Bayesian interpretation of quasi-Bayesian inference in a normal hierarchical model

Jacek Osiewalski[1]

Abstract
In modern parametric statistics and its applications latent variables and random effects are widely used, and their estimation or prediction is of interest. Under some prior assumptions, Bayes formula can be used to obtain their posterior distribution. However, on the sampling-theory grounds, the unknown constants which appear in the prior distribution are estimated by means of the data being actually modelled. We call such approaches quasi-Bayesian; *parametric empirical Bayes* is an important example. In this paper we propose theoretical framework that enables the Bayesian interpretation of incoherent, quasi-Bayesian inference techniques. Our framework amounts to establishing a formal Bayesian model that justifies a quasi-Bayesian "posterior" (resulting from some data-based "prior") as a valid posterior distribution. From such Bayesian model, i.e. the joint distribution of observations and other quantities, one can deduce the true sampling model, that is the conditional distribution of observations, and the true prior (or marginal) distribution of the remaining quantities – latent variables or parameters. Since analytical derivations are possible in very specific cases, this paper presents only a simple, illustrative example based on a normal hierarchical model. It clearly shows that quasi-Bayesian approaches can lead to posterior distributions, which formally correspond to sampling models and prior distributions different than the assumed (declared) ones.

Keywords: *Bayesian statistics, coherent inference, empirical Bayes procedures, random parameters, shrinkage estimation*
JEL Classification: *C11, C18, C51*

1. Introduction

Bayes Theorem, usually used as Bayes formula for density functions of continuous random variables, is a central, important tool of Bayesian statistics, but it is not the only characteristic of this mode of statistical modelling and inference. There are two defining features of the Bayesian approach to statistics. Probabilistic representation of uncertainty about observations (available, missing, future), latent variables (or random parameters) and classical parameters (unknown constants) is the main feature of *Bayesian modelling*, and treating all "unknowns" as random variables is closely related to the concept of subjective probability. Obeying rules of probability calculus is then the main characteristic of *Bayesian inference*. Obviously, Bayes formula is one of these rules (and a very useful one), but following it in isolation from other rules does not mean conducting Bayesian inference.

In modern statistics and its numerous subject areas (like econometrics) latent variables, random effects and other unobservable random quantities are frequently used and their estimation or prediction is usually of particular interest. Bayes formula can be used to obtain their posterior distribution, given appropriate distributional assumptions. Then, the posterior mean can be used

[1] Cracow University of Economics (Faculty of Management, Dept. of Econometrics and Operations Research); 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: eeosiewa@cyf-kr.edu.pl.

as the estimator (or predictor), even within a non-Bayesian approach to statistics. However, the posterior distribution (and the posterior mean) of a latent variable depends on unknown constants (parameters) of the assumed class of marginal (or prior) distributions of this variable. The purely Bayesian solution amounts to treating all unknowns probabilistically and using probability rules on each level of the hierarchical model. On the sampling-theory grounds, however, the unknown constants in the prior distribution are estimated on the basis of the data being actually modelled. Such approach has been popular since 1970s under the name (*parametric*) *empirical Bayes*; see Efron and Morris (1972), Morris (1983) and Casella (1985). Empirical Bayes (EB) methods can be described as incoherent by an orthodox Bayesian, who by coherency means following basic rules of probability. While such description is formally exact and true, it does not provide us with a deeper Bayesian understanding of incoherent inferences that are practically useful and frequently adopted in empirical research.

We propose theoretical framework that enables the purely Bayesian interpretation of incoherent, quasi-Bayesian inference techniques such as EB. Our framework amounts to establishing such formal Bayesian model that justifies a quasi-Bayesian "posterior" (resulting from some data-based "prior") as a valid posterior. From this Bayesian model, i.e. the joint distribution of observations and other quantities, which justifies the posterior in question, one can deduce (at least in principle) the true sampling model, that is the conditional distribution of observations, and the true prior (or marginal) distribution of the remaining quantities – latent variables and parameters. Since analytical derivations are possible only in very specific cases, we present a simple, illustrative example in this paper. However, it clearly shows that incoherence of quasi-Bayesian approaches can lead to posterior distributions, which formally correspond to sampling models and prior distributions different than the assumed (declared) ones. In the next section our general framework is presented. Section 3 is devoted to the Bayesian and quasi-Bayesian (EB type) approaches to dealing with random effects in a normal hierarchical model. Again, our analysis is kept as simple as possible in order to enable fully analytical derivations. Section 4 concludes.

