Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
ISSN NO. 2320-5407
Article DOI:10.21474/IJAR01/7317
DOI URL: http://dx.doi.org/10.21474/IJAR01/7317

RESEARCH ARTICLE

MODEL SELECTION AND COMPARISON OF TIME SERIES MODELS.

P. Mariyappan[1] and P.Arumugam[2].
1. Research scholar Manonmaniyam Sunadaranar university,Thirnelveli,Department of statistics, Mahabararhi Engineering College, Vasudevanur.
2. Asso. Professor in statistics, Annamalai university, Annamalai Nager, Chidhambaram.

Manuscript Info

Manuscript History

Received: 18 April 2018
Final Accepted: 20 May 2018
Published: June 2018

Abstract

In this paper, we study the comparison of Autoregressive moving average (ARMA) and Autoregressive Integrated moving average (ARIMA) model and also study posterior analysis by using the Bayesian approach. A numerical study has been carried out to illustrate the developed model.

Introduction:-

The Bayesian inference holds a distinct advantage over So-called classical statistics in non-standard problems where concepts such as sufficiency or completeness do not apply. That advantages is that the program is unchanged, the prior together with the likelihood produce the posterior. A disadvantage is that conjugate families are not available and so the Bayes theorem must be used numerically, for which approximation and numerical integration techniques are required.

The Autoregressive moving average (ARMA) time series models are quiet non-standard, even if the usual assumption normality is retained. The number of parameters, reflected in the order of the model (p,q), is undetermined. Given (p,q), the parameter space must then be constrained for identifiability reasons. Classical analysis of ARMA models must rely on asymptotic behavior: Consistency, asymptotic Normality, and efficiency. But in time series analysis, important problems with as few as 50 observations.

The choice of a particular econometric model is not prespecified by theory and many competing models can be entrained. Comparing models can be done formally in a bayesian frame work through so-called posterior odds, which is the product of the prior oods and the Bayes factor between any two models in the ratio of the likelihoods integrated out with the corresponding prior and summarizes how the data favor one model over other. Given a set of possible models, this immediately leads to Posterior model probabilities. Rather than choosing a single model, a natural way to deal with model uncertainty is to use the posterior model probabilities to average out the inference (parameters) corresponding to each of the separable models. This is called bayesian model averaging.

The modern time series analysis, considerable attention is paid to the question of the number of times an individual series must be differenced to achieve stationarity. For the ARIMA (p ,d , q) class of models , introduced by Box & Jenkins (1970), the question then is of the number d of unit autoregressive roots in the generating process. Most often in practical applications the choice is between $d = 0$ and $= 1$, though of course choice between $d = 1$ and $d = 2$ can be made along identical lines by working with the first differences of the original time series.

Corresponding Author:- P. Mariyappan
Address:- Research scholar Manonmaniyam Sunadaranar university,Thirnelveli,Department of statistics, Mahabararhi Engineering College, Vasudevanur.

The most commonly applied approach to this issue is through formal hypothesis testing, with the null hypothesis taken as $d = 1$ and the alternative as stationarity in levels$(d = 0)$. In practical applications, the parametric test of Dickey and Fuller (1979) has been employed far more often than any other though the nonparametric approach of Phillips and Perron (1988) is also used on occasions as discussed by for example, Banerjee (1993), the decision on degree of differencing for individual series is an important first step in a methodology for constructing models linking those series.

The prototypical problem is differentiating between a random walk and a stationary first order autoregressive model. Then, for a time series Y_t, the Dickey- Fuller test is based on the least square fit

$$Y_t = \delta + \phi Y_{t-1} + e_t \quad (1)$$

The test statistic is the usual t-ratio associated with $(\hat{\phi} - 1)$, though the null distribution is nonstandard Dickey & Fuller (1979) More generally, equation (1) is a generated adding term in laggeorst differences to account for autocorrelation beyond first order auto-regression. The decision on the level of differencing is then based on the outcome of a test at a significance level that the user must specify. It is not at all clear that this approach accurately reflects the analyst's priorities about d. Indeed, tests where stationarity is the null hypothesis have also been developed (see, for example, Kwiaatkowski (1992), Leybourne and McCabe (1994)). When these tests are applied to the same series, it is common to find that neither null hypothesis- stationarity a unit autoregressive root- can be rejected at the usual significance levels.

