

# Permutation Predictor Tests in Linear Regression

Hye Min Ryu<sup>a</sup>, Min Ah Woo<sup>b</sup>, Kyungjin Lee<sup>c</sup>, Jae Keun Yoo<sup>1,c</sup>

<sup>a</sup>Clinical Research Coordination Center, National Cancer Center

<sup>b</sup>Korea Information and Communication Industry Institute

<sup>c</sup>Department of Statistics, Ewha Womans University

---

## Abstract

To determine whether each coefficient is equal to zero or not, usual  $t$ -tests are a popular choice (among others) in linear regression to practitioners because all statistical packages provide the statistics and their corresponding  $p$ -values. Under smaller samples (especially with non-normal errors) the tests often fail to correctly detect statistical significance. We propose a permutation approach by adopting a sufficient dimension reduction methodology to overcome this deficit. Numerical studies confirm that the proposed method has potential advantages over the  $t$ -tests. In addition, data analysis is also presented.

**Keywords:** Linear regression, non-normality, permutation test, small samples.

---

## 1. Introduction

Regression is a statistical methodology to study the conditional distribution of a response variable  $Y \in \mathbb{R}^1$  given a set of predictors  $\mathbf{X} \in \mathbb{R}^p = (X_1, \dots, X_p)^T$ . The regression seeks an investigation of how the response  $Y$  is changed given  $\mathbf{X}$ . Subsequently, the following model-based regression (*linear regression*) is popular:

$$Y | \mathbf{X} = \alpha + \beta^T \mathbf{X} + \varepsilon, \quad (1.1)$$

where  $\beta = (\beta_1, \dots, \beta_p)^T$  and  $\varepsilon \sim N(0, \sigma^2)$ .

In the model of (1.1), the vector of  $\beta$  represents unknown coefficients and measures the effects of predictors to the regression. For example, if  $\beta_i = 0$ , the corresponding predictor  $X_i$  does not contribute to the regression.

When numerous predictors are considered in the regression, there is a possibility that many unnecessary predictors are involved. Therefore it is best to select significantly important ones among all predictors to the regression because a regression model with smaller set of predictors may be easier to maintain as well as help to increase prediction accuracies. In regression, a procedure to select important predictors and to keep them alone is called *variable selection*.

There exist many variable selection procedures in the context. We briefly explain the classical and most widely used ones in practice. The first one should be  $t$ -tests for testing  $H_0 : \beta_i = 0$  for  $i = 1, \dots, p$ . The constructed test statistic is  $t_i^* = \hat{\beta}_i / \text{se}(\hat{\beta}_i)$ , where  $\hat{\beta}_i$  represents the ordinary least square (OLS) estimate of  $\beta_i$  and  $\text{se}(\hat{\beta}_i)$  does the corresponding standard error. Then, the statistic  $t_i^*$

---

For the corresponding author Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (2012-040077).

<sup>1</sup> Corresponding author: Associate Professor, Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong Soodaemun-Gu, Seoul 120-750, Korea. E-mail: [peter.yoo@ewha.ac.kr](mailto:peter.yoo@ewha.ac.kr)

follows  $t$  distribution with  $(n - p - 1)$  degrees of freedom under the normality and constant variance of  $\varepsilon$ . For testing multiple coefficients equal to zeros and testing  $H_0 : \beta_i = \beta_j = \cdots = \beta_k = 0$  for  $i \neq j \neq \cdots \neq k$ , general  $F$ -tests are often conducted. Let  $M_1$  stand for a current model and  $M_0$  do for a smaller model under the null hypothesis. Then the test statistic and its distribution are as follows:

$$F^* = \frac{(\text{SSE}_0 - \text{SSE}_1)/(df_0 - df_1)}{\text{SSE}_1/df_1} \sim F_{(df_0 - df_1), df_1},$$

where  $\text{SSE}_m$  and  $df_m$  represent residual sums of squares and its degrees of freedom for model  $M_m$ ,  $m = 0, 1$ , respectively.

Another way is to use Akaike's information criteria (AIC) and Schwarz' Bayesian information criteria (BIC). The formulas for AIC and BIC are as follows:  $\text{AIC} = n \log(\text{SSE}) - n \log(n) + 2p$  and  $\text{BIC} = n \log(\text{SSE}) - n \log(n) + [\log(n)]p$ . In both the cases, smaller values indicates more important predictors. The differences between AIC and BIC are placed on the third term. If  $n < 8$ , the term for AIC is larger than that for BIC. Therefore, the BIC tends to favor more parsimonious models according to Kuntner *et al.* (2005).

