SEMI BAYESIAN INFERENCE HIGH AND LOW DIMENSIONAL DATA WITH MULTICOLLINEARITY

OLOYEDE, I.
Department of Statistics, University of Ilorin, Ilorin, Nigeria
E-mail: royalinkonsult@gmail.com **Phone No:** +234-805-304-9890

Abstract
It is generally known that correlation amongst features in high and low dimensional data lead to parameters that artificially insignificance. This study investigates asymptotic properties of some semi-bayesian estimators and compared it with non-bayesian estimator in the presence of multicollinearity. Variational and Empirical Bayes estimators were succinctly compared with ordinary least squares estimator using bias, mean squares error (MSE) and predictive mean squares error (PMSE). The number of iteration was 1000. In high dimensional data, it was found that empirical Bayes Linear Regression (EBLR) outperformed other estimators whereas OLS performed poorly using the PMSE as evaluation criterion. The study found out that in low dimensional data, variational Bayes Linear Regression (VBLR) outperformed other estimators yet OLS performed poorly using the PMSE criterion. Asymptotically, the three estimators were inconsistent but having the same pattern in low dimensional data but they were fairly consistent between the sample sizes 30 to 50 using the bias criterion. The study therefore concluded that empirical Bayes estimator should be adopted in high dimensional data while variational Bayes should be adopted in low dimensional data.
Keywords: variational Bayes, empirical Bayes, high and low dimensional data.

Introduction
It is common knowledge that correlation among the features variables in linear regression(LR) will affect the precision of the estimates, possibly leading to parameters estimates that are artificially statistically insignificant. Barley (1980). This is as a result of inherent instability of inverting a near/non-singular matrix. One common method is to estimate the variance of $\hat{\theta}$ in form of diagonal element of

$$var(\hat{\theta}) = \sigma^2[j'j]^{-1} = \sigma^2\left[\frac{\partial}{\partial\theta_1}, ..., \frac{\partial}{\partial\theta_n}\right]'\left[\frac{\partial}{\partial\theta_1}, ..., \frac{\partial}{\partial\theta_n}\right]^{-1} (1)$$

where j is the gradient matrix, collinearity arises from two sources namely: model and data based collinearity. Model based collinearity arises when columns of j are correlated with each other whereas data based collinearity resulted from the high correlation among the feature variables. Yngve and Erik (2008) examined linear regression with unknown model order selection, their study assumed zero-mean Gaussian noise and distributed coefficient vector. They derived empirical Bayes of maximum a posterior probability model order selection and concluded that their approach outperformed the conventional Akaike Information Criteria and Bayesian Information Criteria.

Empirical Bayes(EB) has taken many forms since its inception from the work of Von (1940) where Robbins extended it with the formulation of nonparametric approach, and George(1985) examined it in view of parametric approach, Empirical Bayes has been applied to various fields of studies such as fire arm probability, Carter and Rolph (1974), and law school admission Robins (1981). In his study, Morris(1983) took Empirical Bayes in to another dimension by formulating parametric Empirical Bayes theory. Hoadley(1981) used EB to solve the problem of quality assurance, revenue sharing issue was examined by the duo of Fay and Herriot (1979) with EB. George(1985) applied parametric EB on assessment of consumer intent and balting average. Morris(1983) reviewed the multi-parameter shrinkage estimators by adopting EB with the consideration of parametric prior distribution. He concluded that EB modeling allowed Statisticians to incorporate into the problem to be solved. This is achieved by viewing the parameter as another stochastic processes with restricted class of distribution.

George(2001) adopted empirical Bayes via Gibbs sampler to depict how hyper-parameters in hierarchical model could be estimated in a way that is computational efficient and statistical valid. He stressed that it should not be taken as an alternative estimator but may be used especially when the full hierarchical cannot be specified. Robin (1981) introduced empirical Bayes as estimator that stands between frequentist and Bayesian approach. Singh(1983) considered EB approach to the square error loss estimation problem in a multiple linear regression. Harvey and Richard(1969)obtained an exact empirical Bayes estimator for the vector β in the general linear regression model. They proposed exact empirical Bayes to proffer solution since MLE could not independently sufficient for the model parameter estimate. Beal(2003) suggested that the use of variational Bayes (VB) technique in high dimensional data which may be computational intensive and costly when adopting Monte Carlo sampling scheme. VB is indirectly optimize an approximation to both model evidence and posterior density. Bishop(2006) adopted VB with automatic relevance determination by assigning an individual hyper-prior to each regression coefficients. This was equally adopted by Jan(2017). Jo-Anne *et-al*(2008) adopted variational Bayes least squares (VBLS)as an application to brain machine interface data and concluded that the real time of VB algorithm demonstrated the suitability for real time interface for both brain and machine. VBLS achieved comparable performance with the baseline study, they opined that VBLS is an iterative statistical method that performed slower than classical one-shot linear least squares technique.

