Senior Thesis in Mathematics

Permutation Tests: A Deep Dive into Applications in Multiple Linear Regression

Author:
Lauren Quesada

Advisor:
Johanna Hardin

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

June 30, 2022

Abstract

A permutation test is a hypothesis test to determine how likely an observed statistic is to come from a null population. In the regression setting, permuting a single covariate (here, "naïve permutation") can create a null relationship between that covariate and the response variable. But there's a catch: while permutation tests are useful for simple linear regression (SLR), their application in multiple linear regression (MLR) is more complex than permuting a single explanatory variable, one at a time. When permuting a single explanatory variable in a MLR model, exchangeability is not preserved unless the variables are independent: the distribution of our permuted variable is not maintained as we randomize it. While current literature serves to inform alternative permutation structures (as opposed to naïve permuting), there lacks sufficient insight quantifying and understanding the degree to which naïve permutation tests are problematic.

We have conducted extensive simulations on multiple approaches to permutation tests in the MLR setting. By varying the model structure, effect size, and sample size, we take a deep dive into understanding various permutation method results in terms of type I error and power. Additionally, we vary whether the model is correctly specified or mis-specified, an additional nuance which informs the variability in error rates that are observed.

Contents

Chapter 1

Statistical Inference

Statistical inference tests begin with a research question. In the case of hypothesis testing, here, permutation tests, the motivating question inquires about the relationship between two quantitative variables. Throughout this paper, the question pertains to the presence (or lack) of a linear relationship between the response variable, Y_i, and one of the explanatory variables, x_{i2}, in the multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

Above, β_0 denotes the intercept, β_1 and β_2 are partial regression coefficients, x_{i1} is another explanatory variable, and ϵ_i are the true errors. The test selected to draw inferences about the linear correlation between Y_i and x_{i2} is the permutation test. Historically, permutation test procedures were inaccessible, cumbersome, and computationally intensive [Küchler, 1999]. However, given their demonstrated robustness in simple linear regression applications and the advancements of modern computing, this paper will advocate for their use. In particular, it will support their extension to multiple linear regression, and quantify their shortcomings in multiple linear regression.

In the permutation tests described here, the test statistic calculated is the t-statistic, which is a choice that will be motivated in the literature section. The hypotheses from which conclusions will ultimately be drawn are:

$$H_o : \beta_2 = 0$$

$$H_a : \beta_2 \neq 0$$

The null hypothesis, H_o, asserts there is no linear relationship in the population between Y_i and x_{i2}. The alternate hypothesis, H_a, concludes the opposite. Using both observed and permuted t-statistic, a conclusion about the hypotheses is made. Conceptually, if there is no linear relationship between Y_i and x_{i2} in the population, the slope corresponding to x_{i2} in the multiple regression model is zero (hence $\beta_2 = 0$).

1.1 Permuting in Simple Linear Regression

Before beginning discussion of permutation tests on multiple linear regression models, it's important to understand their applicability to simple linear regression. Because of the robust nature of the permutation test to non-normal data, it is extremely useful in applications of simple linear regression models. The ultimate goal of this simulation study is that this usefulness can be transferred to multiple linear regression, where more than one explanatory variable is present and there may/may not exist a dependency between the X variables themselves.

1.1.1 Advantages of Permutation Tests

Permutation tests require fewer assumptions about the data than ordinary least squares (OLS). They do not need data or errors to be normally distributed and are useful for any distribution. Moreover, they can draw inferences from small samples sizes. Permutation tests are exact if you perform $n!$ permutations, one for every possible reordering of the data.It can however, be computationally intensive to calculate all $n!$ orderings of observations within a sample. If n is quite large, a substantial, yet smaller, subset of n is used such that $m < n!$. With so many combinations of data, inferences drawn from a permutation test that uses m shuffles is approximately exact, though cannot be truthfully called an exact test. With each shuffling of the data, a different data set is obtained. As values are reassigned, each reordering of the data is a potential outcome in the population. Permutation tests take into account *every single possible ordering of data* given the observed sample if $n!$ permutations are performed.

1.1.2 Type I Error Rate

The motivation for permutation tests is to create a process for differentiating between the null and alternate hypotheses such that type I error rate is controlled. Again, this can be done approximately with $m < n!$ shuffles, or exactly with $n!$ permutations of the sample. The inference test then utilizes the sample of observations to draw conclusions about the behavior of the true population. The results will lead one to either reject the null hypothesis, or fail to reject the null hypothesis. In the first case, the alternate hypothesis (H_a) is accepted. In the second, we fail to accept the alternate hypothesis (H_a), and fail to make any conclusions about the hypotheses. These conclusions are drawn in terms of our hypotheses, though sometimes the conclusions are incorrect. If the inference test of the sample leads us to conclude there is a relationship between Y_i and x_{i2} and there is in fact no linear correlation between the two variables in the population (i.e., the null is true and it is rejected), then the wrong conclusion has been drawn. This rate of false positivity is referred to as the type I error rate, and inference tests aim to control it at a set value, typically $\alpha = 0.05$. Since permutation tests consider every rearrangement of the data, or at least do so approximately, they have the ability to fix the type I error rate at $\alpha \leq 0.05$. Consequently, a permutation test in simple linear regression falsely rejects the null hypothesis no more than 5% of the time.

1.1.3 Permutation Test Assumptions

Permutation tests require fewer assumptions than other inference tests, although they still necessitate certain requirements of data. In particular, the errors in the simple linear regression model must be independent and identically distributed (iid). This is typical for all inference tests, though normal distribution of the error terms is an additional requirement for OLS. The residuals in the sample must not influence each other (independence) *and* come from the same distribution (identically distributed), though a distribution is not specified. Permutation tests also require exchangeability. This ensures the distribution of the permuted variable will be maintained as it is randomized.

Definition 1.1. Exchangeability: The distribution of the permuted variable is maintained as it is randomized [Aldous, 1985].

1.1.4 Power

Opposite of a false positive (type I error), is a false negative. A type II error rate occurs when the inference test suggests the null hypothesis not be rejected for the sample though it is false in relation to the population. If a relationship is not found between Y_i and x_{i2} where one exists in the population (i.e., we don't identify a relationship given the null hypothesis is false), this is a type II error. Conversely, power is the probability of rejecting the null hypothesis when the alternate is true:

$$Power = 1 - P(type\ II\ error\ rate)$$

Although power cannot be controlled, the analyst still hopes for a test with large power. Power is an important measurement of how well an inference test performs, and will be used to compare various permutation methods throughout this paper.

1.1.5 A Permutation Test for a SLR Model

First a basis is provided for the inference test in the simple linear regression case: Given a data set with one explanatory variable to which a linear model has been fit, a permutation test can inquire about the linear relationship between the response variable Y_i and the explanatory variable x_{i1}. Before starting, a significance level will be set at $\alpha = 0.05$. Assumptions to be made are:
1. Exchangeability
2. Errors are iid,
Also, hypotheses will be stated as follows:
$H_o : \beta_1 = 0$
$H_a : \beta_1 \neq 0$,
In the above assignment of the null hypothesis, H_o claims there is no linear relationship between the variables of interest, while the alternate hypothesis (H_a) claims there is. Given the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

where Y_i has been regressed on x_{i1}, β_0 is the intercept of the line, β_1 is the slope of the line, and ϵ_i is the errors, the permutation procedure can begin.

To inquire about the linear relationship between Y_i and x_{i1}, the permutation test proposes a shuffling all observations associated with **either** variable

x_{i1} or x_{i2}. First, a t-statistic is calculated for the raw, observed unshuffled sample:

$$t_{ref} = \frac{b_1}{se(b_1)}$$

in which b_1 is the sample estimate of the population parameter, β_1, and $se(b_1)$ it its standard error. Next, either the Y_i **or** the x_{i1} values are permuted. For this context, the x_{i1} values will be shuffled, but note the results would be the same because either way, all response values will be reassigned to permuted x_{i1} values. For each permutation, the data is reordered and a new linear model is fit:

$$\hat{Y_i} = b_0^* + b_1^* x_{i1}$$

for which a "*" indicates a permutation has been performed, so this is not the original fit, nor the original order of the data. Additionally, a new t-statistic is calculated for each shuffle of the data:

$$t^* = \frac{b_1^*}{se(b_1^*)}$$

After a large number of permutations ($m < n!$ if all $n!$ permutations are too computationally intensive), a distribution of t^* values is achieved (one t^* value for each permutation results in a distribution of t^* values). The original observed t_{ref} value for the sample is then placed in the distribution of many t^* values. If the proportion of $|t^*|$ values greater than or equal to $|t_{ref}|$ is smaller than the predetermined critical level, $\alpha = 0.05$, then the null hypothesis is rejected. There is a significant linear relationship between Y_i and x_{i1}, thus it's concluded $\beta_1 \neq 0$.

In summary, permutation tests operate much like other inference tests. They answer a question about a linear relationships between variables in a data set. Using these tests, conjectures can be answered about the behavior of the population from which the sample comes. More specifically, permutation tests infer how likely our observed data is to occur in the population by comparing the response variable to every possible rearrangement of itself. The inference test quantifies each possible sample using a test statistic, and measures the likelihood of occurrence of our original observed data with a p-value (the aforementioned proportion). Whether this p-value is less than or equal to a significance level α gives a statistician information about the rarity of the observed data under the null hypothesis, and leads to a final conclusion of one of the two hypotheses. With a slew of advantages and demonstrated reliable performance, permutation tests are a pragmatic method of statistical inference test for simple linear regression models.

Chapter 2

Permutation Tests in Multiple Linear Regression

Multiple linear regression introduces multiple explanatory variables into the OLS model. Whereas previously, in simple linear regression models, only x_{i1} was present, multiple linear regression allows for $x_{i1}, x_{i2}, ..., x_{ip}$ explanatory variables where $p \in \mathbb{N}$. For the purposes of simplicity and comprehension, this paper will consider the multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

which is only concerned with two explanatory variables, x_{i1} and x_{i2}.

If the goal is to use a permutation test to assess the significance of the linear coefficients, this model immediately raises a pertinent question: which variable should be permuted? While this decision can be informed by the motivating question of the inference, permuting only the variable of interest (as in SLR) fails to recognize the covariate relationships which are potentially present between the multiple explanatory variables. It is unknown whether Y_i and x_{i1}, Y_i and x_{i2}, or x_{i1} and x_{i2} are correlated. Permutation of any of the individual variables will disrupt these relationships, if present, and consequently prove problematic for inquiries into their relatedness. This paper discusses how a statistician might inquire about one variable without disrupting the covariate relationships, using a permutation test. This paper will concern itself with how to isolate x_{i2} to test for a relationship with y_i **without** the influence of possible dependency on x_{i1}. Three methods of permutation are discussed in the following chapters, including the advantages and pitfalls of their respective procedures and results. Ultimately, a simulation study is presented which not only reproduces results found in literature, but quantifies the deviation of each method from the standard test (the t-test) for an extensive variety of partial regression coefficients, distributions, sample sizes, and covariance structures.

Before beginning an extensive literature review on permutation approaches, this paper presents a summary table of the methods, as well as key differences between them. Going forward, the permutation methods are referred to either by their assigned number (1a, 1b, 2, or 3), or by their seminal author.

Permutation Method:	1a	1b	2	3
Author:	Freedman and Lane	ter Braak	Oja	Manly
Linear Model Fit:	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$
Values Recorded:	• OLS p-value • OLS t-statistic • b_2 • $R_{y_i\|x_{i1}x_{i2}}$ • fitted values for $y_i \sim x_{i1} + x_{i2}$	• OLS p-value • OLS t-statistic • b_2 • $R_{y_i\|x_{i1}}$ • fitted values for $y_i \sim x_{i1}$	• OLS p-value • OLS t-statistic • b_2	• OLS p-value • OLS t-statistic • b_2
Permuted Variable:	$R_{y_i\|x_{i1}x_{i2}}$	$R_{y_i\|x_{i1}}$	x_2	y
Permuted Linear Model Fit:	$R^*_{y_i\|x_{i1}x_{i2}} + b_0 + b_1x_{i1} + b_2x_{i2} \sim y$	$R^*_{y_i\|x_{i1}} + b_0 + b_1x_{i1} \sim y + x_{i2}$	$y \sim x_{i1} + x^*_{i2}$	$y^*_i \sim x_{i1} + x_{i2}$
Values Recorded: (for each permutation)	b^*_2, $se(b^*_2)$	b^*_2, $se(b^*_2)$	b^*_2, $se(b^*_2)$	b^*_2, $se(b^*_2)$
Test Statistic:	$t^* = \frac{b^*_2 - b_2}{se(b^*_2)}$	$t^* = \frac{b^*_2 - 0}{se(b^*_2)}$	$t^* = \frac{b^*_2 - 0}{se(b^*_2)}$	$t^* = \frac{b^*_2 - 0}{se(b^*_2)}$
P-value Calculation:	$\frac{\sum \|t^*\|}{n_{shuffle}} > \|b_2\|$	$\frac{\sum \|t^*\|}{n_{shuffle}} > \|b_2\|$	$\frac{\sum \|t^*\|}{n_{shuffle}} > \|b_2\|$	$\frac{\sum \|t^*\|}{n_{shuffle}} > \|b_2\|$

Table 2.1: This table summarizes the permutation methods to be discussed in later sections, as well as the differences which distinguish them from each other during their application.

