An Analytic Comparison of Permutation Methods for Tests of Partial Regression Coefficients in the Linear Model

Ali SHADROKH

Department of Statistics Science
University of Payam Noor of Thiran, Iran
a_shadrokh@pnu.ac.ir

Gérard d’AUBIGNY

Statistics department, Laboratoire Jean Kuntzman
University of Grenoble, France
Gerard.d-Aubigny@upmf-grenoble.fr

Abstract

Several method of permutation tests have been proposed for testing nullity of a partial regression coefficient in a linear model. These methods were compared in terms of empirical type I error and power by statisticians. One striking result of the simulation based comparison is that the two emerging methods, while previously identified as equivalent formulations of the permutation strategy under the reduced model, did actually produce quite different results. And one of this methods have almost the best result. Some theoretical justification to the empirical findings is given here. We compared estimators and variances of these two methods analytically in the double regression linear model. Our results give mathematical support to the observation obtained by simulation. Furthermore, for the first time we obtained the expected value of the estimators of the variance by the permutaional distribution and we found that one of these estimators is unbiased.

Keywords: Permutation test, p-value permutational, permutational distribution, partial regression
1 Introduction

Let us consider the multiple-regression linear model. Without loss of generality, the situation of interest may be illustrated by the two predictors case:

\[ Y = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \varepsilon \]  

(1)

where one tests for an individual coefficient, say \( H_0 : \beta_2 = 0 \) against the (two-sided) alternative \( H_1 : \beta_2 \neq 0 \). This null hypothesis says that the predictor \( X^2 \) contributes nothing to the multiple linear regression model.

The classical parametric setting makes the following Gaussian assumptions for residuals \( \varepsilon \):

1. a sample of \( n \) observations of the joint distribution of \( (Y, X^1, X^2) \) is available,
2. \( \{ \varepsilon_i, i = 1, 2, ..., n \} \) are i.i.d. and Laplace-Gauss distributed: \( \varepsilon_i \sim LG(0, \sigma^2) \).

When these assumptions are met, the familiar ratio of the estimated coefficient to its standard error (or t-ratio) yields a test statistic for this null hypothesis:

\[ T = (\hat{\beta}_2 - 0)/s(\hat{\beta}_2) \sim St(n - k - 1) \]  

(2)

where \( \hat{\beta}_2 \) is the ordinary least squares (OLS) estimator of the regression coefficient \( \beta_2 \), \( s(\hat{\beta}_2) \) is the standard error of this estimator of \( \beta_2 \), and \( k \) is the number of predictors present in the model. The distribution of \( T \) under \( H_0 \), is a Student law with \( (n - k - 1) \) degrees of freedom.

When the random error term \( \varepsilon \) fails to fulfill some of these assumptions, one needs resort to some other method to carry out statistical inferences. An often used method uses a Gaussian approximation available in the case of large samples. But it gets justified only under the central limit theorem postulates, and a large enough sample size \( n \) is mandatory. In the small sample and non-Gaussian case, one way of proceeding consists in resorting to statistical tests based on the permutation distribution of some statistic like the usual t-ratio.

The first descriptions of permutation tests for linear statistical models, including analysis of variance and regression, can be traced back to the work of Fisher (1935), and Pitman (1937)[11]. Such randomization tests did not receive much attention in practice until much later, when the emergence of widely accessible computer power made feasible such computer intensive methods, by comparison to the traditional normal-theory corresponding tests (Edgington 1995[3]; Manly 1997[8]; Good 2000[5]).
Moreover, while there is general agreement concerning an appropriate method of permutation for exact tests of hypotheses in simple linear regression - see Edgington (1995) and Manly (1991) - this is not so for tests of individual coefficients in the context of multiple linear regression. Several permutation methods have been proposed for testing the significance of each partial regression coefficient in a multiple regression model. They differ on the facet of the classical regression model which is used as a base for designing a permutation strategy, see for example: Freedman and Lane (1983)[4]; Smouse et al. (1986)[13]; Oja (1987)[10]; Manly (1991)[9]; ter Braak (1992)[14]; Kennedy (1995)[7].

Anderson and Legendre (1999)[1] used simulations to compare four competing permutation tests in terms of type I error and power. These methods differ in the type of permutation adopted: Manly (1991)[9] proposes to permute raw data while Freedman and Lane (1983)[4] and Kennedy (1995)[7]) retain permutation of residuals under the reduced model, a fourth strategy is adopted by ter Braak (1992), who considers permutations under the full model.

These authors restricted their attention to methods which do not ignore a potential collinearity between the predictor variables (cf. [1]). They investigated effects of (a) the sample size, (b) the degree of collinearity between the predictor variables, (c) the size of the covariable’s parameter; and (d) the distribution of the added random error. Their simulations revealed that two of these methods (Freedman and Lane; Kennedy), previously considered as equivalent formulations of permutation under the reduced model produced in fact different results.

The type I error resulting from the Freedman and Lane method is always lower than the one obtained by the Kennedy’s method, but both are consistently inflated. No significant differences in power were noticed among the permutation methods favored respectively by Freedman and Lane, Manly and ter Braak, but all had greater power than the normal-theory t-test when error were non-normal. Notice that the case of the Kennedy’s method has not been considered in this comparison of power behavior of the test. The reduced model permutation method (Freedman and Lane’s method) presented the most consistent and reliable results among the methods investigated by Anderson and Legendre (1999) for the permutation test of a partial regression coefficient.

Our study focused on the comparison of the contributions of Kennedy, and of Freedman and Lane respectively. We compared theoretically the resulting estimators, their expected values and variances. Our results yield a mathematical foundation to the empirical findings obtained by Anderson and Legendre. Indeed, we showed that the permutational power function of the method of Freedman and Lane is lower or equal to that of one of Kennedy. Thus, type I error of the Freedman
and Lane method is lower or equal to that obtained by the method of Kennedy. Moreover, the power of the Freedman and Lane method is lower or equal to that obtained by Kennedy.

Furthermore we obtained the permutational variance of the estimator $\hat{\beta}_2$ denoted by $\mathsf{Var}(\hat{\beta}_2)$ and we shows that the approach of Freedman and Lane derives an estimator of $\mathsf{Var}(\hat{\beta}_2)$ denoted by $s^2(\hat{\beta}_2)_F$, which is a permutational unbiased estimator of $\mathsf{Var}(\hat{\beta}_2)$, contrasting with to the estimator $s^2(\hat{\beta}_2)_K$ given by the Kennedy's method.

2 Notations and description of methods

By convention, this paper uses $\pi$ as a subscript of a statistic in a way to denote its transformation by permutation of its sampled values. For example, the notation $Y_\pi$ indicates that the values of a variable $Y$ have been permuted. Complementarity, $\pi$ is placed as a superscript - as in $T^\pi$ - when the statistic $T$ is derived as a function of a set of arguments containing at least one permutated variable.

2.1 The OLS estimators

We follow the presentation given by Gujarati (2004, pp 212-214). First, let us remind the classical linear regression model in the case of two predictors: $Y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \varepsilon$ The OLS estimators are obtained in minimizing the loss function $Q$ defined as the empirical sum of squares:

$$ Q = \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2)^2 $$

$$ = \|\epsilon\|^2 = \|Y - \hat{X}\beta\|^2 \quad \beta = \hat{\beta}((\beta_0, \beta_1, \beta_2), x_i^0 = 1 \forall i = 1, \cdots, n$$

The easiest way to built a solution consists in deriving $Q$ in terms of $\beta$, so as to get the system of three normal equations:

$$ \frac{\partial Q}{\partial \beta} = 0 \Leftrightarrow \begin{cases} \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i^1 + \hat{\beta}_2 x_i^2, \\ \sum_{i=1}^{n} Y_i x_i^1 = \hat{\beta}_0 \sum_{i=1}^{n} x_i^1 + \hat{\beta}_1 \sum_{i=1}^{n} (x_i^1)^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^1 x_i^2, \\ \sum_{i=1}^{n} Y_i x_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} x_i^1 x_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} (x_i^2)^2. \end{cases} $$

