Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 1 (2016), pp. 481-491

http://www.ripublication.com

# On Properties of the difference between two modified $C_p$ statistics in the nested multivariate linear regression models

**S. Boonyim[1], J. Jitthavech[2] and V. Lorchirachoonkul[3]**
*[1-3]School of Applied Statistics, National Institute of Development Administration Bangkok 10240, Thailand.*

**Abstract**

The statistic $D$ which is the difference of the modified $C_p$ statistics in the full model and in the reduced model is developed for variable selection in the contemporaneous multivariate linear regression nested model. The expectation, the variance and the distribution of the statistic $D$ are derived via three theorems and three lemmas.

**Keywords:** Modified $C_p$ statistic, test statistic for variable selection, distribution of the test statistic.

**Mathematics Subject Classification:** 62H10, 62H15

## Introduction

In the system-of-equations model with contemporaneously but no serially correlated disturbance the dependent variable vector $\mathbf{y}$ can be written in the stacked form of the system of $m$ equations model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is the $nm \times 1$ vector consisting of subvectors $\mathbf{y}_i$, the $n \times 1$ dependent variable vector in equation i, $i = 1, 2, \ldots, m$, $\mathbf{X}$ is the $nm \times p_T$ diagonal matrix consisting of diagonal submatrices, $\mathbf{X}_i$, the $n \times p_i$ matrix of independent variables including the constant unit vector in equation $i$ with rank($\mathbf{X}_i$) $= p_i$, $i = 1, 2, \ldots, m$, $p_i$ is the number of parameters in equation $i$, $p_T = \sum_{i=1}^{m} p_i$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1' \boldsymbol{\beta}_2' \ldots \boldsymbol{\beta}_m')'$ is the $p_T \times 1$ parameter vector consisting of subvectors $\boldsymbol{\beta}_i$, the $p_i \times 1$ parameter vector in equation $i$, $i = 1, 2, \ldots, m$, and $n$ is the number of observations in each equation. The $nm \times 1$ random disturbance vector, $\boldsymbol{\varepsilon}$, is assumed to be uncorrelated across observations in the same equation but contemporaneously correlated across observations in different equations and distributed as $N_{nm}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$ where $\boldsymbol{\Omega}_\varepsilon = \boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_n$ and the covariance matrix $\boldsymbol{\Sigma}_\varepsilon$ can be written as

$$\boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix}.$$

The multivariate regression model is a special case of the system-of-equations model where $\mathbf{X}_i = \mathbf{X}$, $p_i = p$, $i = 1, 2, \ldots, m$. The total number of parameters in the multivariate regression model is equal to $p_T = mp$. In a multivariate linear regression model, the modified $C_p$ statistic suggested by Lorchirachoonkul and Jitthavech (2012) becomes

$$M\Gamma = E(\boldsymbol{\varepsilon}'\boldsymbol{\Omega}_y^{-1}\boldsymbol{\varepsilon}) - (n - 2p)tr(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_\varepsilon), \tag{2}$$

where $\boldsymbol{\varepsilon}$ is an $nm \times 1$ vector of random errors in the model and distributed as $N_{nm}(\mathbf{0}, \boldsymbol{\Omega}_\varepsilon)$, $\boldsymbol{\Omega}_y^{-1} = \boldsymbol{\Sigma}_y^{-1} \otimes \mathbf{I}_n$, $\boldsymbol{\Sigma}_y$ is an $m \times m$ covariance matrix of dependent variables. The estimate of $M\Gamma$ in (2) in the full model with the total number of parameters $mp_f$ can be expressed as

$$MC_f = \mathbf{e}_f'\hat{\boldsymbol{\Omega}}_y^{-1}\mathbf{e}_f - (n - 2p_f)tr(\hat{\boldsymbol{\Sigma}}_y^{-1}\mathbf{S}_{e_f}), \tag{3}$$

where $\mathbf{e}_f$ is the $nm \times 1$ vector consisting of $n \times 1$ subvector of residuals in equation $i$, $\mathbf{e}_{f_i}$, $i = 1, 2, \ldots, m$, in the full model, $\hat{\boldsymbol{\Sigma}}_y$ is the estimate of $\boldsymbol{\Sigma}_y$ and $\mathbf{S}_{e_f}$ is the $m \times m$ sample covariance matrix of the residual vectors $\mathbf{e}_{f_i}$'s, which is used as the estimate of $\boldsymbol{\Sigma}_\varepsilon$ provided that the full model is a "good" approximation of the unknown "true" model.