2.0.1 Exchangeability

To maintain exchangeability, a critical assumption for permutation tests, the distribution of the permuted variable must be the same for every permutation under the null hypothesis and the original distribution of the data. When outliers exist in the data, exchangeability is violated with each shuffle of the response variable, as the outlier of the y-value associated with the x-value prevents each permutation from being distributed identically [Anderson and Robinson, 2001]. Anderson and Robinson (2001) use the lack of exchangeability in the outlier of the y-value to argue that the presence of an extreme outlier inhibits the permutation of Y_i values from providing an approximate test. The exchangeability of units under the null hypothesis such that the distribution of the permuted variable is maintained as it is randomized is among the most important aspect of the technical assumptions for an approximate permutation test [Anderson and Robinson, 2001].

2.1 Test Statistic

Without independence across the explanatory variables, it is extremely difficult to inquire and draw conclusions about a sole covariate in relation to the response using the permuted variable. All permutation methods discussed in this paper are compared via a t-statistic. The reason being, the t-statistic is a pivotal statistic which is not influenced by variables it is not designed to test.

Definition 2.1. Pivotal: A test statistic is considered pivotal if it has no reliance on unknown quantities. This means the probability distribution of the statistic only relies on parameters available [for Statistics Education,].

The t-statistic is calculated only in relation to the estimate of our parameter of interest, β_2, and the standard error of the associated statistic. In doing so, the t-statistic will remain unaffected by additional variables in the multiple regression model, such as x_{i1}, or any other linear combination of variables (apart from x_{i2}), since the calculation of the t-statistic only considers the parameter of interest [Manly, 1997].

A non-pivotal statistic such as the slope coefficient b_i is influenced by a non-zero fixed constant when the response variable is permuted. The use of a non-pivotal statistic in permutation testing leads to an inflation of the test statistic itself and a decrease of the type 1 error rate, as confirmed in simulations [Anderson and Robinson, 2001] and [Kennedy and Cade, 1996].

To diminish the inflation and inaccuracy effects of the non-pivotal statistic, this paper proposes the use of a pivotal statistic previously discussed and a change of hypotheses. Additionally, in order to compare various permutation methods, our hypotheses and statistics must align.

Chapter 3

Permutation of the Response Variable

A straightforward approach to permutation tests in multiple linear regression (MLR) entails permuting the response variable, Y_i [Manly, 1997]. Begin with a model:

$$\hat{Y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

which regresses Y_i on x_{i1} and x_{i2} together. OLS is used to generate estimates of β_1 and β_2: b_1 and b_2 [Anderson and Legendre, 1999].

A considerable challenge associated with permuting the response variable is the difficulty in drawing inferences regarding the relationship between Y_i and solely x_{i1} *or* x_{i2} when there exist dependence between the two covariates [Manly, 1997]. Permuting Y_i preserves the relationship between x_{i1} and x_{i2}, should one be present. However, it is only possible to ask questions which pertain to *both* covariates.

3.0.1 Procedure 3

The hypotheses are as follows:

$H_o : \beta_1 = \beta_2 = 0$
$H_a : \text{Either} : \beta_1 \neq 0, \beta_2 \neq 0, \text{or both do not equal zero.}$

Notice for the above hypotheses, one can only test whether there is no relationship between the covariates and Y_i, versus some relationship exists.

No information provided can conclude whether the linear relationship is between Y_i and x_{i1}, Y_i and x_{i2}, or all three variables. If the p-value on the b_2 coefficient is smaller than $\alpha = 0.05$, one concludes there is some linear relationship between the covariates and Y_i such that β_{i1} and β_{i2} aren't both zero, i.e., that the null hypothesis is rejected. Because the focus is on the β_2 parameter, the choice of statistic (b_2) will make the test more sensitive to null deviations of the form $\beta_2 \neq 0$ than null deviations of the form $\beta_1 \neq 0$. However, it is imperative to remember the permutation set-up includes both parameters to be set to zero.

After establishing hypotheses, permute the response variable such that all Y_i values are shuffled and reassigned to new x_{ij} values, and observe the change in calculations of the chosen statistic(s) while keeping variables x_{i1} and x_{i2} unpermuted [Anderson, 2001]. The model becomes:

$$Y_i^* = b_0^* + b_1^* x_{i1} + b_2^* x_{i2}$$

where Y_i^* denotes the permuted response variable. Y_i^* is then regressed on the unpermuted x_{i1} and x_{i2} again to obtain new estimates of b_1 and b_2, b_1^* and b_2^*. The original b_2 is observed in the distribution of b_2^*, and generate a p-value based on the proportion of absolute values of b_2^* found to be greater than or equal to b_2 [Anderson and Legendre, 1999]. If smaller than the predetermined critical value (typically $\alpha = 0.05$),the null hypothesis is rejected.

Furthermore, inference testing by shuffling of Y_i has been criticized when *some* of the remaining X variables have a relationship with Y_i as an approximate test rather than an exact permutation test [Kennedy and Cade, 1996]. Manly (1997) agrees with this assessment of randomizing observations, though many randomization methods are approximate in the multiple linear regression model, with some outperforming others under different technical conditions. Notably, permuting the response variable is exact when the responses (y_i values) are independent of the x_{ij} variables [Anderson and Robinson, 2001]. For a test of partial regression coefficients, if the sample size $n < 10$ and there are no outliers in the covariables, unrestricted permutation of Y_i is recommended [Anderson and Robinson, 2001].

3.0.2 Success of the Permutation of the Response Variable

Simulations have been carried out to observe different data structures, though results of these simulations point to a necessity for more extreme technical conditions, such as major prohibiting outliers or heavily skewed errors [Manly, 1997]. In Manly's (1997) simulation experiment, errors were chosen from a highly non-normal distribution (exponential to the third power), a large outlier was included in the observations, and observations for each covariate were chosen from two uniform distributions with different parameters. Permutation of Y_i values was found to be particularly strong at controlling the type 1 error rate at 5% under the null hypothesis in comparison with t- and F-distributions, as well as randomizing residuals. However, permuting the response variable underperformed in terms of power when compared to the t- and F-distributions. Out of 70,000 tests simulated, roughly 15% more tests were significant using the t- and F-distributions (58.88%) than randomizing observations (43.1%).

3.0.3 Discussion

Anderson and Legendre (1999) note in their simulation study there is no difference in power amongst permutation of residuals and permutation of Y_i values. In a different simulation experiment done by Anderson and Robinson (2001), permutation of the response variable produced a de-stabilized type 1 error rate when the covariate X contained an outlier and with either extremely normal *or* non-normal errors, more so than the randomization of residuals. Conceptually, a large outlier in the data has the ability to heavily influence the partial correlation coefficient relative to the variable being tested, depending on its severity. Under permutation of the response variable, this outlier will be associated with a different Y_i after each shuffle [Anderson and Robinson, 2001].

When the variables in the regression model are uncorrelated, inferences about the relationship between Y_i and x_{i1} *or* Y_i and x_{i2} are much more feasible. Still, in order to easily compare the significance of the t- and F-distributions against randomizing observations, the Y_i observations must be randomly assigned to the x_{ij} observations. A random assignment of Y_i values to x_{ij} values is one key justification for permutation method 3, since it suggests that Y_i is independent of x_{ij} and therefore any Y_i observation is equally likely to occur with any set of x_{ij} observations. [Manly, 1997].

Under permutation of the response variable, we propose a null hypothesis:

$$H_o : \beta_2 = 0$$

This hypothesis is problematic for the aforementioned possibility that $\beta_1 \neq 0$. However, should $\beta_1 = 0$, it is easy to test the proposed null hypothesis. The procedure will remain the same as described above, though a t-statistic will be calculated for the unpermuted dataset, as well as for each permutation of the dataset. For the unpermuted data under the model:

$$\hat{Y_i^*} = b_0^* + b_1^* x_{i1} + b_2^* x_{i2} + \epsilon_i$$

the following t-statistic will be calculated:

$$t_{ref} = \frac{b_2}{se(b_2)}$$

where t_{ref} refers to the reference value for the observed data. After permutation of the response variable, a permuted t-statistic will be calculated from the model

$$Y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

The permuted t-statistic will be denoted

$$t^* = \frac{b_2^*}{se(b_2^*)}$$

This process of permutation and t^* calculation is repeated until there exists a distribution of t^* to which the singular value for t_{ref} can be compared. Just as described previously, a p-value is calculated and if less than 5% of $|t^*|$ fall above $|t_{ref}|$ in the distribution, the null hypothesis is rejected. In other words, we fail to observe a linear relationship between Y_i and x_{i2}.

Chapter 4

Permutation of the Residuals

Both Freedman and Lane (1983) and ter Braak (1992) have proposed a permutation tests which shuffles the residuals in a multiple linear regression model. The former has been dubbed permutation of the residuals under the reduced model; the latter is known as permutation under the full model. Both are similar in their procedures, though slight differences lead to demonstrated differences in power, though theoretical solutions are still needed to provide confirmation of the superiority of one of the methods with respect to power [Anderson and Legendre, 1999].

4.1 Under the Reduced Model

Before inference testing, permutation method 1a requires two conditions be met [Anderson and Legendre, 1999]. First, the data cannot contain extreme outliers. As discussed in the previous section, the presence of extreme outliers can be highly problematic. Second, x_{i1} and x_{i2} cannot be highly collinear. A highly linear relationship between these two variables will likely influence hypothesis testing regarding the relationship between x_{i2} and Y_i.

4.1.1 Procedure 1a

When permuting under the reduced model, the null hypothesis:

$$H_o : \beta_2 = 0$$

is assumed to be true. Given there is no linear relationship between Y_i and x_{i2} under H_o, performing a permutation of the residuals tests the validity of this assumption. Note, a non-linear relationship, or no relationship at all, between Y_i and x_{i2} would reduce our multiple linear regression model from:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

to:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1}$$

Permutation test 1a begins the same as when permuting the response variable, which regresses Y_i on x_{i1} and x_{i2} together. Again, the focus is on β_2 (b_2) and the t-statistic. Recall the observed t-statistic will be referred to as t_{ref} as it refers to the reference value for the observed data. Next, Y_i is regressed **only** on x_{i1}. From this model:

$$Y_i = b_0 + b_1 x_{i1} + R_{y_i|x_{i1}}$$

the estimates of β_0 (b_0), β_1 (b_1), and the residuals ($R_{y_i|x_{i1}}$) are generated. The residuals from this regression ($R_{y_i|x_{i1}}$) are permuted to produce $R^*_{y_i|x_{i1}}$. The permuted residuals replace the unpermuted residuals in the simple linear regression equation, creating new response values as they do so:

$$Y_i^* = b_0 + b_1 x_{i1} + R^*_{y_i|x_{i1}}$$

where Y_i^* represents the response values corresponding to the fitted values plus permuted residuals. This process forces the null hypothesis to be true, since the relationship between Y_i and x_{i2} is broken as the new model omits the presence of the second variable. The new observations Y_i^* are then regressed on covariates x_{i1} and x_{i2} together, similar to the first step. With the new regression model:

$$E(Y_i^*) = \beta_0^* + \beta_1^* x_{i1} + \beta_2^* x_{i2}$$

estimates of the fitted value β_2^* are obtained from the permuted data (b_2^*), as well as the t-statistic :

$$t^* = \frac{b_2^*}{se(b_2^*)}$$

with the consistent $n-3$ degrees of freedom. Finally, beginning with the permutation of the residuals producing $R^*_{y_i|x_{i1}}$, the procedure is repeated multiple times for some $m < n!$, where $n!$ is all possible permutations of the residuals. Note since $m < n!$, the permutation test is not an exact test, but an approximate one. Once a substantial number (m) of t^* has been calculated to create a distribution, t_{ref} is compared to these values [Anderson and Legendre, 1999]. As in general hypothesis testing, if fewer than 5% of t^* values are greater than t_{ref}, the null hypothesis is rejected and it's concluded there is a relationship between Y and x_{i2} (i.e., $\beta_2 = 0$ is forced during permutation under the reduced model and this claim was refuted).

4.1.2 Discussion

It is important to note permutation method 1b, under the reduced model, preserves the relationships between Y_i and x_{i1}, x_{i1} and x_{i2}, but **not** Y_i and x_{i2}. The last is not preserved since $\beta_2 = 0$ was forced when Y_i^* was fitted to the simple linear regression concerning only x_{i1}. Since x_{i1} and x_{i2} were not permuted, their covariance is preserved [Anderson, 2001]. Though, as mentioned in the beginning of Chapter 4, both extreme outliers and high collinearity amongst x_{i1} and x_{i2} is problematic. In the preservation of the relationship between x_{i1} and x_{i2}, there is caution that a high collinearity will influence the results of the permutation test. The method of permutation 1b is optimal since errors may be exchangeable within each variable ($R_{y_i|x_{i1}}$ and $R_{y_i|x_{i2}}$), though not between the two variables (i.e., exchangeability is not preserved under permutation of $R_{y_i|x_{i1}x_{i2}}$ [Kennedy, 1995]. For exchangeability in this application, it was assumed that the errors, rather than the response variable as in permutation of the response variable, are exchangeable under the null hypothesis. Furthermore, possibly the greatest assumption necessary for permutation of the residuals is that the distribution of the residuals in the regression model *approximate* that of the errors in the true population.