It seems natural to explore the Bayesian approach to the comparison of stationary models with those involving a unit autoregressive root, and there has been interest in this possibility in the econometric literature, dating from Sims (1998) and Sims and Uhlig (1991). For the this title we address fundamental issues arising from the practical application of that approach . Our concern is not with asymptotic properties of decision rules, an issue addressed in Phillips and Ploberger (1996), and references therein. Indeed, we view as strength of the bayesian approach that recourse to asymptotics for justification is unnecessary. Instead, we consider inference from the perspective of an analyst with a single time series, requiring posterior odds for a unit root model compared with a stationary competitor.

The practical issues involved can be dealt with through the analysis of the simple case (1), a critical issues is the specification of a prior for the parameter ϕ under stationarity. Possibilities that are sometimes adopted are the uniform prior on (-1,1) and the Jeffreys prior (Berger and Yang, 1994, Uhlig, 1994). It seems to us unreasonable that an analyst could simultaneously hold such vague prior belief while at the same time attaching non-zero probability mass at $\phi = 1$, as would be implied by testing hypotheses with this null. Presumably the same considerations that lead to the suspicion that "true generating model" is a random walk also suggest the likelihood of large positive values for φ under stationarity. Intuitively, a prior density function that approaches infinity as ϕ approaches one seems more plausible. Accordingly, we explore the use of the beta distribution as a prior specification. It emerges from our analysis that it is incumbent on the analyst to give Careful consideration to what constitutes an appropriate prior for ϕ, as this can have substantial impact on the posterior odds for sample sizes commonly analyzed in practice.

A further issue to be faced is the specification of a relatively uninformative prior for the mean of the series, or equivalently for the parameter δ of equation (1), under stationarity. We show how this issue can be circumvented by carrying out the analysis in first differences of the given series. In deriving posterior odds, we work with the exact likelihood, assuming a Gaussian process. A by-product of our analysis is a demonstration that leads to superior decision rules compared with the usual practice of basing tests on the least squares estimation of (1).

Finally, in section 5 of the chapter we briefly demonstrate how our approach can be extended to more general models. Specifically, we consider the comparison of an ARIMA $(p, 1, q)$ model with a stationary ARMA$(p + 1, q)$ competitor.

Bayesian Analysis:-

We are concerned here with comparing the random walk model

$$\mathrm{Y_t} = \mathrm{Y_{t-1}} + e_\mathrm{t} \quad \ldots (2)$$

With the alternative of a stationary first order autoregressive model

$$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + e_t \qquad \ldots (3)$$

Where e_t are iid $N(0, \sigma^2)$ and $|\phi| < 1$, for $t = 0,1,2, \ldots . n$. Bayesian approaches to this and , related problems have considered by several authors, including Phillips(1991), Poirier(1991), Schotman and Van Dijk(1991), Koop(1992), Uhlig(1994), Schotman(1994), Zivot(1994), and Lubrano(1995), and the references therein.

