This is to certify that the

dissertation entitled

Conditional Inference For Incomplete Permutation Bootstraps In Multiple Linear Regression

presented by

Rudolf Bohumil Blazek

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor
Raoul LePage

Date December 18, 1998

Conditional Inference for Incomplete Permutation Bootstraps in Multiple Linear Regression

By

Rudolf Bohumil Blažek

A Dissertation

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Department of Statistics and Probability

December 21, 1998

Abstract

Conditional Inference for Incomplete Permutation Bootstraps in Multiple Linear Regression

By

Rudolf Bohumil Blažek

In this dissertation we develop a method for estimating conditional distributions of estimation errors for coefficients in multiple linear regression. These distributions are conditional on how the observation errors ϵ are presented to the model. The estimation of these distributions is achieved via incomplete permutation bootstrap of the observed residuals. We prove results ensuring that the incomplete permutation bootstrap distributions approximate the desired conditional distributions under very relaxed conditions. In particular, the key assumption needed is for the errors ϵ to have an exchangeable distribution. For the case of independent block permutations in a model with i.i.d. errors from the domain of attraction of a stable law with $\alpha < 2$ we prove an invariance principle assuring correctness of confidence regions based on the incomplete permutation bootstrap. There is an application of these methods to wavelets.

Mým rodičům

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my advisor Dr. LePage for introducing me to the subject studied in this dissertation, for his consistent encouragement, and for all he has taught me during the course of my graduate career. Without his help my studies and research would have been much more difficult.

I would also like to thank all other members of the Department of Statistics and Probability, with special thanks to members of my Guidance Committee, Dr. Gilliland, Dr. Levental, Dr. Salehi, and Dr. Nowak of the Department of Electrical and Computer Engineering. Also Dr. Fotiu of the Computer Laboratory Statistical Consulting Service for his support in my research.

No less important is my appreciation which I would like to express to my colleagues and friends who enlightened my personal life.

I would like to thank my parents for their spiritual support and wisdom they shared with me during all my life. Thanks to my sisters Zuzka and Hanka; and also to Wen-Ying.

Table of Contents

LIST OF FIGURES vi

1 Incomplete Permutation Bootstrap in Multiple Linear Regression 1

1.1 Introduction 1

1.2 Random Permutations in Multiple Linear Regression 9

1.3 Conditional Models 13

1.4 Incomplete Permutation Bootstrap 19

1.5 Independent Block Permutations 22

1.6 Examples 29

1.7 Asymptotics for Independent Block Permutations 39

1.8 Exact Confidence Intervals Based on Incomplete Permutation Tests . . 44

2 Wavelet Expansion Model 47

2.1 Introduction 47

2.2 Modified Discrete Wavelet Transform 48

2.3 Illustration 50

2.4 Conclusions - Incomplete Permutations of Residuals 57

APPENDICES 58

A Matrix formulation of Basic Results 58

BIBLIOGRAPHY 64

List of Figures

1.1 The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with i.i.d. errors from the domain of attraction of a symmetric stable law with $\alpha = 0.5$ (in particular, $\epsilon_i = \delta_i U_i^{-2}$ are reciprocals of squared i.i.d. uniform variables with independent symmetrical random signs). The estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations. 4

1.2 Comparison of the exact and incomplete permutation bootstrap approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with i.i.d. errors from the domain of attraction of a stable law with $\alpha = 0.5$ (see Figure 1.1). All estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations. 5

1.3 The conditional and unconditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. The corresponding independent block and full permutation bootstrap approximations were obtained via Monte–Carlo with 5000 replicas of the corresponding random permutations. 6

1.4 Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. (Data from Figure 1.3.) 7

1.5 Four point linear regression with a single outlier. The conditional and unconditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ and their corresponding independent block and full permutation bootstrap approximations. The approximations were obtained via Monte–Carlo with 5000 replicas of the corresponding random permutations. 34

1.6 The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. The corresponding independent block and full permutation bootstrap approximations were obtained using Monte–Carlo with 5000 replicas of the corresponding random permutations. (Same as Figure 1.3) 35

1.7 Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a four point linear regression with a single outlier. (Data from Figure 1.5.) 36

1.8 Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. (Same as Figure 1.4, see data in Figure 1.6.) 36

1.9 The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with i.i.d. Gaussian errors. The corresponding independent block and full permutation bootstrap approximations were obtained using Monte–Carlo with 2000 replicas of the corresponding random permutations. 37

1.10 The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with i.i.d. errors from the domain of attraction of a symmetric stable law with $\alpha = 0.5$ (in particular, $\epsilon_i = \delta_i U_i^{-2}$ are reciprocals of squared i.i.d. uniform variables with independent symmetrical random signs). The estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations. (Same as Figure 1.1) 38

2.1 Mexican hat mother wavelet $\psi(t) = (1 - t^2)e^{-t^2/2}$. 51

2.2 Wavelets $\psi_{0,-5}$ and $\psi_{0,5}$ based on Mexican hat mother wavelet. 51

2.3 The conditional distribution of the estimation errors for the coefficient θ_2 of $\psi_{0,-5}$, given the order statistics of the errors $\boldsymbol{\epsilon}$(left), and its estimate based on full permutation bootstrap of the observed residuals (right). 54

2.4 The incomplete permutation bootstrap approximation of the distribution of the estimation errors for the coefficient θ_2, conditional on the fact that the errors with large absolute values align with the last $n/2$ components of $\boldsymbol{\epsilon}$. In the simulation study the components of $\boldsymbol{\epsilon}$ have been sorted to satisfy this condition. 54

2.5 The approximation of the conditional distribution of the estimation errors for the coefficient θ_2, given that the errors with large absolute values lie among the first $n/2$ components of $\boldsymbol{\epsilon}$. In the simulation study the components of $\boldsymbol{\epsilon}$ have been sorted to satisfy this condition. 55

2.6 For a vector of errors with randomly arranged components, the block-wise conditional distribution of the estimation errors for the coefficient θ_2, given the order statistics of the errors in each of the two blocks are well recovered by the incomplete permutation bootstrap of the observed residuals. In this particular case the true block-wise conditional distribution appears similar to the conditional distribution given the order statistics of $\boldsymbol{\epsilon}$ (see Figures 2.3 and 2.7 for comparison). 55

2.7 Comparison of conditional distributions of the estimation errors for the coefficient θ_2 and the corresponding incomplete permutation bootstrap recoveries. Full permutation bootstrap was used for the distribution conditional on the order statistics of $\boldsymbol{\epsilon}$. 56

Chapter 1

Incomplete Permutation Bootstrap in Multiple Linear Regression

1.1 Introduction

One of the common goals in multivariate linear models $\boldsymbol{y} = X \cdot \boldsymbol{\beta} + \boldsymbol{\epsilon}$ is to find simultaneous confidence regions for a subset of components of the unknown vector $\boldsymbol{\beta}$. An important consideration is that the relative performance of various estimators and designs should be capable of evaluation *before* taking data.

Introduced as a *descriptive* method in [8], permutation bootstrap of the observed residuals was shown by the authors of [17] and [18] to provide approximate *conditional* confidence intervals for the (*contrast*[1]) components of $\boldsymbol{\beta}$, given the *order statistics* of the components of the vector of errors $\boldsymbol{\epsilon}$. A very important feature of these *conditional* methods is that they may be applied to all but the constant term *without moment assumptions* on exchangeable errors.

The key idea to be developed in this dissertation is that a linear form $\sum v_i \epsilon_{\pi_i}$

[1]A vector $\boldsymbol{v} \in \boldsymbol{R}^d$ will be called a contrast if $\boldsymbol{v} \cdot \mathbf{1} = 0$. Here $\mathbf{1}$ represents a vector with all components equal to 1.

may for outsized errors $\{\epsilon_i\}$ be greatly changed depending upon how the outsized ϵ_{π_i} line up with v_i under a random permutation π uniform over the group Σ_n of all permutations of $\{1, \ldots, n\}$.

We shall show that the estimation of particular conditional distributions (associated with how $\boldsymbol{\epsilon}$ aligns with a contrast $\boldsymbol{v}$) can be obtained via (incomplete) random permutation of the observed residuals over a subgroup Λ of the permutation group Σ_n of indices $\{1, \ldots, n\}$. A very special example of this would be to use independent blockwise permutations[2] uniformly distributed over the indices of vectors $\boldsymbol{v}^1$ and $\boldsymbol{v}^2$, respectively, for a vector

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{pmatrix}$$

where both $\boldsymbol{v}^1$ and $\boldsymbol{v}^2$ are contrasts, and the l_2–norm $\|\boldsymbol{v}^1\|$ is relatively small in comparison to $\|\boldsymbol{v}^2\|$.

It is then of interest to study the conditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ conditional on which block $\boldsymbol{v}^1$ or $\boldsymbol{v}^2$ the extreme values of the errors align with. Figure 1.1 illustrates this with long–tailed errors $\epsilon_i = \delta_i U_i^{-2}$ (where $\delta_1, \ldots, \delta_n$ is a Rademacher sequence and $U_1, \ldots, U_n$ are i.i.d. uniform random variables independent of the random signs δ_i) in a simple linear regression $\boldsymbol{y} = \alpha + \beta\boldsymbol{x} + \pi\boldsymbol{\epsilon}$ with $\boldsymbol{x} = (-10, -9.5, \ldots, 10)'$. Here the centered vector $\boldsymbol{x}$ plays the role of the contrast

[2]The idea of blockwise random permutations appears in many contexts, for example in restricted randomizations (e.g. Lehmann [14, Chapter 5, Section 9]). If the variation among the experimental units is excessive, then a conventional test will have small power. As a remedy Lehmann examines restricted randomization analogous to stratified sampling:

> ... The experimental material is divided into subgroups, which are more homogeneous than the material as a whole, so that within each group the differences among the u's (units) are small. In animal experiments, for example, this can frequently be achieved by a division into litters. Randomization is then applied within each group. ...

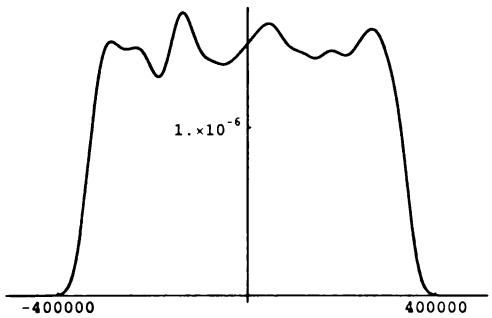
The focus is thus on "matching designs" required for Fisher's exact tests. In particular, the contrasts considered by Lehmann are restricted to indicators of treatment group.