4.2 Under the Full Model

Permutation method 1b *also* implements permutation of the residuals in a linear regression model, although this method does not regress under a reduced model. By permuting the residuals from the full model, method 1b uses the original estimate of β_2, b_2, as part of the permutation test. In doing so, the variance of b_2 is reduced as it is repeatedly estimated in the procedure, increasing the overall power of the test [Anderson and Legendre, 1999].

4.2.1 Procedure 1b

Following the first steps of permutation method 3 and 1a, method 1b proposes a regression of Y_i on both x_{i1} and x_{i2}:

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + R_{y_i|x_{i1}x_{i2}}$$

where $R_{y_i|x_{i1}x_{i2}}$ refers to the residuals from the model on both x_{i1} and x_{i2}. From this regression model, the population parameters $\beta_0, \beta_1, \beta_2, \text{and} \epsilon_i$ are estimated by $b_0, b_1, b_2, \text{and} R_{y_i|x_{i1}x_{i2}}$. Additionally, a t-statistic for the observed data is calculated:

$$t_{ref} = \frac{b_2}{se(b_2)}$$

Next, method 1b permutes $R_{y_i|x_{i1}x_{i2}}$ such that new permuted residuals are achieved and denoted $R^*_{y_i|x_{i1}x_{i2}}$. Adding these back to the fitted model with the population parameter estimates, new response values are achieved:

$$Y_i^* = b_0 + b_1 x_{i1} + b_2 x_{i2} + R^*_{y_i|x_{i1}x_{i2}}$$

Then, Y_i^* is regressed on **both** covariates x_{i1} and x_{i2} (i.e., the full model). After this regression, the new multiple linear regression model is:

$$Y_i^* = b_0^* + b_1^* x_{i1} + b_2^* x_{i2} + R^*_{y_i|x_{i1}x_{i2}}$$

where all coefficients now refer to permuted data. From this, a t-statistic is calculated such that:

$$t^* = \frac{(b_2^* - b_2)}{se(b_2^*)}$$

Notice here that b_2^* is an estimate of β_2^* as b_2 is an estimate of β_2. Thus, t^* inquires about the deviation of a permuted sample estimate (b_2^*) from an unpermuted sample estimate (b_2). By quantifying this deviation, insight is gained about how the sample estimate (b_2) deviates from the true population parameter (β_2). The hypothesis is testing whether the parameter significantly differs from the test statistic, by testing whether the test statistic significantly differs from the permuted test statistic by comparing the test statistic to a distribution of the permuted test statistic. In similar terms, the randomization of the residuals will be approximately equal to the randomization of the true errors in the regression model [Manly, 1997]. Given the null hypothesis is true ($\beta_2 = 0$), the distribution of errors in the full model will approximate the distribution of errors under the null hypothesis [Anderson, 2001]. This being said, if the null hypothesis is true, the full model:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

simplifies to:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1}$$

Following the calculation of t^*, the process is repeated $m < n!$ times, beginning with the permutation of $R_{y_i|x_{i1}x_{i2}}$. Ultimately, the t_{ref} is compared with the distribution of t^* to obtain a p-value. Whether this p-value less than the chosen significance value (α) determines whether the null hypothesis is rejected or not. If it is, the null hypothesis is rejected and consequently, it is concluded that there is a linear relationship between Y_i and x_{i2}.

4.2.2 Discussion

In permuting the residuals under the full model, the covariances among x_{i1} and x_{i2} are preserved. Additionally, if there exist a model defined with more variables which are dependent on x_{i1} or x_{i2}, these relationships are preserved as well [Anderson and Legendre, 1999]. Though ter Braak suggested permutation of the residuals under the full model should have greater power than the same under the reduced model, this was not demonstrated with any significance in simulations [Anderson and Legendre, 1999]. Yet, simulations show both permutation methods of the residuals have asymptotically correct significance levels [Anderson and Legendre, 1999]. Of all three permutation methods described thus far, all produced asymptotically equivalent results and were satisfactory tests for partial regression coefficients. Notably, they all returned greater power and type I error rates approximating α at closer values than the traditional t-test for non-normally distributed data [Anderson and Legendre, 1999]. Regarding both methods of permuting residuals, the procedure under the full model deviated from the reduced model in the presence of an extreme x_{i1} outlier, non-normally distributed errors, and a small sample size. Under these conditions, permutation method 1b provided an inflated type I error rate in simulations. This deviation can be mitigated with an increase in sample size n, though otherwise, permutation method 1b is preferred in under the three aforementioned conditions [Anderson and Legendre, 1999].

An added benefit to method 1b is that the permutation procedure need not be adapted to test for other partial regression coefficients [Anderson, 2001]. To instead inquire about the relationship between Y_i and x_{i1}, the hypothesis can be altered to reflect:

$$H_o : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Consequently, the t-statistic must also be changed such that:

$$t_{ref} = \frac{b_1}{se(b_1)}$$

and:

$$t^* = \frac{b_1^*}{se(b_1^*)}$$

Aside from these minor adjustments, the computationally intensive permutation procedure does not need to be adapted or re-run. Since regardless of the hypotheses, method 1b permutes all residuals under the full model, the only change is the question being asked of the permuted data and the summary statistics which are calculated.

Chapter 5

The Naïve Permutation

The "naïve" permutation is described as such because it is focused on our variable of interest (x_{i2}) rather than the relationships among the variables within the entire model. This permutation method presented by Oja (1987) suggests a permutation of the single covariate of interest, x_{i2}. However, should x_{i1} and x_{i2} be correlated, permutation of x_{i2} removes their correlation structure. Furthermore, inference into the dependency of Y_i and x_{i2} is complicated by the influence of a possible covariate, x_{i1}.

Definition 5.1. Ancillarity The ancillarity principle requires our statistic be independent of the parameters in the assumed model [Ghosh et al., 2010].

The ancillarity principle is violated since the collinearity between x_{i1} and x_{i2} is not conserved [Kennedy, 1995]. A pivotal quanitity is not always a statistic, but when it is, it is called an ancillary statistic. Of the methods described previously in this work, two permute the residuals in the model. Permutation of the residuals under the full model preserves the relationship between Y_i and x_{i1}, while breaking the relationships between Y_i and x_{i2}, as well as between x_{i1} and x_{i2}. The third method shuffled the response variable in the model. Randomization of Y_i preserves the correlation amongst x_{i1} and x_{i2}, while breaking the relationship between Y_i and x_{i1}, as well as Y_i and x_{i2}. The naïve permutation of this section permutes the variable of interest, x_{i2}, in the model. This shuffling maintains the relationship between Y_i and x_{i1}, while breaking the correlation between x_{i1} and x_{i2}, and the relationship between Y_i and x_{i2}.

However, in certain scenarios, permuting one variable is not naïve in the slightest. In randomized experiments, treatment values (x_{i2}) are assigned randomly to subjects [Oja, 1987], while other, for example, demographic characteristics (x_{i1}), are not randomly assigned. Under this condition, x_{i1} and x_{i2} are completely independent. There is no relationship between x_{i1} and x_{i2} to break during permutation. Since subjects are randomly assigned to a treatment level, there is no relationship between the covariates to be preserved. Because randomization of the explanatory variables is inherent to the experimental design, it is possible to permute *only* the variable of interest without violating the exchangeability and ancillarity conditions of the permutation test [Manly, 1997] during shuffling and calculation of the test statistic.

5.0.1 Procedure 2

In permutation method 2, and identical to the others, Y_i is regressed on x_{i1} and x_{i2}:

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2}$$

where b_0, b_1, and b_2 are unknown sample estimates of the population, x_{i2} is a design variable, and x_{i1} is a covariate. From this description, b_2 is estimating β_2, our parameter of interest, while b_0 and b_1 are estimates of "nuisance parameters" β_0 and β_1 [Oja, 1987].

This paper permutes x_{i2} and estimates β_2 using the familiar aforementioned test statistic. Unfortunately, Oja is criticized since "the formulas and calculations are cumbersome" [Collins, 1987]. For the multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

where x_{i2} is assigned randomly to subjects, the statistic proposed in Oja (1987) to test $H_o : \beta_2 = 0$ is:

$$T(x_{i2}^*) = \sum_{i<j<k} \triangle_{ijk}^{y} \triangle_{ijk}^{x_{i2}^*}$$

where:

$$\triangle_{ijk}^{y} = \begin{vmatrix} 1 & 1 & 1 \\ y_i & y_j & y_k \\ x_{i1} & x_{j1} & x_{k1} \end{vmatrix}$$

and the definition of x_{i2} follows similarly,

$$\triangle_{ijk}^{x} = \begin{vmatrix} 1 & 1 & 1 \\ x_{i1} & x_{j1} & x_{k1} \\ x_{i2} & x_{j2} & x_{k2} \end{vmatrix}$$

with x_{i2}^{*} being a random permutation of the variable x_{i2}, though the calculation is notably "awkward" [Collins, 1987]. The estimator of β_{ijk} which will be permuted, b_{ijk}, is defined as:

$$\triangle_{ijk}^{b} = \frac{\triangle_{ijk}^{y}}{\triangle_{ijk}^{x}}$$

Collins (1987) promotes the arguably easier and equivalent test statistic:

$$\begin{aligned} &\sum_{i}\sum_{j}\sum_{k} \triangle_{ijk}^{y} \triangle_{ijk}^{x_{i2}^{*}} \\ = 6(\textstyle\sum y_i x_{i2}^{*})\binom{2}{i1} - (\textstyle\sum y_i x_{i1})(\textstyle\sum x_{i1} x_{i2}^{*}) \\ &\propto y'(I - P_1)x_{i2}^{*} \end{aligned}$$

in which $P_1 = X(X'X)^{-1}$ and $X = [1, x_{i1}]$.Though not equivalent, this paper will only consider the use of the our test statistic in order to be consistent with the analysis done with the other randomization schemes. The use of:

$$t = \frac{b_2}{se(b_2)}$$

allows for equivalent comparisons across all permutation methods described, as the p-value is calculated in relation to this test statistic and the observed t-statistic prior to shuffling, just as with all other randomization methods considered in this research.

Chapter 6

Simulation Study

The purpose of the simulation study is to provide omniscient evidence. To determine whether conclusions drawn from a permutation test performed on a sample are representative of the population, it's necessary to simulate from a known population. By drawing a sample from a known population, it's possible to observe the results of the permutation test on the sample data set and verify whether it accurately describes the behavior of the population. Plainly, if a permutation test is performed on a multiple linear regression model of the observations and deduces there exist a relationship between Y_i and x_{i2}, this can be confirmed or denied by the specifications of the known population (from which the data were sampled).

6.1 Mis-Specification

Often times there exist multiple explanatory variables in a population, though information is colleted on only one or a few of the variables. Mis-specification implies fitting a model to the observed data incorrectly. In practice, mis-specification is difficult to recognize. However, the simulated data can be drawn from a population with k explanatory variables and modeled using only $j < k$ explanatory variables, where $j, k, \in \mathbb{N}^+$. For example, one may have a population model from which a relationship is generated as:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_j x_{ij} + ... + \beta_k x_{ik} + \epsilon_i$$

though in collecting data and thereby mis-specifying the model, a simpler multiple regression model is instead fit to the sample:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + +\beta_j x_{ij}$$

The purpose of simulating a mis-specification is to observe how well a permutation test performs when there exists influence of additional variables which are not present in the OLS regression. Since the model is knowingly incorrect, results report how insightful permutation tests are to the true behavior of the population despite this mis-specification.

6.2 Variants

This simulation study introduces a multitude of variations, as it attempts to be as comprehensive as possible. Four populations were created and sampled from. All variables were generated from a normal distribution with $\mu = 0$ and $\sigma^2 = 1$, though an approachable expansion of this study would include non-normal distributions and errors. An assessment of non-normality would contextualize the simulations done in Anderson and Legendre (1999).

Data Type:	Control	1	2	3
x_{i1}	$\sim N(\mu, \sigma^2)$	$\sim N(\mu, \sigma^2)$	$\sim N(\mu, \sigma^2)$	$\sim N(\mu, \sigma^2)$
x_{i2}	$\sim N(\mu, \sigma^2)$	$\sim N(\mu, \sigma^2) + x_1$	$\sim N(\mu, \sigma^2)$	$\sim N(\mu, \sigma^2)$
x_{i3}	N/A	N/A	N/A	$\sim N(\mu, \sigma^2)$
x_{i4}	N/A	N/A	N/A	$\sim N(\mu, \sigma^2)$
Population	$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$	$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$	$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$	$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$
Linear Model Fit	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$	$y \sim x_{i1} + x_{i2}$
Summary	x_{i1} and x_{i2} are independent	x_{i1} and x_{i2} are correlated	x_{i1} and x_{i2} are interacting	x_{i3} and x_{i4} are important, but missing

To each of the datasets simulated from the four populations, the **same** linear model is applied:

$$\hat{y_i} = b_0 + b_1 x_{i1} + b_2 x_{i2}.$$

Note that the "Control" population specifies that the response variable is a function only of x_{i1} and x_{i2}, which are independent. The "Data Type 1" population specifies y as a function of *only* x_{i1} and x_{i2}, which are correlated. Specifically, x_{i2} is generated from x_{i1}, with some additional, normally distributed noise. The "Data Type 2" population specifies y as a function of x_{i1} and x_{i2} where the covariates interact. The interaction is not specified in the model. Finally, the "Data Type 3" population specifies y as a function of x_{i1} and x_{i2}, as well as x_{i3} and x_{i4}. Note once again, the model is misspecified since the additional variables x_{i3} and x_{i4} are not considered in the linear regression. The purpose of the different data structures is to vary the population from which we sample, and observe how unknown specificity (covariate correlation, interaction, or additional hidden explanatory variables) influences the performance of the permutation test as compared to OLS in multiple linear regression. The number of observations is given by the variable "n_{obs}."