The  $t$  or general  $F$ -tests and the information criteria have quite different aspects to select variables. The former can directly decide which predictors do have null effects to the regression as well as which one is the least important. In addition, contrast effects between predictors are easily tested. This is why the two tests are mostly preferred to usual non-statistical practitioners over the information criteria. However, the latter can provide evidence on which ones are the least important but they cannot properly conduct evaluations of contrast effects. Therefore, the information criteria are typically preferred in use for automatic search procedures for variable selection such as backward, forward, and stepwise selections.

These popular ways for variable selection produce unreliable selection results under relatively smaller sample sizes than the number of parameters and non-normal error. False selections occur mainly due to type-II-error and the null hypothesis of  $H_0 : \beta_i = 0$  is not often rejected (although the hypothesis is not true). Therefore, a set of selected predictors are usually smaller than the true one.

To overcome this deficit, a permutation test is proposed by adopting sufficient dimension reduction. The sufficient dimension reduction is non-parametric and often works well in relatively smaller samples. The proposed permutation approach is expected to improve the problems of existing and popularly used methods discussed above; in addition, statisticians can easily use it and understand the results since the permutation approach provides  $p$ -values.

The paper is organized as follows. In Section 2 we define predictor effect hypothesis and propose permutation predictor effect tests in linear regression. Section 3 is devoted for numerical studies and real data application. Finally we summarize our work in Section 4.

## 2. Permutation Predictor Tests

### 2.1. Predictor effect hypothesis

Variable selection in a regression of  $Y \in \mathbb{R}^1 | \mathbf{X} \in \mathbb{R}^p$  is defined as conditional independences of the response given predictors. To illustrate this, first, partition the original  $p$ -dimensional predictors  $\mathbf{X}$  into  $(\mathbf{X}_H \in \mathbb{R}^h, \mathbf{X}_{H^\perp} \in \mathbb{R}^{(p-h)})$ . The subset  $\mathbf{X}_H$  have no effects to a regression of  $Y | \mathbf{X}$ , if and only if the complement set of  $\mathbf{X}_{H^\perp}$  alone is fully informative to the regression. Equivalently, the conditional distribution of  $Y | \mathbf{X}$  is equal to that of  $Y | \mathbf{X}_{H^\perp}$ , which can be equivalently rephrased as that  $Y$  is independent of  $\mathbf{X}$  given  $\mathbf{X}_{H^\perp}$ . The arguments discussed above can be summarized as the following three equivalent statements: (a) the subset  $\mathbf{X}_H$  has no effects to a regression of  $Y | \mathbf{X}$ ; (b)

$F_{Y|\mathbf{X}}(\cdot) = F_{Y|\mathbf{X}_{\mathbf{H}_\perp}}(\cdot)$ ; (c)  $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{X}_{\mathbf{H}_\perp}$ , where  $F(\cdot)$  and  $\perp\!\!\!\perp$  stands for a cumulative probability function and statistical independence, respectively.

Effects of predictors to regression can be evaluated as subsets of predictors as well as their contrasts. In this sense, variable selection is a special interest of predictor effects to regression. Therefore, we need to generalize variable selection to evaluate predictor effects. Let  $\mathcal{H}$ ,  $\mathbf{H}$ , and  $\mathbf{P}_\mathbf{H}$  represent an  $h$ -dimensional predictor subspace of  $\mathbf{X}$  with  $h \leq p$ , its  $p \times h$  orthonormal basis matrix, and the orthogonal projection operator onto  $\mathcal{H}$  with usual inner-product subspace, respectively. And,  $\mathbf{H}_\perp \in \mathbb{R}^{p \times (p-h)}$  stands for the orthonormal basis matrix for the orthogonal complement of  $\mathcal{H}$ . Then the effects of predictors are defined as the following equivalent statements: (a1)  $\mathbf{P}_\mathbf{H}\mathbf{X}$  has no effects to regression of  $Y | \mathbf{X}$ ; (b1)  $F_{Y|\mathbf{X}}(\cdot) = F_{Y|\mathbf{P}_\mathbf{H}\mathbf{X}}(\cdot)$ ; (c1)  $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{P}_{\mathbf{H}_\perp}\mathbf{X}$ , where  $\mathbf{P}_\mathbf{H} + \mathbf{P}_{\mathbf{H}_\perp} = \mathbf{I}_p$ .

