Original Paper   http://indexmedicus.afro.who.int

Implicit collinearity effect in linear regression: application to basal metabolism rate prediction

H. Noël FONTON [1*], Y.J. Arcadius AKOSSOU [2] and Evelyne LOZES[3]

[1]*Laboratoire d'Etude et de Recherche en Statistique Appliquée et Biométrie,*
Université d'Abomey-Calavi, 01 BP 526, Cotonou, Bénin.
[2]*Faculté d'Agronomie, Université de Parakou, Bénin.*
[3]*Département de Génie, de Biologie Humaine, Ecole Polytechnique Abomey-Calavi, Cotonou, Bénin.*
[*]Corresponding author; E-mail: hnfonton@gmail.com

ABSTRACT

Collinearity of predictor variables is a severe problem in the least square regression analysis. It contributes to the instability of regression coefficients and leads to a wrong prediction accuracy. Despite these problems, studies are conducted with a large number of observed and derived variables linked with a response variable. The aim of this study is to highlight a better understanding of the misleading effect of collinearity introduced by derived variables and the efficiency of alternative methods. Twelve variables selection models were subjected to five parameter estimation methods characterized by their ability to reduce the collinearity effect. The response variable and eight anthropometric variables and two derived variables were collected with 200 children of 5 to 10 years old. We found that the selection methods do not mitigate the collinearity of selected subset variables, the size of selected subset variables depends on the collinearity of data samples and no significant correlation exists between sample and selected subset data collinearities. The analysis show that predictive quality did not improve with the introduction of derived variables. The alternative methods did not result in significant efficiency of prediction quality. We recommend avoiding the introduction of derived variables for the establishment of regression equation for prediction use.
© 2015 International Formulae Group. All rights reserved.

Keywords: Collinearity, prediction, regression, ridge regression, conditional likelihood, basal metabolism rate.

INTRODUCTION

Multivariate regression is often used to make predictions. However, the quality of some established equations is often questioned because this lead to illusions so caution is advisable (Armstrong, 2012). The main reason advanced of poor prediction quality is the selection of explanatory variables in the model. In many studies, a large number of variables are collected regardless their quality as their influence on the response variable. As pointed out by Sauerbrei et al. (2008), the key task is the identification of relevant variables influencing the response variable.

In fact, for a matrix X of predictor variables and y the response variable, the model is : $y = X\beta + \varepsilon$

With β, the unknown vector of regression coefficients and ε the residual vector. The estimated coefficients obtained with the least-square method computed as

$\hat{\beta} = (X'X)^{-1}X'y$ is characterized by a minimum variance according to Gauss-Markov theorem defined as follows:

$$\text{var}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$
$$= \sigma^2 (X'X)^{-1}$$

The estimated mean square error $CM_{\hat{\beta}}$ of β is (Gruber, 1998):

$$CM_{\hat{\beta}} = \hat{\sigma}^2 tr(X'X)^{-1}$$
$$= \hat{\sigma}^2 \sum_{i=1}^{k} \frac{1}{l_i}$$

with $tr(X'X)$ the trace of correlation matrix of X, $\hat{\sigma}^2$ the estimated residual variance, k the number of predictor variables, and l_i the ith eigenvalue. When the predictor variables are correlated, some eigenvalues are closed to zero and increase the $CM_{\hat{\beta}}$. The consequences are the instability of regression coefficients, standard errors on estimates inflated, the change of their sign and consequently inference statistics biased (Tomassone et al., 1992; Tormod and Bjorn-Helge, 2001; Dobson, 2002; Feng-Jenq, 2008; Smith et al., 2009; Woolston et al., 2013; Akossou and Fonton, 2013; Dormann et al., 2013). Some studies pointed out the excessive adjustment quality and optimism for prediction quality (Copas, 1983; Fonton, 1995). It is well known that the reliability of least-square or maximum-likelihood estimators gets worse as linear relationships between predictors become more acute (Lauridsen and Mur, 2006; Yakubu, 2009).