All discussed permutation methods (permutation of the residuals, permutation of the response variable, and naïve permutation) were created as functions, with the data type (population) being their input. The permutation functions run their respective methods on each data type, and summarize the observed value of b_2 in the sample, the observed value of the t-statistic in the sample ("t_{ref}"), the p-value as calculated by OLS ("OLS_{pval}"), and the p-value as calculated by the permutation method ("$perm_{pval}$"). Since the sample is permuted multiple times ("$n_{shuffle}$") to generate a distribution of permuted t-statistics against which to compare t_{ref},

$$perm_{pval} = \frac{\sum[|t^*| > |t_{ref}|]}{n_{shuffle}}$$

where "t^*" represents the calculated t-statistic after each permutation, and of which the distribution of permuted t-statistics is comprised. Simply put, the p-value of the permutation test answers the question: How often was a larger t-statistic seen than the originally observed value after $n_{shuffle}$ permutations of the sample?

For additional variation, population statistics β_1, β_2, β_3, β_4 and n_{obs} were sequenced to provide multiple combinations of coefficient values and sample sizes. β_0 was not sequenced because variation of the intercept value failed to influence permutation results across populations and would have substantially increased computation time had it been included.

Sample Values	Iterations
β_1	0.0, 0.2, 0.4, 0.6, 0.8, 1.0
β_2	0.0, 0.2, 0.4, 0.6, 0.8, 1.0
β_3	0.0, 0.2, 0.4, 0.6, 0.8, 1.0
β_4	0.0, 0.2, 0.4, 0.6, 0.8, 1.0
n_{obs}	15, 30, 45, 60

Table 6.1: The values which set our population are listed above. Every possible combination of these coefficient iterations are used to create many populations, from which samples were taken and permutation tests were performed.

The number of permutations for each sample was fixed at $n_{shuffle} = 100$. For each possible combination of sample values, the data set was permuted 100 times. From this, the summary statistics are compared: OLS versus permutation. Each permutation method is performed on every combination of sample values, for every population generated. For example, Data Type 1 will undergo all four permutation methods separately. During each of these four individual simulations, the permutation functions will iterate through all combinations of sample values, to produce a comprehensive list of summary statistics. This process is repeated for all four data types (populations). In total, sixteen simulations are completed for each combination of parameters in 6.1, with each providing a robust comparison of the permutation method and OLS's performance under different structures of data. The output is how often the null hypothesis is rejected, which will either be a type I error rate or power, depending on the how H_o is specified: whether or not $\beta_2 = 0$

6.3 Results

There are 16 different combinations of "Data Type" and "Permutation Method," all of which will be summarized. However, for each table of graphs which presents these comprehensive results between data and method, the coefficients of the population from which data was generated must also be specified. This paper has opted to display the incrementally increasing values of β_2 along the x-axis of each individual graph, since inquiry focuses on the relationship between x_{i2} and y_i. Thus, it is left to specify β_1, β_3, and β_4 as variable or constant at zero. These remaining coefficients are either specified as non-zero (i.e., $\beta_1 \neq 0$= "$\beta_1! = 0$" in graph table titles), specified as zero (i.e., "$\beta_1 = 0$"), or displayed incrementally by color-coding. When the coefficients are given by color, a line is displayed for each increment of the variable (i.e., pink is $\beta_1 = 0.2$, blue is $\beta_1 = 0.4$), while the line-type (solid or dashed) refers to whether the line is describing the OLS rejection rate or the permutation method rejection rate. In total for these tables of graphs, (see Figures 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12), each individual graph holds twelve lines describing the behavior of the data type and permutation method combination as the coefficients of the population vary. There are twelve lines total to account for the variations of one parameter via six different colored lines, which are color-coded for both the OLS and Permutation Rejection Rates (i.e., there are six solid colored lines and six dashed colored lines to describe how the rejection rates change across different coefficient values). The exception of these twelve lines are Figures 6.1, 6.2, 6.3, and 6.4. For Figures 6.1, 6.2, 6.3, and 6.4, permutation methods and data structures are combined with parameters $\beta_1 = 0$, $\beta_3 = 0$, and $\beta_4 = 0$ for 6.1 and 6.2, and $\beta_1 \neq 0$, $\beta_3 = 0$, and $\beta_4 = 0$ for 6.3, and 6.4. Therefore, the only *specified and incremental* variable parameter is β_2, which is specified along the x-axis. Without variance in the other parameters, or *specified* variance in the case of $\beta_1 \neq 0$ for 6.3, and 6.4, color-coding is not necessary to distinguish between different values. Thus, there are only two lines necessary in these graphs. Along the y-axis, in all of the figures in the rest of this chapter, the rejection rate is plotted, so the line-type simply differentiates between the two comparisons of OLS and permutation p-value calculations. For simplicity, only sample sizes of $n = 15$ and $n = 60$ will be displayed, though simulations were run for all sample size increments listed in the previous section.

6.4 Zero Constant b_1, b_3 and b_4

Below Figures 6.1 and 6.2 display permutation methods across the top, and data types along the side. The rejection rate for each graph is plotted on the y-axis, while the increments of β_2 are plotted along the x-axis. The two line-types in each individual graph refer to whether the line is a representation of the OLS p-value or the permutation p-value. While Figure 6.1 depicts results for a sample size of $n = 15$ from the population, Figure 6.2 depicts results for a sample size of $n = 60$. A horizontal line has been added for $y = 0.05$ to differentiate between size and power of the two approaches. Recall under the null hypothesis (when $\beta_2 = 0$), the best performing test will return a level of significance of 0.05. This describes the ability of the approach to provide an exact test with a fixed size of 0.05, given the null hypothesis is accepted. When the contrary is true and $\beta_2 \neq 0$, it becomes a question of which approach is rejecting the null hypothesis at a higher power (i.e., more often). It is possible to not only visualize how permutation tests compare to OLS via size and power, but additionally how various permutation methods compare to OLS via size and power under different populations of data. Furthermore, one can ask: If given a larger sample size from the data (i.e., $n = 60$ rather than $n = 15$), do these observations comparing OLS and permutation tests change?

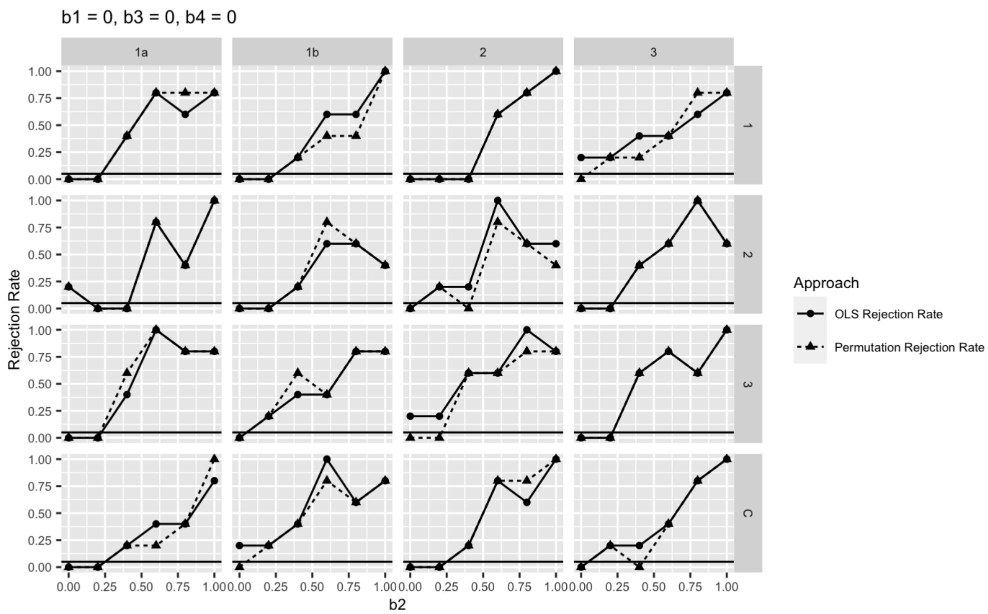

Figure 6.1: For sample size $n = 15$.
Power: With β_2 as the only non-zero variable, both approaches are quite variable across methods of permutation and data structures. However, both typically deviate the same, which insinuates this variation in rejection rate is a consequence of small sample size and zero coefficients for β_1, which describes a variable in the fitted model (x_{i1}).
Size: Both approaches appear to be underestimating *and* overestimating the size for $\beta_2 = 0$, though OLS appears to make an overestimation in error more frequently than any permutation test. Their variability again, points to a necessity for a larger sample size.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0	0	0.2	0.2	0	0	0	0
0	1b	0	0	0	0	0	0	0.2	0
0	2	0	0	0	0	0.2	0	0	0
0	3	0.2	0	0	0	0	0	0	0
0.2	1a	0	0	0	0	0	0	0	0
0.2	1b	0	0	0	0	0.2	0.2	0.2	0.2
0.2	2	0	0	0.2	0.2	0.2	0	0	0
0.2	3	0.2	0.2	0	0	0	0	0.2	0.2
0.4	1a	0.4	0.4	0	0	0.4	0.6	0.2	0.2
0.4	1b	0.2	0.2	0.2	0.2	0.4	0.6	0.4	0.4
0.4	2	0	0	0.2	0	0.6	0.6	0.2	0.2
0.4	3	0.4	0.2	0.4	0.4	0.6	0.6	0.2	0
0.6	1a	0.8	0.8	0.8	0.8	1	1	0.4	0.2
0.6	1b	0.6	0.4	0.6	0.8	0.4	0.4	1	0.8
0.6	2	0.6	0.6	1	0.8	0.6	0.6	0.8	0.8
0.6	3	0.4	0.4	0.6	0.6	0.8	0.8	0.4	0.4
0.8	1a	0.6	0.8	0.4	0.4	0.8	0.8	0.4	0.4
0.8	1b	0.6	0.4	0.6	0.6	0.8	0.8	0.6	0.6
0.8	2	0.8	0.8	0.6	0.6	1	0.8	0.6	0.8
0.8	3	0.6	0.8	1	1	0.6	0.6	0.8	0.8
1	1a	0.8	0.8	1	1	0.8	0.8	0.8	1
1	1b	1	1	0.4	0.4	0.8	0.8	0.8	0.8
1	2	1	1	0.6	0.4	0.8	0.8	1	1
1	3	0.8	0.8	0.6	0.6	1	1	1	1

Table 6.2: Here, the rejection rates for OLS and permutation methods are compared side-by-side via data type, with the coefficents of β_2 and the method of permutation listed in the first two columns. These are the values plotted in Figure 6.1. Here, the caption of Figure 6.1 is observed with exact rejection rate values.

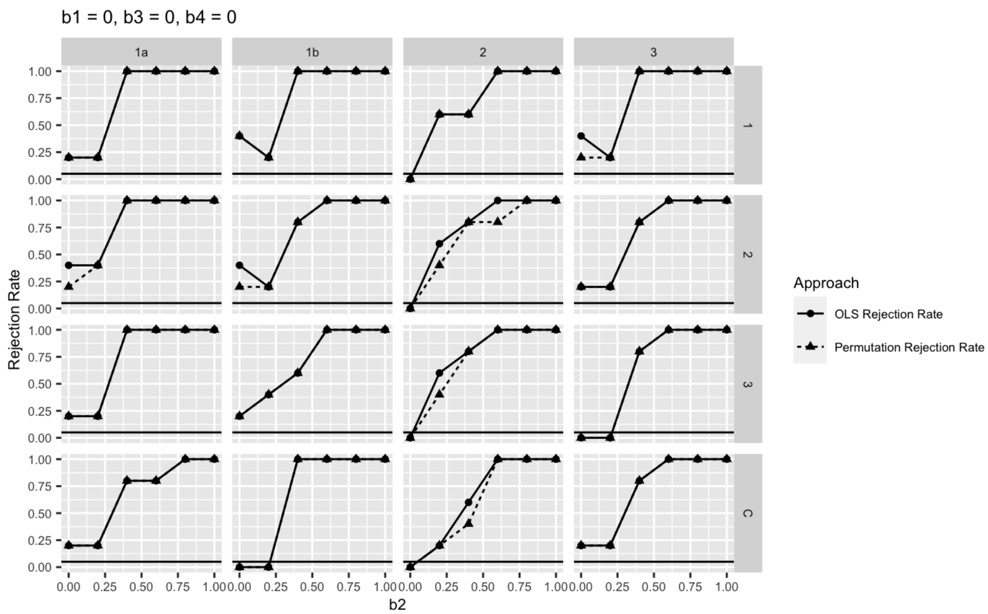

Figure 6.2: For sample size $n = 60$.

