Bayesian method for solving the problem of multicollinearity in regression

Article · October 2018

DOI: 10.16929/as/1823.135

Adedayo Adepoju
University of Ibadan

Oluwadare O. Ojo
Federal University of Technology, Akure

Afrika Statistika

Afrika Statistika
Vol. 13 (3), 2018, pages 1823 – 1834.
DOI: http://dx.doi.org/10.16929/as/1823.135

ISSN 2316-090X

Bayesian method for solving the problem of multicollinearity in regression

Adedayo A. Adepoju[1,*] **and Oluwadare O. Ojo**[2]

[1]Department of Statistics, University of Ibadan, Nigeria
[2]Department of Statistics, Federal University of Technology Akure, Nigeria

Received : July 13, 2018. Accepted : November 22, 2018

Abstract. . The popular method of estimation in regression, Ordinary Least Squares (OLS) often displays inefficiency especially with large variances and wide confidence intervals thereby making precise estimate difficult when there is strong multicollinearity. Bayesian method of estimation is expected to improve the efficiency of estimated regression model when there is relevant prior information and belief of situation being modelled is available. This study however provided an alternative approach to OLS when there is almost perfect multicollinearity while its performance were compared with the aid of simulation approach to OLS estimator. Results of the simulation study indicate that with respect to Mean Squared Error (MSE) criterion and other criteria, the proposed method perform better than OLS.

Key words: Multicollinearity, Regression, Standard Error, Simulation.
AMS 2010 Mathematics Subject Classification : 62F15, 62GO5, 62H10.

---

*Corresponding author Adedayo A. Adepoju: pojuday@yahoo.com
Oluwadare O. Ojo : ojooo@futa.edu.ng

Résumé. (French) La méthode d'estimation par régression populaire, la méthode des moindres carrés ordinaires (MCO), est souvent peu efficace, en particulier lorsque les variances sont grandes et que les intervalles de confiance sont larges, ce qui rend difficile une estimation précise lorsque la multicolinéarité est forte. La méthode d'estimation bayésienne devrait améliorer l'efficacité du modèle de régression estimé lorsqu'il existe des informations préalables pertinentes et lorsque l'on est convaincu de la situation à modéliser. Cette étude a toutefois fourni une approche alternative à la méthode des moindres carrés ordinaires lorsqu'il y a une multicolinéarité presque parfaite, tandis que ses performances ont été comparées à l'aide de l'approche de simulation de l'estimateur MCO. Les résultats de l'étude de simulation indiquent qu'en ce qui concerne le critère d'erreur quadratique moyenne (MSE) et d'autres critères, la méthode proposée donne de meilleurs résultats que la méthode MCO.

1. Introduction

Multicollinearity is a violation of assumption of regression model. It occurs when the regressors are correlated. This violation can be a serious problem when there is a near perfect correlation, in the sense that regression coefficients of X variables although may be determinate but possess large standard errors, which means that parameters cannot be estimated with great precision. If the correlation between the regressors is perfect, the parameters of the regression model can be indeterminate while the standard errors are infinite Gujarati (1995), Belsley *et al.* (1980)

Some of the solutions in literature to multicollinearity are addition of new data, transforming of variables using suitable transformations, the method of principal component by the reducing the number of regressors Jeffers (1967), Jollife (1972), Mansfield *et al.* (1977) and Miller (1990). Other solution to the problem of multicollineairty, is the use of ridge estimator by Hoerl and Kennard (1970), Duzan and Shariff (2015)and Iguernane (2016), but all the methods are classical methods and Dreeze (1962) argued that classical inferences have shortcomings in that; the available information on parameters is ignored.

However, the use of Bayesian estimation method to solve the problem of multicollinearity in regression model is not common due to its complexity in terms of computation and prior information. Recently, some Bayesian works on multicollinearity in regression are Curtis and Ghosh (2011) and Ijarchelo *et al.* (2016).