Hereafter, any of the three statements given in (a1), (b1), and (c1) will be called *predictor effect hypothesis*. Let us illustrate the hypothesis. If one selects  $\mathbf{H}$  as the  $i$ th  $p$ -dimensional canonical basis vector,  $\mathbf{e}_i$  with the  $i$ th element equal to one and zeros elsewhere,  $\mathbf{P}_\mathbf{H}\mathbf{X}$  is equal to  $X_i$  in order to evaluate whether or not  $X_i$  is significantly important to regression. If one chooses  $\mathbf{H} = (0.5, -0.25, -0.25, 0, \dots, 0)^T$ , it is to determine whether or not the effect of  $X_1$  is equal to the average of the effects of  $X_2$  and  $X_3$ . The former example is a case of variable selection and the latter is of the contrasts.

## 2.2. Permutation predictor effect tests in linear regression

Under linear regression, effects of each predictor are measured by the corresponding regression coefficients. So, the non-parametric statements of predictor effect hypotheses are evaluated through the regression coefficients and we can formulate the evaluation as a procedure of parametric hypothesis tests.

Under linear regression in (1.1),  $\mathbf{P}_\mathbf{H}\mathbf{X}$  has no effects to regression of  $Y | \mathbf{X}$ , if and only if  $\mathbf{P}_\mathbf{H}\boldsymbol{\beta} = 0$ , equally  $\mathbf{H}^T\boldsymbol{\beta} = 0$ . Therefore, the effect of  $\mathbf{P}_\mathbf{H}\mathbf{X}$  is easily evaluated through testing the hypotheses of  $H_0 : \mathbf{H}^T\boldsymbol{\beta} = 0$  versus  $H_1 : \mathbf{H}^T\boldsymbol{\beta} \neq 0$ . Hereafter, for notational conveniences, we define that  $\mathbf{X}_\mathbf{H} = \mathbf{P}_\mathbf{H}\mathbf{X}$  and  $\mathbf{X}_{\mathbf{H}_\perp} = \mathbf{P}_{\mathbf{H}_\perp}\mathbf{X}$ . Popular  $t$ -tests allow us to conduct these hypothesis tests easily, but they have limits in use under relatively smaller samples and non-normal errors.

The conditional statement (c1) in Section 2.1 has the same form as the key statement used in sufficient dimension reduction (SDR). Its main idea is to replace the original  $p$ -dimensional predictor  $\mathbf{X}$  by lower-dimensional linearly-transformed predictors  $\eta^T\mathbf{X}$  with  $\eta \in \mathbb{R}^{p \times m}$  and  $m < p$  such that

$$Y \perp\!\!\!\perp \mathbf{X} | \eta^T\mathbf{X}. \quad (2.1)$$

The minimal subspace among all possible subspaces spanned by the columns of  $\eta$  satisfying (2.1) and its dimension are called the *central subspace* and the *structural dimension*. For more about sufficient dimension reduction, readers can refer Cook (1998).

The same representation of (c1) and (2.1) offer a possibility to adopt methodologies used in SDR. Here we use a permutation approach to test the null hypothesis of  $\mathbf{H}^T\boldsymbol{\beta} = 0$ . The permutation tests in SDR were suggested in Cook and Weisberg (1991) and formalized in Yin and Cook (2002) to estimate the structural dimension. However, predictor effect hypotheses must be tested here and do not have a direct application to our case.

The key part in the permutation test procedure proposed by Yin and Cook (2002) is to permute  $\hat{\eta}_\perp^T\mathbf{X}$  to generate permuted samples, where the  $p \times (p - m)$  matrix  $\hat{\eta}_\perp$  is the orthonormal complement matrix of  $\hat{\eta} \in \mathbb{R}^{p \times m}$  such that  $\hat{\eta}_\perp^T\hat{\eta}_\perp = \mathbf{I}_{(p-m)}$  and  $\hat{\eta}_\perp^T\hat{\eta} = 0$ . Therefore, we have to find appropriate objects randomly permuted for adapting the permutation tests into our problem. For this, based on the

null hypothesis of  $\mathbf{H}^T \mathbf{X}$ , the original predictor  $\mathbf{X}$  is partitioned into  $(\mathbf{X}_H = \mathbf{H}^T \mathbf{X}, \mathbf{X}_{H_\perp} = \mathbf{H}_\perp^T \mathbf{X})$ . Then, since  $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{H}_\perp^T \mathbf{X}$  under the null hypothesis, we can randomly permute  $\mathbf{X}_H$ . We will denote permuted samples of  $\mathbf{X}$  as  $\mathbf{X}^{\text{perm}} = (\mathbf{X}_H^{\text{perm}}, \mathbf{X}_{H_\perp})$ .

