Journal of Applied Statistics, Vol. 26, No. 1, 1999, 97–101

Remarks on the 'Bayesian' method of moments

SEYMOUR GEISSER[1] **& TEDDY SEIDENFELD**[2], [1]*University of Minnesota, Minneapolis, USA and* [2]*Carnegie Mellon University, Pittsburgh, PA, USA*

SUMMARY *Zellner has proposed a novel methodology for estimating structural parameters and predicting future observables based on two moments of a subjective distribution and the application of the maximum entropy principle—all in the absence of an explicit statistical model or likelihood function for the data. He calls his procedure the 'Bayesian method of moments' (BMOM). In a recent paper in this journal, Green and Strawderman applied the BMOM to a model for slash pine plantations. It is our view that there are inconsistencies between BMOM and Bayesian (conditional) probability, as we explain in this paper.*

1 Introduction

Zellner (1996) proposes a novel methodology for estimating structural parameters and predicting future observables based on two moments of a subjective distribution and the application of the maximum entropy principle—all in the absence of an explicit statistical model or likelihood function for the data. He calls his procedure the 'Bayesian method of moments' (BMOM). It is our view that there are inconsistencies between his approach and Bayesian (conditional) probability, and that the procedure is misnamed as being 'Bayesian'. A more appropriate designation might be the 'maximum entropy method of moments' (MEMOM). We give the reasons for our view for only the simplest case he considers. Other cases suffer from the same predicament, however.

2 BMOM

For the simplest BMOM case, we start with the structural model

$$y_i = \theta + u_i, \qquad i = 1, \ldots, n \tag{1}$$

Correspondence: T. Seidenfeld, Carnegie Mellon University, Baker Hall 135, Pittsburgh, PA 15213-3890, USA.

where only the y_i terms are observed. Let $y^{(n)}$ denote these observations. Both θ and u_i are unknown, and no statistical model, i.e. no likelihood function, is specified for them apart from what equation (1) entails is impossible.

Green and Strawderman (1996) adapt Zellner's BMOM inference and this simple regression model to an agricultural 'growth and yield' problem. Correctly, we think, they emphasize the importance of giving Bayesian (full) distributions for parameters and future observable quantities of interest, instead of providing merely point estimates for these terms, as has been done in the literature that they cite. However, the inconsistencies of Zellner's BMOM approach with Bayesian inference that we point out for equation (1) (see later) are also present in their reasoning.

We identify two steps in Zellner's derivation of BMOM probabilities. The first step involves Assumptions I and II, which fix two moments of the 'posterior' distribution of the error term $\bar{\mathbf{u}}$ as a function of the observed sample $y^{(n)}$ and a new parameter σ^2. This first step induces 'posterior' first moments on each of the two parameters (θ, σ^2) and a conditional second moment on θ, given σ^2. Let

$$\bar{y}_n = n^{-1}\sum_{i=1}^{n} y_i$$

and

$$s_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$$

From Assumption I, we have that $E[\theta|y^{(n)}] = \bar{y}_n$ and, from Assumption II, we have that $E[\sigma^2|y^{(n)}] = s_n^2$ and then that $\mathrm{Var}[\theta|\sigma^2, y^{(n)}] = \sigma^2/n$. Zellner also applies these two moment assumptions to the predictive probability $p(y_{n+1}|y^{(n)})$, so that

$$E[y_{n+1}|\sigma^2, y^{(n)}] = \bar{y}_n$$

and

$$\mathrm{Var}[y_{n+1}|\sigma^2, y^{(n)}] = \frac{\sigma^2(n+1)}{n}$$

The second step for arriving at the BMOM probability is to use the MAXENT principle, in order to fix exact distributions with these moments as constraints. Thus, Zellner obtains

$$p(\theta|\sigma^2, y^{(n)}) = \sqrt{\frac{n}{2\pi\sigma^2}}\, e^{-\frac{n}{2\sigma^2}(\theta-\bar{y}_n)^2} \tag{2}$$

$$p(\sigma^2|y^{(n)}) = \frac{1}{s_n^2}\, e^{-\frac{\sigma^2}{s_n^2}} \tag{3}$$

and

$$p(y_{n+1}|y^{(n)}) = \frac{1}{\sqrt{2}s_n'}\, e^{-\frac{\sqrt{2}}{s_n'}|yn+1-\bar{y}_n|} \tag{4}$$

where

$$s_n'^2 = \left(\frac{n+1}{n}\right) s_n^2.$$

As a first warning that this inference is suspicious, we note that the BMOM distribution of σ^2 does not converge to σ^2 as n increases. In other words, let $z_n = \sigma^2/s_n^2$. Then, the BMOM density satisfies $p(z_n|y^{(n)}) = e^{-z_n}$ for all n. To put this in perspective, note that $E[\sigma^2|y^{(n)}] = s_n^2$. However, this BMOM distribution of σ^2 has its median at $(\ln 2)s_n^2 \approx 0.7s_n^2$, which is bounded away from its mean s_n^2, and its quartiles are approximately $0.3s_n^2$ and $1.4s_n^2$. We think that this is a relatively minor anomaly, which we set aside here in order to discuss two non-Bayesian aspects of BMOM inferences.