The modified $C_p$ statistic in (2) is different from the original definition of Mallows $C_p$ statistic, and the extended definition of Mallows $C_p$ statistic in a single equation model to a multivariate regression model (Sparks et al., 1983; Fujikoshi and Satoh, 1997; Yanagihara and Satoh, 2010). The unknown error covariance matrix of the true model in the original definition is replaced by the known covariance matrix of dependent variables as shown in (2). The rest of the paper is organized as follows. The hypothesis testing for variable selection under the modified $C_p$ statistic in (2) in the contemporaneous multivariate linear regression nested model is introduced in the next section. Section 3 presents the derivation of the statistical properties and the distribution of the difference of the modified $C_p$ statistics in the full model and in a reduced model in the multivariate linear regression. The final section is conclusions.

## Hypothesis Testing

In variable selection in the multivariate linear regression model in (1) under the $M\Gamma$ criterion in (2), an explanatory variable can be eliminated from the model specification if its discard does not increase the value of the statistic significantly. Otherwise, such explanatory variable is retained in the model. In other words, a variable or a number of variables can be eliminated from the model specification if the null hypothesis is not rejected at a significant level $\alpha$ .

$$H_0 : M\Gamma_r = M\Gamma_f \text{ against } H_a : M\Gamma_r > M\Gamma_f \ , \tag{4}$$

where $M\Gamma_f$ and $M\Gamma_r$ are the values of the modified $C_p$ statistics of the full model and the reduced model respectively. In order to test the hypothesis (4), it is sufficient to consider whether the difference

$$D = MC_r - MC_f , \tag{5}$$

is significantly different from zero. From (3), it is obvious that the difference $D$ can be expressed as

$$D = (\mathbf{e}_r - \mathbf{e}_f)'\hat{\mathbf{\Omega}}_y^{-1}(\mathbf{e}_r - \mathbf{e}_f) - 2d\,tr(\hat{\mathbf{\Sigma}}_y^{-1}\mathbf{S}_{e_f}), \tag{6}$$

where $\mathbf{e}_f$ and $\mathbf{e}_r$ are the residual vectors in the full and reduced models respectively, $d$ is the difference in the number of parameters in the full and reduced models, $d = p_f - p_r$ .

## Properties of the Statistic $D$

In this section, the expectation, the variance and the distribution of the statistic $D$ are investigated. The statistic $D$ in (6) consists of two terms. The first term in the quadratic form can be expressed in terms of a linear combination of the products of two correlated normal vectors as

$$(\mathbf{e}_r - \mathbf{e}_f)'\hat{\mathbf{\Omega}}_y^{-1}(\mathbf{e}_r - \mathbf{e}_f) = \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\sigma}_y^{ij}(\mathbf{e}_{r_i} - \mathbf{e}_{f_i})'(\mathbf{e}_{r_j} - \mathbf{e}_{f_j}) = \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\sigma}_y^{ij}\mathbf{d}_i'\mathbf{d}_j , \tag{7}$$

where $\mathbf{d}_i = \mathbf{e}_{r_i} - \mathbf{e}_{f_i}$ and $\mathbf{d}_j = \mathbf{e}_{r_j} - \mathbf{e}_{f_j}$ . Similarly, $tr(\hat{\mathbf{\Sigma}}_y^{-1}\mathbf{S}_{e_f})$ which is a part of the second term of the statistic $D$ in (6) can be written in terms of a linear combination of the product of two correlated normal vectors as (7),

$$tr(\hat{\mathbf{\Sigma}}_y^{-1}\mathbf{S}_{e_f}) = \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\sigma}_y^{ij} s_{e_{f_{ij}}} = \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\sigma}_y^{ij} \frac{\mathbf{e}'_{f_i}\mathbf{e}_{f_j}}{(n - p_f)}, i, j = 1, 2, ..., m, \tag{8}$$

where $s_{e_{f_{ij}}} = \frac{\mathbf{e}'_{f_i}\mathbf{e}_{f_j}}{(n-p_f)}, i=1,2,...,m; j=1,2,...,m.$ Noted that both terms in the LHSs of (7) and (8) are in the similar form of weighted sum of the product of twocorrelated vectors which may be represented by $\mathbf{a}'\mathbf{Q}\mathbf{a}$ where $\mathbf{a}$ is a $nm\times 1$ vector consisting of $m$ correlated subvectors of size $n\times 1$, $\mathbf{a}_i$, with zero mean and variance $\sigma_{a_i}^2\mathbf{I}_n$, $i=1,2,\ldots,m$, and $\mathbf{Q}$ is an $nm\times nm$ matrix.

