Simultaneous Confidence Intervals Using the Restriced Scheffe Method and Sequential Testing Procedures for Single-step Methods

By

Yimin Zhang

Bachelor of Science
Dalian University of Technology
Dalian, P.R.China
2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2014

Simultaneous Confidence Intervals Using the Restriced Scheffe Method and Sequential Testing Procedures for Single-step Methods

Dissertation Approved:

Dr. Melinda McCann
Adviser

Dr. Ibrahim Ahmad

Dr. Lan Zhu

Dr. Anthony Kable

Name: YIMIN ZHANG

Date of Degree: MAY,2014

Title of Study: SIMULTANEOUS CONFIDENCE INTERVAL USING THE RESTRICTED SCHEFFE METHOD AND SEQUENTIAL TESTING PROCEDURES FOR SINGLE-STEP METHODS

Major Field: STATISTICS

Abstract: Multiple comparison problems arise when a set of inferences are considered simultaneously. When a separate inference approach is used to answer multiple related questions, it can cause many false positives and lead to erroneous conclusions. Multiple comparison procedures are used to deal with this issue and properly control for multiplicity through some joint measure of erroneous inference. A plethora of measures for error rates, as well as many methods addressing them, have been developed, although in this dissertation, we restrict our attention to familywise error rate (FWE), the probability of making at least one false positive among all the hypotheses.

Casella and Strawderman (1980) constructed a Scheffe-type confidence band for multiple regression over a restricted range of the predictor variables, the restricted Scheffe method. We propose applying this restricted Scheffe method to a discrete set of multiple comparisons. This proposed method requires minimal assumptions on the distribution of the estimated parameter vector and gives a less conservative solution than Scheffe's method. A rectangle embedding approach was introduced also by Casella and Strawderman (1980) to find appropriate restricted ranges for practical problem settings. However, this approach, which was developed for regression-type problems, encompasses a large excess of comparisons, and consequently causes rather conservative critical values. A new minimal cone approach is developed to address this issue by utilizing the discreteness of the comparisons to obtain the optimal restricted cone-shaped range.

Sequentially rejective methods are another type of multiple comparison methods in which the current step result depends on the test results of previous steps. We propose three sequentially rejective methods based on single-step methods, all of which control the FWE, to achieve improved power in multiple testing. We suggest modifications to the critical values such that the modified critical values are monotone at all times. To facilitate computation, two of the three proposed methods are developed to modify monotone critical values along the rejection path. Moreover, new sequentially rejective methods are developed from applying these modifications to the hybrid method proposed by McCann and Edwards (2000), and they are shown to often outperform Holms procedure and the stepdown Sidaks method.

ACKNOWLEDGMENTS

Foremost, I thank the members of my committee, who have all been professors of mine. Dr. McCann, my advisor, deserves special thanks. To her, I owe my career. This August will mark five years since I met Dr. McCann, and in that time, I have grown as a scholar and as a person through my interactions with her. Hers was my first class in Oklahoma State, and I had the pleasure of learning about the world of statistics from her in four classes. Over the past two years, she has been instrumental in helping me reach this point in my career. Her forensic scrutiny of my technical writing has been invaluable. She has always found the time to propose consistently excellent improvements. I owe a great debt of gratitude to you, Dr. McCann.

I would like to express my deep gratitude to the other members in my dissertation committee, Dr. Ahmad, Dr. Zhu and Dr. Kable, for reading previous drafts of this dissertation and providing many valuable comments that improved the presentation and contents of this dissertation. During the process, they have provided, with kindness, their insight and suggestions, which are precious to me. I would also like to thank Dr. Payton, my previous committee member, for offering thorough feedback on an earlier version of this dissertation.

My thanks also go to the other faculty members, Dr. Goad, Dr. Habiger and Dr. Liang, who have provided me with a tremendous graduate education: they have taught me how to think about statistical problems and shown me how to approach my work as a statistical scientist.

A special mention for Dr. Darcy, my first teaching supervisor and a lifelong friend, for his insightful advice many times during my graduate school career. He was and remains one of my best role models for a researcher, mentor, and teacher.

To my friends and colleagues: Ningning Wang, Xuesong Li, Ole Forsberg, Nicholas Kaukis, David Watts, Thagendra Timsina, Sayed Mostafa, Ana Tehrani, Ryan Lusk, Fran Mihura, Renee Flasch, Lichi Lin, Su Chen, Jordan Crabbe, Jack Knorr, Xue Huang, Qin Pan, Kyuseong Wee, and many others, thank you for making life in the doctoral program bearable, and yes, even fun.

Last, but not least, I would like to thank my parents, Hong Yun and Luqing Zhang, for the love and encouragement they have shown me throughout my life. Nothing I have accomplished would have been possible without them. My endlessly patient boyfriend, Rohit, deserves special thanks for his unfailing love, support and sticking with me throughout the process. I may never be able to repay the debt I owe him. I love all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Multiple comparisons arise when we consider a set of inferences simultaneously. Analogous to the type I error in a single hypothesis test, in multiple comparisons, the familywise error rate (FWE) measures the probability of making at least one incorrect assertion among all the hypotheses. There are basically two types of FWE control. Controlling FWE in the weak sense is to control the probability of making at least one rejection given that all the hypotheses are true. Controlling the FWE in the strong sense, which is more stringent, is to control the supremum of FWE over any possible configuration of the true and false null hypotheses. In practice, we do not know which null hypotheses are true and which are false. So, we should protect against all the scenarios where all or some null hypotheses are true to fully control the FWE, i.e. to control the FWE in the strong sense. Hence, all the multiple comparison methods that we investigate in this dissertation will provide strong control of the FWE.

Researchers often pursue multiple comparisons that provide inferences related to their research conclusions. These research conclusions are often made based on the entirety of results from the comparisons. In most situations, a false positive could lead to an incorrect conclusion. For example, one may want to conduct all pairwise comparisons on multiple treatment means to rank the means, if possible. However, even one falsely significant difference between two means could change the overall rankings. Thus, an accurate overall conclusion tolerates no false positives and consequently FWE control is appropriate.

Consider the following example. Blazer et al. (1985) reports results of a survey of residents in the Piedmont region of North Carolina where the population was nearly equally distributed between urban and rural counties. There were 3798 subjects chosen from randomly sampled households over 5 counties in the Piedmont region. One adult participant was randomly selected from each sampled household. Rural/urban

residence was determined by county. In this survey, each subject went through a 2-hour interview and the interview was utilized to diagnose each subject as positive or negative for each of nine psychiatric disorders. The main goal of this study was to explore differences in the prevalence of the 9 psychiatric disorders between urban and rural residents. Therefore, there were 9 comparisons of interest. The researcher wanted to reach an overall conclusion on which of the 9 psychiatric disorders varied in prevalence between urban and rural residents. To control the probability of occurrence of any false positive, the researchers would have needed to control the FWE in the strong sense, which is to control the probability of mistakenly claiming that any psychiatric disorder differed in rural/urban prevalence when that psychiatric disorder actually did not. Since the probability of at least one error is obviously greater than the probability of an error on a single comparison, making single comparisons individually at the specified error rate does not control the FWE, and could lead to a greatly inflated FWE in some cases.

In the above example, the researchers might wish to only determine diagnosis prevalence equality or inequality. In some situations, researchers go further and construct confidence intervals for the prevalence differences. Depending on these different inferential objectives, various multiple comparison methods may be appropriate. For example, a pharmaceutical company faces two issues: passing drugs through the regulatory agency and marketing drugs. For a combination drug to pass the regulatory agency, it requires convincing evidence that the combination drug is better than every one of its subcombinations. The desired task is to detect directional differences, in the direction that the combination drug must be more effective than all of its subcombinations. Thus a declaration of directional differences suffices; confidence intervals for the differences are not needed. On the other hand, for marketing purposes, when comparing several drugs with the most effective one on the market, the desired task is to evaluate practical differences and see if the superiority any drug gains over the most effective one is useful enough to convince the company that the new drug is

worth marketing. Thus, two inferential tasks in multiple comparisons are capturing significant directional differences and providing interval estimation to assess practical differences. For instance, we may make 0-1 (yes-no) decisions based on directional differences, or construct confidence intervals that provide the additional information needed to assess practical differences.

As mentioned, there are different types of multiple comparison methods to address these different inferential tasks. Westfall et al. (1999) referred to single-step methods as methods that make an equivalent multiplicity adjustment for all the tests or confidence intervals of interest. As a result, each test or confidence interval is calculated without taking the other test results into account. In stepwise methods, on the other hand, the current step result depends on the test results of previous steps. Consequently, simultaneous confidence intervals cannot generally be derived from a stepwise method. For instance, generally, stepwise methods that test the equality of all pairwise treatment means fail to provide a set of simultaneous confidence intervals for mean differences. However, if only hypothesis testing is of concern and simultaneous confidence intervals are not needed, then stepwise methods can increase the power of the testing procedure, while still controlling the familywise error rate.

To pursue increased power, single-step methods can often be adapted into stepwise methods, for example step-down methods. For simplicity, suppose the rejection region of a test is based on large values of a test statistic. Step-down methods begin by testing the joint null hypotheses using a single-step method and continue rejecting more hypotheses in the subsequent steps. At each step, the single-step test is conducted on the current collection of unrejected hypotheses. If any rejection occurs, in the next step the testing will be based on an updated collection with the newly rejected hypotheses removed. These sequential rejections are realized by decreasing the critical values, as a result of the previous rejections, for the remaining hypotheses at each subsequent step. Once a step is reached where no rejections occur, the method stops. As an example, the Holms' method is a stepdown extension of the Bonferroni

method. Similarly, we propose modifying other single-step methods that have not yet been utilized to obtain step-down procedures. The Hunter-Worsley method, the tube method, the hybrid method and the restricted Scheffe method are four single-step methods that often perform well in terms of simultaneous confidence intervals for various situations. They all provide strong control of the FWE. However, the hybrid method incorporates the tube method and the Hunter-Worsley method and consequently always outperform them. Thus, we propose obtaining step-down versions of the hybrid method and the restricted Scheffe method. These methods will control the FWE and should provide increased power for various situations.

As mentioned, we consider several powerful FWE-controlling single-step methods for modification into stepwise procedures. The Hunter-Worsley method is a single-step method that arises from the Hunter-Worsley inequality (Hunter, 1976; Worsley, 1982). The Hunter-Worsley inequality states that the probability that at least one of a set of events occurs is no greater than the sum of the probabilities of the individual events, which gives the Bonferroni inequality, minus the sum of probabilities of the pairwise events taken over a spanning tree of all pairwise intersection events. (The definition of a spanning tree will be provided in Chapter 2.) Given that the Hunter-Worsley inequality sharpens the upper bound in the Bonferroni inequality, it is also referred to as an improved Bonfferoni inequality. Notice that the Hunter-Worsley inequality involves the probabilities of the pairwise intersection events and thus incorporates information about the pairwise correlations, which is something that the Bonferroni method fails to consider.

Another single-step method available for procuring a step-down procedure is the "tube" method presented in McCann and Edwards (1996). It provides a new confidence bound for the multivariate-t distribution, which utilizes the results in Naiman (1986) and bounds the joint error probability as the surface of the union of disks with a capped tube. The tube method also guarantees dominance over the Scheffe (1959) method.

A hybrid method that incorporates the Hunter-Worsley method and the tube method and improves on both of them, is a single-step method proposed by McCann and Edwards (2000). In the paper of McCann and Edwards (2000), they utilized the work of Uusipaikka (1984) and Naiman (1986) to express the probability that simultaneous coverage for confidence intervals fails as an expectation of the surface area of a union of disks on the unit sphere within a random angular distance of a vector. Conveniently, many conservative methods, including the Hunter-Worsley method and the tube method, can be considered using different upper bounds for the surface area. The hybrid method, in particular, chooses to use the minimum of the upper bounds of the Hunter-Worsley and the tube methods, which provides a less conservative critical value. As a result, the hybrid method always outperforms the Hunter-Worsley and the tube methods, and hence the Bonferroin method and Scheffe's method as well.

Finally, we consider the restricted Scheffe method which was developed by Casella and Strawderman (1980). This is a single-step method that gives a sharper confidence bound than the Scheffe method. It restricts the range of predictor variables in a regression function into a cone, rather than allowing for any values of the predictor variables, and therefore provides a narrower confidence band for the regression function over the restricted region. In order to utilize the restricted Scheffe method for practical examples, Casella and Strawderman (1980) further introduced a rectangle embedding approach in their paper. This approach first assumes the range of predictor variables has a hyper-rectangle shape, then finds a cone such that the hyper-rectangle can be embedded into the cone, and then the restricted Scheffe method is applied to this cone.

When we considered modifying the restricted Scheffe method to obtain a stepdown method, we realized that no previous publication in the literature has investigated applying the restricted Scheffe method to construct simultaneous confidence intervals for a finite set of comparisons. Unfortunately in the finite comparisons situation, the

rectangle embedding approach often fails to find a cone of the form required by the restricted Scheffe method to enclose the coefficients of all the comparisons. In this case, even the tightest hyper-rectangle turns out to be an overexpansion of the convex hull of actual coefficients in the sense that after transformation the hyper-rectangle set is too large to fit into a cone with the structure specified in the restricted Scheffe method. In light of this situation, we propose obtaining the exact minimal cone from the convex hull of the coefficients of all the comparisons, completely avoiding the troublesome hyper-rectangle. We will refer to this as the minimal cone approach and refer to the restricted Scheffe method utilizing the minimal cone approach as the minimal restricted Scheffe method. As this approach has not been considered, we also focus on obtaining simultaneous confidence intervals via the minimal restricted Scheffe method and evaluating its performance for a finite set of comparisons with other appropriate single-step methods.

Now consider stepdown methods in general. Although the mechanism of stepdown methods provides the possibility of improved power, it is crucial to verify that these stepdown methods control the FWE. Otherwise the gain in power could actually be at the cost of increasing the FWE. Goeman and Solari (2010) proposed a general sequentially rejective multiple testing framework that sheds light on proving strong FWE control of stepwise methods. Specifically, they show that the sequentially rejective framework strongly controls the familywise error if it fulfills the criterion of monotonicity of the critical values, *the monotonicity condition*, and a limited form of weak familywise error control in each single step, *the single-step condition*. These two conditions are referred to as the sequential rejection conditions. Since stepdown methods fit perfectly into the sequentially rejective framework, we are inspired to justify the strong FWE control of stepdown methods by means of the sequential rejection conditions. At each step, our stepdown methods are reduced to single-step methods which all control the FWE in the strong sense, and consequently the single-step condition easily follows. However, the true challenge comes from showing that

the monotonicity condition holds. Although simulations have empirically vindicated monotonicity of the critical values for the stepdown methods considered, theoretical proofs of monotonicity, especially for the critical values from single-step methods that take correlations between the hypotheses into account, has proved considerably difficult. In view of this difficulty and with the addition of the subset pivotality assumption, we suggest a modification to the critical values such that the modified critical values are monotone at all times, while still satisfying the single-step condition. We refer to the stepdown methods based on these modified critical values as forced stepdown methods, as we are essentially forcing monotonicity. To facilitate computation, two alternative methods are likewise proposed that force monotone critical values along the rejection path and hence are referred to as the forced pathwise methods. All these various modifications are utilized to develop sequential tests from the hybrid method, all of which control the FWE in the strong sense.

In summary, we propose applying the minimal cone approach to the restricted Scheffe method for constructing simultaneous confidence intervals for a finite set of comparisons, and compare its performance with various other single-step methods. Additionally, forced stepdown methods and forced pathwise methods are proposed to develop sequential methods utilizing the hybrid procedure and the minimal restricted Scheffe method. The performance of these sequential methods is compared to that of the appropriate competitors.

The remaining chapters are organized as follows. In Chapter 2 we review various single-step methods and the restricted Scheffe method in the context of the general multiple comparison problem. Then we illustrate the application of the restricted Scheffe method to practical problems using the rectangle embedding approach and introduce the minimal cone approach to resolve the issues that occur in applying the restricted Scheffe method to a finite set of comparisons. After that, we conduct a simulation study to compare simultaneous confidence intervals constructed by the minimal restricted Scheffe method with other appropriate single-step methods. In

Chapter 3 we discuss the concept of the sequentially rejective framework and the sequential rejection conditions, and propose the forced step-down method and the two forced pathwise methods. We show that under the subset pivotality assumption these procedures all control the FWE in the strong sense. Next, we present the results from investigating the performance of the forced sequential procedures utilizing the hybrid method and the minimal restricted Scheffe method with other stepwise competitors. A summary follows in Chapter 4.

2 Minimal Cone Approach

2.1 Single-step methods

McCann and Edwards (1996) considered a general multiple comparison model with k parameters $\boldsymbol{\beta'} = (\beta_1, \beta_2, \beta_3, ..., \beta_k)$. Suppose it is of interest to estimate $\boldsymbol{C\beta}$, namely p linear combinations of β, where $\boldsymbol{C}$ is a $p \times k$ known matrix with rank r. An estimator $\boldsymbol{b}$ of $\boldsymbol{\beta}$ is available and distributed as multivariate normal $N(\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$ with σ^2 unknown. Without loss of generality, assume that $\boldsymbol{V}$ is a known positive definite matrix. Let s^2 denote the usual unbiased estimator of σ^2 with degrees of freedom ν, where $\nu s^2/\sigma^2 \sim \chi^2_\nu$ and is independent of $\boldsymbol{b}$. Denote the correlation matrix of $\boldsymbol{Cb}$ by $\boldsymbol{R} = [r_{ij}]$ which is a $p \times p$ matrix and derived from the covariance matrix $\boldsymbol{CVC'}$.

Suppose we want to construct $100 \times (1-\alpha)\%$ simultaneous confidence intervals for $\boldsymbol{C\beta}$ of the form

$$\boldsymbol{c_j'b} \pm ds(\boldsymbol{c_j'Vc_j})^{1/2}, \qquad j = 1, 2, ..., p, \tag{2.1}$$

where $\boldsymbol{c_j'}$ is the j^{th} row of $\boldsymbol{C}$.

The following section details several single-step methods that give conservative simultaneous confidence intervals of the form (2.1) with various values for d.

First define $Row(\boldsymbol{A})$ to be the set of all row vectors of a matrix $\boldsymbol{A}$. Then many classic multiple comparisons methods have provided various conservative critical values d in (2.1), such that

$$P(S_{\boldsymbol{\beta}}(Row(\boldsymbol{C})) := \{\boldsymbol{b} : |\boldsymbol{y'b} - \boldsymbol{y'\beta}| \leqslant ds(\boldsymbol{y'Vy})^{1/2}, \forall \boldsymbol{y} \in Row(\boldsymbol{C})\}) \geq 1 - \alpha, \tag{2.2}$$

where for any set $E \subseteq \mathbb{R}^k$, $S_{\boldsymbol{\beta}}(E)$ denotes the event $\{\boldsymbol{b} : |\boldsymbol{y'b} - \boldsymbol{y'\beta}| \leqslant ds(\boldsymbol{y'Vy})^{1/2}, \forall \boldsymbol{y} \in E\}$.

For example, the Bonferroni method follows from the Bonferroni inequality

$$1 - P(S_{\boldsymbol{\beta}}(Row(\boldsymbol{C})) = P(|\boldsymbol{c_j'}(\boldsymbol{b} - \boldsymbol{\beta})| > ds(\boldsymbol{c_j'}V\boldsymbol{c_j})^{1/2} \text{ for at least one } j)$$
$$\leqslant \sum_{j=1}^{p} P(|\boldsymbol{c_j'}(\boldsymbol{b} - \boldsymbol{\beta})| > ds(\boldsymbol{c_j'}V\boldsymbol{c_j})^{1/2}). \tag{2.3}$$

In fact, the Bonferroni method chooses the critical value d_{BON} as the solution to an equation setting (2.3) equal to a predetermined level α. Since $\frac{\boldsymbol{c_j'b} - \boldsymbol{c_j'\beta}}{s(\boldsymbol{c_j'Vc_j})^{1/2}}, j = 1, 2, ..., p$, are identically distributed as t_ν, the Bonferroni critical value d_{BON} is equal to $t_{\frac{\alpha}{2p},\nu}$, the upper $100 \times (1 - \frac{\alpha}{2p})^{\text{th}}$ quantile of a t distribution with ν degrees of freedom.

Very closely related to the Bonferroni method is Sidak's method. The following Sidak equation is derived by assuming independence of all the p tests of our context.

$$P(S_{\boldsymbol{\beta}}(Row(\boldsymbol{C})) = \{\boldsymbol{b} : |\boldsymbol{c_j'b} - \boldsymbol{c_j'\beta}| \leqslant ds(\boldsymbol{c_j'Vc_j})^{1/2}, \forall j = 1, 2, \ldots p\})$$
$$= \prod_{j=1}^{p} P(|\boldsymbol{c_j'}(\boldsymbol{b} - \boldsymbol{\beta})| \leqslant ds(\boldsymbol{c_j'}V\boldsymbol{c_j})^{1/2}). \tag{2.4}$$

Without the independent assumption, Sidak (1967) showed that the equality (2.4) becomes the inequality "$\geqslant$", when $\frac{\boldsymbol{c_j'b} - \boldsymbol{c_j'\beta}}{s(\boldsymbol{c_j'Vc_j})^{1/2}}, j = 1, 2, ..., p$, are multivariate normal random variables with means 0. Jogdeo (1977) extended Sidak's inequality to a larger collection of distributions including multivariate F and t distributions, where the latter coincides with the distribution of the test statistics in our model. Thus, Sidak's critical value results from equating the lower bound (2.4) to $(1 - \alpha)$ and is given by $t_{\frac{1-(1-\alpha)^{1/p}}{2},\ \nu}$.

Naiman (1986) employed the result of Uusipaikka (1984) and expressed the probability that coverage fails for the intervals (2.1) as the probability that a random vector uniformly distributed on the unit sphere in $\mathbb{R}^r$ space, falls within a random distance from at least one of the vectors $\boldsymbol{a_j}$, where $\boldsymbol{a_j}$'s, $j = 1, 2, ..., p$, are the rows of a $p \times r$ matrix $\boldsymbol{A}$ such that $\boldsymbol{R} = \boldsymbol{AA'}$. Let $\boldsymbol{z}$ be a random variable distributed as $N_r(\boldsymbol{0}, I)$. Now the probability that simultaneous coverage of the intervals (2.1) fails,

can be given in the following form,

$$\begin{aligned}
&P(|\boldsymbol{c_j'}(\boldsymbol{b}-\boldsymbol{\beta})| > ds(\boldsymbol{c_j'}V\boldsymbol{c_j})^{1/2} \text{ for at least one } j) \\
&= P(\cup_{j=1}^{p}[|\boldsymbol{a_j'}\boldsymbol{z}/Q| > d]) \\
&= P(\cup_{j=1}^{p}[|\boldsymbol{a_j'}\boldsymbol{u}| > dT]) \\
&= \int_0^{\frac{1}{d}} P(\cup_{j=1}^{p}[|\boldsymbol{a_j'}\boldsymbol{u}| > td]) f_T(t)\, dt, \qquad (2.5)
\end{aligned}$$

where $\boldsymbol{u} = \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|}$, $T = \frac{Q}{\|\boldsymbol{z}\|}$, with $\|.\|$ being vector length, and $Q^2 \backsim \chi^2_\nu/\nu$. Note that T is independent of $\boldsymbol{u}$ and $rT^2 \backsim F_{\nu,r}$. Geometrically, each set of $\{\boldsymbol{u}||\boldsymbol{a_j'}\boldsymbol{u}| > td\}$ is a disk involving all points on the unit sphere that are within the angular distance $\arccos(td)$ of $\pm\boldsymbol{a_j}$.

In most situations, it is fairly difficult to obtain the exact expression of $B(t) \equiv P(\cup_{j=1}^{p}[|\boldsymbol{a_j'}\boldsymbol{u}| > td])$. Thus, appropriate upper bounds may be employed to replace $B(t)$ in (2.5), leading to various conservative single-step methods (McCann and Edwards, 2000). Precisely, the critical values of various conservative single-step methods can be expressed as the solutions to an equation that sets (2.5), with an appropriate upper bound $B^*(t)$ for $B(t)$, equal to a desired error probability, i.e.

$$\int_0^{\frac{1}{d}} P(B^*(t)) f_T(t)\, dt = \alpha, \qquad (2.6)$$

where $B^*(t) \geqslant B(t)$, for any t between 0 and $\frac{1}{d}$. As examples, the Bonferroni method can be expressed as the solution to (2.6), using

$$B^*_{\text{Bon}}(t) = \sum_{j=1}^{p} P(|\boldsymbol{a_j'}\boldsymbol{u}| > td) = pF_{r-1,1}[\frac{(td)^{-2}-1}{r-1}].$$

2.1.1 The Tube Method

McCann and Edwards (1996) presented a tube method using the result of Naiman (1986). Note that its critical value can be expressed as the solution to (2.6) using

$$B^*(t) = \frac{\Lambda}{\pi} F_{r-2,2}[\frac{2((td)^{-2}-1)}{r-2}] + F_{r-1,1}[\frac{(td)^{-2}-1}{r-1}], \tag{2.7}$$

where $\Lambda = \sum_{j=1}^{p-1} \arccos(|r_{j,j+1}|)$. From the geometric perspective, (2.7) covers the surface area of the union of the disks $\{|\boldsymbol{a}_j'\boldsymbol{u}| > td\}, j = 1, 2, ..., p$, with the surface area of a single long tube of length Λ and the two half disks that cap the tube. In fact, the tube method uses a sharper bound $B^*_{\text{tube}}(t) = \min((2.7), 1)$ so it always outperforms Scheffe's method.

2.1.2 The Hunter-Worsley Method

The Hunter-Worsley inequality provides as an upper bound for the probability of a union event, the sum of the probabilities of individual events minus the sum of the probabilities of a set of pairwise intersection events, where the pair combinations come from a spanning tree with each event a node. A spanning tree of p nodes is a connected graph with $(p-1)$ branches. That implies that in a spanning tree each node is connected to every other node through a unique branch and the tree contains no cycles. It is easy to see that the Hunter-Worsley inequality always provides a sharper bound than the Bonferroni inequality. In particular, the sharpest bound in the Hunter-Worsley inequality occurs when the chosen spanning tree maximizes the sum of probabilities of pairwise intersection events of all the spanning trees. This sharpest upper bound is utilized in the Hunter-Worsley method for $B^*(t)$ and the critical value can thus be represented as the solution to (2.6) with

$$B^*_{\text{HW}}(t) = \sum_{j=1}^{p} P(\cup_{j=1}^{p}[|\boldsymbol{a}_j'\boldsymbol{u}| > td]) - \sum_{(i,j)\in\tau} P[(|\boldsymbol{a}_i'\boldsymbol{u}| > td) \cap (|\boldsymbol{a}_j'\boldsymbol{u}| > td)], \tag{2.8}$$

where τ is the minimal spanning tree of the nodes $\{1, 2, ..., p\}$ with branch weights of $\arccos|r_{ij}|$. The Hunter-Worsley bound $B^*_{\mathrm{HW}}(t)$ can be interpreted as the sum of the surface area of the disks $\{|\boldsymbol{a}'_j\boldsymbol{u}| > td\}, j = 1, 2, ..., p$, with the total surface area of the 2-way intersections subtracted from this sum, where the intersected disks are the ones whose pair indices appear in the minimal spanning tree.

2.1.3 The Hybrid Method

McCann and Edwards (2000) introduced three different hybrid methods that further sharpen the upper bounds of the tube and/or Hunter-Worsley methods. These three hybrid methods refine the upper bounds by exploring the relationship among the three covering shapes, tube, disk and "disk minus intersection", and selecting whichever shape provides a sharper surface area.

Specifically, the first hybrid method takes the minimum of the upper bounds in the tube and Hunter-Worsley methods and has the conservative bound

$$B^*_{\mathrm{H}}(t) = \min(B^*_{\mathrm{tube}}(t),\ B^*_{\mathrm{HW}}(t)).$$

Hence, this method is referred to as the minAll method. As a result, the minAll method always outperforms the tube and Hunter-Worsley methods.

The second hybrid method, also known as the Hunter-Worsley capped tubes method, enables the comparison between the tube and Hunter-Worsley methods on a smaller scale. It starts with the Bonferroni upper bound which is the sum of the surface area of p disks. Then the method searches each branch from the shortest to the longest along the minimal spanning tree and replaces the surface area of a disk with the surface area of a disk minus the intersection, or a tube between the two nodes of the branch provided that the tube gives a smaller surface area and neither node has been attached to more than one of the previously connected tubes. By this refinement of the upper bound, the Hunter-Worsley capped tubes method always outperforms the Hunter-Worsley method and also outperforms the tube method in

all simulations, although an analytical proof of its superiority over the tube method is not available.

The third hybrid method is referred to as the capped tubes method, as it is similar to the Hunter-Worsley capped tubes method, except that the surface area of a disk is only compared to the applied tube for possible replacement, instead of to the minimum of the surface area of a disk minus the intersection and the tube. Thus the Hunter-Worsley method is not included in this method. However, since the "disk minus intersection" expression brings in double integration to the final calculation of critical values, avoiding it in the capped tubes method greatly facilitates the computation. As a cost, the capped tubes method is not guaranteed to outperform either the tube method or the Hunter-Worsley method.

Of all the three hybrid methods, we chose to investigate the minAll method, as it performs nearly identically to the Hunter-Worsley capped tubes method but requires slightly less computational time. Hence in our context "the hybrid method" refers to the minAll method.

2.1.4 Scheffe's Method

Recall that Scheffe's method takes into account any possible values for the c_j's in the entire $\mathbb{R}^p$ space, and thus it has the upper bound

$$B^*_{SCH}(t) = 1.$$

From the simple form of $B^*_{SCH}(t)$, Scheffe's critical value in (2.2) and (2.6) can be readily derived as $(rF_{\alpha,r,\nu})^{\frac{1}{2}}$. Let $\boldsymbol{\gamma}$ be an arbitrary basis of the vector space generated by $C\boldsymbol{\beta}$. There is a unique $p \times r$ matrix $\boldsymbol{B}$ of full column rank such that $\boldsymbol{C\beta} = \boldsymbol{B\gamma}$. We rewrite $\boldsymbol{\gamma} = \boldsymbol{B}^+\boldsymbol{C\beta}$, where $\boldsymbol{B}^+$ is the generalized inverse matrix of $\boldsymbol{B}$. Then, $\widehat{\boldsymbol{\gamma}}$, defined as $\boldsymbol{B}^+\boldsymbol{C}\widehat{\boldsymbol{\beta}}$, has a multivariate normal distribution, $N(\boldsymbol{\gamma}, \boldsymbol{B}^+\boldsymbol{CVC'B'}^+)$.

Now for any set $F \subset \mathbb{R}^r$, let $S_{\boldsymbol{\gamma}}(F)$ be the event

$$\{\widehat{\boldsymbol{\gamma}} : |\boldsymbol{x'}\widehat{\boldsymbol{\gamma}} - \boldsymbol{x'\gamma}| \leqslant ds(\boldsymbol{x'B^{+}CVC'B'^{+}x})^{1/2}, \forall \boldsymbol{x} \in F\}.$$

Then the Scheffe confidence intervals can be obtained from the following well-known equation,

$$\begin{aligned} P(S_{\boldsymbol{\gamma}}(\mathbb{R}^r) &= P\{\widehat{\boldsymbol{\gamma}} : |\boldsymbol{x'}\widehat{\boldsymbol{\gamma}} - \boldsymbol{x'\gamma}| \leqslant ds(\boldsymbol{x'B^{+}CVC'B'^{+}x})^{1/2}, \forall \boldsymbol{x} \in \mathcal{R}^r\}) \\ &= P(F_{r,\nu} \leqslant d^2/r). \end{aligned} \tag{2.9}$$

Now since $Row(\boldsymbol{B})$ is a subset of $\mathbb{R}^r$, given d equal to $d_{SCH} \equiv (rF_{\alpha,r,\nu})^{\frac{1}{2}}$ we have

$$P(S_{\boldsymbol{\beta}}(Row(\boldsymbol{C}))) = P(S_{\boldsymbol{\gamma}}(Row(\boldsymbol{B}))) \geq P(S_{\boldsymbol{\gamma}}(\mathbb{R}^r)) = 1 - \alpha. \tag{2.10}$$

Notice, the conservativeness of the Scheffe critical value d_{SCH} results from relaxing the coefficient set $Row(\boldsymbol{B})$ to the entire $\mathbb{R}^r$. Another less conservative critical value is derived by seeking a similar equation to (2.9) with constraints on $\mathbb{R}^r$.

2.1.5 The Restricted Scheffe Method

The restricted Scheffe confidence intervals were introduced in the paper by Casella and Strawderman (1980). They replaced $\mathbb{R}^r$ with a constrained range, which is actually defined with respect to transformed variables of the foregoing chosen basis $\boldsymbol{\gamma}$, and found the exact expression for the coverage probability gain over (2.9), as a result of the constrained range. Specifically, let $\boldsymbol{P}$ denote the matrix of orthonormal eigenvectors of $\boldsymbol{(B^{+}CVC'B'^{+})^{-1}}$, and let $\boldsymbol{D}$ denote the diagonal matrix of its eigenvalues. So $\boldsymbol{(B^{+}CVC'B'^{+})^{-1}}$ is factorized as

$$\boldsymbol{(B^{+}CVC'B'^{+})^{-1} = PDP'}. \tag{2.11}$$

Then we define

$$\boldsymbol{Z} = \boldsymbol{BPD^{-1/2}}, \tag{2.12}$$
$$\boldsymbol{\eta} = \boldsymbol{D^{1/2}P'\gamma}.$$

So $\boldsymbol{Z\eta} = \boldsymbol{B\gamma} = \boldsymbol{C\beta}$. The least squares estimator $\widehat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$, has a multivariate normal distribution $N(\boldsymbol{\eta}, \sigma^2\boldsymbol{I})$. Henceforth, $\boldsymbol{Z}$ and $\boldsymbol{\eta}$ will refer to the transformed problem, and $\boldsymbol{B}$ and $\boldsymbol{\gamma}$ will refer to the original problem.

For the transformed problem, given any set $G \subset \mathbb{R}^r$, we denote the event

$$\{\widehat{\boldsymbol{\eta}} : |\boldsymbol{z'}\widehat{\boldsymbol{\eta}} - \boldsymbol{z'\eta}| \leqslant ds(\boldsymbol{z'z})^{1/2}, \forall z \in G\}$$

by $S_{\boldsymbol{\eta}}(G)$.

The classic result originally given by Casella and Strawderman (1980) and corrected by Sa and Edwards (1993), shows that

$$\begin{aligned} &P(S_{\boldsymbol{\eta}}(C_z) := \{\widehat{\boldsymbol{\eta}} : |\boldsymbol{z'}\widehat{\boldsymbol{\eta}} - \boldsymbol{z'\eta}| \leqslant ds(\boldsymbol{z'z})^{1/2}, \forall z \in C_z\}) \\ &= P(F_{r,\nu} \leqslant \frac{d^2}{r}) + \int_{d^2/r}^{d^2/rb^2} P(F_{s,r-s} \leqslant \frac{r-s}{s}\lambda_1^2(t)) f_{r,\nu}(t),\, dt \end{aligned} \tag{2.13}$$

where C_z has the form,

$$C_z = \{\boldsymbol{z} : \sum_{i=1}^{s} z_i^2 \geq q^2 \sum_{i=s+1}^{r} z_i^2\}, \tag{2.14}$$

q is a fixed constant and $\lambda_1(t) = \frac{d(t-d^2)^{1/2} - abt}{a^2t - d^2}$ with $a = q(1+q^2)^{-1/2}$, and $b = (1+q^2)^{-1/2}$.