3 Non-Bayesian nature of BMOM: Global model

Of course, the BMOM joint posterior density for the two parameters is just the product

$$p(\theta, \sigma^2|y^{(n)}) = p(\theta|\sigma^2, y^{(n)})p(\sigma^2|y^{(n)}) \tag{5}$$

Next, consider the Bayesian condition

$$p(y_{n+1}|y^{(n)}) = \int\int p(y_{n+1}|\theta, \sigma^2, y^{(n)})p(\theta, \sigma^2|y^{(n)})\, d\theta\, d\sigma^2 \tag{6}$$

Equations (4) and (5) identify two of the three terms in equation (6) with only the likelihood $p(y_{n+1}|\theta, \sigma^2, y^{(n)})$ not yet explicitly given. However, the integral equation (6) has a unique solution:

$$p(y_{n+1}|\theta, \sigma^2, y^{(n)}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta - y_{n+1})^2}. \tag{7}$$

First, let us consider the existence of a global Bayesian model for BMOM inference, by working backwards from the joint BMOM posterior of equation (5). In other words, according to Bayes, there must exist some likelihood $L(y^{(n)}|\theta, \sigma^2)$ and prior $p(\theta, \sigma^2)$ where

$$p(\theta, \sigma^2|y^{(n)}) \propto L(y^{(n)}|\theta, \sigma^2)p(\theta, \sigma^2) \tag{8}$$

although these may not be unique. Assume that we are willing to make the predictive assumptions I and II for all n. Then, because the solution in equation (7) is unique for all $n > 1$, we extract a joint likelihood which is, in fact, the independent and identically distributed (i.i.d.) $N(\theta, \sigma^2)$ statistical model for y_i:

$$L(y^{(n)}|\theta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta - y_i)^2}. \tag{9}$$

However, if we use the BMOM posterior of equation (5) and induced likelihood of equation (9) to identify the prior $p(\theta, \sigma^2)$ in equation (8), then we discover that, unfortunately, the ratio

$$p(\theta, \sigma^2) \propto \frac{p(\theta, \sigma^2|y^{(n)})}{L(y^{(n)}|\theta, \sigma^2)} \tag{10}$$

depends on the sample statistics $\bar{y}_n$ and s_n^2; therefore, it changes for each n, which means that it is not a prior in the Bayesian sense. Hence, either there is no global Bayesian model for BMOM inference, or the predictive assumptions I and II are allowed only for certain samples. Notably, which values of n may these be?

4 Non-Bayesian nature: Local model

The Bayesian approach yields coherent solutions for updating according to the laws of conditional probability. Thus, to have even a local Bayes model, the BMOM must satisfy

$$p(\theta, \sigma^2 | y^{(n+1)}) = \frac{p(y_{n+1} | y^{(n)}, \theta, \sigma^2) p(\theta, \sigma^2 | y^{(n)})}{p(y_{n+1} | y^{(n)})} \tag{11}$$

On the right-hand side of equation (11), the two terms in the numerator are as in equations (7) and (5) respectively, and the denominator is from equation (3). In solving the right-hand side of equation (11), we do not obtain the BMOM posterior that corresponds to equation (5) evaluated at $y^{(n+1)}$. In other words, the right-hand side of equation (11) does not yield

$$\sqrt{\frac{n+1}{2\pi\sigma^2}} e^{-\frac{n+1}{2\sigma^2}(\theta - \bar{y}_{n+1})^2} \times \frac{1}{s_{n+1}^2} e^{-\frac{\sigma^2}{s_{n+1}^2}}, \tag{12}$$

which is the BMOM solution to $p(\theta, \sigma^2 | y^{(n+1)})$ under the very same assumptions that lead to the three terms that constitute the right-hand side of equation (11).

