Economics Bulletin

Volume 34, Issue 2

On the implicit uniform BIC prior

Richard Startz
UC Santa Barbara

Abstract

I show how to find the uniform prior implicit in using the Bayesian Information Criterion to consider a hypothesis about a single normally distributed parameter. The ratio of the width of the implicit prior to the standard deviation of the parameter estimate is $\sqrt{2\pi n}$ for large samples.

Richard Startz, Department of Economics, University of California, Santa Barbara, startz@ucsb.edu
Citation: Richard Startz, (2014) "On the implicit uniform BIC prior", *Economics Bulletin*, Vol. 34 No. 2 pp. 766-771.
Contact: Richard Startz - startz@ucsb.edu.
Submitted: February 26, 2014. **Published:** April 04, 2014.

1. Introduction

The Bayesian Information Criterion (Raftery (1986a, 1986b, 1995)), or BIC, is widely used for Bayesian hypothesis testing because it offers a large sample approximation to the Bayes factor that is easily computed and that does not require an explicit statement of a prior. It may be useful to turn the problem around and ask what prior is implied by the BIC within an easy to interpret class.[1] Specifically, consider the problem of a point null hypothesis for a coefficient that is approximately normally distributed, a linear regression coefficient in a large sample being a common example. Ask then how wide a uniform interval around the null is implied as a prior by the BIC. It turns out that that the width of the interval is well-approximated by $\sqrt{2\pi n}$ times the standard error of the regression coefficient.

The Bayes factor, which is the posterior odds ratio assuming equal prior probabilities for the null and the alternative, is the ratio of the marginal likelihoods of the models being compared. The usual Bayesian procedure maps a likelihood function and a prior density into the marginal likelihood. Here we reverse the mapping. Given a value of the BIC and a likelihood, we compute a prior density. The implied prior from reversing the mapping is not unique; there are many prior specifications for the alternative which yield the same value for the marginal likelihood. Any prior for which the exact Bayes factor gives the same result as the BIC approximation is equally valid as an implicit prior. The implicit uniform prior may be particularly easy to interpret.[2]

Suppose the point null is $H_0: \theta = \theta_0$, then using Bayes theorem $p(H_0|\hat{\theta}) = p(\hat{\theta}|H_0) \times p(H_0)/p(\hat{\theta})$, where $\hat{\theta}$ is the estimated coefficient. Suppose the alternative is $H_A: \theta \neq \theta_0$ and we want to find a prior under the alternative of an interval of width w spread evenly around the null, $\pi(\theta|H_A) \sim U\left[\theta_0 - \frac{w}{2}, \theta_0 + \frac{w}{2}\right]$, so $p(H_A|\hat{\theta}) = p(\hat{\theta}|H_A) \cdot p(H_A)/p(\hat{\theta}) = \int_{\theta_0 - \frac{w}{2}}^{\theta_0 + \frac{w}{2}} p(\hat{\theta}|\theta)\pi(\theta|H_A)d\theta \cdot p(H_A)/p(\hat{\theta})$. Suppose the estimated coefficient $\hat{\theta} \sim N(\theta, \sigma^2)$, define the Studentized coefficient $z = (\hat{\theta} - \theta_0)/\sigma$, and let $\phi(\cdot)$ and $\Phi(\cdot)$ stand for the standard normal pdf and cdf respectively. Then (as shown in the Supplemental Material) the log Bayes factor favoring the null is

$$\begin{aligned}\log p(H_0|\hat{\theta})/p(H_A|\hat{\theta}) &= \log B_{0A} \\ &= \log \phi(z) - \log\left[\Phi\left(\frac{w}{2\sigma} - z\right) - \Phi\left(-\frac{w}{2\sigma} - z\right)\right] + \log\left(\frac{w}{\sigma}\right)\end{aligned} \tag{1}$$

Kass and Raftery (1995) give the BIC for n data points for this situation as

---

[1] Weakliem (1999) raises objections to the BIC on the grounds that the implicit prior may not be close to a prior that an investigator would choose. In a response to Weakliem, Raftery (1999) argues that the BIC "is sufficiently conservative, perhaps too conservative." Berger and Pericchi (2001) compare other objective Bayes factor methods to the BIC. For a comparison of model selection using the BIC and other criteria see Dey et. al. (2008).