They concluded that VBLS offered a new technique of high statistical robustness. They claimed that VBLS affects the quality of function approximation and might be towards overfitting. The purpose of this study is to extend further the parametric Empirical Bayes in an attempt to estimate the model parameters via Gibbs sampler and made comparison with variational Bayes and ordinary least squares.

Material and Methods

Let $y = X\beta + e$ be linear regression equation with $\boldsymbol{y}$ is an $\boldsymbol{n} \times \mathbf{1}$ vector of observations that is stochastic, X is $n \times p$ matrix of feature variables that is non-stochastic. β is $p \times \mathbf{1}$ vector of unknown parameters with $\boldsymbol{e}$ as residual error. Thus $\boldsymbol{y}|\beta \sim (X\beta;\, (\sigma^2 + \tau^2)I)$ while$e|\beta \sim N(\mathbf{0};\, \sigma_i^2\boldsymbol{\Omega})$**.** The empirical bayes approach is defined by specifying (1) in hierarchical form:

$y_i|\theta_i \sim N(\theta_i;\, \sigma^2)$

(1)

$\theta_i\beta_i \sim N(\beta X_i;\, \tau^2)$

(2)

the likelihood probability density function is of the form:

$$f(\beta, \sigma^2|X, y) = \frac{1}{\sqrt{(2\pi(\sigma^2 + \tau^2)}} e - \frac{1}{2(\sigma^2 + \tau^2)}(y - X\beta)'(y - X\beta) \qquad (3)$$

The cumulative distribution function is derived as

$$F(\beta,\sigma^2|X,y)=\prod_{i=1}^{n}(2\pi(\sigma^2+\tau^2))^{-\frac{n}{2}}e-\frac{1}{2\pi(\sigma^2+\tau^2)}(y-X\beta)'(y-X\beta) \quad (4)$$

The density of normal prior is

$$\Pi(\beta|y)=\frac{1}{2\pi\sigma^2}e^{-}\frac{y^2}{2\sigma^2} \quad (5)$$

The joint probability

$$\Pi(\beta,\sigma^2|X,y)=\prod_{i=1}^{n}(2\pi(\sigma^2+\tau^2))^{-\frac{n}{2}}e-\frac{1}{2\pi(\sigma^2+\tau^2)}(y-X\beta)'(y-X\beta)\times\frac{1}{2\pi\sigma^2}e^{-}\frac{y^2}{2\sigma^2} \quad \textit{\textbf{(6)}}$$

Obtaining the mean of Empirical Bayes as:

$$\sigma^2(\frac{\partial log(m(X))}{\partial y}-\frac{\partial log(h(X))}{\partial y}) \quad (7)$$

Take log of both sides

$$=-\frac{n}{2}ln(2\pi)-\frac{n}{2}ln(\sigma^2+\tau^2)-\frac{1}{2(\sigma^2+\tau^2)}(y-X\beta)'(y-X\beta)+\frac{1}{2}ln(2\pi)-\frac{1}{2}ln\sigma^2-\frac{1}{2\sigma^2}y^2 \quad (8)$$

Differentiate eq (6) with respect to ***y***

$$\frac{\partial l}{\partial y}=\frac{-1}{2(\sigma^2+\tau^2)}2(y-X\beta)-\frac{1}{2\sigma^2}2y \quad (9)$$

follow (6) we have

$$\sigma^2(-\frac{(y-X\beta)}{\sigma^2+\tau^2}+\frac{y}{\sigma^2}) \quad (10)$$

$$\frac{-y\sigma^2+X\beta\sigma^2+y(\sigma^2+\tau^2)}{\sigma^2+\tau^2} \quad (11)$$

$$\frac{-y\sigma^2+X\beta\sigma^2+y\sigma^2+y\tau^2}{\sigma^2+\tau^2} \quad (12)$$

$$\frac{X\beta\sigma^2+y\tau^2}{\sigma^2+\tau^2} \quad (13)$$

$$\theta_i(\beta_i)=\frac{\sigma^2}{\sigma^2+\tau^2}X\beta+\frac{\tau^2}{\sigma^2+\tau^2}y \quad (14)$$