$\boldsymbol{v}$ with $\boldsymbol{v}^1$ consisting of the 7 smallest and 7 largest components of $\boldsymbol{x}$ (see (1.40) on page 32). In the second row the figure shows the distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ conditional on the event that the components of $\boldsymbol{\epsilon}$ with large absolute values align themselves with the block $\boldsymbol{v}^1$ of extreme values of the vector $\boldsymbol{x}$, while the third row illustrates the conditional distribution given that the extreme values of $\boldsymbol{\epsilon}$ align instead with the block $\boldsymbol{v}^2$. The full permutation bootstrap approximation and the unconditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ are shown for comparison in the first row. Figure 1.2 compares the true conditional distributions to their incomplete permutation bootstrap approximations.

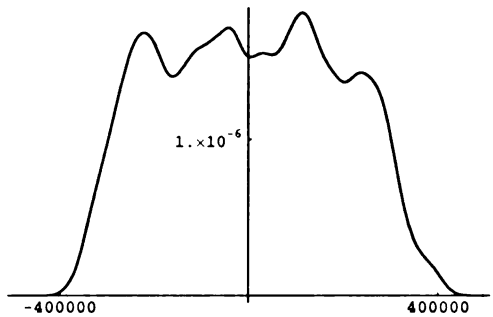
For the same model Figure 1.3 provides a second illustration with a "pulse" error i.e. containing a single outlier $\boldsymbol{\epsilon} = (a, 0, \ldots, 0)'$. The second and third rows show the distributions of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ conditional on whether the single large error aligns with the block $\boldsymbol{v}^1$ or $\boldsymbol{v}^2$, respectively. (In other words, the last two rows show the conditional distributions $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \,|\, \pi(1) \in A_1$ and $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \,|\, \pi(1) \in A_2$ where the sets A_1 and A_2 contain indices of components in $\boldsymbol{v}$ corresponding to $\boldsymbol{v}^1$ and $\boldsymbol{v}^2$, respectively.) The first row of Figure 1.3 illustrates the full permutation bootstrap approximation of the unconditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$. Figure 1.4 offers a more detailed comparison of the incomplete permutation approximations of the corresponding conditional probabilities.

Motivated by the above idea we extend results of [17] to create *conditional* confidence regions given $\Lambda\pi$, where Λ is a subgroup of Σ_n and $\Lambda\pi = \{\nu\pi : \nu \in \Lambda\}$. We achieve this by the means of *incomplete* permutation bootstrap of the observed residuals, i.e. bootstrap based on random permutations distributed uniformly over Λ, instead of Σ_n as in [8, 17, 18]. These are shown in row 3 of Figure 1.3 and in Figure 1.4 with Λ permuting the indices in each block A_1 and A_2 among themselves.

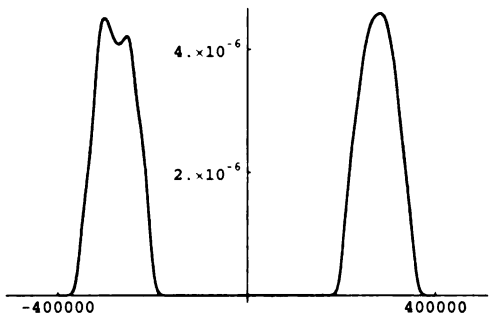
Figure 1.1: The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\epsilon$ in a simple linear regression with i.i.d. errors from the domain of attraction of a symmetric stable law with $\alpha = 0.5$ (in particular, $\epsilon_i = \delta_i U_i^{-2}$ are reciprocals of squared i.i.d. uniform variables with independent symmetrical random signs). The estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations.



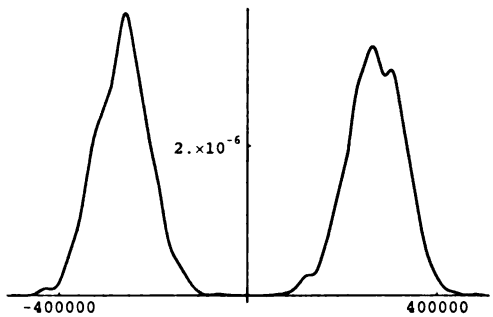
Unconditional distribution of $\boldsymbol{v} \cdot \pi\epsilon$



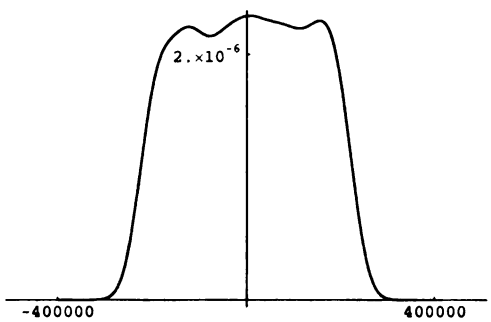
Full permutation bootstrap approximation of the distribution of $\boldsymbol{v} \cdot \pi\epsilon$



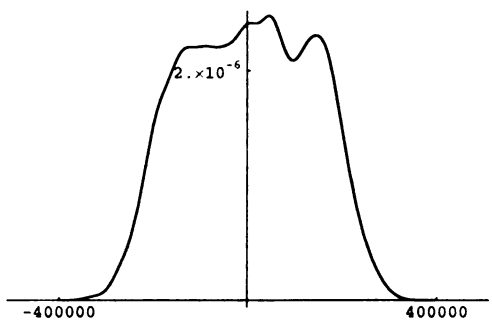
Conditional distribution $\mathcal{L}_1$ of $\boldsymbol{v} \cdot \pi\epsilon$ given that extreme values of $\pi\epsilon$ align with $\boldsymbol{v}^1$



Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_1$

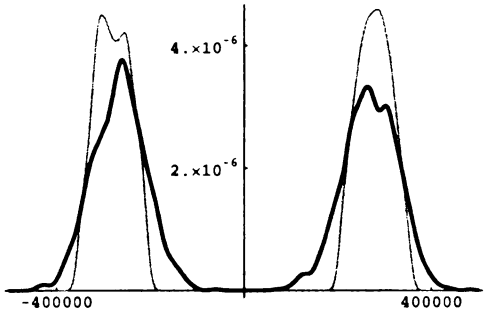


Conditional distribution $\mathcal{L}_2$ of $\boldsymbol{v} \cdot \pi\epsilon$ given that extreme values of $\pi\epsilon$ align with $\boldsymbol{v}^2$

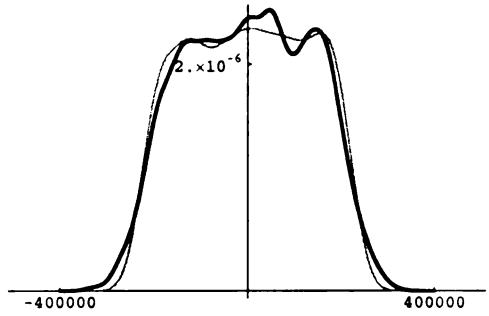


Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_2$

Figure 1.2: Comparison of the exact and incomplete permutation bootstrap approximated conditional distributions of the errors $v \cdot \pi\epsilon$ in a simple linear regression with i.i.d. errors from the domain of attraction of a stable law with $\alpha = 0.5$ (see Figure 1.1). All estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations.



Comparison of the conditional distribution $\mathcal{L}_1$ and its approximation



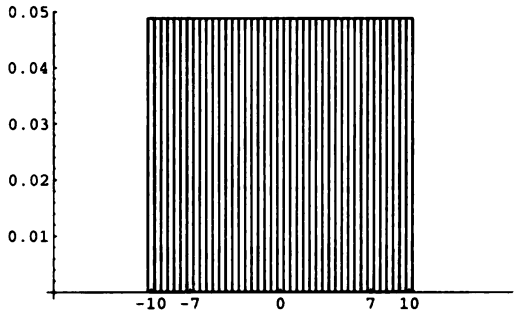
Comparison of the conditional distribution $\mathcal{L}_2$ and its approximation

The fact that the permutation is incomplete generally introduces a *conditional* bias to the least squares estimator of β. This bias is not present in the case $\Lambda = \Sigma_n$ studied in [17]. We introduce a *conditional* bias adjusted estimator for which we obtain in Theorem 1.3.4 below a *design* formula (1.19) for the relative size of the mean square of its (remaining) *conditional* bias (RMSB).

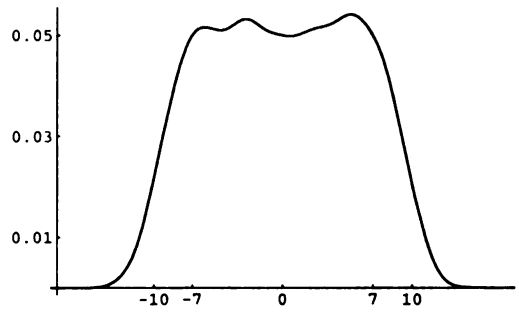
Aside from bias a key issue is the relative size of the mean squared *discrepancy* (RMSD) between the exact $\Lambda\pi$–*conditional* distribution of the bias adjusted estimator and its approximation based on incomplete permutation bootstrap of the observed residuals. Theorem 1.4.1 below establishes an exact *design* formula (1.23) for the relative size of the mean square discrepancy.

In Theorem 1.4.3 we also obtain a *design* formula (1.27) for the relative reduction in the size of our *conditional* confidence regions (based on the bias adjusted estimator and *incomplete* permutation bootstrap) versus the full permutation bootstrap in cases where both the RMSB and RMSD are small.

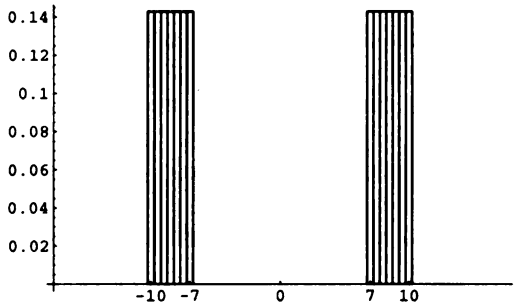
Figure 1.3: The conditional and unconditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. The corresponding independent block and full permutation bootstrap approximations were obtained via Monte–Carlo with 5000 replicas of the corresponding random permutations.