Notice that in (2.13) the first term is exactly the same as $P(S_{\boldsymbol{\gamma}}(\mathbb{R}^r))$ and the second term is the amount of gain from the constraint C_z over Scheffe's bound. If $Row(Z) \subseteq C_z$ for some values s and q, a new critical value d_{RSCH} (RSCH stands for restricted Scheffe), can be obtained from equating (2.13) to $(1-\alpha)$. From (2.13),

it implies that d_{RSCH} is smaller than $d_{SCH} = (rF\alpha, r, \nu)^{1/2}$ and consequently the restricted Scheffe method constructs narrower confidence bands or intervals.

Casella and Strawderman (1980) pointed out that usually it was impossible to obtain the explicit constraint C_x in the original problem corresponding to C_z, unless $\boldsymbol{PD'P}$ is nearly orthogonal. Notwithstanding, an implicit relationship between C_z and C_x is available. Consider a compact expression of C_z

$$C_z = \{\boldsymbol{z} : \boldsymbol{z'Mz} \geq 0\}, \tag{2.15}$$

where $\boldsymbol{M}$ is a diagonal matrix with the first s entries 1's and the last $r - s$ entries $-q^2$. Then the pertaining constrained set C_x in the original problem can be written as

$$C_x = \{\boldsymbol{x} : \boldsymbol{x'PD^{-1}MP'x} \geq 0\}.$$

However, this relationship between C_z and C_x does not adequately address the question, i.e., given a C_x of interest, how to find a C_z in which C_x is contained. Instead, a more practical approach, the rectangle embedding approach, was proposed by Casella and Strawderman (1980).

2.2 Restricted Scheffe Method For a Finite Number of Comparisons

We want to apply the restricted Scheffe method, that was intended for a continuous regression function, to a finite number of comparisons and construct simultaneous confidence intervals. In adopting the restricted Scheffe method, a constraint cone C_z is required that contains the discrete set from the finite comparisons in the transformed problem.

2.2.1 Rectangle Embedding Approach

Casella and Strawderman (1980) introduced a rectangle embedding approach which assumes that the original set of interest has a hyperrectangle form,

$$R_x = \{l_1 \leqslant x_1 \leqslant u_1,\ l_2 \leqslant x_2 \leqslant u_2, ...,\ l_r \leqslant x_r \leqslant u_r\}, \tag{2.16}$$

where $l_1, ..., l_r$ and $u_1, ..., u_r$ are specified constants. This assumption not only has a wide applicability to different problems, but also facilitates searching for all the sets C_z of the form (2.14) that contain the transformed R_x. The conservative critical value of the rectangle embedding approach is the smallest critical value of all those obtained from (2.13) restricted to those cones C_z that cover the transformed R_x.

To find these covering cones, C_z, Casella and Strawderman (1980) showed that R_x is convex, and that it retains its convexity after the transformation in (2.12). Let R_z be the transformed R_x. Then the vertices of R_z are the images of the vertices of R_x. That is, given a vertex $\boldsymbol{x^\nu}$ of R_x, e.g. $\boldsymbol{x^\nu} = (l_1, l_2, l_3 \ldots, u_r)'$ or $(l_1, u_2, u_3 \ldots, u_{r-1}, l_r)'$, $\boldsymbol{z^\nu} = \boldsymbol{D^{-\frac{1}{2}} P' x^\nu}$ is a vertex of R_z. In implementing the search for C_z, Casella and Strawderman (1980) only deal with the vertices of R_z. In fact, for each i^{th} coordinate z_i, $1 \leqslant i \leqslant r$, they define

$$\begin{aligned} z_i^{\max} &= \max |z_i|, \\ z_i^{\min} &= \begin{cases} 0, & \text{if } \min z_i < 0 < \max z_i; \\ \min |z_i|, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.17}$$

Then it is shown that for any integer s between 1 and $r-1$, if the value

$$\frac{(\sum_{i=1}^{s} z_i^{\min})^2}{(\sum_{i=s+1}^{r} z_i^{\max})^2}$$

is nonzero, it is the largest q^2 such that the region of the form

$$\{\sum_{i=1}^{s} z_i^2 \geqslant q^2 \sum_{i=s+1}^{r} z_i^2\} \tag{2.18}$$

contains R_z. Otherwise, there exists no region of the form (2.18) that encloses R_z.

The rectangle embedding approach is intended for constructing confidence bands where the regions of interest are continuous. However, when it comes to a finite number of comparisons, the following example shows that the rectangle embedding approach can fail. That is because there exists no s and q such that any C_z could contain the rectangle envelope of the discrete set from the finite comparisons. (Note the rectangle envelope of a set X is the smallest hyperrectangle set that contains X.)

Suppose we have 5 parameters of interest, $\boldsymbol{\beta'} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, where the last parameter represents the control group. Assume there exists an estimator $\boldsymbol{b'}$ of $\boldsymbol{\beta'}$ with a multivariate normal distribution, $N(\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$, where $\boldsymbol{V}$ is a correlation matrix with a first-order autoregressive structure with $\phi = 0.5$. In this scenario, we are concerned with the linear combinations $\boldsymbol{C\beta}$, where the matrix

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Since $\boldsymbol{C}$ is of full row rank, we reparameterize by letting $\boldsymbol{\gamma} = \boldsymbol{B^{-1}C\beta}$, where $\boldsymbol{B}$ is a nonsingular matrix and without loss of generality, we choose

$$\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ -1 & 1 & 2 & 3 \\ -1 & -1 & 1 & 2 \\ -1 & -1 & -1 & 1 \end{pmatrix}. \tag{2.19}$$

Then the estimator $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ is $\boldsymbol{B^{-1}Cb}$, with $\widehat{\boldsymbol{\gamma}} \backsim N(\boldsymbol{B^{-1}C\beta}, \sigma^2\boldsymbol{B^{-1}CVC'B'^{-1}})$. Now, we seek a value d such that

$$P(S_\gamma(Row(\boldsymbol{B})) = \{\widehat{\boldsymbol{\gamma}} : |\boldsymbol{x'}\widehat{\boldsymbol{\gamma}} - \boldsymbol{x'\gamma})| \leqslant ds(\boldsymbol{x'B^{-1}CVC'B'^{-1}x})^{1/2},$$
$$\forall \boldsymbol{x} \in Row(\boldsymbol{B})\}) = 1 - \alpha.$$

We adopt the rectangle embedding approach, fit the rows of $\boldsymbol{B}$ into a rectangle and find the C_z that contains the transformed rectangle and gives the smallest critical value d from the restricted Scheffe equation (2.13).

Notice that,

$$R_x = \{-1 \leqslant x_1 \leqslant 1, \quad -1 \leqslant x_2 \leqslant 2, \quad -1 \leqslant x_3 \leqslant 3, \quad 1 \leqslant x_4 \leqslant 4\}$$

is the smallest rectangle that contains $Row(\boldsymbol{B})$. Let R_z denote the image of R_x after transformation as in (2.12). R_x passes on its convexity to R_z, so the vertices of R_z are exactly the images of the vertices of R_x. Hence, we obtain all the vertices of R_z by implementing the transformation on the vertices of R_x, which are left-multiplied by $\boldsymbol{D^{-\frac{1}{2}}P'}$, where $\boldsymbol{PDP'}$ is the eigenvalue decomposition of $\boldsymbol{B^{-1}CVC'B'^{-1}}$. Then we collect the maximum and minimum values of the vertices of R_z for each coordinate according to (2.17). It turns out $z_i^{min} = 0, \text{for } i = 1, 2, 3, 4$. That means that there exists no C_z that R_z could be embedded into. We obtain the same results when choosing different values for ϕ.

It is noteworthy that given most choices of $\boldsymbol{\gamma}$ and the resulting $\boldsymbol{B}$, the rectangle embedding approach fails to find any cones C_z that cover the transformed $Row(B)$. Even if there exist some values of $\boldsymbol{\gamma}$ and $\boldsymbol{B}$ where covering cones are available from the rectangle embedding approach, many of these cones turn out to be too large to give a competitive critical value d. However, it is the rectangle envelope that prohibits the derivation of an appropriate cone to obtain a sharp critical value. Consequently, we seek a new approach that eliminates the need for the rectangle envelope.

The drawback of the rectangle embedding approach arises particularly in some scenarios such as all pairwise comparisons, where the number of comparisons, p, is strictly larger than the number of linearly independent comparisons r. When we choose a basis $\boldsymbol{\gamma}$ of the vector space generated by $\boldsymbol{C\beta}$, $\boldsymbol{\gamma}$ can be any r linearly independent combinations out of the p comparisons and all the p comparisons are then expressed in terms of $\boldsymbol{\gamma}$ with a $p \times r$ coefficient matrix $\boldsymbol{B}$. As p grows, we can imagine that the rectangle envelope of $Row(\boldsymbol{B})$ will become larger and larger, and hence it becomes even more difficult to find a cone C_z that contains the transformed image of the expanding rectangle envelope.

From the forgoing example, it appears that the failure of the rectangle embedding approach is due to the overexpansion of $Row(\boldsymbol{B})$ into a rectangle. In the situation of continuous problems, the whole rectangle set in the original problem is of interest and the rectangle embedding approach allows us to find the exact covering C_z in the transformed problem. However, when we deal with a finite number of comparisons, the set of interest is discrete. If we still apply the rectangle embedding approach and throw the discrete set into a larger rectangle envelope, most of the time the rectangle envelope will overly expand the discrete set causing excessively large covering cones, and consequently rather conservative critical values. To resolve this overexpansion issue, we can instead find exact covering cones for the discrete set. Searching for the exact covering C_z's, we can utilize the discreteness of $Row(\boldsymbol{B})$, examine the transformed image of each element in $Row(\boldsymbol{B})$, and find the exact smallest C_z that contains all those finite elements. This leads to a new approach, the minimal cone approach.

2.2.2 Minimal Cone Approach

In the minimal cone approach, we look for the exact minimal cone C_z in the sense that of all the C_z's that cover the transformed $Row(\boldsymbol{B})$, the minimal cone is the one that leads to the smallest critical value. This approach applies whenever a finite number of comparisons are considered. The cone C_z is determined by two factors, s

and q^2. The minimal C_z with the least critical value could result from any possible value of s. Thus for all $s = 1, 2, ..., r-1$, we compute the largest q^2 from each element in the transformed $Row(\boldsymbol{B})$, which would give the smallest cone in which that element is enclosed. Then pick the smallest of all those q^2 as the final q^2 such that its resulting C_z contains all the elements in the transformed $Row(\boldsymbol{B})$. From (2.13), the critical value with respect to each constraint C_z with varying s can be obtained. Of all those locally minimal constraint cones, the one that corresponds to the smallest critical value is the optimal C_z.

Now we illustrate the minimal cone approach with the previous example. First, we transform the matrix B in the original problem into

$$\boldsymbol{Z} = \boldsymbol{BPD}^{-\frac{1}{2}} = \begin{pmatrix} 1.1147 & -0.6251 & 0.1001 & 0.4814 \\ 0.6393 & -1.0852 & -0.3963 & -0.0806 \\ 0.1871 & -1.1049 & 0.4935 & -0.0232 \\ -0.1873 & -0.7302 & -0.0134 & 0.6569 \end{pmatrix},$$

where $r = \text{rank}(\boldsymbol{B}) = 4$.

For each $s = 1, 2, 3$, and each row of $\boldsymbol{Z}$, we compute

$$q^2_{m,s} := \frac{\sum_{i=1}^{s} z^2_{mi}}{\sum_{j=s+1}^{r} z^2_{mj}} \qquad m = 1, 2, ..., p,$$

where z_{mk} is the $(m,k)^{\text{th}}$ entry of $\boldsymbol{Z}$. Then $\min_{1 \leqslant m \leqslant p} q^2_{m,s}$ is the largest q^2 with respect to s, such that all the rows of $\boldsymbol{Z}$ are a subset of $C_z(s, q^2)$, which is the cone defined in (2.18) with parameters s and q^2. With degrees of freedom of 60, we can solve for the critical values from (2.13) by setting

$$P(S_{\boldsymbol{\eta}}(C_z(s, \min_{1 \leqslant m \leqslant p} q^2_{m,s}))) = 1 - \alpha, \quad s = 1, 2, ..., r-1.$$

The minimal cone is the C_z with the exact $(s, \min_{1\leqslant m\leqslant p} q^2_{m,s})$ that gives the smallest critical value of all the 3 sets of $(s, \min_{1\leqslant m\leqslant p} q^2_{m,s})$. The final results are shown in Table 1.

s	1	2	3
q^2	0.0239	1.3165	1.3175
d	3.1758	3.0819	3.1556

Table 1: Cones $C_z(s, q^2)$ and critical values d

Apparently, $C_z(2, 1.3165)$ is the optimal C_z and the associated critical value $d = 3.0819 < d_{SCH} = (rF_{\alpha,r,\nu})^{\frac{1}{2}} = 3.178185$.

With a finite number of comparisons, the minimal cone approach is capable of finding the optimal cone C_z that contains $Row(\boldsymbol{Z})$ and gives the smallest critical value, provided that such a cone exists. The overexpansion issue associated with the rectangle envelopes that results in no covering cones or a large covering cone, is completely resolved by the minimal cone approach. As for moderate-scale multiple comparisons problems, it is clearly computationally feasible to utilize the minimal cone approach.

To summarize, we give a general algorithm of obtaining critical values via the restricted Scheffe method utilizing the minimal cone approach (the minimal restricted Scheffe method). In the context of general multiple comparisons that we defined in Chapter 2, we want to construct conservative $100 \times (1-\alpha)\%$ simultaneous confidence intervals for $\boldsymbol{C\beta}$ of the form

$$\boldsymbol{c_j'b} \pm ds(\boldsymbol{c_j'Vc_j})^{1/2}, \qquad j = 1, 2, ..., p.$$

In order to find such d,

1. Choose a basis $\boldsymbol{\gamma}$ of the vector space generated by $\boldsymbol{C\beta}$. Then $\boldsymbol{\gamma}$ has the dimension r, which is the rank of $\boldsymbol{C}$.

2. Express $\boldsymbol{C\beta}$ as linear combinations of $\boldsymbol{\gamma}$ and obtain the coefficient matrix $\boldsymbol{B}$, i.e. $\boldsymbol{C\beta} = \boldsymbol{B\gamma}$.

3. Calculate the matrix $\boldsymbol{Z}$ in the transformed problem. That is, $\boldsymbol{Z} = \boldsymbol{BPD}^{-\frac{1}{2}}$, where $\boldsymbol{PDP'}$ is the eigenvalue decomposition of $(\boldsymbol{B}^{+}\boldsymbol{CVC'B'}^{+})^{-1}$ and V is the known correlation matrix of $\boldsymbol{b}$.

4. Compute

$$q_{m,s}^2 := \frac{\sum_{i=1}^{s} z_{mi}^2}{\sum_{j=s+1}^{r} z_{mj}^2} \qquad m = 1, 2, ..., p, \;\; s = 1, 2, ..., r-1,$$

   where z_{mk} is the $(m,k)^{\text{th}}$ entry of $\boldsymbol{Z}$.

5. Find $q_s^2 = \min_{1 \leqslant m \leqslant p} q_{m,s}^2$ for each $s = 1, 2, ..., r-1$.

6. Calculate the critical values d_s from equating (2.13) to $(1-\alpha)$, where α is a predetermined error rate.

7. $\min_{1 \leqslant s \leqslant r-1} d_s$ is the critical value given by the minimal restricted Scheffe method.

Note that in the above algorithm, there are two factors that are not uniquely determined, the basis $\boldsymbol{\gamma}$ and the eigendecomposition of $(\boldsymbol{B}^{+}\boldsymbol{CVC'B'}^{+})^{-1}$. Next we are going to explore how different choices in these two factors would affect the critical values of the minimal restricted Scheffe method.

2.2.3 Eigendecomposition Effect

In the restricted Scheffe method, the restriction is defined with respect to the transformed problem. To make the transformation, we have the eigendecomposition of the inverse matrix of $(\boldsymbol{B}^{+}\boldsymbol{CVC'B'}^{+})$. So $(\boldsymbol{B}^{+}\boldsymbol{CVC'B'}^{+})^{-1}$ is factorized as

$$(\boldsymbol{B}^{+}\boldsymbol{CVC'B'}^{+})^{-1} = \boldsymbol{PDP'}, \tag{2.20}$$

Notice that in the eigendecomposition (2.20), if all the eigenvalues are distinct, $\boldsymbol{D}$ is unique subject to the order of eigenvalues on the diagonal line. Then with a fixed order of eigenvalues in $\boldsymbol{D}$, $\boldsymbol{P}$ is unique subject to the signs of columns. That is, $\boldsymbol{P}$ can be replaced by $\boldsymbol{PI^*}$, where

$$\boldsymbol{I^*} = \begin{bmatrix} \pm 1 & & & \\ & \pm 1 & & \\ & & \ddots & \\ & & & \pm 1 \end{bmatrix}.$$

However, since $\boldsymbol{I^*}$ is commutative with any diagonal matrix,

$$\boldsymbol{Z^* = BPI^* D^{-\frac{1}{2}} = BPD^{-1/2} I^* = ZI^*},$$

which results in at most opposite signs to the columns of $\boldsymbol{Z}$. In the minimal cone approach, we only utilize the absolute values of the elements of $\boldsymbol{Z}$ and thus the choice of variant forms of $\boldsymbol{P}$ does not affect the ultimate minimal cone.

Now let us investigate the effect of variant forms of $\boldsymbol{D}$ on the minimal cone. Assume $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are two diagonal matrices of the eigenvalues of $\boldsymbol{(B^+CVC'B'^+)^{-1}}$ in different orders on the diagonal line. Then there exists a permutation matrix $\boldsymbol{Q}$ such that $\boldsymbol{QQ' = I}$ and

$$\boldsymbol{\tilde{D} = QDQ'}.$$

Thus, the eigendecomposition can be expressed as

$$\boldsymbol{PDP' = PQ'QDQ'QP' = (PQ')\tilde{D}(PQ')' = \tilde{P}\tilde{D}\tilde{P}'},$$

where $\boldsymbol{\tilde{P} = PQ'}$. Using these two eigendecomposition representations to define the transformed problem (2.12), we have $\boldsymbol{Z = BPD^{-1/2}}$ and

$$\boldsymbol{\tilde{Z} = B\tilde{P}\tilde{D}^{-1/2}}$$

$$= \boldsymbol{BPQ'QD^{-1/2}Q'}, \quad \text{since } \tilde{\boldsymbol{D}}^{\boldsymbol{-1/2}} = \boldsymbol{QD^{-1/2}Q'},$$
$$= \boldsymbol{BPD^{-1/2}Q'}. \tag{2.21}$$

Note that $\tilde{\boldsymbol{Z}}$ is equal to $\boldsymbol{Z}$ right multiplied by $\boldsymbol{Q'}$ which is also a permutation matrix. That is to say, $\tilde{\boldsymbol{Z}}$ is composed of the shuffled columns of $\boldsymbol{Z}$. From the minimal cone approach, we know that a switch of columns in $\boldsymbol{Z}$ would almost certainly lead to a different resulting minimal cone and hence a different critical value. Therefore, different orders of eigenvalues in $\boldsymbol{D}$ usually result in distinct critical values from the minimal cone approach.

In most math/stat software, the eigenvalue decomposition routine by default gives the diagonal matrix with eigenvalues placed in decreasing order from top to bottom. However, simulation results show that among all the diagonal matrices with the eigenvalues in permuted orders, this default diagonal matrix is not always the one that yields the smallest critical value via the minimal cone approach. In fact, no evidence shows that the minimal critical value is related to a particular order of eigenvalues in the diagonal matrix. Since there are $(r!)$ permutations of the eigenvalues, where r is the dimension of $\boldsymbol{(B^{+}CVC'B'^{+})^{-1}}$, we apply the minimal cone approach to all $(r!)$ possible $\tilde{\boldsymbol{Z}}$ in (2.21), obtain corresponding minimal cones and critical values, and select the smallest critical value. In practice, as $r!$ increases rapidly with r, an allowed maximum number of permutations is set to control the computing time. If $r!$ exceeds this maximum threshold, we randomly select only the threshold amount of eigenvalue permutations into comparison. Since this permuting technique uniformly improves the critical values obtained from the previous algorithm, we include it as part of the algorithm, by adding

8. Go to step 3 and let $\boldsymbol{D}$ choose a different permuted order of the eigenvalues on the diagonal line. Do not proceed until the maximum number of permutations is reached.

9. Select the minimum of all the critical values obtained from the permuted orders of the eigenvalues in $\boldsymbol{D}$.

It is noteworthy that $\boldsymbol{D}$ does not necessarily have distinct eigenvalues. Hence, the eigenvectors in $\boldsymbol{P}$ associated with duplicate eigenvalues are not unique. Lack of enough scenarios that produce duplicate eigenvalues, it remains unclear how the choices of the associated eigenvectors, jointly with the effect of eigenvalue permutations, change the critical value of the minimal restricted Scheffe method.

2.2.4 Effect of Basis Change

We have presented the minimal cone approach based on $\boldsymbol{C\beta}$'s representation with respect to an arbitrary basis $\boldsymbol{\gamma}$. A question arises: Would a different choice of the basis for $\boldsymbol{C\beta}$ result in a different C_z and consequently a different critical value? To answer this question, we investigate elementary linear operations on the basis, including the switching operation, the multiplication operation and the addition operation, as any linear transformation is a composite of these three elementary operations.

Switching operation

A switching operation interchanges two rows and is represented by left multiplication with an elementary matrix

$$\boldsymbol{Q_1} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}.$$

Let $\widetilde{\boldsymbol{\gamma}}$ be a basis with two components of $\boldsymbol{\gamma}$ switched. Then,

$$\boldsymbol{C\beta} = \boldsymbol{B\gamma} = \widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\gamma}}, \qquad \textit{where } \widetilde{\boldsymbol{B}} = \boldsymbol{BQ_1} \textit{ and } \widetilde{\boldsymbol{\gamma}} = \boldsymbol{Q_1\gamma}.$$

With $\boldsymbol{\gamma}$, the diagonalized coefficient matrix $\boldsymbol{Z} = \boldsymbol{BPD}^{-\frac{1}{2}}$, where

$$(\boldsymbol{B}^{+}\boldsymbol{CVC}'\boldsymbol{B}'^{+})^{-1} = \boldsymbol{PDP}'$$

is the eigenvalue decomposition.

Now, $\widetilde{\boldsymbol{\gamma}}$ follows $N(\boldsymbol{Q_1\gamma}, \boldsymbol{Q_1B}^{+}\boldsymbol{CVC}'\boldsymbol{B}'^{+}\boldsymbol{Q}_1')$ and

$$\begin{aligned}(\boldsymbol{Q_1B}^{+}\boldsymbol{CVC}'\boldsymbol{B}'^{+}\boldsymbol{Q}_1')^{-1} &= \boldsymbol{Q_1}(\boldsymbol{B}^{+}\boldsymbol{CVC}'\boldsymbol{B}'^{+})^{-1}\boldsymbol{Q}_1' \\ &= \boldsymbol{Q_1PDP}'\boldsymbol{Q}_1' \\ &= (\boldsymbol{Q_1P})\boldsymbol{D}(\boldsymbol{Q_1P})'.\end{aligned}$$

Since $\boldsymbol{Q}_1P$ is an orthonormal matrix, the eigenvalue decomposition is

$$(\boldsymbol{Q_1B}^{+}\boldsymbol{CVC}'\boldsymbol{B}'^{+}\boldsymbol{Q}_1')^{-1} = \widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{P}}', \qquad \textit{where } \widetilde{\boldsymbol{P}} = \boldsymbol{Q_1P} \textit{ and } \widetilde{\boldsymbol{D}} = \boldsymbol{D}.$$

Then

$$\begin{aligned}\widetilde{\boldsymbol{Z}} = \widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}} &= \boldsymbol{BQ_1Q_1PD}^{-\frac{1}{2}} \\ &= \boldsymbol{BPD}^{-\frac{1}{2}}, \qquad \textit{since } \boldsymbol{Q_1Q_1} = \boldsymbol{I}, \\ &= \boldsymbol{Z}.\end{aligned}$$

Therefore, the switching operation on the basis actually does not make any change in the transformed problem. That is to say that the minimal restricted Scheffe method is invariant under the switching operation.

Multiplication operation and addition operation

The multiplication operation multiplies a row by a nonzero scalar and the addition operation adds a row multiplied by a scalar to another row. These two operations are represented by left multiplication with the following two elementary matrices,

respectively.

$$\boldsymbol{Q_2} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \ \boldsymbol{Q_3} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & a & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} (a \neq 0).$$

Unfortunately, the previous invariance property does not extend to the multiplication or addition operations. Because there is no explicit form of eigenvalue decomposition for $(\boldsymbol{Q_2 B^+ C V C' B'^+ Q_2'})^{-1}$ or $(\boldsymbol{Q_3 B^+ C V C' B'^+ Q_3'})^{-1}$, it remains as an open question how exactly the basis change from $\boldsymbol{Q}_2$ or $\boldsymbol{Q}_3$ affects the minimal restricted Scheffe critical values.

In addition to the three elementary operations, we conducted an extensive investigation into arbitrary nonsingular matrices that cause basis change. However, there is no sufficient evidence for a perceptible pattern of an optimal basis that gives the smallest critical value from the minimal restricted Scheffe method. In practice, it is impossible to search an infinite number of possible bases for an optimal basis. On the other hand, the facts suggest that the change of basis has an important impact on the minimal cone critical values. Hence, we introduce adopting the minimal restricted Scheffe method with a certain number of varying bases and pick the minimum of the critical values from these bases. Consequently, the algorithm is revised by including

10. Go to step 1 and modify $\boldsymbol{\gamma}$ by left-multiplying $\boldsymbol{\gamma}$ by a randomly generated nonsingular matrix. Do not proceed until the allowed maximum number of varying bases is reached.

11. Select the minimum critical values obtained from all the considered bases.

2.3 Simulation: Comparing simultaneous confidence intervals

In this section, we present a simulation study to compare simultaneous confidence intervals constructed by the mininal restricted Scheffe method with other appropriate single-step methods. The single-step competitors include the hybrid method, Sidak's method, and Scheffe's method. Since the single-step methods we consider have all been proven to be conservative and utilize a common standard error, the performance is only determined by the critical values provided by the different methods. Hence we evaluate the performance of these confidence intervals in terms of relative efficiency. The relative efficiency of the minimal restricted Scheffe method to another single-step method is defined to be the squared ratio of their corresponding critical values. This ratio approximates the ratio of sample sizes needed by the two methods to have intervals of equal length. Hence, substracting the relative efficiency from 1 approximately represents the percentage of sample-size savings of the minimal restricted Scheffe method relative to the other method for achieving equal-length-intervals.

Specifically, our investigation includes the following factor settings for the general multiple comparison model:

1. $\boldsymbol{V}$ matrix.

   $\boldsymbol{V}$ is chosen to be a first-order autoregressive correlation matrix with parameter ϕ, i.e. $\boldsymbol{V} = [v_{ij}]$ for $v_{ij} = \phi^{|i-j|}, \phi = 0, 0.5, 0.9$.

2. $\boldsymbol{C}$ matrix. $\boldsymbol{C}$ is chosen to obtain the following sets of comparisons:

   (a) All pairwise comparisons of $\beta_1, .., \beta_k$(MCA), $p = \frac{k(k-1)}{2}$.

   (b) Multiple comparisons with a control (MCC), $p = k - 1$.

   (c) All of the parameters individually (SMM, the studentized maximum modulus), $p = k$.

3. Number of parameters k. We consider $k = 4, 10$.

4. Error degrees of freedom, ν. We consider $\nu = 5, 15, 60, 100$.

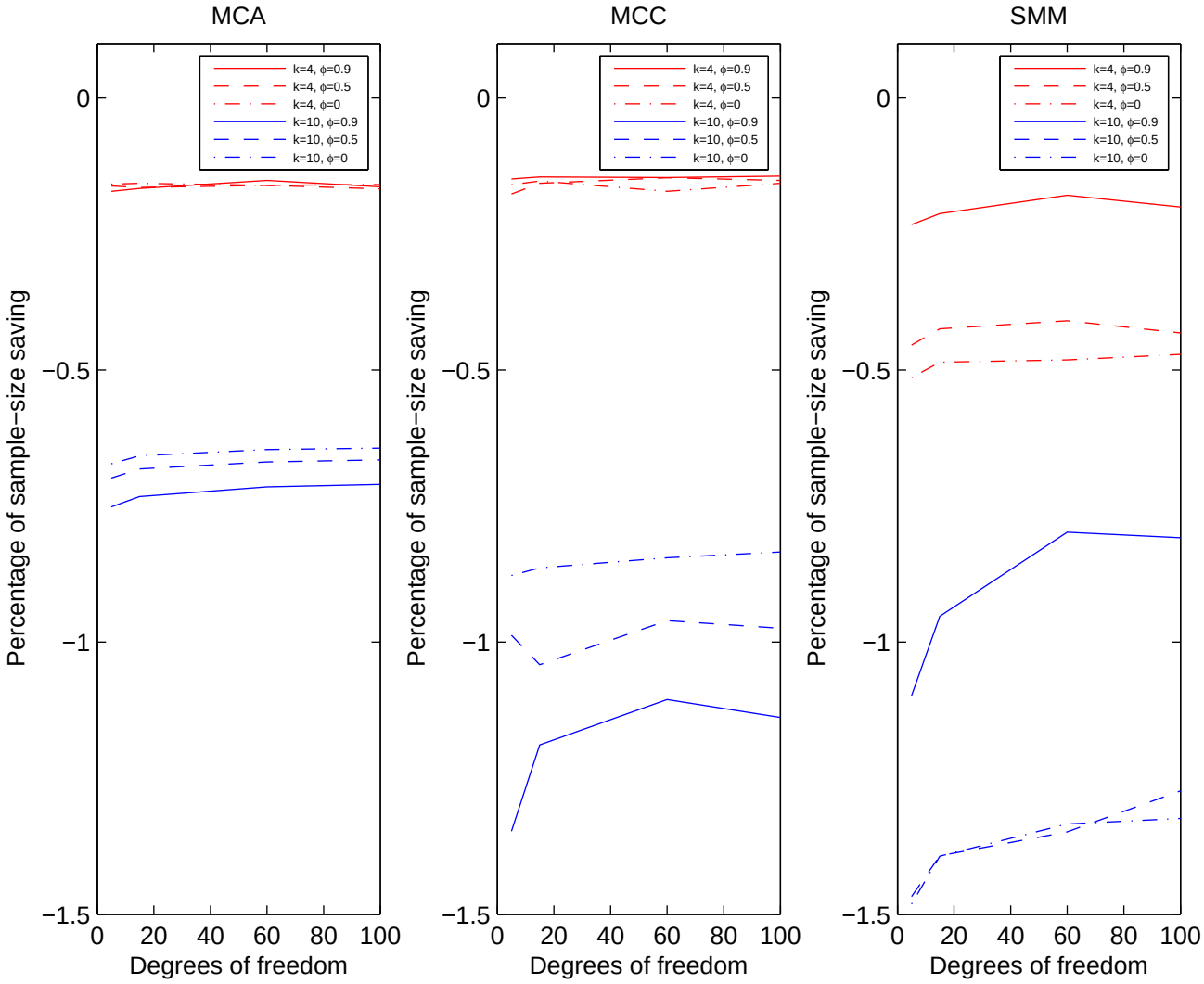
5. Nominal FWE level, α. We consider $\alpha = 0.05, 0.1$.

From the simulation results, the percentages of sample-size savings of the minimal restricted Scheffe method share very similar patterns when $\alpha = 0.05$ and 0.1. Hence, we present results only for $\alpha = 0.05$. In Chapter 2, we discussed the two effects in our algorithm that influence the minimal restricted Scheffe critical values, the eigenvalue permutation and the basis change. In simulations, an eigenvalue permutation incurs almost the same cost in computation as that of one basis change. Empirical results show that allowing for more than 2 permutation iterations does not necessarily increase the chance of obtaining a smaller critical value than when we only permute the eigenvalue order twice for a given basis. Hence, we fix the number of permutations at two and allocate all the remaining computational resources for varying bases. The minimal restricted Scheffe critical values and computing time in this simulation study are obtained based on 100 basis variations. To generate each basis, we collect the first r linearly independent vectors in $\boldsymbol{C\beta}$ and transform them into a randomly selected basis via multiplying them by a non-singular matrix of which the elements are independently and identically drawn from the standard normal distribution. Similar results were observed when we used uniform distributions. It is noteworthy that although less conservative minimal restricted Scheffe critical values can be attained with an additional investment in basis iterations, they do not fundamentally change the main pattern of performance in the following simulation results.

2.3.1 The minimal restricted Scheffe method and the hybrid method

Figure 1 provides the percentages of sample-size savings of the minimal restricted Scheffe method in all the considered settings relative to the hybrid method. Note that all the percentages are below zero, which means that the minimal restricted Scheffe method is outperformed by the hybrid method. Moreover, a negative saving percentage tells the percentage of extra sample sizes that the minimal restricted Scheffe method has to pay for achieving equal-length intervals to the hybrid method.