Power: There is significantly less variation observed in both OLS and permutation rejection rates than in Figure 6.1. Also, with many data types and permutation methods achieving 100% power at $\beta_2 = 0.6$, whereas few did, if at all, in Figure 6.1, an increase in power is present much more rapidly than with $n = 15$.

Size: The majority of graphs appear to have a level of significance greater than 0.05, suggested an inflated type I error rate There are considerably more errors in controlling size when $n = 60$. Mainly, overestimation errors occur in both approaches. Other research saw variation in type I error rates decrease as sample size increased across permutation methods with non-zero β_2 and correlation between x_{i1} and x_{i2} at zero, as is the case in this particular simulation [Anderson and Legendre, 1999]. While permutation method 2, the naïve permutation, does not display an inflated type I error rate and this is notable, the concern is primarily the lack of consensus among OLS rejection rates across permutation methods. OLS is performed on each data type across permutation methods, so the variation seen in the solid line of rejection rates horizontally is chance, rather than an indicator of notable variation. Still, the variation of the OLS rejection rate gives valuable insight into the consistency of this method and its performance. With a sample size $n = 60$ and no collinearity among variables, the OLS rejection rate appears to perform on par or worse than the rejection rate of any of the four permutation methods.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.2
0	1b	0.4	0.4	0.4	0.2	0.2	0.2	0	0
0	2	0	0	0	0	0	0	0	0
0	3	0.4	0.2	0.2	0.2	0	0	0.2	0.2
0.2	1a	0.2	0.2	0.4	0.4	0.2	0.2	0.2	0.2
0.2	1b	0.2	0.2	0.2	0.2	0.4	0.4	0	0
0.2	2	0.6	0.6	0.6	0.4	0.6	0.4	0.2	0.2
0.2	3	0.2	0.2	0.2	0.2	0	0	0.2	0.2
0.4	1a	1	1	1	1	1	1	0.8	0.8
0.4	1b	1	1	0.8	0.8	0.6	0.6	1	1
0.4	2	0.6	0.6	0.8	0.8	0.8	0.8	0.6	0.4
0.4	3	1	1	0.8	0.8	0.8	0.8	0.8	0.8
0.6	1a	1	1	1	1	1	1	0.8	0.8
0.6	1b	1	1	1	1	1	1	1	1
0.6	2	1	1	1	0.8	1	1	1	1
0.6	3	1	1	1	1	1	1	1	1
0.8	1a	1	1	1	1	1	1	1	1
0.8	1b	1	1	1	1	1	1	1	1
0.8	2	1	1	1	1	1	1	1	1
0.8	3	1	1	1	1	1	1	1	1
1	1a	1	1	1	1	1	1	1	1
1	1b	1	1	1	1	1	1	1	1
1	2	1	1	1	1	1	1	1	1
1	3	1	1	1	1	1	1	1	1

Table 6.3: As mentioned previously in Figure 6.2, there is a large increase in power across permutation methods, as evidenced by the "1's" displayed in Table 6.3, which refer to a 100% rejection rate. While this was hardly achieved for sample size $n = 15$, as seen in Table 6.2, here both the OLS and permutation methods are rejecting the null hypothesis with 100% confidence for a range of intercepts, beginning with $\beta_2 = 0.4$ and continuing until $\beta_2 = 1$. However, permutation and OLS still fail to minimize the type I error rate, as evidenced by the values in the first four rows of the table (where $\beta_2 = 0$, thus the null hypothesis is true), of which none are the intended $\alpha = 0.05$. Instead, most of these values are inflated at values of 0.2 and 0.4, and consequently say both methods are incorrectly rejecting the null hypothesis at high rates (20-40% of the time).

6.5 Variable β_1, Zero Constant β_3 and β_4

In the previous subsection, 6.4, while β_2 was varied (i.e., $\beta_2 = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$), β_1, β_3, and β_4 were held constant at zero. Here, both β_2 **and** β_1 are varied, while β_3 and β_4 still remain held constant at zero as before. Then, the results were observed across permutation methods and data types as done in subsection 6.4. The values of β_1 are not displayed in the plot as β_2 is along the x-axis. Instead, β_1 simply does not equal zero, though the exact value of

β_1 (either $0.0, 0.2, 0.4, 0.6, 0.8$ or 1.0) is unknown.

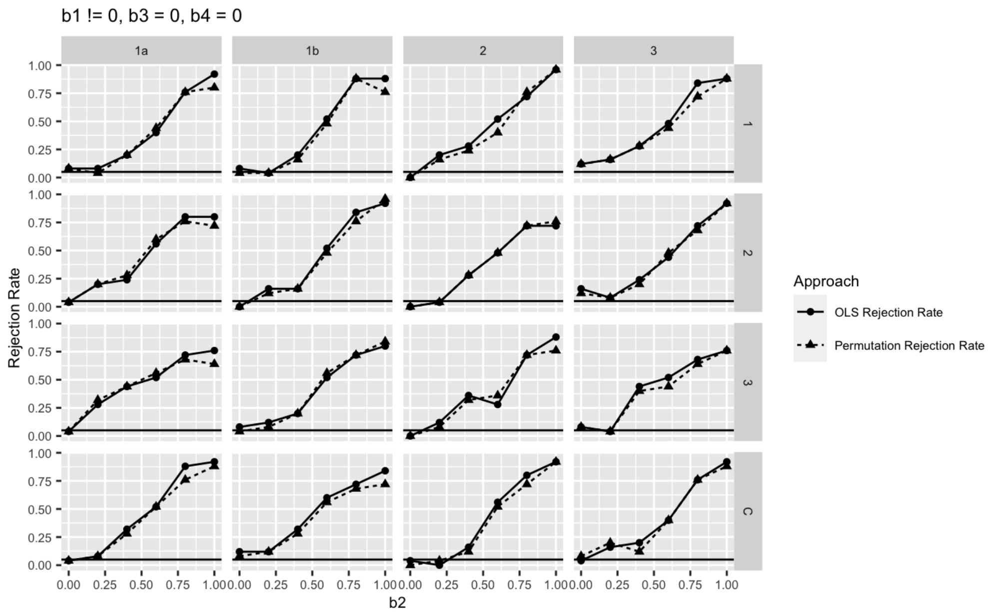

Figure 6.3: For sample size $n = 15$.

Power: it appears most methods of permutation tests slightly under-perform in comparison to OLS, across different data structures. However, permutation method 1a (that promoted by Freedman and Lane) seems to barely outperform OLS for data types 1, 2, and 3, until β_2 has a larger coefficient (i.e., when $\beta_2 = 1$). Both approaches (OLS and permutation method 1a, 1b, 2, and 3) correctly fail to to reject the null hypothesis when $b_2 = 0$ for all data types.

Size: all tests appear to report very similar results, though Table 6.4 shows the slight differences.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.08	0.08	0.04	0.04	0.04	0.04	0.04	0.04
0	1b	0.08	0.04	0	0	0.08	0.04	0.12	0.08
0	2	0	0	0	0	0	0	0.04	0
0	3	0.12	0.12	0.16	0.12	0.08	0.08	0.04	0.08
0.2	1a	0.08	0.04	0.2	0.2	0.28	0.32	0.08	0.08
0.2	1b	0.04	0.04	0.16	0.12	0.12	0.08	0.12	0.12
0.2	2	0.2	0.16	0.04	0.04	0.12	0.08	0	0.04
0.2	3	0.16	0.16	0.08	0.08	0.04	0.04	0.16	0.2
0.4	1a	0.2	0.2	0.24	0.28	0.44	0.44	0.32	0.28
0.4	1b	0.2	0.16	0.16	0.16	0.2	0.2	0.32	0.28
0.4	2	0.28	0.24	0.28	0.28	0.36	0.32	0.16	0.12
0.4	3	0.28	0.28	0.24	0.2	0.44	0.4	0.2	0.12
0.6	1a	0.4	0.44	0.56	0.6	0.52	0.56	0.52	0.52
0.6	1b	0.52	0.48	0.52	0.48	0.52	0.56	0.6	0.56
0.6	2	0.52	0.4	0.48	0.48	0.28	0.36	0.56	0.52
0.6	3	0.48	0.44	0.44	0.48	0.52	0.44	0.4	0.4
0.8	1a	0.76	0.76	0.8	0.76	0.72	0.68	0.88	0.76
0.8	1b	0.88	0.88	0.84	0.76	0.72	0.72	0.72	0.68
0.8	2	0.72	0.76	0.72	0.72	0.72	0.72	0.8	0.72
0.8	3	0.84	0.72	0.72	0.68	0.68	0.64	0.76	0.76
1	1a	0.92	0.8	0.8	0.72	0.76	0.64	0.92	0.88
1	1b	0.88	0.76	0.92	0.96	0.8	0.84	0.84	0.72
1	2	0.96	0.96	0.72	0.76	0.88	0.76	0.92	0.92
1	3	0.88	0.88	0.92	0.92	0.76	0.76	0.92	0.88

Table 6.4: Based on the graphs in Figure 6.3, it appeared permutation method 1a was the only method to correctly fail to reject the null hypothesis with more confidence than OLS (i.e., have a higher power/rejection rate for $\beta_2 \neq 0$). However, a closer examination of Table 6.4 shows permutation method 1a only has a consistently equal or higher rejection rate than OLS when $\beta_2 = 0.6$ for all data types. As the analysis of simulation results continues, it will become more obvious that $\beta_2 = 0.6$ appears to be a threshold to confidently reject the null hypothesis for both OLS and permutation methods.

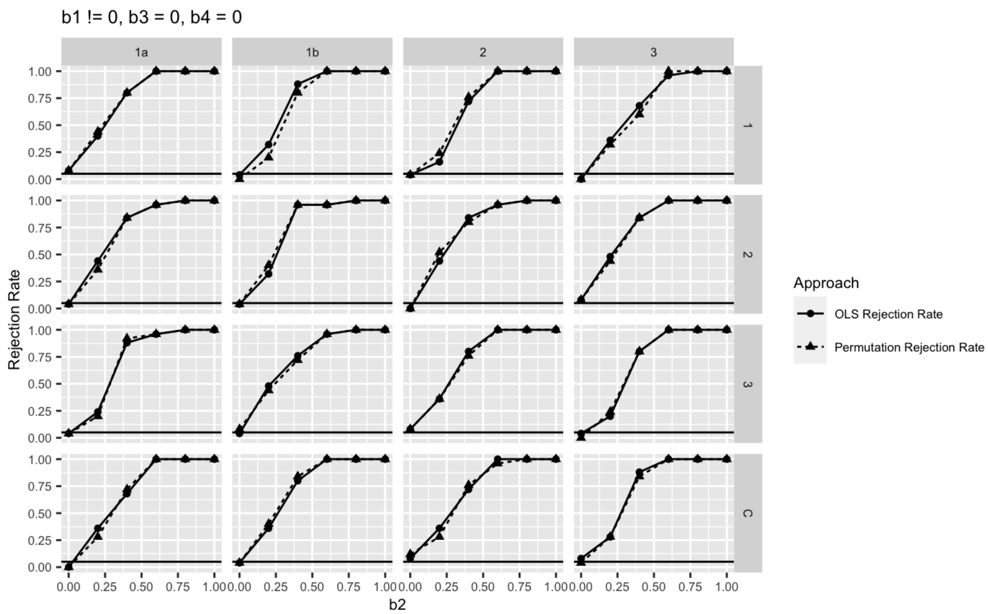

Figure 6.4: For sample size $n = 60$.
Power: Much of the same trends are seen as in Figure 6.3 for sample size $n = 15$, though both approaches reach a higher power more quickly (i.e., for lower coefficients of β_2) than previously. Comparison of the two approaches are much closer in this subsection 6.5 than in subsection 6.4 regardless of β_2.
Size: Again, all tests appear to report very similar results, though numerical values of the size given in Table 6.5 later will showcase the slight differences.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.08	0.08	0.04	0.04	0.04	0.04	0	0
0	1b	0.08	0.04	0	0	0.08	0.04	0.04	0.04
0	2	0	0	0	0	0	0	0.08	0.12
0	3	0.12	0.12	0.16	0.12	0.08	0.08	0.08	0.04
0.2	1a	0.08	0.04	0.2	0.2	0.28	0.32	0.36	0.28
0.2	1b	0.04	0.04	0.16	0.12	0.12	0.08	0.36	0.4
0.2	2	0.2	0.16	0.04	0.04	0.12	0.08	0.36	0.28
0.2	3	0.16	0.16	0.08	0.08	0.04	0.04	0.28	0.28
0.4	1a	0.2	0.2	0.24	0.28	0.44	0.44	0.68	0.72
0.4	1b	0.2	0.16	0.16	0.16	0.2	0.2	0.8	0.84
0.4	2	0.28	0.24	0.28	0.28	0.36	0.32	0.72	0.76
0.4	3	0.28	0.28	0.24	0.2	0.44	0.4	0.88	0.84
0.6	1a	0.4	0.44	0.56	0.6	0.52	0.56	1	1
0.6	1b	0.52	0.48	0.52	0.48	0.52	0.56	1	1
0.6	2	0.52	0.4	0.48	0.48	0.28	0.36	1	0.96
0.6	3	0.48	0.44	0.44	0.48	0.52	0.44	1	1
0.8	1a	0.76	0.76	0.8	0.76	0.72	0.68	1	1
0.8	1b	0.88	0.88	0.84	0.76	0.72	0.72	1	1
0.8	2	0.72	0.76	0.72	0.72	0.72	0.72	1	1
0.8	3	0.84	0.72	0.72	0.68	0.68	0.64	1	1
1	1a	0.92	0.8	0.8	0.72	0.76	0.64	1	1
1	1b	0.88	0.76	0.92	0.96	0.8	0.84	1	1
1	2	0.96	0.96	0.72	0.76	0.88	0.76	1	1
1	3	0.88	0.88	0.92	0.92	0.76	0.76	1	1

Table 6.5: With the increase in sample size to $n = 60$, many of the type 1 error rates displayed in the first four rows of the table for $\beta_2 = 0$ are expected to oscillate around $\alpha = 0.05$. Furthermore, the OLS rejection rates for different permutation methods are expected to be the same in their data types, since OLS is the same method used across different permutation methods 1a, 1b, 2, and 3. However, there is variation ranging from 0.04 to 0.16 in type 1 error rates for OLS in each respective data type when $\beta_2 = 0$. Similarly, the power of OLS should be relatively close as we proceed down the table to non-zero values of β_2. This occurrence alludes to substantial variation in our simulation results, and calls for more extensive simulations to reach consensus among these values for OLS.