The first equation in this system fixes the estimator $\hat{\beta}_0$ of the intercept as a function of data and of $\hat{\beta}_1$ and $\hat{\beta}_2$. The two other equations simplify by substitution of $\hat{\beta}_0$, \ldots
and using the dot convention to denote empirically centered variables:
\[
\sum_{i=1}^{n} \hat{Y}_i \hat{x}_i^1 = \hat{\beta}_1 \sum_{i=1}^{n} (\hat{x}_i^1)^2 + \hat{\beta}_2 \sum_{i=1}^{n} \hat{x}_i^1 \hat{x}_i^2, \tag{5}
\]
\[
\sum_{i=1}^{n} \hat{Y}_i \hat{x}_i^2 = \hat{\beta}_1 \sum_{i=1}^{n} \hat{x}_i^1 \hat{x}_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} (\hat{x}_i^2)^2. \tag{6}
\]
This system may be written more compactly by defining
\[
s_j^2 = \sum_{i=1}^{n} (\hat{x}_i^j)^2, \quad s_{Yj} = \sum_{i=1}^{n} \hat{Y}_i \hat{x}_i^j \quad j = 1, 2, \quad s_{12} = \sum_{i=1}^{n} \hat{x}_i^1 \hat{x}_i^2, \quad \Delta = s_1^2 s_2^2 - s_{12}^2. \tag{7}
\]
Thus, \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are obtained by inverting this system (cf. [6]):
\[
\hat{\beta}_1 = (s_{Y1} s_2^2 - s_{Y2} s_{12}) / \Delta, \quad \hat{\beta}_2 = (s_{Y2} s_1^2 - s_{Y1} s_{12}) / \Delta \tag{8}
\]
Moreover, these expressions express \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) as two deterministic linear functions of the vector \( \hat{Y} \). So, one can easily derive the sampling moments of these estimators. Namely, they are unbiased and if \( \sigma^2 \) denotes the variance of \( \varepsilon \),
\[
\var(\hat{\beta}_1) = \sigma^2 s_2^2 / \Delta, \quad \var(\hat{\beta}_2) = \sigma^2 s_1^2 / \Delta, \quad \cov(\hat{\beta}_1, \hat{\beta}_2) = -\sigma^2 s_{12} / \Delta \tag{9}
\]
These expressions show that the variances of these estimators as well as their covariance depend on the strength of correlation between predictors, since. \( \Delta^2 = s_1 s_2 (1 - r_{12}^2) \). Gujarati observes too (page 238, ex. 7.3) that the estimator 8 may be written in an equivalent way:
\[
\hat{\beta}_1 = \sum_{i=1}^{n} \hat{Y}_i (\hat{x}_i^1 - b_{12} \hat{x}_i^2) / \sum_{i=1}^{n} (\hat{x}_i^1 - b_{12} \hat{x}_i^2)^2 \text{ which uses the estimation } b_{12} \text{ of the slope parameter in the simple linear regression of } \hat{X}^1 \text{ on } \hat{X}^2, \quad \text{i.e. } b_{12} = s_{12} / s_{22}. \text{ Of interest too for the sequel is the classical decomposition of the simple linear regression coefficient as a sum of direct and indirect relations, see e.g. Wonnacott and Wonnacott (1979, pp 95-97),}
\]
\[
\hat{\beta}_{Y1} = \hat{\beta}_1 + \hat{\beta}_2 \hat{\beta}_{21} \tag{10}
\]
where \( \hat{\beta}_{Y1} \) (respectively \( \hat{\beta}_{21} \)) is the OLS estimator of the simple regression coefficient in the linear regression of \( Y \) (resp. of \( X^2 \)) on \( X^1 \), while \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are the partial regression coefficients of \( X^1 \) and \( X^2 \) in the bivariate linear regression of \( Y \) on these two regressors.

By convention, this paper uses \( \pi \) as a subscript of a statistic in a way to denote its transformation by permutation of its sampled values. For example, the notation \( Y^\pi \) indicates that the values of a variable \( Y \) have been permuted. Complementarity, \( \pi \) is placed as a superscript - as in \( T^\pi \) - when the statistic \( T \) is derived as a function of a set of arguments containing at least one permuted variable.
2.2 The Shuffle-all-predictors approach

Freedman and Lane method (1983) proposed a testing based on permutations of the residuals calculated under the reduced model of the full linear regression one. More precisely, Freedman and Lane suggested to employ the following procedure for bivariate regression:

1. The variable $Y$ is regressed - by an ordinary least squares (or OLS) method - on the full set of predictor variables (here $X^1$ et $X^2$) in a way to derive the OLS estimates of the parameters of the model. One particularly obtains an estimator $\hat{\beta}_2$ of the parameter $\beta_2$. A $t$-ratio is then calculated on the observed sample for testing the hypothesis $H_0 : \beta_2 = 0$. This produces the reference value $t_{\text{ref}} = b_2/s(\hat{\beta})$, where $b_2$ is the value of $\hat{\beta}$ on the observed sample.

2. The variable $Y$ is regressed on $X^1$ alone according to the model

$$Y = a_0 + \alpha_1 x^1 + \epsilon_{Y,1}$$

(11)

to obtain the prediction $\hat{Y}_{Y,1} = a_0 + \alpha_1 X^1$ and values of the residuals $\epsilon_{Y,1}$, where $a_0$ and $\alpha_1$ are the OLS estimates of $a_0$ and $\alpha_1$ respectively.

3. The residuals $\epsilon_{Y,1}$ from the regression Step 2 are permuted, producing $\epsilon_{Y,1}^\pi$.

4. A new pseudo-variable $Y^{F,\pi}$ is calculated by adding the permuted values of the residuals to the fitted values obtained in Step 2, namely,

$$Y^{F,\pi} = b_0 + b_1 x^1 + b_2 x^2 + \epsilon_{Y,1}^\pi.$$  

5. The new variable $Y^{F,\pi}$ is regressed on all predictor variables simultaneously according to model

$$Y^{F,\pi} = \beta_0^{F,\pi} + \beta_1^{F,\pi} X^1 + \beta_2^{F,\pi} X^2 + \epsilon^\pi$$

(12)

to obtain the predicted equation $Y^{F,\pi} = b_0^{F,\pi} + b_1^{F,\pi} x^1 + b_2^{F,\pi} x^2$ and a value $t_{2}^{F,\pi} = b_2^{F,\pi}/s(\hat{\beta}_2^{F,\pi})$ of the $t$-ratio for testing $H_0 : \beta_2 = 0$. Where, $b_2^{F,\pi}$ et $s(\hat{\beta}_2^{F,\pi})$ are respectively the OLS estimates for $\beta_2^{F,\pi}$ et $\sigma(\hat{\beta}_2^{F,\pi})$ on the permuted data.

6. Steps 2 through 5 are repeated a large number of time until obtain a permutation distribution of $t^{F,\pi}$.

7. For a two-tailed test, the absolute value of the reference value $t_{\text{ref}}$ is placed in the permutation distribution of absolute value of $t_{\text{ref}}$. The p-value is calculated as the proportion of values in this distribution greater than or equal to the absolute value of $t_{\text{ref}}$. 

2.2.1 Statistical properties of the reduced model permutational tests

**Definition 2.1.**

1) \( s(X^k, Y^F) = \frac{1}{n} \sum_i x_i^k (Y^{F(i)} - \hat{Y}^F) \)
2) \( s(X^k, e^{Y1}) = \frac{1}{n} \sum_i x_i^k e^{Y1(i)} \), \( k = 1, 2 \)
3) \( \Delta_1 = s_2^2 s(X^1, e^{Y1}) - s(X^1, e^{Y1})s_{12} \)
4) \( \Delta_2 = s_1^2 s(X^2, e^{Y1}) - s(X^2, e^{Y1})s_{12} \)
5) \( a_1 = s_{Y1}/s_1^2 \) (from 11)

**Property 2.1.** The Freeman and Lane’s approach characterized by the regression model:

\[
Y^{F,\pi(i)} = \beta_0^{F,\pi} + \beta_1^{F,\pi} x_1^i + \beta_2^{F,\pi} x_2^i + \epsilon_i. \tag{13}
\]

verifies the following identities:

1) \( \hat{Y}^{F,\pi} = \bar{Y} \)
2) \( Y^{F,\pi(i)} - \hat{Y}^{F,\pi} = a_1(x_i^1 - \bar{x}^1) + e^{Y1(i)} \) \( \forall i = 1, 2 \)
3) \( s(X^j, Y^{F,\pi}) = a_1 s(X^1, X^j) + s(X^j, e^{Y1}) \) \( j = 1, 2 \)

**Lemma 2.1.** The OLS estimator of regression coefficient of Freedman and Lane model (13) verify:

1) \( \hat{\beta}_1^{F,\pi} = a_1 + \Delta_1/\Delta \)
2) \( \hat{\beta}_2^{F,\pi} = \Delta_2/\Delta \)
3) \( \hat{\beta}_0^{F,\pi} = a_0 - (\Delta_1\bar{x}^1 + \Delta_2\bar{x}^2)/\Delta \)

**Proof:**

1) The estimator OLS for coefficient regression \( \beta_1^{F,\pi} \) is (by 8):

\[
\hat{\beta}_1^{F,\pi} = s_2^2 s(X^1, Y^{F,\pi}) - s(X^2, Y^{F,\pi})s_{12}/\Delta. \]

According to property (2.1) and by substitution of \( s(X^1, Y^{F,\pi}) \), \( s(X^2, Y^{F,\pi}) \) one obtains: \( \hat{\beta}_1^{F,\pi} = a_1 + \Delta_1/\Delta \)

2) In an analogous way OLS estimator of \( \hat{\beta}_2^{F,\pi} \) (by 8) is:

\[
\hat{\beta}_2^{F,\pi} = [s_1^2 s(X^2, Y^{F,\pi}) - s(X^1, Y^{F,\pi})s_{12}]/\Delta = \Delta_2/\Delta.
\]

3) It is known that OLS estimator of \( \beta_0^{F,\pi} \) is: \( \hat{\beta}_0^{F,\pi} = \bar{Y} - \hat{\beta}_1^{F,\pi}\bar{x}^1 - \hat{\beta}_2^{F,\pi}\bar{x}^2 \). From the \( \hat{\beta}_1^{F,\pi} \) and \( \hat{\beta}_2^{F,\pi} \) one obtains by substitution: \( \hat{\beta}_0^{F,\pi} = a_0 - (\Delta_1\bar{x}^1 + \Delta_2\bar{x}^2)/\Delta \).