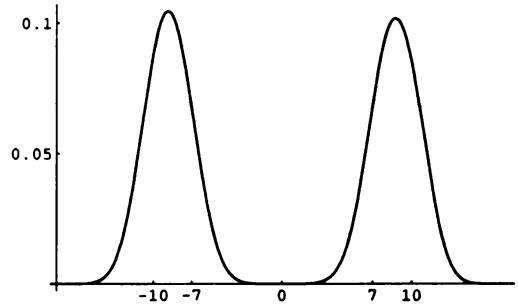
Unconditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



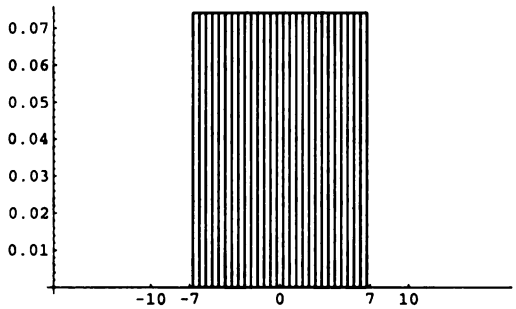
Full permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



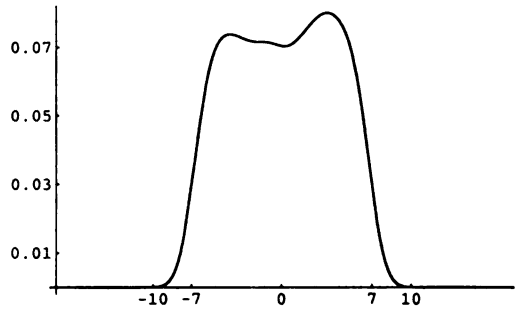
Distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$



Incomplete permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$

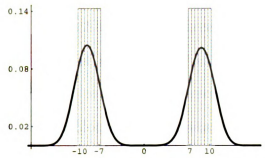


Distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$

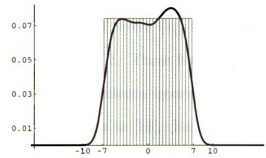


Incomplete permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$

Figure 1.4: Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. (Data from Figure 1.3.)



Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$ and its approximation



Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$ and its approximation

We refine these results in Section 1.5 where we obtain a formula (1.33) for the relative $\Lambda\pi$–*conditional* mean square discrepancy for the case of independent block permutations of block-wise contrasts in which the estimator is unbiased. In this section we also obtain simplified versions of the formulas for the case of independent block permutation bootstrap in simple linear regression. The reduced formulas and the performance of the independent block permutation bootstrap in simple linear regression are illustrated by a numerical simulation and comparison in Example 1.6.1.

In addition to developing these incomplete permutation bootstrap confidence regions we also extend the idea of Quade (1973, [20]) and the results from [19] and [9] to find an *exact* one-dimensional confidence interval (1.52) for a single coefficient of the regression model (1.1), conditional on the *incomplete* random permutations of the errors (Proposition 1.8.1). Unfortunately, the pivot employed in this exact method seems not to generalize to simultaneous confidence regions for several coefficients. At present it would seem that our incomplete permutation

method provides the only viable solution for the case of simultaneous confidence regions, short of a fully exhaustive exact permutations test method.

Although our results are not limited to the case of errors attracted to stable laws, we establish limit theorems for the special case of errors $\{\epsilon_i\}$ belonging to the domain of attraction of an α–stable law with $\alpha < 2$. These limit theorems describe the *conditional* convergence of *conditional* confidence intervals based on incomplete block–wise permutations, founded on results in [18]. For such errors the sampling distribution of the size of the *conditional* confidence intervals tends to be small compared to the size of unconditional confidence intervals.

In Chapter 2 we discuss an application of these results to the problem of estimating the coefficients of a wavelet expansion of an L_2 function which can, from the practical point of view, be considered a special case of model (1.1). The special form of the matrix X in the wavelet expansion model suggests that instead of the full permutation bootstrap one should consider using incomplete permutation bootstrap with a specifically selected permutation subgroup Λ.

In wavelet expansion of an L_2 function f the vector $\boldsymbol{y}$ of the sampled values of the function f tends to be of a large dimension n, therefore the limiting theory of the (incomplete) permutation bootstrap will apply to the discrete approximation of the original wavelet expansion problem.

Therefore these modified bootstrap methods can potentially be applied with good results to wavelet shrinkage (see [5]) to construct simultaneous conditional confidence intervals for the wavelet coefficients not only in the case of normal errors as considered by the authors in [5], but also in cases with ill behaved errors for which the central limit theorems fail. An important example of such a situation is the case of errors attracted to a stable law with index $\alpha < 2$ mentioned above.

1.2 Random Permutations in Multiple Linear Regression

Consider a linear regression model

$$\boldsymbol{y} = X \cdot \boldsymbol{\beta} + \pi\boldsymbol{\epsilon} \tag{1.1}$$

where $\boldsymbol{y}_{n\times 1}$ is an observed real vector, $X_{n\times d}$ is a known matrix with real components, $\boldsymbol{\beta}_{d\times 1}$ is a vector of unknown real parameters, and π is a random permutation with distribution uniform over the collection Σ_n of all permutations over $\{1,\ldots,n\}$. The vector of errors is introduced into the model via its random π-presentation $\pi\boldsymbol{\epsilon}$ obtained from $\boldsymbol{\epsilon}$ by applying π to the indices of the components of $\boldsymbol{\epsilon}$. The unknown real vector $\boldsymbol{\epsilon}_{n\times 1}$ itself is considered to be non-random.

We will assume that $n \geq 1$, X has full rank d and that either its first column is the vector $\mathbf{1}$ with all components equal to 1, or alternatively that all the columns of X are orthogonal to $\mathbf{1}$.

Typically, the goal is to estimate the joint unconditional distribution of the estimation errors $\hat{\beta}_i - \beta_i$ for several indices $i \geq 1$. Here $\hat{\boldsymbol{\beta}}$ represents the least squares estimator

$$\hat{\boldsymbol{\beta}} = V \cdot \boldsymbol{y} = \left((X'X)^{-1}X'\right) \cdot \boldsymbol{y}$$

.

We will be concerned with a modification of the above goal which seeks instead to estimate particular *conditional* distributions of the estimation errors. Denoting the rows of the matrix V by $\boldsymbol{v}_i$, $i = 1,\ldots,d$, allows us to rewrite the estimation errors as $\hat{\beta}_i - \beta_i = \boldsymbol{v}_i \cdot \pi\boldsymbol{\epsilon}$ and for $i \geq 2$ consider a more general problem of estimating the joint distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ for several contrast vectors $\boldsymbol{v}$ which are generally not assumed to be rows of the matrix V.

Note that this is a generalization since if $\mathbf{1}$ is the first column of X, then for each $i \geq 2$ the vectors $\boldsymbol{v}_i$ are contrasts. Indeed, for all $i \geq 2$ the vectors $\boldsymbol{v}_i$ are orthogonal to the first column $\mathbf{1}$ of the matrix X since VX is equal to the identity matrix. In the case when all columns of X are orthogonal to $\mathbf{1}$ the definition of V clearly guarantees that $\boldsymbol{v}_i$ are contrasts for all $i \geq 1$.

1.2.1 Basic Properties

Let us first consider a few basic properties of random permutations. Hereafter we will assume that π is a random permutation with distribution uniform over the group Σ_n of all permutations over $\{1, \ldots, n\}$.

For an n-dimensional vector $\boldsymbol{\epsilon}$ and a permutation $\rho \in \Sigma_n$ we will by $\rho\boldsymbol{\epsilon}$ denote the vector obtained from $\boldsymbol{\epsilon}$ by applying ρ to the indices of its components. The i-th component of $\rho\boldsymbol{\epsilon}$ will be naturally denoted by $\epsilon_{\rho(i)}$.

We will use the usual notation $E^{\mathcal{F}}$ to denote the expectation conditional on a σ–algebra $\mathcal{F}$. If Z is a random variable, then $E^{\sigma(Z)}$ will for simplicity be written as E^{Z}.

In this section we will take advantage of the following two facts. For any measurable function f, any n-dimensional real vector $\boldsymbol{\epsilon}$, and for all $i, j = 1, \ldots, n$, where $i \neq j$, it holds that

$$Ef(\epsilon_{\pi(i)}) = \frac{1}{n}\sum_{k=1}^{n} f(\epsilon_k) \tag{1.2}$$

$$E^{\pi(i)} f(\epsilon_{\pi(j)}) = \frac{1}{n-1}\sum_{k=1}^{n} f(\epsilon_k) - \frac{1}{n-1} f(\epsilon_{\pi(i)}) \tag{1.3}$$

To see that (1.2) holds write $Ef(\epsilon_{\pi(i)}) = E\sum_{k=1}^{n} 1_{\{\pi(i)=k\}} f(\epsilon_k)$ and observe that for all i and k the probability $P(\pi(i) = k)$ is equal to $\frac{1}{n}$. Similarly, for every i and

l we have that $E^{\pi(i)} f(\epsilon_{\pi(j)}) = E^{\pi(i)} \sum_{k \neq l} 1_{\{\pi(j)=k\}} f(\epsilon_k)$ a.s. on the set $\{\pi(i) = l\}$, and that for all $k \neq l$ and $i \neq j$ $P(\pi(j) = k \mid \pi(i) = l) = \frac{1}{n-1}$. Therefore (1.3) holds as well.

Lemma 1.2.1 (Lemma 1 in [17]) *Let $\boldsymbol{\epsilon}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$ and let π be a random permutation with distribution uniform over Σ_n. Then for every contrast vector $\boldsymbol{u}$ in $\mathbb{R}^n$ it holds that*

$$E(\boldsymbol{u} \cdot \pi\boldsymbol{\epsilon})(\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}) = \frac{1}{n-1}(\boldsymbol{u} \cdot \boldsymbol{v}) \left\|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\right\|^2, \tag{1.4}$$

where $\bar{\boldsymbol{\epsilon}}$ denotes the vector $\mathbf{1}(\boldsymbol{\epsilon} \cdot \mathbf{1})/n$ with all components equal to the average of the components of $\boldsymbol{\epsilon}$.

Proof: First notice that for any vector $\boldsymbol{z} \in \mathbb{R}_n$ the equation (1.2) implies that $E(\boldsymbol{z} \cdot \pi\boldsymbol{\epsilon}) = \bar{\boldsymbol{z}} \cdot \bar{\boldsymbol{\epsilon}}$. Consequently we can write

$$\begin{aligned} E\left[\boldsymbol{u} \cdot \pi(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}})\right]\left[\boldsymbol{v} \cdot \pi(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}})\right] &= E(\boldsymbol{u} \cdot \pi\boldsymbol{\epsilon})(\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}) - (\bar{\boldsymbol{u}} \cdot \bar{\boldsymbol{v}}) \left\|\bar{\boldsymbol{\epsilon}}\right\|^2 \\ &= E(\boldsymbol{u} \cdot \pi\boldsymbol{\epsilon})(\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}) \end{aligned}$$

since $\boldsymbol{u}$ is a contrast. Therefore we can assume, without loss of generality, that $\boldsymbol{\epsilon}$ is also a contrast since for an arbitrary vector $\boldsymbol{\epsilon}$ this modified lemma can be applied to the vector $\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}$ which is a contrast.