Despite the relevance of the collinearity, what kind of interest including in the predictor variables some variables computed from others and therefore introducing multicollinearity? It is most commonly used to include in a prediction equation all variables which are known or derived to affect the response variable. The reason is supported by biological concepts or theories frequently used as predictor variables in regression analyses. While the inferences based on ordinary least-squares regression were strongly influenced by the collinearity, some alternative methods were proposed to reduce the effect of collinearity on the prediction quality of regression equations. Many of them are based on the selection of variables and the other on the estimation of regression coefficients. They are ranging from clustering of predictors, threshold-based pre-selection, through latent variable methods to shrinkage (Dormann et al., 2013) and the conditional likelihood method to control selection bias (Miller, 1990).

The aim of this paper is to explain the misleading effect of introducing derived predictors and the capacity of some alternatives methods to reduce collinearity effect. The following hypothesis are tested: (1) the collinearity of data samples governs those of selected subset variables, (2) the size of selected subset variables depends on the collinearity of data samples, (3) the efficiency of alternative methods depends on the data structure, (4) the quality of prediction is improved with the introduction of derived variables.

MATERIALS AND METHODS
Study data and sampling

The data was obtained from a study to establish regression equations to predict the basal metabolism rate (*BMR*) of children. Two hundred children from a primary school in southern Benin (West Africa) aged between 5 and 10 years were sampled. The first data configuration (configuration I) involved age and seven anthropometric variables: weight, height, mid-upper arm circumference, folds of biceps, triceps, subscapularis and suprailiac. These eight predictor variables are essential to understanding the basal metabolic rate (Lovegrove, 2003). The second data configuration (configuration II) consisted of

ten independent variables, those of the first configuration and two derived variables. The two derived variables *BMI* and *MA* were supported by biological concepts theories. *BMI* is the Body Mass Index of QUETELET which is the ratio between the weight and the square of the height. The variable *MA* is derived from the relationship between the folds of triceps and the circumference of the forearm.

Two sampled sizes n of data were chosen: $n = 15$ and $n = 30$ for k/n located at 2/6 and 4/6. Twenty independent samples of data were generated for each data configuration and size. This sampling was done using the command SAMPLE of MINITAB 14.1 software. For each sampled data, the collinearity index was computed and observations excluded (prediction observations) were used to assess prediction quality. The collinearity index (IC) is the average of the inverse of the eigenvalues l_i of the correlation matrix of k predictor variables and computed as:

$$IC = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \frac{1}{l_i}}$$

It is also the trace of the inverse of the correlation matrix divided by the number of variables. This design was adopted because the response, the sample size and the sampling variations govern the impact of collinearity on the parameter estimations (Woolson et al., 2013).

Prediction models

Twelve selection models were subjected to five coefficients estimation methods, giving sixty prediction models. The twelve selection models consisted of four selection methods at three probability levels of selection: 15%, 5% and 1.5% which correspond respectively to $t = 1.5$, $t = 2$ and $t = 2.5$ of Student's t parameter. Two

commonly used selection methods: the stepwise (sw) and forward (fw) methods were used; their selection models were designated respectively by $sw1.5$, $sw2$, $sw2.5$ and $fw1.5$, $fw2$ $fw2.5$. The two other selection methods used were alternative to the least-squares method based on a parameter that controls the amount of shrinkage (Bondell and Reich, 2007; Ejaz et al., 2011). They are based on testing the significance of regression coefficients with the *t-test*. For these, a completed regression including all the k potential predictors was performed using the value d_1 proposed by Hoerl et al. (1975) and, secondly, the value d_2 proposed by Lawless and Wang (1976). The parameters d_1 and d_2 were computed as follows:

$$d_1 = p\hat{\sigma}^2 / \left(\hat{\beta}_A' \hat{\beta}_A\right) \quad \text{and}$$

$$d_2 = p\hat{\sigma}^2 / \left(\beta_A' X_A' X_A \hat{\beta}_A\right) \text{ with } \hat{\sigma}^2 \text{ the}$$

estimated residual variance of the least-squares regression. The significance tests were based on three probability levels ($t = 1.5$, $t = 2$, $t = 2.5$) and non-significant variables were discarded. Those selection models were designated by $hkb1.5$, $hkb2$ and $hkb2.5$ in the case of d_1 and $lw1.5$, $lw2$, and $lw2.5$ for d_2.