Curtis and Ghosh (2011) in their work proposed a Bayesian model that accounted for correlation among the predictors by simultaneously performing selection and clustering of the predictors, dirichlet process and variable selection priors were used for regression coefficient while redundant predictors were removed from the models; they concluded that Bayes method proposed did not outperformed all other methods in all situations but often the best in high collinearity. Ijarchelo *et al.* (2016) developed a Bayesian regression procedure for variable selection under collinearity of parameters using a Zellner's g-prior. Their results showed that a strong collinearity may lead to a multimodal posterior distribution over models in which joint summaries are more appropriate than marginal summaries. They concluded that their posterior distribution were not available in closed form and that can make the problem of multicollinearity become computationally challenging.

All the aforementioned Bayesian methods are variable selection method, but as noted by Lee *et al* (2015) on the use of variable selection methods, that if some explanatory variable are throws out in a regression model, others might not have explanatory power on the dependent variable and this may lead to difficulty in assessing the effect of regressors on the dependent variable.

Herein, we propose a Bayesian estimation procedure with the use of an informative prior. The use of proposed method permits easy computation of many posterior features of interest in regression to overcome the problem of multicollinearity.

The structure of the remainder of this paper is as follows. In section 2, the regression model and the method of Ordinary Least Squares (OLS) will be reviewed. Section 3 provides an overview of Bayesian procedure using an informative prior in regression model in the presence of Multicollinearity. Section 4; provide a simulation where numerical studies are conducted. For comparative purposes, the performance of proposed Bayesian estimation procedure is compared to OLS in Section 5. Section 6 concludes.

2. Regression Model and OLS

The Normal Regression model is given by:

$$y = x\theta + \epsilon \tag{1}$$

Where y and x are the observed data on the n x 1 vector of dependent and n x k matrix of explanatory variables of the regression respectively. θ is the k x 1 vector of parameters to be estimated and ϵ is an error term which is normally distributed with mean zero and constant σ^2 and x values are independent of the error term. In order to estimate the parameters in (1), the popular Classical estimator, OLS for estimating the regression parameters is given by:

$$\hat{\theta} = (x'x)^{-1}x'y \tag{2}$$

While the confidence interval can be obtained as:

$$\hat{\theta} \pm t_{1-\alpha/2,N-K}SE(\hat{\theta}) \tag{3}$$

Where

$$SE(\hat{\theta}) = \sqrt{S^2(x'x)^{-1}}$$

And

$$S^2 = \frac{y'y - \hat{\theta}'x'y}{n-k} \tag{4}$$

It could be observed from the equations (3) and (4) that parameter θ heavily depends on x'x.

3. Bayesian Estimation Procedures

In order to ameliorate the problem of multicollinearity, Bayesian method of estimation is given in this section. Bayesian approach can be expressed through the following relationship which can be written as:

$$P(\theta|y) \propto P(\theta)P(y|\theta) \tag{5}$$

$P(y|\theta)$ is the likelihood function.
$P(\theta)$ is the prior density distribution and $P(\theta|y)$) is the posterior distribution
The likelihood is written as follows;

$$P(y|\theta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} exp[-\frac{h}{2}(y - x\theta)'(y - x\theta)] \tag{6}$$

For convenience, it is better to write (6) in terms of Ordinary Least Squares (OLS) estimator:

$$\begin{aligned}
(y - x\theta)'(y - x\theta) &= (y - x\theta + x\hat{\theta} - x\hat{\theta})'(y - x\theta + x\hat{\theta} - x\hat{\theta}) && (7)\\
&= (y - x\hat{\theta})'(y - x\hat{\theta}) + (\hat{\theta} - \theta)'x'x(\hat{\theta} - \theta) && (8)\\
&= SSE + (\hat{\theta} - \theta)'x'x(\hat{\theta} - \theta) && (9)
\end{aligned}$$

Hence, the likelihood is written as:

$$P(y|\theta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} exp[-\frac{h}{2}\{SSE + (\hat{\theta} - \theta)'x'x(\hat{\theta} - \theta)\}] \tag{10}$$

Prior distribution. Priors play a defining role in Bayesian inference which can take any form and are also meant to reflect any information the researcher has before seeing the data. However, it is common to choose particular classes of priors that are easy to interpret or which would make computation easier Koop (2003). Natural conjugate priors typically belong to such class. The likelihood function in (10) suggests a prior in form of Normal distribution for$\theta|h$ and a Gamma distribution for h. The name of such prior which is a product of Gamma and a conditional Normal is called a Normal-Gamma distribution.
Based on the above premise, it follows that:

$$\theta|h \sim N(\theta_0, h^{-1}Q_0) \tag{11}$$

Equation (11) can also be written as:

$$P(\theta|h) = \frac{h^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}|Q_0|^{\frac{1}{2}}}\{exp[-\frac{h}{2}(\theta - \theta_0)'(Q_0)^{-1}(\theta - \theta_0)]\} \tag{12}$$