Observe that for a contrast $\boldsymbol{\epsilon}$ the right hand side in (1.3), with the identity function in place of f, becomes $\frac{-1}{n-1}\epsilon_{\pi(i)}$. Then, using equations (1.2) and (1.3) and the fact that $\boldsymbol{u}$ is a contrast, we obtain

$$\begin{aligned} E(\boldsymbol{u} \cdot \pi\boldsymbol{\epsilon})(\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}) &= \sum_{i=1}^{n} u_i v_i E\epsilon_{\pi(i)}^2 + \sum_{i \neq j} u_i v_j E\epsilon_{\pi(i)} E^{\pi(i)} \epsilon_{\pi(j)} \\ &= \frac{1}{n} \left\|\boldsymbol{\epsilon}\right\|^2 (\boldsymbol{u} \cdot \boldsymbol{v}) - \frac{1}{n-1} \sum_{i \neq j} u_i v_j E\epsilon_{\pi(i)}^2 \end{aligned}$$

$$= \frac{1}{n}\|\epsilon\|^2(\boldsymbol{u}\cdot\boldsymbol{v}) - \frac{1}{n(n-1)}\|\epsilon\|^2[(\boldsymbol{u}\cdot\mathbf{1})(\boldsymbol{v}\cdot\mathbf{1}) - (\boldsymbol{u}\cdot\boldsymbol{v})]$$

$$= \frac{1}{n-1}\|\epsilon\|^2(\boldsymbol{u}\cdot\boldsymbol{v}).$$

□

Corollary 1.2.2 *Let $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{\epsilon}$ be vectors in $\mathbb{R}^n$ and let π be a random permutation with distribution uniform over Σ_n. Then*

$$E(\boldsymbol{u}\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) = \frac{1}{n-1}([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\boldsymbol{v})\|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 + (\bar{\boldsymbol{u}}\cdot\bar{\boldsymbol{v}})\|\bar{\boldsymbol{\epsilon}}\|^2 \quad (1.5)$$

$$= \frac{1}{n-1}([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot[\boldsymbol{v}-\bar{\boldsymbol{v}}])\|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 + (\bar{\boldsymbol{u}}\cdot\bar{\boldsymbol{v}})\|\bar{\boldsymbol{\epsilon}}\|^2. \quad (1.6)$$

Proof: To prove (1.5) write

$$\begin{aligned} E(\boldsymbol{u}\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) &= E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) + E(\bar{\boldsymbol{u}}\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) \\ &= E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) + (\bar{\boldsymbol{u}}\cdot\boldsymbol{\epsilon})(\bar{\boldsymbol{v}}\cdot\boldsymbol{\epsilon}) \quad (1.7) \\ &= \frac{1}{n-1}([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\boldsymbol{v})\|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 + (\bar{\boldsymbol{u}}\cdot\bar{\boldsymbol{v}})\|\bar{\boldsymbol{\epsilon}}\|^2 \end{aligned}$$

by Lemma 1.2.1 with $\boldsymbol{u}-\bar{\boldsymbol{u}}$ playing the role of the contrast vector.

To see that (1.6) holds notice that in (1.7)

$$E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon})(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) = E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon})([\boldsymbol{v}-\bar{\boldsymbol{v}}]\cdot\pi\boldsymbol{\epsilon})$$

since $E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon})(\bar{\boldsymbol{v}}\cdot\pi\boldsymbol{\epsilon}) = (\bar{\boldsymbol{v}}\cdot\bar{\boldsymbol{\epsilon}})E([\boldsymbol{u}-\bar{\boldsymbol{u}}]\cdot\pi\boldsymbol{\epsilon}) = 0$.

□

Corollary 1.2.3 *Let $\boldsymbol{\epsilon}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$ and let π be a random permutation with distribution uniform over Σ_n. Let W be a subspace in $\mathbb{R}^n$ with either $\mathbf{1} \in W$*

or $\mathbf{1} \perp W$. *Then for any contrast vector* $\boldsymbol{u}$ *in* $\mathbb{R}^n$

$$E(\boldsymbol{u} \cdot (\pi\epsilon)/_W)(\boldsymbol{v} \cdot (\pi\epsilon)/_W) = \frac{1}{n-1}(\boldsymbol{u}/_W \cdot \boldsymbol{v}/_W) \, \|\epsilon - \bar{\epsilon}\|^2 . \tag{1.8}$$

Here $\boldsymbol{u}/_W$ *denotes the projection of the vector* $\boldsymbol{u}$ *on the space* W.

Proof: We can clearly write $(\boldsymbol{u} \cdot (\pi\epsilon)/_W)(\boldsymbol{v} \cdot (\pi\epsilon)/_W) = (\boldsymbol{u}/_W \cdot \pi\epsilon)(\boldsymbol{v}/_W \cdot \pi\epsilon)$ where $\boldsymbol{u}/_W$ is a contrast. Hence a direct application of Lemma 1.2.1 implies that (1.8) holds.

To see that $\boldsymbol{u}/_W$ is a contrast recall that either $\mathbf{1} \in W$ or $\mathbf{1} \perp W$ and observe that because of $\boldsymbol{u}$ being a contrast $\boldsymbol{u}/_W \cdot \mathbf{1} = \boldsymbol{u} \cdot \mathbf{1}/_W = 0$ in both cases.

□

1.3 Conditional Models

To simplify the notation of projections we will use $\boldsymbol{z}/_X$ to denote the projection of a vector $\boldsymbol{z} \in \mathbb{R}^n$ on the subspace generated either by a set X of vectors in $\mathbb{R}^n$, or by the columns of a matrix X. Further, $X^\perp$ will denote the appropriate orthogonal complement of the corresponding space and $\boldsymbol{z}^\perp$ will stand for $\boldsymbol{z}/_{X^\perp}$ whenever the meaning of X is implicitly clear from the context. Similarly, $z \perp X$ will indicate that the vector z is orthogonal either to the space X or to the column space of the matrix X.

Before we introduce the conditional model, let us briefly mention a few important properties of the permutations and random permutations being used. Consider a subgroup Λ of the group Σ_n of all permutations over $\{1, \ldots, n\}$ and recall that for every permutation $\nu \in \Lambda$ there is a unique inverse permutation in Λ, denoted by ν^{-1}, such that $\nu\nu^{-1}$ results in the identity permutation. In addition, if λ is a

random permutation with distribution uniform over Λ then also λ^{-1} is distributed uniformly over Λ. We will also take advantage of the fact that for such a random permutation λ and for any vector $\boldsymbol{v} \in \mathbb{R}^n$ it holds that

$$\lambda E\lambda \boldsymbol{v} = E\lambda \boldsymbol{v}. \tag{1.9}$$

The equation (1.9) follows from the fact that the application of any fixed permutation $\gamma \in \Lambda$ leaves the components of $E\lambda\boldsymbol{v}$ unaffected. To see this we can write $\gamma E\lambda \boldsymbol{v} = \frac{1}{|\Lambda|}\sum_{\nu\in\Lambda}\gamma\nu\boldsymbol{v} = E\lambda\boldsymbol{v}$, where $|\Lambda|$ represents the cardinality of Λ. The last equality holds because Λ is a group of permutations and hence for each δ in Λ there exists a unique $\nu \in \Lambda$ such that $\delta = \gamma\nu$.

Assumption: *For all that follows we require that λ be uniformly distributed over a subgroup Λ of Σ_n and independent of π which is distributed uniformly over Σ_n.*

We will now turn our attention to a conditional version of the model (1.1). Observe that the conditional distribution $\pi \mid \Lambda\pi$ of π given the σ-algebra generated by the random set–valued mapping $\Lambda\pi = \{\nu\pi : \nu \in \Lambda\}$ satisfies

$$\pi \mid \Lambda\pi = \lambda\pi \mid \pi. \tag{1.10}$$

Under the $\Lambda\pi$–conditional model for the errors $\pi\boldsymbol{\epsilon}$ the estimator $\boldsymbol{v}\cdot\boldsymbol{y}$ may be conditionally biased with conditional bias

$$E^{\Lambda\pi}(\boldsymbol{v}\cdot\boldsymbol{y} - \boldsymbol{v}\cdot X\boldsymbol{\beta}) = E^{\pi}(\lambda\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}) = (E\lambda\boldsymbol{v})\cdot\pi\boldsymbol{\epsilon}. \tag{1.11}$$

In an attempt to reduce the effect of any $\Lambda\pi$–conditional bias we propose using a bias–adjusted estimator

$$(\boldsymbol{v}\cdot\boldsymbol{y})_{\text{adj}} = \boldsymbol{v}\cdot\boldsymbol{y} - (E\lambda\boldsymbol{v})\cdot(\pi\boldsymbol{\epsilon})^{\perp}, \tag{1.12}$$

where, of course, $(\pi\boldsymbol{\epsilon})^{\perp} = \boldsymbol{y}^{\perp} = \boldsymbol{y}/_{X^{\perp}}$ is a statistic.

The remaining conditional bias of the adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ and its unconditional mean square are described in the following two lemmas.

Lemma 1.3.1 *The remaining $\Lambda\pi$–conditional bias of the bias–adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ is equal to*

$$\begin{aligned} \mathrm{B}_{\text{adj}}(\Lambda, \pi) &= \left(E\lambda\boldsymbol{v} - E\lambda(E\lambda\boldsymbol{v})^{\perp}\right) \cdot \pi\boldsymbol{\epsilon} && (1.13) \\ &= E\left(\lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right) \cdot \pi\boldsymbol{\epsilon} && (1.14) \end{aligned}$$

Proof: The $\Lambda\pi$–conditional bias of $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ can be written as

$$\begin{aligned} & E^{\Lambda\pi}\left(\boldsymbol{v} \cdot \boldsymbol{y} - (E\lambda\boldsymbol{v}) \cdot (\pi\boldsymbol{\epsilon})^{\perp} - \boldsymbol{v} \cdot X\boldsymbol{\beta}\right) \\ &\qquad = E^{\Lambda\pi}\left(\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} - (E\lambda\boldsymbol{v}) \cdot (\pi\boldsymbol{\epsilon})^{\perp}\right) \\ &\qquad = E^{\pi}\left(\boldsymbol{v} \cdot \lambda\pi\boldsymbol{\epsilon} - (E\lambda\boldsymbol{v}) \cdot (\lambda\pi\boldsymbol{\epsilon})^{\perp}\right) \\ &\qquad = E^{\pi}\left(\lambda\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} - \lambda(E\lambda\boldsymbol{v})^{\perp} \cdot \pi\boldsymbol{\epsilon}\right) \qquad (1.15) \\ &\qquad = E^{\pi}\left(\lambda\boldsymbol{v} - \lambda E\lambda\boldsymbol{v} + \lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right) \cdot \pi\boldsymbol{\epsilon} \\ &\qquad = E\left(\lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right) \cdot \pi\boldsymbol{\epsilon} \end{aligned}$$

which proves both statements of the lemma as (1.13) follows directly from (1.15). In the last equality above we have used (1.9).