Instead, the right-hand side of equation (11) yields

$$\sqrt{\frac{n+1}{2\pi\sigma^2}} e^{-\frac{n+1}{2\sigma^2}(\theta - \bar{y}_{n+1})^2} \times \frac{1}{\sigma s_n \sqrt{\pi}} e^{\frac{-\left(\frac{1}{\sqrt{2}} s_n |\bar{y} - y_{n+1}| - \sigma^2 \sqrt{\frac{n+1}{n}}\right)^2}{\frac{n+1}{n}\sigma^2 s_n^2}}. \tag{13}$$

Evidently, the conflict between the BMOM rule and Bayesian updating relates to the (marginal) posterior for the parameter σ^2. In short, either there is no local Bayes model of BMOM inference, or the predictive assumptions I and II cannot be made for even a future sample of size 1.

5 Concluding remarks

The reader may better understand the non-Bayesian aspects of BMOM probability by remembering the two steps that Zellner takes in deriving it. The first step is taken with Assumptions I and II, which fix moments of the posterior (θ, σ^2) parameter distribution, and of the predictive distribution for y_{n+1}, all as a function of the observed sample $y^{(n)}$. Specifically, from Assumption I, we have that $E[\theta | y^{(n)}] = \bar{y}_n$ and, from Assumption II, we have that $E[\sigma^2 | y^{(n)}] = s_n^2$ and then that $\mathrm{Var}[\theta | \sigma^2, y^{(n)}] = \sigma^2/n$. Applied to the predictive probability $p(y_{n+1} | y^{(n)}]$, the two assumptions yield $E[y_{n+1} | \sigma^2, y^{(n)}] = \bar{y}_n$ and $\mathrm{Var}[y_{n+1} | \sigma^2, y^{(n)}] = \sigma^2(n+1)/n$. Recall, also, that the second step in arriving at the BMOM probability is to use the MAXENT principle to fix exact distributions with these moments as constraints.

The first step, by itself, is not in conflict with Bayesian theory, as the following analysis shows. Specifically, for coherence of conditional expectations (in fact, as a consequence of the law of total probability), we require that, for each random variable X, we have (assuming that $E[X|y^{(n)}]$ exists)

$$E[E[X|y^{(n+1)}]\,|y^{(n)}] = E[X|y^{(n)}] \tag{14}$$

Now, from Assumption I applied to the posterior at the two sample sizes, i.e.

$$E[\theta|y^{(n)}] = \bar{y}_n, \qquad E[\theta|y^{(n+1)}] = \bar{y}_{n+1} \tag{15}$$

we obtain

$$E[y_{n+1}|y^{(n)}] = \bar{y}_n \tag{16}$$

just as is needed for consistency with the first predictive moment about y_{n+1}.

The consequences of the two versions of Assumptions II are straightforward too. Assume that

$$E[\sigma^2|y^{(n)}] = s_n^2, \qquad E[\sigma^2|y^{(n+1)}] = s_{n+1}^2 \tag{17}$$

Now, expanding s_{n+1}^2, write

$$s_{n+1}^2 = s_n^2\frac{n-1}{n} + \frac{1}{n+1}[y_{n+1} - \bar{y}_n]^2 \tag{18}$$

Then, because $E[s_{n+1}^2|y^{(n)}]$ must equal s_n^2 by the 'law', we have the simple result

$$s_n^2 = \frac{n}{n+1}\,E[(y_{n+1} - \bar{y}_n)^2|y^{(n)}] \tag{19}$$

From before, we have that $E[y_{n+1}|y^{(n)}] = \bar{y}_n$. Thus, we obtain

$$s_n^2 = \frac{n}{n+1}\mathrm{Var}\,[y_{n+1}|y^{(n)}] \tag{20}$$

However, we have by BMOM

$$\mathrm{Var}\,[y_{n+1}|y^{(n)}] = \frac{n+1}{n}s_n^2 \tag{21}$$

just as is needed for coherence.

Thus, Zellner's use of sample moments to fix the BMOM moments for the parameters is coherent. However, in light of the results about global and local non-Bayesian aspects of the BMOM probability, we conclude that, here, it is the applications of the MAXENT principle which are the source of conflict between the BMOM and Bayes rule for updating.

Acknowledgement

This work was supported in part by NIGMS Grant 25271.

REFERENCES

GREEN, E. J. & STRAWDERMAN, W. E. (1996) A Bayesian growth and yield model for slash pine plantations, *Journal of Applied Statistics*, 23, pp. 285–299.

ZELLNER, A. (1996) Bayesian method of moments (BMOM) analysis of mean and regression models. *In:* J. C. Lee *et al.* (Eds) *Modelling and Prediction*, pp. 61–74 (New York, Springer).