And also,

$$P(h) = \frac{1}{\Gamma(\frac{v_0}{2})(\frac{2S_0^{-2}}{v_0})^{\frac{v_0}{2}}} h^{\frac{v_0-2}{2}} exp(-\frac{hv_0}{2S_0^{-2}}) \tag{13}$$

Where,

$$\Gamma(\frac{v_0}{2})(\frac{2S_0^{-2}}{v_0})^{\frac{v_0}{2}} \; is\ the\ integrating\ constant.$$

In the distribution of (12) and (13), θ_o denotes the prior mean for parameter θ, Q_0 is the un-scaled variance-covariance matrix for parameterθ, S_0^{-2} is the prior mean of gamma density function for the model precision h and v_0 is the prior degree of freedom of gamma distribution for the model precision h.

Hence, equations (12) and (13), the natural conjugate prior for θ and h can be simply written as:

$$\begin{aligned} P(\theta, h) &= \frac{h^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}|Q_0|^{\frac{1}{2}}}\{exp[-\frac{h}{2}(\theta-\theta_0)'(Q_0)^{-1}(\theta-\theta_0)]\} \\ &\times \frac{1}{\Gamma(\frac{v_0}{2})(\frac{2S_0^{-2}}{v_0})^{\frac{v_0}{2}}} h^{\frac{v_0-2}{2}} exp(-\frac{hv_0}{2S_0^{-2}}) \end{aligned}$$

$$\begin{aligned} P(\theta, h) &= \frac{h^{\frac{v_0+k}{2}}-1}{(2\pi)^{\frac{k}{2}}|Q_0|^{\frac{1}{2}}\Gamma(\frac{v_0}{2})(\frac{2S_0^{-2}}{v_0})^{\frac{v_0}{2}}}\{exp[-\frac{h}{2}(\theta-\theta_0)'(Q_0)^{-1}(\theta-\theta_0) \\ &+\frac{v_0}{S_0^{-2}}]\} \end{aligned} \tag{14}$$

Equation (14) can also be written as:

$$\theta, h \sim NG(\theta_0, Q_0, S_0^{-2}, V_0) \tag{15}$$

Equation (15) above means that the distribution of the prior, $P(\theta, h)$ for θ and h is a multivariate Normal-Gamma.

N.B: The symbol "o" under the parameters are the priors, while symbol represented by $*$ over the parameters indicate the posterior parameters.
Multiplying (10) and (15), gives the joint posterior distribution as:

$$\theta, h|y \sim NG(\theta^*, Q^*, S_0^{-2}, V^*) \tag{16}$$

Since both the prior and posterior distributions are Normal-Gamma, conjugacy was established.

Hence, the hyper-parameters given in (16) are:

$$Q^* = (Q_0^{-1} + x'x)^{-1} \tag{17}$$

$$\theta^* = Q^*(Q_0^{-1}\theta_0 + x'x\hat{\theta}) \tag{18}$$

$$v^* = N + v_0 \tag{19}$$

Equations (17), (18) and (19) are the estimators for un-scaled variance-covariance matrix which is a k x k matrix, posterior mean and degree of freedom of posterior, respectively.
While the Sum of Squares of Error (SSE) and Variance of the error of the model in (1) can also be given respectively as:

$$SSE = (vS^2)_0 + vS^2 + (\hat{\theta} - \theta_0)'[Q_0 + (x'x)^{-1}]^{-1}(\hat{\theta} - \theta_0) \tag{20}$$

$$S^{2\ *} = \frac{(vS^2)_0 + vS^2 + (\bar{\theta} - \theta_0)'[Q_0 + (x'x)^{-1}]^{-1}(\bar{\theta} - \theta_0)}{v^0} \tag{21}$$