Next, to conduct predictor effect tests through the random permutations, we need to define a statistic to evaluate whether or not  $\mathbf{H}^T \beta = 0$ . For this we think of the following reasoning. If the null hypothesis is true, we expect that the distances between two vectors of  $\hat{\beta}$  and  $\hat{\beta}_H = \mathbf{P}_{H_\perp} \hat{\beta}$  must be relatively close. After normalizing  $\hat{\beta}$  and  $\hat{\beta}_H$  to have unit length, the following distance is measured between the two:

$$\hat{\lambda}_H = 1 - \hat{\beta}^T (\hat{\beta}_H \hat{\beta}_H^T) \hat{\beta}. \quad (2.2)$$

To understand  $\hat{\lambda}_H$ , we consider two extreme cases. The first extreme case is that  $\beta = \beta_H$ . That is, a subspace spanned by  $\beta$  is equivalent to a subspace of the orthogonal complement of  $\mathcal{H}$ . If so, the value of  $\beta^T (\beta_H \beta_H^T) \beta$  after normalization is equal to one, and hence the value of  $1 - \beta^T (\beta_H \beta_H^T) \beta$  becomes zero. The second extreme case is  $\beta = \mathbf{P}_H \beta$ . A subspace spanned by  $\beta$  is a subspace of  $\mathcal{H}$ . Then, we have  $\beta^T (\beta_H \beta_H^T) \beta = 0$ , and hence the value of  $1 - \beta^T (\beta_H \beta_H^T) \beta$  is equal to one. From the two extreme cases,  $\hat{\lambda}_H$  should vary from zero and one. The first extreme case holds, if the null hypothesis is true. So relatively smaller values of  $\hat{\lambda}_H$  or ones close to zero imply that the null hypothesis is true. However, if  $\hat{\lambda}_H$  is relatively larger or close to one, then the null hypothesis is indicated to be false.

Here  $\hat{\lambda}_H$  in (2.2) is proposed as a statistic to test  $H_0 : \mathbf{H}^T \beta = 0$ . We compute  $\hat{\lambda}_H$  from the original sample and permuted samples that will be respectively denoted as  $\hat{\lambda}_H^{\text{ref}}$  and  $\hat{\lambda}_H^{\text{perm}}$ . Then an empirical sampling distribution of  $\hat{\lambda}_H^{\text{ref}}$  is constructed by  $\hat{\lambda}_H^{\text{perm}}$  computed from each permuted sample. Finally,  $p$ -values to test  $H_0 : \mathbf{H}^T \beta = 0$  are calculated as the percentages of  $\hat{\lambda}_H^{\text{perm}} > \hat{\lambda}_H^{\text{ref}}$ . We summarize the procedure as follows:

#### Algorithm of permutation predictor tests

1. Obtain the OLS estimate  $\hat{\beta}$  of  $\beta$  from the original sample of  $Y$  and  $\mathbf{X}$ .
2. Based on  $H_0 : \mathbf{H}^T \beta = 0$ , partition  $\mathbf{X}$  into  $(\mathbf{X}_H = \mathbf{H}^T \mathbf{X}, \mathbf{X}_{H_\perp} = \mathbf{H}_\perp^T \mathbf{X})$  with  $\mathbf{H}_\perp^T \mathbf{H} = 0$ .
3. Compute  $\hat{\beta}_H = \mathbf{P}_{H_\perp} \hat{\beta}$  and normalize  $\hat{\beta}$  and  $\hat{\beta}_H$ . Then calculate  $\hat{\lambda}_H^{\text{ref}}$ .
4. Permute  $\mathbf{X}_H$  randomly and construct permuted predictors of  $\mathbf{X}^{\text{perm}} = (\mathbf{X}_H^{\text{perm}}, \mathbf{X}_{H_\perp})$ .
5. Obtain the OLS estimate  $\hat{\beta}^{\text{perm}}$  from the permuted samples of  $Y$  and  $\mathbf{X}^{\text{perm}}$ .
6. Compute  $\hat{\beta}_H^{\text{perm}} = \mathbf{P}_{H_\perp} \hat{\beta}^{\text{perm}}$  and normalize  $\hat{\beta}^{\text{perm}}$  and  $\hat{\beta}_H^{\text{perm}}$ . Then calculate  $\hat{\lambda}_H^{\text{perm}}$ .
7. Repeat steps 4 to 6  $K$  times.
8. Calculate the percentages of  $\hat{\lambda}_H^{\text{perm}} > \hat{\lambda}_H^{\text{ref}}$  among all  $\hat{\lambda}_H^{\text{perm}}$ s, and report the percentages as  $p$ -values for testing  $H_0 : \mathbf{H}^T \beta = 0$ .