The regression coefficients of the best subsets, *A*, selected by the selection models were estimated by the ordinary least-squares, designated by *LS*:

$$\hat{\beta}_A = (X_A' X_A)^{-1} X_A' y$$

The following three shrinkage methods were the ridge regression and simple shrinkage. The ridge regression was computed as:

$$\hat{\beta}_A(d_i) = (X_A' X_A + d_i I)^{-1} X_A' y \ (d_i \geq 0)$$

The two ridge regression methods used were designated respectively by *HKB* and *LW*. The

fourth method was the simple shrinkage method based on the estimator of James and Stein (Stein, 1962) identify as *JS*:

$$\hat{\beta}_A(JS) = c\hat{\beta}_A$$

with $c = 1 - \dfrac{(p-2)(n-p-1)}{(n-p+1)pF_{obs}}$, F_{obs} the

ratio of regression mean square and residual mean square, $\hat{\beta}_A$ the least-squares estimation and p the number of selected subset variables. The latter estimation method was the conditional likelihood estimation (ML). It was based on the adjustment the regression coefficients to the estimated selection bias (Miller, 1990; Fonton, 1995).

Statistical parameters and analyses

A total of 240 subsets of variables were selected and 1200 regression equations were established by data structure. For each set of sample data the collinearity index IC, the number p of selected subset variables A and their collinearity IC_A were determined. The correlation coefficient between IC and IC_A was tested according to the null hypothesis $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$. For each data structure, the sizes of selected subset variables were compared using Generalized Linear Model according to IC and IC_A. Those two analyses were intended to address the two first hypotheses.

The statistical parameter used to evaluate the predictive accuracy of regression equations was the mean square error of prediction (*MSEP*), computed as follows:

$$MSEP = \sqrt{\sum_{i=1}^{n_P} (\hat{y}_i - y_i)^2 / n_P}$$

with y_i the observed value and $\hat{y}_i$ the predicted value for the observation i, n_p the number of prediction observations. For each

data structure, ANOVA was used to compare the regression methods. The later statistical methods were used to address the hypotheses 3 and 4.

RESULTS
Descriptive analysis of data

The descriptive statistics were presented in Table 1. The average metabolic rate (*BMR*) was 2.55 with a standard deviation of 0.38 and collinearity index of the predictors of 7.2. The correlation coefficient between the response variable *BMR* and each predictor variable ranged from -0.04 to 0.62. The correlation coefficients of *BMR* with *BMI* and *MA* were 0.40 and 0.54, respectively. The coefficient of determination R^2 of the completed model (including all the predictors' variables) was 0.45 for the two data configurations (I and II). The introduction of derived variables did not increase adjustment quality of the completed model.

Analysis of collinearity index

The averages of IC were 19.5 ($\pm$ 0.7) and 8.1 ($\pm$ 0.1), respectively for sample sizes $n = 15$ and $n = 30$ for data configuration I, while for the configuration II, it increased to 1896.4 ($\pm$ 39.5) and 410.0 ($\pm$ 8.0) for $n = 15$ and $n = 30$ respectively (Table 2). For both configurations, the collinearity decreased with the increase in sample size (n).

For all data structures, the correlation coefficient $r(IC, IC_A)$ between the collinearity index of sampled data (IC) and the collinearity index of selected subsets (IC_A) was $r = 0.007$ with a probability value of 0.125. A similar result was found with the ordinary stepwise selection method ($r = 0.060$ and $p_{value} = 0.518$). We obtained for any significant correlation between the collinearity of sampled data and

the collinearity of selected subset variables whatever the data structure irrespective of $k = 8$ and $n = 30$. The correlation coefficient $r(IC, IC_A)$ and the ratio k/n were inversely proportional (Figure 1). This explained that when k/n increased the correlations decreased and changed the sign. For the data structure $k/n = 8/30$ (low value of k/n), $r(IC, IC_A)$ was 0.3, while for the other structures it was close to 0.