□

Note that if in the following lemma $\mathbf{1} \perp X$ then the assumption that $\boldsymbol{v}$ is a contrast is not necessary. This fact can be seen from the proof of the lemma. However, for other results below we will have to require that $\boldsymbol{v}$ be a contrast even if $\mathbf{1} \perp X$.

Lemma 1.3.2 *If $\boldsymbol{v} \in \mathbb{R}^n$ is a contrast, then the mean square of the conditional*

bias $\mathrm{B}_{\mathrm{adj}}(\Lambda, \pi)$ *is*

$$E(\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi))^2 = \frac{1}{n-1}\left\|E\lambda\boldsymbol{v} - E\lambda(E\lambda\boldsymbol{v})^{\perp}\right\|^2 \|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 \tag{1.16}$$

$$= \frac{1}{n-1}\left\|E\lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right\|^2 \|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 . \tag{1.17}$$

Proof: We will first show that the vector $E\lambda\left[(E\lambda\boldsymbol{v})/_X\right]$ is a contrast. If $\boldsymbol{u}$ is a contrast then so is a vector obtained from $\boldsymbol{u}$ by permuting its components and hence also $E\lambda\boldsymbol{u}$ is a contrast. In addition, the projection $\boldsymbol{u}/_X$ of a contrast $\boldsymbol{u}$ is a contrast since $\boldsymbol{u}/_X \cdot \mathbf{1} = \boldsymbol{u}\cdot\mathbf{1}/_X = 0$ in both cases when $\mathbf{1}$ is the first column of X or if $\mathbf{1}\perp X$. Therefore $(E\lambda\boldsymbol{v})/_X$ and consequently also $E\lambda\left[(E\lambda\boldsymbol{v})/_X\right]$ are contrast vectors.

Hence the assertion of the lemma is a direct consequence of Lemma 1.2.1 applied to $E\left(E\left(\lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right)\cdot\pi\boldsymbol{\epsilon}\right)^2$ with the contrast $E\lambda\left[(E\lambda\boldsymbol{v})/_X\right]$ playing the role of the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$.

□

The next lemma and theorem provide the means for comparing the mean square of the remaining $\Lambda\pi$–conditional bias $\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi)$ against the mean squared error of the bias adjusted estimator $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$.

Lemma 1.3.3 *For a contrast vector* $\boldsymbol{v}\in\mathbb{R}^n$ *the mean squared error of the bias–adjusted estimator* $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$ *equals*

$$\mathrm{MSE}_{\mathrm{adj}} = \frac{1}{\mathrm{n}-1}\left\|\boldsymbol{v} - (\mathrm{E}\lambda\boldsymbol{v})^{\perp}\right\|^2 \|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\|^2 . \tag{1.18}$$

Proof: First write

$$\mathrm{MSE}_{\mathrm{adj}} = E\left((\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}} - \boldsymbol{v}\cdot X\boldsymbol{\beta}\right)^2$$

$$
\begin{aligned}
&= E\left(\boldsymbol{v}\cdot\boldsymbol{y} - (E\lambda\boldsymbol{v})\cdot(\pi\boldsymbol{\epsilon})^{\perp} - \boldsymbol{v}\cdot X\boldsymbol{\beta}\right)^2 \\
&= EE^{\Lambda\pi}\left(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon} - (E\lambda\boldsymbol{v})\cdot(\pi\boldsymbol{\epsilon})^{\perp}\right)^2 \\
&= EE^{\pi}\left(\boldsymbol{v}\cdot\lambda\pi\boldsymbol{\epsilon} - (E\lambda\boldsymbol{v})\cdot(\lambda\pi\boldsymbol{\epsilon})^{\perp}\right)^2 \\
&= EE^{\lambda}\left(\left[\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}\right]\cdot\pi\boldsymbol{\epsilon}\right)^2.
\end{aligned}
$$

Similarly as in the proof of Lemma 1.3.2 we will show that, for every fixed value of the random permutation λ, the vector $\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}$ is a contrast. Observe that it suffices to prove that $(E\lambda\boldsymbol{v})^{\perp}$ is a contrast since then it is clear that $\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}$ is nothing but a random permutation of a difference of two contrast vectors, and hence itself is a contrast, although random.

As noted in the proof of Lemma 1.3.2, if $\boldsymbol{v}$ is a contrast then so are $E\lambda\boldsymbol{v}$ and it's projection to the subspace $X^{\perp}$ which, similarly as X itself, either contains the vector $\mathbf{1}$ or is perpendicular to it.

A straightforward application of Lemma 1.2.1 with $\boldsymbol{u}$ and $\boldsymbol{v}$ both equal to the contrast vector $\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}$ then yields that for each fixed value of λ

$$
\begin{aligned}
E^{\lambda}([\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}]\cdot\pi\boldsymbol{\epsilon})^2 &= \frac{1}{n-1}\left\|\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^{\perp}\right\|^2\|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 \\
&= \frac{1}{n-1}\left\|\boldsymbol{v} - (E\lambda\boldsymbol{v})^{\perp}\right\|^2\|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2
\end{aligned}
$$

which completes the proof.

□

Theorem 1.3.4 *Let $\boldsymbol{\epsilon}$ be a vector in $\mathbb{R}^n$ and let π be a random permutation with distribution uniform over Σ_n. Let $X_{n\times d}$ be a matrix of rank d with the first column equal to $\mathbf{1}$ or such that $\mathbf{1}\perp X$. Assume that λ is a random permutation distributed uniformly over a subgroup Λ of Σ_n and independent of π. Then for any contrast*

vector $\boldsymbol{v}$ *in* $\mathbb{R}^n$ *the ratio of the mean square of the remaining* $\Lambda\pi$*–conditional bias* $\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi)$ *to the mean squared error of the bias–adjusted estimator* $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$*, i.e. the relative mean squared bias, is*

$$RMSB = \frac{E(\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi))^2}{\mathrm{MSE}_{\mathrm{adj}}} = \frac{\left\|E\lambda\left[(E\lambda\boldsymbol{v})/_X\right]\right\|^2}{\left\|\boldsymbol{v}-(E\lambda\boldsymbol{v})^{\perp}\right\|^2}. \tag{1.19}$$

Proof: The assertion of the theorem follows directly from (1.17) and (1.18). □

An extension of Theorem 1.3.4 to the multi–dimensional case can be obtained as follows.

Corollary 1.3.5 *Let* r *be a positive integer and let* $\{a_k\}_{k=1}^r$ *and* $\{b_k\}_{k=1}^r$ *be two sets of real numbers. Then for* r *contrast vectors* $\{\boldsymbol{v}_k\}_{k=1}^r$ *it holds that*

$$\frac{\sum_{k=1}^r a_k E(\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi,\boldsymbol{v}_k))^2}{\sum_{k=1}^r b_k \mathrm{MSE}_{\mathrm{adj}}(\boldsymbol{v}_k)} = \frac{\sum_{k=1}^r a_k \left\|E\lambda[(E\lambda\boldsymbol{v}_k)/_X]\right\|^2}{\sum_{k=1}^r b_k \left\|\boldsymbol{v}_k-(E\lambda\boldsymbol{v}_k)^{\perp}\right\|^2} \tag{1.20}$$

Proof: Equations (1.17) and (1.18) imply that

$$\begin{aligned}
\frac{\sum_{k=1}^r a_k E(\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi,\boldsymbol{v}_k))^2}{\sum_{k=1}^r b_k \mathrm{MSE}_{\mathrm{adj}}(\boldsymbol{v}_k)} &= \frac{\sum_{k=1}^r a_k E[E^{\Lambda\pi}((\boldsymbol{v}_k\cdot\boldsymbol{y})_{adj}-\boldsymbol{v}_k\cdot X\boldsymbol{\beta})]^2}{\sum_{k=1}^r b_k E((\boldsymbol{v}_k\cdot\boldsymbol{y})_{adj}-\boldsymbol{v}_k\cdot X\boldsymbol{\beta})^2} \\
&= \frac{\frac{1}{n-1}\left\|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\right\|^2 \sum_{k=1}^r a_k \left\|E\lambda[(E\lambda\boldsymbol{v}_k)/_X]\right\|^2}{\frac{1}{n-1}\left\|\boldsymbol{\epsilon}-\bar{\boldsymbol{\epsilon}}\right\|^2 \sum_{k=1}^r b_k E\left\|\lambda\boldsymbol{v}_k-\lambda(E\lambda\boldsymbol{v}_k)^{\perp}\right\|^2} \\
&= \frac{\sum_{k=1}^r a_k \left\|E\lambda[(E\lambda\boldsymbol{v}_k)/_X]\right\|^2}{\sum_{k=1}^r b_k \left\|\boldsymbol{v}_k-(E\lambda\boldsymbol{v}_k)^{\perp}\right\|^2}.
\end{aligned}$$

□

In general it is desirable that (1.19) or (1.20) be small so that the bias component of the error of the bootstrap approximation developed below is relatively small when compared to the sampling variation of $\boldsymbol{v}\cdot\boldsymbol{y}$.

1.4 Incomplete Permutation Bootstrap

The $\Lambda\pi$–conditional distribution of the estimation errors of the bias–adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ satisfies

$$[(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}} - \boldsymbol{v} \cdot X\beta] \mid \Lambda\pi = (\lambda[\boldsymbol{v} - (E\lambda\boldsymbol{v})^{\perp}] \cdot \pi\boldsymbol{\epsilon}) \mid \pi \tag{1.21}$$

as noted in the proof of Lemma 1.3.1.

To approximate the conditional distribution (1.21) of the errors we will propose using

$$\lambda[\boldsymbol{v} - (E\lambda\boldsymbol{v})^{\perp}] \cdot (\pi\boldsymbol{\epsilon})^{\perp} \mid \boldsymbol{y} \tag{1.22}$$

where the unknown presentation of the errors $\pi\boldsymbol{\epsilon}$ is replaced by the observed residuals $(\pi\boldsymbol{\epsilon})^{\perp}$. The proposal will by justified by the fact that both the mean squared remaining bias of the bias–adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ and the mean squared difference of the random variables in (1.21) and (1.22) will in some cases be small when scaled by the mean squared error of the bias–adjusted estimator.