2. Bayesian interpretation of "posteriors" resulting from data-based "priors"

By quasi-Bayesian inferences we mean approaches where Bayes formula is used mechanically, outside the fully probabilistic Bayesian context that guarantees coherence. Consider the conditional density of observations $p(y|\omega) = g(y;\omega)$, corresponding to some parametric statistical model, and the prior density $p(\omega)$ from some parametric family. Then the Bayesian inference relies on the posterior density function $p(\omega|y) \propto p(y|\omega)p(\omega)$ that, again, belongs to some parametric family. Assuming that instead of specifying the prior hyper-parameters, one estimates them using the actual data and inserts to the formula for the posterior density. It leads to the "posterior" density $p^*(\omega|y)$ that amounts to using the data-based "prior" $p^*(\omega) = f(\omega; y)$, which cannot be the marginal distribution of parameters (and other unknowns, like latent variables). Thus, the density $p^*(\omega|y) \propto p(y|\omega)p^*(\omega) \propto g(y;\omega)f(\omega;y)$ is not the posterior density in the original model with initially assumed $p(y|\omega)$ and $p(\omega)$. However, such $p^*(\omega|y)$ is a member of the same parametric family as $p(\omega|y)$, so it is a well-defined probability density function and it can be the true, formal posterior in a completely different Bayesian model. The main question we pose here is as follows: what are the true building blocks (sampling model and prior) coherently justifying $p^*(\omega|y)$ that was initially obtained from a data-based "prior"? It would be useful to know hidden assumptions underlying formal Bayesian inference based on $p^*(\omega|y)$.

Specifying the marginal distribution of the parameters with the use of the actual data is a fundamental form of incoherence, and $p^*(\omega|y)$ obtained through Bayes formula corresponds then to some statistical model and prior assumptions, which have to be discovered. In order to obtain $p^*(\omega|y)$ as the posterior density, we consider the joint distribution of observations and parameters that is characterised by the density function:

$$\tilde{p}(y,\omega) \propto p(y|\omega)p^*(\omega) \propto g(y;\omega)f(\omega;y),$$

which can be decomposed in two ways:

$$\tilde{p}(y,\omega) = \tilde{p}(\omega|y)\tilde{p}(y) = \tilde{p}(y|\omega)\tilde{p}(\omega).$$

Note that $\tilde{p}(\omega|y)$ and $\tilde{p}(y|\omega)$ are probability density functions if and only if $\tilde{p}(y)$ and $\tilde{p}(\omega)$ are densities of σ-finite measures. Then $\tilde{p}(\omega|y) = p^*(\omega|y)$ by construction, since

$$\tilde{p}(\omega|y) \propto \tilde{p}(y,\omega) \propto g(y;\omega)f(\omega;y).$$

Also note that the joint density $\tilde{p}(y,\omega)$ represents the Bayesian model corresponding to the sampling density $\tilde{p}(y|\omega)$ and the prior density $\tilde{p}(\omega)$. In this Bayesian model we obtain $p^*(\omega|y)$ as the formal (true) posterior density.

3. Quasi-Bayesian and Bayesian analysis of hierarchical models

Now we consider a statistical model with hierarchical structure which can be the starting point for explanation and justification of the EB approach. However, in order to use purely analytical tools and obtain closed-form solutions, only normal distributions with known variances and covariances are examined here. More general priors that correspond to the ones in the basic EB literature are left for future research. Hierarchical Bayesian estimation of a more general random parameters regression type model is presented in Greene (2008, section 18.8); however, it cannot be examined analytically.

The hierarchical structure of a statistical model amounts to assuming the conditional distribution of observations $p(y|\theta) = g(y;\theta)$ ($y \in Y$, $\theta \in \Theta$) where the parameters are in fact latent random variables with some distribution dependent on deeper parameters (treated as unknown constants on non-Bayesian grounds); its density is denoted as $f_0(\theta;\alpha)$, $\alpha \in A \subseteq \mathbb{R}^s$. Then the joint distribution of observations and latent variables (with α fixed) can be written and decomposed in the following way:

$$p(y|\theta)f_0(\theta;\alpha) = g(y;\theta)f_0(\theta;\alpha) = f_1(\theta|y;\alpha)\,h(y;\alpha),$$

where $h(y; \alpha)$ and $f_1(y|\theta; \alpha)$ are the densities of the marginal distribution of observations and the conditional distribution of latent variables, respectively. Of course, Bayes formula describes the relation between all four density functions:

$$f_1(\theta|y; \alpha) = g(y; \theta) f_0(\theta; \alpha)/h(y; \alpha) \propto g(y; \theta) f_0(\theta; \alpha).$$