[2] Raftery (1999) offers the unit information prior (UIP) as an implicit prior for the BIC, writing "roughly speaking, the prior distribution contains the same amount of information as would, on average, a single observation." Using the notation in the next paragraph, UIP for this problem would be $N\left(\hat{\theta}, \frac{\sigma^2}{n}\right)$. Note that the UIP is centered around the estimated coefficient rather than the null.

$$BIC = -z^2 + \log n \tag{2}$$

Note that for both the Bayes factor and the BIC the data enters only through the statistic z.

Since the BIC approximates twice the log Bayes factor, equating two times equation (1) with equation (2) lets one solve for the value of w implicit in the BIC. The equation simplifies a little since $\log \phi(z) = -\frac{1}{2}\log 2\pi - \frac{1}{2}z^2$, so

$$-\log 2\pi - 2\log\left[\Phi\left(\frac{w}{2\sigma} - z\right) - \Phi\left(-\frac{w}{2\sigma} - z\right)\right] + 2\log\left(\frac{w}{\sigma}\right) = \log n \tag{3}$$

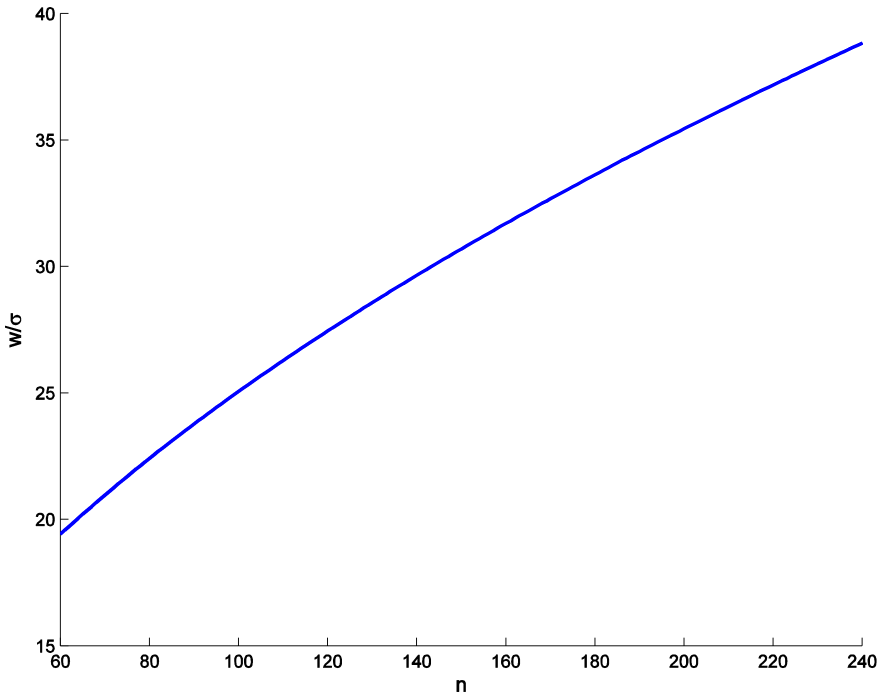
Equation (3) defines the width of the implicit BIC prior relative to the standard deviation of the estimated coefficient as a function of n and z. The values of w/σ solving equation (3) are plotted in Figure 1 for a range of values of n and $z \in [0,5]$.[3] (For $z \in [0,5]$ the values of w/σ are identical to the eye, so only one curve is apparent.)

Figure 1 illustrates two points. First, it is often said that the BIC is conservative. This is seen to be true in the sense that with only 60 observations the implied prior covers almost 10 standard deviations on either side of θ_0.[4] Second, for large values of w/σ, $\Phi\left(\frac{w}{2\sigma}\right) - \Phi\left(-\frac{w}{2\sigma}\right) \approx 1$ so $w/\sigma \approx \sqrt{2\pi n}$. In fact even for $n = 60$ the approximation is essentially perfect.

[3] Note that the Bayes factor favoring the null, equation (1), is not necessarily monotonic in w/σ and when regarded as a function of the width of the prior has a minimum value in favor of the null (for a given z). Equivalently, in what may be a curiosum, the uniform prior implies a maximum value of the Bayes factor in favor of the alternative. This means that for very large values of z relative to n there may not be a solution to equation (3), or there may be two solutions. However, for large n relative to z there is always a single solution, as presented in Figure 1

[4] Another way to see that the BIC is conservative is to note that for the UIP, the standard deviation of the prior is $\sqrt{n}$ times as wide as the standard deviation of $\hat{\theta}$.