**Lemma 1.** Let $\mathbf{a}_i$ and $\mathbf{a}_j$ be the $n\times 1$ correlated vectors distributed as $N_n(\mathbf{0},\sigma_{a_i}^2\mathbf{I}_n)$ and $N_n(\mathbf{0},\sigma_{a_j}^2\mathbf{I}_n)$ respectively, $\rho_{a_{ij}}$ is a correlation coefficient between $\mathbf{a}_i$ and $\mathbf{a}_j$, $\mathbf{z}_i$ is an $n\times 1$ vector distributed as $N_n(\mathbf{0},\sigma_{a_i}^2\mathbf{I}_n)$, independent of $\mathbf{a}_i$, and $\tilde{\mathbf{a}}_j$ is defined as

$$\tilde{\mathbf{a}}_j = \rho_{a_{ij}}\frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{a}_i + \sqrt{(1-\rho_{a_{ij}}^2)}\frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{z}_i . \quad (9)$$

Then $\mathbf{a}_j$ and $\tilde{\mathbf{a}}_j$ are identically distributed as $N_n(\mathbf{0},\sigma_{a_j}^2\mathbf{I}_n)$.

**Proof of Lemma 1.** Since $\mathbf{a}_i$ and $\mathbf{z}_i$ are independent normal random vectors with the same zero mean and variance $\sigma_{a_i}^2\mathbf{I}_n$, the expectation of $\tilde{\mathbf{a}}_j$ is equal to zero,

$$E(\tilde{\mathbf{a}}_j) = E\left(\rho_{a_{ij}}\frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{a}_i + \sqrt{(1-\rho_{a_{ij}}^2)}\frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{z}_i\right)$$
$$= \mathbf{0},$$

and the variance of $\tilde{\mathbf{a}}_j$ can be expressed as

$$Var(\tilde{\mathbf{a}}_j) = \rho_{a_{ij}}^2\frac{\sigma_{a_j}^2}{\sigma_{a_i}^2}\sigma_{a_i}^2\mathbf{I}_n + (1-\rho_{a_{ij}}^2)\frac{\sigma_{a_j}^2}{\sigma_{a_i}^2}\sigma_{a_i}^2\mathbf{I}_n$$
$$= \sigma_{a_j}^2\mathbf{I}_n .$$

Therefore, $\tilde{\mathbf{a}}_j$ is distributed as $N_n(\mathbf{0},\sigma_{a_j}^2\mathbf{I}_n)$, the identical distribution as $\mathbf{a}_j$.

**Lemma 2.** Let $\mathbf{a}$ be a $nm\times 1$ vector consisting of $m$ correlated subvectors, $\mathbf{a}_i$'s, of size $n\times 1$ distributed as $N_n(\mathbf{0},\sigma_{a_i}^2\mathbf{I}_n)$, $\rho_{a_{ij}}$ be a correlation coefficient between $\mathbf{a}_i$ and $\mathbf{a}_j$ and $\mathbf{Q} = \Sigma\otimes\mathbf{I}_n$ where $\mathbf{\Sigma}$ is an $m\times m$ matrix. Thus,

$$\mathbf{a}'\mathbf{Q}\mathbf{a} = \sum_{i=1}^{m} u_i\frac{\mathbf{a}_i'\mathbf{a}_i}{\sigma_{a_i}^2}, \quad (10)$$

where $u_i = \sum_{j=1}^{m} \sigma_{ij}\rho_{a_{ij}}\sigma_{a_i}\sigma_{a_j}$ , and $\sigma_{ij}$ is the $ij^{th}$ element of the matrix $\mathbf{\Sigma}$.

**Proof of Lemma 2.** The quadratic term $\mathbf{a}'\mathbf{Q}\mathbf{a}$ can be expressed as

$$\mathbf{a}'\mathbf{Q}\mathbf{a} = \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{ij}\mathbf{a}_i'\mathbf{a}_j \,. \tag{11}$$

By Lemma 1, the random vector $\mathbf{a}_j$ in (11) can be replaced by the random vector $\tilde{\mathbf{a}}_j$ in (9). Then (11) becomes

$$\mathbf{a}'\mathbf{Q}\mathbf{a} = \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{ij}\rho_{a_{ij}} \frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{a}_i'\mathbf{a}_i + \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{ij}\sqrt{(1-\rho_{a_{ij}}^2)}\,\frac{\sigma_{a_j}}{\sigma_{a_i}}\mathbf{a}_i'\mathbf{z}_i.$$