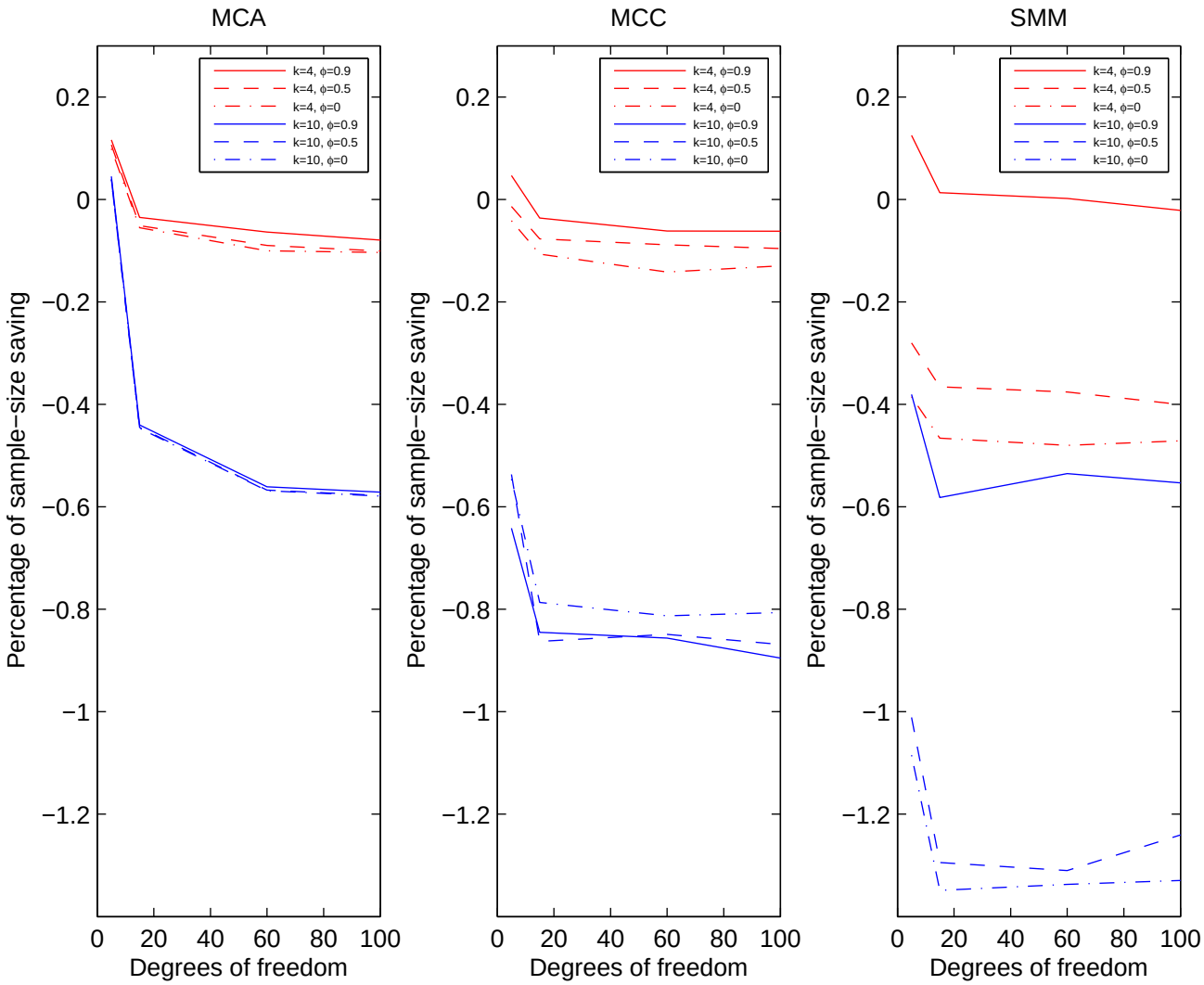
Figure 1: Sample-size saving percentage of the minimal restricted Scheffe method to the hybrid method.



In Figure 1, the sample-size saving percentage of the minimal restricted Scheffe method relative to the hybrid method drops rapidly as the number of parameters, k, grows. That is to say, the performance of the minimal restricted Scheffe method relative to the hybrid method becomes poorer with the increase of k. The values of the correlation parameter, ϕ, do not seem to make as much difference in the percentage of sample-size savings in MCA and MCC, as in SMM, where the saving percentage increases substantially as ϕ rises. When k is small, the degrees of freedom do not appear to play a role in the sample-size saving percentage of the minimal restricted

Scheffe method relative to the hybrid method in any of the three scenarios. However, when k is large, a slight increasing trend in saving percentages is detected in some cases of MCC and SMM, as the degrees of freedom grow.

Figure 2: Sample-size saving percentage of the minimal restricted Scheffe method to Sidak's method.



2.3.2 The minimal restricted Scheffe method and Sidak's method

Figure 2 plots the percentages of sample-size savings of the minimal restricted Scheffe method in all the considered settings relative to Sidak's method. Similar to the results relative to the hybrid method, the sample-size saving percentage of the minimal restricted Scheffe method relative to Sidak's method decreases as the number

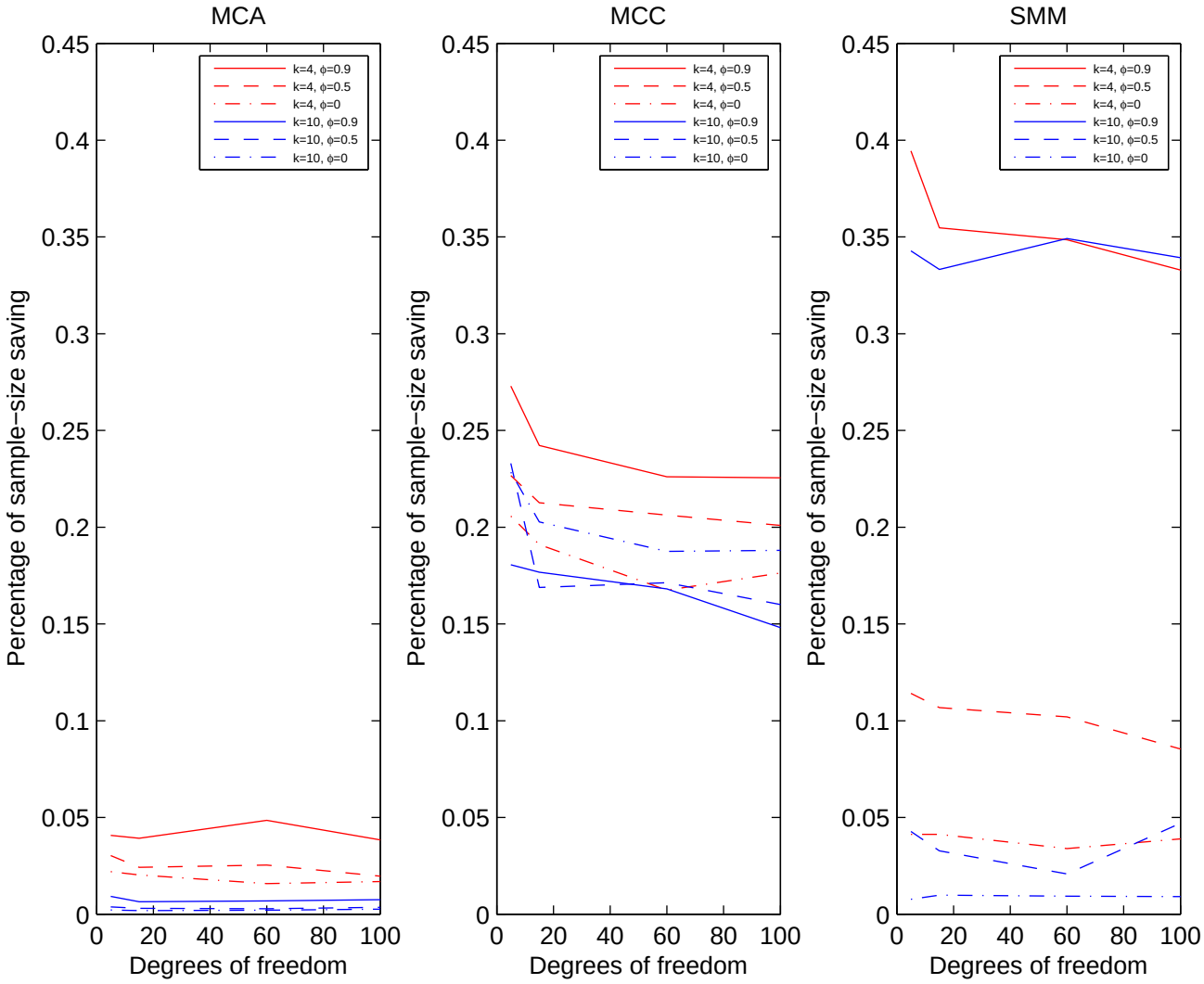
of parameters, k, increases. Degrees of freedom contribute to the variation in the percentages of sample-size savings: lower degrees of freedom lead to a higher saving percentage relative to Sidak's method. Particularly, the minimal restricted Scheffe method can outperform Sidak's method and reaches up to 10% savings in sample size relative to Sidak's method in MCA, when the degrees of freedom are low. With the same low degrees of freedom, similar patterns of the sample-size savings in the minimal restricted Scheffe method relative to Sidak's method are observed for some SMM and MCC cases, given a small number of parameters and a high correlation parameter.

Note that in SMM, the percentage of sample-size savings relative to Sidak's method improves as the correlation parameter ϕ increases. This is because, in SMM, the covariance matrix of $\boldsymbol{C\beta}$ is reduced to $\boldsymbol{V}$. A high value of ϕ directly reflects the strong correlation between the comparisons, and consequently forms a tight cone constraint which leads to a small minimal restricted Scheffe critical value, whereas Sidak's method makes no use of the correlation information. A similar monotone trend also exists in MCA and MCC when k is small, but it is not as clear as in SMM, since $\boldsymbol{C}$, in addition to $\boldsymbol{V}$, also plays a role in the correlation values of the comparisons. In addition, when k is large, the percentage of sample-size savings is almost constant in MCA, regardless of the correlation values. This is attributed to the increased number of comparisons in MCA as k rises. As a result, the minimal restricted Scheffe critical values closely approach Scheffe's critical values and the variation caused by ϕ in the saving percentages relative to Sidak's method becomes minor.

2.3.3 The minimal restricted Scheffe method and Scheffe's method

Figure 3 gives the percentages of sample-size savings of the minimal restricted Scheffe method in all the considered settings relative to Scheffe's method. All the saving percentages are positive, as the minimal restricted Scheffe method uniformly improves on Scheffe's method for a finite number of comparisons. The effects of the

Figure 3: Sample-size saving percentage of the minimal restricted Scheffe method to Scheffe's method.



numbers of parameters and the correlation parameter on the sample-size saving percentages of the minimal restricted Scheffe method relative to Scheffe's method share a similar pattern with those relative to Sidak's method. The highest saving percentage is achieved in SMM when the correlation parameter $\phi = 0.9$. This results from the effect of extreme correlations between comparisons. In general, the percentage of sample-size savings relative to Scheffe's method decreases as the number of comparisons increases. Hence, in MCC, the sample-size saving percentages are relatively high. With regards to degrees of freedom, the percentages of sample-size savings in

MCA appear insensitive to the change in degrees of freedom, whereas in MCC and SMM, the saving percentages show a slightly decreasing trend with the increase in degrees of freedom.

2.3.4 Computing time

Figures 4 and 5 plot the execution time (in seconds) of computing the minimal restricted Scheffe critical values and the hybrid critical values in MATLAB R2012b. The computing time for the minimal restricted Scheffe method is obtained based on 100 basis variations. In Figures 4 and 5, the computing time of both the minimal restricted Scheffe critical values and the hybrid critical values increases as the number of parameters, k, increases. In fact, the computing time of the minimal restricted Scheffe critical values is associated with r, as the minimal cone approach examines $(r-1)$ locally minimal cones to pick the optimal one; The computing time of the hybrid critical values is associated with the number of comparisons, as the Hunter-Worsley method, which takes most of the computing time in the hybrid method, makes use of pairwise correlations between comparisons. The computing time for the minimal restricted Scheffe method and the hybrid method is comparable in MCA, when $k = 4$. However, the minimal restricted Scheffe method is much more computationally costly than the hybrid method in other cases. The computing time of Sidak's critical values and Scheffe's critical values are not presented, as it is in the magnitude of 10^{-4} seconds, almost negligible relative to the minimal restricted Scheffe method and the hybrid method. To compute the minimal restricted Scheffe critical values, we choose 100 basis variations, because empirically the effect of basis change is significant when the number of basis variations is below 100 and becomes subtle as the number of basis change goes beyond 100. In practice, researchers can adjust the number of basis iterations in the minimal restricted Scheffe method depending on the computational resources.

Figure 4: Computing time for the minimal restricted Scheffe method.

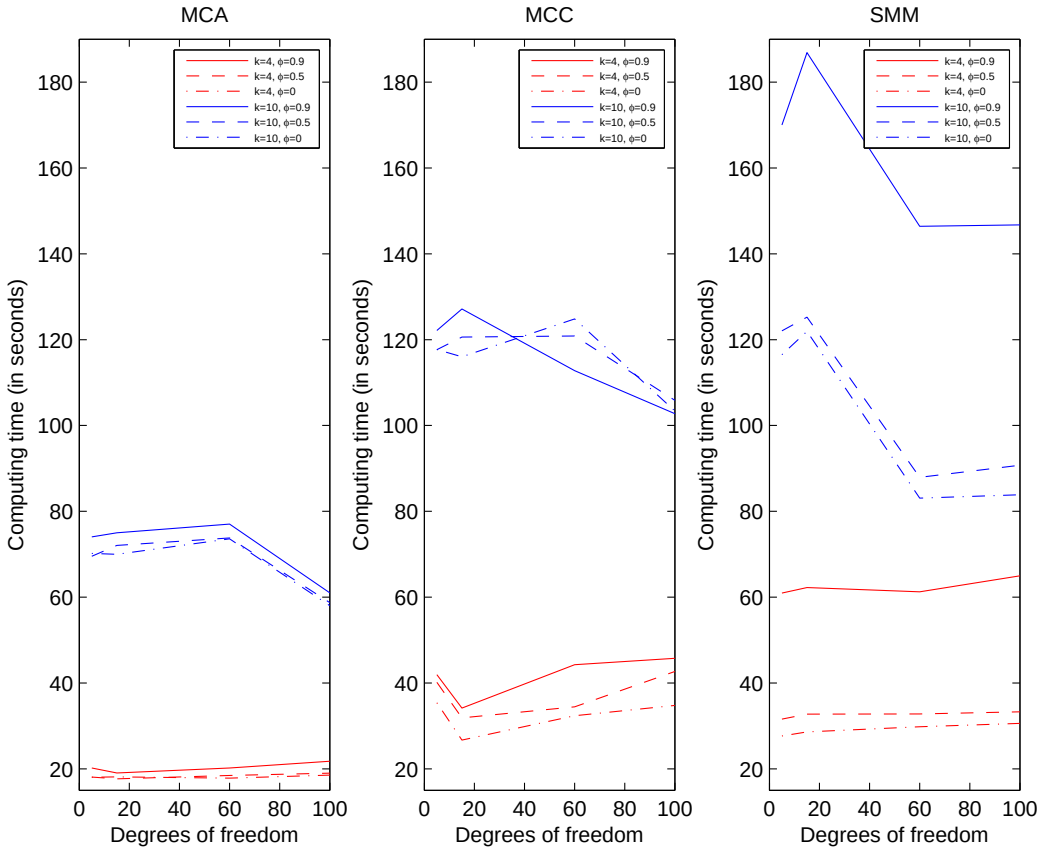
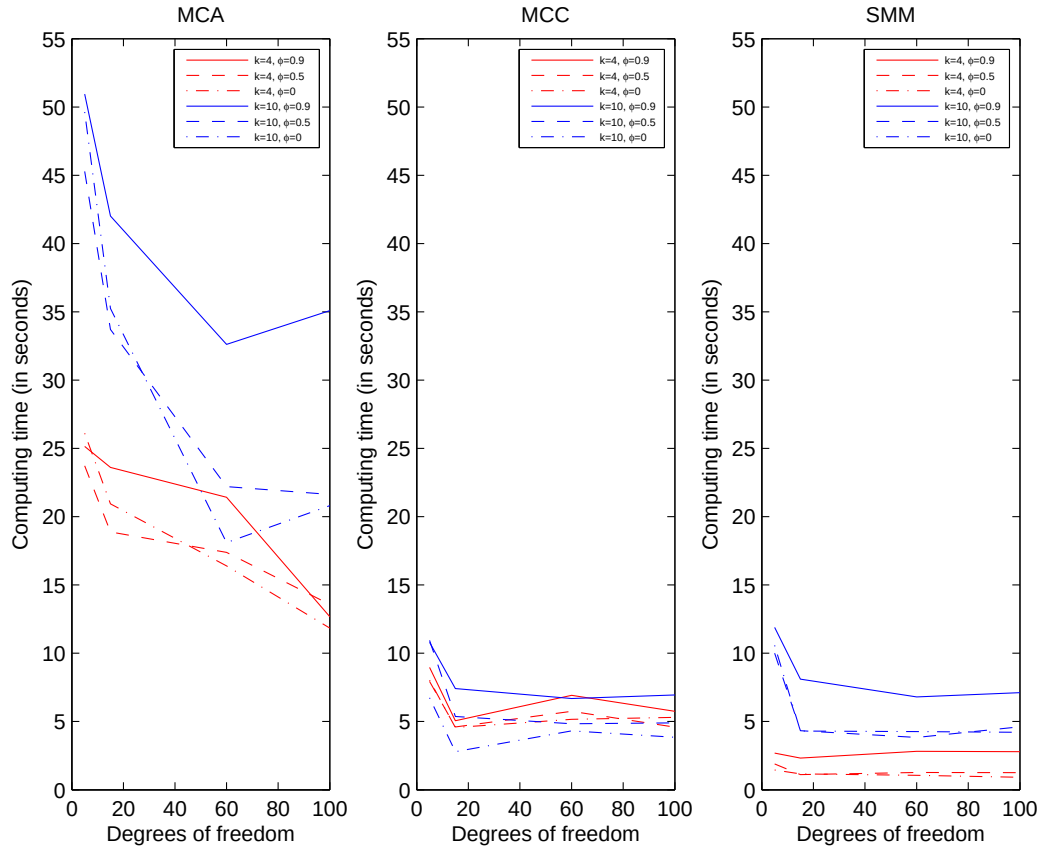


Figure 5: Computing time for the hybrid method.



3 Forced Sequential Methods

3.1 Sequentially Rejective Framework

In simultaneous testing, stepwise methods are often employed to obtain more power than single-step methods when simultaneous confidence intervals are not of interest. Particularly, single-step methods can be adapted into stepdown methods which are methods that sequentially reject the most significant hypotheses. Romano and Wolf (2005) constructed a general stepdown procedure from single-step methods with both exact and random critical values. We restrict our focus to exact critical values. Suppose there is only one common critical value for all the hypotheses in the single-step methods and, without loss of generality, that a large test statistic indicates evidence against the null hypothesis. Stepdown procedures begin by testing the joint null hypotheses that all hypotheses are true. If the maximum test statistic is a large value, reject its corresponding hypothesis; otherwise, stop. After a hypothesis is rejected, treat the remaining hypotheses as a new family by rejecting for large values of the maximum of the remaining test statistics, and so on. At each step, the critical value that the largest test statistic is compared with comes from the single-step test on the intersection of the hypothesis that remain. This procedure was further generalized in the paper by Romano and Wolf (2010) by relaxing the assumption of having identical critical values for all hypotheses. In order to conform to the monotonicity condition on the critical values, Romano and Wolf (2005) adjusted the critical value to be the maximum of the critical values from single-step tests on any set that is a subset of the current remaining hypotheses. Goeman and Solari (2010) generalized the work of Romano and Wolf (2005, 2010) to logically related hypotheses and showed that their proposed sequentially rejective framework is a more versatile approach to multiple comparisons. The sequentially rejective framework was shown to encompass the stepdown framework of Romano and Wolf (2005) as a special case,

as well as a wide range of other stepwise methods. More importantly, Goeman and Solari (2010) gave sequential rejection conditions on which the sequentially rejective framework strongly controls FWE.

3.1.1 Sequential Rejection Conditions

Goeman and Solari (2010) supposed that a statistical model is structured on a parameter space Ω. Each $\omega \in \Omega$ is a parameter configuration that fully determines a probability measure P_ω on a common outcome space $\mathbb{S}$. Let $\mathcal{H}$ be a collection of all null hypotheses. Under a probability measure P_ω, define $\mathcal{T}(\omega)$ and $\mathcal{F}(\omega)$ as the collection of all true null hypotheses and the collection of all false null hypotheses, respectively. For each $H \in \mathcal{H}$, suppose $M(H)$ is a subset of Ω such that $M(H) = \{\omega \in \Omega \mid H \in \mathcal{T}(\omega)\}$. Then given $\mathcal{T}(\omega) \supseteq \mathcal{I}$, where $\mathcal{I}$ is a subset of $\mathcal{H}$, $\omega \in \bigcap_{H\in\mathcal{I}} M(H)$.

For illustration, suppose under treatment i, a random sample $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ is taken from a normal distribution $N(\mu_i, \sigma^2),\ i = 1, 2, 3$. In this statistical model, the parameter space Ω of $(\mu_1, \mu_2, \mu_3, \sigma)$ is set to be $\mathbb{R}^3 \times \mathbb{R}^+$. Hence, $\omega = (0, 3, 3, 1) \in \Omega$ is a specific parameter configuration and defines a probability measure P_ω. Assume all pairwise comparisons among the treatment means are of interest, i.e., $\mathcal{H} = \{H_1, H_2, H_3\}$ and

$$H_1 : \mu_1 = \mu_2;\ H_2 : \mu_2 = \mu_3;\ H_3 : \mu_1 = \mu_3.$$

Then given $\omega = (0, 3, 3, 1)$, it follows that $\mathcal{T}(\omega) = \{H_2\}$, $\mathcal{F}(\omega) = \{H_1, H_3\}$ and $\omega \in M(H_2) = \{(\mu_1, \mu_2, \mu_3, \sigma) \in \Omega \mid \mu_2 = \mu_3\} \subset \Omega$.

In general, assume that $T_H : \mathcal{H} \to \mathbb{R}$ is a test statistic function for each null hypothesis $H \in \mathcal{H}$ and, without loss of generality, that large values of T_H are evidence against H. To formulate the sequentially rejective framework, define a critical value function $\mathbf{d} = \{d_H\}_{H\in\mathcal{H}}$ where each d_H maps from the power set $2^{\mathcal{H}}$ of all subsets of

$\mathcal{H}$ to $\mathbb{R} \cup \{-\infty, \infty\}$. Given the rejection set $\mathcal{R}$ from previous steps, the set

$$\{H \in \mathcal{H} \setminus \mathcal{R} : T_H \geqslant d_H(\mathcal{R})\}.$$

indicates the null hypotheses that would be rejected in the next step.

The sequentially rejective framework in the paper by Goeman and Solari (2010) is executed in the following fashion. Let $\mathcal{R}_i$ denote the collection of null hypotheses rejected after step i, $i = 0, 1, ...$ Start with

$$\mathcal{R}_0 = \emptyset.$$

Iteratively,

$$\mathcal{R}_{i+1} = \mathcal{R}_i \cup \{H \in \mathcal{H} \setminus \mathcal{R}_i : T_H \geqslant d_H(\mathcal{R}_i)\},$$

where $\mathcal{R}_\infty = \lim_{i \to \infty} \mathcal{R}_i$ is the final set of all rejected null hypotheses. In a nutshell, a sequentially rejective framework subsequently chooses hypotheses to reject on the basis of the previously rejected set. Equivalently, critical values are adjusted after each step, based on the newly rejected hypotheses. Now we quote the theorem proven by Goeman and Solari (2010) which shows that under two conditions the sequentially rejective framework controls FWE in the strong sense.

Lemma 1. *(Sequential rejection conditions).*

Suppose that for every $\mathcal{R} \subseteq \mathcal{S} \subseteq \mathcal{H}$ and every $H \in \mathcal{H} \setminus \mathcal{S}$,

$$d_H(\mathcal{R}) \geqslant d_H(\mathcal{S}) \tag{3.22}$$

and that for every $\omega \in \Omega$,

$$P_\omega(\bigcup_{H \in \mathcal{T}(\omega)} \{T_H \geqslant d_H(\mathcal{F}(\omega))\}) \leqslant \alpha. \tag{3.23}$$

Then, for every $\omega \in \Omega$,

$$P_\omega(\mathcal{R}_\infty \subseteq \mathcal{F}(\omega)) \geqslant 1 - \alpha.$$

Condition (3.23), the single-step condition, requires a weak form of FWE control in the case where exactly all the false hypotheses and none of the true hypotheses have been rejected. Condition (3.22), the monotonicity condition, ensures that if no false rejection occurs in the weak control case, there must be no false rejections in the cases with fewer hypotheses rejected previously than the weak control case. Consequently Conditions (3.22) and (3.23) are sufficient to guarantee the strong FWE control of the sequentially rejective framework.

Note that condition (3.22) immediately implies that $\forall H \in \mathcal{H} \setminus \mathcal{R}_i$,

$$d_H(\mathcal{R}_i) \geqslant d_H(\mathcal{R}_{i+1}). \tag{3.24}$$

That is, the sequential rejection conditions guarantee nonincreasing critical values at each step. However, it is important to be aware that condition (3.24), in place of condition (3.22), does not suffice to control FWE in the strong sense.

3.2 Forced Sequential Methods

3.2.1 Forced Sequentially Rejective Method

The sequential rejection conditions provide a way to verify the strong FWE control, when developing a step-down procedure based on a single-step procedure. Of the two sequential rejection conditions, the single-step condition is immediately justified by the strong FWE control property of the single-step tests that the stepdown methods are developed from. However, these single-step tests do not clearly exhibit the monotonicity condition. In fact, for most of the single-step tests that we have considered, it is rather difficult to give a rigorous proof of monotonicity. That leads us to speculate on forcing the monotonicity condition in stepdown methods. It turns

out that under a certain assumption, if we modify the critical values in stepdown methods by taking them to be the minimum of the critical values from single-step tests on any set of which the current set of hypotheses can be a subset, the monotonicity condition can be fulfilled, as well as the single-step condition. We refer to the sequentially rejective framework based on those modified critical values as the forced sequentially rejective method.

Recall the two sequential rejection conditions include the single-step condition and the monotonicity condition. We are going to show that if single-step tests have FWE control in the weak sense, under some assumptions, we can reassign critical values to satisfy the monotonicity condition, while still preserving the single-step condition.

First, let us introduce two assumptions.

Assumption 1. *(Subset pivotality condition)*

For any subset $\mathcal{I} \subset \mathcal{H}$, $\widetilde{\omega}(\mathcal{I})$ *is the least favorable parameter configuration in terms of FWE, among all* $\omega \in \bigcap_{H\in\mathcal{I}} M(H)$ *if*

$$\sup_{\omega\in\bigcap_{H\in\mathcal{I}} M(H)} P_\omega(\bigcup_{H\in\mathcal{I}}\{T_H > d\}) = P_{\widetilde{\omega}(\mathcal{I})}(\bigcup_{H\in\mathcal{I}}\{T_H > d\}),$$

for any real value d. $P_{\widetilde{\omega}(\mathcal{I})}$ *is thus called the least favorable measure for* $\mathcal{I}$.

Given that $\widetilde{\omega}(\mathcal{I})$ *exists for any* $\mathcal{I} \subseteq \mathcal{H}$, *the subset pivotality condition requires that there is an* $\omega^* \in \bigcap_{H\in\mathcal{H}} M(H)$ *such that for any* $\mathcal{I} \subseteq \mathcal{H}$, *the joint distribution of* $\{T_H, H \in \mathcal{I}\}$ *under* $P_{\widetilde{\omega}(\mathcal{I})}$ *is the same as that under* P_{ω^*}.

This means that, under the subset pivotality condition, an ultimate least favorable measure exists and dominates all the least favorable measures.

Assumption 2. *If* $d_H(\mathcal{R})$ *is a critical value of the single-step method applied to* $\mathcal{H} \setminus \mathcal{R}$, *then 1)* $d_H(\mathcal{R})$ *is identical for any* $H \in (\mathcal{H} \setminus \mathcal{R})$ *and 2)* $\forall\omega \in \bigcap_{H\in\mathcal{H}\setminus\mathcal{R}} M(H)$, $P_\omega(\bigcup_{H\in\mathcal{H}\setminus\mathcal{R}}\{T_H > d_H(\mathcal{R})\}) \leqslant \alpha$. *(That is to say, in a single-step method, individual test statistics are compared with a common critical value and the single-step method has FWE control in the weak sense.)*

Moreover, for any $\mathcal{R} \subseteq \mathcal{H}$, since $d_H(\mathcal{R})$ gives the same critical value for all hypotheses $H \in (\mathcal{H} \setminus \mathcal{R})$ after the hypotheses in $\mathcal{R}$ have been rejected, let $d(\mathcal{R})$ be shorthand for $d_H(\mathcal{R})$.

Now under these two assumptions, we define the forced sequentially rejective method (FSR) to be a sequentially rejective framework based on the modified critical value function collection $d^*_{\mathrm{FSR}}(.)$ such that $\forall \mathcal{R} \subseteq \mathcal{H}$,

$$d^*_{\mathrm{FSR}}(\mathcal{R}) = \min\{d(\mathcal{U}), \forall \mathcal{U} \subseteq \mathcal{R} \subseteq \mathcal{H}\}. \tag{3.25}$$

It is noteworthy that from the definition, FSR is applicable to any single-step method that satisfies Assumption 2.

Theorem 1. *Under the above two assumptions, the sequential rejection conditions hold for the forced sequentially rejective method.*

Proof. It suffices to verify the monotonicity condition and the single-step condition of the forced sequentially rejective method. First, the monotonicity condition is trivial, since it is easy to see $\forall \mathcal{R} \subseteq \mathcal{S} \subseteq \mathcal{H}$,

$$d^*_{\mathrm{FSR}}(\mathcal{S}) = \min\{d(\mathcal{U}), \forall \mathcal{U} \subseteq \mathcal{S} \subseteq \mathcal{H}\} \leqslant d^*_{\mathrm{FSR}}(\mathcal{R}), \quad \text{by } \mathcal{R} \subseteq \mathcal{S}.$$

For any $\omega_0 \in \Omega$, let $\mathcal{T}(\omega_0)$ and $\mathcal{F}(\omega_0)$ be the collections of true null hypotheses and false null hypotheses, respectively. By the definition of $d^*_{\mathrm{FSR}}(.)$ in (3.25), there exists an $\mathcal{R}_0 \subseteq \mathcal{F}(\omega_0)$ such that $d^*_{\mathrm{FSR}}(\mathcal{F}(\omega_0)) = d(\mathcal{R}_0)$. Then,

$$\begin{aligned}
&P_{\omega_0}\Big(\bigcup_{H \in \mathcal{T}(\omega_0)} \{T_H > d^*_{\mathrm{FSR}}(\mathcal{F}(\omega_0))\}\Big) \\
&\leqslant \sup_{\omega \in \bigcap_{H \in \mathcal{T}(\omega_0)} M(H)} P_\omega\Big(\bigcup_{H \in \mathcal{T}(\omega_0)} \{T_H > d^*_{\mathrm{FSR}}(\mathcal{F}(\omega_0))\}\Big) \\
&= P_{\widetilde{\omega}(\mathcal{T}(\omega_0))}\Big(\bigcup_{H \in \mathcal{T}(\omega_0)} \{T_H > d(\mathcal{R}_0)\}\Big)
\end{aligned}$$

$$= P_{\omega^*}\Big(\bigcup_{H\in\mathcal{T}(\omega_0)}\{T_H > d(\mathcal{R}_0)\}\Big), \qquad \text{for some } \omega^* \text{ from the subset pivotality condition,}$$

$$\leqslant P_{\omega^*}\Big(\bigcup_{H\in\mathcal{H}\setminus\mathcal{R}_0}\{T_H > d(\mathcal{R}_0)\}\Big), \qquad \text{since } \mathcal{T}(\omega_0) = \mathcal{H}\setminus\mathcal{F}(\omega_0) \subseteq \mathcal{H}\setminus\mathcal{R}_0,$$

$$\leqslant \alpha, \text{ by the weak FWE control of the single-step method that is applied to } \mathcal{H}\setminus\mathcal{R}.$$

Since ω_0 is arbitrarily chosen from Ω, the single-step condition holds for $d^*_{\mathrm{FSR}}(.)$. $\square$

Therefore, Lemma 1 and Theorem 1 guarantee that under Assumptions 1 and 2, FSR controls the FWE in the strong sense.

For ease in obtaining the critical values of FSR in (3.25), we represent $d^*_{\mathrm{FSR}}(.)$ in the following stepwise manner,

$$\begin{aligned} d^*_{\mathrm{FSR}}(\mathcal{R}_n) =& \min\{d^*_{\mathrm{FSR}}(\mathcal{R}_{n-1}), d(\mathcal{U}) : \forall\mathcal{U} \in (2^{\mathcal{R}_n}\setminus 2^{\mathcal{R}_{n-1}})\}, \\ & \text{and} \quad d^*_{\mathrm{FSR}}(\emptyset) := d(\emptyset). \end{aligned} \tag{3.26}$$

where $\mathcal{R}_n$ is the set of hypotheses rejected before step n and $2^{\mathcal{R}_n}$ is a power set of all the subsets of $\mathcal{R}_n$.

Expression(3.26) makes use of the critical value in the previous step $d^*(\mathcal{R}_{n-1})$, avoiding the redundant comparisons that expression(3.25) may indicate, and hence it gives an algorithm for finding new critical values at each step.

Following the critical values in(3.26), it is easy to determine how many evaluations of $d(\mathcal{R})$ are required to obtain $\{d^*_{\mathrm{FSR}}(R_n)\}_{n=0}^{\infty}$. If we count the computing time of $d(\mathcal{R})$ as 1 unit, for any $\mathcal{R}\subseteq\mathcal{H}$, then the total computation time for the critical values of FSR is $2^{[\mathcal{R}_\infty]}$, where [.] is the count of components in a set and $\mathcal{R}_\infty$ is the set of all rejected hypotheses when FSR terminates.

3.2.2 Equivalence of the forced stepdown method and FSR

Now given common critical values for all the hypotheses at each given step, consider two possible rejection procedures that specify the ways sequential rejections are implemented.

P1 : At each step with n null hypotheses, if there are totally $r(r > 0)$ significant signals, reject all the r signals and move on to next step with the remaining $(n - r)$ hypotheses. Otherwise, the procedure ends.

P2 : At each step with n null hypotheses, if there is any significant signal, only reject the signal with the largest test statistic and move on to next step with the remaining $(n - 1)$ hypotheses. Otherwise, the procedure ends.

Notice that the forced sequentially rejective method defined in the previous section follows P1. Now we define an analogous method that shares the same critical value function with FSR but follows P2 and refer to this method as the forced stepdown method (FSD). Then by the following theorem, we can show that FSD is equivalent to FSR.

Theorem 2. *Suppose a critical value function satisfies the monotonicity condition and Assumption 2. Let* $\mathcal{R}^1_\infty$,$\mathcal{R}^2_\infty$ *be the eventual rejection sets from the two rejection procedures P1 and P2, based on the same critical value function. Then* $\mathcal{R}^1_\infty = \mathcal{R}^2_\infty$.

Proof. Suppose $\mathcal{R}^1_\infty \neq \mathcal{R}^2_\infty$. Then at least one of the following two assertions is true, $\mathcal{R}^1_\infty \setminus \mathcal{R}^2_\infty \neq \emptyset$ or $\mathcal{R}^2_\infty \setminus \mathcal{R}^1_\infty \neq \emptyset$.

Without loss of generality, assume $\mathcal{R}^2_\infty \setminus \mathcal{R}^1_\infty \neq \emptyset$. That is, $\mathcal{R}^2_\infty \supsetneqq \mathcal{R}^1_\infty \cap \mathcal{R}^2_\infty$. Then, of all the sequential steps of procedure 2, there is a step n such that

$$R^2_n \subseteq (\mathcal{R}^1_\infty \cap \mathcal{R}^2_\infty) \subsetneq \mathcal{R}^2_{n+1}, \tag{3.27}$$

where $\mathcal{R}^i_n$ is the collection of the null hypotheses rejected after step n in manner i, $i = 1, 2$.