In regression modelling, the coefficient on the regressors, θ is usually a primary focus, and a measure of marginal effect of the regressors on the dependent variable. The posterior mean, $E(\theta|y)$ is the point estimate, and $v(\theta)$ is a metric for measuring the uncertainty associated with the point estimate.
Since the interest is on θ, we integrate out h in (16) to obtain the marginal posterior for θ. Applying the rule of probability we have:

$$E(\theta|y) = \int\int \theta P(\theta, h|y)\partial h \partial\theta = \int \theta P(\theta|y)\partial\theta \tag{22}$$

Where,

$$P(\theta|y) = \int P(\theta, h|y)\partial h \tag{23}$$

Hence, equation (23) becomes:

$$P(\theta|h) = \frac{v^{\frac{n}{2}}\Gamma(\frac{v^0+k}{2})}{\pi^{\frac{k}{2}}\Gamma(\frac{v^0}{2})}|S^{2\ *}Q^*|^{-\frac{1}{2}}[v^* + (\theta - \theta^*)'(S^{2\ *}Q^*)^{-1}(\theta - \theta^*)]^{\frac{v^0+K}{2}} \tag{24}$$

Equation (24) follows a t-distribution which can also be written as:

$$\theta|y \sim t(\theta^*, S_0^{-2\ *}, Q^*, V^*) \tag{25}$$

And from the definition of t-distribution, the mean and variance can be obtained as:

$$E(\theta|y) = \theta^* \tag{26}$$

$$v(\theta) = \frac{SSE}{v^0 - 2}Q^* \tag{27}$$

Equation (26) and (27) are mean and variance estimators used to obtain the values for parameter, θ for different degree of multicollinearity. SE (θ^*) is the standard error of Bayesian estimator of θ^* which can also be obtained as:

$$SE(\theta^*) = \sqrt{v(\theta^*)} \tag{28}$$

The Credible interval for estimators of Bayesian in the same way we have confidence interval in the classical is given by:

$$\theta^* \pm t_{1-\alpha/2}v^*SE(\theta^*) \tag{29}$$

4. Simulation and Prior specification

The data experiment is set up using the Data Generating Process (DGP) below:

$$y = 17 + 8.5x_1 + 5.0x_2 + 2.0x_3 + \epsilon \tag{30}$$

The error term, $\epsilon \sim N(0,1)$ and the explanatory variables generate the dependent variable. Since the degree of collinearity among regressors (X's) is of central importance, the works of Alkhamisi *et al.* (2006), Kibria (2003), Kibria and Banik (2016) will be used in generating x's using the following equation:

$$x_{ij} = (1-\rho^2)^{1/2} x^*_{ij} + \rho x_{ip}, \quad i = 1, 2, ..., n, \ j = 1, 2, ..., p$$

Where x^*_{ij} is the independent standard normal pseudo-random numbers, ρ is the correlation between any two x's.

Prior specification

$$v_0 = 4, \qquad V_0 = \begin{pmatrix} 2.4 & 0 & 0 & 0 \\ 0 & 6\times 10^{-7} & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.6 \end{pmatrix}$$

$$S_0^{-2} = 1.5, \qquad \theta_0 = \begin{pmatrix} 15 \\ 10 \\ 5.5 \\ 2.5 \end{pmatrix}$$

5. Results and Discussion

The objective of the work is to compare performances of our approach with OLS and work of based on the strength of multicollinearity. Here, two different degrees of correlation between variables considered are ρ =0.80 and 0.95. The sample sizes selected are N= 30, 200 and 300. The results from Bayesian and OLS methods for different degree of multicollinearity using Standard Error (SE), Confidence/Credible intervals and Mean Squared Error (MSE)as criteria for comparison are presented in this section. In the tables below, estimators are represented as; Bayesian Informative Prior (BIP) for our estimator used in this work and OLS for Ordinary Least Squares. Tables 1-6 show the Standard Error estimates and CI of the estimators while the MSE of the estimators are also reported in Table 7

Table 1: ρ = 0.95 when the sample size, N=30.

Parameters	Estimators	Standard Error	CI
θ_0	OLS	0.6965	(16.5354,19.3985)
	BIP	0.3642	(15.7856, 17.2657)
θ_1	OLS	2.8858	(3.3379,15.2017)
	BIP	0.0009	(9.9981, 10.0019)
θ_2	OLS	4.9174	(-7.8681, 12.3478)
	BIP	0.4546	(4.2310, 6.0787)
θ_3	OLS	5.2168	(-9.7692, 11.6775)
	BIP	0.8081	(-0.5084, 2.7760)

Table 2: ρ = 0.80 when the sample size, N=30.