Confidence regions and hypotheses tests based on the distribution (1.22) will under some circumstances and with high probability closely approximate the conditional forms based on (1.21). Typically we have to require that (1.19) and (1.23) below both be small which depends only on X and Λ. These regions and tests may, as a practical matter, be constructed via Monte–Carlo simulations.

The following theorem will allow us to investigate how well (1.22) approximates the target distribution (1.21) as it establishes an upper bound on the Mallows metric between these two distributions (see e.g. [3] for details on Mallows metric). For a fixed contrast vector $\boldsymbol{v}$ and a random permutation λ uniform over Λ let $True_{\Lambda} = \lambda[\boldsymbol{v} - (E\lambda\boldsymbol{v})^{\perp}] \cdot \pi\boldsymbol{\epsilon}$ and $BS_{\Lambda} = \lambda[\boldsymbol{v} - (E\lambda\boldsymbol{v})^{\perp}] \cdot (\pi\boldsymbol{\epsilon})^{\perp}$. These two specifically

chosen *dependent* random variables have conditional distributions (1.21) and (1.22), respectively.

Theorem 1.4.1 *The mean squared difference of* $True_\Lambda$ *and* BS_Λ *scaled by the mean squared error* $\mathrm{MSE}_{\mathrm{adj}}$ *of the bias-adjusted estimator* $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$*, in other words, the relative mean squared discrepancy, is*

$$RMSD = \frac{E(BS_\Lambda - True_\Lambda)^2}{\mathrm{MSE}_{\mathrm{adj}}} = \frac{E\left\|[\lambda\boldsymbol{v} - E\lambda\boldsymbol{v}]/_X\right\|^2}{\|\boldsymbol{v} - (E\lambda\boldsymbol{v})^\perp\|^2}. \tag{1.23}$$

Proof: To show (1.23) we will first use Corollary 1.2.3 to observe that the numerator can be written as

$$\begin{aligned}
&E(BS_\Lambda - True_\Lambda)^2 \\
&\quad = E[\lambda\boldsymbol{v}\cdot(\pi\boldsymbol{\epsilon})^\perp - \lambda(E\lambda\boldsymbol{v})^\perp\cdot(\pi\boldsymbol{\epsilon})^\perp - \lambda\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon} + \lambda(E\lambda\boldsymbol{v})^\perp\cdot\pi\boldsymbol{\epsilon}]^2 \\
&\quad = E[\lambda\boldsymbol{v}\cdot(\pi\boldsymbol{\epsilon})/_X - \lambda(E\lambda\boldsymbol{v})^\perp\cdot(\pi\boldsymbol{\epsilon})/_X]^2 \\
&\quad = EE^\lambda[(\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^\perp)\cdot(\pi\boldsymbol{\epsilon})/_X]^2 \\
&\quad = \frac{1}{n-1}E\left\|[\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})]/_X\right\|^2\|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 \\
&\quad = \frac{1}{n-1}E\left\|[\lambda\boldsymbol{v} - E\lambda\boldsymbol{v}]/_X\right\|^2\|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2.
\end{aligned}$$

The role of the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in Corollary 1.2.3 is, for each fixed value of λ, played by $\lambda\boldsymbol{v} - \lambda(E\lambda\boldsymbol{v})^\perp$ which has been shown to be a contrast in the proof of Lemma 1.3.3.

The form of the denominator in (1.23) can be obtained from the assertion of Lemma 1.3.3, hence (1.23) holds.

□

By minimizing (1.23) we can control how closely the incomplete permutation bootstrap method approximates the conditional distribution (1.21).

Corollary 1.4.2 (Part a) is Proposition 1 in [17]) *Assume that all requirements of Theorem 1.4.1 are satisfied for* $\Lambda = \Sigma_n$. *Then the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *has no conditional bias and the following hold:*

a) If $\mathbf{1}$ *is the first column of* X *then the relative mean square discrepancy becomes*

$$RMSD = \frac{E(\boldsymbol{v} \cdot \lambda\pi\boldsymbol{\epsilon} - \boldsymbol{v} \cdot \lambda(\pi\boldsymbol{\epsilon})^{\perp})^2}{E(\boldsymbol{v} \cdot \lambda\pi\boldsymbol{\epsilon})^2} = \frac{d-1}{n-1}. \tag{1.24}$$

b) If $\mathbf{1} \perp X$ *then the relative mean square discrepancy becomes*

$$RMSD = \frac{E(\boldsymbol{v} \cdot \lambda\pi\boldsymbol{\epsilon} - \boldsymbol{v} \cdot \lambda(\pi\boldsymbol{\epsilon})^{\perp})^2}{E(\boldsymbol{v} \cdot \lambda\pi\boldsymbol{\epsilon})^2} = \frac{d}{n-1} \tag{1.25}$$

Proof: Let $\{\boldsymbol{w}_i\}_{i=1}^d$ be an orthonormal basis of the space generated by the columns of X. Both assertions (1.24) and (1.25) follow directly from (1.23) and Lemma 1.2.1 applied to $E(\boldsymbol{v} \cdot \lambda\boldsymbol{w}_i)^2$ for each i.

To prove (1.25) first observe that if $\Lambda = \Sigma_n$ then $E(\lambda\boldsymbol{v}) = \bar{\boldsymbol{v}} = 0$ since $\boldsymbol{v}$ is a contrast vector. This and Lemma 1.2.1 yield that the numerator on the outmost right hand side in (1.23) becomes

$$\begin{aligned} E\left\|(\lambda\boldsymbol{v})/_X\right\|^2 &= \sum_{i=1}^{d} E(\boldsymbol{v} \cdot \lambda\boldsymbol{w}_i)^2 \\ &= \sum_{i=1}^{d} \frac{1}{n-1} \|\boldsymbol{v}\|^2 \|\boldsymbol{w}_i - \overline{\boldsymbol{w}_i}\|^2 \qquad (1.26) \\ &= \frac{d}{n-1} \|\boldsymbol{v}\|^2 . \end{aligned}$$

The last equality holds because of the fact that $\{\boldsymbol{w}_i\}_{i=1}^d$ is an orthonormal basis and that $\overline{\boldsymbol{w}_i} = 0$ for every $i = 1, \ldots, d$ since $\mathbf{1} \perp X$.

The proof of (1.24) is essentially identical if we assume that $\boldsymbol{w}_1$ is equal to a multiple of $\mathbf{1}$ because then $\overline{\boldsymbol{w}_i} = 0$ for every $i = 2, \ldots, d$, and the first summand of

$\sum_{i=1}^{d} E(\boldsymbol{v} \cdot \lambda \boldsymbol{w}_i)^2$ in (1.26) becomes $E(\boldsymbol{v} \cdot \lambda \boldsymbol{w}_1)^2 = 0$. Note that such a basis exists since in this case $\mathbf{1}$ is assumed to be the first column of X.

□

We will use the previous result to compare the performance of our incomplete permutation bootstrap to the performance of the full permutation bootstrap. The latter has been shown in [17] to perform in some cases better than the unconditional approach, yielding narrower confidence regions. Therefore it is sensible to determine in which situations our incomplete permutation bootstrap brings an additional performance improvement.

Corollary 1.4.3 *Let $\boldsymbol{v}$ be a contrast vector and let BS_Λ, BS_{Σ_n}, $True_\Lambda$, and $True_{\Sigma_n}$ be as above. Then the ratio of the incomplete bootstrap mean square discrepancy versus the mean square discrepancy under the full permutation bootstrap model can be expressed as*

$$\frac{E(BS_\Lambda - True_\Lambda)^2}{E(BS_{\Sigma_n} - True_{\Sigma_n})^2} = \frac{n-1}{d-1} E \left\| (\lambda \boldsymbol{v} \ - E\lambda \boldsymbol{v} \ )/_X \right\|^2, \tag{1.27}$$

where $\boldsymbol{v} \ = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.

Proof: According to the proof of Theorem (1.4.1) and formula (1.26) we can write

$$\frac{E(BS_\Lambda - True_\Lambda)^2}{E(BS_{\Sigma_n} - True_{\Sigma_n})^2} = \frac{E \left\| (\lambda \boldsymbol{v} - E\lambda \boldsymbol{v})/_X \right\|^2}{\frac{d-1}{n-1} \|\boldsymbol{v}\|^2}$$

which implies (1.27).

□

1.5 Independent Block Permutations

In the previous sections we obtained results for the unconditional relative mean square bias (RMSB) of the bias adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ and the unconditional

relative mean square discrepancy (RMSD) between the true conditional distribution of its errors and the incomplete permutation bootstrap approximation.

In this section we will consider the relative $\Lambda\pi$*–conditional* mean square discrepancy between the true and incomplete permutation bootstrap distributions in a special case of independent block permutations of a vector $\boldsymbol{v}$ consisting of blocks of contrasts. One of the advantages of this special case is that the estimator $\boldsymbol{v}\cdot\boldsymbol{y}$ becomes conditionally unbiased which implies that the true conditional distribution (1.21) of the estimation errors and the incomplete permutation bootstrap approximation (1.22) turn out to be

$$\lambda\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon}\mid\pi, \text{ and}$$

$$\lambda\boldsymbol{v}\cdot(\pi\boldsymbol{\epsilon})^{\perp}\mid\boldsymbol{y}$$

respectively.

For a partition $\mathcal{A}=(A_k)_{k=1}^K$ of $\{1,\ldots,n\}$, a vector $\boldsymbol{u}$ in $\mathbb{R}^n$, and every $k\in\{1,\ldots,K\}$ we will define a vector $[\![\boldsymbol{u}]\!]^k=(u_i)_{i\in A_k}$ representing the $k^{-\text{th}}$ block of components of the vector $\boldsymbol{u}$. The cardinalities of the sets A_k will be denoted by $n_k=|A_k|$, $k=1,\ldots,K$.

We will say that a random permutation λ *consists of* K *independent block permutations* if for some fixed partition $\mathcal{A}$ as above and any vector $\boldsymbol{u}\in\mathbb{R}^n$ the random permutation λ satisfies

$$[\![\lambda\boldsymbol{u}]\!]^k=\lambda_k[\![\boldsymbol{u}]\!]^k,\quad k=1,\ldots,K, \tag{1.28}$$

where $\lambda_1,\lambda_2,\ldots,\lambda_K$ are independent random permutations with distributions uniform over the full permutation groups $\Sigma_{n_1},\ldots,\Sigma_{n_K}$, respectively.