### 3. Simulation Studies and Data Analysis

#### 3.1. Simulation studies

For all numerical studies, five dimensional predictors  $\mathbf{X} = (X_1, \dots, X_5)^T$  were generated from multivariate normal distributions with mean 0 and the correlation matrix  $\Sigma$ , whose diagonal and off-diagonal elements are (respectively) equal to one and  $\rho$  for  $\rho = 0.3, 0.5, 0.7$ .

The following artificial model was constructed:

$$Y | \mathbf{X} = X_1 + X_2 + 0.5X_3 + \varepsilon,$$

where a random error  $\varepsilon$  is distributed as one of the three distributions of standard normal distribution,  $\chi^2$  distributions with two degrees of freedom, and  $t$  distributions of three degrees of freedom. In addition we considered 15, 30, 50 and 100 sample sizes and the model was iterated 500 times.

By construction, the model has the following values for the regression coefficients:

$$\beta_1 = \beta_2 = 1; \quad \beta_3 = 0.5; \quad \beta_4 = \beta_5 = 0.$$

For this model, we tested the following seven hypothesis:

$$\begin{aligned} H_0 : \beta_i = 0 & \quad \textit{versus} \quad H_1 : \beta_i \neq 0, \quad i = 1, \dots, 5; \\ H_0 : \beta_1 = \beta_2 & \quad \textit{versus} \quad H_1 : \beta_1 \neq \beta_2; \\ H_0 : \beta_2 = \beta_3 & \quad \textit{versus} \quad H_1 : \beta_2 \neq \beta_3. \end{aligned}$$

For  $i = 1, 2, 3$ , the null hypothesis of  $\beta_i = 0$  is false, while it is true for  $i = 4, 5$ . Also, the null hypothesis of  $\beta_1 = \beta_2$  is true, but that of  $\beta_2 = \beta_3$  is false.

To report the numerical studies, the percentages of the rejections of each null hypothesis were computed with level 5%. Therefore, in the case that the null hypotheses are true, the rejection percentages indicates the observed levels, and the tests are more desirable, if the percentages are close to 5%, which is the nominal level. However, if the null hypotheses are false, the rejection percentages represent the observed powers and the tests become more powerful as they are closer to 100%.

To test  $H_0 : \beta_i = 0$ , we used usual  $t$ -tests in linear regression and the proposed permutation tests for  $\mathbf{H} = \mathbf{e}_i \in \mathbb{R}^5$ ,  $i = 1, \dots, 5$ . As results, we report the rejection percentages of  $H_0 : \beta_3 = 0$  and  $H_0 : \beta_4 = 0$ , because the asymptotic behaviors of testing  $H_0 : \beta_i = 0$  for  $i = 1, 2$  by the proposed permutation tests and  $t$ -tests are similar to those of  $H_0 : \beta_3 = 0$ . Omitting the results for  $H_0 : \beta_5 = 0$  is the same reason. Tables 1–3 summarize these results. In addition, the permutation tests for  $\mathbf{H} = (1, -1, 0, 0, 0)^T$  and  $\mathbf{H} = (0, 1, -1, 0, 0)^T$  and general  $F$ -tests were conducted to test  $H_0 : \beta_1 = \beta_2$  and  $H_0 : \beta_2 = \beta_3$ . We summarize these numerical studies in Table 4 and Table 6.