The model of regression of the variable \( e^{Y1(i)} \) en \( x^1 \) et \( x^2 \):

\[
e^{Y1(i)} = \beta_0^1 + \beta_1^1 x_1^i + \beta_2^1 x_2^i + \epsilon_i \quad (i = 1, 2, ..., n) \tag{14}
\]

where \( \text{Var}(\epsilon_i|x^1, x^2) = (\sigma')^2 \) and \( \mathbb{E}(\epsilon_i|x^1, x^2) = 0 \), verify:

**Proposition 2.1.** 1. the OLS estimators of \( \beta_0^{F,\pi} \), \( \beta_1^{F,\pi} \) et \( \beta_2^{F,\pi} \) are:

a) \( \hat{\beta}_2^{F,\pi} = \Delta_2/\Delta = \hat{\beta}_2^2 \)

b) \( \hat{\beta}_1^{F,\pi} = \Delta_1/\Delta = \hat{\beta}_1^{F,\pi} - a_1 \)

c) \( \hat{\beta}_0^{F,\pi} = -(\Delta_1\bar{x}^1 + \Delta_2\bar{x}^2)/\Delta = \hat{\beta}_0^{F,\pi} - a_0 \)

2. The OLS observed residuals of the model (14) are equal to the residuals of the model of Freedman and Lane, i.e.: \( \hat{\epsilon} = \hat{\epsilon}^F \)
3. The OLS estimators of variance the estimators of regression coefficients the model (14) are equal to the OLS estimators of variance the estimators of regression coefficients the Freedman and Lane model, i.e.:

\[ \overline{\text{Var}}(\hat{\beta}_k^{\pi}) = \overline{\text{Var}}(\hat{\beta}_k^{\pi}) \quad k = 0, 1, 2 \]

**Proof:** In ([12])

**Lemma 2.2.** The residuals of Freedman and Lane model (13) verify:

\[
\sum_{i=1}^{n} (\hat{\epsilon}_i^F)^2 = \sum_{i=1}^{n} (e_{x(i)}^{Y_1})^2 + ns_1^2 \Delta_1^2 / \Delta^2 + ns_2^2 \Delta_2^2 / \Delta^2 + 2ns_{12} \Delta_1 \Delta_2 / \Delta^2 \\
- 2ns(X_1, e_{x(i)}^{Y_1}) \Delta_1 / \Delta + s(X_2, e_{x(i)}^{Y_1}) \Delta_2 / \Delta \]

**Proof:** Since \( \hat{Y}_i^1 = \hat{Y} + a_1(x_i^1 - \bar{x}^1) \), the Freedman and Lane model allows to obtain

\[ \hat{\epsilon}_i^F = Y^{F,i}_\pi - \hat{Y}_\pi^F = (\hat{Y}_i^1 + e_{x(i)}^{Y_1}) - (\hat{\beta}_0 + \hat{\beta}_1 x_i^1 + \hat{\beta}_2 x_i^2) = e_{x(i)}^{Y_1} + \hat{Y} + a_1(x_i^1 - \bar{x}^1) - \hat{\beta}_0 - \hat{\beta}_1 x_i^1 - \hat{\beta}_2 x_i^2 = e_{x(i)}^{Y_1} + a_1(x_i^1 - \bar{x}^1) - \hat{\beta}_1(x_i^1 - \bar{x}^1) - \hat{\beta}_2(x_i^2 - \bar{x}^2) \]

since \( \hat{Y}_\pi^F = \hat{Y} \) and in addition \( \hat{\beta}_0 = \hat{Y} - \hat{\beta}_1 \bar{x}^1 - \hat{\beta}_2 \bar{x}^2 \). But we have:

\[ \hat{\beta}_1^{F,i} = a_1 + \Delta_1 / \Delta \text{ and } \hat{\beta}_2^{F,i} = \Delta_2 / \Delta. \]

Consequently:

\[ \hat{\epsilon}_i^F = e_{x(i)}^{Y_1} - \Delta_1 / \Delta (x_i^1 - \bar{x}^1) - \Delta_2 / \Delta (x_i^2 - \bar{x}^2) \]

Thus:

\[
\sum_{i=1}^{n} (\hat{\epsilon}_i^F)^2 = \sum_{i=1}^{n} (e_{x(i)}^{Y_1})^2 + \Delta_1^2 / \Delta^2 \sum_{i=1}^{n} (x_i^1 - \bar{x}^1)^2 + \Delta_2^2 / \Delta^2 \sum_{i=1}^{n} (x_i^2 - \bar{x}^2)^2 \\
+ 2(\Delta_1 \Delta_2 / \Delta^2) \sum_{i=1}^{n} (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2) - 2\Delta_1 / \Delta \sum_{i=1}^{n} (x_i^1 - \bar{x}^1)e_{x(i)}^{Y_1} \\
- 2\Delta_2 / \Delta \sum_{i=1}^{n} (x_i^2 - \bar{x}^2)e_{x(i)}^{Y_1}; \text{ and since: } \sum_{i=1}^{n} (e_{x(i)}^{Y_1})^2 = \sum_{i=1}^{n} (e_{x(i)}^{Y_1})^2
\]

\[
\sum_{i=1}^{n} (\hat{\epsilon}_i^F)^2 = \sum_{i=1}^{n} (e_{x(i)}^{Y_1})^2 + n\Delta_1^2 / \Delta s \Delta_1^2 / \Delta + n\Delta_2^2 / \Delta s \Delta_2^2 / \Delta + 2n((\Delta_1 \Delta_2) / (\Delta^2 s_{12}) \\
- 2n (\Delta_1 / \Delta s(X_1, e_{x(i)}^{Y_1}) + \Delta_2 / \Delta s(X_2, e_{x(i)}^{Y_1}))
\]

### 2.3 The Shuffle-the-conditioning-predictors approach

Kennedy (1995) presented a permutation method which he considered as producing results practically identical to the Freedman and Lane (1983) approach, when retaining a pivotal quantity as a test statistic. Indeed, his numerical experiments showed that the estimator of \( \beta_2 \) by Freedman and Lane method and his were undistinguishable, and he deduced that the two methods were identical. The rationale
for Kennedy’s method also consists in permutating the residuals under one reduced regression model, as in the case of the Freedman and Lane’s approach. From a computational point of view, it proceeds by the following steps:

1-3 The same first three steps are done as in the Freedman and Lane method.

4. The variable $X^2$ is regressed on $X^1$ alone according to the model

$$X^2 = \gamma_0 + \gamma_1 X^1 + \epsilon^{2.1}$$  \hspace{1cm} (15)

to obtain the OLS estimate of $\gamma_0$, $\gamma_1$ and the residuals the respective $g_0$ and $g_1$ and $\epsilon^{2.1}$.

5. The permuted residuals $\epsilon^Y_\pi$ from step 3 are regressed on $\epsilon^{2.1}$ according to the model:

$$\epsilon^Y_\pi = \beta^K_2 \epsilon^{2.1} + \epsilon_k$$  \hspace{1cm} (16)

to obtain an estimate $b^K_2$ and a value $t^K_2 = b^K_2 / s(\hat{\beta}^K_2)$. Here, $t^K_2$ is calculated with $(n - 3)$ degrees of freedom, as in the Freedman and Lane method.

5-6 The same last two steps are done as in the Freedman and Lane method.

### 2.3.1 Statistical properties of the Shuffle-the-conditioning-predictors approach

**Lemma 2.3.** 1) $\sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} \epsilon_{i}^{2.1} = n\Delta_2 / s_1^2$  

2) $\sum_{i=1}^{n} (\epsilon_{i}^{2.1})^2 = n\Delta / s_1^2$

**Proof:** 1) A direct consequence of the definition $\epsilon_{i}^{2.1}$ is:

\[
\sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} \epsilon_{i}^{2.1} = \sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} (x_{i}^2 - g_0 - g_1 x_{i}^1) = \sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} (x_{i}^2 - \bar{x}^2 - d_1(x_{i}^1 - \bar{x}^1))
\]

\[
= \sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} (x_{i}^2 - \bar{x}^2) - s_{12}/s_1^2 \sum_{i=1}^{n} \epsilon_{\pi(i)}^{Y,1} (x_{i}^1 - \bar{x}^1)
\]

\[
= ns(X^2, \epsilon_{\pi(i)}^{Y,1}) - ns_{12}/s_1^2 s(X^1, \epsilon_{\pi(i)}^{Y,1})
\]

\[
= n (s_1^2 s(X^2, \epsilon_{\pi(i)}^{Y,1}) - s_{12} s(X^1, \epsilon_{\pi(i)}^{Y,1}) / s_1^2) = n\Delta_2 / s_1^2
\]  \hspace{1cm} (17)