Obtain the mean of EB

$$\frac{\sigma^2}{\sigma^2+\tau^2}X(X'X)^{-1}X'y+\frac{\tau^2}{\sigma^2+\tau^2}y \quad (15)$$

$$\frac{1}{\sigma^2+\tau^2}[X(X'X)^{-1}X'\sigma^2+\tau^2]y \quad (16)$$

represent idempotent matrix $(\boldsymbol{X'X})^{-1}\boldsymbol{X'}$by H

$$\frac{1}{\sigma^2+\tau^2}[H\sigma^2+\tau^2]y \quad (17)$$

$$\frac{\tau^2}{\sigma^2+\tau^2}[H\frac{\sigma^2}{\tau^2}+1]y \quad (18)$$

In Variational Bayes: Let $\boldsymbol{\beta}, \boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ be the coefficients of interest, precision and noise respectively due to availability of analytical posterior $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}|\boldsymbol{y}, \boldsymbol{X})$, variational approximation approach $\boldsymbol{q}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \boldsymbol{q}_{\boldsymbol{\beta}}(\boldsymbol{\beta}), \boldsymbol{q}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}), \boldsymbol{q}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda})$ is adopted.
Thus

$$p(\alpha) = Ga(\alpha|a_o, b_o) \quad (19)$$

$$p(\beta|\alpha) = N_d(\beta|0, \alpha^{-1}|d \times d) \quad (20)$$

$$p(e) = Ga(e|c_o, d_o) \quad (21)$$

$$p(y, X|\beta) = N(y, X|\beta' X, e^{-1}) \quad (22)$$

Data Analysis, Results and Interpretation
This section focuses on the analysis of low and high dimensional data considering the three estimators.

Table 1 showing Comparison of Variational Bayes, OLS and Empirical Bayes Based on Absolute Bias in Different Sample Sizes

samples	estimators	β_1	β_2	β_3	β_4	β_5	β_6
	ols	0.1372	0.4271	0.0344	0.0658	0.2832	0.0350
20	vblr	0.1198	0.4563	0.0011	0.0285	0.3236	0.0073
	eblr	0.1248	0.4452	0.0067	0.0250	0.3130	0.0118
	ols	0.5722	0.1246	0.0644	0.0033	0.0214	0.7102
30	vblr	0.5618	0.1293	0.0195	0.0185	0.0109	0.7094
	eblr	0.5718	0.1230	0.0386	0.0072	0.0084	0.7131
	ols	0.0158	0.0429	0.2420	0.5879	0.0405	0.5715
50	vblr	0.0019	0.0088	0.2481	0.5869	0.0124	0.5370
	eblr	0.0105	0.0312	0.2427	0.5868	0.0290	0.5589
	ols	0.0228	0.8268	0.8334	0.0837	0.0315	0.0173
100	vblr	0.0150	0.8259	0.8180	0.0730	0.0189	0.0080
	eblr	0.0208	0.8260	0.8313	0.0821	0.0286	0.0149
	ols	0.0069	0.0107	0.4620	0.4213	0.6050	0.2099
200	vblr	0.0025	0.0037	0.4687	0.4173	0.6126	0.2133
	eblr	0.0065	0.0100	0.4628	0.4211	0.6052	0.2108
	ols	0.0143	0.1564	0.0107	0.0031	0.5733	0.0089
500	vblr	0.0109	0.1556	0.0123	0.0009	0.5766	0.0070
	eblr	0.0140	0.1565	0.0105	0.0030	0.5736	0.0087
	ols	0.5377	0.1913	0.0074	0.0156	0.1349	0.0028
1000	vblr	0.5353	0.1902	0.0074	0.0137	0.1362	0.0016
	eblr	0.5376	0.1914	0.0073	0.0155	0.1350	0.0028

Both small and large sample sizes were considered in the study in order to verify the properties of consistency and efficiency in low dimensional datasets, the sample observation varies from 20 to 1000. It was observed in $\hat{\boldsymbol{\beta}}_1$ parameter that VBLS outperformed other estimators across all the sample sizes with minimum bias, in $\hat{\boldsymbol{\beta}}_2$, it equally outperformed other estimators with minimum bias from sample sizes 50 to 1000, in $\hat{\boldsymbol{\beta}}_3$ the three estimators shared the performances. In $\hat{\boldsymbol{\beta}}_4$, the study found out that VBLS outperformed other estimators with minimum bias from sample sizes 100 to 1000 so also in b6 where it outperformed in sample sizes 20 to 100. Thus VBLS has the overall performances.