Much of the published discussion of this problem has concerned the choice of price for ϕ and μ. Schotman and Van Dijk (1991) point out that the lack of identification of μ in the random walk model means that a uniform prior cannot be used for this parameter. Several different priors are considered by the cited authors, many of whom model an interaction between ϕ and μ. Schotman (1994), in a useful chapter that reviews much of the work in this area, concludes that prior assumptions about the dependence between ϕ and μ have a marked effect on the posterior location of ϕ. In many practical applications, however, the analyst may be reluctant to specifyan informative prior on μ. Indeed, it is difficult to contemplate a situation where the analyst simultaneously feels able to specify a sharp prior for μ while entertaining non-zero prior probability for the random walk model where that parameters is undefined. Our concern here is with the case where the analyst's inclination is to use an improper perior for μ. This causes no difficulty in deriving a posterior for ϕ under the stationary autoregressive model. However, when improper priors are used for parameters occurring in one model and not the other posterior odds ratios are undefined. (See O'Hagan, 1995, for a general discussion and Schotman & Van Dijk(1991), for analysis in the context of the present problem). In our approach we remove this as a problem by formulating both of the above models in terms of the first differences $W_t = Y_t - Y_{t-1.}$ so that the random walk model and first order autoregressive model we consider here are

$$M_1: \; W_t = e_t$$

$$M_2: \; W_t - \phi W_{t-1} = e_t - e_{t-1}$$

Given a sample $W = (W_1, W_2, \ldots \ldots W_n)$, the bayesian comparision of the two models proceeds by computing the posterior model probabilities, which are given by Bayes theorem as

$$P(M_i/W) = \frac{P(M_i)P(W/M_i)}{\sum_{j=1}^{2} P(M_j)P(j)} \qquad \ldots (4)$$

In (4), $P(M_i)$ is the probability assigned to model $M_{i.}$

$P(W/M_i) = \int_{-1}^{1}\int_{0}^{\infty} P(\phi, \sigma/M_i)P(W/\phi, \sigma, M_i)d\sigma d\phi$ is the integrated joint density of (ϕ, σ, W),

Where $P(\phi, \sigma/M_i)$ is the joint prior density for the parameters, and $P(W/\phi, \sigma, M_i)$ is the likelihood.

Posterior Analysis:-

The likelihood for M_i can be written in terms of the difference W as

$$P(W/\sigma, M_1) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\, exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n} {W_t}^2\right\}$$

For model M_2 the likelihood can be shown to be

$$P(W/\sigma, M_2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} A^{\frac{1}{2}}\, exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n+2} {u_t}^2\right\}$$

Where

$$A = [! + n(1-\phi)(1+\phi)^{-1}]^{-1},$$

$$u_1 = -AC,$$

$$u_2 = \phi(1-\phi^2)^{-\frac{1}{2}} AC,$$

$$u_3 = W_1 - (1+\phi)^{-1}AC, \quad \text{and}$$

$$u_t = W_{t-2} + (1-\phi)\sum_{j=1}^{t-3} W_j - (1+\phi)^{-1}AC; t = 4, \ldots \ldots, n+2$$

Where

$$C = (1-\phi)\sum_{t=1}^{n} W_t + (1-\phi)^2 \sum_{t=1}^{n} (n-t)W_t.$$

This is a special case of the result in Newbold(1974).

Adopting the usual no informative prior for σ, we write $P(\phi,\sigma) = \sigma^{-1}P(\phi)$. The joint density of (ϕ,σ,W) for M_2 can then be written as

$$P(\phi,\sigma,W/M_2) = \frac{1}{\sigma^{n+1}(2\pi)^{\frac{n}{2}}} A^{\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n+2} {u_t}^2\right\} P(\phi).$$

Integrating this with respect to σ gives

$$P(\phi,W/M_2) = (2\pi)^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right) 2^{(n-2)/2} A^{\frac{1}{2}}\left[\sum_{t=1}^{n+2} {u_t}^2\right]^{-\frac{n}{2}} P(\phi)$$

and then

$$P(W/M_2) = \int_{-1}^{1} P(\phi, W/M_2)\, d\phi. \qquad .. (5)$$

It is straightforward to show that the integrated joint density in the case of M_1 is given by

$$P(W/M_1) = (2\pi)^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right) 2^{(n-2)/2} \left[\sum_{t=1}^{n} {W_t}^2\right]^{-\frac{n}{2}}$$