2) In the same way, it results from the definitions of $\epsilon_{i}^{2.1}$ that:

\[
\sum_{i=1}^{n} (\epsilon_{i}^{2.1})^2 = \sum_{i=1}^{n} (x_{i}^2 - \bar{x}^2 - g_1(x_{i}^1 - \bar{x}^1))^2 = \sum_{i=1}^{n} (x_{i}^2 - \bar{x}^2 - s_{12}/s_1^2 (x_{i}^1 - \bar{x}^1))^2
\]

\[
= ns^2(X^2) + ns_{12}^2/s_1^2 - 2ns_{12}^2/s_1^2 = n(s_2^2 - s_{12}^2/s_1^2) = n\Delta / s_1^2
\]  \hspace{1cm} (18)
Using matrixial argument Kennedy (1995) shows that the OLS estimator of $\beta_2$
by the Freedman and Lane method and his method are identical, that is:
$\hat{\beta}_{2, \pi}^{F} = \hat{\beta}_{2}^{K, \pi}$. Indeed, for Freeman-Lane model (13) the OLS estimator for $\beta_2^{F, \pi}$ is:
$\hat{\beta}_2^{F, \pi} = \Delta_2 / \Delta$ and for Kennedy’s model (16), the OLS estimator for $\beta_2^{K, \pi}$ is: $\hat{\beta}_2^{K, \pi} = \sum_{i=1}^{n} e^{Y_1(i)} / \sum_{i=1}^{n} (e^{21}_i)^2$. So that by lemma 2.2 we have: $\hat{\beta}_{2}^{K, \pi} = \Delta_2 / \Delta = \hat{\beta}_{2}^{F, \pi}$
henceforth, we denote it by $\hat{\beta}_2^2$,

**Lemma 2.4.** The residuals of Kennedy’s model (16) verify:

$$\sum_{i=1}^{n} (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^{n} (e_i^{Y, 1})^2 - (n \Delta_1^2) / (s_1^2)$$

**Proof:** By proposition (2.1):

$$\hat{\epsilon}_i^K = e_i^{Y, 1}_\pi - \hat{\beta}_2 \pi e_i^{2, 1} = e_i^{Y, 1}_\pi - \Delta_2 / \Delta ((x_i^2 - \bar{x}^2) - s_{12} / s_1^2 (x_i^1 - \bar{x}^1))$$

$$= e_i^{Y, 1}_\pi - \Delta_2 / \Delta (x_i^1 - \bar{x}^1) + (\Delta_2 s_{12}) / (\Delta s_1^2) (x_i^1 - \bar{x}^1)$$

and thus

$$\sum_{i=1}^{n} (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^{n} (e_i^{Y, 1})^2 + \Delta_2 / \Delta \sum_{i=1}^{n} x_i^2 - \bar{x}^2 + \Delta_2 / \Delta \sum_{i=1}^{n} s_{12} / s_1^2 \sum_{i=1}^{n} x_i^1 - \bar{x}^1$$

$$- 2 \Delta_2 / \Delta \sum_{i=1}^{n} x_i^2 - \bar{x}^2 e_i^{Y, 1}_\pi + 2 (\Delta_2 s_{12}) / (\Delta s_1^2) \sum_{i=1}^{n} x_i^1 - \bar{x}^1 e_i^{Y, 1}_\pi$$

$$- 2 (\Delta_2 s_{12}) / (\Delta s_1^2) \sum_{i=1}^{n} (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2) \quad \text{(by } \sum_{i=1}^{n} e_i^{Y, 1}_\pi = \sum_{i=1}^{n} e_i^{Y, 1})$$

$$\sum_{i=1}^{n} (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^{n} (e_i^{Y, 1})^2 + n \Delta_2 / \Delta \sum_{i=1}^{n} x_i^2 - \bar{x}^2 + n \Delta_2 / \Delta \sum_{i=1}^{n} s_{12}^2 / s_1^2 + 2 n \Delta_2 / \Delta s(X^2, e_i^{Y, 1}_\pi)$$

$$+ 2 n (\Delta_2 s_{12}) / (\Delta s_1^2) s(X^1, e_i^{Y, 1}_\pi) - 2 n (\Delta_2 s_{12}) / (\Delta s_1^2)$$

$$= \sum_{i=1}^{n} (e_i^{Y, 1})^2 + n \Delta_2 / \Delta (s_2^2 + [s_{12}]^2 / s_1^2 - 2 [s_{12}]^2 / s_1^2)$$

$$- 2 n \Delta_2 / \Delta (s(X^2, e_i^{Y, 1}_\pi) - s_{12} / s_1^2 s(X^1, e_i^{Y, 1}_\pi)) = \sum_{i=1}^{n} (e_i^{Y, 1})^2 - n \Delta_1^2 / (s_1^2)$$

\[ \square \]

3 Main Results

Thus $\sum_{i=1}^{n} (\hat{\epsilon}_i^K)^2 \neq \sum_{i=1}^{n} (\hat{\epsilon}_i^F)^2$ and, the Student T-ratios derived by the Freedman
and Lane and Kennedy’s methods differ:

$$T_2^{F, \pi} = \hat{\beta}_2 / s^F(\hat{\beta}_2^2) \neq \hat{\beta}_2^2 / s^K(\hat{\beta}_2^2) = T_2^{K, \pi}$$

(19)
since $S^F(\hat{\beta}^\pi) \neq S^K(\hat{\beta}^\pi)$ where $s^2F(\hat{\beta}^\pi) = (s^2F)^2/(n\Delta)$ is the estimators of $\text{Var}(\hat{\beta}^\pi)$ obtained by the Freedman while the Kennedy’s method leads to $s^2K(\hat{\beta}^\pi) = s^2K^2/(n\Delta)$. Moreover, $s^2F = \sum_{i=1}^n (\hat{\varepsilon}_i^F)^2/(n-3)$ and $s^2K = \sum_{i=1}^n (\hat{\varepsilon}_i^K)^2/(n-3)$, are respectively the OLS estimators of $(\sigma^F)^2$ and $(\sigma^K)^2$. The comparison of the residual variance obtained by each methods can averted since:

**Lemma 3.1.** \[
\sum_{i=1}^n (\hat{\varepsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\varepsilon}_i^F)^2 = n[s(X^1, e_{\pi}^Y)]^2/s_1^2 \geq 0
\]

**Proof:** From 2.2 and 2.4 we have:

\[
\sum_{i=1}^n (\hat{\varepsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\varepsilon}_i^F)^2 = -nS_1^2\Delta_2^2/\Delta - nS_2^2\Delta_1^2/\Delta^2 - nS_1^2\Delta_1^2/\Delta^2 - 2n(\Delta_1\Delta_2)/\Delta^2S_{12} + 2n(S(X^2, e_{\pi}^Y)\Delta_2/\Delta + S(X^1, e_{\pi}^Y))\Delta_1/\Delta = n(-\Delta_2^2(\Delta + S_1^2S_2^2)) - n\Delta_1^2[S_1^2]^2 - 2n\Delta_1\Delta_2S_1^2S_{12})/(\Delta^2S_1^2) + (2n\Delta S_2^2[\Delta_2S(X^2, e_{\pi}^Y) + \Delta_1S(X^1, e_{\pi}^Y)])/(\Delta^2S_1^2)
\]

by substitution $\Delta$, $\Delta_1$ and $\Delta_2$ one obtain:

\[
\sum_{i=1}^n (\hat{\varepsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\varepsilon}_i^F)^2 = n[S(X^1, e_{\pi}^Y)]^2/S_1^2 \geq 0
\]

**Theorem 3.1.** \[
s^2K(\hat{\beta}^\pi) - s^2F(\hat{\beta}^\pi) = [s(X^1, e_{\pi(i)}^Y)]^2/((n-3)\Delta) \geq 0
\]

**Proof:** By definition,

\[
s^2K(\hat{\beta}^\pi) - s^2F(\hat{\beta}^\pi) = s_1^2n\sum_{i=1}^n (\hat{\varepsilon}_i^K)^2/(n(n-3)\Delta) - s_1^2n\sum_{i=1}^n (\hat{\varepsilon}_i^F)^2/(n(n-3)\Delta) = s_1^2/(n(n-3)\Delta)(\sum_{i=1}^n (\hat{\varepsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\varepsilon}_i^F)^2)
\]

so from lemma (3.1) \[
[s(X^1, e_{\pi(i)}^Y)]^2/((n-3)\Delta) \geq 0
\]

Thus, for each permutation the absolute value of the $t$-ratio derived by the Freedman and Lane’s method is greater or equal to the corresponding one obtained by Kennedy’s approach. \[
s^2F(\hat{\beta}^\pi) \leq s^2K(\hat{\beta}^\pi) \implies |T_2^{K,\pi}| = |\hat{\beta}^\pi|/s^K(\hat{\beta}^\pi) \leq |\hat{\beta}^\pi|/s^F(\hat{\beta}^\pi) = |T_2^{F,\pi}| \implies |T_2^{K,\pi}| \leq |T_2^{F,\pi}|
\]

As a consequence the dispersion of the permutational distribution of $|T_2^{F,\pi}|$ is greater or equal to that of $|T_2^{K,\pi}|$, which means making these two statistics stochastically ordered:
Theorem 3.2. The distribution function of $|T_2^{F,\pi}|$ is lower or equal to that of $|T_2^{K,\pi}|$, i.e.: 
\[ F_{|T_2^{F,\pi}|}(t) \leq F_{|T_2^{K,\pi}|}(t) \quad \forall t \]

\textbf{Proof:} In ([12]),

The observed permutational p-value associated with each method is the percentage of values of the permutational statistics (i.e. $t_2^{F,\pi}$ and $t_2^{K,\pi}$) which are larger than $t_{ref}$. The next Corollary shows that the permutational p-value of the method of Freedman and Lane ($p_2^{F,\pi}$) is higher or equal to that of the method of Kennedy ($p_2^{K,\pi}$). In fact permutational power function of the method of Kennedy is higher or equal to that of the method of Freedman and Lane. One deduces that the type I error of the Freedman and Lane’s method is lower or equal to that of the Kennedy’s method and vice versa (the opposite), the power of test obtained by the Freedman and Lane method is less or equal than that of the Kennedy’s method.