Asymptotic Consistency(Absolute Bias) of Variational Bayes, OLS and Empirical. The study succinctly observed similar pattern amongst the three estimators deemed with respect to their asymptotic consistency (absolute bias), considering $\hat{\boldsymbol{\beta}}_1$ parameter, absolute bias increased from sample size 20 to 30, at sample size 50 the study observed decrease absolute bias and thus increased at sample size 100, decreased at sample size 200, as the sample sizes increases to 500, the study observed an increase in absolute bias up to sample size 1000. Considering $\hat{\boldsymbol{\beta}}_2$ parameter, absolute bias decreased from sample size 20 to 50, at sample size 100 the study observed increase and decrease at sample size 200, therefore increases at sample size 500, as the sample sizes increases to 500, the study observed an increase in absolute bias and as well at sample size 1000. Considering $\hat{\boldsymbol{\beta}}_3$ parameter, absolute bias increased from sample size 20 to 100, at sample size 200 the study observed decreases and continue to decrease as sample size increases from 500 to 1000 Considering $\hat{\boldsymbol{\beta}}_4$ parameter, absolute bias decreases from sample size 20 to 30, at sample size 50 the study observed increases and decreases at sample size 100, therefore increases at sample size 200, so the absolute bias kept on decreasing and increasing as the sample sizes increases from 500 to 1000. Considering $\widehat{\boldsymbol{\beta}}_5$ parameter, the absolute bias decreased from sample size 20 to 30, at sample size 50 the study observed increase and decrease at sample size 100, therefore increases at sample size 200, as the sample sizes increases from 500 to 1000, and the study observed a decrease in absolute bias of all the three estimators. Considering $\widehat{\boldsymbol{\beta}}_6$ parameter, the absolute bias increases as the sample sizes increases from 20 to 30. The study also observed a decrease as the sample sizes increases from 50 to 100. Thereafter, increases at sample size 200 but further decreases from sample sizes 500 to 1000. It can be concluded from the study that all the three estimators were in consistent due to the presence of multicollinearity.

Table 2: Showing Comparison of VBLR, EBLR and OLS based on Mean Squares Error criterion

samples	estimators	β_1	β_2	β_3	β_4	β_5	β_6
	ols	1.5493	1.7608	1.5645	1.4493	1.5301	1.5579
20	vblr	0.2229	0.4346	0.1939	0.2093	0.2881	0.2300
	eblr	0.3287	0.5344	0.3081	0.3207	0.3947	0.3292
	ols	1.1612	0.8345	0.8273	0.9077	0.8575	1.2475
30	vblr	0.4509	0.1604	0.1380	0.1318	0.1467	0.6182
	eblr	0.6444	0.3211	0.3123	0.3383	0.3327	0.7828
	ols	0.4887	0.5324	0.5569	0.8657	0.4893	0.7700
50	vblr	0.1005	0.1140	0.1644	0.4642	0.1045	0.3778
	eblr	0.2851	0.3095	0.3494	0.6486	0.2850	0.5712
	ols	0.2914	0.9873	1.0018	0.2895	0.2805	0.3149
100	vblr	0.0942	0.7742	0.7759	0.0905	0.0867	0.1068
	eblr	0.2288	0.9199	0.9337	0.2283	0.2200	0.2490
	ols	0.1980	0.2114	0.4366	0.4041	0.5828	0.2700
200	vblr	0.0961	0.0979	0.3288	0.2859	0.4794	0.1568
	eblr	0.1781	0.1899	0.4146	0.3811	0.5611	0.2480
	ols	0.1583	0.1915	0.1616	0.1587	0.5009	0.1508
500	vblr	0.1086	0.1368	0.1111	0.1094	0.4512	0.1033
	eblr	0.1524	0.1853	0.1556	0.1529	0.4949	0.1452
	ols	0.4375	0.1639	0.1313	0.1572	0.1769	0.1498
1000	vblr	0.4078	0.1389	0.1069	0.1285	0.1491	0.1215
	eblr	0.4348	0.1618	0.1290	0.1545	0.1742	0.1473

From table 2 above, three estimators were compared in a semi Bayesian paradigm, the study found out that VBLR outperformed other estimators with minimum MSE in all the sample sizes considered in the study.