Since $\mathbf{z}_i$ is independent of $\mathbf{a}_i$, we get

$$\mathbf{a}'\mathbf{Q}\mathbf{a} = \sum_{i=1}^{m} u_i \frac{\mathbf{a}_i'\mathbf{a}_i}{\sigma_{a_i}^2},$$

where $u_i = \sum_{j=1}^{m} \sigma_{ij}\rho_{a_{ij}}\sigma_{a_i}\sigma_{a_j}$. (12)

By Lemma 2, it is obvious that (7) and (8) can be respectively re-written as

$$(\mathbf{e}_r - \mathbf{e}_f)'\hat{\mathbf{\Omega}}_y^{-1}(\mathbf{e}_r - \mathbf{e}_f) = \sum_{i=1}^{m}\sum_{j=1}^{m} w_{1i}\mathbf{d}_i{}'\mathbf{d}_i, \tag{13}$$

$$tr(\hat{\mathbf{\Sigma}}_y^{-1}\mathbf{S}_{e_f}) = \sum_{i=1}^{m} w_{2i}\frac{\mathbf{e}_{f_i}'\mathbf{e}_{f_i}}{\sigma_{e_{f_i}}^2}, \tag{14}$$

where $w_{1i} = \sum_{j=1}^{m} \hat{\sigma}_y^{ij}\rho_{d_{ij}}\sigma_{d_i}\sigma_{d_j}$, (15)

and $w_{2i} = \sum_{j=1}^{m} \hat{\sigma}_y^{ij}\dfrac{\rho_{e_{f_{ij}}}\sigma_{e_{f_i}}\sigma_{e_{f_j}}}{n-p_f}$. (16)

Substituting (13) and (14) into (6) yields

$$D = \sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}_i'\mathbf{d}_i}{\sigma_{d_i}^2} - 2d\sum_{i=1}^{m} w_{2i}\frac{\mathbf{e}_{f_i}'\mathbf{e}_{f_i}}{\sigma_{e_{f_i}}^2}. \tag{17}$$

**Theorem 1.** Let $\mathbf{d}_i$ and $\mathbf{e}_{f_i}$ be distributed as $N_n(\mathbf{0},\sigma_{d_i}^2\mathbf{I}_n)$ and $N_n(\mathbf{0},\sigma_{e_{f_i}}^2\mathbf{I}_n)$, $w_{1i}$ and $w_{2i}$ be as defined in (15) and (16) respectively. Then the distribution of $\sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}_i'\mathbf{d}_i}{\sigma_{d_i}^2}$ can be approximated by $R_1 \sim \gamma_1\chi_{f_1}^2$, where $\gamma_1$ and $f_1$ are chosen so that the first two cumulants of the distributions are equal. Similarly, the distribution of $\sum_{i=1}^{m} w_{2i} \frac{\mathbf{e}'_{f_i}\mathbf{e}_{f_i}}{\sigma^2_{e_{f_i}}}$ can be approximated by $R_2 \sim \gamma_2 \chi^2_{f_2}$, under the same criteria, where

$$\gamma_1 = \frac{\sum_{i=1}^{m}\left(w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2\right)}{\sum_{i=1}^{m} w_{1i}},\ f_1 = \frac{n\left(\sum_{i=1}^{m} w_{1i}\right)^2}{\sum_{i=1}^{m}\left(w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2\right)},\ \gamma_2 = \frac{\sum_{i=1}^{m}\left(w_{2i}^2 + 2\sum_{j>i}^{m} w_{2i}w_{2j}\rho_{ef_{ij}}^2\right)}{\sum_{i=1}^{m} w_{2i}},$$

$$f_2 = \frac{n\left(\sum_{i=1}^{m} w_{2i}\right)^2}{\sum_{i=1}^{m}\left(w_{2i}^2 + 2\sum_{j>i}^{m} w_{2i}w_{2j}\rho_{ef_{ij}}^2\right)},$$

$\rho_{d_{ij}}$ is the correlation coefficient between $\mathbf{d}_i$ and $\mathbf{d}_j$, and $\rho_{e_{f_{ij}}}$ is the correlation coefficient between $\mathbf{e}_{f_i}$ and $\mathbf{e}_{f_j}$.