Figure 1: Implicit width of uniform prior for BIC, $z \in [0,5]$



Finally, for many estimators the standard deviation of the estimate is inversely proportional to the square root of the number of observations so the implicit BIC prior is constant for more than a modest number of observations. For the sample mean as the canonical example, if $y_i = \theta + \epsilon_i$ and $\epsilon_i \sim i.i.d.\ N(0, \sigma_\epsilon^2)$ so $\sigma^2 = \sigma_\epsilon^2/n$, then $w \approx \sqrt{2\pi\sigma_\epsilon^2}$.

2. An Example

Economists are interested in whether the stock market is "weak form efficient," meaning whether the market return can be predicted by lagged returns. Economists expect the coefficient on lagged returns to be close to zero—zero being the value under weak form efficiency—for two different reasons. First, a large coefficient suggests the existence of profit opportunities that one would expect to be competed away in a well-functioning market. Second, economic theory suggests rates of return should be relatively stable. At the very least rates of return should not be explosive, so the coefficient on lagged returns should be less than 1.0 in absolute value. Table 1 shows the results of regressing the S&P 500 return on the lagged return for both monthly and daily data.[5] The frequentist results "weakly reject" market efficiency with the monthly data and strongly reject using the daily data. In contrast, the Bayes factors computed from the BIC are

[5] The underlying data is the series SP500 from the St. Louis Federal Reserve Economic Database (FRED). The daily return is $\log(SP500_t) - \log(SP500_{t-1})$. The monthly return uses SP500 on the last day of each month.

much more conservative about rejecting the market efficiency. The Bayes factor from the monthly data provides positive support for market efficiency while the Bayes factor from the daily data indicates a toss-up. As a substantive point the $U[-1.25,1.25]$ implicit priors for the BIC are overly conservative, as they include coefficients outside the $[-1,1]$ stationary interval which are not credible. Thus one benefit of computing implicit BIC priors is as an aid in deciding in a particular instance whether more explicit prior information should be brought to bear.

Table 1: BIC and implicit prior for S&P 500 returns regression

Data	S&P 500 returns, monthly 1957M03 2012M08	S&P 500 returns, daily 1/04/1957 8/30/2012
observations	666	14,014
Coefficient on lagged return (*t*-statistic) [standard error] *p*-value	0.067 (1.74) [0.039] <0.082>	0.026 (3.09) [0.008] <0.002>
Bayes factor from BIC	5.67	1.00
w	2.50	2.51
$^{w}/_{std.\ err.}$	64.7	296.7

3. Summary

Since computation of the BIC does not account for the prior, any number of implicit priors are consistent with the BIC. However, backing out the uniform prior implied for a single normally distributed estimator is straightforward and may shed light on whether use of the BIC is consistent with what might be an investigator's true priors.

4. References

Berger, J. and L. Pericchi (2001) "Objective Bayesian methods for model selection: Introduction and comparison," in *Model Selection* (P. Lahiri, ed.) 135-207. IMS, Beachwood, OH.

Dey, T., H. Ishwaran, and J. S. Rao (2008) "An In-Depth Look at Highest Posterior Model Selection," *Econometric Theory*, Vol. 24, 377-403.

Kass, R.E. and A.E. Raftery (1995) "Bayes Factors," *Journal of the American Statistical Association*, Vol. 90, No. 430, 773-796.

Raftery, A.E. (1986a) "Choosing Models for Cross-Classifications," *American Sociological Review*, Vol. 51, No. 1,145-146.

________ (1986b) "A Note on Bayes Factors for Log-linear Contingency Table Models With Vague Prior Information," *Journal of the Royal Statistical Society, Series B,* vol. 48. pp. 249-250.

________ (1995) "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111-163.

________ (1999) "Bayes Factors and BIC: Comment on 'A Critique of the Bayesian Information Criterion for Model Selection,'" *Sociological Methods & Research*, 27(3), 411-427.

Weakliem, D.L. (1999) "A Critique of the Bayesian Information Criterion for Model Selection," *Sociological Methods & Research*, 27(3), 359–397.