6.6 Incremental β_1, Zero Constant β_3 and β_4

To further investigate the influence of β_1, the increments of the coefficient have been color-coded for each graph in the table. Now, rather than just assessing the influence of a variable β_1 as in subsection 6.5, this paper will look at the effect when $\beta_1 = 0.2, 0.4, 0.6, 0.8$ and 1.0 individually. By separating rejection rates via the β_1 coefficient, one can analyze the performance of both inference approaches under extremely specific conditions and interactions of β_2 and β_1, while β_3 and β_4 remain constant at zero.

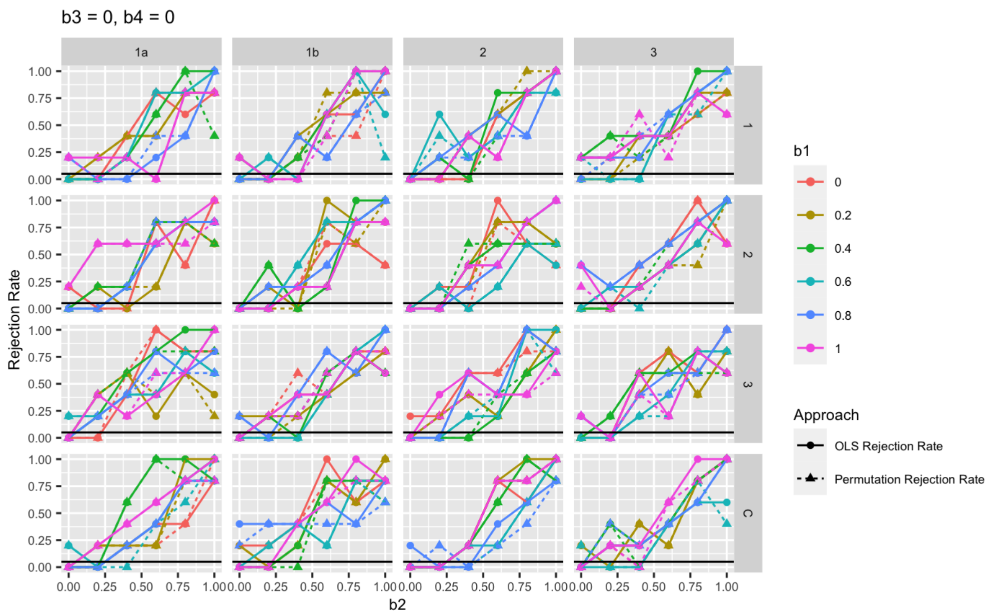

Figure 6.5: For sample size $n = 15$.

Power: It's interesting to still see a lot of the same general trends of the rejection rates as in Figure 6.3. This suggests sample size is a stronger indicator of rejection rate trends than β_1, though variability amongst β_1 amidst a changing β_2 is still very evident. In particular for permutation method 1a and OLS, $\beta_1 = 0.4$ appears to achieve the highest power the quickest (for the lowest values of β_2) for data types 1, 3, and C. Still, it is difficult to draw conclusions given the variability of the results for different values of β_1. However, higher coefficients of β_1 and β_2 in combination generally lead to higher rejection rates for OLS and across permutation methods. Meaning, no combination of coefficients for β_1 and β_2, permutation methods, or data types, seems to be consistently underperforming.

Size: Both OLS and permutation appear to either return type 1 errors of 0 for smaller values of β_1 or overestimate size at $\alpha = 0.05$ for larger values of β_1. The latter is typically seen for values where $\beta_1 = 0.8$ or $\beta_1 = 1.0$.

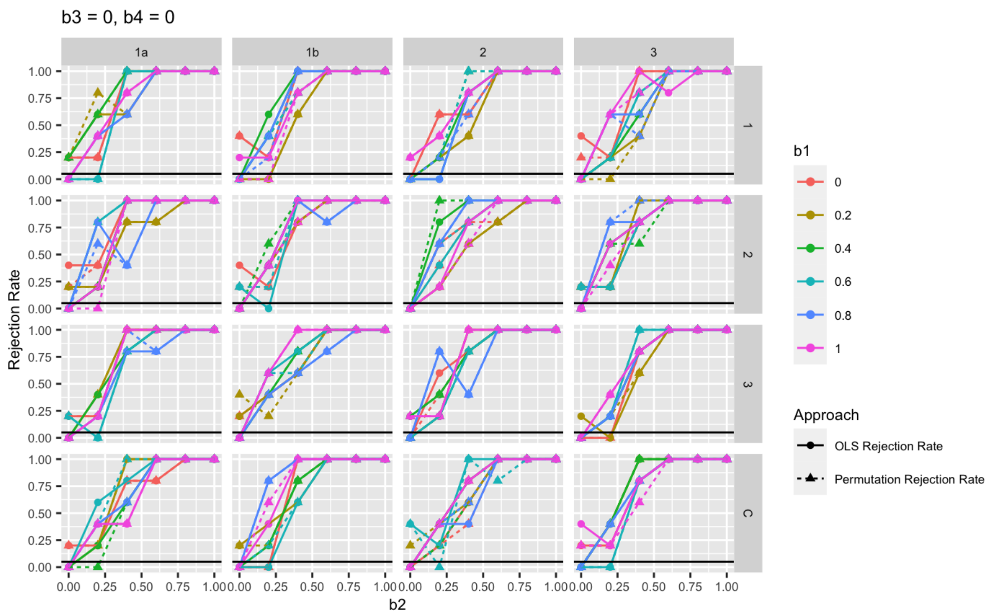

Figure 6.6: For sample size $n = 60$.

Power: Once again, there are obvious similarities between this table of graphs, Figure 6.3, and Figure 6.4. Power grows fastest across permutation methods for data type 3 when $\beta_1 = 1$, which is significant given the data structure of 3 involves important but missing variables x_{i3} and x_{i4} in the fitted model. Data structure 2 was created with a x_{i2} which was dependent on x_{i1}, and it is evident permutation method 2 (the naïve permutation) quickly grows in power for both approaches.

Size: There don't appear to be certain values of β_1 that contribute to an under or overestimation of size across permutation methods and data types, both approaches are noticeably weak at controlling size for all data types. The discrepancy in type 1 error rates was likely present in Figure 6.5 but less noticeable with the overall increased variability. It's interesting there was a decrease in the variation of rejection rates when $\beta_2 \neq 0$ but not type 1 error rates (when $\beta_2 = 0$), leading one to ponder how to specifically decrease variability in type 1 error rates if not with an increase in sample size.

6.7 Incremental β_3, Zero Constant β_1 and β_4

Rather than continue with β_1 as a variable coefficient, it will now be held constant at zero with β_4, while instead $\beta_3 \neq 0$. While β_3 is included in data structures 2 and 3, β_4 is only included in the generation of x values for data type 3, though it remains in the structure of the y values for both data type 2 and 3. The coefficient β_4 has the same affect as b_3 in data type 3: both are important since variables x_{i3} and x_{i4} are present in the data structure, though they're missing from the fitted model $y \sim x_{i1} + x_{i2}$. However, β_3 is present in data type 2 as well, as it concerns the interactive term $x_{i1} * x_{i2}$ in the population. For this reason, this section has elected to look at variable version of β_3 closely, rather than β_4.

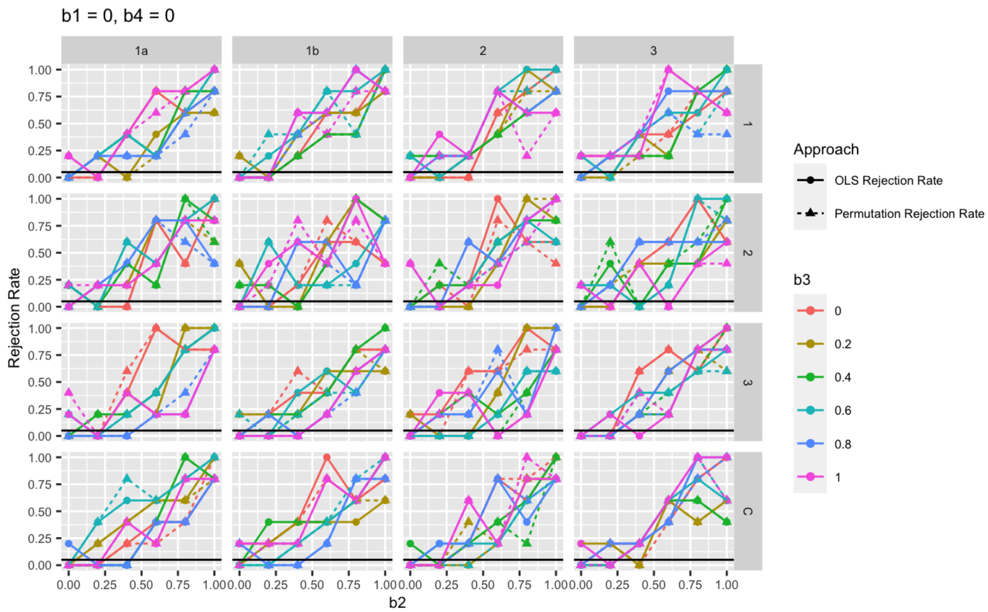

Figure 6.7: For sample size $n = 15$.
Power: Across permutation methods for data structure 3 and C, the rejection rates of both approaches look extremely similar to that seen in Figure 6.5. Across all graphs, red rejection rates ($\beta_3 = 0$) for both approaches seem nearly identical to Figure 6.5. In data structure 2 particularly, the rejection rates of OLS and permutation appear highly dependent on the value of β_3 across permutation methods (i.e., appear extremely variable). Power descreases for $\beta_2 = 1.0$, as evidenced by a decrease in power for many of the graphs, suggesting difficulty correctly rejecting the null hypothesis under these circumstances for both OLS and permutation methods after the $\beta_2 = 0.6$ threshold is reached as mentioned in Table 6.4.
Size: There is clear inconsistency in the ability of both approaches to control size as β_3 varies, pointing to a necessity for a larger sample size.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.04	0.04	0.08	0.12	0.08	0.12	0.04	0
0	1b	0.04	0.04	0.12	0.12	0.04	0.08	0.08	0.04
0	2	0.08	0.08	0.08	0.08	0.04	0.04	0.04	0
0	3	0.08	0.08	0.08	0.08	0	0	0.08	0
0.2	1a	0.16	0.16	0.12	0.12	0.04	0.04	0.12	0.12
0.2	1b	0.04	0.08	0.24	0.2	0.12	0.08	0.16	0.12
0.2	2	0.16	0.12	0.04	0.08	0.16	0.12	0.08	0
0.2	3	0.12	0.12	0.16	0.24	0.08	0	0.08	0.08
0.4	1a	0.28	0.28	0.36	0.36	0.24	0.24	0.28	0.32
0.4	1b	0.44	0.4	0.28	0.32	0.16	0.12	0.24	0.2
0.4	2	0.2	0.2	0.24	0.24	0.2	0.2	0.24	0.28
0.4	3	0.32	0.32	0.28	0.28	0.2	0.28	0.16	0.16
0.6	1a	0.36	0.28	0.52	0.52	0.28	0.28	0.44	0.44
0.6	1b	0.6	0.56	0.44	0.4	0.4	0.32	0.44	0.44
0.6	2	0.64	0.64	0.44	0.44	0.28	0.28	0.36	0.36
0.6	3	0.56	0.52	0.32	0.32	0.4	0.36	0.56	0.52
0.8	1a	0.68	0.64	0.84	0.8	0.6	0.64	0.72	0.72
0.8	1b	0.76	0.64	0.72	0.64	0.6	0.6	0.6	0.68
0.8	2	0.76	0.6	0.84	0.76	0.48	0.48	0.64	0.64
0.8	3	0.76	0.64	0.56	0.6	0.76	0.72	0.76	0.76
1	1a	0.84	0.84	0.76	0.68	0.92	0.92	0.88	0.84
1	1b	0.88	0.88	0.72	0.72	0.8	0.8	0.88	0.88
1	2	0.8	0.8	0.84	0.88	0.84	0.84	0.84	0.84
1	3	0.84	0.72	0.8	0.76	0.88	0.76	0.72	0.64

Table 6.6: There is a clear display of low rejection rates when $\beta_2 \neq 0$ across data types and permutation methods. Notably, power never surpasses 0.88 for any scenario, and permutation methods appear to significantly underperform OLS for data type 3 when $\beta_2 = 1.0$. Additionally, inflated type 1 error rates are present for permutation methods 1a and 1b for data structures 2 and 3.