**Proof of Theorem 1.** Since $\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}$ is distributed as chi-squared with $n$ degrees of freedom, it is obvious that the expectation of $\sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}$ can be written as

$$E\left(\sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}\right) = n\sum_{i=1}^{m} w_{1i}\ .$$

The variance of $\sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}$ can be written as

$$Var\left(\sum_{i=1}^{m} w_{1i}\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}\right) = \sum_{i=1}^{m} w_{1i}^2 Var\left(\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}\right) + 2\sum_{i=1}^{m}\sum_{j>i}^{m} w_{1i}w_{1j}Cov\left(\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}, \frac{\mathbf{d}'_j\mathbf{d}_j}{\sigma^2_{d_j}}\right)$$

$$= 2n\sum_{i=1}^{m}\left(w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2\right),$$

since $Var\left(\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}\right) = 2n,\ i = 1, 2\ldots, m$ and $\rho_{d_{ij}}^2$ is the correlation coefficient of $\frac{\mathbf{d}'_i\mathbf{d}_i}{\sigma^2_{d_i}}$ and $\frac{\mathbf{d}'_j\mathbf{d}_j}{\sigma^2_{d_j}}$ (Isserlis, 1981).

The distribution of the weighted sum of correlated chi-squared variable $\sum_{i=1}^{m} w_{1i}\dfrac{\mathbf{d}_i'\mathbf{d}_i}{\sigma_{d_i}^2}$ can be approximated by $R_1 \sim \gamma_1\chi_{f_1}^2$ (Brown, 1975; Makambi, 2003 and Hou, 2005) where the first two cumulants of the two distributions are equal. Thus, equating the first two cumulants of $\sum_{i=1}^{m} w_{1i}\dfrac{\mathbf{d}_i'\mathbf{d}_i}{\sigma_{d_i}^2}$ and $R_1$ yields

$$\gamma_1 f_1 = n\sum_{i=1}^{m} w_{1i}, \tag{18}$$

$$2\gamma_1^2 f_1 = 2n\sum_{i=1}^{m}\left( w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2 \right). \tag{19}$$

Solving (18) and (19) yields

$$f_1 = \frac{n\left(\sum_{i=1}^{m} w_{1i}\right)^2}{\sum_{i=1}^{m}\left( w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2 \right)}, \tag{20}$$

$$\gamma_1 = \frac{\sum_{i=1}^{m}\left( w_{1i}^2 + 2\sum_{j>i}^{m} w_{1i}w_{1j}\rho_{d_{ij}}^2 \right)}{\sum_{i=1}^{m} w_{1i}}. \tag{21}$$

Similarly, it can be shown that

$$f_2 = \frac{n\left(\sum_{i=1}^{m} w_{2i}\right)^2}{\sum_{i=1}^{m}\left( w_{2i}^2 + 2\sum_{j>i}^{m} w_{2i}w_{2j}\rho_{e_{f_{ij}}}^2 \right)}, \tag{22}$$

$$\gamma_2 = \frac{\sum_{i=1}^{m}\left( w_{2i}^2 + 2\sum_{j>i}^{m} w_{2i}w_{2j}\rho_{e_{f_{ij}}}^2 \right)}{\sum_{i=1}^{m} w_{2i}}, \tag{23}$$

where $\rho_{e_{f_{ij}}}^2$ is the correlation coefficient of $\dfrac{\mathbf{e}_{f_i}'\mathbf{e}_{f_i}}{\sigma_{ef_i}^2}$ and $\dfrac{\mathbf{e}_{f_j}'\mathbf{e}_{f_j}}{\sigma_{ef_j}^2}$ (Isserlis, 1981). □

**Lemma 3.** $\mathbf{d}'\mathbf{W}_1\mathbf{d}$ and $\mathbf{e}_f'\mathbf{W}_2\mathbf{e}_f$ are independent where $\mathbf{d}$ and $\mathbf{e}_f$ are the $nm\times 1$ vectors consisting of $m$ subvectors $\mathbf{d}_i$, distributed as $N_n(\mathbf{0},\sigma_{d_i}^2\mathbf{I}_n)$, and $m$ subvectors $\mathbf{e}_{f_i}$, distributed as $N_n(\mathbf{0},\sigma^2_{e_{f_i}}\mathbf{I}_n)$ respectively, $i=1,2,...,m$, $\mathbf{W}_1$ and $\mathbf{W}_2$ are the $nm\times nm$ diagonal matrices with the $i^{th}$ diagonal element $\frac{w_{1i}}{\sigma^2_{d_i}}$ and $\frac{w_{2i}}{\sigma^2_{e_{f_i}}}$ respectively.