Parameters	Estimators	Standard Error	CI
θ_0	OLS	0.6793	(15.7968, 18.5894)
	BIP	0.3371	(15.6903, 17.0605)
θ_1	OLS	1.5259	(5.4896, 11.7626)
	BIP	0.0008	(9.9983, 10.0017)
θ_2	OLS	2.4797	(-0.6799, 9.5144)
	BIP	0.4078	(4.4804, 6.1378)
θ_3	OLS	2.6355	(-2.8534, 7.9813)
	BIP	0.7238	(0.3521, 3.2941)

Tables 1 and 2 report the SE and CI for both the Bayesian (BIP) and OLS methods when the degree of multicollinearity are ρ = 0.80 and 0.95 for sample size of 30. The SE of the estimators for parameters when, ρ = 0.80 are bigger than when ρ = 0.95 which means then lower the degree of multicollinearity the better the estimates. Results obtained from the tables 1 show that Bayesian method gives better performances than the OLS estimator having a minimum SE for all the parameters considered. The CI also reveals that the proposed Bayesian method has a narrower CI than the OLS method of estimation.

Table 3: ρ = 0.95 when the sample size, N=200.

Parameters	Estimators	Standard Error	95%CI
θ_0	OLS	0.2174	(16.5865, 17.4247)
	BIP	0.2125	(16.5879, 17.2018)
θ_1	OLS	1.0457	(5.9640, 10.0887)
	BIP	0.0008	(9.9983, 10.0016)
θ_2	OLS	1.5373	(-0.5523, 5.5111)
	BIP	0.3689	(3.7979, 5.2524)
θ_3	OLS	1.5661	(2.1171, 8.2944)
	BIP	0.5015	(-0.6434, 1.3343)

Table 4: ρ = 0.80 when the sample size, N=200.

Parameters	Estimators	Standard Error	CI
θ_0	OLS	0.2050	(16.2912, 17.0996)
	BIP	0.1660	(16.2988, 16.9536)
θ_1	OLS	0.4811	(7.3150, 9.2127)
	BIP	0.0008	(9.9985, 10.0015)
θ_2	OLS	0.7675	(4.2616, 7.2887)
	BIP	0.3240	(4.2634, 5.5410)
θ_3	OLS	0.7885	(0.2960, 3.4062)
	BIP	0.4334	(-0.4631, 1.2459)

Table 5: ρ = 0.95 when the sample size, N=300.

Parameters	Estimators	Standard Error	CI
θ_0	OLS	0.1783	(16.6980, 17.3999)
	BIP	0.0858	(16.6082, 16.9460)
θ_1	OLS	0.8281	(7.2806, 10.5401)
	BIP	0.0008	(9.9984, 10.0016)
θ_2	OLS	1.3195	(2.3793, 7.5728)
	BIP	0.3455	(5.1697, 6.5293)
θ_3	OLS	1.2601	(0.4608, 5.4207)
	BIP	0.4380	(3.0036, 4.7276)

Table 6: ρ = 0.80 when the sample size, N=300.

Parameters	Estimators	Standard Error	CI
θ_0	OLS	0.1744	(16.2166, 16.9030)
	BIP	0.1482	(16.3663, 16.9497)
θ_1	OLS	0.4130	(7.0599, 8.6854)
	BIP	0.0008	(9.9984, 10.0015)
θ_2	OLS	0.6380	(4.2917, 6.8028)
	BIP	0.3068	(4.0674, 5.2748)
θ_3	OLS	0.6512	(2.2425, 4.8056)
	BIP	0.3952	(0.3285, 1.8838)

Tables 3-6 also present the SE and CI for both the Bayesian and OLS estimators when the degree of multicollinearity are ρ = 0.80 and 0.95 for sample sizes of 200 and 300. The SE for sample sizes of 200 and 300 reduces compared to when the sample is 30. However, Bayesian method also has minimum value for SE and compact CI compared to OLS for sample sizes of 200 and 300.