Thus, in order to make inferences on latent variables (given observations) Bayes formula is used. However, within non-Bayesian approaches, like EB, no prior distribution is assumed for the deeper parameters α, which are estimated using (for example) the maximum likelihood principle applied to the density $h(y; \alpha)$, which is considered as a function of α (for any given y). Then the estimate of α, e. g.,

$$\hat{\alpha} = \hat{\alpha}_{ML} = \arg\max L(\alpha; y) = \arg\max h(y; \alpha), \quad \alpha \in A,$$

is inserted into the posterior density of latent variables. So such quasi-Bayesian approach uses

$$p^*(\theta|y) = f_1(\theta|y, \hat{\alpha}) \propto p(y|\theta) f_0(\theta; \hat{\alpha}),$$

i.e. the “posterior” of θ corresponding to the “prior” with hyper-parameter based on y. This cannot be a formal Bayesian approach, although Bayes formula has been used at an earlier stage. Thus, it is called quasi-Bayesian.

Now let us consider the Bayesian hierarchical model (BHM)

$$p(y, \omega) = p(y, \theta, \alpha) = p(y|\theta)\, p(\theta|\alpha)\, p(\alpha),$$

where $\omega = (\theta, \alpha)$, $p(\alpha)$ is the prior density for $\alpha \in A$, and conditional independence: $y \perp \alpha|\theta$ (characteristic for hierarchical models) leads to $p(y|\omega) = p(y|\theta)$. We use the same notation $p(y|\theta) = g(y; \theta)$ and $p(\theta|\alpha) = f_0(\theta; \alpha)$ as in the quasi-Bayesian case. Basic rules of probability calculus lead to the decomposition of our Bayesian model; this decomposition serves making inferences on $\omega = (\theta; \alpha)$:

$$p(y, \theta, \alpha) = p(y)\, p(\theta, \alpha|y) = p(y)\, p(\alpha|y)\, p(\theta|y, \alpha),$$

where

$$p(\theta|y, \alpha) = \frac{p(y|\theta) p(\theta|\alpha)}{p(y|\alpha)} = \frac{g(y; \theta) f_0(\theta; \alpha)}{h(y; \alpha)} = f_1(\theta|y; \alpha),$$

$$p(\alpha|y) = \frac{p(y|\alpha) p(\alpha)}{p(y)} = \frac{h(y; \alpha) p(\alpha)}{p(y)}, \quad p(y) = \int_A p(y|\alpha) p(\alpha)\, d\alpha.$$

Note that Bayes formula has been used twice: for latent variables θ (given parameters α) and for parameters α themselves. Again, according to probability rules, the marginal density of latent variables is a continuous mixture

$$p(\theta|y) = \int_A f_1(\theta|y; \alpha) p(\alpha|y)\, d\alpha,$$

which is the basis of Bayesian inference on θ. Clearly, uncertainty about α is now fully taken into account, contrary to the quasi-Bayesian case, where unknown α is simply replaced by its point estimate.

In order to provide a coherent Bayesian interpretation of the quasi-Bayesian "posterior" $p^*(\theta|y) = f_1(\theta|y, \hat{\alpha})$ in any particular case, we should consider the joint density (Bayesian model) $\tilde{p}(y, \theta)$ that formally leads to $\tilde{p}(\theta|y) = p^*(\theta|y)$. Then the sampling density $\tilde{p}(y|\theta)$ and the prior density $\tilde{p}(\theta)$ Bayesianly justify quasi-Bayesian inference based on $p^*(\theta|y)$. Now we use a simple normal hierarchical model as a purely analytical example of our approach. Firstly, we present the strict (coherent) Bayesian analysis. Secondly, quasi-Bayesian results are given and their Bayesian interpretation is derived. Let θ_i denote an unobservable characteristic, randomly distributed over n observed units ($i = 1,\ldots, n$). Since θ_i is specific for the i-th unit, it can be called an individual effect. Let $x_{ij} \sim iiN(\theta_i, c_0)$ (for $j = 1,\ldots, m$) denote m independent measurements of θ_i, with c_0 known. Let e_m denote the $m \times 1$ vector of ones. Then $y_i = \frac{1}{m}e'_m x_i = \bar{x}_i$ is a sufficient statistic (for given θ_i) and n average measurements grouped in $y = (y_1 \ldots y_n)'$ are independent and normally distributed given $\theta = (\theta_1 \ldots \theta_n)'$, i.e. $y_i|\theta \sim iiN(\theta_i, c)$, where $c = c_0/m$. As it has been assumed, the unobserved parameters or individual effects are random (thus they are latent variables), independent and normally distributed with the same unknown mean α and the same known variance d, i.e. $\theta_i \sim iiN(\alpha, d)$. Thus in this example