Take any $r \in \mathcal{R}^2_{n+1} \setminus (\mathcal{R}^1_\infty \cap \mathcal{R}^2_\infty)$. Then $r \notin \mathcal{R}^2_n$ and $r \in \mathcal{R}^2_{n+1}$. It follows that,

$$T_r > d(\mathcal{R}^2_n),$$

where T_r is the test statistic of hypothesis r and $d(\mathcal{R}^2_n)$ is the critical value for all the remaining hypotheses after $\mathcal{R}^2_n$ has been rejected.

Since $r \in \mathcal{R}^2_{n+1} \setminus (\mathcal{R}^1_\infty \cap \mathcal{R}^2_\infty)$, $r \in \mathcal{R}^2_{n+1} \subset \mathcal{R}^2_\infty$ and thus $r \notin \mathcal{R}^1_\infty$.

On the other hand, (3.27) implies that

$$R^2_n \subseteq (\mathcal{R}^1_\infty \cap \mathcal{R}^2_\infty) \subseteq \mathcal{R}^1_\infty.$$

As a result of the monotonicity condition,

$$d(\mathcal{R}^1_\infty) \leqslant d(\mathcal{R}^2_n) < T_r.$$

This contradicts the fact that $r \notin \mathcal{R}^1_\infty$. Hence, the supposition is false and $\mathcal{R}^2_\infty \setminus \mathcal{R}^1_\infty = \emptyset$.

It follows that $\mathcal{R}^1_\infty = \mathcal{R}^2_\infty$. □

From the monotonicity condition of $d^*_{\mathrm{FSR}}(.)$, Theorem 2 implies that FSD is equivalent to FSR, as they terminate with the same final set of rejections. Moreover, it is verified in Theorem 1 that under Assumptions 1 and 2, the forced sequentially rejective method satisfies the sequential rejection conditions and hence it provides strong control of the FWE. Consequently, with the same assumptions FSD controls the FWE in the strong sense. In fact, the strong control of FWE in FSD can be alternatively proven from a theorem due to Romano and Wolf (2005). In this paper they used the theorem to derive a modified stepdown method which will be discussed in section 3.2.4.

Analogous to (3.26), we represent the critical value function $d^*_{\text{FSD}}(.)$ for FSD utilizing the rejection procedure P2, as

$$d^*_{\text{FSD}}(\mathcal{R}_n) = \min\{d^*_{\text{FSD}}(\mathcal{R}_{n-1}), d(\mathcal{U} \cup (\mathcal{R}_n \setminus \mathcal{R}_{n-1})) : \forall \mathcal{U} \in 2^{\mathcal{R}_{n-1}}\}$$
$$\text{and} \quad d^*_{\text{FSD}}(\emptyset) := d(\emptyset). \tag{3.28}$$

Note that (3.28) can be considered as a special case of (3.26) when $\mathcal{R}_n \setminus \mathcal{R}_{n-1}$ only includes one hypothesis. Thus, obtaining all the critical values of FSD requires $2^{[\mathcal{R}_\infty]}$ evaluations of $d(\mathcal{R})$. This coincides with the total computation time of FSR, as they have the identical final set of rejections.

With regards to choosing between FSR and FSD, we let the practitioners determine which rejection method they prefer. However, given that FSR is equivalent to FSD in all concerning aspects, henceforth we refer to both methods as the forced stepdown method (FSD) and use $d^*(.)$ to refer to either $d^*_{\text{FSR}}(.)$ or $d^*_{\text{FSD}}(.)$.

3.2.3 Forced Pathwise Methods

From section 3.2.2, the computation time for the critical values in the forced stepdown method grows exponentially with the total number of rejections. In order to reduce the computational complexity of FSD, in this section we propose two alternative procedures that both force critical values to be monotone along the series of rejections in each realization path. Based on the adoption of two different rejection procedures, we refer to them as the forced pathwise sequentially rejective method (FPSR) and the forced pathwise stepdown method (FPSD), respectively. Concisely, the forced pathwise sequentially rejective method rejects all the significant signals at each step, whereas the forced pathwise stepdown method only rejects the most significant signal at each step.

In general, given a sequence of rejection sets $(\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_{n-1}, \mathcal{R}_n)$ in the current realization path, the critical value for the remaining hypotheses in the two forced

pathwise methods can be represented uniformly by

$$\tilde{d}(\mathcal{R}_n) = \min\{d(\mathcal{R}_m), 0 \leqslant m \leqslant n\},$$
$$\tilde{d}(\emptyset) = d(\emptyset). \tag{3.29}$$

As a result of the uniform expression (3.29) of the critical values, from now on, we use the term "the forced pathwise methods" to refer to both FPSR and FPSD. For example, with the computing time of $d(\mathcal{R})$ as unit 1, $\forall \mathcal{R} \subseteq \mathcal{H}$, the total computational time for the critical values of the forced pathwise methods is the count of components in the final sequence of rejection sets $\{\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_n, \ldots, \mathcal{R}_\infty\}$.

For any sequence of rejection sets $(\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_{n-1}, \mathcal{R}_n)$,

$$\tilde{d}(\mathcal{R}_n) = \min\{d(\mathcal{R}_m), 0 \leqslant m \leqslant n\} \geqslant d^*(\mathcal{R}_n),$$

where $d^*(\mathcal{R}_n)$ is the critical value of FSD defined in 3.25. That is, the critical values of the forced pathwise methods are never less than those of FSD. Thus under Assumptions 1 and 2, the forced pathwise methods have strong control of the FWE.

Note that the two forced pathwise methods have been shown to share most characteristics. However, they do differ in the pathwise context and consequently the two forced pathwise methods could have different critical value sequences. In fact, from (3.29), we can see that given a complete rejection path for FPSR, at each step the critical values of FPSR are uniformly greater than or equal to those of FPSD, because the former are the minimums from the previous steps in the path whereas the latter take the minimums from not only the previous steps but also the individual connection stairs that FPSD takes between the steps. Consequently, FPSD could end up with more rejections than FPSR and hence is potentially more powerful. On the other hand, the loss of power in FPSR is the cost for a smaller computational time than FPSD when more than one hypothesis is rejected at any step.

Now consider a special case. If the critical values of single-step methods satisfy the monotonicity condition, the forced sequentially rejective method and the forced stepdown method reduce to their forced pathwise counterparts, respectively. In this case, the critical value functions in the forced pathwise methods come straight from the single-step method and hence are the same for FPSR and FPSD. By Theorem 2, the forced pathwise methods have the common final set of rejections. Therefore, in this scenario FSD, FPSD and FPSR all lead to the same rejections. In view of computational complexity, however, FPSR could save considerable computational time in comparison to FSD and FPSD. In practice, given a single-step method satisfying Assumption 2, often times the simulation results indicate that the monotonicity condition applies to the critical value function, even though the analytic proof is intractable. In this situation, FPSR may be a preferable choice over FSD and FPSD, in the sense that it potentially saves sufficient computational time without necessarily sacrificing much power or even any.

3.2.4 Comparison of Sequential Methods

Note that if Assumption 2, the subset pivotality condition, holds, our proposed forced stepdown method is guaranteed to control the FWE in the strong sense when applied to any single-step method that satisfies Assumption 2. Alternatively, Romano and Wolf (2005) constructed a modified stepdown method (RWSD) that has strong control of the FWE under Assumption 2 and does not require Assumption 1. To acquire monotonicity, they adjusted the critical value function as follows,

$$d^{**}(\mathcal{R}_n) = \max\{d(\mathcal{U}) : \forall \mathcal{U} \supseteq \mathcal{R}_n\}.$$

Notice that $d^{**}(\mathcal{R}_n)$ is always greater than or equal to $d(\mathcal{R}_n)$ defined in 3.25. This conservativeness is the cost for eliminating the subset pivotality assumption which Romano and Wolf (2005) felt was too restrictive. However, Westfall and Troendle

(2008) stated "the subset pivotality condition is easily satisfied in many cases, including the general multivariate regression model with location-shift multivariate (possibly nonnormal) distributions". Certainly, the general multiple comparisons model that we considered in Chapter 2 is one of these cases. Then under the subset pivotality condition,

$$d^{**}(\mathcal{R}_n) \geqslant d(\mathcal{R}_n) \geqslant \tilde{d}(\mathcal{R}_n) \geqslant d^{*}(\mathcal{R}_n),$$

and this implies that the Romano and Wolf (2005) modified stepdown method sometimes leads to less rejections than the forced pathwise stepdown method and hence is potentially less powerful than all of our three proposed forced methods.

Now consider the computational time of RWSD. At the initial step

$$d^{**}(\emptyset) = \max\{d(\mathcal{U}) : \forall \mathcal{U} \subseteq \mathcal{H}\},$$

which is the maximum critical value taken over all the possible rejection sets and hence needs $(2^{[\mathcal{H}]} - 1)$ evaluations of $d(\mathcal{R})$. After the computation of $d^{**}(\emptyset)$, the critical values in the follow-up step are determined completely based on the already evaluated $d(\mathcal{R})$ and do not require any more computations. Therefore, the total computing time of RWSD is fixed at $(2^{[\mathcal{H}]} - 1)$. Notice this is greater than $2^{[\mathcal{R}_\infty]}$, the total computing time of FSD, the most time-consuming one of the three proposed forced methods. Only when all the null hypotheses are rejected, FSD uses the same computing time as RWSD, since no further critical values are calculated. Therefore, the Romano and Wolf (2005) modified stepdown method is computationally more intense than all of the proposed forced methods.

Now we provide an illustration of the construction of the critical values for the three proposed forced methods and the Romano and Wolf (2005) modified stepdown method via Figures 6 and 7. Suppose $\mathcal{H} = \{H_1, H_2, H_3, H_4\}$. Given any $\mathcal{R} \subset \mathcal{H}$, $d(\mathcal{R})$ is the critical value of a single-step method applied to all the remaining hypotheses in $\mathcal{H} \backslash \mathcal{R}$. Figure 6 shows a full diagram of possible sequential rejection paths, where each

node represents the current collection of rejections. All rejection paths start with no rejections. A rejection path can cover each level of the diagram or skip over multiple levels to reach a lower level. Every time a path moves to a next level, a new hypothesis is rejected. The termination of a rejection path can happen at any level. Without loss of generality, assume $T_{H_1} > T_{H_2} > T_{H_3} > T_{H_4}$. Supposing that H_1 and H_2 have already been rejected, Figure 7 provides the remaining possibilities of a full rejection path and illustrates how the critical values for the various methods are determined for the remaining hypotheses $\{H_3, H_4\}$. Let $\{a_1, a_2, \ldots, a_k\}$ be a shorthand for the set $\{H_{a_1}, H_{a_2}, \ldots, H_{a_k}\}$. Now, given the rejection set $\{1, 2\}$, the critical values of the four sequential procedures are demonstrated as follows,

I) FSD critical value

$$d^*(\{1,2\}) = \min[d(\emptyset), d(\{1\}), d(\{2\}), d(\{1,2\})] = \min[d^*(\{1\}), d(\{2\}), d(\{1,2\})].$$

II) RWSD critical value

$$d^{**}(\{1,2\}) = \max[d(\{1,2\}), d(\{1,2,3\}), d(\{1,2,4\})].$$

III) FPSD critical value

Suppose the current rejection sequence is $\emptyset \to \{1\} \to \{1,2\}$,

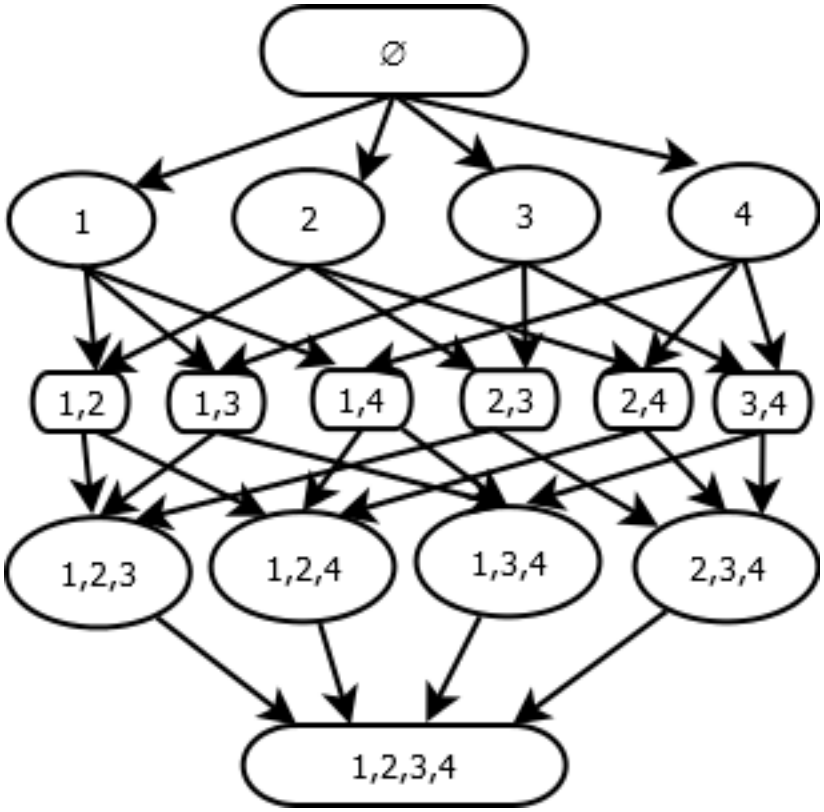
$$\tilde{d}(\{1,2\}) = \min[d(\emptyset), d(\{1\}), d(\{1,2\})] = \min[\tilde{d}(\{1,2\}, d(\{1,2\})].$$

IV) FPSR critical value

Suppose the current rejection sequence is $\emptyset \to \{1,2\}$,

$$\tilde{d}(\{1,2\}) = \min[d(\emptyset), d(\{1,2\})].$$

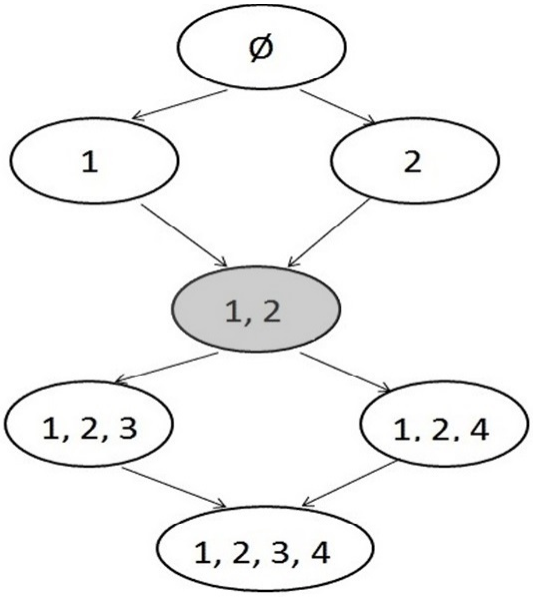
Figure 6: Full sequential rejection set diagram



i Each node represents the current rejection collection.

ii Arrows from node to node indicate directions of rejection paths.

Figure 7: Partial sequential rejection set diagram



i Node of $\{1, 2\}$ is the occurring rejection collection considered.

ii Arrows from node to node indicate directions of rejection paths.

Finally, in Table 2, we summarize the relationship between three proposed forced methods and the Romano and Wolf (2005) modified stepdown method in terms of FWE, power and computing time. The results in Table 2 are subject to Assumptions 1 and 2.

Table 2: Comparisons of the sequential methods

Criterion	FSD	FPSD	FPSR	RWSD
Strong FWE control	✓	✓	✓	✓
Power[1]	1	2	3	4
Computational consumption	2	3	4	1

[1] Ordinal numbers indicate descending strength of the criteria.

Now we consider applying our proposed forced procedures in the context of the general multiple comparisons model. Specifically, we modify the hybrid method to obtain all three proposed forced methods FSD, FPSD and FPSR. These three forced sequential methods control the FWE in the strong sense and are potentially more powerful than their single-step counterpart. In addition, with the subset pivotality condition, all the three forced sequential methods potentially outperform the corresponding modified stepdown methods suggested by Romano and Wolf (2005), in terms of both power and computing time. The performance of these above-mentioned sequential methods will be investigated in detail in the following simulation study.

3.3 Simulation: Multiple testing with sequential methods

This section presents a simulation study that explores the performance of these sequential methods in multiple tesing. We utilized the critical values of the hybrid method to obtain the three proposed forced procedures FSD, FPSD, FPSR and the Romano and Wolf (2005) modified stepdown procedure. We chose the hybrid method because it outperforms the other single-step methods most of the time in the first simulation study. We compared the performance of these sequential methods with the stepdown Bonferroni method, also known as Holm's procedure, and the stepdown Sidak method in the context of SMM.

There have been some simulation-based methods that provide good approximations for exact multivariate-t quantiles, such as the algorithm by Genz and Bretz (2009) based on Quasi-Monte Carlo methods. Since these simulation-based methods are nearly exact, they outperform all the conservative methods that we have discussed. However, we do not consider modifying these simulation-based methods to obtain stepdown methods, because these estimated critical values, based on simulated data, are not always guaranteed to control the FWE.

Following the general multiple comparisons model, we generate a vector estimate $\hat{\boldsymbol{\beta}}$ from a multivariate normal distribution with mean vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_k)$ and

covariance matrix $\sigma^2\boldsymbol{V}$. An estimate s^2 of σ^2 is generated from $\frac{\sigma^2\chi^2_\nu}{\nu}$, independently from $\hat{\boldsymbol{\beta}}$, where ν is the degrees of freedom.

Each null hypothesis is $H_j : \beta_j = 0$, and each alternative hypothesis is two-sided. The performance criteria are empirical FWE, any-pair power, all-pairs power, the average proportion of true positives, and computational time, where any-pair power is referred to as the probability of rejecting at least one of the false hypotheses, and all-pairs power is referred to as the probability of rejecting all false hypotheses, see Ramsey (1978). The specific levels of additional factors concerning both the null and alternative hypotheses are given below.

1. $\boldsymbol{\beta}$ values. The chosen levels are

   (a) $\boldsymbol{\beta} = \mathbf{0}$.

   (b) Half of the β_j's are equal to 0 and half of the β_j's are equal to 1.5.

   (c) Half of the β_j's are equal to 0 and half of the β_j's are equal to 3.

   (d) Half of the β_j's are equal to 1.5 and half of the β_j's are equal to -1.5.

   (e) Half of the β_j's are equal to 3 and half of the β_j's are equal to -3.

   (f) All of the β_j's are equal to 1.5.

   (g) All of the β_j's are equal to 3.

2. $\boldsymbol{V}$ matrix.

   (a) Choose $\boldsymbol{V}$ to be a first-order autoregressive correlation matrix with parameter ϕ, i.e. $\boldsymbol{V} = [v_{ij}]$ for $v_{ij} = \phi^{|i-j|}, \phi = 0, 0.5, 0.9$.

   (b) Let $\boldsymbol{V} = [v_{ij}]$ for $v_{ij} = \phi^{I\{i \neq j\}}$, where $I\{.\}$ is an indicator function. $\phi = 0, 0.5, 0.9$.

   (c) Generate a random $\boldsymbol{V}$ from a Wishart distribution $W_k(\Sigma, df)$, where Σ is a first-order autoregressive correlation matrix with $\phi = 0.3, 0.8$ and df, the degrees of freedom, is chosen to be 5 and 60.

3. $\boldsymbol{C}$ matrix. We consider SMM, i.e. $\boldsymbol{C} = \boldsymbol{I}$.

4. The variance $\sigma^2 = 1$.

5. Number of parameters, k. We consider $k = 4, 10$.

6. Error degrees of freedom, ν. We consider $\nu = 30, 100$.

7. Nominal FWE level, α. We consider $\alpha = 0.05, 0.1$.

Figures 8-13 show the results of the empirical FWE and power of several methods in the four scenarios for the mean vector $\boldsymbol{\beta}$. All the results are based on $10,000$ repetitions. Of all the stepdown procedures developed from the hybrid method, we include only the FPSR, as the FPSR uses the least computing time and the other three stepdown procedures always have the same rejections as the FPSR. The identical rejections from all these stepdown procedures are attributed to the monotonicity of the hybrid critical values which was empirically observed, but not theoretically proven. In order to show the power improvement of the stepdown procedure, we also compare the single-step hybrid method with its stepdown version. The results in Figures 8-13 are presented for $\alpha = 0.05$. We did not proceed with the comprehensive results for $\alpha = 0.1$, as the preliminary results exhibited similar patterns to those when $\alpha = 0.05$. These preliminary results for $\alpha = 0.1$ are provided in Appendix B.

Figures 8-10 provide the simulated FWE from the three structures of $\boldsymbol{V}$ matrix, for the case with all or half of the hypotheses being true. Particularly, the FWEs in Figure 10 are the averages of the empirical FWEs obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions. So, the standard error is $\sqrt{\frac{.05\times 0.95}{10{,}000}} \approx 0.002$, for the simulated FWEs in Figure 8 and Figure 9, and $\frac{.0002}{\sqrt{5}} \approx 0.0009$, for the obtained FWEs in Figure 10. We can see that the results in Figures 8-10 demonstrate the control of the FWE in all four considered methods, as all the FWEs are either below α or within 2 standard errors of α. Specifically, when $\phi = 0$, in the least favorable case with all the null hypotheses being true, the simulated

Figure 8: Empirical FWEs for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods when $\boldsymbol{V} = AR1(\phi)$

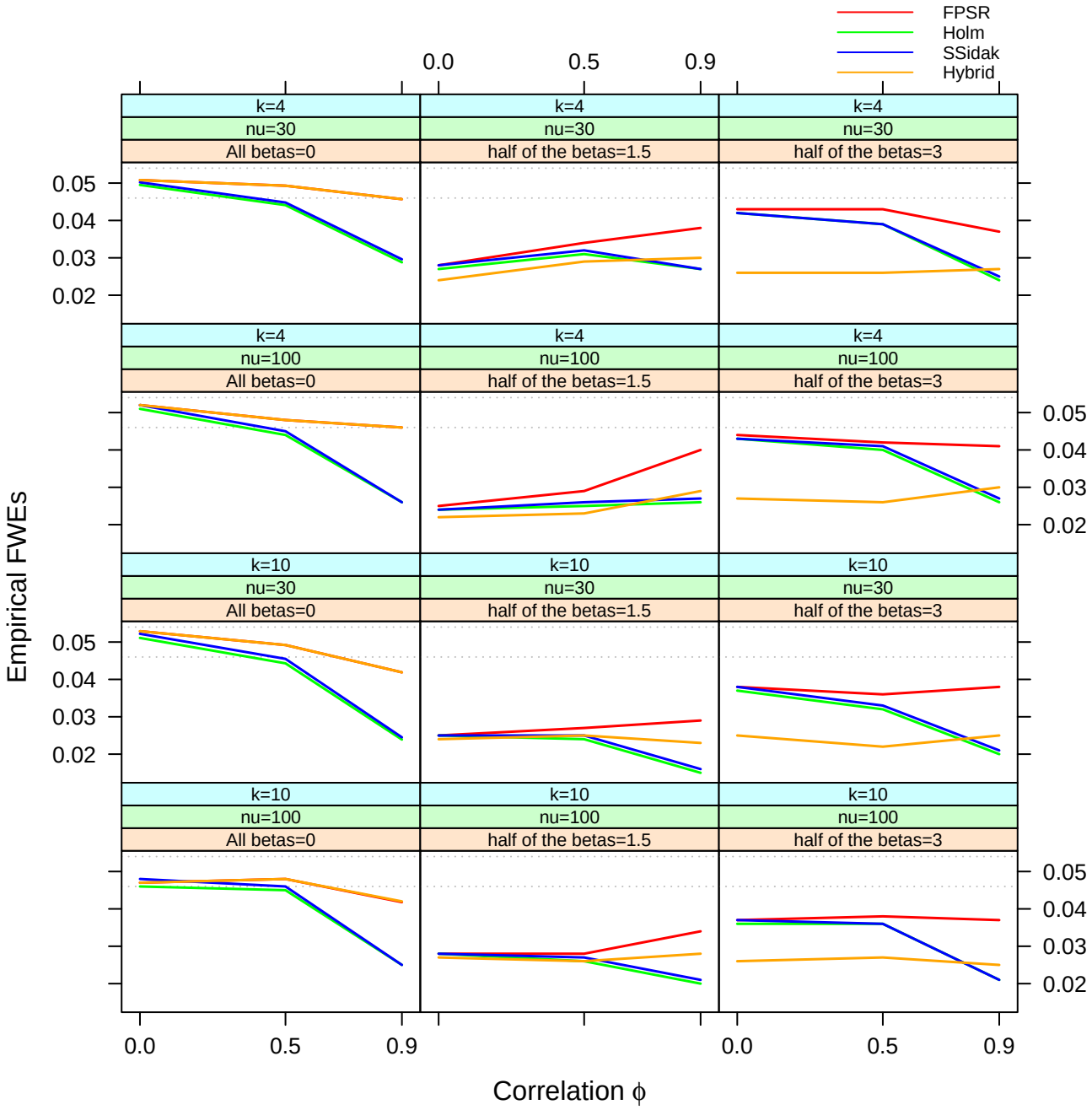


Figure 9: Empirical FWEs for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

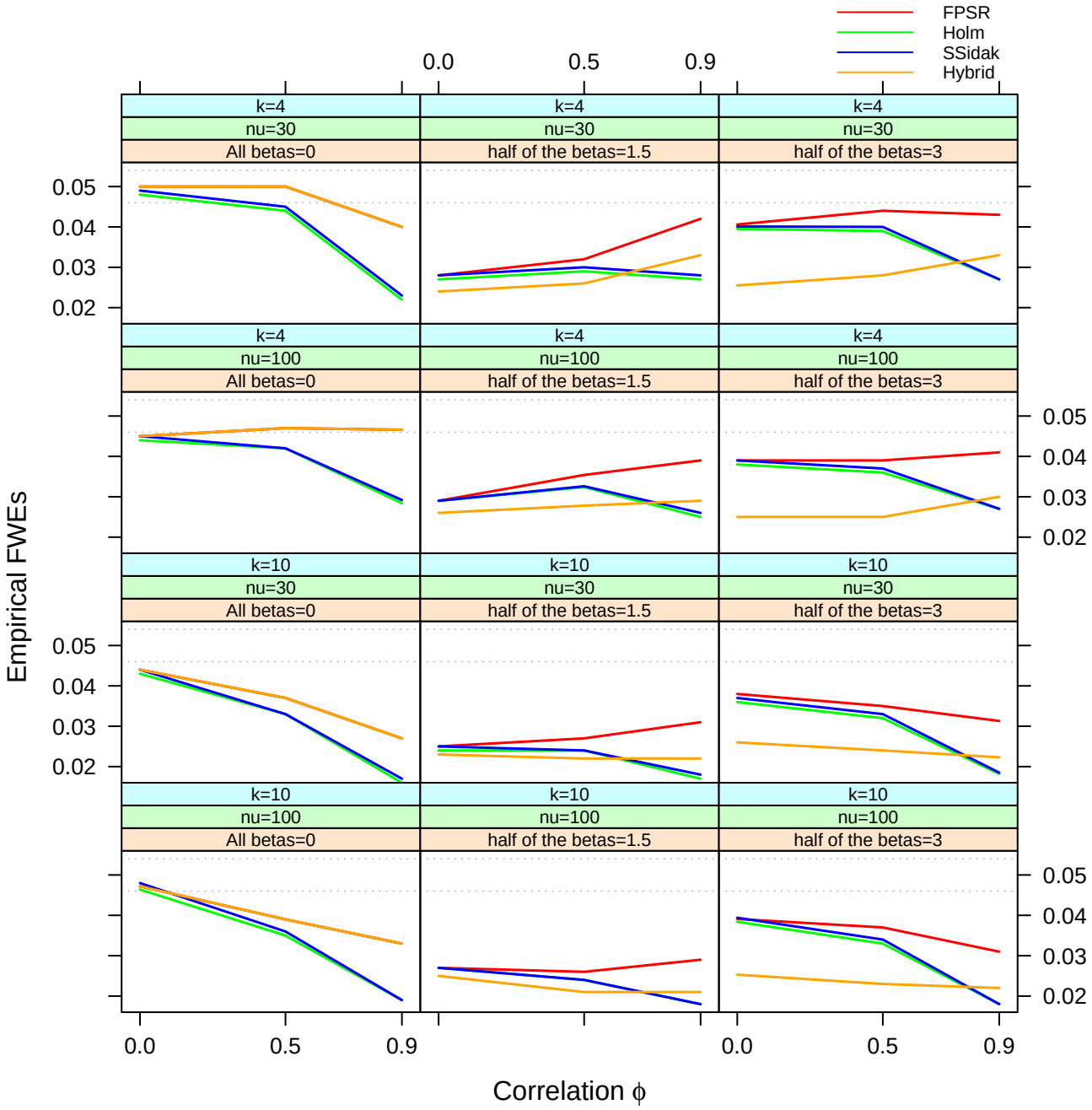
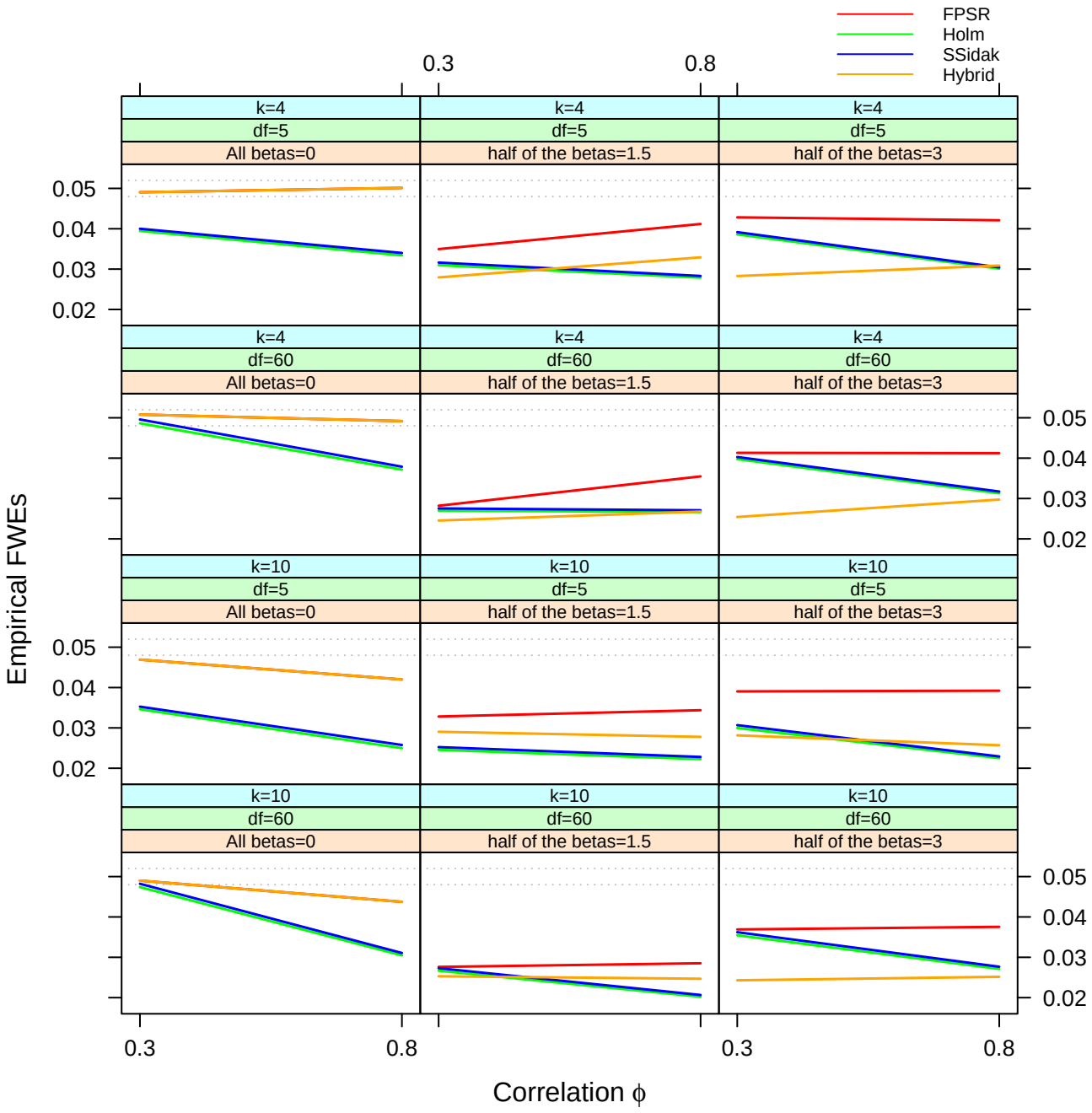


Figure 10: Empirical FWEs for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.



FWEs are similar for the stepdown Sidak, hybrid FPSR and Holm's methods. For large values of ϕ, the simulated FWE for the hybrid FPSR remains controlled at α and is much closer to the nominal level than those of the stepdown Sidak and Holm's methods. Note that the single-step hybrid method has identical FWEs with the hybrid FPSR, when all the null hypotheses are true. That is because, any (false) rejections occur in the hybrid FPSR if and only if some hypotheses are rejected in the first step by the single-step hybrid method. From Figures 8-10, we see that the comparative patterns of the empirical FWEs in the four methods are consistent across all three structures of the $\boldsymbol{V}$ matrix. Analogously, the following simulation results of the power performance apply to all the three structures of $\boldsymbol{V}$ matrix. Therefore, we chose to only present the results when $\boldsymbol{V}$ has the compound symmetry structure, and the results with other two structures of $\boldsymbol{V}$ matrix can be found in Appendix A.

Figures 11-13 present the three measurements of powers for the four compared methods, under three different configurations of false hypotheses. In these figures, although the three power measurements show different behaviors as the correlation changes, as far as the four considered methods are concerned, the powers of any-pair and all-pairs basically have consistent patterns with the average power. The only inconsistency occurs in the comparison between the hybrid method, and the stepdown Sidak or Holm's methods. Often times, while the average power of the hybrid method is lower than those of the stepdown Sidak or Holm's methods, the any-pair power exhibits the converse relation. This may be due to the circumstances where some hypotheses are rejected by the hybrid method whereas meanwhile no rejections occur in the stepdown Sidak or Holm's methods, because the hybrid method is usually more powerful than the Sidak and Bonferroni methods. Sometimes, for a large value of ϕ, if the average power of the hybrid method is higher than those of the stepdown Sidak or Holm's methods, the all-pair power may show a reverse trend. This is probably because in the cases of "any-pair" rejections by the stepdown Sidak

Figure 11: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to 0, when $\boldsymbol{V} = AR1(\phi)$.