Equation (4) can now be used to obtain the posterior model probabilities as

$$P(M_1/W) = \frac{1}{1+K}$$

and

$$P(M_2/W) = \frac{K}{1+K}$$

$$K = \frac{\int_{-1}^{1} A^{\frac{1}{2}}[\sum_{t=1}^{n+2} {u_t}^2]^{-\frac{n}{2}} P(\phi) d\phi}{\left[\sum_{t=1}^{n} {W_t}^2\right]^{-\frac{n}{2}}}$$

We note here that if the conjugate prior

$$P(\sigma) \propto \frac{1}{\sigma^{\nu_{0+1}}} exp\left\{-\frac{\nu_0 {c_0}^2}{2\sigma^2}\right\}, \qquad \nu_0 > 0$$

Is adopted for σ, we obtain the same expression for $P(M_i/W)$ but now with

$$K = \frac{\int_{-1}^{1} A^{\frac{1}{2}}[\nu_0 {c_0}^2 + \sum_{t=1}^{n+2} {u_t}^2]^{-\frac{n+\nu_0}{2}} P(\phi) d\phi}{\left[\nu_0 {c_0}^2 + \sum_{t=1}^{n} {W_t}^2\right]^{-\frac{n+\nu_0}{2}}}$$

It should be noted that the analysis of first differences involves no information loss about the parameter of the autoregressive model compared with an analysis of levels with an improper prior on μ. To see this, suppose that inference is based on $Y = (Y_0, Y_1, Y_2 \ldots\ldots Y_n)$ and consider the transformation, with Jacobian one, to (Y_0, W). Then, we have

$$\begin{aligned} P(\phi,\mu,Y) &\propto P(\phi)P(\sigma)P(Y/\mu,\phi,\sigma) \\ &= P(\phi)P(\sigma)P(Y_0, W/\phi, \mu, \sigma) \\ &= P(\phi)P(\sigma)P(Y_0/W,\phi,\mu,\sigma)P(W/\phi, \sigma) \end{aligned}$$

Now, μ only appears in $P(Y_0/W, \phi, \mu, \sigma)$, which necessarily takes the form

$$P(Y_0/W, \phi, \mu, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}}[f(\phi)]^{-\frac{1}{2}}exp\left[-\frac{[Y_0 - \mu - g(\phi, W)]^2}{2\sigma^2 f(\phi)}\right]$$

For particular function $f(\phi)$ and $g(\phi, W)$ that need not be specified. It then follows on integrating out μ that

$$P(\phi, \sigma/Y) = \int P(\phi, \mu, \sigma/Y)\, d\mu \propto P(\phi)P(\sigma)P(W/\phi, \sigma) \propto P(\phi, \sigma/W).$$

Hence the posterior density of (ϕ, σ) is the same whether it is based on the levels Y ,What is lost in working with first differences is the opportnuity to specify have prior on μ or more generally on (μ, ϕ). We have noted, a considerable debate has ensured over the choice of prior for ϕ.For example ,Berger & Yang (1994) considered reference priors, Poirier(1991) considered a proper Gaussian prior. All of these authors allocated some prior probability to $\phi > 1$. Our view is conservative in that we believe, along with Schotman & Van Dijk(1991), Sims(1991) and Kwiaatkowski(1992), that evidence of explosive roots is indicative of alternative type of model not considered here. In any event we believe that explosive behaviour of the type implied by allowing some prior probability for $\phi > 1$ is not seen in time series arsing in practice and so we explicitly exclude this type of behaviour by only considering priors for $|\phi| < 1$. Another important consideration is how much prior information is actually available for ϕ. Can uniform prior really be considered a sensible choice if the investcator seriously beliveves that a random walk model could provide a reasonable explanation of the behaviour of a time series?. It becomes to use to be inconsistent to attach non-zero perior probability mass to $\phi = 1$on the one hand, while at the same time adopting a uniform prior when $\phi < 1$ on the other.Surely, if $\phi = 1$ is a likely value, then ϕ close to 1 is a perior more plausible than ϕ distant from 1.We must bear in mind that acceptance of a uniform prior implies that the perior belief