$$p(y|\theta) = f_N^n(y|\theta, cI_n), \quad f_0(\theta;\alpha) = f_N^n(\theta|\alpha e_n, dI_n).$$

where $f_N^k(.|b, C)$ denotes the density function of the k-variate normal distribution with mean vector and covariance matrix C.

We can decompose the product $p(y|\theta) f_0(\theta; \alpha)$ into $f_1(\theta|y; \alpha)\, h(y; \alpha)$, where

$$h(y;\alpha) = \int_{\mathbb{R}^n} p(y|\theta) f_0(\theta;\alpha)\, d\theta = f_N^n(y|\alpha e_n, (c+d)I_n)$$

is the density function of the marginal distribution of the observation vector (given α), and

$$f_1(\theta|y;\alpha) = f_N^n\left(\theta \left| \frac{d^{-1}}{c^{-1}+d^{-1}}\,\alpha e_n + \frac{c^{-1}}{c^{-1}+d^{-1}} y,\ \frac{1}{c^{-1}+d^{-1}} I_n \right.\right)$$

is the posterior density of the vector of random effects (given α), with the mean

$$E(\theta|y;\alpha) = w \cdot \alpha e_n + (1-w)\cdot y, \quad w = \frac{d^{-1}}{c^{-1}+d^{-1}} = \frac{c}{c+d} \in (0,1).$$

Note that the posterior precision (the inverse of posterior variance) is the sum of the sample precision c^{-1} and the prior precision d^{-1}, and the posterior mean is a weighted average of the vector of prior means and the observation vector – with weights equal to the share of prior or sample precision in the posterior (or final) precision. Thus $E(\theta|y; \alpha)$ is a point (in $\Theta = \mathbb{R}^n$) that lies on the line segment between $(\alpha\ \alpha\ ...\ \alpha)'$ and $(y_1\ y_2\ ...\ y_n)$.

It is worth stressing that the conditional density $f_1(\theta|y;\alpha)$ follows from Bayes formula for any fixed α, so to this point the presented approach obeys coherence. However, the deeper parameter (the prior mean α) is unknown, so there are two possible ways of treating it. On the sampling theory grounds (like in the EB approach) some point estimate $\hat{\alpha}$ is inserted into $f_1(\theta|y;\alpha)$, which results in $p^*(\theta|y) = f_1(\theta|y,\alpha=\hat{\alpha})$. In our example we get

$$p^*(\theta|y) = f_N^n\left(\theta|\hat{\theta}_{EB}, \frac{1}{c^{-1}+d^{-1}}I_n\right),$$

where

$$\hat{\alpha} = \bar{y} = \frac{1}{n}e_n'y, \quad \hat{\theta}_{EB} = w\bar{y}e_n + (1-w)y.$$

Three points are worth mentioning. Firstly, uncertainty about α is not fully taken into account. Secondly, $p^*(\theta|y)$ is not the posterior density, because the conditional prior mean has been replaced by the sample average. Thus, coherency is violated despite the use of Bayes formula at the initial step of this statistical procedure (which can be called quasi-Bayesian). Thirdly, within the sampling theory approach, $\hat{\theta}_{EB}$ is a natural point estimate of the vector of random effects. It has the "shrinking" property, since the measurement average y_i corresponding to θ_i is "shrunk" towards the overall average of measurements. So, in $\hat{\theta}_{EB}$ all observations are used to estimate θ_i, not only observations related to this particular effect. While incoherence is a crucial deficiency from the Bayesian point of view, shrinkage estimators have interesting sampling properties. It is therefore important to provide Bayesian interpretation and justification of conducting inference on the basis of $p^*(\theta|y)$.