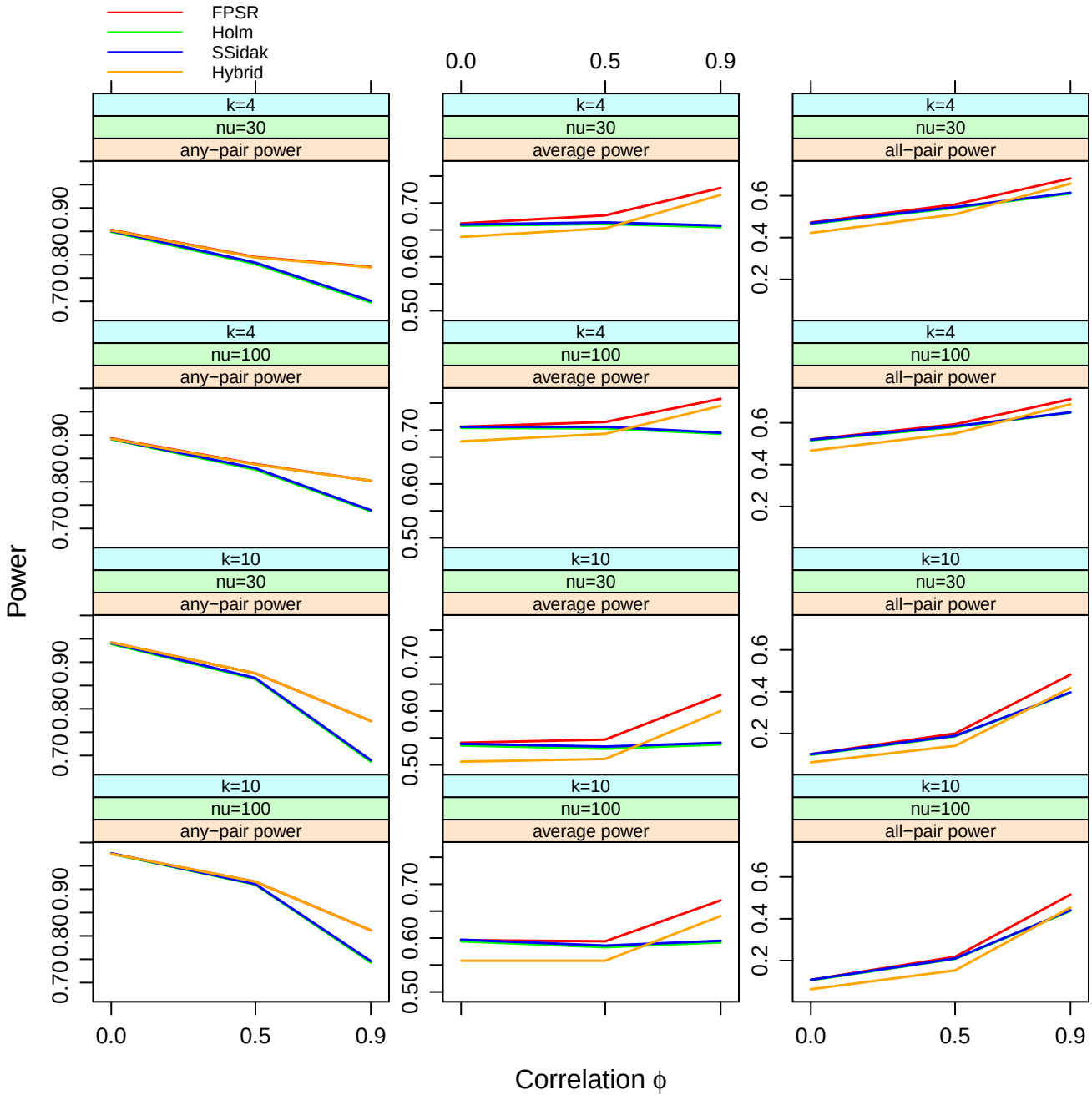


Figure 12: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to -3, when $\boldsymbol{V} = AR1(\phi)$.

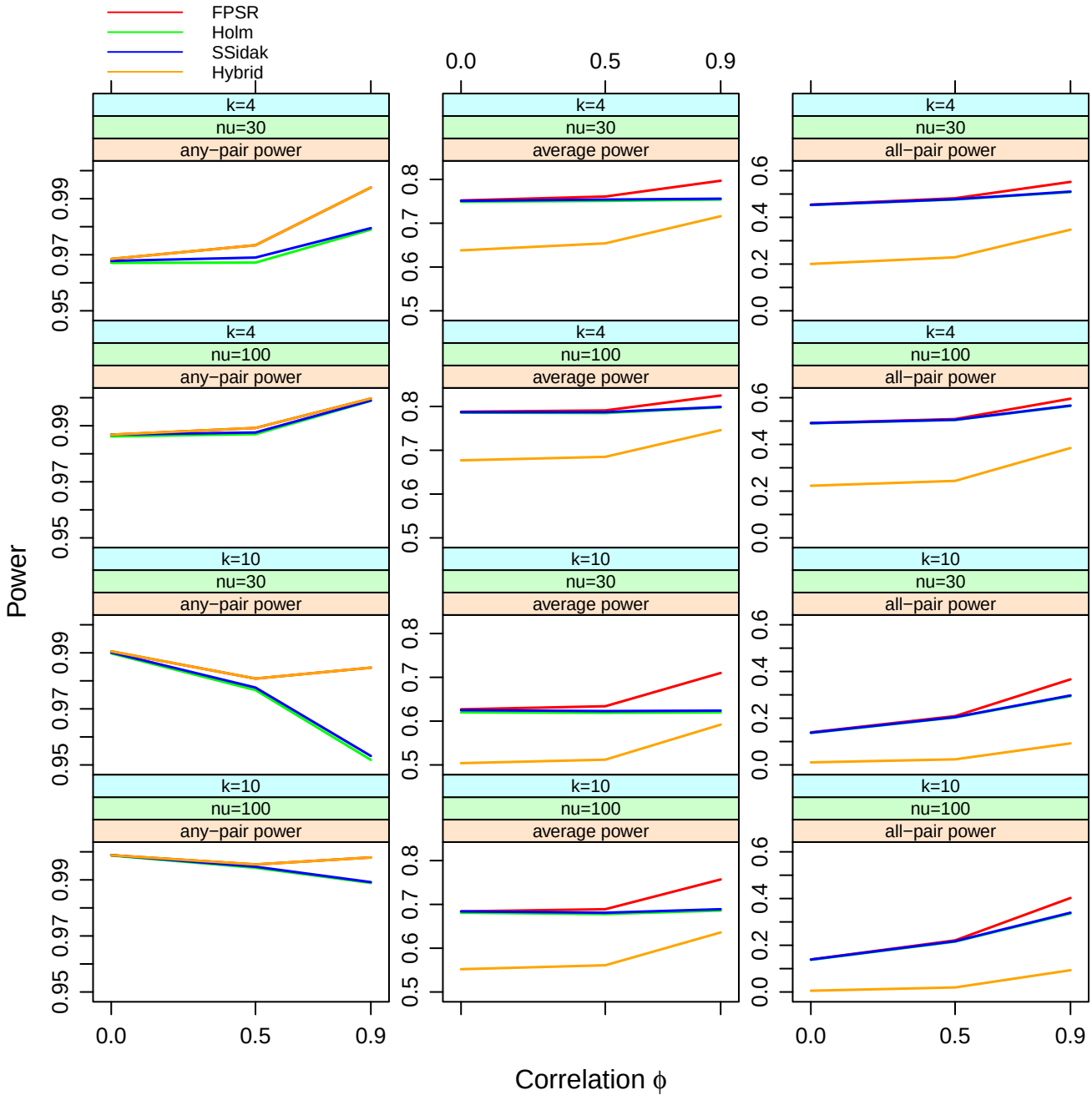
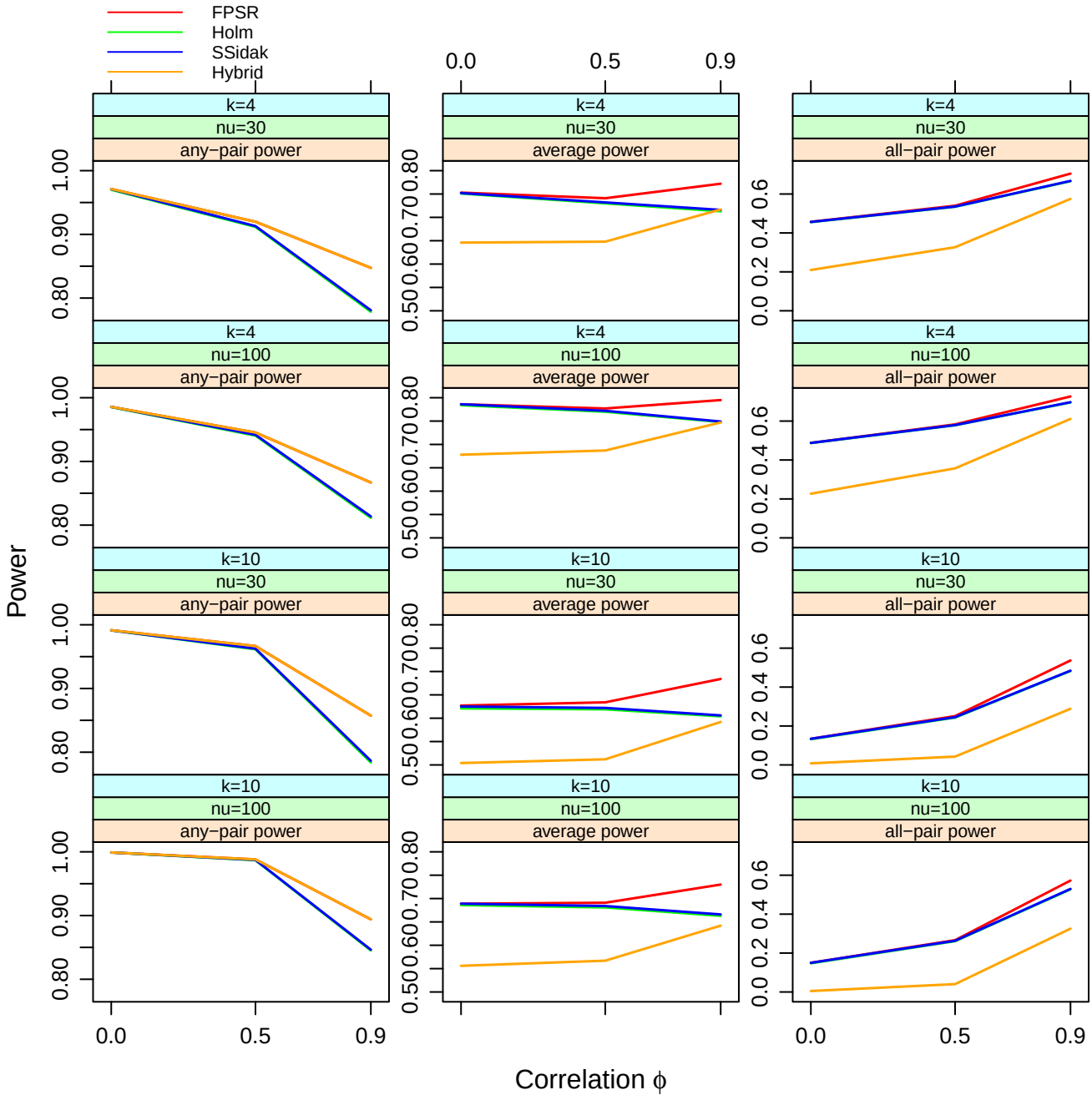


Figure 13: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 3, when $\boldsymbol{V} = AR1(\phi)$.



or Holm's methods, the stepdown nature make them more likely to reject all false hypotheses than the single-step hybrid method.

From Figures 11-13, we can see that the hybrid FPSR is usually the most powerful method of the four, no matter which power measurement is concerned. Only when the correlation is extremely low and the degrees of freedom is large, is the hybrid FPSR outperformed by the stepdown Sidak method. The reason for this exception is that in those cases, the test statistics $\frac{c_j'\hat{\boldsymbol{\beta}}}{sc_j'\boldsymbol{V}c_j}$ are almost independent. Hence, Sidak's method is virtually exact and apparently less conservative than the hybrid method. The performance of the stepdown procedures is closely related to that of their single-step counterparts. Since for very low correlation and large degrees of freedom, Sidak's method is less conservative than the hybrid method, in the same cases the stepdown Sidak method outperforms the hybrid FPSR. Likewise, when the correlation is high, as the hybrid method provides a substantial gain in power over the Bonferroni method and Sidak method, the power advantage of the hybrid FPSR relative to Holm's procedure and the stepdown Sidak method is significant. In Figures 11-13, we did not include the results for the case of $\beta_j = 1.5$ in the false hypotheses but included them in Appendix B, as they show very similar patterns to the results with $\beta_j = 3$.

With regards to computation, although the hybrid FPSR requires much more computing time than Holm's procedure and the stepdown Sidak method, in practice it is computationally feasible. For example, in Matlab R2012b, one repetition with the hybrid FWER takes about 3 seconds if $k = 4$, and 10 seconds if $k = 10$.

From the simulation results in Chapter 2, we know that the minimal restricted Scheffe method is outperformed by the hybrid method in all the considered scenarios. Hence, when the minimal restricted Scheffe method is used to develop sequential methods, these sequential methods are less powerful than the corresponding methods utilizing the hybrid method. Here, to show the power difference of the sequential methods developed from these two single-step methods, in Table 3 we present the results of average power in the scenario where $k = 4, \nu = 100$ and $\boldsymbol{V} = AR1(0.9)$.

This is also the scenario where we obtained the highest relative efficiency of the minimal restricted Scheffe method to the hybrid method.

In the simulation, we utilize the unique minimal restricted Scheffe critical values at given steps when modifying them to obtain the critical values for each of the four sequential methods. This means that, with a given set of hypotheses, the minimal restricted Scheffe critical value is obtained subject to the same sets of eigenvalue permutations and basis variations, when being modified to develop each of the four sequential methods. Hence, the comparisons of the four resulting sequential methods exibit the patterns illustrated in Table 2. Additionally, our simulation results show that the critical values of the minimal restricted Scheffe method agree with the monotonicity condition almost all the time and hence the forced sequential methods almost always have the same rejections. With a random choice of eigenvalue permutation and basis variation, very rarely an extremely small critical value does occur that breaches monotonicity. In that case, however, the differences in power within the four sequential methods based on the minimal restricted Scheffe method are almost negligible, compared with the power differences of the sequential methods using the hybrid method and the minimal restricted Scheffe method. Therefore, we choose to compare the least time-consuming FPSR utilizing the minimal restricted Scheffe method with the hybrid FPSR in Table 3. Note that the FPSR based on the minimal restricted Scheffe method controls the FWE more conservatively than the hybrid FPSR, and consequently exhibit lower power in all the considered configurations of the false hypotheses.

Table 3: Empirical FWEs and average proportion of true positives for hybrid FPSR *(hFPSR)* and minimal restricted Scheffe FPSR *(mFPSR)* when $k = 4, \nu = 100$ and $\boldsymbol{V} = AR1(0.9)$.

FWE *(hFPSR)*	FWE *(mFPSR)*	Average Power *(hFPSR)*	Average Power *(mFPSR)*
all $\beta_j = 0$			
0.047	0.028	-	-
half of the $\beta_j = 1.5$			
0.038	0.032	0.207	0.161
half of the $\beta_j = \pm 1.5$			
-	-	0.215	0.169
all $\beta_j = 1.5$			
-	-	0.240	0.204

4 Summary

Researchers often conduct large experiments designed to answer multiple related questions. A separate inferences approach to these questions can cause many false positives and lead to erroneous conclusions that may have negative consequences. Multiple comparison procedures are intended to take into account and properly control for multiplicity through some joint measure of erroneous inference. The family-wise error rate (FWE) is the probability of making at least one false positive among all the hypotheses. Research conclusions are often made based on the entirety of results from the comparisons. When the requirement of the simultaneous correctness of all inferences must be satisfied, FWE is the appropriate choice for error control. In this dissertation, we are concerned with FWE control. There are different types of multiple comparison methods to control the FWE. Single-step methods make an equivalent multiplicity adjustment for all the comparisons and are primarily used to construct simultaneous confidence intervals. Hence, for single-step methods, we propose a minimal cone approach, which is applied to the restricted Scheffe method, for constructing simultaneous confidence intervals for a finite number of comparisons.

Our simulation results indicate that, among the single-step competitors, the hybrid method outperforms the minimal restricted Scheffe method in all the considered scenarios; The minimal restricted Scheffe method is, however, less conservative than Sidak's method for some cases when the degrees of freedom are low. For a finite number of comparisons, the minimal restricted Scheffe method always outperforms Scheffe's method and the improvement over Scheffe's method is considerable, if the correlations between comparisons are high or the number of comparisons is small. The minimal restricted Scheffe method is usually more computationally costly than

the other considered methods. However, in practice, researchers can adjust the number of basis iterations in the minimal restricted Scheffe method depending on the computational resources.

Stepwise methods are another type of multiple comparison methods in which the current step result depends on the test results of previous steps. If only hypothesis testing is of concern and simultaneous confidence intervals are not needed, stepwise methods can substantially increase the power of the testing procedure. With regards to stepwise methods, we develop new forced sequential methods that utilize single-step methods to achieve improved power in multiple testing. At each step, our sequential methods are reduced to single-step tests which control the FWER. We suggest modifications to the critical values such that the modified critical values are monotone at all times. To facilitate computation, two of the three stepwise methods are proposed to modify monotone critical values along the rejection path. Moreover, new sequential methods are developed from applying these modifications to the hybrid method proposed by McCann and Edwards (2000).

To develop sequential procedures for the hybrid method, we suggest the forced pathwise sequentially rejective method (FPSR), as in the simulation results it has the same rejections as the other forced sequential methods but uses the least computing time. In the simulation study, the hybrid FPSR exhibits the strong control of FWE in all different settings of correlation structures between the test statistics. In terms of power, the hybrid FPSR outperforms other competitor, except the stepdown Sidak method for some cases when Sidak's method is essentially exact. The power advantage of the hybrid FPSR becomes particularly substantial when the correlations between the test statistics are high. Additionally, from the simulation results, the hybrid FPSR is computationally feasible; for instance, in MATLAB R2012b, one replication with the hybrid FPSR takes about 10 seconds when testing 10 parameters in SMM.

The minimal restricted Scheffe method and the forced sequential methods are designed to improve the performance of simultaneous confidence intervals or the power

of multiple hypothesis testing, subject to strong FWE control. The minimal restricted Scheffe method requires minimal assumptions on the distribution of the estimated parameter vector, and the discrete set of comparisons can be any linear combinations of the parameter vector. Meanwhile, the forced sequential procedures are applicable to any single-step methods under mild conditions and show the potential to uniformly improve power over their single-step counterparts. Therefore, one of our future plans is to choose, apart from the hybrid method, another single-step method that may not exhibit monotone critical values, and utilize it via the proposed sequential methods. From the perspectives of power and computing time, it will be interesting to compare these sequential methods that may end up with different rejections.

BIBLIOGRAPHY

Blazer, D., L. Georg, R. Landerman, M. Melville, M. Woodbury, K. Manton, K. Jordan, and B. Locke (1985). Psychiatric disorders: A rural/urban comparison. *Arch Gen Psychiatry 42*, 651–656.

Casella, G. and W. Strawderman (1980). Confidence bands for linear regression with restricted predictor variables. *Journal of the American Statistical Association 75*, 862–868.

Genz, A. and F. Bretz (2009). Computation of multivariate normal and t probabilities, volume 195 of lecture notes in statistics. Springer.

Goeman, J. and A. Solari (2010). The sequential rejection principle of familywise eror control. *The Annals of Statistics 38*, 3782–3810.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*, 65–70.

Hsu, J. (1996). Multiple comparisons: Theory and methods. Chapman & Hall/CRC.

Hunter, D. (1976). An upper bound for the probability of a union. *Journal of Applied Probability 13*, 597–603.

Jogdeo, K. (1977). Association and probability inequalities. *Annals of Statistics 5*, 495–504.

McCann, M. and D. Edwards (1996). A path length inequality for the multivariate-t distribution, with applications to multiple comparisons. *Journal of the American Statistical Association 91*, 211–216.

McCann, M. and D. Edwards (2000). A hybrid method for improved critical points for multiple comparisons. *Communications in Statistics - Simulation and Computation 29*, 703–722.

Naiman, D. (1986). Conservative confidence band in curvilinear regression. *The Annals of Statistics 14*, 896–906.

Ramsey, P. H. (1978). Power differences between pairwise multiple comparisons. *Journal of the American Statistical Association 73*(363), 479–485.

Romano, J. and M. Wolf (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*, 94–108.

Romano, J. and M. Wolf (2010). Balanced control of generalized error rates. *The Annals of Statistics 38*, 598–633.

Sa, P. and D. Edwards (1993). Tables of the casella and strawderman (1980) critical points for use in multiple comparisons with a control in response surface methodology. *Communications in Statistics - Simulation and Computation 22*, 765–777.

Scheffe, H. (1959). The analysis of variance. New York, NY: John Wiley & Sons.

Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association 62*, 626–633.

Uusipaikka, E. (1984). Exact confidence bands on certain restricted sets in the general linear model. Unpublished manuscript.

Westfall, P., R. Tobias, D. Rom, R. Wolfinger, and Y. Hochberg (1999). Multiple comparisons and multiple tests: Using sas. Gary, NC: SAS Institute.

Westfall, P. and J. Troendle (2008). Multiple testing with minimal assumptions. *Biometrical Journal 50*, 745–755.

Worsley, K. (1982). An improved bonferroi inequality and applications. *Biometrika 69*, 297–302.

A Appendix

Figure A.14: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to 0, when $\boldsymbol{V} = AR1(\phi)$.

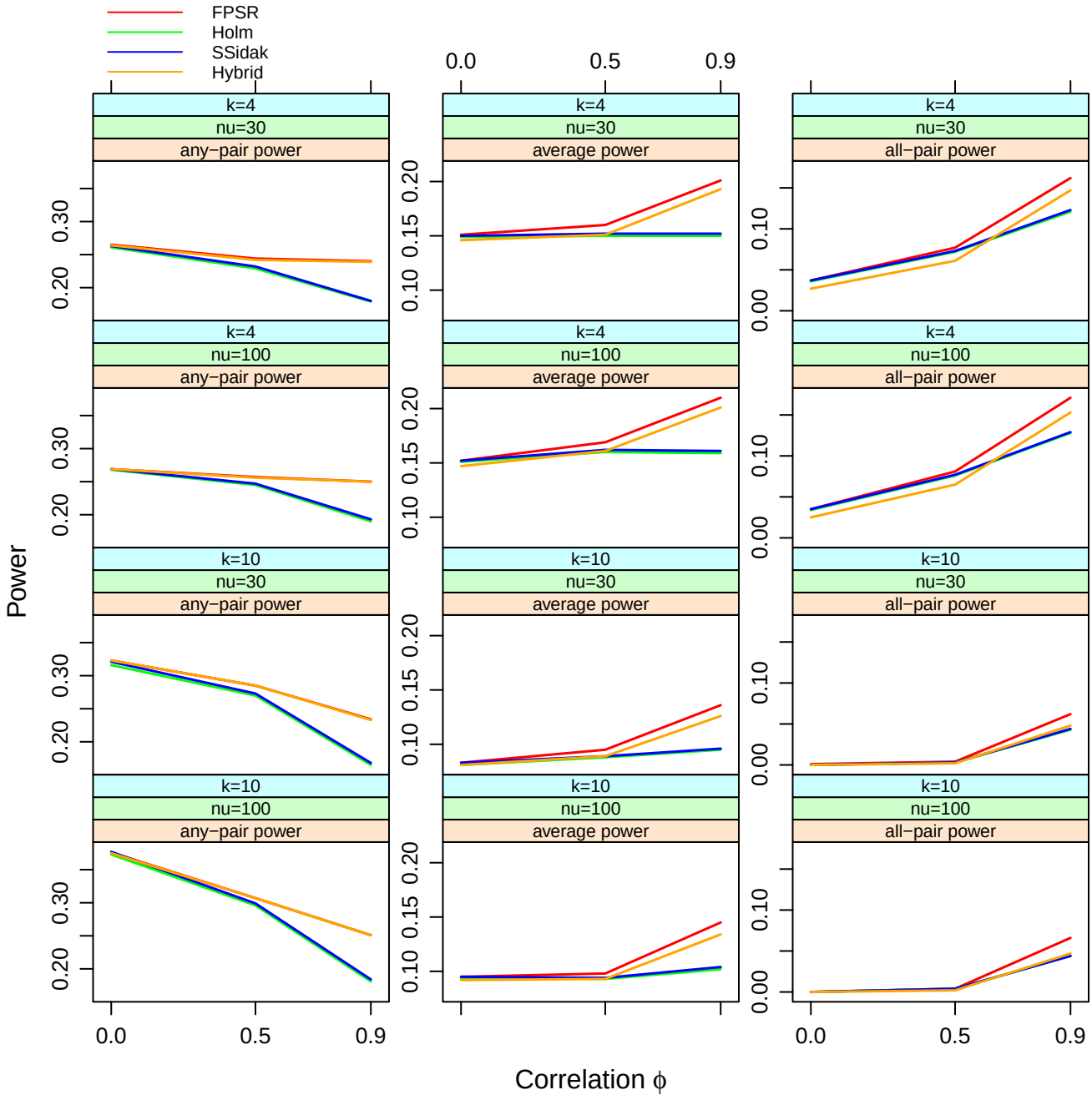


Figure A.15: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to -1.5, when $\boldsymbol{V} = AR1(\phi)$.

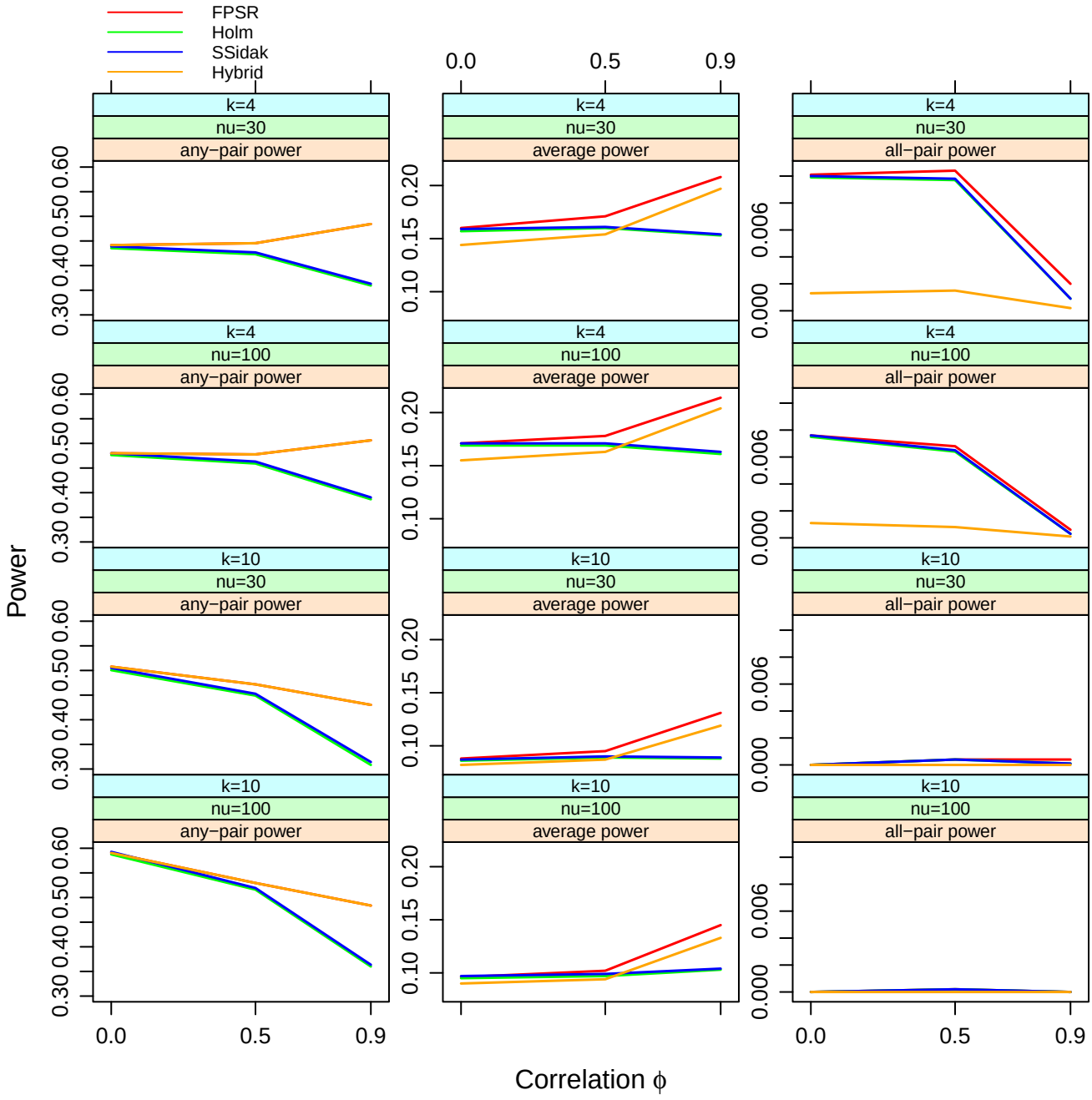


Figure A.16: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 1.5, when $\boldsymbol{V} = AR1(\phi)$.

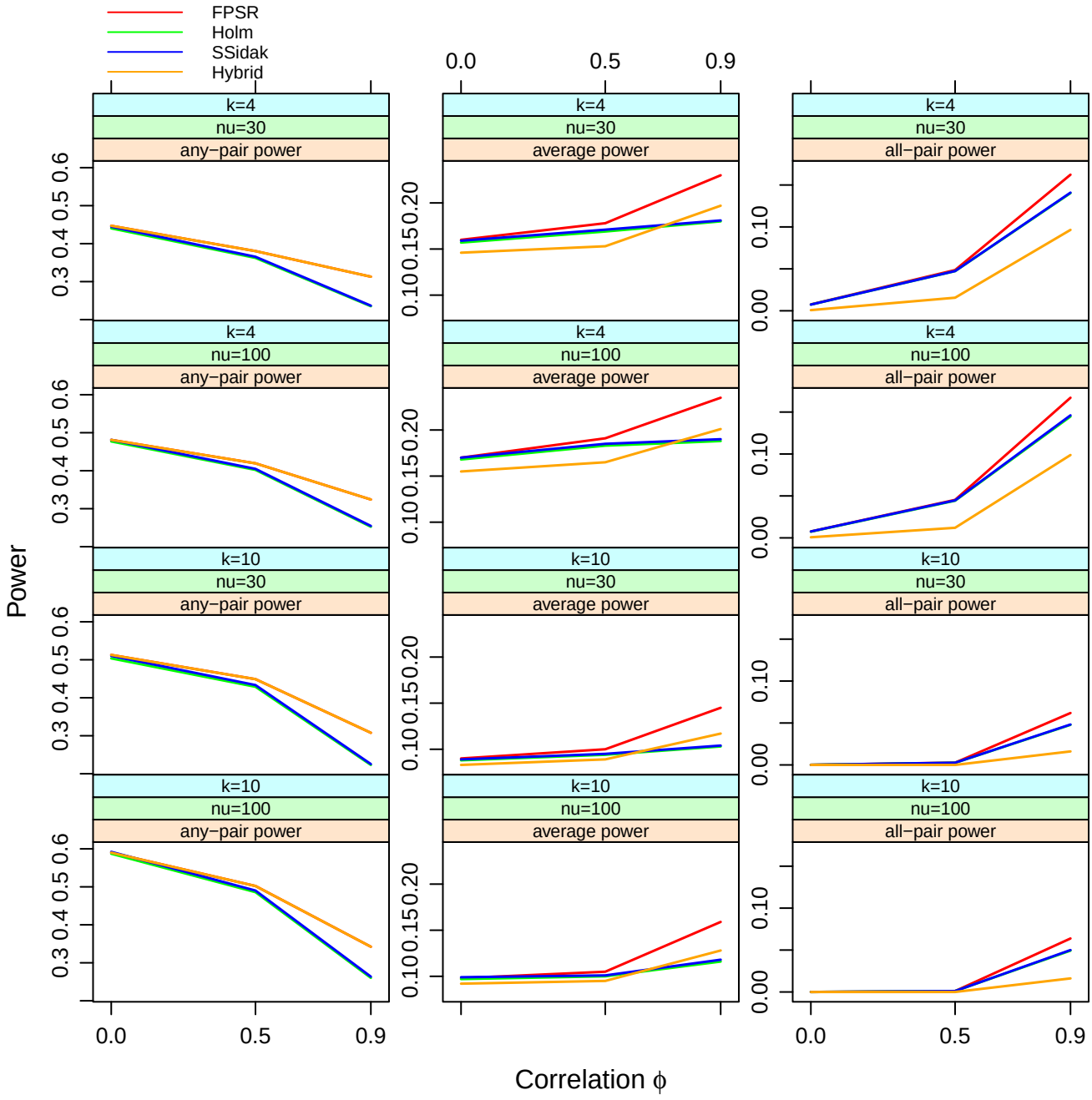


Figure A.17: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to 0, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

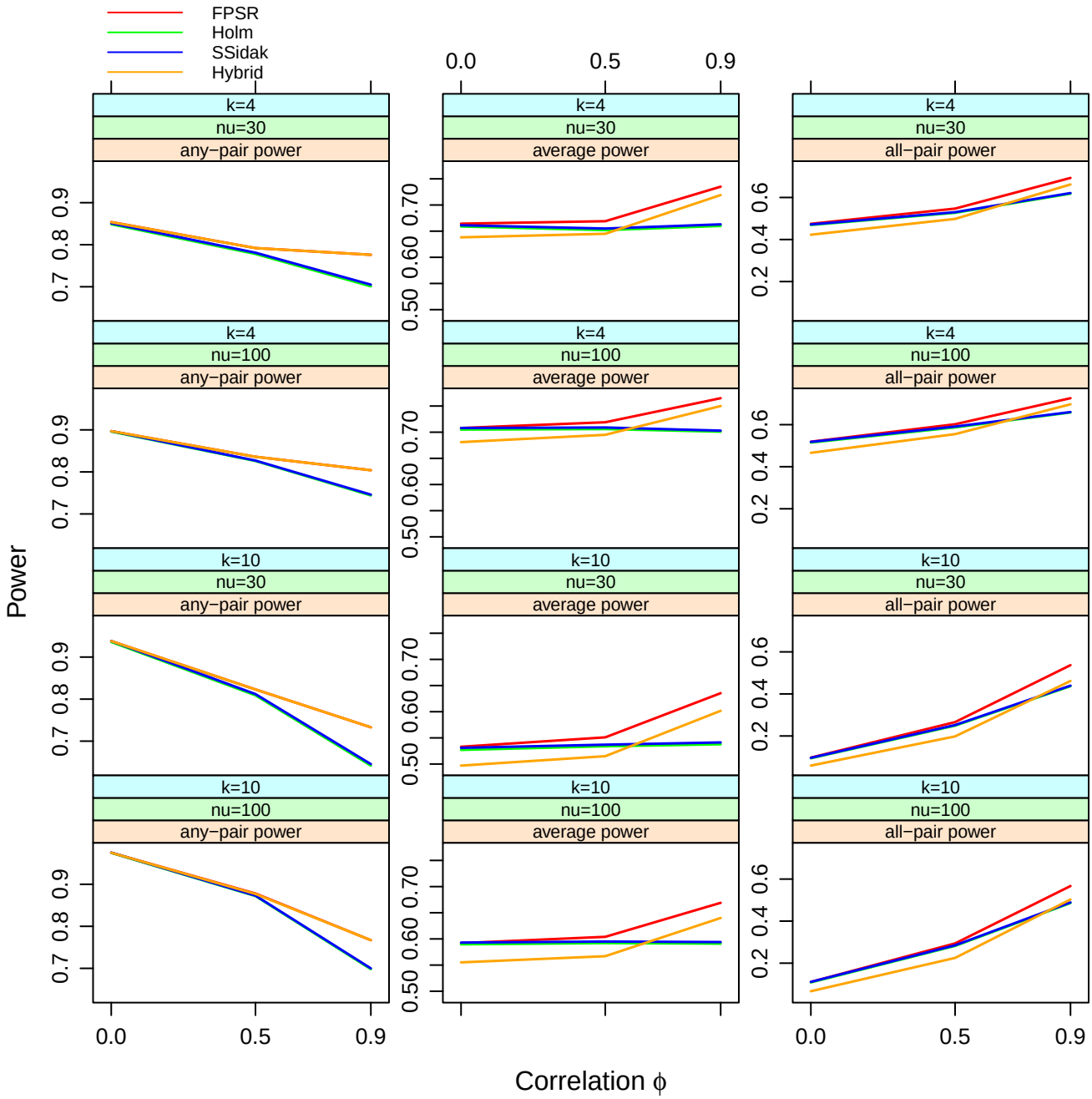


Figure A.18: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to -3, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