\textbf{Corollary 3.1.} For testing $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$, the permutational p-value of the Freedman and Lane method is greater or equal to that of the Kennedy’s method: $p_2^{F,\pi} \geq p_2^{K,\pi}$, and the permutational power function of the Kennedy’s method is higher or equal to that of Freedman and Lane’s method. Moreover:

1) The type I error of the Freedman and Lane’s method is lower or equal to that of the Kennedy’s method.

2) The power of test of Freedman and Lane method is lower or equal to that of the Kennedy’s method.

\textbf{Proof:} According to the theorem (3.2):

\[ p_2^{F,\pi} = Pr(|T_2^{F,\pi}| > t_{obs}) \geq Pr(|T_2^{K,\pi}| > t_{obs}) = p_2^{K,\pi} \quad \square \quad (20) \]

We know that the hypothesis $H_0 : \beta_2 = 0$ is rejected when the observed p-value is smaller than the test size $\alpha$. The probability of rejecting $H_0 : \beta_2 = 0$ by the Freedman and Lane’s method is thus lower or equal to that which one obtains by the Kennedy’s method i.e.:

\[ Pr(P_2^{F,\pi} < \alpha) \leq Pr(P_2^{K,\pi} < \alpha) \quad (21) \]

Therefore, the permutational power function of the method of Freedman and Lane is lower or equal to that of the Kennedy’s method. If $C$ is the critical region of the test:

\[ \eta^{F,\pi}(\beta_2) = Pr^{F,\pi}(C) \leq Pr^{K,\pi}(C) = \eta^{K,\pi}(\beta_2) \quad (22) \]

Consequently, for data generated under the hypothesis $H_0 : \beta_2 = 0$, the relation above says that the type I error of the Freedman and Lane method is lower than
that of Kennedy’s approach.

\[ Pr_{H_0}(P_2^{F, \pi} < \alpha) \leq Pr_{H_0}(P_2^{K, \pi} < \alpha) \]  

(23)

These results give a mathematical justification to the results of simulations of Andersson and Legendre (1999). On the other hand, when the data are generated under the hypothesis \( H_1 : \beta_2 \neq 0 \) the relation above says that the power of the test of the Freedman and Lane’s method is lower or equal to that of the Kennedy’s method.

\[ Pr_{H_1}(P_2^{F, \pi} < \alpha) \leq Pr_{H_1}(P_2^{K, \pi} < \alpha) \]  

(24)

However, When Andersson and Legendre compare the power of test of the permutations methods by simulation the method of Kennedy is not considered. \( \square \)

More precisely, the theorem (3.1) says that the magnitude of

\[ s^2K(\hat{\beta}_2^\pi) - s^2F(\hat{\beta}_2^\pi) = s_1^2s_2^2(e_{Y,1})[\rho(X^1, e_{\pi}^Y)]^2/((n - 3)\Delta) \geq 0 \]  

(25)

depends on the square of the correlation of \( X^1 \) and \( e_{\pi}^Y \). Since the empirical variance is invariant by permutation. So, two tests strategy have the same power only if this correlation is null, a strong exogeneity conclusion.

**Theorem 3.3.** When the sample size grows to infinity, the difference between the two estimators of \( \text{Var}(\hat{\beta}_2^\pi) \), tends to zero.

**Proof:** When \( n \) grows to infinity assuming that all the limits exist, classical arguments show that the matrix:

\[
\begin{pmatrix}
  s_1^2 & s_{12} \\
  s_{12} & s_2^2
\end{pmatrix}
\]

converges toward

\[
\begin{pmatrix}
  \sigma_1^2 & \sigma_{12} \\
  \sigma_{12} & \sigma_2^2
\end{pmatrix},
\]

\( s^2(e_{2,1}) \) converges towards \( (\sigma_{2,1})^2, \Delta \) converges toward \( \Delta^* = \sigma_1^2\sigma_2^2 - (\sigma_{12})^2 > 0 \) and \( \rho(X^1, X^2) \) exists then:

\[
\lim_{n \to +\infty} (s^2K(\hat{\beta}_2^\pi) - s^2F(\hat{\beta}_2^\pi)) = \lim_{n \to +\infty} s_1^2s_2^2(e_{Y,1})[\rho(X^1, e_{\pi}^Y)]^2/((n - 3)\Delta)
\]

\[
= \sigma_1^2\sigma_2^2[\rho(X^1, e_{\pi}^Y)]^2/\Delta^* \times \lim_{n \to +\infty} 1/(n - 3) = 0
\]

So, the two estimators are asymptotically equal. This theorem shows in particular that the \( t \)-ratio statistics of the Freedman and Lane’s method and Kennedy’s method become indistinguishable when the sample size is large, showing very close type I error and power values, in agreement with the results of simulations published by Anderson and Legendre (1999).
4 Permutational Properties of Estimation

This section shows that while estimate \( \text{Var}(\hat{\beta}_2^\pi) \) is known to the two approaches, it lead to different permutational properties.

This section is devoted to the calculation of the permutational expectation and permutational variance of \( \hat{\beta}_2^\pi \) like to the permutational expectation of the estimators of \( \text{Var}(\hat{\beta}_2^\pi) \) built according to the methods of Freedman and Lane and Kennedy. Here, we use the results of Czaplewski et Al. ([2]) to calculate the first two permutational moments of a random variable of interest. In this section, the notation \( \mathcal{E}(.) \) and \( \mathcal{V}(.) \) distinguishes permutational moments from the ones \( \mathbb{E}(.) \) and \( \text{Var}(.) \).

4.1 Statistical properties of variables permuted

Let \( \pi \) denote a permutation of the set \( I \) of \( n \) elements and Suppose \( \pi(i) \) a permuted index of the \( I_n \) unit and \( x_{\pi(i)} \) the associated permuted random variable.

Lemma 4.1. The permutational expectation of \( X_{\pi(i)} \) and \( (X_{\pi(i)})^2 \) calculated on all the \( n! \) possible permutations of \( I \) are (\( \forall i \)):

1. \( \mathcal{E}(x_{\pi(i)}) = 1/n \sum_{p=1}^{n} x_p = \bar{x} \)
2. \( \mathcal{E}(x_{\pi(i)})^2 = 1/n \sum_{p=1}^{n} x_p^2 = \bar{x}^2 \) \( \square \)

Proof: 1. By definition of the expectation, we have:

\[ \mathcal{E}(x_{\pi(i)}) = \frac{1}{n} \sum_{p=1}^{n} x_{\pi(i)} = \frac{1}{n} \sum_{p=1}^{n} x_p = \bar{x} \]

where \( p(x_{\pi(i)} = x_p) \) is the probability with which permuted variable \( x_{\pi(i)} \) takes the observed value \( x_p \) (\( 1 \leq p \leq n \)) among \( n! \) possible permutation. Among these \( n! \) possibles permutation, there exists \( (n-1)! \) permutations for which \( x_{\pi(i)} = x_p \) and thus \( p(x_{\pi(i)} = x_p) = (n-1)!/n! = 1/n \). Consequently:

\[ \mathcal{E}(x_{\pi(i)}) = 1/n \sum_{p=1}^{n} x_p = \bar{x} \]

2. In a similar way is shown, that: \( \mathcal{E}(x_{\pi(i)})^2 = 1/n \sum_{p=1}^{n} x_p^2 = \bar{x}^2 \) \( \square \)

In bivariate case we consider a sample of \( n \) observations of the couple \( Z = (X,Y) \). Let \( z = (x,y) \) the observation of this random vector on the real sample of size \( n \) \( z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \) is defined by: \( z_i = (x_i, y_i) = 1, 2, ..., n \). Any permutation \( \pi \) defined on \( I_n \) induces a permutation of the coordinates of the vector \( z \). This permutation is also noted \( \pi : (x_{\pi(i)}, y_{\pi(i)}) \)

\[ z = (z_1, z_2, ..., z_n) \xrightarrow{\pi} z_{\pi} = (z_{\pi(1)}, z_{\pi(2)}, ..., z_{\pi(n)}) \]

(26)