We seek for the sampling density $\tilde{p}(y|\theta)$ and the prior density $\tilde{p}(\theta)$ that lead to the Bayesian model $\tilde{p}(y,\theta) = \tilde{p}(y|\theta)\tilde{p}(\theta)$ characterised by the joint density of the form

$$\begin{aligned}\tilde{p}(y,\theta) &= p(y|\theta)p(\theta|\alpha=\hat{\alpha}) \\ &\propto g(y;\theta)f_0(\theta;\hat{\alpha}) = f_N^n(y|\theta, cI_n)f_N^n(\theta|\bar{y}e_n, dI_n)\end{aligned}$$

which results in

$$p^*(\theta|y) = f_1(\theta|y,\alpha=\hat{\alpha}) \propto g(y;\theta)f_0(\theta;\hat{\alpha})$$

as the true posterior $\tilde{p}(\theta|y)$. Elementary calculations show that

$$\tilde{p}(y,\theta) \propto f_N^n\left(y|\theta, \left(\frac{1}{c}I_n + \frac{1}{dn}e_ne_n'\right)^{-1}\right)\exp\left(-\frac{1}{2d}\theta'M\theta\right),$$

$$\tilde{p}(\theta) = \int_Y \tilde{p}(y,\theta)\,dy \propto \exp\left(-\frac{1}{2d}\theta'M\theta\right), \quad M = I_n - \frac{1}{n}e_ne_n',$$

$$\tilde{p}(y|\theta) = \frac{\tilde{p}(y,\theta)}{\tilde{p}(\theta)} = f_N^n\left(y|\theta, c\left(I_n - \frac{c}{n(c+d)}e_ne_n'\right)\right).$$

Since M is an idempotent singular matrix, the true prior is improper, but σ-finite. It is informative, as it favours approximate equality $\theta_1 \approx ... \approx \theta_n$. In fact, this prior deserves more attention. Consider the non-singular linear transformation of θ into $(\bar{\theta}, \eta)$, where $\bar{\theta} = e_n'\theta/n$ and $\eta_i = \theta_i - \bar{\theta}$ $(i = 1, \ldots, n-1)$. Since $\theta' M \theta = \eta'(I_{n-1} + e_{n-1}e_{n-1}')\eta$, $\tilde{p}(\theta)$ leads to $\tilde{p}(\bar{\theta}, \eta) = \tilde{p}(\bar{\theta})\tilde{p}(\eta)$ with $\tilde{p}(\bar{\theta})$ constant and $\tilde{p}(\eta) = f_N^{n-1}\left(\eta|0, d\left(I_{n-1} - \frac{1}{n}e_{n-1}e_{n-1}'\right)^{-1}\right)$. The prior of $\bar{\theta} \in \mathbb{R}$ is improper uniform, but η, the vector of $n-1$ deviations $\theta_i - \bar{\theta}$, is *a priori* normally distributed around 0. The sampling density $\tilde{p}(y|\theta)$ is different from $p(y|\theta)$. The true conditional distribution is normal, like the initially declared one, but it assumes that the observations are equally correlated (instead of being independent). The true sampling covariance matrix leads to the same correlation coefficient for each pair of observations:

$$\widetilde{Corr}(y_i, y_j|\theta) = -\frac{c}{(n-1)c + nd} \quad (i \neq j),$$

which tends to zero when n increases; $\tilde{p}(y|\theta)$ practically coincides with $p(y|\theta)$ when n is sufficiently large. However, we cannot use the standard Bayesian asymptotic argument to say that the prior does not matter when n is large, because θ is of dimension n, which is not fixed. Thus the full Bayesian justification of $p^*(\theta|y)$, even asymptotic, requires considering the prior specification as well. Intuitively, $\tilde{p}(\theta)$ is a very reasonable prior. It explains shrinking through giving equal random effects the highest prior chance without introducing any prior information about the average value of all n random effects. In order to show how such prior distribution of individual effects can appear within the fully Bayesian approach, we now consider the Bayesian normal hierarchical model, which introduces one more level – the normal prior distribution of α.