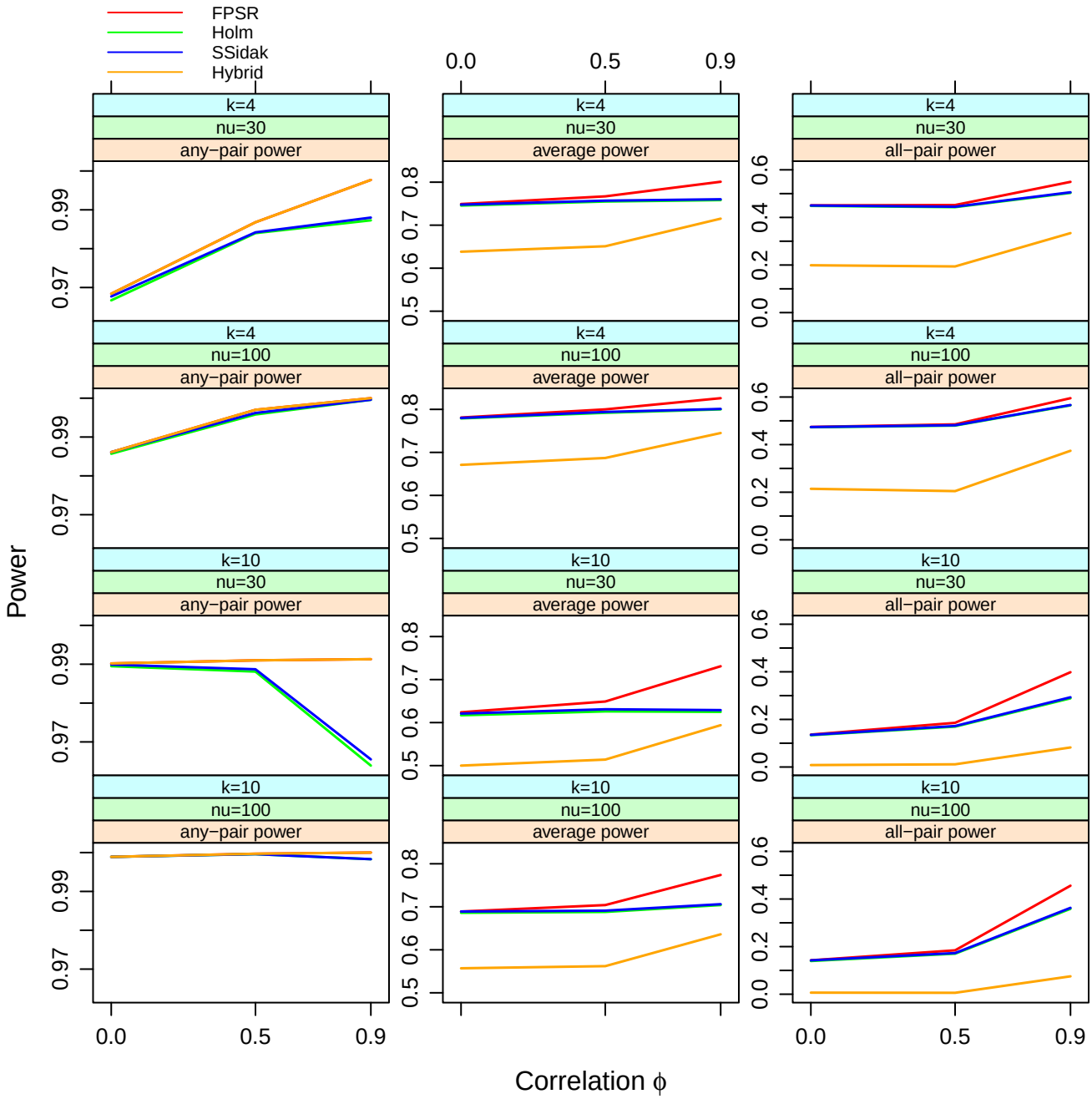


Figure A.19: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 3, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

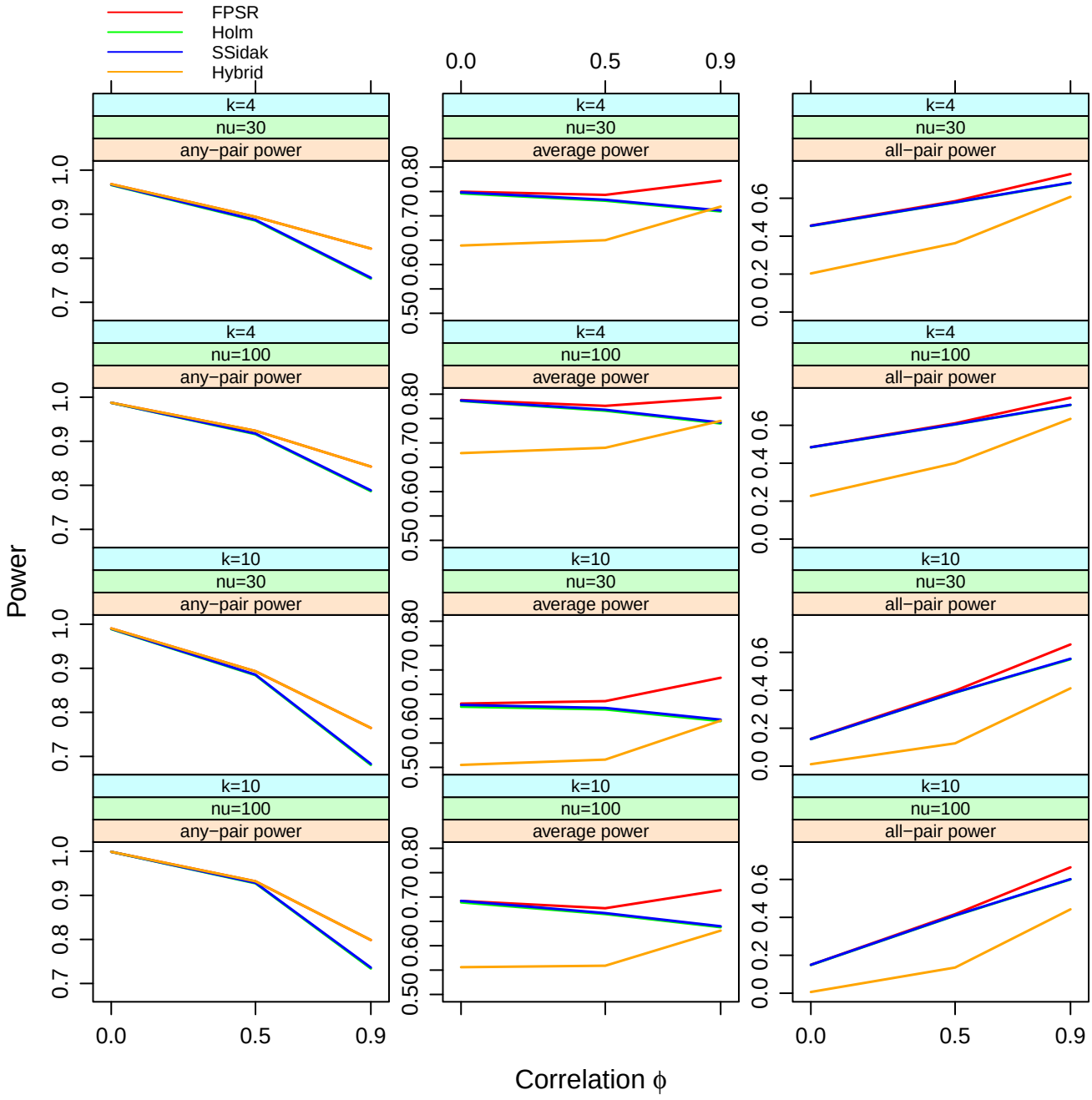


Figure A.20: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to 0, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

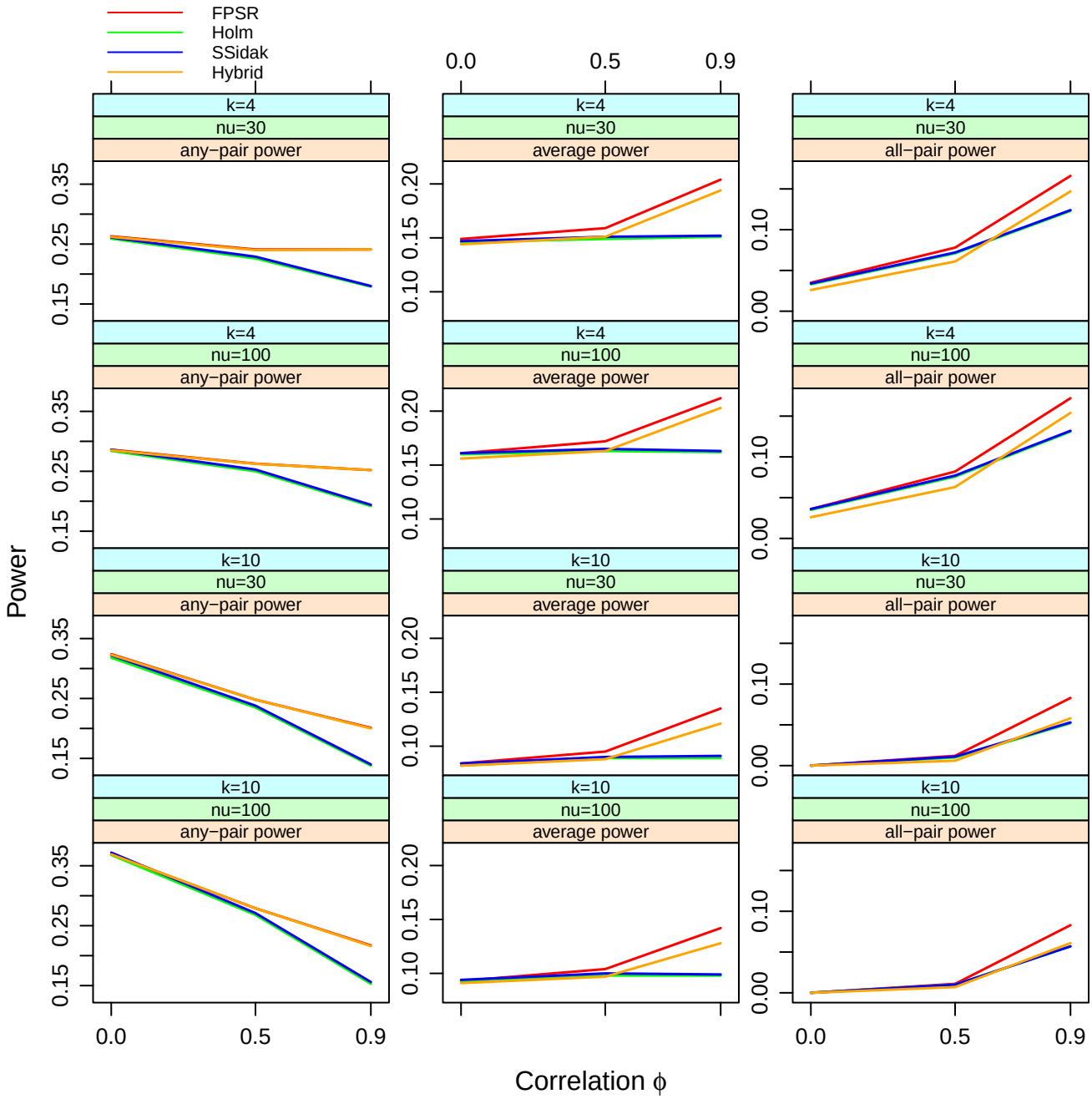


Figure A.21: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to -1.5, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

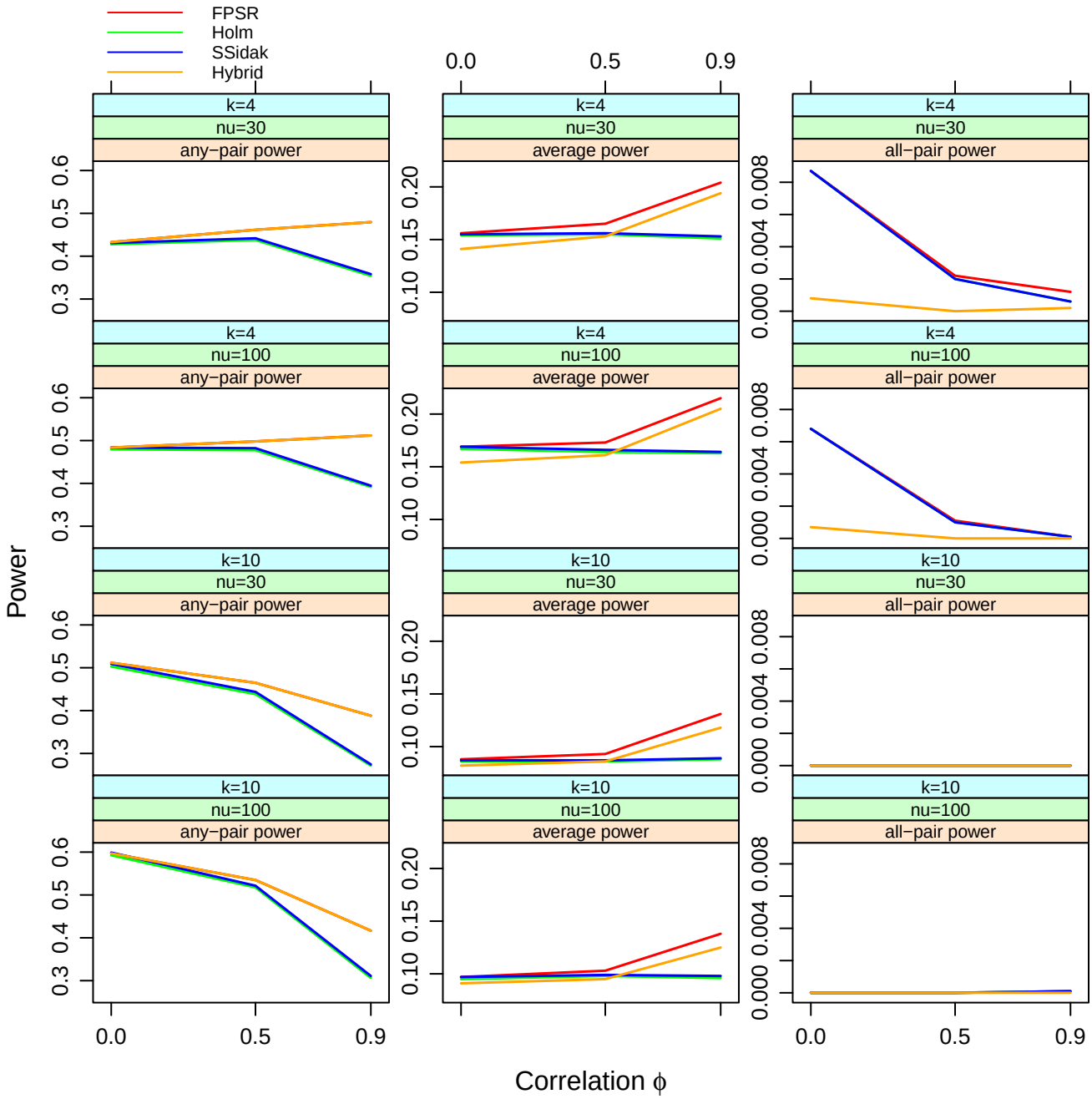


Figure A.22: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 1.5, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

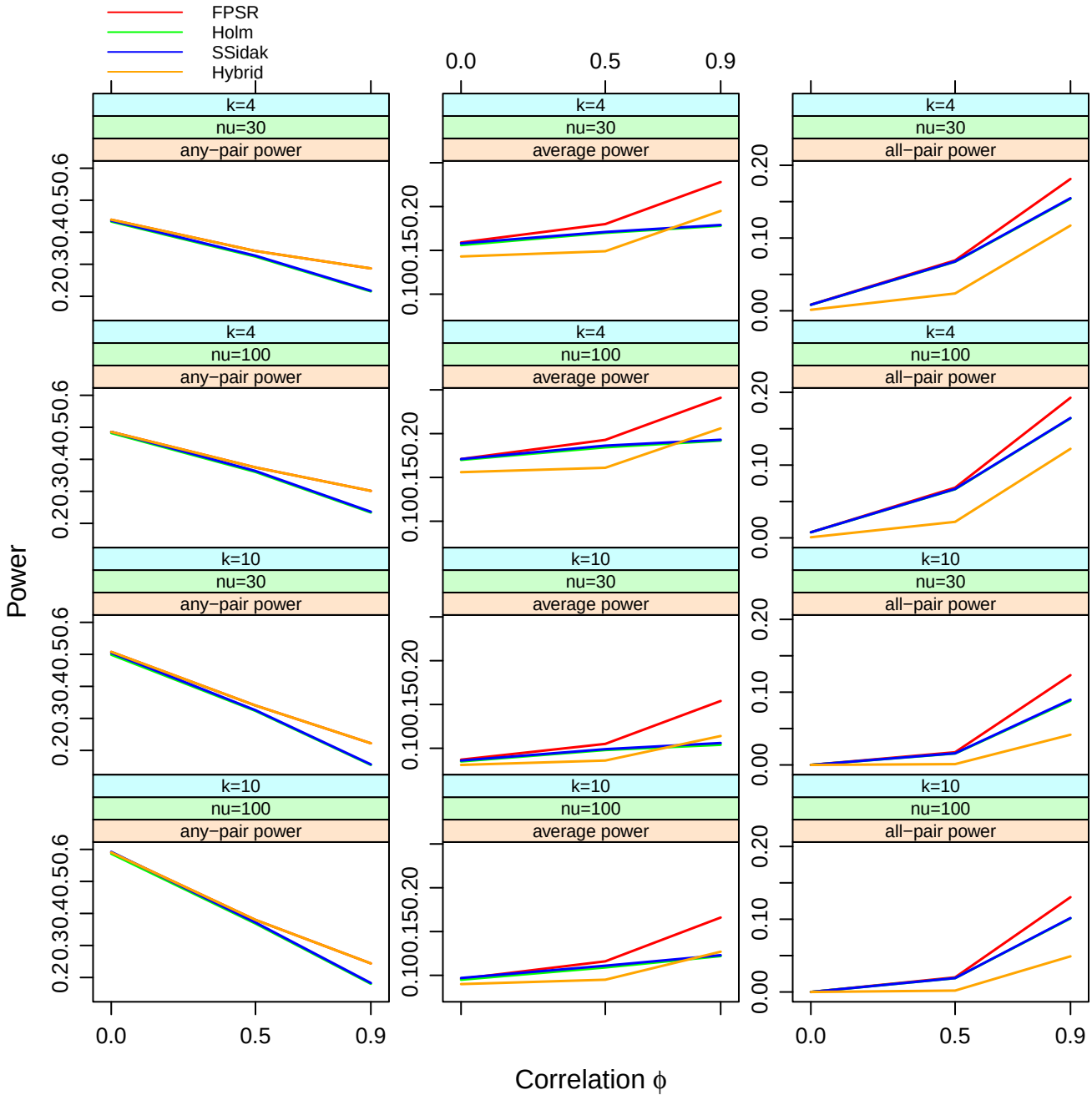
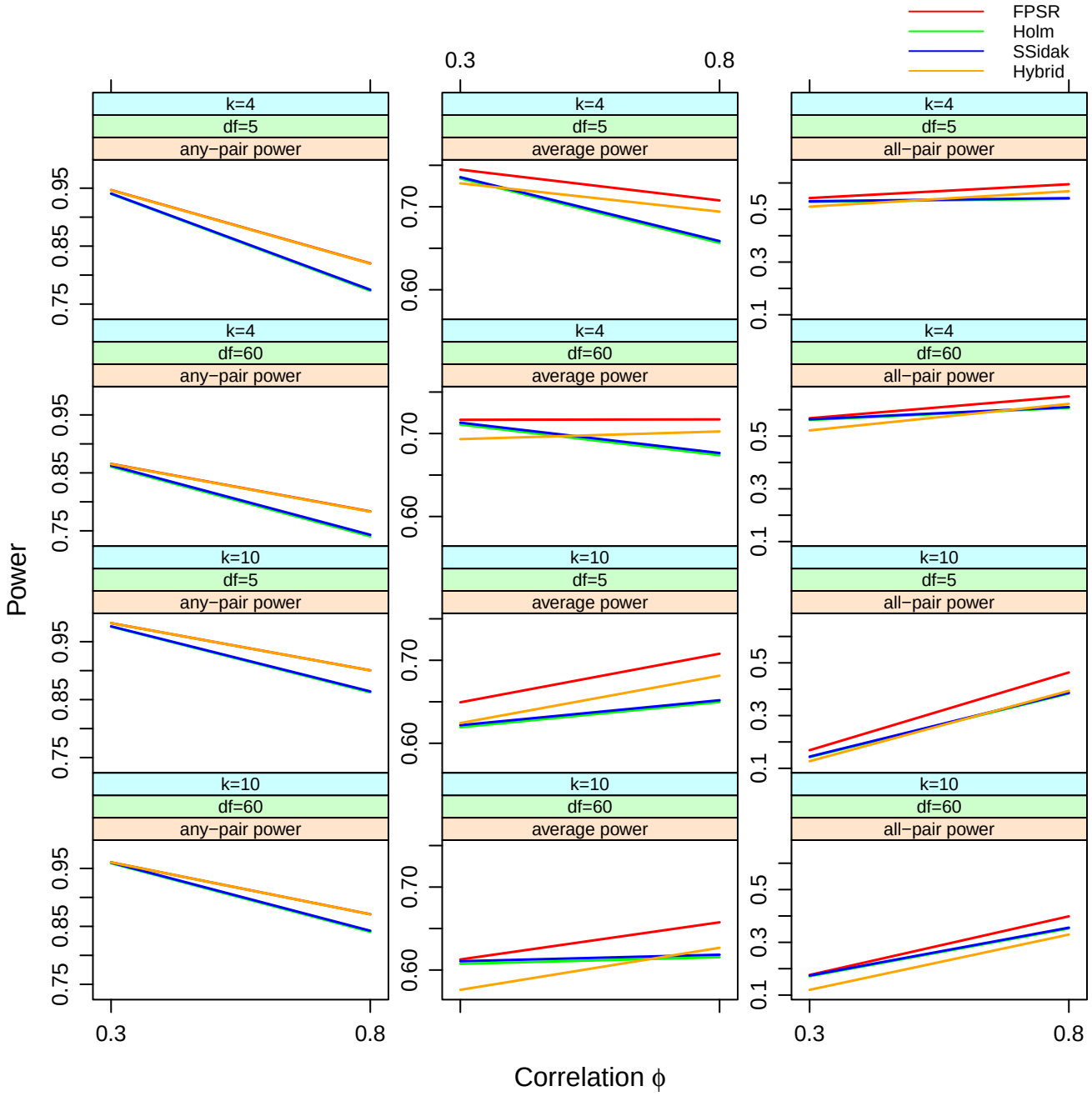
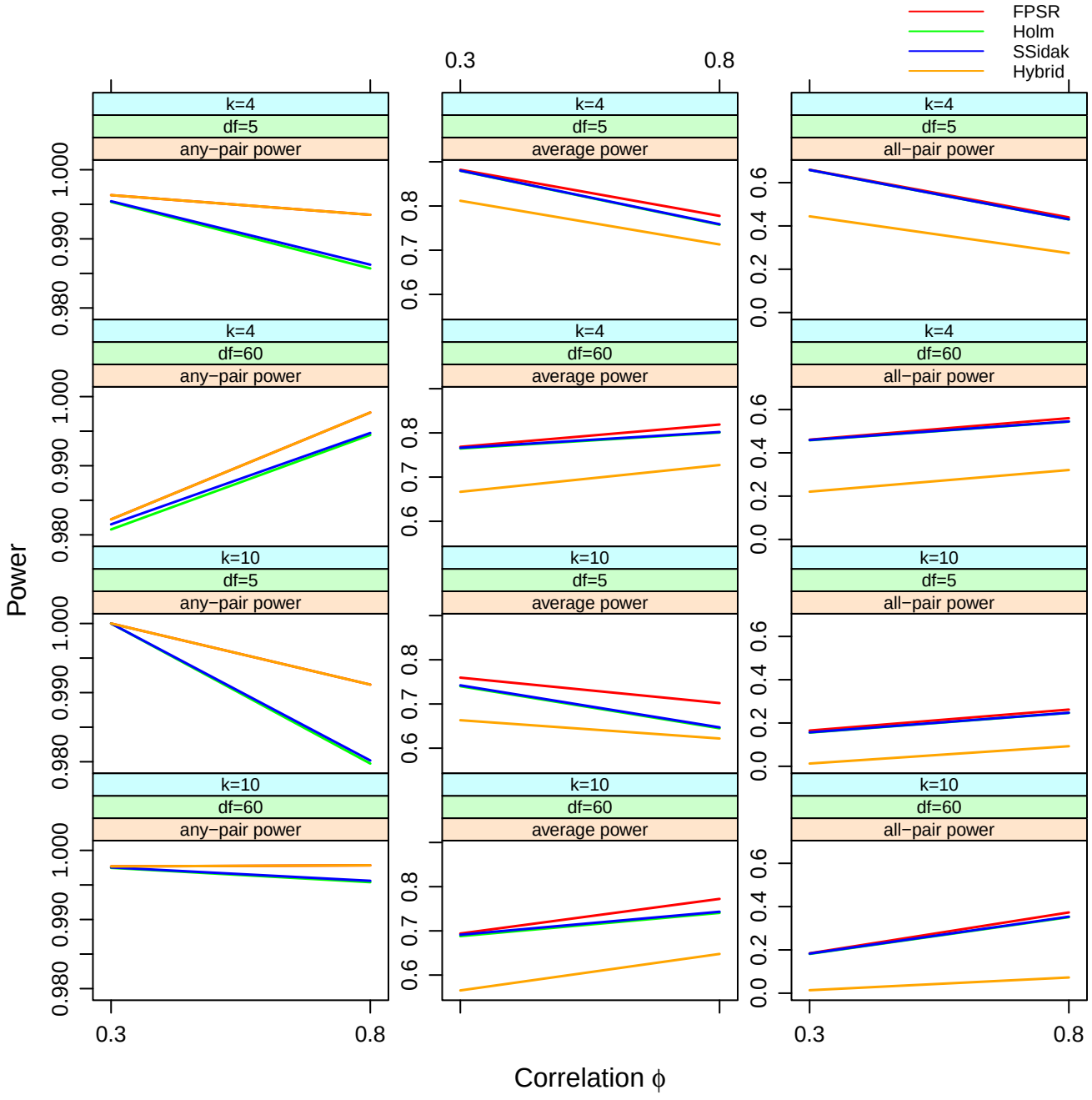


Figure A.23: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to 0, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



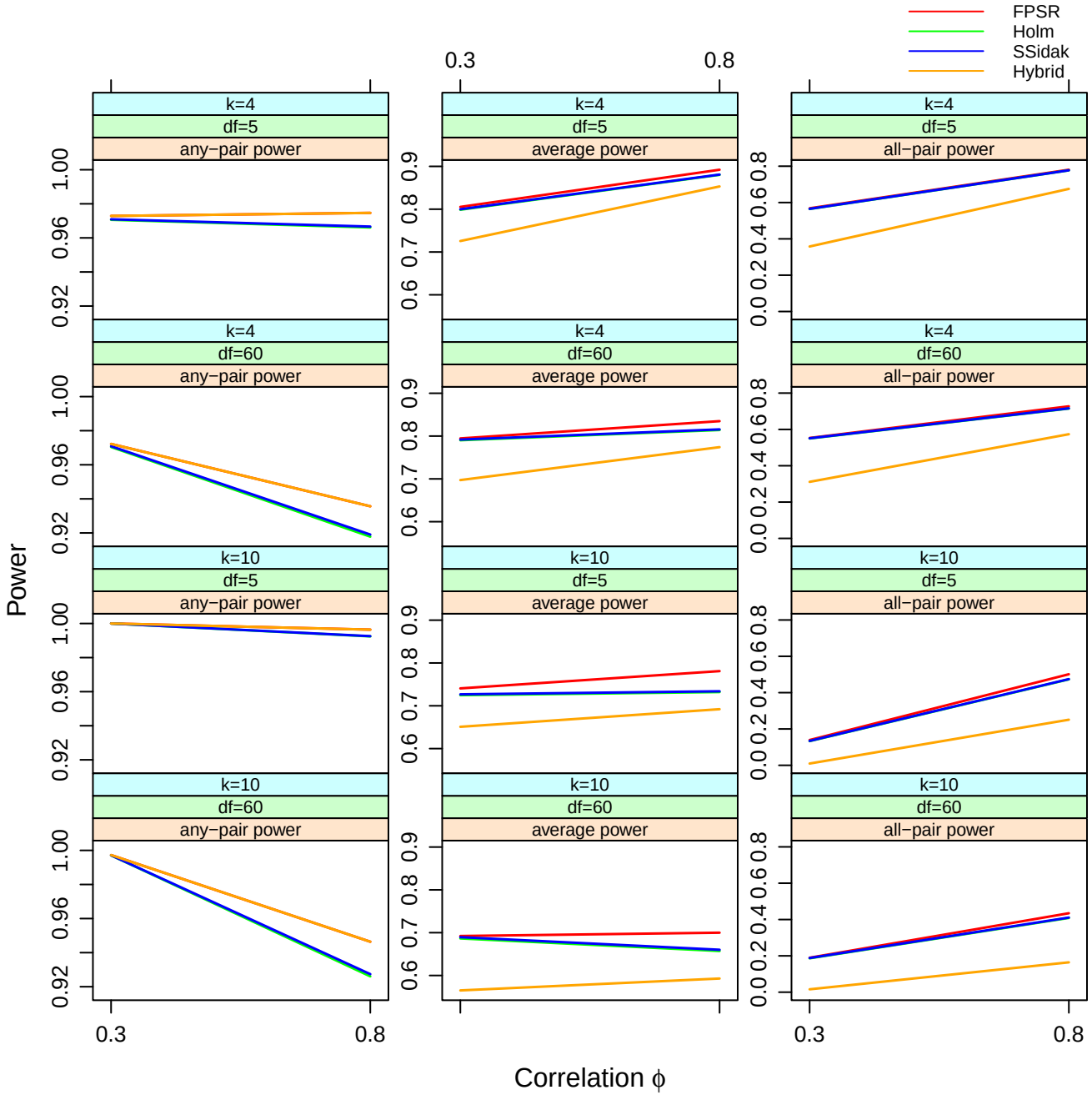
[a]The power measurements shown in Figure A.26 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.24: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to -3, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



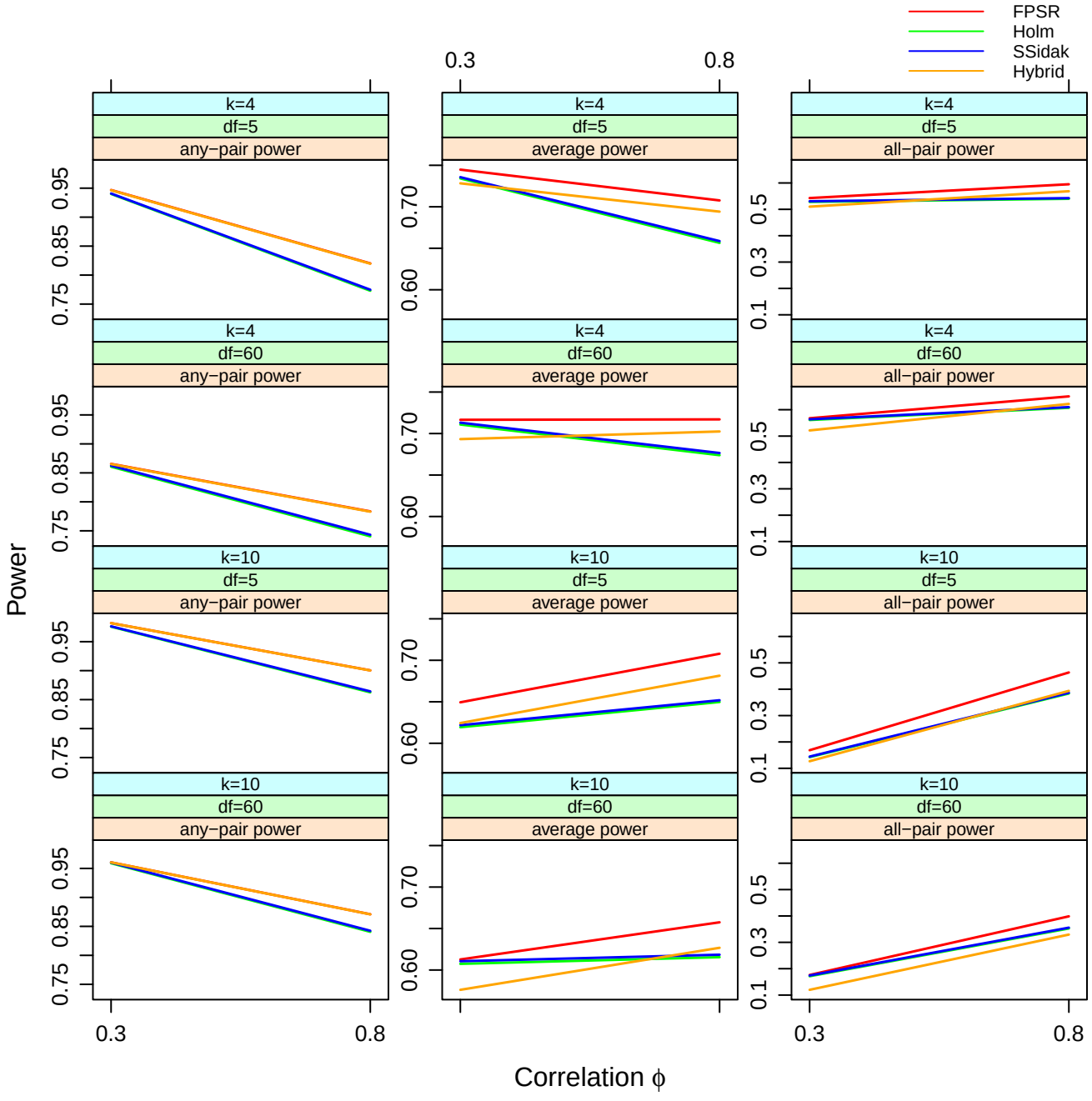
[a]The power measurements shown in Figure A.27 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.25: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 3, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