**Proof of Lemma 3.** The residual vectors $\mathbf{e}_{f_i}$ and $\mathbf{e}_{r_i}$ in the full model and in the reduced model can be respectively expressed in term of $\mathbf{y}_i$ (Graybill, 1976),

$$\mathbf{e}_{f_i}=(\mathbf{I}_n-\mathbf{M}_f)\mathbf{y}_i, \quad (24)$$

$$\mathbf{e}_{r_i}=(\mathbf{I}_n-\mathbf{M}_r)\mathbf{y}_i, \quad (25)$$

where $\mathbf{M}_f=\mathbf{X}_f(\mathbf{X}'_f\mathbf{X}_f)^{-1}\mathbf{X}'_f$ and $\mathbf{M}_r=\mathbf{X}_r(\mathbf{X}'_r\mathbf{X}_r)^{-1}\mathbf{X}'_r$. From (24) and (25), the difference vector $\mathbf{d}_i$ can be written as

$$\mathbf{d}_i=(\mathbf{M}_f-\mathbf{M}_r)\mathbf{y}_i. \quad (26)$$

Consider the term $\mathbf{M}_r\mathbf{M}_f$ by partitioning the matrix $\mathbf{X}_f$ into $\mathbf{X}_r$ and $\mathbf{X}_d$.

$$\mathbf{M}_r\mathbf{M}_f=\mathbf{X}_r(\mathbf{X}'_r\mathbf{X}_r)^{-1}\mathbf{X}'_r\begin{bmatrix}\mathbf{X}_r & \mathbf{X}_d\end{bmatrix}\begin{bmatrix}(\mathbf{X}'_r\mathbf{X}_r)^{-1} & \mathbf{0}\\ \mathbf{0} & (\mathbf{X}'_d\mathbf{X}_d)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{X}_r'\\ \mathbf{X}'_d\end{bmatrix}=\mathbf{M}_r, \quad (27)$$

since $\mathbf{X}_r$ and $\mathbf{X}_d$ are independent. Similarly,

$$\mathbf{M}_f\mathbf{M}_r=\mathbf{M}_r. \quad (28)$$

Then, by (27), (28) and the idempotent property of $\mathbf{M}_f$ and $\mathbf{M}_r$, the quadratic term $\mathbf{d}'\mathbf{W}_1\mathbf{d}$ and $\mathbf{e}'_f\mathbf{W}_2\mathbf{e}_f$ can be expressed in term of y as

$$\mathbf{d}'\mathbf{W}_1\mathbf{d}=\sum_{i=1}^{m}\frac{w_{1i}}{\sigma^2_{d_i}}\mathbf{y}'_i(\mathbf{M}_f-\mathbf{M}_r)(\mathbf{M}_f-\mathbf{M}_r)\mathbf{y}_i$$

$$=\sum_{i=1}^{m}\frac{w_{1i}\sigma^2_{y_i}}{\sigma^2_{d_i}}\tilde{\mathbf{y}}'_i(\mathbf{M}_f-\mathbf{M}_r)\tilde{\mathbf{y}}_i, \quad (29)$$

$$\mathbf{e}'_f\mathbf{W}_2\mathbf{e}_f=\sum_{i=1}^{m}\frac{w_{2i}}{\sigma^2_{f_i}}\mathbf{y}'_i(\mathbf{I}_n-\mathbf{M}_f)(\mathbf{I}_n-\mathbf{M}_f)\mathbf{y}_i$$

$$=\sum_{i=1}^{m}\frac{w_{2i}\sigma^2_{y_i}}{\sigma^2_{f_i}}\tilde{\mathbf{y}}'_i(\mathbf{I}_n-\mathbf{M}_f)\tilde{\mathbf{y}}_i, \quad (30)$$

where $\tilde{\mathbf{y}}_i$ is a standardized form of $\mathbf{y}_i$ with variance $\mathbf{I}_n$. From (29) and (30), it can be said that if $\tilde{\mathbf{y}}'_i(\mathbf{M}_f-\mathbf{M}_r)\tilde{\mathbf{y}}_i$ and $\tilde{\mathbf{y}}'_i(\mathbf{I}_n-\mathbf{M}_f)\tilde{\mathbf{y}}_i$ are independent, then $\mathbf{d}'\mathbf{W}_1\mathbf{d}$ and $\mathbf{e}'_f\mathbf{W}_2\mathbf{e}_f$ are independent. Again, by (27) and the idempotent property of $\mathbf{M}_f$, it can be concluded that

$$(\mathbf{M}_f-\mathbf{M}_r)(\mathbf{I}_n-\mathbf{M}_f)=\mathbf{0},$$

which leads to the conclusion that $\tilde{\mathbf{y}}_i'(\mathbf{M}_f - \mathbf{M}_r)\tilde{\mathbf{y}}_i$ and $\tilde{\mathbf{y}}_i'(\mathbf{I}_n - \mathbf{M}_f)\tilde{\mathbf{y}}_i$ are independent (Graybill, 1976). □