Lemma 4.2. \( z = (x,y) \) a bivariate random vector observed, defined on \( (\omega, \mathcal{F}, P) \). Suppose \( z \) a sample of size \( n \) drawn according to the law from \( z \) subscripted by unit \( I_n \). Suppose \( (x_{\pi(i)}, y_{\pi(i)}) \) the associated permuted random variable. Then, the
permutationelles expectation of \((x_{\pi(i)}y_{\pi(i)})\) and \((x_{\pi(i)})^2(y_{\pi(i)})^2\), calculated on all the \(n!\) possible permutations of \(z\), are: 1. \(E(x_{\pi(i)}y_{\pi(i)}) = 1/n \sum_{p=1}^{n} x_p y_p = \bar{x}\bar{y}\)
2. \(E((x_{\pi(i)})^2(y_{\pi(i)})^2) = 1/n \sum_{p=1}^{n} x_p^2 y_p^2 = \bar{x}^2\bar{y}^2\)

**Proof:** 1. By definition of the expectation, we have:

\[ E(x_{\pi(i)}y_{\pi(i)}) = \sum_{k=1}^{n} x_k y_k p(x_{\pi(i)} = x_k, y_{\pi(i)} = y_k) \]

where \(p(x_{\pi(i)} = x_p, y_{\pi(i)} = y_p) = (n-1)!/n!\), \((1 \leq p \leq n)\) and thus:

\[ E(x_{\pi(i)}y_{\pi(i)}) = \frac{1}{n} \sum_{p=1}^{n} x_p y_p = \bar{x}\bar{y} \]

2. In a similar way is shown: \(E((x_{\pi(i)})^2(y_{\pi(i)})^2) = 1/n \sum_{p=1}^{n} x_p^2 y_p^2 = \bar{x}^2\bar{y}^2\)

**Lemma 4.3.** Let us consider the random permutation \(\pi\), defined on the \(x = (x_1, x_2, ..., x_n)\),

\[ x = (x_1, x_2, ..., x_n) \xrightarrow{\pi} x^\pi = (x_{\pi(1)}, ..., x_{\pi(i)}, ..., x_{\pi(j)}, ..., x_{\pi(n)}) \]

Then, for any couple \(i, j = 1, \cdots, n, i \neq j\), the permutational expectation of \((x_{\pi(i)}x_{\pi(j)})\) calculated on the \(n!\) possible permutations \(\pi\), verify:

\[ E(x_{\pi(i)}x_{\pi(j)}) = n\bar{x}^2/(n-1) - \bar{x}^2/(n-1) \neq E(x_{\pi(i)}E(x_{\pi(j)}) = \bar{x}^2 \]

**Proof:** \(E(x_{\pi(i)}x_{\pi(j)}) = \sum_{p=1}^{n} \sum_{q=1}^{n} x_p x_q p(x_{\pi(i)} = x_p, x_{\pi(j)} = x_q) \) where:

\[ p(x_{\pi(i)} = x_p, x_{\pi(j)} = x_q) = p(x_{\pi(j)} = x_q|x_{\pi(i)} = x_p)p(x_{\pi(i)} = x_p) \]

In addition, we know that when the margin \(X_{\pi(i)}\) bivariate variable \((x_{\pi(i)}, x_{\pi(j)})\) takes value \(x_p\), the variable \(x_{\pi(j)}(j \neq i)\) cannot take this value \(x_p\), since \(x_p\) is taken by the variable \(x_{\pi(i)}\). Among the \(n!\) possible permutations, there thus exists \((n-1)!\) such as \(x_{\pi(i)} = x_p\) about it. Then, \(p(x_{\pi(i)} = x_p) = (n-1)!/n! = 1/n\). So only \((n-1)\) permutations among \(n!\) possible permutations are acceptable for \((n-1)\) random variable \((x_{\pi(k)}), (k \neq i)\) and among these \((n-1)!\) permutations acceptable there exist \((n-2)!\) permutation such as \(x_{\pi(j)} = x_q(q \neq p)\). Consequently:

\[ p(x_{\pi(j)} = x_q|x_{\pi(i)} = x_p) = (n-2)!/(n-1)! = 1/(n-1) \]

and thus:

\[ p(x_{\pi(i)} = x_p, x_{\pi(j)} = x_q) = p(x_{\pi(j)} = x_q|x_{\pi(i)} = x_p)p(x_{\pi(i)} = x_p) = 1/(n(n-1)) \]

consequently:

\[ E(x_{\pi(i)}x_{\pi(j)}) = 1/(n(n-1)) \sum_{p=1}^{n} \sum_{q=1}^{n} x_p x_q = 1/(n(n-1)) \sum_{p=1}^{n} x_p \left( \sum_{q=1}^{n} x_q - x_p \right) \]

\[ = n\bar{x}^2/(n-1) - \bar{x}^2/(n-1) \]

\(\blacksquare\)

**An analytic comparison of permutation methods**

871
Corollary 4.1. If \( \pi \) is a random permutation defined on the set of observed value \( x = \{ x_1, x_2, \cdots, x_n \}, i \neq j = 1, 2, \cdots, n \), then:

1. \( \mathcal{E}((x_{\pi(i)})^2, (x_{\pi(j)})^2) = n \left( \frac{\overline{x}^2}{n - 1} - \frac{x^2}{n} \right) \)
2. \( \mathcal{E}((x_{\pi(i)})^2, x_{\pi(j)}) = n \frac{\overline{x}^2 \bar{x}}{(n - 1) - \frac{x^3}{n}} \)

Proof: We can obtain by sub situation in lemma 4.3

4.2 Permutational moments of \( \hat{\beta}_2^\pi \)

We notice from (2.2.1) that we must know the permutational expectation of the statistics \( s(X^1, e_{\pi Y}^1) \) and \( s(X^2, e_{\pi Y}^1) \) in order to calculate the permutational expectation of \( \hat{\beta}_2^\pi \), this results from the next lemma.

Lemma 4.4. For any permutation \( \pi \) defined on \( I_n \), the vector of the residuals \( e_{\pi Y}^1 \)

verify: 1. \( \mathcal{E}(s(X^1, e_{\pi Y}^1)) = 0 \)
2. \( \mathcal{E}(s(X^2, e_{\pi Y}^1)) = 0 \)

Proof: 1. By definition \( \mathcal{E}(s(X^1, e_{\pi Y}^1)) = \mathcal{E} \left( \sum_{i=1}^{n} (x_i^1 - \bar{x}^1) e_{\pi(i)}^Y \right) = \sum_{i=1}^{n} \bar{x}^1 e_{\pi(i)}^Y \)

2. In the same way, we can shown that: \( \mathcal{E}(s(X^2, e_{\pi Y}^1)) = 0 \)

4.3 The permutational variance of \( \hat{\beta}_2^\pi \)

The evaluation of the permutational sampling variance of \( \hat{\beta}_2^\pi \) necessitates the use of one technical lemmas.

Lemma 4.5. Let \( e_{\pi Y}^1 \) be the vector of residuals permuted and suppose that \( X^1 \) is of nonnull empirical sample variance. One has:

1. \( \sum_{i=1}^{n} (e_i^Y)^2 = n \left( s_Y^2 - [s_{Y1}]^2 / s_1^2 \right) \)
2. \( \sum_{j \neq i=1}^{n} e_i^Y e_j^Y = -n \left( s_Y^2 - [s_{Y1}]^2 / s_1^2 \right) \)
Proof:

1. \( \sum_{i=1}^{n} (e_{i}^{Y,1})^2 = \sum_{i=1}^{n} (Y_i - a_0 - a_1 x_{1i})^2 = \sum_{i=1}^{n} (Y_i - \bar{Y} - a_1(x_{1i} - \bar{x}^1))^2 = \frac{n}{2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + a_1^2 \sum_{i=1}^{n} (x_{1i} - \bar{x}^1)^2 - 2a_1 \sum_{i=1}^{n} \sum_{i \neq j} (x_{1i} - \bar{x}^1)(Y_i - \bar{Y}) = n s_Y^2 + n \frac{s_{Y,1}^2}{s_1^2} - 2n s_{Y,1} s_{Y} / s_1 = n (s_Y^2 - [s_{Y,1}]^2 / s_1^2) \)

2. \( \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} e_{i}^{Y,1} e_{j}^{Y,1} = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (Y_i - a_0 - a_1 x_{1i})(Y_j - a_0 - a_1 x_{1j}) = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} ((Y_i - \bar{Y}) - a_1(x_{1i} - \bar{x}^1))((Y_j - \bar{Y}) - a_1(x_{1j} - \bar{x}^1)) = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} ((Y_i - \bar{Y} - Y_j + \bar{Y}) + a_1^2 + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (x_{1i} - \bar{x}^1)(x_{1j} - \bar{x}^1) - 2a_1 \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (Y_i - \bar{Y})(x_{1j} - \bar{X}) = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (Y_i - \bar{Y})(x_{1j} - \bar{X}^1) = -n s_Y^2 - n \frac{s_{Y,1}^2}{s_1^2} s_1^2 + 2n s_{Y,1} s_{Y} / s_1 = -n (s_Y^2 - [s_{Y,1}]^2 / s_1^2) \)

Lemma 4.6. If \( e_{\pi,1}^{Y} \), the vector permuted residual and \( X^1 \), is of nonnull empirical variance, one has:

1. \( \mathcal{E}(e_{\pi(i)}^{Y,1})^2 = s_Y^2 - [s_{Y,1}]^2 / s_1^2 \)