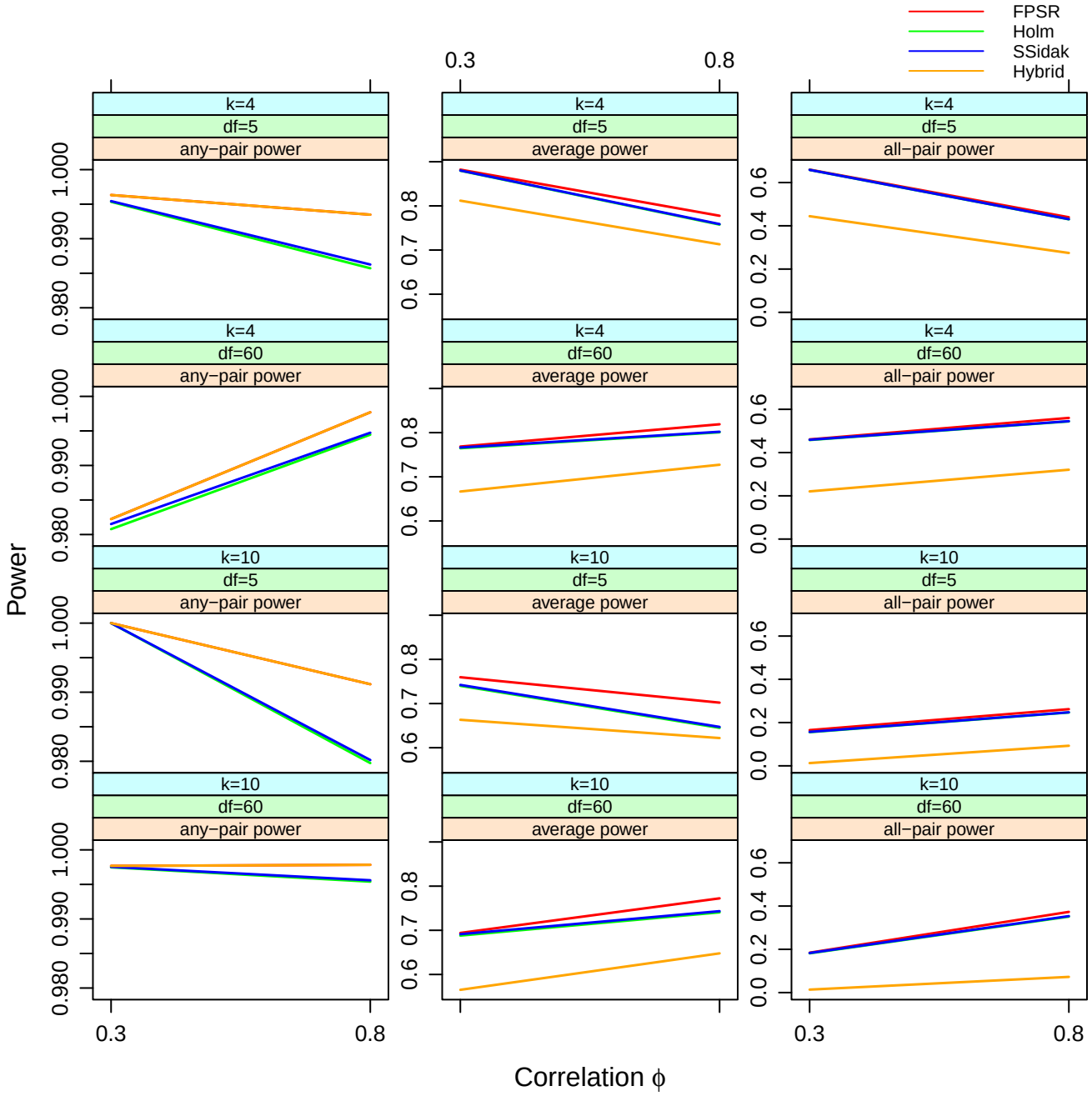
[a]The power measurements shown in Figure A.28 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.26: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to 0, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



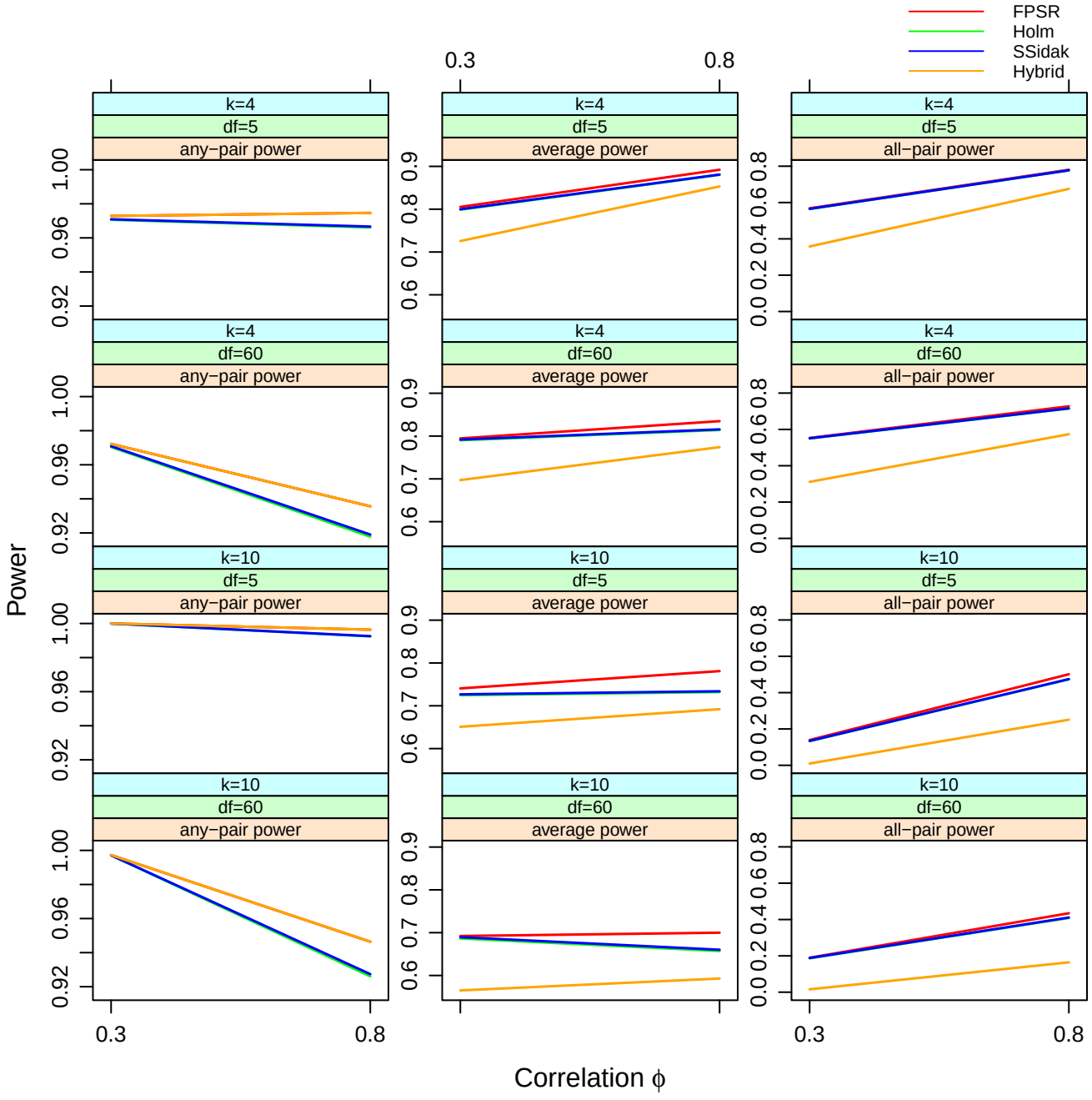
[a]The power measurements shown in Figure A.26 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.27: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 1.5 and half β_j's equal to -1.5, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



[a]The power measurements shown in Figure A.27 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.28: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 1.5, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



[a]The power measurements shown in Figure A.28 are the averages of the empirical power measurements obtained with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.29: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 0, when $\boldsymbol{V} = AR1(\phi)$.

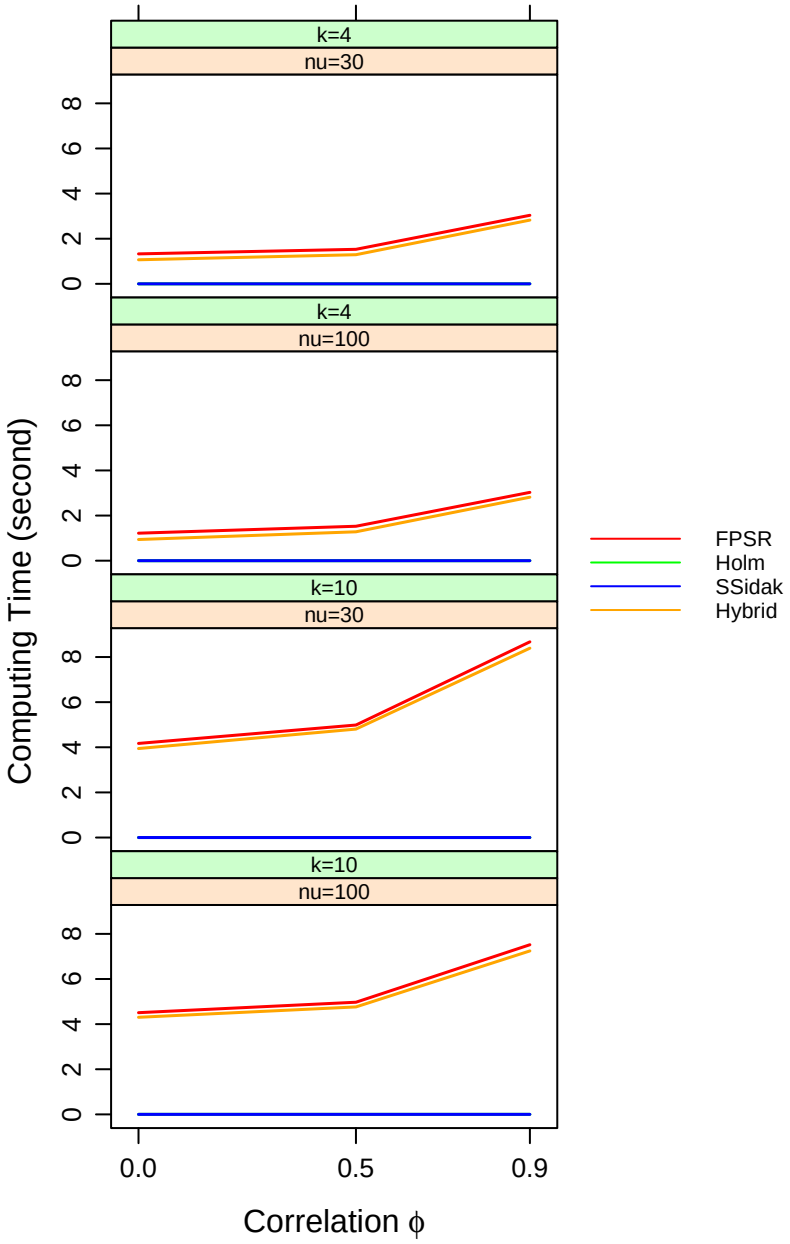


Figure A.30: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a nonzero value, when $\boldsymbol{V} = AR1(\phi)$.

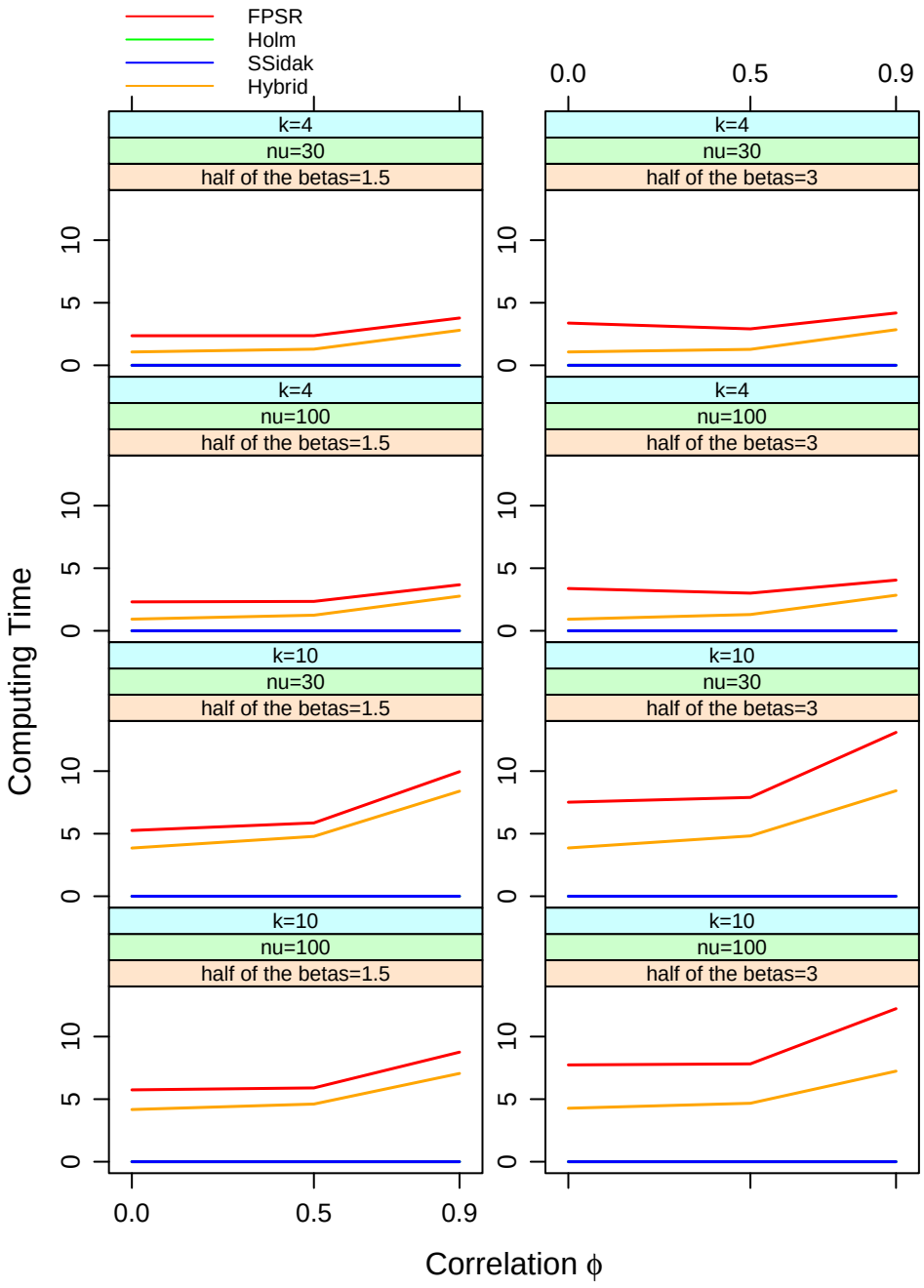


Figure A.31: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a positive nonzero value and half β_j's having the negative nonzero value, when $\boldsymbol{V} = AR1(\phi)$.

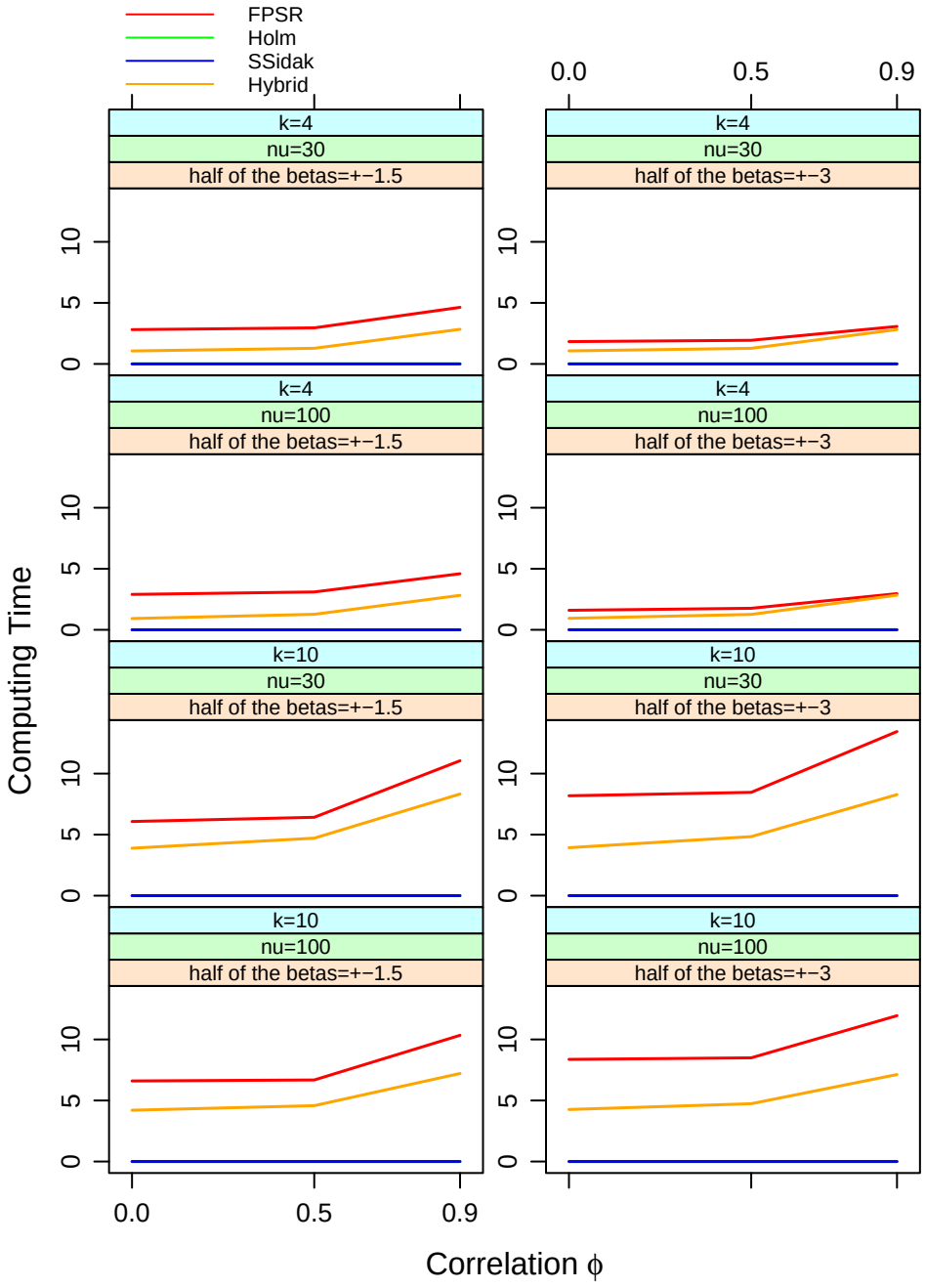


Figure A.32: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's having a positive value, when $\boldsymbol{V} = AR1(\phi)$.

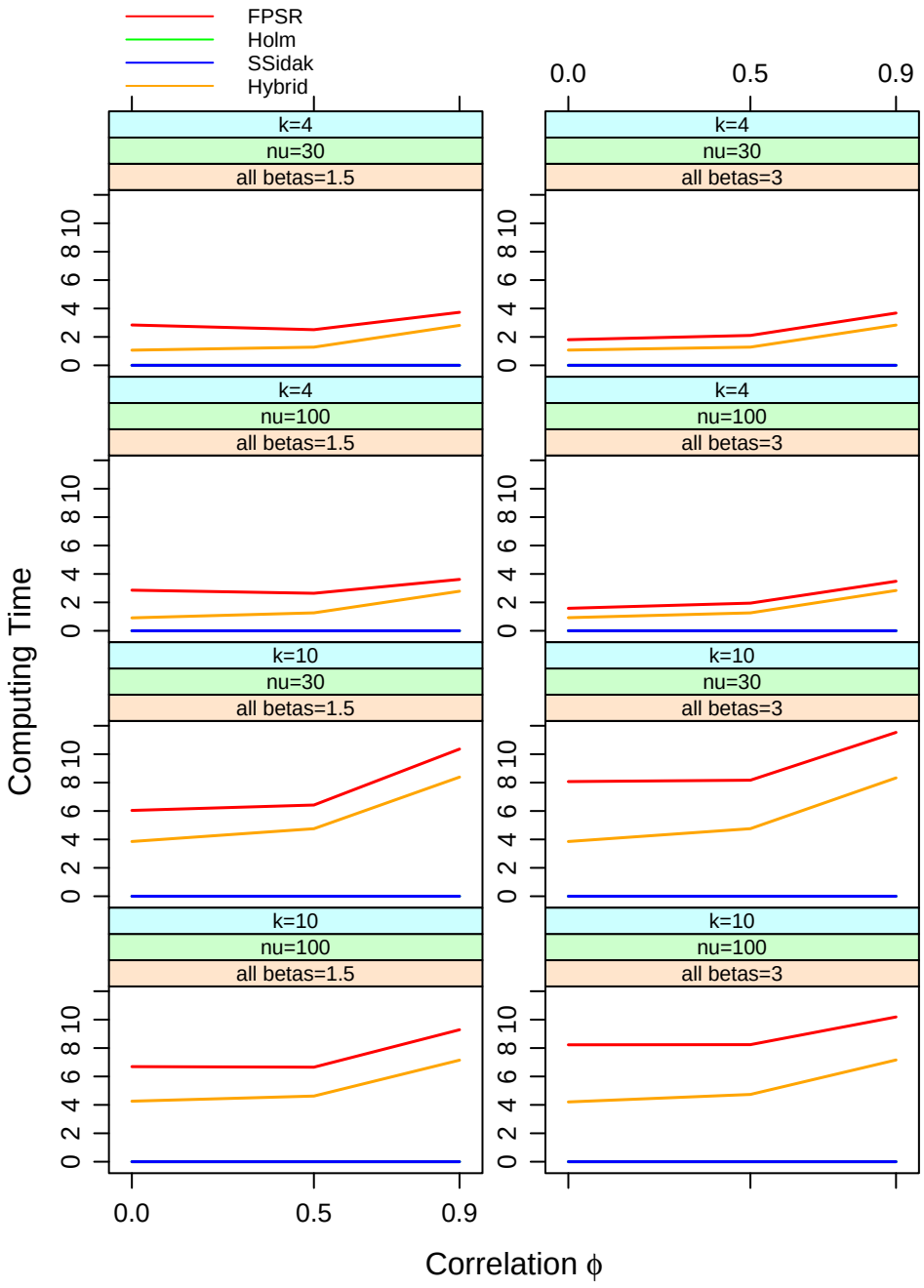


Figure A.33: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 0, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

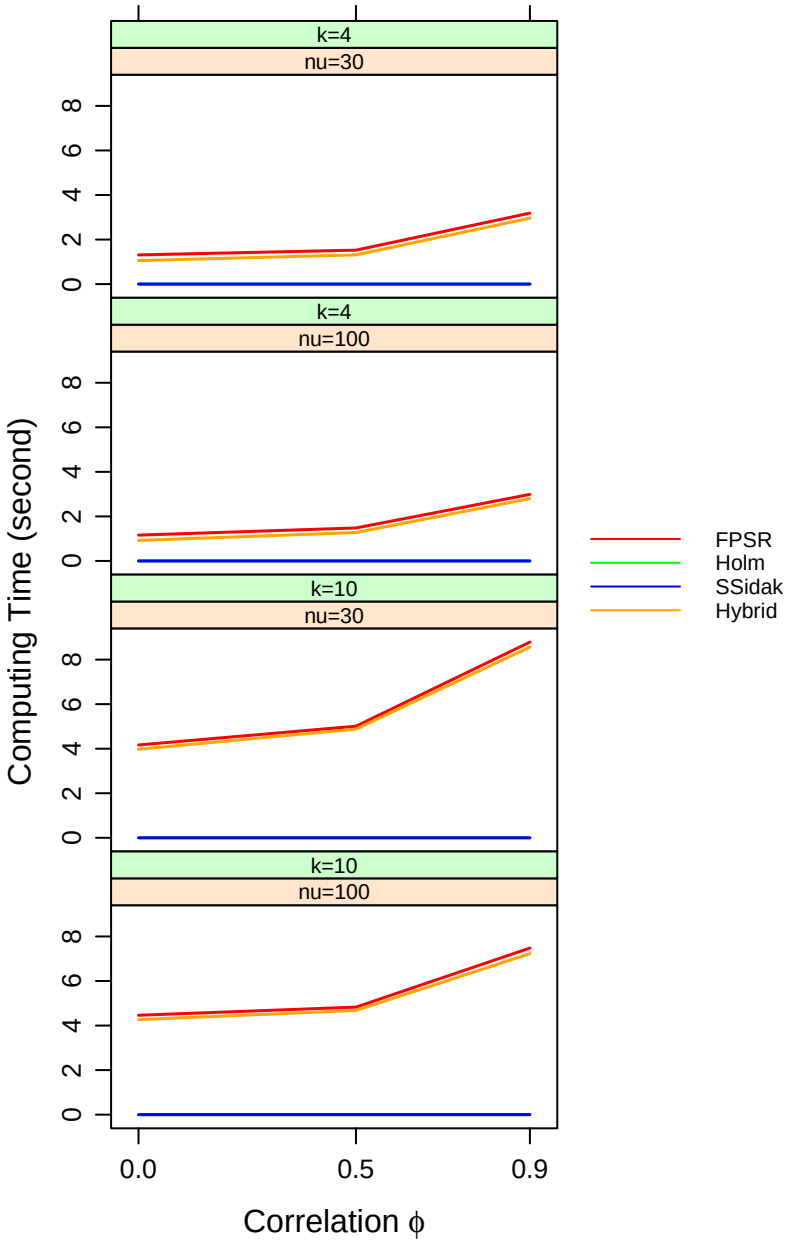


Figure A.34: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a nonzero value, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

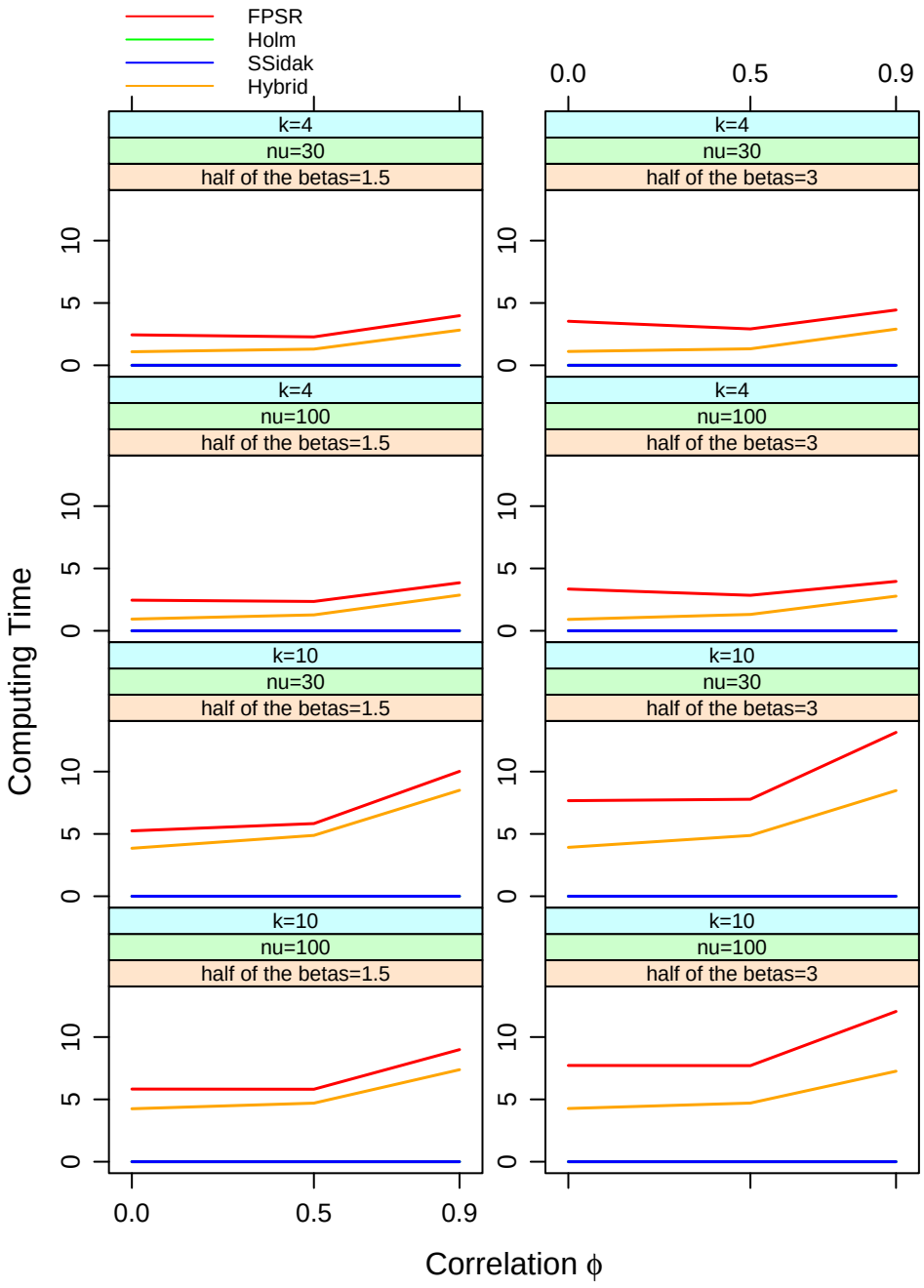


Figure A.35: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a positive nonzero value and half β_j's having the negative nonzero value, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

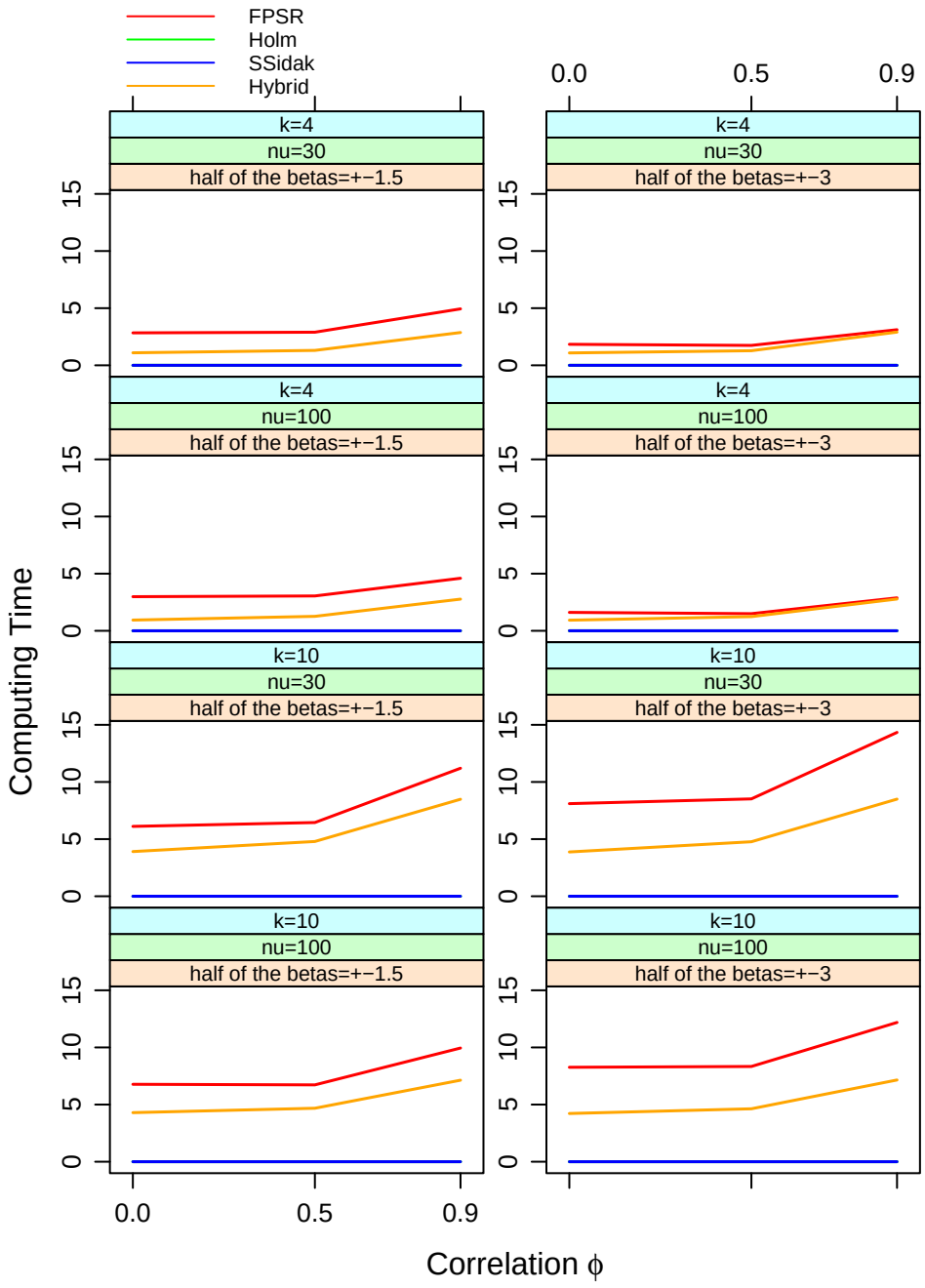


Figure A.36: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's having a positive value, when $\boldsymbol{V} = [v_{ij}]$, for $v_{ij} = \phi^{I\{i \neq j\}}$.