Tables 1–3 represent the characteristic patterns observed in other simulations. Regardless of distributions of random errors and correlation among the predictors, the proposed permutation tests are more powerful than usual  $t$ -tests. Also the power estimations in usual  $t$ -tests are greatly affected, especially with smaller samples, by the distributions of the random errors, while they are not in the permutation tests. Overall, with smaller samples, the proposed permutation tests clearly outperforms the  $t$ -tests, and reaches 100% faster than the  $t$ -tests. The permutation tests are affected by the correlation among the predictors so that their higher correlations make the observed levels worse, while usual  $t$ -tests are consistent to the correlation in the estimation of the nominal level. In the permutation samples, effects of unimportant predictors tend to be overestimated, and this can be partial reason that the permutation tests have larger observed levels than actual levels. From the simulation, we learn that the permutation tests often tend to reject the hypothesis of  $\beta_i = 0$ . This implies that the permutation tests would keep more variables than usual  $t$ -tests in practice. With smaller samples, this should be a good property, in sense that it provides strong evidence that the corresponding predictors are not important, if the tests do not reject  $H_0 : \beta_i = 0$ ,  $i = 1, \dots, p$ .

According to Tables 4–6, there are no notable differences between the permutation tests and general  $F$ -tests in the estimations of level and power.

Table 1: Percentages of rejection of  $H_0 : \beta_i = 0$  for  $i = 3, 4$  with  $\varepsilon \sim N(0, 1)$ ; Perm: permutation tests;  $t$ :  $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$	
	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$
$n = 15$	40.0	31.0	10.0	3.0	50.0	24.0	10.0	2.0	53.0	15.0	29.0	6.0
$n = 30$	65.0	53.0	8.0	7.0	66.0	40.0	11.0	5.0	71.0	34.0	21.0	6.0
$n = 50$	84.0	77.0	6.0	5.0	87.0	74.0	9.0	3.0	78.0	50.0	20.0	6.0
$n = 100$	100	100	5.0	7.0	100	96.0	8.2	4.0	98.0	80.0	17.0	4.0

Table 2: Percentages of rejection of  $H_0 : \beta_i = 0$  for  $i = 3, 4$  with  $\varepsilon \sim \chi_1^2$ ; Perm: permutation tests;  $t$ :  $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$	
	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$
$n = 15$	31.0	25.0	12.0	3.0	46.0	21.0	17.0	5.0	44.0	12.0	36.0	4.0
$n = 30$	56.0	41.0	6.0	7.0	59.0	33.0	19.0	7.0	64.0	24.0	26.0	6.0
$n = 50$	66.0	51.0	6.0	5.0	65.0	40.0	12.0	3.0	64.0	23.0	25.0	5.0
$n = 100$	94.0	92.0	7.0	4.0	89.0	82.0	11.0	7.0	82.0	52.0	17.0	5.0

Table 3: Percentages of rejection of  $H_0 : \beta_i = 0$  for  $i = 3, 4$  with  $\varepsilon \sim t_3$ ; Perm: permutation tests;  $t$ :  $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$		$\beta_3 = 0$		$\beta_4 = 0$	
	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$
$n = 15$	29.0	17.0	9.0	6.0	22.0	5.0	17.0	1.0	38.0	9.0	27.0	7.0
$n = 30$	40.0	24.0	7.0	3.0	45.0	24.0	13.0	4.0	44.0	11.0	23.0	4.0
$n = 50$	64.0	45.0	7.0	5.0	60.0	40.0	10.0	4.0	60.0	27.0	22.0	4.0
$n = 100$	84.0	73.0	6.0	5.0	82.0	67.0	9.0	2.0	78.0	52.0	17.0	2.0

In the coordinate tests, the permutation tests and the  $t$ -tests produced quite different results, while the permutation tests and general  $F$ -tests were similar in the contrast tests. The reason why both the approaches produced similar results for the contrast tests in linear regression can be thought of as follows. For the coordinate effect tests, both the test statistics can be computed from one OLS fit computed under the alternative hypothesis, although it can be computed from the fits under the null and alternative hypotheses for usual  $t$ -tests. However, the two tests use the same information, so permutation tests yield more desirable results. However, for the contrast tests general  $F$ -tests should use information acquired from the two fits under the null and alternative hypotheses, while the permutation tests have to use information only under the alternative hypothesis. Usage of the supplemental information obtained from the null hypothesis in general  $F$ -tests is a possible reason why the permutation tests and general  $F$ -tests showed similar results in numerical studies.