Analysis of the sizes of selected subsets

The size of selected subsets decreased when the significance selection level increases, irrespective of the selection method and data configuration. The sizes p of the selected subsets variables ranged from 0 to 5 for the configuration I, and from 0 to 8 for configuration II. The percentages of data samples leading to $p = 0$ were 6.8% and 16.7% for configuration I, 16.7% and 20.8% for configuration II, respectively for $n = 30$ and $n = 15$ (Figure 2.a). The case $p = 0$ was characterized by strong collinearity of data samples mainly for $n = 15$ as showed by the Figures 2.b and 2.c; the lead regression equation was the mean model $y = \bar{y}$. The subsets with $p = 1$ were the most represented (Figure 2.a) with 45.0% and 52.5% for configuration I and 37.5% and 41.7% for configuration II, respectively for $n = 15$ and $n = 30$. The same trend was observed with $p = 2$. It might be assumed that the size of selected subsets is inversely proportional to the collinearity of data sample. The ANOVA performed for the comparison of p showed a significant difference for IC whatever the data structure except for $k = 8$ and $n = 30$ as presented in Table 3. A significant difference between sizes of select

subset variables for IC_A was also observed whatever the data structure.

The analysis of variance for $p \geq 2$ showed no-significant difference between the sizes of subsets of variables selected regardless to the data structure. About 70% of the data samples were characterized by $p = 0$ and $p = 8$ with the shrinkage selection methods (Figure 3). The stepwise method was characterized by $p \leq 5$ and more represented (39%) with $p = 1$. This was followed by the forward selection method with 38%. The selection method of Lawless and Wang was more representative for $2 \leq p \leq 5$ (41% to 63%).

Prediction quality of established equations

The analysis of prediction quality showed that the *MSEP* were ranging from 0.27 to 0.77 (Table 4). On average, for the same configuration ($k = 8$ or $k = 10$), the prediction quality was better when *n* is higher. Also, for the same number of observations *n*, the best predictive value was obtained with the configuration of low collinearity.

The maximum value of *MSEP* was obtained with $k = 10$, $n = 15$ about twice the standard error. This indicates the introduction of additional variability resulting in the consequence of instability of regression coefficients. For this structure 12.2% of 1200 equations were characterized by an additional variability, while it was only 8.8%, 0% and 0.5%, respectively for $k = 8$ $n = 15$, $k = 8$ $n = 30$ and $k = 10$ $n = 30$. The selection method presenting more variability was hkb while for parameters estimation it was JS method.

As presented in Table 5, models for predicting basal metabolism highlighted a significant difference between data structures except $k = 8$ $n = 30$. Irrespective to all data structure the difference was statistically

significant for selection models. The interaction between selection models and estimation methods was not significant irrespective of the data structure.

For all models, improved predictive quality of data configuration I compared to those containing derived variables were 5.10% and 1.50%, respectively for $n = 15$ and $n = 30$. The analysis revealed that the best predictive model equation was $Lw * lw1.5$. Comparing this regression model to $LS * sw1.5$, the improvements of predictive quality were 3.0%, 2.6%, 7.3% and 1.6%, respectively for $k = 8\ n = 15$, $k = 8\ n = 30$, $k = 10\ n = 15$ and $k = 10\ n = 30$.

Table 1: Summary statistics of eight predictor variables, two derived predictor variables (*BMI* and *MA*) and the response variable (*BMR*) and the correlation coefficients (r) between the response variable and the predictor variables.

Variables	Mean	Standard deviation	Correlation coefficient (r)
Age (months)	90.2	16.7	0.454
Weight (kg)	20.9	4.0	0.631
Height (cm)	119.8	9.2	0.598
Circumference of forearm (cm)	16.7	1.4	0.470
Folds of biceps ($10g/mm^2$)	4.0	0.8	-0.035
Folds of triceps ($10g/mm^2$)	6.3	1.4	-0.048
Folds of subscapularis ($10g/mm^2$)	5.3	1.0	0.238
Folds Suprailiac ($10g/mm^2$)	3.04	0.6	0.159
*BMI (*Index of QUETELET*)*	14.4	1.1	0.395
MA	17.5	2.9	0.542
BMR	2.55	0.38	

Table 2: Statistics (mean, minimum, maximum and $\hat{\sigma}_{\overline{IC}}$) of the collinearity index for sampled data for different data structures.