Asymptotic Efficiency (Mean Squares Error) of Variational Bayes, OLS and Empirical Bayes Estimators

The study observed from the $\widehat{\boldsymbol{\beta}}_1$ parameter that EBLR was efficient between the sample sizes 30 and 500, but the MSE misbehaves as sample size increases to 1000; OLS depicted efficiency between the sample sizes 20 and 500 and equally misbehaved at sample size 1000; VBLR was efficient between sample sizes 30 and 100 and suddenly change at sample size 200, therefore kept increasing until sample size 1000. The study observed from the $\widehat{\boldsymbol{\beta}}_2$ parameter that EBLR was efficient between the sample sizes 20 and 50, but the MSE misbehaves at sample size 100, but it's efficient at sample sizes 100 and 1000; OLS showed efficiency between the sample sizes 20 and 50 and equally misbehave at sample size 100. But it's efficient at sample

sizes 100 and 1000; VBLR was efficient between sample sizes 20 and 50 and suddenly change at sample size 100, but it's inefficient as sample sizes increases 100 and 1000.
The study observed from the $\widehat{\boldsymbol{\beta}}_3$ parameter that EBLR was inefficient between the sample sizes 20 and 50, but the MSE misbehaves at sample size 100, but it's efficient at sample sizes 100 and 1000 ;OLS showed efficiency between the sample sizes 20 and 50 and equally misbehave at sample size 100. But it's efficient at sample sizes 100 and 1000; VBLR was inefficient between sample sizes 20 and 100 and suddenly change at sample size 100, but it's efficient as sample sizes increases from 200 to 1000.

The study observed from the $\widehat{\boldsymbol{\beta}}_4$ parameter that EBLR was inefficient between the sample sizes 20 and 1000; OLS showed efficiency between the sample sizes 20 and 100 and equally misbehave at sample size 200; VBLR shows inefficiency between the sample sizes 20 and 1000.
The study observed from the $\widehat{\boldsymbol{\beta}}_5$ parameter that EBLR was efficient between the sample sizes 20 and 100, but the MSE misbehaves at sample size 200, but it's efficient as sample size increases from 500 and 1000; OLS showed efficiency between the sample sizes 20 and 100 and equally misbehave at sample size 200. But it's efficient as sample size increases from 500 and 1000; VBLR was efficient between sample sizes 20 and 100 and suddenly change at sample size 200, but kept decreasing until sample size 1000. The study observed from the $\widehat{\boldsymbol{\beta}}_6$ parameter that EBLR was inefficient between the sample sizes 20 and 1000; OLS depicted efficiency between the sample sizes 20 and 1000; VBLR portrayed inefficiency between the sample sizes 20 and 1000.

Comparison of Variational Bayes, OLS and Empirical Bayes Based on Predictive Mean Squares Error on Low and High Dimensional Data

Table 3 showing results from high dimensional data

Samples	Est	Test	train
	ols	1.7492	0.5593
20	vblr	1.1697	0.7717
	eblr	1.1890	0.6562
	ols	1.4080	0.7152
30	vblr	1.1283	0.8639
	eblr	1.1785	0.7534
	ols	1.1985	0.8296
50	vblr	1.0724	0.9174
	eblr	1.1215	0.8403
	ols	1.1004	0.9078
100	vblr	1.0528	0.9477
	eblr	1.0814	0.9095
	ols	1.0492	0.9530
200	vblr	1.0348	0.9697
	eblr	1.0451	0.9533
	ols	1.0096	0.9832
500	vblr	1.0074	0.9875
	eblr	1.0091	0.9832
	ols	1.0105	0.9931
1000	vblr	1.0099	0.9945
	eblr	1.0103	0.9931