Note that previously we have required that λ be distributed uniformly over the

permutation subgroup Λ, therefore the condition (1.28) also enforces the form of Λ. Such a subgroup Λ will be called a *K-block permutation group.*

Below we describe the $\Lambda\pi$–conditional Mean Square Error of the estimator $\boldsymbol{v}\cdot\boldsymbol{y}$ and the $\Lambda\pi$–conditional Mean Square Discrepancy of the conditional distribution of its estimation errors from the proposed incomplete permutation bootstrap approximation under the assumption that $E\lambda\boldsymbol{v}=\mathbf{0}$. Note that the requirement $E\lambda\boldsymbol{v}=\mathbf{0}$ is in the case of independent block permutations equivalent to the condition that $E\lambda_k[\![\boldsymbol{v}]\!]^k=\overline{[\![\boldsymbol{v}]\!]^k}=\mathbf{0}$, i.e. that $[\![\boldsymbol{v}]\!]^k$ is a contrast vector for every $k=1,\ldots,K$.

It is also easy to see that if $E\lambda\boldsymbol{v}=0$ then the estimator $\boldsymbol{v}\cdot\boldsymbol{y}$ is conditionally unbiased as suggested above since according to (1.11) on page 14 the $\Lambda\pi$–conditional bias of $\boldsymbol{v}\cdot\boldsymbol{y}$ equals $(E\lambda\boldsymbol{v})\cdot\pi\boldsymbol{\epsilon}$.

Lemma 1.5.1 *Assume that λ consists of K independent block permutations and that $\boldsymbol{v}\in\mathbb{R}^n$. If $E\lambda\boldsymbol{v}=\mathbf{0}$ then the $\Lambda\pi$–conditional mean squared error of the estimator $\boldsymbol{v}\cdot\boldsymbol{y}$ equals*

$$\mathrm{MSE}^{\Lambda\pi}=E^{\Lambda\pi}(\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon})^2=\sum_{k=1}^{K}\frac{1}{n_k-1}\left\|[\![\boldsymbol{v}]\!]^k\right\|^2\left\|[\![\pi\boldsymbol{\epsilon}]\!]^k-\overline{[\![\pi\boldsymbol{\epsilon}]\!]^k}\right\|^2. \tag{1.29}$$

Proof: Using Lemma 1.2.1 with $[\![\boldsymbol{v}]\!]^k$ playing the role of the contrast vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ we can write

$$\begin{aligned}\mathrm{MSE}^{\Lambda\pi} &= E^{\Lambda\pi}(\boldsymbol{v}\cdot\boldsymbol{y}-\boldsymbol{v}\cdot X\boldsymbol{\beta})^2\\ &= E^{\pi}(\lambda\boldsymbol{v}\cdot\pi\boldsymbol{\epsilon})^2\\ &= E^{\pi}\left(\sum_{k=1}^{K}\lambda_k[\![\boldsymbol{v}]\!]^k\cdot[\![\pi\boldsymbol{\epsilon}]\!]^k\right)^2\\ &= \sum_{k=1}^{K}E^{\pi}\left([\![\boldsymbol{v}]\!]^k\cdot\lambda_k[\![\pi\boldsymbol{\epsilon}]\!]^k\right)^2\end{aligned} \tag{1.30}$$

$$= \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![v]\!]^k \right\|^2 \left\| [\![\pi\epsilon]\!]^k - \overline{[\![\pi\epsilon]\!]^k} \right\|^2.$$

In the equality (1.30) we used the independence of the random permutations $\{\lambda_k\}_{k=1}^K$ and the fact that $E^{\Lambda\pi}\lambda_k[\![v]\!]^k \cdot [\![\pi\epsilon]\!]^k = \overline{[\![v]\!]^k} \cdot E^{\Lambda\pi}[\![\pi\epsilon]\!]^k = 0$ for every $k = 1, \ldots, K$.

□

Lemma 1.5.2 *Under the assumptions of Lemma 1.5.1 the condition* $E\lambda v = 0$ *implies that the* $\Lambda\pi$*-conditional mean squared discrepancy between the true conditional distribution of the errors of the estimator* $v \cdot y$ *and its incomplete permutation bootstrap approximation is equal to*

$$E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2 = \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![v]\!]^k \right\|^2 E^{\pi} \left\| [\![(\lambda\pi\epsilon)/_X]\!]^k - \overline{[\![(\lambda\pi\epsilon)/_X]\!]^k} \right\|^2, \tag{1.31}$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

Proof: Similarly as in the proof of Lemma 1.5.1 we will use the independence of $\lambda_1, \ldots, \lambda_K$, the fact that $E^{\Lambda\pi}\lambda_k[\![v]\!]^k \cdot [\![(\pi\epsilon)/_X]\!]^k = \overline{[\![v]\!]^k} \cdot E^{\Lambda\pi}[\![(\pi\epsilon)/_X]\!]^k = 0$ for every $k \in \{1, \ldots, K\}$, and Lemma 1.2.1 with u and v replaced by the contrasts $[\![v]\!]^k$ to obtain

$$\begin{aligned}
E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2 &= E^{\Lambda\pi}\left(\lambda v \cdot (\pi\epsilon)^\perp - \lambda v \cdot \pi\epsilon\right)^2 \\
&= E^{\Lambda\pi}\left(\lambda v \cdot (\pi\epsilon)/_X\right)^2 \\
&= E^{\Lambda\pi}\left(\sum_{k=1}^{K} \lambda_k [\![v]\!]^k \cdot [\![(\pi\epsilon)/_X]\!]^k\right)^2 \\
&= \sum_{k=1}^{K} E^{\Lambda\pi}\left([\![v]\!]^k \cdot \lambda_k [\![(\pi\epsilon)/_X]\!]^k\right)^2
\end{aligned}$$

$$= \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![v]\!]^k \right\|^2 E^{\Lambda\pi} \left\| [\![(\pi\epsilon)/_X]\!]^k - \overline{[\![(\pi\epsilon)/_X]\!]^k} \right\|^2$$

$$= \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![v]\!]^k \right\|^2 E^{\pi} \left\| [\![(\lambda\pi\epsilon)/_X]\!]^k - \overline{[\![(\lambda\pi\epsilon)/_X]\!]^k} \right\|^2.$$

□

Corollary 1.5.3 *Assume that* $\mathbf{1}$ *is the first column of the matrix* X *and that* λ *is as in Lemma 1.5.1. Let* $\{\boldsymbol{w}_j\}_{j=1}^{d}$ *be an orthonormal basis of the column space of the matrix* X *such that* $\boldsymbol{w}_1 = \mathbf{1}/\sqrt{n}$. *If* $E\lambda\boldsymbol{w}_j = 0$ *for all* $j = 2, \ldots, d$ *then for any* $\boldsymbol{v} \in \mathbb{R}^n$ *with* $E\lambda\boldsymbol{v} = \mathbf{0}$ *the* $\Lambda\pi$*-conditional mean squared discrepancy between the true conditional distribution of the errors of the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *and its incomplete permutation bootstrap approximation can be written as*

$$E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2 = \tag{1.32}$$

$$= \sum_{i,j=2}^{d} \left[\left(\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![v]\!]^k \right\|^2 \left\| [\![w_i]\!]^k \right\|^2 \right) \right.$$

$$\left. \left(\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\pi\epsilon]\!]^k - \overline{[\![\pi\epsilon]\!]^k} \right\|^2 \left([\![w_i]\!]^k \cdot [\![w_j]\!]^k \right) \right) \right],$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

Proof: The assertion follows from Lemma 1.2.1 and 1.5.2. In (1.31) we can write

$$E^{\Lambda\pi} \left\| [\![(\pi\epsilon)/_X]\!]^k - \overline{[\![(\pi\epsilon)/_X]\!]^k} \right\|^2 = E^{\Lambda\pi} \left\| \sum_{i=2}^{d} (\pi\epsilon \cdot \boldsymbol{w}_i) [\![w_i]\!]^k \right\|^2$$

$$= \sum_{i,j=2}^{d} [\![w_i]\!]^k \cdot [\![w_j]\!]^k E^{\Lambda\pi}(\pi\epsilon \cdot \boldsymbol{w}_i)(\pi\epsilon \cdot \boldsymbol{w}_j),$$

for every $k \in \{1, \ldots, K\}$, and then use Lemma 1.2.1 to observe that for every i and j in $\{2 \ldots, d\}$

$$\begin{aligned} E^{\Lambda\pi}(\pi\boldsymbol{\epsilon} \cdot \boldsymbol{w}_i)(\pi\boldsymbol{\epsilon} \cdot \boldsymbol{w}_j) &= E^{\pi}\left(\sum_{l=1}^{K} [\![\lambda\pi\boldsymbol{\epsilon}]\!]^l \cdot [\![\boldsymbol{w}_i]\!]^l\right)\left(\sum_{l=1}^{K} [\![\lambda\pi\boldsymbol{\epsilon}]\!]^l \cdot [\![\boldsymbol{w}_j]\!]^l\right) \\ &= \sum_{l=1}^{K} E^{\pi}\left([\![\boldsymbol{w}_i]\!]^l \cdot \lambda_l [\![\pi\boldsymbol{\epsilon}]\!]^l\right)\left([\![\boldsymbol{w}_j]\!]^l \cdot \lambda_l [\![\pi\boldsymbol{\epsilon}]\!]^l\right) \\ &= \sum_{l=1}^{K} \frac{1}{n_l - 1} [\![\boldsymbol{w}_i]\!]^l \cdot [\![\boldsymbol{w}_j]\!]^l \left\| [\![\pi\boldsymbol{\epsilon}]\!]^l - \overline{[\![\pi\boldsymbol{\epsilon}]\!]^l} \right\|^2. \end{aligned}$$

In the second equality we use the independence of $\lambda_1, \ldots, \lambda_K$ and the fact that $E^{\pi}\lambda_l [\![\boldsymbol{w}_i]\!]^l \cdot [\![\pi\boldsymbol{\epsilon}]\!]^l = \overline{[\![\boldsymbol{w}_i]\!]^l} \cdot E^{\Lambda\pi} [\![\pi\boldsymbol{\epsilon}]\!]^l = 0$ for every $l = 1, \ldots, K$.