	Data structures			
	$k = 8$		$k = 10$	
	$n = 15$	$n = 30$	$n = 15$	$n = 30$
Mean	19.5	8.1	1896.4	410.0
Minimum	6.3	4.9	406.1	215.3
Maximum	67.5	13.5	3189.5	943.4
$\hat{\sigma}_{\overline{IC}}$	0.7	0.1	39.5	8.0

k is the number of predictors and n is the number of observations, and $\hat{\sigma}_{\overline{IC}}$ the standard error of mean.

Table 3: Comparison of subset sizes (p) for collinearity indexes of sampled data (IC) and subset variables (IC_A).

		Data structures			
		$k = 8, n = 15$	$k = 8, n = 30$	$k = 10, n = 15$	$k = 10, n = 30$
IC	F_{value}	4.50	1.79	2.70	2.69
	P_{value}	0.001***	**0.18 ns**	0.009***	0.013**
IC_A	F_{value}	9.83	39.27	35.63	7.32
	P_{value}	0.000***	0.000***	0.000***	0.000***

ns = not significant, ** : high significant difference, *** : very high significant difference.

Table 4: Prediction mean square error by data structure.

k	n	Mean	Minimum	Maximum
8	15	0.324	0.278	0.446
8	30	0.301	0.273	0.373
10	15	0.340	0.269	0.766
10	30	0.306	0.268	0.417

k is the number of predictor variables and n the number of observations.

Table 5: F_{values} and their significant tests (P_{values}) for the comparison of regression models estimation methods, selection models, selection methods and significant levels.

k	n	Regression models	Estimation methods	Selection models	Selection methods	Significant levels
8	15	2,80 (***)	1,95 (ns)	12,46(***)	19,10 (***)	20,68 (***)
8	30	0,69 (ns)	0,79 (ns)	2,95 (***)	2,12 (ns)	5,48 (**)
10	15	2,02 (***)	2,33 (ns)	8,34 (***)	27,48 (***)	3,74 (*)
10	30	6,04 (***)	1,72 (ns)	27,68 (***)	94,80 (***)	2,33 (ns)