Table 4 showing results from low dimensional dataset with Predictive MSE

Samples	Est	Test	Train
	OLS	1.4536	4.3141e-30
$20_{1}00$	VBLR	1.2435	1.1026
	EBLR	1.2434	0.8191
	OLS	1.8067	6.2025e-30
$50_{1}00$	VBLR	1.2292	1.1724
	EBLR	1.2182	0.6671
	OLS	4.9634	1.1962e-29
$100_{1}00$	VBLR	1.1869	1.1772
	EBLR	1.1915	0.5624
	OLS	1.8222	7.8593e-30
$100_{2}00$	VBLR	1.2406	1.1762
	EBLR	1.2298	0.6666
	OLS	1.3627	8.0259e-30
$100_{5}00$	VBLR	1.2125	1.1951
	EBLR	1.2053	0.8945
	OLS	1.2669	9.3884e-30
$100_{1}000$	VBLR	1.1971	1.1839
	EBLR	1.1951	1.0136

Thus n-samples is the first samples while n-features is the second samples. The study observed the behavioural pattern by the three estimators (OLS, EBLR and VBLR) deemed in the study. The pattern was observed based in the high dimensional data where the number of features (independent variables) were more than observations size categories were considered, variational Bayes predicted well and outperformed other estimators with minimum predictive mean square error for testing dataset when the visual expression of this estimators are depicted in the presented in the appendix A. Empirical Bayes performed when sample observations were 20, 50 with 100 features. Variational Bayes outperformed when the sample observation was set at 100, with 100 features. 100 observations with 200, 500 and 1000 features. OLS predicted poor in testing dataset throughout the sample observation considered.

Conclusion

The study compared three algorithm (Ordinary Least Square, Variational Bayes and Empirical Bayes) where there is presence of multicollinearity both at low dimensional and high dimensional data. Asymptotic behavioural pattern of each of the estimators were equally observed. The study therefore concluded that there is similar patterns amongst the three estimators asymptotically for all the parameter estimates, as the sample size increases the asymptotic properties of consistency and efficiency subsist. Our approach can be applied to further studies in the area of simultaneous equation and other econometric models.

References

Beal, M. J. (2003). *Variational algorithms for approximate bayesian inference.* citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.9951.

Bishop C. M. (2006). *Pattern recognition and machine learning.* New York: Springer,

Carter, G., & Rolph, J. E. (1974). Empirical bayes methods applied to estimating fire alarm Probabilities. *Journal of the American Statistical Association*, 69,880-885.

Fay, R. E., & Herriot, R. A. (1979). Estimate income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

George, C. (1985). An introduction to empirical Bayes data analysis in the American Statistician, 39(2), 83-87. http://www.jstoriorg/stable/2682801.

George, C. (1992). Illustration empirical bayes methods, chemometrical and intelligent laboratory systems, 16, 107-125, Elsevier science publisher B.V. Amsterdon.

George, C. (2001). Empirical bayes gibbs sampling. *Biostatistics,* 2(4), 485- 500.

Harry, F. M., & Richard, G. K. (1969). Empirical bayes estimators in a multiple linear regression model. *Biometrics*, 56(20), 367.

Jan, D. (2017). Variational bayesian inference for linear and logistic regression, arVir: 1310.5438v3 [stat.inc], Department Des Neurosciences fundamentales, universite de Geneve, 1211 Geneve.

Jo-Anne, T., Aaron, D., Kenji, Y., Toshinor, Y., Donna, H., Shinji, K., Lauren, S., John, K., Mitsuo, K., Peter, S., & Stefan, S. (2009). Variational bayesian least squares an application to brain machine interface data. *Neural Network*, 21(8), 1112-1131. da:10.1016/j. neunet

Hoadley, B. (1981). Quality management plan (QMP). *Bell System Technical Journal,* 60, 215-273.

Moris, C. N. (1983). Parametric empirical bayes inference theory and applications. *Journal of American Statistical Association,* 78(381).

Natarian, R., & McCulloch, C. E. (1995). A note on the existence of the posterior distribution for a class of mixed models for binomial responses. *Biometrical,* 82, 639-643.

Robins, H. (1983). Some thoughts on empirical bayes estimation. *Annals of Statistics*, 11, 713-723.

Rubin, D. (1981). Using empirical bayes techniques in the law school validity studies. *Journal of the American Statistical Association,* 75, 801-827.

Singh, R. (1983). Empirical bayes estimation in a multiple linear regression model. *Ann. Inst. Statistics Math,* 37, Part A, 71-86.

Yngve, S., & Erik, G. L. (2008). Empirical bayes linear regression with unknown model order. *Journal Digital Signal Processing Archive,* 18(2), 236-248 Academic Press, Inc. Orlando, FL, USA doi>10.1016/j.dsp.2007.03.005.