### 3.2. Data analysis

In this section, the proposed method is applied to a cloud seeding data introduced in Cook and Weisberg (1999, p.313). The data was collected to judge the success or failure of cloud seeding designed to increase rainfall. The original data contains seven variables; action ( $A$ ) with 1 if seeded, or 0 otherwise; a time trend; the suitability ( $S$ ) for seeding score; cloud over in the experimental area; prewetness ( $P$ ); echo motion category with 1 or 2, a measure of the type of clouds present; rainfall ( $R$ ) in the target area following action. For illustration purpose, we consider the following regression:

$$E\{\log R | (A, S, P)\} = \alpha + \beta_1 A + \beta_2 \log P + \beta_3 S + \beta_4 AS + \beta_5 A \log P. \quad (3.1)$$

Table 4: Percentages of rejections of  $H_0 : \beta_1 = \beta_2$  and  $H_0 : \beta_2 = \beta_3$  with  $\varepsilon \sim N(0, 1)$ ; Perm:permutation tests;  $t$ : $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$	
	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF
$n = 15$	9.0	8.0	10.0	7.0	3.0	3.0	5.0	5.0	7.0	5.0	8.0	9.0
$n = 30$	6.0	4.0	35.0	31.0	5.0	6.0	26.0	25.0	9.0	9.0	17.0	16.0
$n = 50$	3.0	2.0	49.0	52.0	8.0	9.0	37.0	35.0	5.0	6.0	20.0	19.0
$n = 100$	5.0	5.0	79.0	79.0	5.0	8.0	67.0	65.0	9.0	9.0	46.0	45.0

Table 5: Percentages of rejections of  $H_0 : \beta_1 = \beta_2$  and  $H_0 : \beta_2 = \beta_3$  with  $\varepsilon \sim \chi_1^2$ ; Perm:permutation tests;  $t$ : $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$	
	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF
$n = 15$	6.0	6.0	11.0	16.0	8.0	10.0	13.0	12.0	5.0	6.0	6.0	5.0
$n = 30$	3.0	3.0	22.0	23.0	6.0	7.0	19.0	17.0	3.0	2.0	12.0	9.0
$n = 50$	4.0	5.0	34.0	34.0	2.0	1.0	29.0	29.0	4.0	5.0	21.0	20.0
$n = 100$	7.0	7.0	53.0	52.0	5.0	5.0	51.0	49.0	1.0	1.0	35.0	33.0

Table 6: Percentages of rejections of  $H_0 : \beta_1 = \beta_2$  and  $H_0 : \beta_2 = \beta_3$  with  $\varepsilon \sim t_3$ ; Perm:permutation tests;  $t$ : $t$ -test

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
	$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$		$\beta_1 = \beta_2$		$\beta_2 = \beta_3$	
	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF	Perm	GF
$n = 15$	2.0	4.0	9.0	14.0	3.0	2.0	9.0	7.0	6.0	6.0	8.0	5.0
$n = 30$	7.0	6.0	15.0	17.0	3.0	3.0	17.0	14.0	8.0	5.0	10.0	10.0
$n = 50$	4.0	5.0	29.0	26.0	8.0	7.0	23.0	25.0	7.0	8.0	17.0	16.0
$n = 100$	5.0	4.0	47.0	46.0	6.0	6.0	37.0	36.0	6.0	6.0	26.0	27.0

Table 7:  $P$ -values for  $H_0 : \beta_i = 0, i = 1, \dots, 5$  for the model described in (3.1); Perm: permutation tests;  $t$ :  $t$ -test

	$\beta_1 = 0$		$\beta_2 = 0$		$\beta_3 = 0$		$\beta_4 = 0$		$\beta_5 = 0$	
	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$	Perm	$t$
$p$ -values	0.02	0.67	0.03	0.75	0.02	0.21	0.42	0.19	0.19	0.65

In model (3.1), the two variables of  $R$  and  $P$  are transformed for diagnostic purpose, and all main effects and two-way interactions between  $A$  and  $S$  and between  $A$  and  $\log P$  are presented.

Now fit the model and the corresponding  $p$ -values for testing  $H_0 : \beta_i = 0, i = 1, \dots, 5$ , are summarized in Table 7.

Table 7 indicates that the permutation tests determine that all main effects are significant with level 5%, while usual  $t$ -tests conclude that all effects under consideration are not significant. To determine which results are more reliable, we investigate a scatter plot matrix of  $S, \log P$  and  $\log R$  only corresponding action equal to one (Figure 1). In the scatter plot matrix, the red solid lines represents the ordinary least square fits and the red dotted lines do LOWESS fits with a smoothing parameter fixed at  $2/3$ . If there are no conditional association of rainfall given action, suitability and prewetness, the scatter plot must show no relations on  $\log R | \log P$  and  $\log R | S$ . However, we can clearly observe the linear relationship in  $\log R | S$ , and it can be concluded that the results from usual  $t$ -tests are not reliable.