$P(-1 < \phi \leq 0) = 0.5$ must be plausible.In what follows we use for purposes of comparision the uniform perior for ϕ, together with two, sharper, beta periors with densities

$$P(\phi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta-1}}(1+\phi)^{\alpha-1}(1-\phi)^{\beta-1}, \qquad |\phi| < 1$$

Numerical Study:-

We generated 1000 samples of 100 observations from first order autoregressive processes with ϕ= 1, 0.95, 0.9, 0.85 and 0.8, the empirical distributions function of $P(M_1/W)$ for each of the three priors for each value ϕ: We also computed the proportion of the samples for which $P(M_1/W) > 0.5$; that is the proportion of the time that the rand m walk model would have been chosen. The results are presented in Table 1.1.

It is clear that, when the "true" process is stationary autoregressive, the very sharp prior, beta 2, performs extremely well and would clearly be the best of the three in such a situation. However, the opposite is the case for a random walk generating process. The sharp prior weight to those values of ϕ that are very close to 1 than dose the uniform prior. The random walk is therefore a much more plausible model for the "true" random walks in the uniform prior case. The message here is clear; priors matter in this case and the investigator should think carefully before blindly following a "formula" of noninformative or uniform priors for ϕ, which clearly favour the unit root model even when the process generating the data is stationary with $\phi = 0.90$. Of course, appropriate prior specification does require careful thought. For example values of ϕ very close to one are a prior more plausible for high frequency data then for low frequency data.

The shapes of the empirical distribution functions are interesting. For the beta2 prior, these functions exhibit very steep growth around the median. The implication is that only very rarely when this prior is adopted will the posterior odds favoring a particular model be either very high or very low. This relative posterior uncertainty results, of course, from the shape of

The beta 2 prior. With 100 observations it would be difficult to distinguish with great certainty between a random walk and a stationary first order autoregressive model in which the parameter was drawn at random from the beta2 distribution. Somewhat greater posterior certainty can result if the alternative to a random walk is a stationary first order autoregressive model with parameter drawn from the more disperses beta1 distribution. The results of table 1.1, clearly demonstrate that, for sample of 100 observations, the posterior odds favouring a particular model, and the outcome of any decision rule based on those odds, can strongly depend on the prior for the autoregressive parameter. The dependence of posterior on prior is sometimes viewed as a shortcoming of the

Table 1.1:-Proportion of Time $P(M_1/W) > 0.5$; for series of 100 observations

Prior for ϕ			
ϕ	**Uniform**	**beta1**	**beta2**
1	0.966	0.877	0.679
0.95	0.825	0.556	0.246
0.90	0.552	0.182	0.019
0.85	0.210	0.020	0.000
0.80	0.039	0.000	0.446

Bayesian approach, compared with alternative “objective” approaches to statistical inference. We take the opposite stance and claim that the flexibility in specifying a prior is strength in that approach to our problem. In effect, the prior for the parameter ϕ is a part of our model M_2. The comparison is then between a random walk model and a stationary first order autoregressive model for which nature first draws the autoregressive parameter from some distribution. An analysis who believes that this parameter is more likely than not to be close to one could and should incorporate this belief into prior distribution. We feel that attaching labels such as “objective” and “uninformative” when analysis based on a prior such as the uniform is misleading. Indeed, such a prior carries the “information” that the autregression viewed as an alternative to the random walk is a model in which the autoregressive parameter is just as likely to be negative as positive. It seems absurd to simultaneously hold this view while attaching probability mass to the belief that this parameter is exactly one. The classical hypothesis testing parading does not truly provide an objective criterion for deciding between models. Indeed, in the context of our problem, that approach is tantamount to imposing a uniform prior for ϕ, and manipulating the prior odds for two models to achieve a particular desired significance level.