**Theorem 2.** The expectation and the variance of the statistic $D$ can be expressed as

$$\mu_D = n\sum_{i=1}^{m}\left[\hat{\sigma}_y^{ii}\left(\sigma_{d_i}^2 - \frac{2d}{n-p_f}\sigma_{ef_i}^2\right) + 2\sum_{j>i}^{m}\hat{\sigma}_y^{ij}\left(\rho_{d_{ij}}\sigma_{d_i}\sigma_{d_j} - \frac{2d}{n-p_f}\rho_{ef_{ij}}\sigma_{ef_i}\sigma_{ef_j}\right)\right],$$

$$\sigma_D^2 = 2n\sum_{i=1}^{m}\left[\sigma_{d_i}^2 S_{d_i}^2 + \left(\frac{2d\sigma_{e_{f_i}}}{n-p_f}\right)^2 S_{e_{f_i}}^2\right]$$

$$+4n\sum_{i=1}^{m}\sum_{j>i}^{m}\left(\sigma_{d_i}\sigma_{d_j}\rho_{d_{ij}}^2 S_{d_i}S_{d_j} + \left(\frac{2d}{n-p_f}\right)^2\sigma_{e_{f_i}}\sigma_{e_{f_j}}\rho_{ef_{ij}}^2 S_{ef_i}S_{ef_j}\right),$$

where $S_{d_i} = \sum_{k=1}^{m}\hat{\sigma}_y^{ik}\rho_{d_{ik}}\sigma_{d_k}$, $S_{d_j} = \sum_{k=1}^{m}\hat{\sigma}_y^{jk}\rho_{d_{jk}}\sigma_{d_k}$, $S_{ef_i} = \sum_{k=1}^{m}\hat{\sigma}_y^{ik}\rho_{ef_{ik}}\sigma_{ef_k}$ and $S_{ef_j} = \sum_{k=1}^{m}\hat{\sigma}_y^{jk}\rho_{ef_{jk}}\sigma_{ef_k}$.

**Proof of Theorem 2.** By Theorem 1, the expectation of the statistic $D$ in (17) can be expressed as

$$\mu_D = \gamma_1 f_1 - 2d\gamma_2 f_2. \tag{31}$$

Substituting $\gamma_1, f_1, \gamma_2$ and $f_2$ from (20)-(23) into (31) gives

$$\mu_D = n\sum_{i=1}^{m}\left[w_{1i} - 2dw_{2i}\right]. \tag{32}$$

Substituting (15) and (16) into (32) gives

$$\mu_D = n\sum_{i=1}^{m}\left[\sum_{j=1}^{m}\hat{\sigma}_y^{ij}\rho_{d_{ij}}\sigma_{d_i}\sigma_{d_j} - 2d\sum_{j=1}^{m}\hat{\sigma}_y^{ij}\frac{\rho_{e_{f_{ij}}}\sigma_{e_{f_i}}\sigma_{e_{f_j}}}{n-p_f}\right]$$

$$= n\sum_{i=1}^{m}\left[\hat{\sigma}_y^{ii}\left(\sigma_{d_i}^2 - \frac{2d}{n-p_f}\sigma_{ef_i}^2\right) + 2\sum_{j>i}^{m}\hat{\sigma}_y^{ij}\left(\rho_{d_{ij}}\sigma_{d_i}\sigma_{d_j} - \frac{2d}{n-p_f}\rho_{ef_{ij}}\sigma_{ef_i}\sigma_{ef_j}\right)\right]. \tag{33}$$

By Theorem 1 and Lemma 3, the variance of statistic $D$ can be written as

$$\sigma_D^2 = 2f_1\gamma_1^2 + 8d^2 f_2\gamma_2^2. \tag{34}$$

Again, substituting $\gamma_1, f_1, \gamma_2$ and $f_2$ from (20)-(23) into (34) yields

$$\sigma_D^2 = 2n\sum_{i=1}^{m}\left[w_{1i}^2 + 4d^2 w_{2i}^2 + 2\sum_{j>i}^{m}\left(w_{1i}w_{1j}\rho_{d_{ij}}^2 + 4d^2 w_{2i}w_{2j}\rho_{e_{f_{ij}}}^2\right)\right]. \tag{35}$$