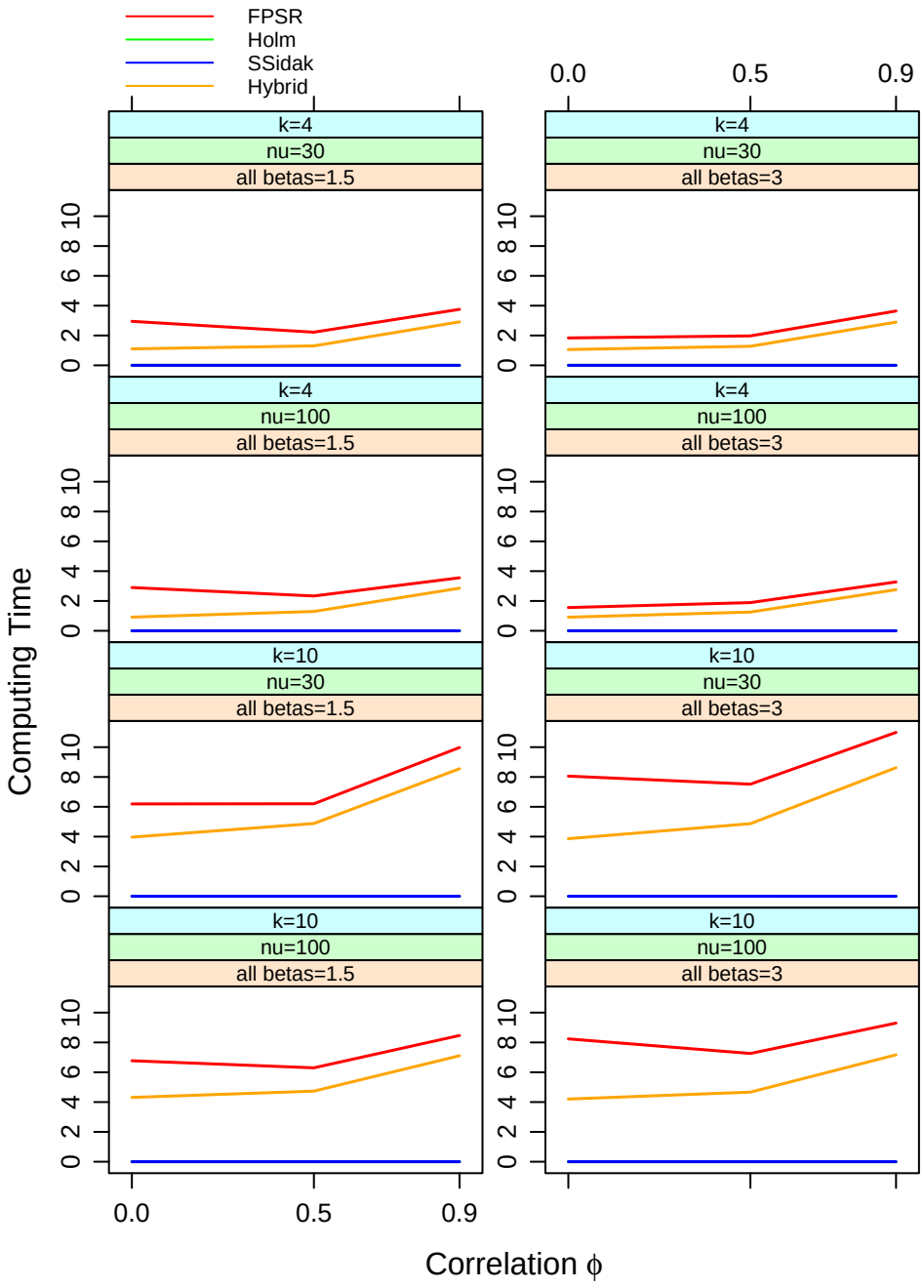
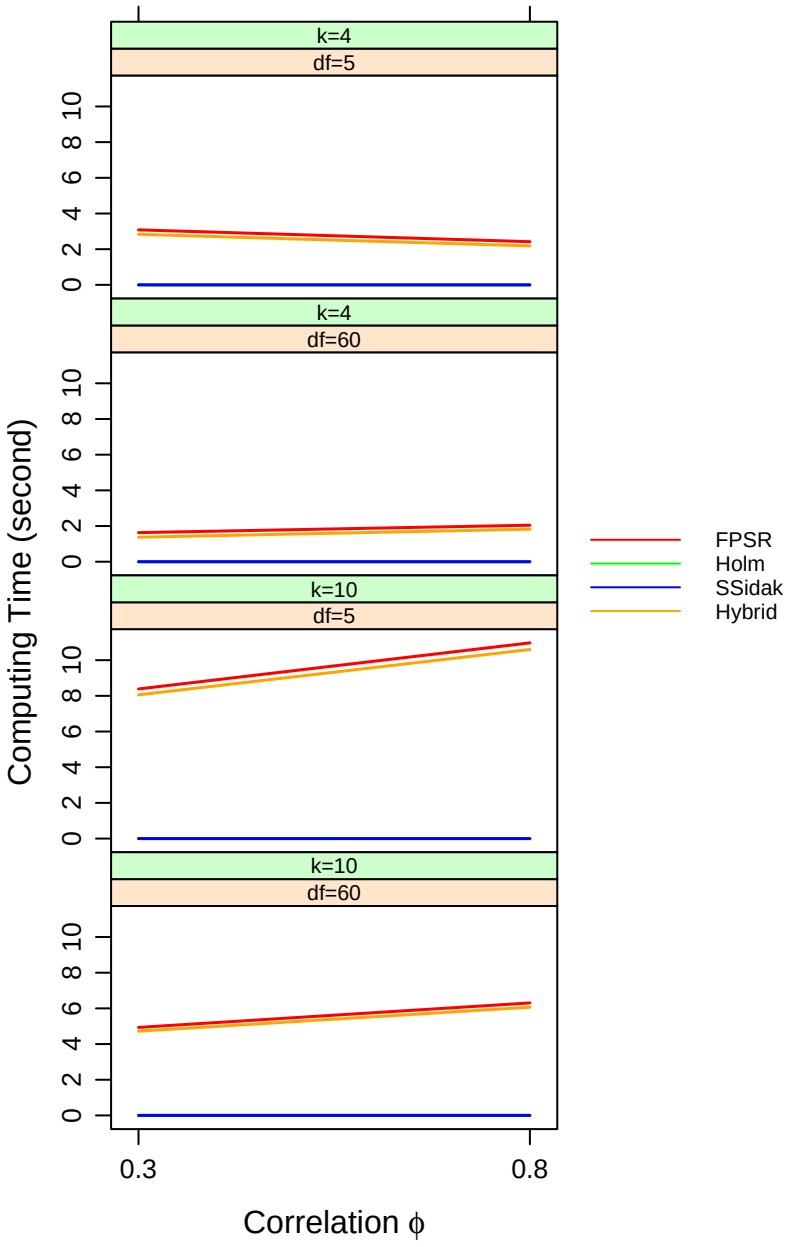
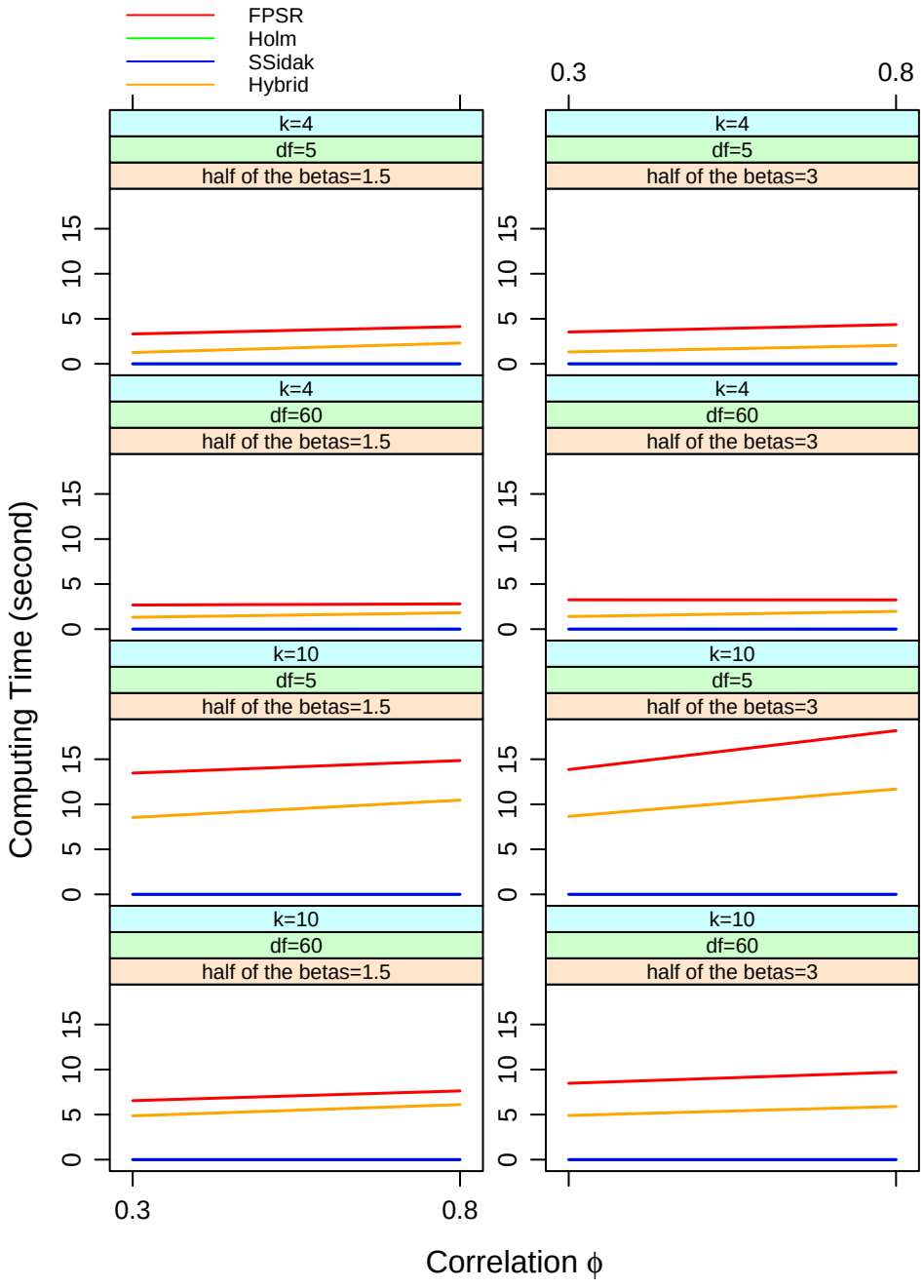


Figure A.37: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 0, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



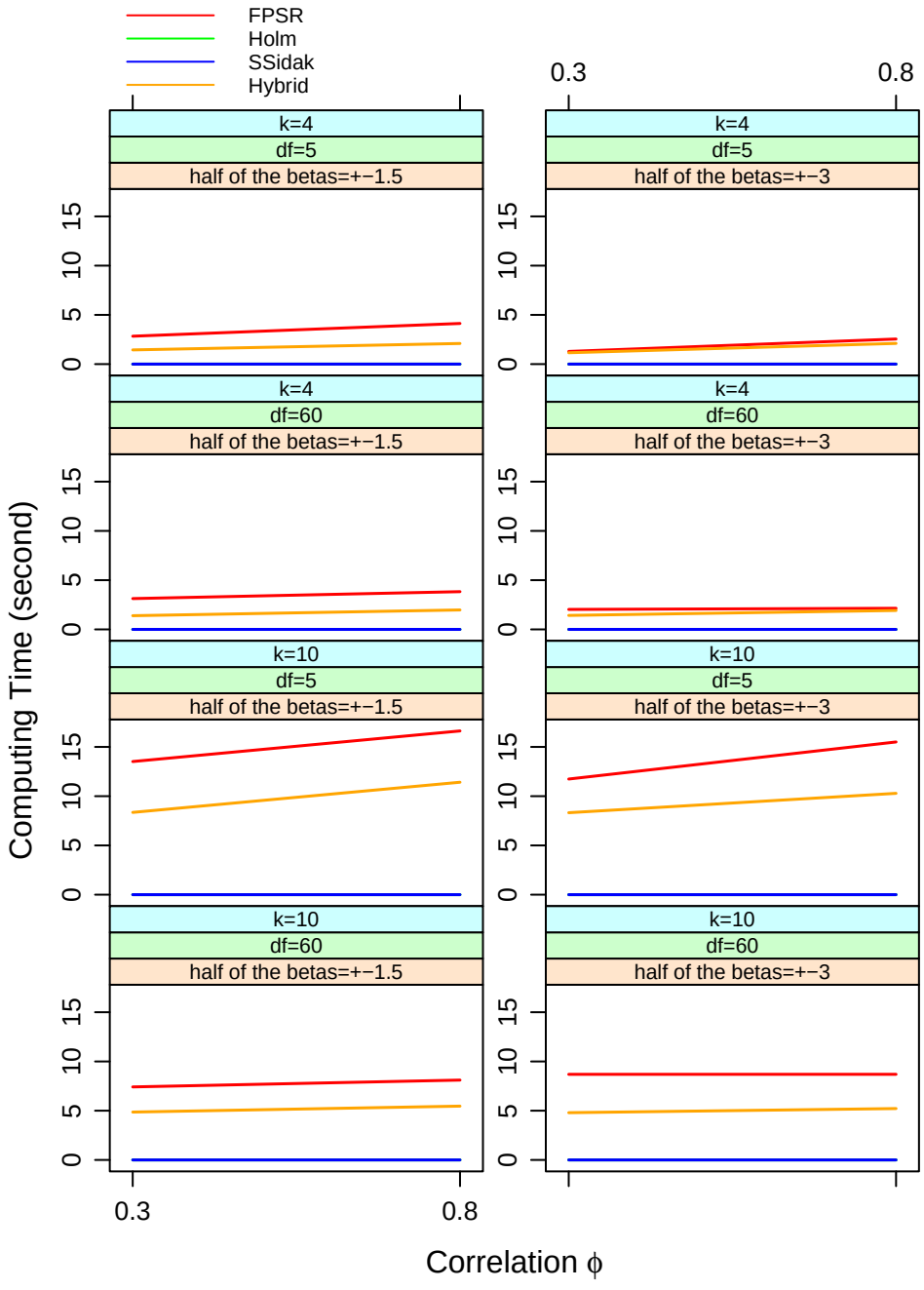
[a]The computing time shown in Figure A.37 are the averages of the obtained computing time with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.38: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a nonzero value, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



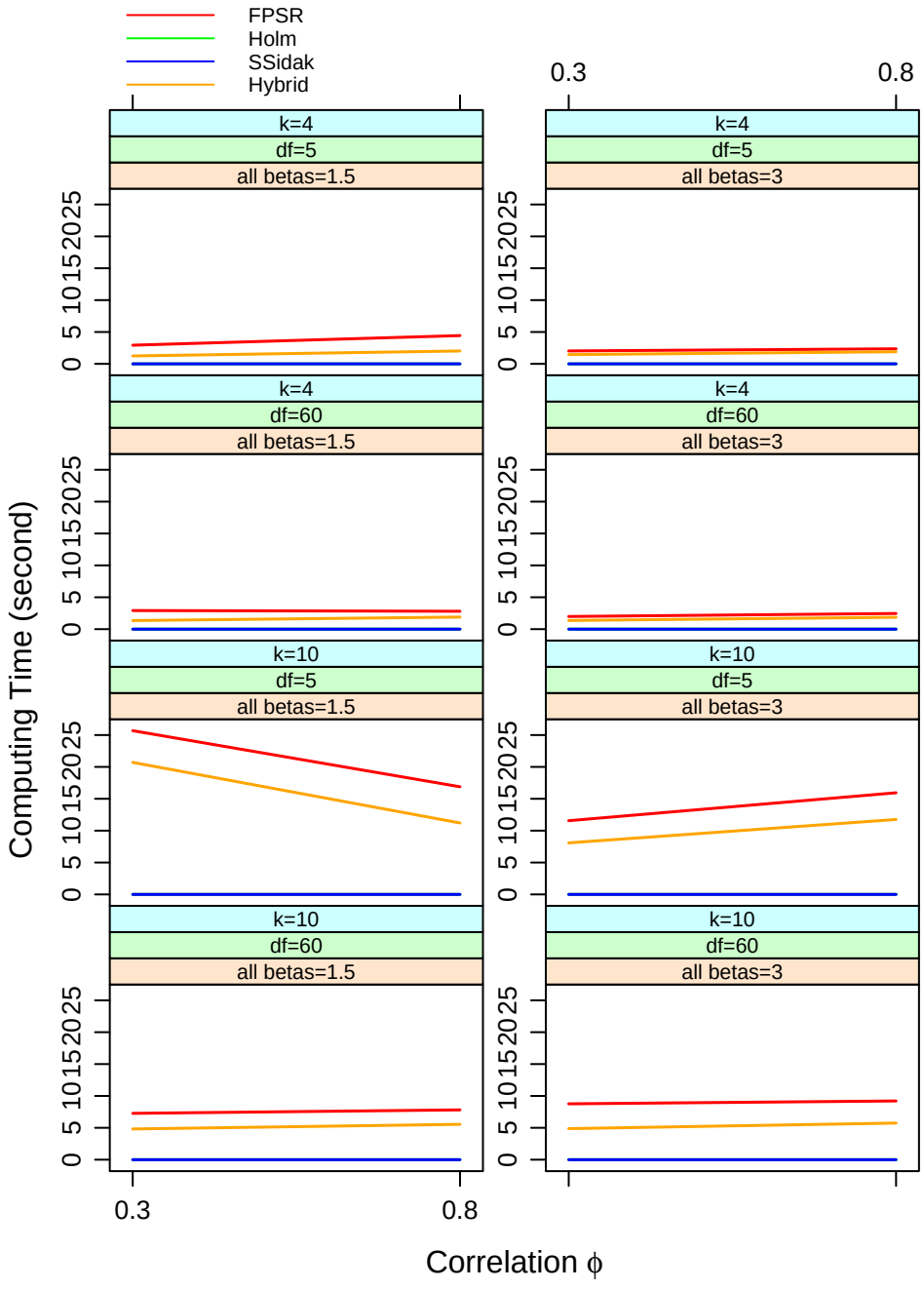
[a]The computing time shown in Figure A.38 are the averages of the obtained computing time with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.39: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's having a positive nonzero value and half β_j's having the negative nonzero value, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



[a]The computing time shown in Figure A.39 are the averages of the obtained computing time with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

Figure A.40: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's having a positive value, when $\boldsymbol{V}$ matrices are randomly generated from the Wishart distributions.[a]



[a]The computing time shown in Figure A.40 are the averages of the obtained computing time with five different randomly generated $\boldsymbol{V}$ matrices from the Wishart distributions.

B Appendix

Figure B.41: Empirical FWEs for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods when $\boldsymbol{V} = AR1(\phi)$, $k = 4$ and $\alpha = 0.1$.

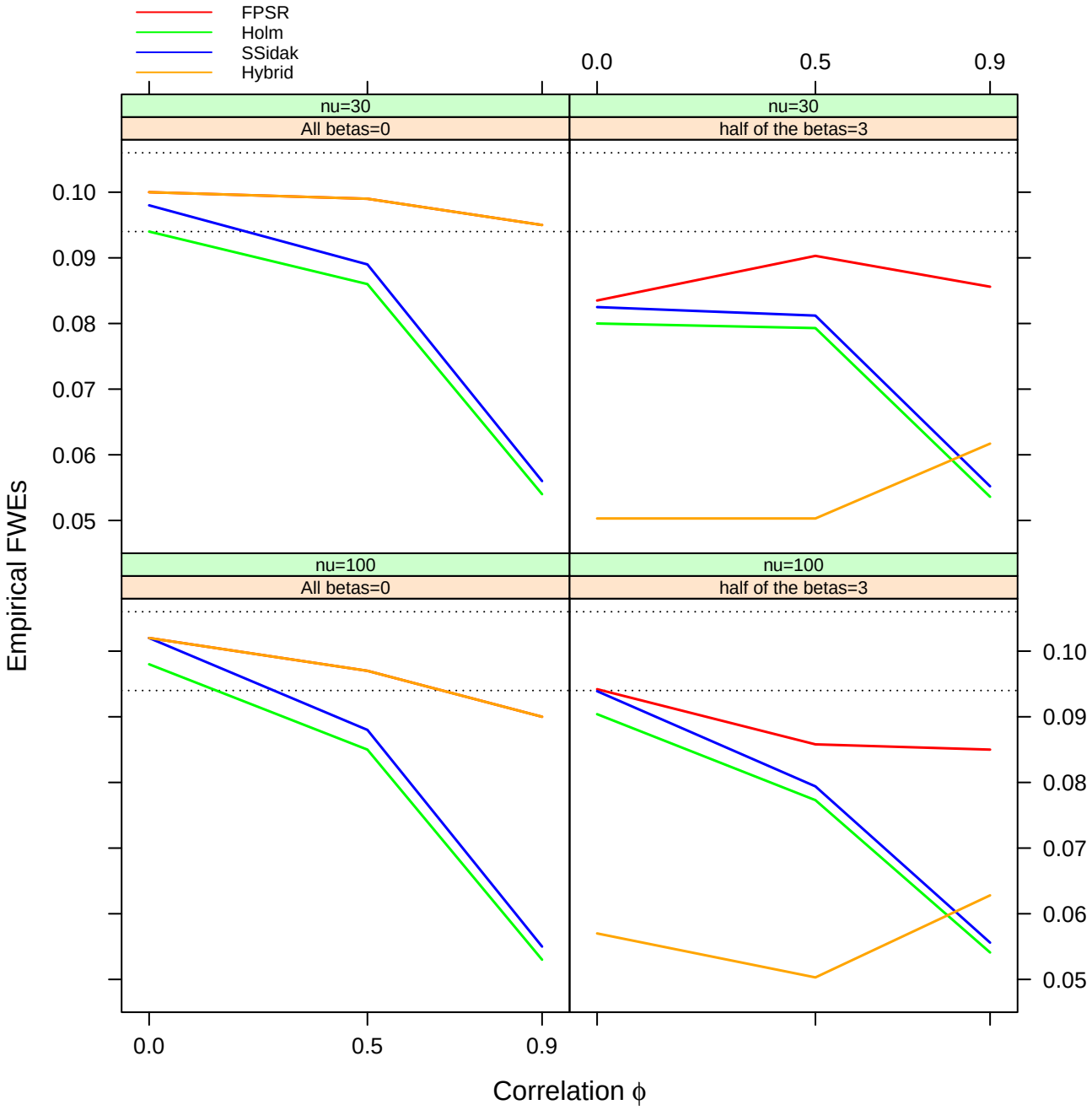


Figure B.42: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with half β_j's equal to 3 and half β_j's equal to 0, when $\boldsymbol{V} = AR1(\phi)$, $k = 4$ and $\alpha = 0.1$..

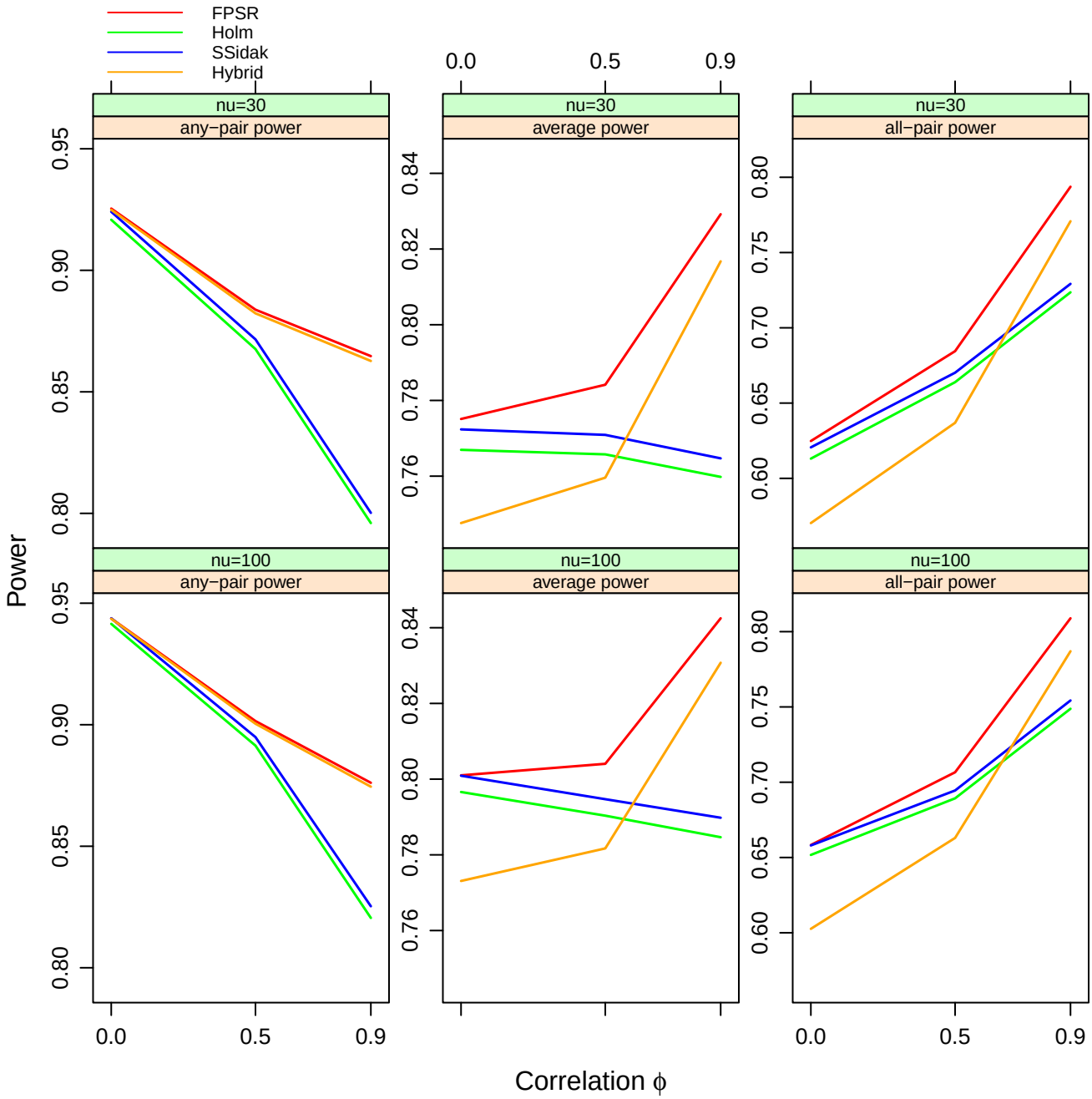


Figure B.43: Any-pair power, average proportion of true positives, and all-pair power for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods with all β_j's equal to 3, when $\boldsymbol{V} = AR1(\phi)$, $k = 4$ and $\alpha = 0.1$..

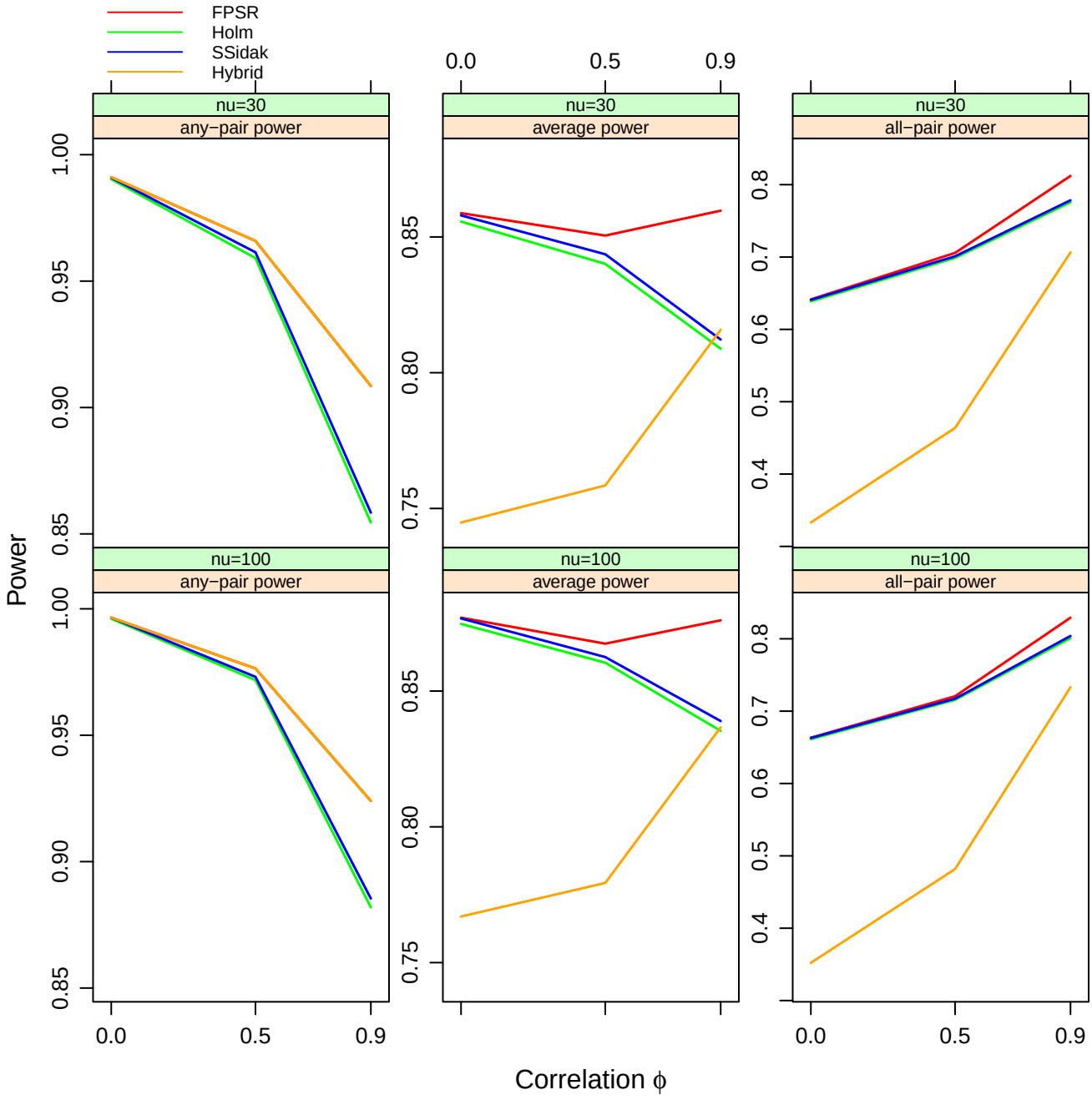
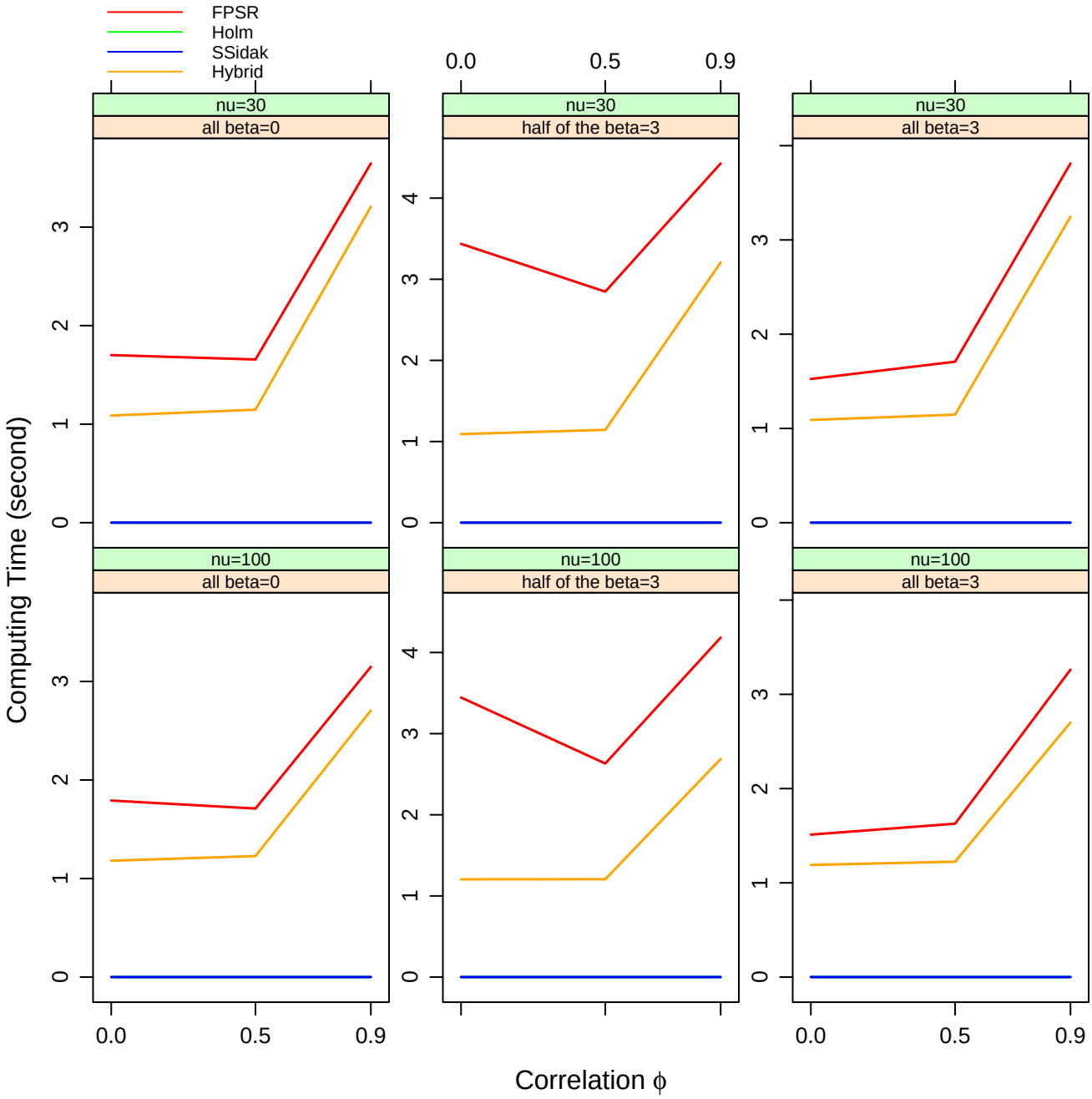


Figure B.44: Computing time for hybrid FPSR, Holm's procedure, stepdown Sidak and single-step hybrid methods when $\boldsymbol{V} = AR1(\phi)$, $k = 4$ and $\alpha = 0.1$..



VITA

YIMIN ZHANG

Candidate for the Degree of

Doctor of Philosophy

Thesis: SIMULTANEOUS CONFIDENCE INTERVALS USING THE RESTRICED SCHEFFE METHOD AND SEQUENTIAL TESTING PROCEDURES FOR SINGLE-STEP METHODS

Major Field: STATISTICS

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Statistics at Oklahoma State University, Stillwater, Oklahoma in May, 2014.

Completed the requirements for the Bachelor of Science in Computational Mathematics at Dalian University of Technology, Dalian, China in June, 2009.

Experience:

Instructor (Jan 2013 – May 2014) at Oklahoma State University: Be solely responsible for preparing and delivering lectures for STAT2013, create homework, quizzes and examinations, consult with students during office hours and via email, and manage all course grades.

Graduate Research Assistant (Aug 2012 – Dec 2012) at Oklahoma State University: Developed R programs for constructing and outputting simultaneous confidence intervals in the general multiple comparison setting and summarized the latest statistical literature on new theories and methods in multiple comparisons.

Graduate Teaching Assistant (Aug 2009 – Jul 2012) at Oklahoma State University: Graded homework assignments, and held office hours to answer questions on course content and aid in comprehension of course material.

Professional Memberships:

American Statistical Association, Institute of Mathematical Statistics