After removing the two two-way interactions from model (3.1), refit and define that  $\hat{\beta}^T \mathbf{X} = \hat{\beta}_1 A + \hat{\beta}_2 \log P + \hat{\beta}_3 S$ . Then, as a summary plot of the regression, we report a scatter plot in Figure 2 of  $\log R$  versus  $\hat{\beta}^T \mathbf{X}$  with imposing LOWESS line with the same parameter configurations as Figure 1. Figure 2 indicates that we can observe a quadratic relation between  $\hat{\beta}^T \mathbf{X}$  and  $\log R$ .

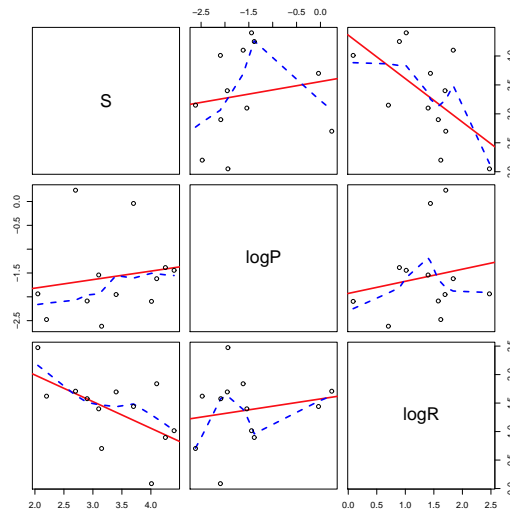


Figure 1: A scatter plot matrix of  $S$ ,  $\log P$  and  $\log R$  only with action equal to one; solid red line, OLS line; blue dotted line, LOWESS line

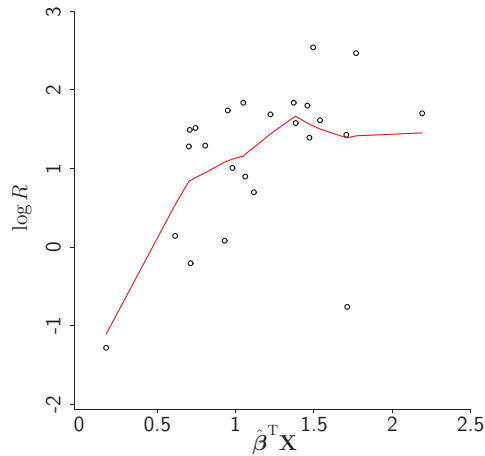


Figure 2: A scatter plot of  $\hat{\beta}^T \mathbf{X}$  and  $\log R$

#### 4. Discussions

To determine whether each coefficient is equal to zero or not, usual  $t$ -tests are popular choice among others in linear regression to usual practitioner, because all statistical packages provide the statistics and their corresponding  $p$ -values. Under smaller samples (especially with non-normal errors) the tests often fail to detect the statistical significance correctly. To overcome this deficit, we propose a permutation approach by adopting a sufficient dimension reduction methodology.

For this, first we define general predictor hypothesis that contain contrasts. To implement the proposed permutation approach, first, a reference statistics is computed through measuring the distances between the OLS estimates and restricted OLS estimates under user-selected predictor hypothesis. By linearly transformed predictors following the predictor hypothesis and permuting them, we obtain permutation samples. From each permutation sample, the same statistics as the reference one are



computed. Then the percentages that the permuted statistics are bigger than or equal to the reference one are reported as  $p$ -values to just the predictor hypothesis. According to numerical studies, the proposed method provides higher powers than usual  $t$ -tests regardless of sample sizes and distributions of errors; however, it tends to produce higher observed levels.

The proposed permutation tests are not always a winner over usual  $t$ -tests; however, it provides additional reliable evidences with smaller samples to determine whether or not each regression coefficient is equal to zero. Therefore, if one uses both tests, it is out of question that they can make more correct decisions. The R codes for the proposed tests are available upon requests.

## References

- Cook, R. D. (1998) *Regression Graphics*, Wiley, New York.
- Cook, R. D. and Weisberg, S. (1991). Discussion of sliced inverse regression for dimension reduction by K.C. Li. *Journal of the American Statistical Association*, **86**, 328–332.
- Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*, Wiley, New York.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models*, 5th edition, McGraw-Hill/Irwin, New York.
- Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional  $k$ th moment in regression. *Journal of Royal Statistical Society, Series B*, **64**, 159–175.

*Received February 19, 2013; Revised March 18, 2013; Accepted March 25, 2013*