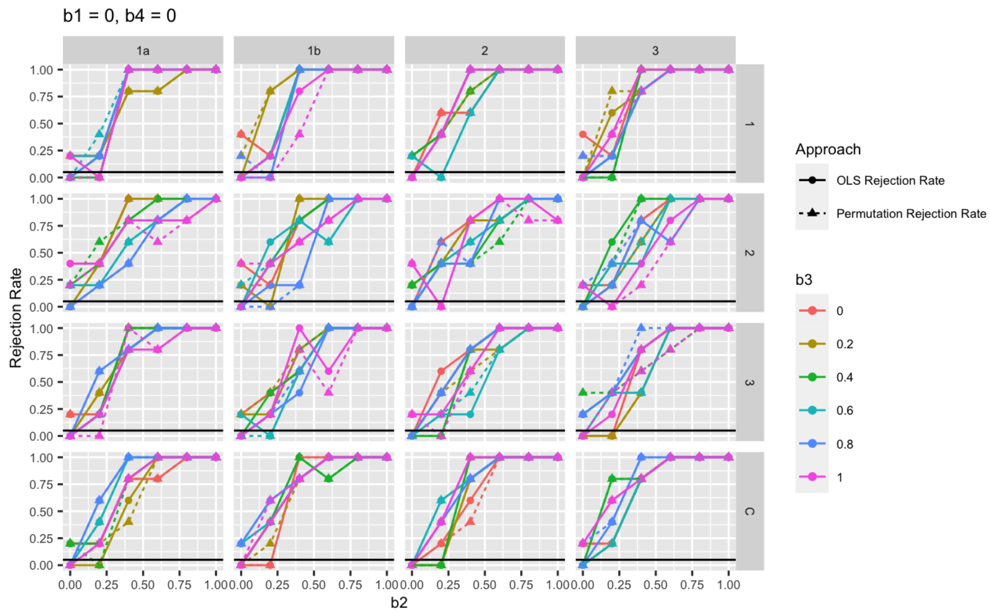

Figure 6.8: For sample size $n = 60$.

Power: Many of the graphs look quite similar to Figure 6.6, aside from data structures 2 and 3, which coincidentally are the two which contain β_3 in their structures. For these two structures, the data requires a higher β_2 coefficient to achieve the same power as the other structures across values of β_3.

Size: Once again, there is a lot of variety displayed across all graphs surrounding size. Type 1 error rates are particularly inflated for data structures 1, 2, and 3, though variability is displayed for all combinations of data types and permutation methods.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.08	0	0.16	0.12	0	0	0.04	0.04
0	1b	0	0.08	0.04	0.16	0.08	0.04	0.08	0.08
0	2	0.12	0.12	0.16	0.16	0.04	0.04	0	0
0	3	0	0.04	0.04	0.12	0.08	0.12	0.04	0.08
0.2	1a	0.12	0.16	0.32	0.36	0.32	0.28	0.28	0.28
0.2	1b	0.28	0.24	0.32	0.24	0.2	0.2	0.44	0.44
0.2	2	0.32	0.32	0.32	0.36	0.24	0.2	0.28	0.28
0.2	3	0.28	0.32	0.28	0.24	0.28	0.32	0.44	0.44
0.4	1a	0.96	0.96	0.72	0.68	0.84	0.88	0.84	0.8
0.4	1b	0.96	0.88	0.68	0.68	0.68	0.72	0.84	0.84
0.4	2	0.88	0.88	0.6	0.6	0.64	0.64	0.88	0.88
0.4	3	0.88	0.84	0.64	0.64	0.64	0.6	0.84	0.84
0.6	1a	0.96	0.96	0.88	0.8	0.96	0.96	1	1
0.6	1b	1	1	0.88	0.88	0.92	0.88	0.96	0.96
0.6	2	1	1	0.88	0.84	0.92	0.92	1	1
0.6	3	1	1	0.88	0.84	1	0.92	1	1
0.8	1a	1	1	0.96	0.96	1	1	1	1
0.8	1b	1	1	1	1	1	1	1	1
0.8	2	1	1	1	0.96	1	1	1	1
0.8	3	1	1	1	1	1	1	1	1
1	1a	1	1	1	1	1	1	1	1
1	1b	1	1	1	1	1	1	1	1
1	2	1	1	0.96	0.96	1	1	1	1
1	3	1	1	1	1	1	1	1	1

Table 6.7: The higher powers from the increase in sample size are seen concretely in the lower half of Table 6.7 for both OLS and permutation methods, though data type 2 does not reach 100% power as the other structures do until $\beta_2 > 0.6$. This being said, higher power is still achieved for lower coefficients of β_2 than previously seen, such as for values of $\beta_2 = 0.4$.

6.8 Variable β_1, Incremental β_3, and Zero Constant β_4

This next section will continue to look at incremental influences of β_3 across data structures and permutation methods. Additionally, β_1 will now be included as a variable such that $b_1 \neq 0$, although β_4 will continue to be held constant at zero.

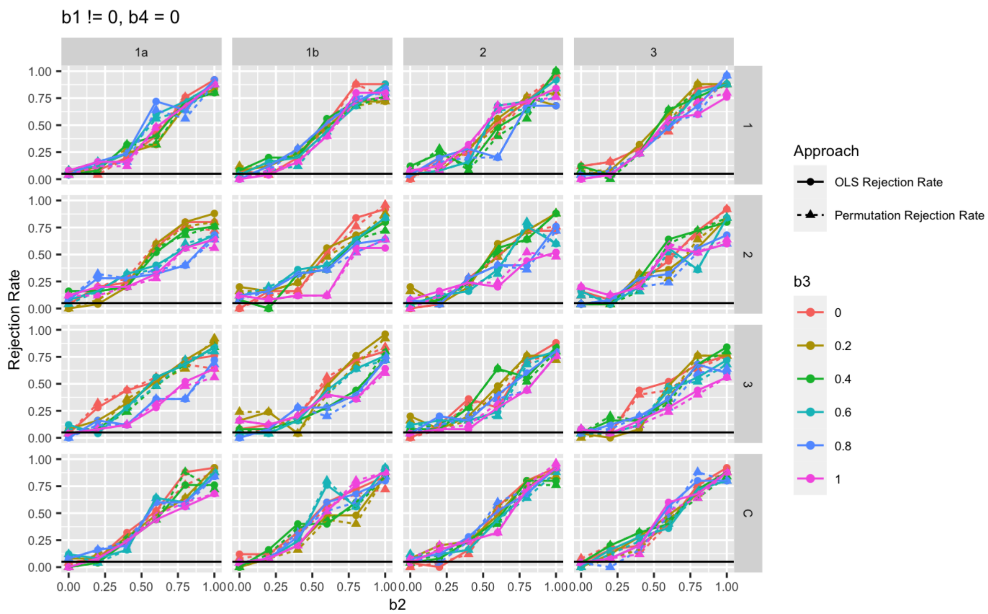

Figure 6.9: For sample size $n = 15$.

Power: Finally, but not conducive to our narrative, there is clear agreement across permutation methods and data structures. All appear to struggle with confidently rejecting the null hypothesis, though it becomes easier as β_2 increases, regardless of b_3 values. However, in data structure 2, this rejection becomes particularly hard for both approaches as β_3 increases in size, likely because it represents an interaction term which concerns both x_{i1} and x_{i2}, both of which now have variable coefficients in this display. The same is witnessed for data structure 3 across permutation methods, which has the same considerably difficulty as data structure 2 for larger coefficients of β_3.

Size: There appears to be increased agreement on controlling size across all graphs. However, data structures 2 and 3 appear to have less consensus around the y-intercept of 0.05 at $\beta_2 = 0$, though still notably stronger than previous tables of graphs.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.056	0.048	0.088	0.064	0.056	0.048	0.056	0.04
0	1b	0.04	0.04	0.128	0.112	0.088	0.104	0.024	0.024
0	2	0.064	0.064	0.088	0.064	0.096	0.088	0.064	0.08
0	3	0.048	0.048	0.096	0.088	0.056	0.048	0.024	0.024
0.2	1a	0.128	0.12	0.168	0.168	0.104	0.088	0.088	0.088
0.2	1b	0.128	0.104	0.112	0.12	0.112	0.104	0.096	0.088
0.2	2	0.144	0.144	0.088	0.08	0.12	0.104	0.136	0.128
0.2	3	0.048	0.048	0.064	0.072	0.072	0.08	0.128	0.096
0.4	1a	0.24	0.224	0.24	0.248	0.224	0.224	0.224	0.232
0.4	1b	0.216	0.216	0.272	0.264	0.168	0.176	0.264	0.256
0.4	2	0.232	0.2	0.224	0.24	0.176	0.184	0.232	0.224
0.4	3	0.256	0.24	0.256	0.208	0.144	0.144	0.256	0.2
0.6	1a	0.504	0.464	0.432	0.424	0.456	0.432	0.52	0.528
0.6	1b	0.48	0.464	0.36	0.344	0.376	0.336	0.552	0.544
0.6	2	0.512	0.48	0.424	0.392	0.408	0.392	0.44	0.464
0.6	3	0.552	0.544	0.488	0.448	0.384	0.336	0.48	0.456
0.8	1a	0.688	0.656	0.608	0.6	0.592	0.576	0.632	0.656
0.8	1b	0.728	0.712	0.624	0.592	0.52	0.488	0.616	0.584
0.8	2	0.704	0.672	0.592	0.608	0.616	0.568	0.736	0.728
0.8	3	0.744	0.76	0.544	0.552	0.624	0.592	0.72	0.712
1	1a	0.864	0.864	0.728	0.688	0.784	0.76	0.816	0.8
1	1b	0.8	0.776	0.728	0.744	0.76	0.76	0.856	0.864
1	2	0.824	0.832	0.728	0.712	0.792	0.76	0.888	0.872
1	3	0.872	0.88	0.752	0.752	0.696	0.688	0.848	0.84

Table 6.8: With this table, it becomes obvious that though type 1 error rates are finally approaching $\alpha = 0.05$, they remain multiples of 0.004, which is an odd cause for concern and suggests the simulations may not be averaging correctly to return the rejection rate. Additionally, there is more agreement across data types between OLS and permutation methods than among OLS, which is a greater cause for concern since OLS is not being manipulation with data structures as the permutation methods varies among 1a, 1b, 2 or 3. As in Table 6.6 none of the rejection rates surpass 0.89 when $\beta_2 = 1.0$.

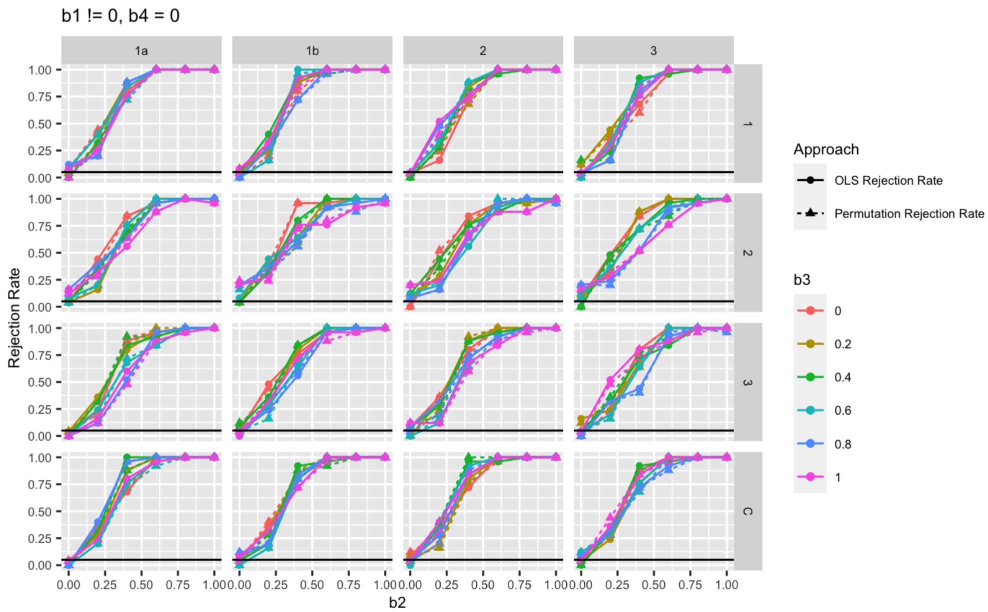

Figure 6.10: For sample size $n = 60$.
Power: All methods achieve a higher power for lower coefficients of β_2 than previously in Figure 6.9 for sample size $n = 15$. The most variability is seen in data structures 2 and 3 across permutation methods.
Size: There is noticeably more consensus of size at 0.05 for data structure 3 and C, with the same variability seeming to be present for all other data structures.