ns = not significant, * : significant difference, ** : high significant difference, *** : very high significant difference.
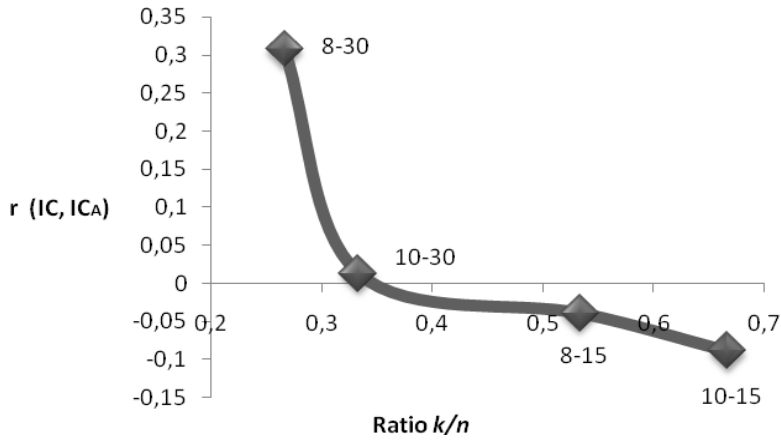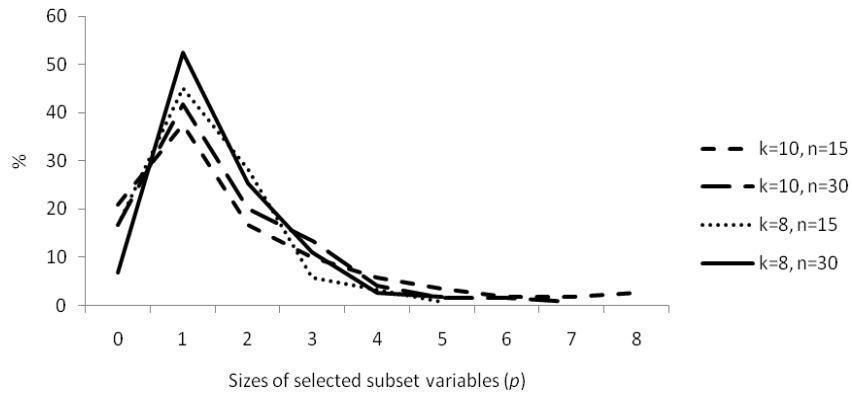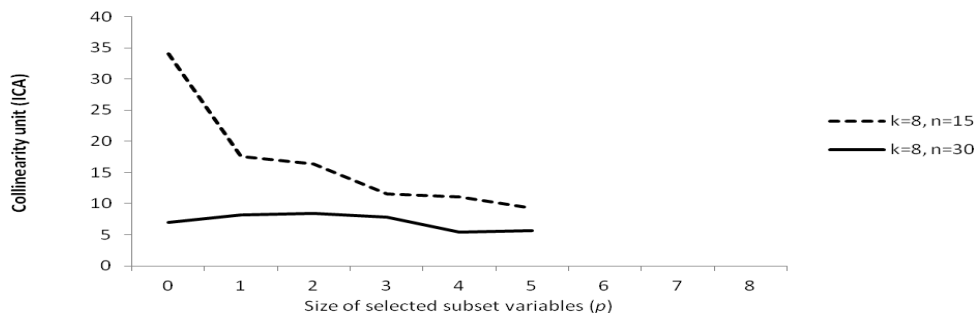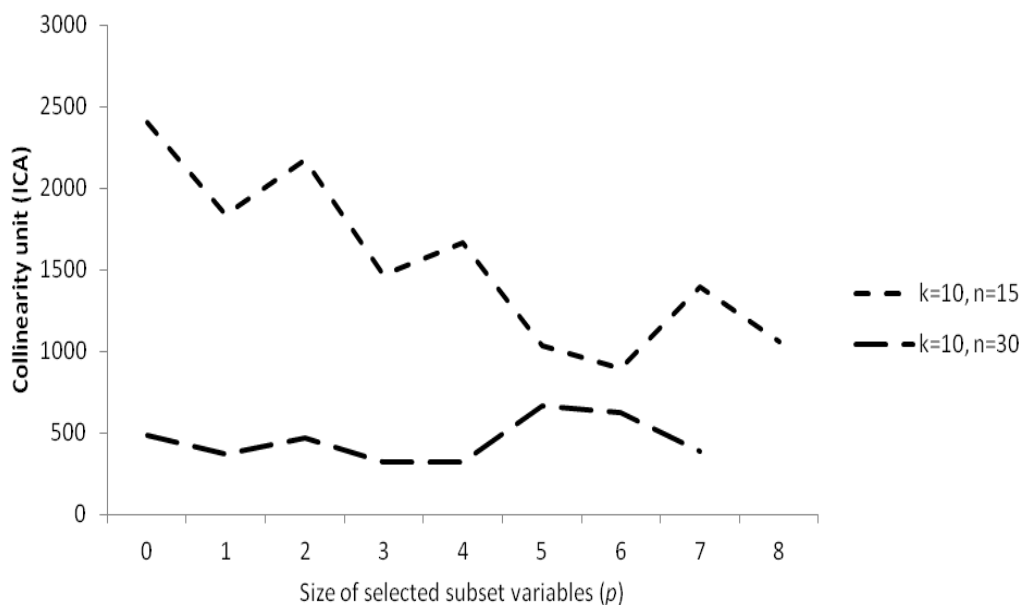
Figure 1: Trend of the correlation coefficients between collinearity indexes of data samples and subset selected variables according to the ratio *k/n*.

a) Percentage of size (p) of selected subset variables by data structure.