2. \( \mathcal{E}(e_{\pi(i)}^{Y,1} e_{\pi(j)}^{Y,1}) = -\left( s_Y^2 - [s_{Y,1}]^2 / s_1^2 \right) / (n - 1) \)

Proof:

1. According to Lemma (4.1), one has: \( \mathcal{E}(e_{\pi(i)}^{Y,1})^2 = 1 / n \sum_{i=1}^{n} (e_{i}^{Y,1})^2 \) lemma (4.5) allows to write: \( \mathcal{E}(e_{\pi(i)}^{Y,1})^2 = s_Y^2 - [s_{Y,1}]^2 / s_1^2 \)

2. According to Lemma (4.3), one has moreover:

\( \mathcal{E}(e_{\pi(i)}^{Y,1} e_{\pi(j)}^{Y,1}) = n (e_Y^1)^2 / (n - 1) - \sum_{p=1}^{n} (e_Y^1)^2 / (n(n - 1)) \) and it is known that \( e_Y = 0 \). Consequently, the following result is a consequence of the Lemma (4.5):

\( \mathcal{E}(e_{\pi(i)}^{Y,1} e_{\pi(j)}^{Y,1}) = -\left( s_Y^2 - [s_{Y,1}]^2 / s_1^2 \right) / (n - 1) \)

Lemma 4.7. It is supposed that \( s_1^2 \neq 0 \). The variable \( e_{\pi,1}^{Y} \) verify:

1. \( \mathcal{E}(s(X^1, e_{\pi,1}^{Y}))^2 = s_1^2 (s_Y^2 - [s_{Y,1}]^2 / s_1^2) / (n - 1) \)

2. \( \mathcal{E}(s(X^2, e_{\pi,1}^{Y}))^2 = s_2^2 (s_Y^2 - [s_{Y,1}]^2 / s_1^2) / (n - 1) \)

3. \( \mathcal{E}(s(X^1, e_{\pi,1}^{Y}) s(X^2, e_{\pi,1}^{Y})) = s_{12} (s_Y^2 - [s_{Y,1}]^2 / s_1^2) / (n - 1) \)
Proof:
1. $\mathcal{E}(s(X^1, e_{\pi}^{Y_1}))^2 = \mathcal{E}(1/n \sum_{i=1}^{n} (X_{i1} - \bar{X}^1)e_{\pi(i)}^{Y_1})^2 = \mathcal{E}(1/n^2 \sum_{i=1}^{n} (X_{i1} - \bar{X}^1)^2 e_{\pi(i)}^{Y_1})^2$

$+ \mathcal{E}(1/n^2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_{i1} - \bar{X}^1)(X_{j1} - \bar{X}^1)e_{\pi(i)}^{Y_1}e_{\pi(j)}^{Y_1}) = 1/n^2 \sum_{i=1}^{n} (X_{i1} - \bar{X}^1)^2 \mathcal{E}(e_{\pi(i)}^{Y_1})^2$

$+ 1/n^2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_{i1} - \bar{X}^1)(X_{j1} - \bar{X}^1)\mathcal{E}(e_{\pi(i)}^{Y_1}e_{\pi(j)}^{Y_1}) = \frac{1}{n} s_2^2(\bar{X}^1)^2 \mathcal{E}(e_{\pi(i)}^{Y_1})^2 - \mathcal{E}(e_{\pi(i)}^{Y_1})^2$]
and lemma (4.6) ensure that: $\mathcal{E}(s(X^1, e_{\pi}^{Y_1}))^2 = s_1^2(s_1^2 - [s_1^2]^2)/(n - 1)$
2. In a similar way, one can show that:
$\mathcal{E}(s(X^2, e_{\pi}^{Y_1}))^2 = s_2^2(s_2^2 - [s_2^2]^2)/(n - 1)$
3. $\mathcal{E}(s(X^1, e_{\pi}^{Y_1}, \pi) s(X^2, e_{\pi}^{Y_1}, \pi)) = \mathcal{E}(1/n \sum_{i=1}^{n} (X_{i1} - \bar{X}^1)e_{\pi(i)}^{Y_1} \frac{1}{n} \sum_{i=1}^{n} (X_{2i} - \bar{X}^2)e_{\pi(i)}^{Y_1})$

$= \mathcal{E}(1/n^2 \sum_{i=1}^{n} (X_{i1} - \bar{X}^1)(X_{2i} - \bar{X}^2) e_{\pi(i)}^{Y_1})$

$+ \mathcal{E}(1/n^2 \sum_{i=1}^{n} \sum_{j \neq i}^{n} (X_{i1} - \bar{X}^1)(X_{2j} - \bar{X}^2) e_{\pi(i)}^{Y_1} e_{\pi(j)}^{Y_1})$

$= \frac{s_{12}}{n} \mathcal{E}(e_{\pi(i)}^{Y_1})^2 - \frac{s_{12}}{n} \mathcal{E}(e_{\pi(i)}^{Y_1} e_{\pi(j)}^{Y_1}) = \frac{s_{12}}{n} \mathcal{E}(e_{\pi(i)}^{Y_1})^2 - \mathcal{E}(e_{\pi(i)}^{Y_1})^2$

by using lemma 4.6 one obtains thus:
$\mathcal{E}(s(X^1, e_{\pi}^{Y_1}) s(X^2, e_{\pi}^{Y_1})) = \frac{s_{12}}{n} (s_1^2 - [s_2^2]^2)/(n - 1)$

Proposition 4.2. In the (non trivial) case where $\Delta \neq 0$, the permutational variance of $\beta_2^\pi$ is:
$\text{Var}(\hat{\beta}_2^\pi) = (s_1^2 s_2^2 - [s_1^2]^2)/(n - 1)$

Proof: By proposition (4.1) $\text{Var}(\hat{\beta}_2^\pi) = \mathcal{E}(\hat{\beta}_2^\pi)^2 - \mathcal{E}^2(\hat{\beta}_2^\pi) = \mathcal{E}(\hat{\beta}_2^\pi)^2$ thus:
$\Delta \mathcal{E}(\hat{\beta}_2^\pi)^2 = \mathcal{E} \left[ s_1^2 s(X^2, e_{\pi}^{Y_1}) - s(X^1, e_{\pi}^{Y_1}) s_{12} \right]^2 = (s_1^2)^2 \mathcal{E} [s(X^2, e_{\pi}^{Y_1})]^2$

$+ (s_{12})^2 \mathcal{E} [s(X^1, e_{\pi}^{Y_1})]^2 - 2s_1^2 s_{12} \mathcal{E} [s(X^1, e_{\pi}^{Y_1}) s(X^2, e_{\pi}^{Y_1})]$ so by lemma (4.7)
$\text{Var}(\hat{\beta}_2^\pi) = (s_1^2 s_2^2 - [s_1^2]^2)/(n - 1)$

This equation shows that the permutational variance of the estimator $\hat{\beta}_2^\pi$ is a simple function of the empirical moments of the vector of the observations $(X^1, X^2, Y)$ and is invariant by permutation of the observed data set. We are now in a position to compare this value with the OLS estimator of $\text{Var}(\hat{\beta}_2^\pi)$, for the Friedman and Lane’s method and the one of Kennedy.

4.4 Permutational expectation of the estimators of $\text{Var}(\hat{\beta}_2^\pi)$

Proposition (4.2) showed that the permutational variance of $\hat{\beta}_2^\pi$ is independent of $\pi$ (cf equation 4.3). As already known, the OLS estimator of $\text{Var}(\hat{\beta}_2^\pi)$ resulting from the Friedman and Lane method differs from the Kennedy’s one. We to study
their permutational law, with help of the permutational expectation of the estimators of \( \text{Var}(\hat{\beta}_2^\pi) \) provided by the two methods. Then, they are compared to the permutational variance of \( \hat{\beta}_2^\pi \).

### 4.4.1 The Freedman and Lane permutation method

The approach of Freedman and Lane, is characterized by the regression equation (13) and we know that: \( \text{Var}(\hat{\beta}_2^\pi)_F = s_2^2(\sigma^2F)/(n\Delta) \). The OLS estimator of \( \text{Var}(\hat{\beta}_2^\pi)_F \) is: \( \text{Var}(\hat{\beta}_2^\pi)_F = s_2^2(\hat{\beta}_2^\pi)_F = s_2^2s^2F/(n\Delta) \). In a way to calculate the permutational expectation of \( s^2(\hat{\beta}_2^\pi)_F \), we notice first that:

\[
\mathcal{E}(s^2(\hat{\beta}_2^\pi)_F) = \mathcal{E}(s_2^2\sum_{i=1}^n (\hat{e}_i^F)^2/(n(n-3)\Delta) = s_2^2/(n(n-3)\Delta). \sum_{i=1}^n \mathcal{E}(\hat{e}_i^F)^2 \tag{27}
\]

and so, one needs know \( \sum_{i=1}^n \mathcal{E}(\hat{e}_i^F)^2 \) to calculate the permutational expectation of \( s^2(\hat{\beta}_2^\pi)_F \). But, lemma (2.3) says that:

\[
\mathcal{E}(\sum_{i=1}^n (\hat{e}_i^F)^2) = \sum_{i=1}^n (\hat{e}_i^{Y1})^2 + n s_1^2/\Delta^2 \mathcal{E}(\Delta_1^2) + n s_2^2/\Delta^2 \mathcal{E}(\Delta_2^2) + 2n s_{12}/\Delta^2 \mathcal{E}(\Delta_1 \Delta_2) - 2n/\Delta \mathcal{E}(\Delta_1 s(X^1, e_{\pi(i)}^{Y1})) + \mathcal{E}(\Delta_2 s(X^2, e_{\pi(i)}^{Y1}))/\Delta
\]