Table 7: MSE of estimators for ρ = 0.95

Sample sizes	OLS	BIP
30	2.5599	0.8123
200	4.2136	1.3060
300	2.8584	1.1877

Results from Table 7 show the MSE of the estimators when the degree of multicollinearity is ρ = 0.95 for all the sample sizes. The MSE of Bayesian method of estimation are smaller than the OLS estimator for all the sample sizes considered.

6. Conclusion

In this work, Bayesian method of estimation with the use of informative prior (conjugate) in the presence of multicollinearity for linear regression model was proposed. The performance of proposed Bayesian method was compared with OLS. In order to facilitate comparison between the two methods, three different data sets were simulated with two multicollinearity levels. The criteria used for evaluation of performance of the estimators are the Standard Error (SE),Confidence/ Credible Intervals (CI)and MSE.

The performance of the estimators in terms of Standard Error (SE) shows that there was an increase in SE due to the increase in the degree of correlation especially when the sample size is small. As the sample sizes increase, the performance of both the OLS and BIP improve. The Bayesian method, BIP is more precise than OLS estimator having the minimum SE for both small and large samples while the average deviation from the true parameter as measured by MSE also showed that the MSE of Bayesian estimator is relatively smaller than the OLS

In terms of stability of all the estimators, the results from the Credible and Confidence Intervals (CI) of estimators show that, Bayesian method, BIP has a narrower CI of parameter estimates and also the most stable estimator compared to OLS.

The results suggested that the proposed Bayesian method using a natural conjugate prior; outperformed the OLS method.
Therefore, Bayesian method of estimation is suitable in handling multicollinearity especially when degree of multicollinearity is high and when there is sufficient prior information.

References

Alkhamisi, M. A., Khalaf, G. and Shukur, G. 2006. Some modifications for choosing ridge parameter. *Communications in Statistics A.*, 35:1–16.

Belsley, D. A., Kuh, E., and Welsch, R. E., 1980. Regression Diagnostics:Identifying Inflential Data and Sources of Collinearity. *Wiley.*

Curtis, S. M and Ghosh, S.K. 2011. A Bayesian approach to Multicollinearity and the simultaneous selection and clustering of predictors in Linear Regression. *Journal of statistical theory and practice* 5.4:715-735.

Dreze, J. 1962. The Bayesian Approach to Simultaneous Equations Estimation, Memorandom 67. Techonological Institute, Northwestern University.

Duzan, H. and Shariff, N. S. 2015. Ridge Regression for solving the Multicollinearity Problem: Review of Methods and Models, *Journal of Applied Sciences.*

Gujarati, D. N. 1995. Basic econometrics. New York, NY: *McGraw-Hill.*

Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression: application to non-orthogonal Problems. *Technometrics* 12:69–82.

Iguernane, M. 2016. A Good Choice of Ridge Parameter with Minimum Mean Squared Error. *Journal of Biometrics & Biostatistics.* 7:2.

Ijarchelo, S. M., Afereydoon, K. and Zamanzadeh, L. 2016. Bayesian variable selection under collinearity of parameters. *Reserach journal of applied sciences* 11.7:428-438.

Jeffers, J.N. 1967. Two case studies in the application of principal component analysis. *Appl.Statist.*, 16, 225-236.

Jolliffe, LT. 1972. Discarding variables in a principal component analysis, I: Artificial data. *Appl. Statist.*, 21, 160-173

Kibria, B. M. G. 2003. Performance of some new ridge regression estimators. *Communications in Statistics- Theory and Methods* 32:419–435.

Kibria, B. M. G. and Banik, S. 2016. Some ridge regression estimators and their performances. *Journal of Modern Applied statistical methods.*15.1 206-238.

Koop, G. 2003. Bayesian Econometrics.*John Wiley and sons Ltd.* U.K.

Lee, H, Park, Y and Lee, S. 2015. Principal Components Regression by Principal Component Selection. *Communications for Statistical Applications and Methods*,22.2:173-180.

Mansfield, E.R., Webster, J.T., Gunst, R.F. 1977. An analytic variable selection technique for principal component regression. *Appl. Statist.*, 36, 34-40.

Miller, A.J. 1990. Subset Selection in Regression, *Chapman and Hall,* London.