non-normality is seen to diminish with increasing sample size.

Table 2: Forecast Variance for Noise and Demand (underlined) Edgeworth wiht Non-normality λ for Observed Constant Process.

T	Var(X)	Forecast Variance for $(T+1)^{th}$ period			
		Normal	Edgeworth		
	$1/\gamma T$	(0.0, 0.0)	(0.3, 0.5)	(0.4, 1.5)	(0.5, 2.4)
1	150.0	21.43	25.66	30.42	19.28
			<u>15.40</u>	<u>11.44</u>	<u>6.47</u>
2	75.0	18.73	21.63	23.60	21.79
			<u>12.49</u>	<u>7.83</u>	<u>1.98</u>
5	30.0	13.64	14.58	15.71	16.45
			<u>10.90</u>	<u>8.06</u>	<u>4.95</u>
10	15.0	9.38	9.55	10.15	10.42
			<u>8.56</u>	<u>7.40</u>	<u>6.35</u>

REFERENCES

[1] Bansal, A.K., (1977a) "Robustness of Bayes Estimator for the Mean of Normal Population with Non-normal Prior," *Communication in Statistics*, Series A (To appear).

[2] ____________ (1977b) "On Bayesian and Testing of Mean in Non-normal Samples," unpublished.

[3] Barton, D.E. and Dennis, K.E., (1952) "The Conditions under which Gram-Charlier and Edgeworth Curves Are Positive, Definite and Unimodal," *Biometrika* 39, 425.

[4] Box, G.E.P. and Tiao, G.C., (1973) *Bayesian Inference in Statistical Analysis*, Addison-Wesley.

[5] Finucan, H.M., (1971) "Posterior Precision for Non-normal Distributions", *Jour. Roy. Statist. Soc.* B 33, 95–97.