b) Sample data mean collinearity index by subset size for configuration I

c) Sample data mean collinearity index by subset size for configuration II

Figure 2: Characterization of selected subset variables: relation between the percentage of size of subsets selected variables and the collinearity index (IC) by data structure.
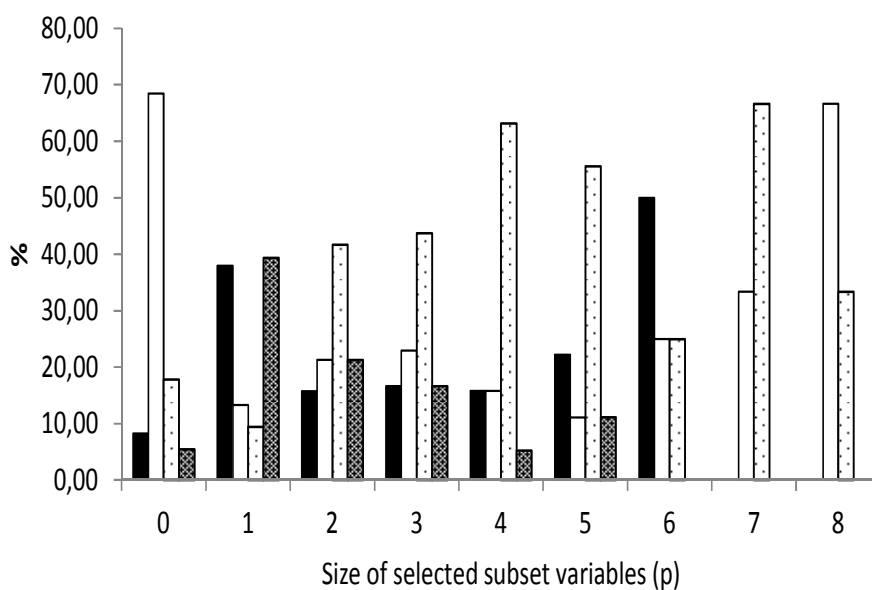

Figure 3: Percentage of selection models for each size of selected subsets variables for all data structures confused.

DISCUSSION

The results showed no relationship according to collinearity index between data samples and the selected subsets variable unless for $k = 8$ $n = 30$. The collinearity index of the selected subset could be considered as the effect of the selection method because shrinkage methods were based on the monitoring the collinear variables. This situation might stem from the fact that the collinearity is a parameter that allows testing the degree of robustness of those methods. The proposed ridge regression (hkb and *lw*) is a technique for coping with difficulties arising from multicollinearity (McCabe, 1978) by shrinking the ordinary least square estimator towards zero (Tibshirani, 1996). On the contrary, classical selection methods such as stepwise and forward selection have been less affected by this increased collinearity.

The collinearity of data samples do not determined the collinearity of selected subset variables. The ratio k / n was inversely proportional to the correlation between IC and IC_A. When n increases for a given value of k, the correlation increases. With respect to the collinearity of data samples, the comparison of the sizes of the subsets of selected variables showed the existence of differences. For $p \leq 1$ data samples were characterized by high collinearity. It can then be assumed that the size of selected subset variables depends on the collinearity of predictors variables. Nevertheless, the results showed that for $p > 1$, the data samples collinearity were statistically equal.

The accuracy of regression models for the establishment of equations is strongly influenced by the selection model. That was the wrong trajectory of regression with the stepwise methods as pointed out by Harell (2001) and Meloun et al. (2002). This explained the promotion of new techniques combining variables selection and shrinkage of coefficients of penalized regression (Efron et al. 2004). The prediction quality did not improve with the introduction of derived variables. The *MSEP* of configuration II was 5.10% higher compared to configuration I (lowest collinearity). The shrinkage methods improved the prediction quality more than the LS for highest muticollinearity and lowest number of observations. This is consistent with penalised regressions in presence of multicollinearity and where the number of predictors is greater (Avalos, 2009). Hastie et al., (2009) confirmed this result and concluded that shrinkage of coefficients towards zero leads to smaller prediction error due to decreased variance. In the same way, Arezoo and Habshah (2009) proposed two robust estimators in order to address multicollinearity problems. Those methods are more complex for a simple user of statistical tools. We think that the solution is not quite easy with an implicit multicollinearity.

Conclusion

In the light of the foregoing, the introduction of derived variables affects the size of selected subset variables with no significant efficiency in prediction. So, if possible, we recommend avoiding any such process for prediction use.

REFERENCES

Akossou YJ, Fonton NH. 2013. Empirical comparison of three estimators of collinearity. *Int. J. Math. Comput.*, **20**(3): 66-76.

Armstrong, JC. 2012. Illusions in Regression Analysis. *Int. J. Forecast.,* **28**(3): 689-694.