b2	method	ols_rej_1	perm_rej_1	ols_rej_2	perm_rej_2	ols_rej_3	perm_rej_3	ols_rej_C	perm_rej_C
0	1a	0.072	0.04	0.088	0.088	0.008	0.008	0.016	0.016
0	1b	0.032	0.032	0.112	0.136	0.032	0.056	0.056	0.048
0	2	0.032	0.032	0.12	0.112	0.056	0.048	0.024	0.048
0	3	0.048	0.072	0.112	0.112	0.056	0.048	0.056	0.056
0.2	1a	0.312	0.296	0.288	0.272	0.24	0.208	0.288	0.28
0.2	1b	0.288	0.264	0.368	0.336	0.296	0.28	0.24	0.256
0.2	2	0.384	0.368	0.264	0.256	0.232	0.184	0.328	0.288
0.2	3	0.288	0.256	0.336	0.328	0.32	0.328	0.304	0.336
0.4	1a	0.824	0.8	0.68	0.688	0.688	0.688	0.88	0.848
0.4	1b	0.888	0.864	0.688	0.64	0.712	0.736	0.824	0.816
0.4	2	0.824	0.768	0.68	0.696	0.776	0.76	0.872	0.88
0.4	3	0.832	0.832	0.672	0.672	0.656	0.64	0.8	0.776
0.6	1a	1	1	0.96	0.96	0.912	0.92	0.984	0.976
0.6	1b	1	0.984	0.92	0.928	0.984	0.968	0.992	0.976
0.6	2	0.992	0.992	0.92	0.936	0.928	0.928	0.992	1
0.6	3	0.992	1	0.912	0.88	0.928	0.952	0.976	0.96
0.8	1a	1	1	1	1	0.992	0.992	1	1
0.8	1b	1	1	0.976	0.96	0.992	0.992	1	1
0.8	2	1	1	0.968	0.968	1	0.992	1	1
0.8	3	1	1	0.976	0.984	1	0.992	1	1
1	1a	1	1	0.992	0.992	1	1	1	1
1	1b	1	1	0.992	0.992	1	1	1	1
1	2	1	1	0.992	0.992	1	1	1	1
1	3	1	1	1	1	1	0.992	1	1

Table 6.9: Type 1 error rates for data type 1 underestimate, while they are inflated for data type 2, and relatively the same for data types 3 and C. In other words, permutation methods are minimizing type 1 error at roughly the same rate, and data structure seems a stronger indicator of their variability in performance. Once again, higher rejection rates are achieved for lower coefficients of β_2, such as $\beta_2 = 0.8$ and $\beta_2 = 0.6$ than in Table 6.8.

6.9 Incremental β_3, Variable β_1 and β_4

Finally, this section will introduce a variable β_4 into the data structure. Since data type 3 is the only one which includes the variable x_{i4} in its population, it's expected that all other data structures will provide similar results to Figure 6.9 and Figure 6.10, given some normal variability is expected between reproductions of these simulations.

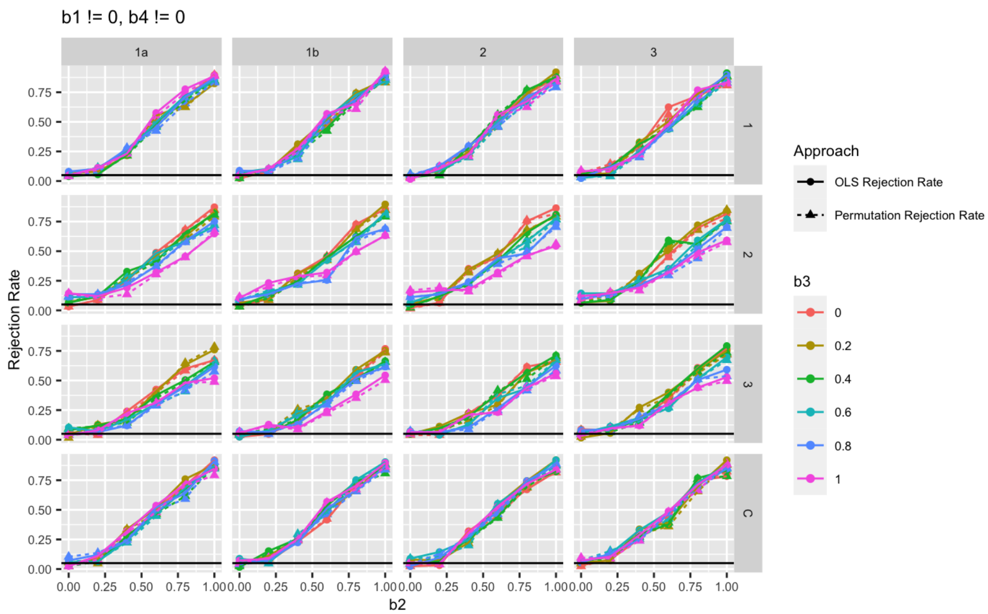

Figure 6.11: For sample size $n = 15$.

Power: The addition of a fourth varying coefficient served to reduce deviation across the data structures. Presumably, it allowed for more iterations with the same results for those which do not contain β_4, and this resulted in a stronger consensus across rejection rates for different β_2 and β_3. Even in data structure 3, which contains x_{i4} in its population, there is less variation in power between coefficient sizes of β_3 and β_2. Lower values for coefficient β_3, such as $\beta_3 = 0.2$ and 0.4 are rejecting at higher rates across data type 2 and 3.

Size: Allowing for variation of β_4 also improved the ability of both approaches to approximate a rejection rate of 0.05 across values of β_3 when $\beta_2 = 0$.

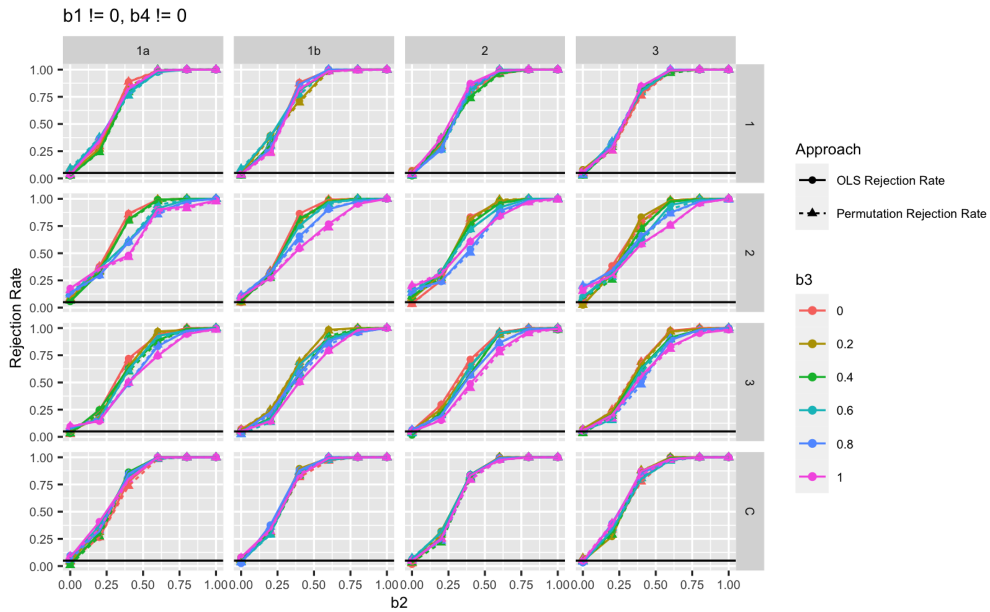

Figure 6.12: For sample size $n = 60$.

Power: As with all other scenarios, an increase in sample size leads to an increase in rejection rate for lower values of β_2. Although there is more variability displayed in power for data structures 2 and 3, there is overall much more consensus in rejection rates across all variables. But if this paper is nitpicking, permutation methods appear particularly poorer for method 2 and data structure 3 with large coefficients of β_3, just as it did in Figure 6.10. The previous outperformance of rejection rates for $\beta_3 = 0.2$ and 0.4 in comparison to $\beta_3 0.6, 0.8,$ and 1.0 for data types 2 and 3 mentioned in Figure 6.11 is no longer present.

Size: Size does not appear to have improved in controlling tests at a rejection rate of 0.05 despite the increase in sample size to $n = 60$.

Chapter 7

Discussion

Recall, the purpose of this paper is to discern if the naïve permutation (permutation method 2) is in fact as problematic as described in literature. Permutation method 2 follows logic in the sense that the user permutes the variable of interest, x_{i2}, in the inference test. Other permutation methods described involve permutation of the response variable and residuals in order to inquire about x_{i2}. For this reason, this paper criticizes these methods as less accessible and straightforward than that of the naïve permutation. In order to dissect this argument, all permutation methods described were run in comparative simulations, as was OLS. The results of these simulation studies allow this paper to discuss the validity of the pertinent question regarding the naïve permutation. The method can be compared not only across permutation methods, but across the popular OLS method.

For $\beta_2 = 0$ in the tables of the previous section, it's intended that the rejection rate approximate the significance level $\alpha = 0.05$. In comparison to OLS and other methods, permutation method 2 consistently returns a size of 0 until the scenario described in Figure 6.7. Once the variation of β_3 is introduced, permutation method 2 performs at the same level or worse (for control data structure) than OLS. This trend continues into Figure 6.9 where the naïve permutation method is in fact better than OLS for data types 2 and 3. With two variable coefficients, permutation method 2 is increasingly better at approximating the size at 0.05, and this continues to be true as sample size increases.

Beginning with $\beta_2 = 0.2$, the discussion switches to whether permutation method 2 is better than OLS at rejecting the null hypothesis. With all coefficients aside form β_2 set at zero, OLS sometimes outperforms permutation method 2, while other times, the opposite is the case. With only five reps, there is some obvious variation in the values returned by OLS and the permutation test. Neither differ significantly from the other in this scenario until an increase in sample size, where OLS outperforms permutation method 2 under data types 2 and 3. In the scenario given by Figure 6.3 where β_1 is variable, OLS either equates or outperforms the naïve permutaion for both sample sizes across data types. The same outperformance is seen amplified in Figure 6.7. However in Figure 6.9, the difference between these two approaches becomes hundredths of a decimal place for both sample sizes.

The other coefficients continue on with many similar patterns to those described above. Since the variation of the OLS rejection rates and consequently, likely the permutation rejection rates as well suggests the necessity for more extensive simulations and reproducible results, this paper will cease to discuss further simulated scenarios (i.e., non-normally distributed data and errors, larger sample sizes, and smaller increments in coefficient size, to name a few suggestions for research).

Finally, this paper suggests it would be helpful to know exact values of the variable β_1, β_3, and β_4, rather than simply $\beta_i \neq 0$, so the result of their interaction with an incremental β_2 can be assessed for patterns in rejection rates. In the rejection rates, multiples of 0.04, then 0.04, were seen. This hints at an issue in averaging across simulations. Furthermore, this paper expected to see more consistency in OLS rejection rates across permutation methods for each data type. The fact that this consensus was not present points to errors in the simulation of results, since OLS rejection rates were more consistent with permutation rejection rates. The goal of this paper was to find evidence of how the naïve permutation, permutation method 2, varied from other permutation methods and OLS rejection rates across different scenarios of data structures. However, permutation method 2 did not seem to perform substantially worse than other permutation methods. In fact, permutation methods in general did not appear to significantly underperform in comparison to OLS in this simulation study. This conclusion suggests an easily accessible permutation method is nearly as reliable as the favored OLS in hypothesis testing.

Bibliography

[Aldous, 1985] Aldous, D. J. (1985). Exchangeability and related topics. In Hennequin, P. L., editor, *École d'Été de Probabilités de Saint-Flour XIII — 1983*, pages 1–198, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Anderson, 2001] Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(3):626–639.

[Anderson and Legendre, 1999] Anderson, M. J. and Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62(3):271–303.

[Anderson and Robinson, 2001] Anderson, M. J. and Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, 43(1):75–88.

[Collins, 1987] Collins, M. F. (1987). A permutation test for planar regression. *Australian & New Zealand Journal of Statistics*, 29:303–308.

[for Statistics Education,] for Statistics Education, T. I. Pivotal statistic.

[Ghosh et al., 2010] Ghosh, M., Reid, N., and Fraser, D. A. S. (2010). Ancillary statistics: A review.

[Kennedy, 1995] Kennedy, P. E. (1995). Randomization tests in econometrics. *Journal of Business Economic Statistics*, 13(1):85–94.

[Kennedy and Cade, 1996] Kennedy, P. E. and Cade, B. S. (1996). Randomization tests for multiple regression. *Communications in Statistics - Simulation and Computation*, 25(4):923–936.

[Küchler, 1999] Küchler, I. (1999). Data driven statistical methods. peter sprent, chapman & hall, london 1998. no. of pages: 406. price. £35. isbn 0-412-79540-x. *Statistics in Medicine*, 18(22):3153–3154.

[Manly, 1997] Manly, B. F. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall.

[Oja, 1987] Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Australian Journal of Statistics*, 29(1):91–100.