Although our interest is not in the usual hypothesis testing framework, the results in Table 1.1 can be interpreted within that framework. In that context, the results for the uniform prior are interesting. The implication is that our decision rule has the properties of a set of size 0.039, and power 0.446 when $\phi = 0.90$.The usual Dickey-Fuller test of the random walk null against a stationary first order autoregressive model is based on least squares estimation o the autoregressive model. However, it has recently been recognized that tests of considerably more power can be achieved through alternative estimators, including maximum likelihood. Interesting simulation results are given by Pantula (1994). In this chapter for samples of 100 observations, with $\phi = 0.90$, it is reported that the usual Dickey-Fuller test at the 5% level has power 0.311. While a test based on maximum likelihood estimation has power 0.644. It appears from the results of table 1.1 that our Bayesian approach also captures these powers gains over the usual Dickey-Fuller test. Presumably that outcome arises through our use of the exact likelihood function of model M_2.

ARIMA$(p, 1, q)$ versus Stationary ARMA$(p + 1, q)$:-

The approach of the previous section can extended to allow comparison of an ARIMA model with a stationary ARMA model. As one possible generating process, we consider the ARIMA$(p, 1, q)$ model

$$\omega(B)(1 - B)Y_t = \theta(B)e_t \qquad \dots (6)$$

where

$$\omega(B) = 1 - \omega_1(B) - \cdots - \omega_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

And B is the backshift operator. It is assumed that p and q are given, and that the conditions for stationary and invariability are satisfied. The stationary alternative is the ARMA (p+1, q) model

$$\left[1 - (\phi + \omega_1)B - \sum_{j=2}^{p}\left(\omega_j - \omega_{j-1}\right)B^j + \omega_p B^{p+1}\right](Y_t - \mu) = \theta(B)e_t \qquad \dots (7)$$

This reduces to equation (6) when $\phi = 1$. Again, we work with first differences $W_t = Y_t - Y_{t-1}$; so from equation (6) and (7) the models to be analyzed are

$$M_1: \omega(B)W_t = \theta(B)e_t$$

and

$$M_2: \left[1-(\phi+\omega_1)B-\sum_{j=2}^{p}(\omega_j-\omega_{j-1})B^j+\omega_p B^{p+1}\right]W_t=\theta(B)(1-B)e_t$$

The posterior model probabilities again follow from equation (4), where now

$$P(W/M_i)=\iiint P(\phi,\gamma,\sigma/M_i)\,P(W/\phi,\gamma,\sigma,M_i)d\sigma d\gamma d\phi \qquad \ldots (8)$$

Where γ denotes the vector of parameters$\left(\omega_i,\theta_j\right)$. The likelihoods can be determined in terms of the elements of γ, either through the technique in Newbold (1974), or the computationally more efficient algorithm of Ansley (1979).We propose for M_2 the prior density

$$P(\phi,\gamma,\sigma/M_2)=P(\phi)P(\gamma)P(\sigma)$$

The same principles as in the previous section apply to the prior for ϕ, while as we saw there either a no informative or conjugate prior can be employed for σ. We would propose a no informative prior for γ the autoregressive and moving average parameters of equation (6) following, for example Box & Jenkins (1970), Monahan (1983), and Marriott & Smith (1992). This seems reasonable in practice, as typically one would expect the analyst to have little genuine prior information about these parameters. Of course, the same priors should be applied to M_1 as to M_2, expect that in the former ϕ is taken to be one. Finally, we not that the general form of the ARMA likelihood is

$$P(W/\phi,\gamma,\sigma,M_i)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}f(\phi,\gamma)exp\left[-\frac{g(\phi,\gamma)}{2\sigma^2}\right]$$

So that analytic integration over σ in equation (8) is possible. The numerical integration over γ can easily be accomplished using the Bayes integration rules employed by Marriott & Smith (1992).