If we assume that

$$p(y|\theta) = f_N^n(y|\theta, cI_n),\, p(\theta|\alpha) = f_N^n(\theta|\alpha e_n, dI_n),\, p(\alpha) = f_N^1(\alpha|a, v),$$

we can write

$$p(\theta) = \int_{-\infty}^{+\infty} p(\theta|\alpha)p(\alpha)\, d\alpha = f_N^n(\theta|ae_n, dI_n + ve_ne_n'),$$

$$p(\theta|y) \propto p(y|\theta)p(\theta) = f_N^n(y|\theta, cI_n)p(\theta)$$

or, equivalently,

$$p(\theta|y) = \int_{-\infty}^{+\infty} p(\theta|y, \alpha)p(\alpha|y)\, d\alpha = \int_{-\infty}^{+\infty} f_1(\theta|y; \alpha)p(\alpha|y)\, d\alpha,$$

where

$$p(\alpha|y) = f_N^1\left(\alpha\left|\left(\frac{n}{c+d} + \frac{1}{v}\right)^{-1}\left(\frac{n}{c+d}\bar{y} + \frac{a}{v}\right), \left(\frac{n}{c+d} + \frac{1}{v}\right)^{-1}\right.\right).$$

Since we use the Bayesian approach, where all unknown elements of the statistical model are random variables, it is important to distinguish fixed and random (individual) effects in a Bayesian sense. We use the definition of Koop, Osiewalski and Steel (1997), who call effects *fixed* if they are marginally independent and *random* if they are not. In our example effects θ_i are only conditionally independent (given α), but they are marginally dependent – as it is clear from the non-diagonal covariance matrix of the marginal prior $p(\theta)$. This definition can be extended to σ-finite measures (improper priors). If $p(\theta) = p(\theta_1)p(\theta_2) \dots p(\theta_n)$, then θ_i are fixed effects, otherwise they are random. Note that $\tilde{p}(\theta) \propto \exp\left(-\frac{1}{2d}\theta' M\theta\right)$ cannot be presented as a product of σ-finite measures of individual θ_i, therefore this improper prior describes random effects – similarly as the proper prior $p(\theta)$ does.

Finally, we obtain the marginal posterior density of random effects:

$$p(\theta|y) = f_N^n\left(\theta|(1-w)y + w\left(\frac{n}{c+d} + \frac{1}{\nu}\right)^{-1}\left(\frac{n}{c+d}\bar{y} + \frac{a}{\nu}\right)e_n,\right.$$
$$\left.\frac{cd}{c+d}I_n + w^2\left(\frac{n}{c+d} + \frac{1}{\nu}\right)^{-1} e_n e_n'\right).$$

If $\nu^{-1} = 0$, then $p(\alpha) \approx const$, the marginal prior of θ is $p(\theta) \approx \tilde{p}(\theta)$ and the posterior is:

$$p(\theta|y) \approx f_N^n\left(\theta|\hat{\theta}_{EB}, \frac{c}{c+d}\left(dI_n + \frac{c}{n}e_n e_n'\right)\right).$$

Note that the flat prior of α leads to the informative improper marginal prior $\tilde{p}(\theta)$ obtained earlier. However, the above presented posterior $p(\theta|y)$ has a different covariance matrix than $\tilde{p}(\theta|y) = p^*(\theta|y) = f_N^n\left(\theta|\hat{\theta}_{EB}, \frac{cd}{c+d}I_n\right)$; the additional term $\frac{c^2}{n(c+d)}e_n e_n'$ in the posterior covariance matrix reflects uncertainty about α. If both ν and n are large enough, then our fully Bayesian and quasi-Bayesian posteriors approximately coincide: $p(\theta|y) \approx p^*(\theta|y)$ and, thus, small-sample incoherence does not matter.

4. Concluding remarks

In order to somehow validate the use of data-based "priors" (met in incoherent, quasi-Bayesian approaches), a formal method has been proposed in this paper. It amounts to defining and examining the Bayesian model that coherently generates the same posterior distribution as the "posterior" obtained in the original sampling model coupled with the data-based "prior". Although the method is general, only in simple cases it can lead to the closed forms of the true sampling model and the true prior. Therefore, our example includes a normal sampling model with a normal prior structure, always with known variance. The unknown (estimated) prior variance case, fundamental for the parametric empirical Bayes, is left for future research.

As a by-product of our research, we have obtained an important prior structure that favours parameters' equality. This prior is completely uninformative (improper uniform) about the parameters' average, but it is quite informative (proper normal) about the deviations from the average.

Acknowledgement

The author acknowledges support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39, 83–87.

Efron, B. & Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators – Part II: the empirical Bayes case. *Journal of the American Statistical Association*, 67, 130–139.

Greene, W.H. (2008). *Econometric Analysis* (Sixth Edition), Pearson, Upper Saddle River NJ.

Koop, G., Osiewalski, J. & Steel, M.F.J (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, 76, 77–105.

Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications (with discussion). *Journal of the American Statistical Association*, 78, 47–65.