Substituting (15) and (16) into (35) gives

$$\sigma_D^2 = 2n\sum_{i=1}^{m}\left[\sigma_{d_i}^2 S_{d_i}^2 + \left(\frac{2d\sigma_{e_{f_i}}}{n-p_f}\right)^2 S_{e_{f_i}}^2\right] + 4n\sum_{i=1}^{m}\sum_{j>i}^{m}\left(\sigma_{d_i}\sigma_{d_j}\rho_{d_{ij}}^2 S_{d_i}S_{d_j} + \left(\frac{2d}{n-p_f}\right)^2 \sigma_{e_{f_i}}\sigma_{e_{f_j}}\rho_{ef_{ij}}^2 S_{ef_i}S_{ef_j}\right), \quad (36)$$

where $S_{d_i}$, $S_{d_j}$, $S_{ef_i}$ and $S_{ef_j}$ are as defined in Theorem 2.

**Theorem 3.** The distribution of the statistic $D$ is a gamma distribution with shape parameter $\mu_D^2/\sigma_D^2$ and scale parameter $\sigma_D^2/\mu_D$.

**Proof of Theorem 3.** Since the statistic $D$ in (17) is a linear combination of two independent Chi-squared variables, the distribution of $D$ can be approximated by $\gamma\chi_f^2$ where $f$ and $\gamma$ are chosen so that the first two cumulants of the two distributions are equal (Brown, 1975; Makambi, 2003 and Hou, 2005). By Theorem 2, it is obvious that $f = 2\mu_D^2/\sigma_D^2$ and $\gamma = \sigma_D^2/(2\mu_D)$. Therefore, it can be concluded that the distribution of the statistic $D$ is a gamma distribution with shape parameter $\mu_D^2/\sigma_D^2$ and scale parameter $\sigma_D^2/\mu_D$. □

By Theorem 2, the standardized statistic $D$ can be written as $T_D = \frac{D-\mu_D}{\sigma_D}$ which, from the central limit theorem, converges to N(0,1) as $n \to \infty$.

## Conclusions

In this paper, three theorems and three lemmas are proved to provide the statistical properties of the statistic $D$ which can be used in variable selection in the contemporaneous multivariate linear regression nested model. This is another alternative of variable selection by testing the hypothesis (4) instead of direct comparison of the modified $C_p$ statistics. Without the hypothesis testing, the backward procedure of variable elimination suggests to retain any variable in the model specification if its elimination results in the modified $C_p$ statistic higher than $MC_f$ even in the case of slight increase. But by hypothesis testing, a variable can be eliminated from the model specification if its elimination does not cause the rejection of the null hypothesis. It can be concluded that the number of independent variables in the final model by using the hypothesis testing in variable elimination is at most equal to the number of independent variables in the final model by the conventional variable elimination. The proposed concept can be easily extended for variable selection by the forward and stepwise procedures of variable selection.

## References

[1] Brown, M.B., 1975, A method for combining non-independent, one-sided tests of significance. *Biometrics,* 31, 987-992.

[2] Fujikoshi, Y., Satoh, K., 1997, Modified AIC and $C_p$ in multivariate Linear Regression. *Biometrika*, 84(3), 707-716.

[3] Graybill, F.A., *Theorem and Application of the Linear model*. Duxbury, Canada, 1976.

[4] Hou, C.D., 2005, A simple approximation for the distribution of the weighted combination of nonindependent or independent probabilities. *Statistics and Probability Letters*, 73, 179-187.

[5] Isserlis, L., 1981, On a formula for the product-moment coefficient of any order of a normal frequency distribution on in any number of variables. *Biometrika*, 12, 134-139.

[6] Lorchirachoonkul, V., Jitthavech, J., 2012, A modified $C_p$ statistic in a system-of-equations model. *Journal of Statistical Planning and Inference*,142, 2386-2394.

[7] Makambi, K.H., 2003, Weighted inverse chi-square method correlated significance tests. *Journal of Applied Statistics*, 30, 225-234.

[8] Sparks, R.S., Troskie, L., Coutsourides, D., 1983, The multivariate $C_p$. *Communications in Statistics-Theory and Methods*, 12(15), 1775-1793.

[9] Yanagihara, H., Satoh, K., 2010, An unbiased $C_p$ criterion for multivariate ridge regression. *Journal of Multivariate Analysis*, 101, 1226-1238.