Avalos M. 2009. Selection de variables avec Lasso dans la régression logistique conditionnelle. *41ᵉᵐᵉˢ Journées de Statistique*, Société Française de Statistique, Bordeau France.

Arezoo B, Habshah M. 2009. Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. *J. Math. Stat.*, **5**(4): 311-321.

Bondell HD, Reich BJ. 2007. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics,* **64**: 115-121.

Copas JB. 1983. Regression, prediction and shrinkage. *J. R. Stat. Soc.,* **B45**: 311-354.

Dobson AJ. 2002. *An Introduction to Generalized Linear Models.* Chapman and Hall: New York.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Garcia Marquéz JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D Lautenbach S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography,* **36**(1): 27-46.

Efron B, Hastie T, Johnstone I, Tibshirani R. 2004. Least angle regression. *Ann. Statist.*, **32**(2): 407-499.

Ejaz SA, Enayetun R, Shakhawat H. 2011. Absolute penalty estimation. In *International Encyclopedia of Statistical Science*, Miodrag L. (ed). Springer-Verlag: Berlin Heidelberg.

Fenq-Jenq L. 2008. Solving multicollinearity in the process of fitting regression model using nested estimate procedure. *Quality and Quantity*, **42**: 417-426.

Fonton NH. 1995. *Comparaison des méthodes de prédiction en régression linéaire multiple.* Thèse de doctorat, Faculté Universitaire d'Agronomie de Gembloux, Belgique, 230pp + annexes.

Fonton N, Palm R. 1998. Comparaison empirique des méthodes de prédiction en régression linéaire multiple. *Rev. Stat. Appl.*, **46**(3): 53-64.

Graham MH. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology,* **84**: 2809-2815.

Gruber MJH. 1998. Improving efficiency by shrinkage: *The James Stein and Ridge Regression Estimators.* Marcel Becker: New York.

Harrell FEJr. 2001. *Regression Modeling Strategies – With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer: New York.

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2ⁿᵈ edn). Springer: New York.

Hoerl AE, Kennard RW, Badwin KF. 1975. Ridge regression: some simulations. *Comm. Stat., A***4**: 105-123.

McMabe GPJr. 1978. Evaluation of regression coefficient estimates usingαacceptability. *Technometrics*, **20**(2): 131-139.

Meloun M, Militky J, Hill M, Brereton RG. 2002. Crucial problems in regression modeling and their solutions. *Analyst* **127**: 433-450.

Miller AJ. 1990. *Subset Selection in Regression. Monographic on Statistics and Applied Probability*. Chapman and Hall: New York.

Lauridsen J, Mur J. 2006. Multicollinearity in cross-sectional regressions. *J. Geo. Syst.*, **8**(4): 317-333.

Lawless JF, Wang P. 1976. A simulation study of ridge and other regression estimators. *Comm. Stat., A***5**: 307-323.

Lovegrove BG. 2003. The influence of climate on the basal metabolism rate of small mammals: a slow-fast metabolic continuum. *J. Comp. Physiol., B***173**: 87-112.

Sauerbrei W, Hollader N, Buchholz A. 2008. Investigation about a screening step in model selection. *Statistics and Computing*, **18**(2): 195-208.

Smith A et al. 2009. Confronting collinearity: comparing methods for disentangling the effects of habitat loss and fragmentation. *Landscape Ecol.*, **24**: 1271-1285.

Stein CM. 1962. Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Stat. Soc., B***24**: 265-296.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J. R. Soc. B* **58**: 267-288.

Tomassone R, Audrin S, Lesquoy de Turckheim E, Millier C. 1992. *La Régression: Nouveaux Regards sur une Ancienne Méthode Statistique* (2nd edn). Masson: Paris.

Tormod N, Bjorn-Helge M. 2001. Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics*, **15**: 413-426.

Woolston A, Tu YK, Gilthorpe MS, Bacter PD. 2013. Measuring the impact of collinearity in epidemiology research. *Int. J. Stat. Prob.*, **2**(2): 1-11.

Yakubu A. 2009. Fixing collinearity instability in the estimation of body weight from morpho-biometrical traits of West African dwarf goats. Trakia *Journal of Sciences*, **7**(2): 61-66.