Conclusion:-

If often occurs in empirical work that an investigator wants to assess the strength of evidence in the data for particular stationary model compared with a model with a unit autoregressive root. The use of Bayes theorem to compute posterior odds provides a natural and attractive mechanism for such model comparison. Presumably this issue arises when an investigator believes that an autoregressive root is either large and positive, or precisely equal to one. The Bayesian approach allows such prior belief to be incorporated into the analysis. We have demonstrated how Bayesian calculations can be carried out, noting the importance of the analyst giving careful thought to the question of what might be an appropriate prior. Of course, one possible strategy is to compute model posterior probabilities for a range of priors. This should allow the analyst insight into the impact of the prior, and the extent to which data are able to distinguish between, for example, a random walk and a first order autoregressive model whose parameter is drawn from a distribution with substantial mass close to one. Such insight is of course not possible from the testing of a null hypothesis at some arbitrary significance level. As a by –product of our investigation, we noted the superiority of decision rules when the exact likelihood function is employed to calculate posterior probabilities, and strongly recommend the use of exact likelihood for this problem, whether a Bayesian or classical hypothesis testing approach is used.

References:-

1. Aitkin, M. (1991) Posterior Bays factors (with discussion). "Journal of the Royal Statistical Society." Series B, 53, 111- 142.
2. Allen, D.M.(1974) The relationship between variable selection and prediction. "Technometrics," 16, 125-127.
3. Bayarri, M.J.(1999) Discussion of Bayesian Model Choice: What and Why?. In bayesian Statistics 6, Bernardo, J.M., Berger, J.O., David., A.P. and Smith, A.F.M (eds). Oxford: Oxford university Press 357-359.
4. Berger, J.O & Yang, R.Y. (1994). Noninformative priors and Bayesian for the AR(1) model. Econometrics theory. 10. 461-482.
5. Berger,J.O and Perichi, L.R. (1998) Accurate and stable Bayesian Model selection: The median Intrinsic Bayes Factor . Sankhya, B, 60, 1-18.
6. Bernardo, Jose M. (1999): Nested Hypothesis Testing: The Bayesian Reference Criterion" (with discussion), in J.Bernardoetal. (eds.): Bayesian Statistics 6: Proceedings of the Sixth Valencia Meeting,101-130 Oxford: Oxford University Press.

7. Billio M,Monfort A,Robert CP (1999) .”Bayesian estimation of switching ARMA models. Journal of Econometrics, Vo l93, pp 229-255
8. Broemeling,L.D. and Shaarawy,S.(1988). Time series: A Bayesian Analysis in Time Series Domain,In Bayesian Analysis of Time Series and Dynamic Models, (ed.J.C.Spall), pp. 1-21. Marcel Dekker Inc., New York.
9. Bruce E. Hansen (2005). “Challenges for econometric Model Selection”. Econometric Theory. Vol 21, 60 – 68.
10. Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. J. Am. Statist. Ass. 74, 153-160; correction, 75 (1980), 765.
11. Hedibert Freitas Lopes and Mike West (2004). “Bayesian Model Assessment in Factor Analysis.” Statistica sinica 14, 41-67.
12. Jan Springer (2011). “What is a Bayesian Model Selection Procedure,”
13. John F.Monahan (1983). “Fully Bayesian Analysis of ARMA time series Models”. Journal of Econometrics 21. 307 – 331.
14. Lindley, D.V. (1991) Discussion of Posterior Bayes Factor. Journal of the Royal statistical Society,
15. Philippe, A. (2001). “Bayesian model selection for autoregressive moving average Processes”. Technical report, IRMA, Lille.
16. Ricardo S.Ehlers and Stephen P.Brooks(2003). Bayesian analysis of order Uncertainty in ARIMA models Wadaed Uturbey (2006). Identification of ARMA “Models by Bayesian methods applied to stream flow data.” 9th International conference on probabilistic methods Applied to paver systems KTH Stockholm, Sweden June 11- 15.
17. Wasserman, L. (2000). Bayesian model selection and model averaging. J. Math.PsyChology 44, 92 - 107.