Then, the expectation \( \sum_{i=1}^n (\hat{e}_i^F)^2 \) is a function of those of \( \Delta_1, \Delta_2, s(X^2, e_{\pi(i)}^{Y1}) \) and \( s(X^1, e_{\pi(i)}^{Y1}) \):

**Lemma 4.8.** For any data set such as \( s_2^2 \neq 0 \), the vector \( e_{\pi}^{Y1} \) of permuted residuals in the regression of \( Y \) on \( X^1 \), satifies:

1. \( \mathcal{E}(\Delta_1^2) = \Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) s_2^2/(n-1) \)
2. \( \mathcal{E}(\Delta_2^2) = \Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) s_2^2/(n-1) \)
3. \( \mathcal{E}(\Delta_1 \Delta_2) = -\Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) s_{12}^2/(n-1) \)
4. \( \mathcal{E}(\Delta_1 s(X^2, e_{\pi(i)}^{Y1})) = \Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) / (n-1) \)
5. \( \mathcal{E}(\Delta_2 s(X^1, e_{\pi(i)}^{Y1})) = \Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) / (n-1) \)

**Proof:**

\[
\begin{align*}
1. \mathcal{E}(\Delta_1^2) &= \mathcal{E}[s_2^2s(X^1, e_{\pi(i)}^{Y1}) - s_{12}s(X^2, e_{\pi(i)}^{Y1})]^2 \\
&= [s_2^2]^2\mathcal{E}[s(X^1, e_{\pi(i)}^{Y1})]^2 + [s_{12}]^2\mathcal{E}[s(X^2, e_{\pi(i)}^{Y1})]^2 - 2s_2^2s_{12}\mathcal{E}[s(X^1, e_{\pi(i)}^{Y1})s(X^2, e_{\pi(i)}^{Y1})] \\
\text{but the lemma 4.7 makes it possible to write} \\
&= 1/(n-1) (s_Y^2 - [s_{Y1}]^2/s_1^2) ([s_2^2]^2 s_Y^2 + [s_{12}]^2 s_2^2 - 2s_2^2[s_{12}]^2) \\
&= \Delta. (s_Y^2 - [s_{Y1}]^2/s_1^2) s_2^2/(n-1) \\
\end{align*}
\]
2. An analogous argument shows that: $\mathcal{E}(\Delta_2^2) = \Delta \left(s_Y^2 - [s_{Y1}]^2/s_1^2 \right) s_1^2/(n-1)$

3. It remains to know the conditional expectation of the products. But,

$$
\mathcal{E}(\Delta_1\Delta_2) = \mathcal{E}[s^2(X^1)s(X^2, e^{Y_1, \pi}) - s(X^1, X^2)s(X^1, e^{Y_1, \pi})] \\
\times \left[s^2(X^2)s(X^1, e^{Y_1, \pi}) - s(X^1, X^2)s(X^2, e^{Y_1, \pi})\right] \\
= s^2(X^1)s^2(X^2)\mathcal{E}[s(X^1, e^{Y_1, \pi})s(X^2, e^{Y_1, \pi})] \\
- s(X^1, X^2)s^2(X^2)\mathcal{E}[s(X^1, e^{Y_1, \pi})]^2 - s^2(X^1)s(X^1, X^2)\mathcal{E}[s(X^2, e^{Y_1, \pi})]^2 \\
+ [s(X^1, X^2)]^2\mathcal{E}[s(X^1, e^{Y_1, \pi})s(X^2, e^{Y_1, \pi})]
$$

so, lemma 4.8 implies after some substitution and algebraic reorganization:

$$
\mathcal{E}(\Delta_1\Delta_2) = -\Delta \left(s_Y^2 - [s_{Y1}]^2/s_1^2 \right) s_1^2/(n-1)
$$

4. $\mathcal{E}(\Delta_1 s(X^2, e^{Y_1, \pi})) = \mathcal{E}[s^2(X^1)s(X^2, e^{Y_1, \pi}) - s(X^1, X^2)s(X^1, e^{Y_1, \pi})] \\
\times \left[s^2(X^2)s(X^1, e^{Y_1, \pi}) - s(X^1, X^2)s(X^2, e^{Y_1, \pi})\right] \\
= s^2(X^1)\mathcal{E}[s(X^2, e^{Y_1, \pi})]^2 - s(X^1, X^2)\mathcal{E}[s(X^1, e^{Y_1, \pi})s(X^2, e^{Y_1, \pi})] \\
\text{by lemma 4.7:} \quad \mathcal{E}(\Delta_1 s(X^2, e^{Y_1, \pi})) = \Delta \left(s_Y^2 - [s_{Y1}]^2/s_1^2 \right) / (n-1)
$$

5. It may be shown in a similar way that

$$
\mathcal{E}(\Delta_2 s(X^1, e^{Y_1})) = \Delta \left(s_Y^2 - [s_{Y1}]^2/s_1^2 \right) / (n-1)
$$

**Proposition 4.3.** In the (non trivial) case $\Delta \neq 0$, $\Delta_1 \neq 0$, the permutational expectation of $s^2(\hat{\beta}_2^\pi)$ is equal to the permutational variance of $\hat{\beta}_2^\pi$:

$$
\mathcal{E}(s^2(\hat{\beta}_2^\pi)) = (s_Y^2 s_1^2 - [s_{Y1}]^2) / (\Delta(n-1)) = \text{Var}(\hat{\beta}_2^\pi)
$$

**Proof:** We use the lemma (4.8) to obtain by substitution and after simplification

$$
\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2 = (s_Y^2 - [s_{Y1}]^2/s_1^2) n(n-3) / (n-1).
$$

As a consequence:

$$
\mathcal{E}(s^2(\hat{\beta}_2^\pi)) = s_1^2 / (n(n-3)\Delta). \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2 = \text{Var}(\hat{\beta}_2^\pi)
$$

Consequently, the estimator of $\text{Var}(\hat{\beta}_2^\pi)$ used in the approach of Freedman and Lane is permutationally unbiased.

### 4.4.2 The Kennedy’s permutation method

The approach suggested by Kennedy is based on the equation (16) and the notation $(\sigma^K)^2 = \text{Var}(\epsilon^K)$ shows that: $\text{Var}(\hat{\beta}_2^\pi) = s_1^2 (\sigma^K)^2 / n\Delta$. The OLS estimator of $\text{Var}(\hat{\beta}_2^\pi)$ is: $\text{Var}(\hat{\beta}_2^\pi) = s^2(\hat{\beta}_2^\pi) = s^2(X^1)s^{2K} / n\Delta$. Here, as in the section (4.4.1), we calculate the permutational expectation of $\text{Var}(\hat{\beta}_2^\pi)$, by using the identity:

$$
\mathcal{E}(s^2(\hat{\beta}_2^\pi)) = \mathcal{E}(s_Y^2 \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 / (n(n-3)\Delta)) = s_1^2 / (n(n-3)\Delta) \mathcal{E}(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2).
$$
Therefore, the derivation of the permutational expectation of \( s^2(\hat{\beta}_2^\pi) \) depends on that of \( \mathcal{E}(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2) \). But, according to the lemma 2.4, \( \mathcal{E}(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2) = \sum_{i=1}^n (e_i^{Y,1})^2 - (n\mathcal{E}(\Delta_i^2))/(s^2_1\Delta) \). So, we are in a position to establish the following result:

**Proposition 4.4.** In the non trivial case \( \Delta \neq 0 \), we have:

\[
\mathcal{E}(s^2(\hat{\beta}_2^\pi)) = (s^2_1s^2_Y - [s_1Y]^2)(n - 2)/((n - 3)(n - 1)\Delta) = (n - 2)/(n - 3)\text{Var}(\hat{\beta}_2^\pi)
\]

**Proof:** By definition, \( \mathcal{E}(s^2(\hat{\beta}_2^\pi)) = s^2_1/(n(n - 3)\Delta)\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 \). One then obtains from lemma 4.8: \( \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 = (s^2_Y - s_1Y/s^2_1)n(n - 2)/(n - 1) \) which results in:

\[
\mathcal{E}(s^2(\hat{\beta}_2^\pi)) = s^2_1/(n(n - 3)\Delta)\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K) = ((n - 2)/(n - 3))\text{Var}(\hat{\beta}_2^\pi)
\]

This result shows that the approach of Kennedy derives an estimator of \( \text{Var}(\hat{\beta}_2^\pi) \) denoted by \( s^2(\hat{\beta}_2^\pi) \), which is a permutationally biased estimator of \( \text{Var}(\hat{\beta}_2^\pi) \), contrasting with to the estimator \( s^2(\hat{\beta}_2^\pi)_F \) given by the Freedman and Lane’s method.

**References**


Received: June, 2009