□

Theorem 1.5.4 *Let λ be a random permutation consisting of K independent block permutations. If $\boldsymbol{v} \in \mathbb{R}^n$ satisfies $E\lambda\boldsymbol{v} = \mathbf{0}$ then the ratio of the $\Lambda\pi$-conditional mean squared discrepancy between the true conditional distribution of the errors of the estimator $\boldsymbol{v} \cdot \boldsymbol{y}$ and its incomplete permutation bootstrap approximation versus its $\Lambda\pi$-conditional mean squared error, in other words the relative* $\Lambda\pi$-conditional *mean square discrepancy of $\boldsymbol{v} \cdot \boldsymbol{y}$ can be expressed as*

$$\frac{E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2}{E^{\Lambda\pi}(\lambda\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2} = \frac{\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 E^{\pi} \left\| [\![(\lambda\pi\boldsymbol{\epsilon})/_X]\!]^k - \overline{[\![(\lambda\pi\boldsymbol{\epsilon})/_X]\!]^k} \right\|^2}{\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 \left\| [\![\pi\boldsymbol{\epsilon}]\!]^k - \overline{[\![\pi\boldsymbol{\epsilon}]\!]^k} \right\|^2}, \tag{1.33}$$

where $True_\Lambda$ and BS_Λ are as in Theorem 1.4.1.

Proof: The result is a direct consequence of Lemmas 1.5.1 and 1.5.2.

□

Corollary 1.5.5 *Under the assumptions and using the notation of Corollary 1.5.3 we obtain that if* $\mathbf{1}$ *is the first column of* X *then*

$$\frac{E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2}{E^{\Lambda\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2} = \left(\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 \left\| [\![\pi\boldsymbol{\epsilon}]\!]^k - \overline{[\![\pi\boldsymbol{\epsilon}]\!]^k} \right\|^2\right)^{-1} \tag{1.34}$$

$$\sum_{i,j=2}^{d} \left[\left(\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 \left\| [\![\boldsymbol{w}_i]\!]^k \right\|^2\right)\right.$$

$$\left.\left(\sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\pi\boldsymbol{\epsilon}]\!]^k - \overline{[\![\pi\boldsymbol{\epsilon}]\!]^k} \right\|^2 \left([\![\boldsymbol{w}_i]\!]^k \cdot [\![\boldsymbol{w}_j]\!]^k\right)\right)\right],$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

Proof: The assertion follows from Lemma 1.5.1 and Corollary 1.5.3.

□

As a corollary of Theorem 1.5.4 we obtain a formula for the case of full permutation bootstrap in which the relative $\Lambda\pi$–conditional mean squared discrepancy becomes equal to the unconditional RMSD described in Corollary 1.4.2 above and in Proposition 1 of [17].

Corollary 1.5.6 *If* λ *is a random permutation with distribution uniform over* Σ_n *and* $\boldsymbol{v} \in \mathbb{R}^n$ *is a contrast vector then the* $\Lambda\pi$*–conditional mean squared discrepancy becomes*

$$\frac{E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2}{E^{\Lambda\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2} = \frac{d-1}{n-1} \tag{1.35}$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

Proof: In the assertion of Corollary 1.5.5 take $K = 1$ and write, using the facts that $[\![\boldsymbol{u}]\!]^1 = \boldsymbol{u}$ for any vector $\boldsymbol{u}$ and that $\overline{\boldsymbol{w}_i} = \mathbf{0}$ for $i \geq 2$,

$$\frac{E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2}{E^{\Lambda\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2} = \left(\frac{1}{n-1} \left\| \boldsymbol{v} \right\|^2 \left\| \pi\boldsymbol{\epsilon} - \overline{\pi\boldsymbol{\epsilon}} \right\|^2\right)^{-1}$$

$$\sum_{i,j=2}^{d} \left(\frac{1}{n-1} \|v\|^2 \|w_i\|^2 \right) \left(\frac{1}{n-1} \|\pi\epsilon - \overline{\pi\epsilon}\|^2 (w_i \cdot w_j) \right)$$

$$= \sum_{i,j=2}^{d} \frac{1}{n-1} \|w_i\|^2 w_i \cdot w_j = \frac{d-1}{n-1}.$$

□

1.6 Examples

To illustrate the previous results we will consider a simple linear regression model

$$y_{n\times 1} = \alpha + \beta x_{n\times 1} + \pi\epsilon_{n\times 1}$$

which is a special case of model (1.1) with $\beta_{2\times 1} = (\alpha, \beta)'$ and $X_{n\times 2} = (1, x)$. The random permutation π is assumed to be distributed uniformly over Σ_n while the unknown vector ϵ is assumed to be non-random as in model (1.1).

Under the simple linear regression model the matrix $V = (X'X)^{-1}X'$ has two rows $v_1 = 1/n$ and $v_2 = x/\|x\|^2$. Although the following example is formulated in terms of a general blockwise contrast vector v, we will keep in mind that one would most often be concerned with the $\Lambda\pi$–conditional distribution of $v_2 \cdot \pi\epsilon = \beta - \hat{\beta}$ and its incomplete permutation bootstrap approximation. As before, $\hat{\beta}$ denotes the least squares estimator of β and Λ is a K–block permutation subgroup of Σ_n.

Example 1.6.1 (Single outlier) *Assume that the random permutation λ consists of K independent block permutations and that the vector of errors $\epsilon \in \mathbb{R}^n$ is of a very simple form, namely that $\epsilon = (a, 0, \ldots, 0)'$ for some a.*

Define a sequence of disjoint *events $\{E_k\}_{k=1}^K$ as $E_k = \left\{ \left\| [\![\pi\epsilon]\!]^k \right\|^2 = a^2 \right\}$ for every $k = 1, \ldots, K$. Then for any $v \in \mathbb{R}^n$ with $E\lambda v = 0$ the following hold:*

1. *The estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *is* $\Lambda\pi$*–conditionally unbiased.*

2. *The* $\Lambda\pi$*–*conditional *Mean Squared Error of the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *becomes*

$$E^{\Lambda\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2 = a^2 \sum_{k=1}^{K} \frac{1}{n_k} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 1_{E_k}. \tag{1.36}$$

3. *The ratio of the* $\Lambda\pi$*–*conditional *Mean Squared Error of the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *versus its* unconditional *Mean Square Error is equal to*

$$\frac{E^{\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2}{\mathrm{MSE}_{\mathrm{adj}}} = \frac{1}{\|\boldsymbol{v}\|^2} \sum_{k=1}^{K} \frac{n}{n_k} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 1_{E_k}. \tag{1.37}$$

4. *The* $\Lambda\pi$*–conditional mean squared discrepancy between the true conditional distribution of the errors of the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *and its incomplete permutation bootstrap approximation can be written as*

$$\begin{aligned} & E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2 = \qquad (1.38) \\ & = \frac{a^2}{\|\boldsymbol{x}\|^4} \left(\sum_{k=1}^{K} \frac{1}{n_k} \left\| [\![\boldsymbol{x}]\!]^k \right\|^2 1_{E_k} \right) \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 \left\| [\![\boldsymbol{x}]\!]^k - \overline{[\![\boldsymbol{x}]\!]^k} \right\|^2, \end{aligned}$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

5. *The ratio of the* $\Lambda\pi$*–conditional mean squared discrepancy between the true conditional distribution of the errors of the estimator* $\boldsymbol{v} \cdot \boldsymbol{y}$ *and its incomplete permutation bootstrap approximation versus its* $\Lambda\pi$*–conditional mean squared error, in other words its relative* $\Lambda\pi$*–*conditional *mean square discrepancy is*

$$\begin{aligned} & \frac{E^{\Lambda\pi}(BS_\Lambda - True_\Lambda)^2}{E^{\Lambda\pi}(\lambda \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon})^2} = \qquad (1.39) \\ & = \frac{1}{\|\boldsymbol{x}\|^4} \left(\sum_{k=1}^{K} \frac{\left\| [\![\boldsymbol{x}]\!]^k \right\|^2}{\left\| [\![\boldsymbol{v}]\!]^k \right\|^2} 1_{E_k} \right) \sum_{k=1}^{K} \frac{1}{n_k - 1} \left\| [\![\boldsymbol{v}]\!]^k \right\|^2 \left\| [\![\boldsymbol{x}]\!]^k - \overline{[\![\boldsymbol{x}]\!]^k} \right\|^2, \end{aligned}$$

where $True_\Lambda$ *and* BS_Λ *are as in Theorem 1.4.1.*

Proof: In light of the fact that for every $k = 1, \ldots, K$

$$\left\| [\![\pi\epsilon]\!]^k - \overline{[\![\pi\epsilon]\!]^k} \right\|^2 = a^2 \frac{n_k - 1}{n_k} 1_{E_k}$$

it is easy to see that (1.36) is a direct consequence of Lemma 1.5.1.

Equation (1.38) follows from Lemma 1.5.2. Similarly as in the proof of Corollary 1.5.3 (with K=2 and $E\lambda \boldsymbol{w}_2$ not necessarily equal to 0) we can in (1.31) write

$$\begin{aligned} E^{\Lambda\pi} \left\| [\![(\pi\epsilon)/_X]\!]^k - \overline{[\![(\pi\epsilon)/_X]\!]^k} \right\|^2 &= E^{\Lambda\pi} \left\| (\pi\epsilon \cdot \boldsymbol{w}_2) \left([\![\boldsymbol{w}_2]\!]^k - \overline{[\![\boldsymbol{w}_2]\!]^k} \right) \right\|^2 \\ &= \left\| [\![\boldsymbol{w}_2]\!]^k - \overline{[\![\boldsymbol{w}_2]\!]^k} \right\|^2 E^{\Lambda\pi} (\pi\epsilon \cdot \boldsymbol{w}_2)^2, \end{aligned}$$

and

$$\begin{aligned} E^{\Lambda\pi} (\pi\epsilon \cdot \boldsymbol{w}_2)^2 &= \sum_{m=1}^{K} E^{\pi} \left(\sum_{l=1}^{K} 1_{E_m} [\![\lambda\pi\epsilon]\!]^l \cdot [\![\boldsymbol{w}_2]\!]^l \right)^2 \\ &= \sum_{m=1}^{K} E^{\pi} 1_{E_m} \left([\![\pi\epsilon]\!]^m \cdot \lambda_k [\![\boldsymbol{w}_2]\!]^m \right)^2 \\ &= \sum_{m=1}^{K} \frac{1}{n_m} a^2 \left\| [\![\boldsymbol{w}_2]\!]^m \right\|^2 1_{E_m}. \end{aligned}$$

where $\boldsymbol{w}_2 = \boldsymbol{x} / \|\boldsymbol{x}\|$.

Clearly (1.36) and (1.38) together yield (1.39). Recall that the events $E_1, \ldots, E_k$ are disjoint.

Equation (1.37) is a consequence of (1.36) and Lemma 1.3.3, according to which the mean square error of the estimator is equal to

$$\mathrm{MSE_{adj}} = \mathrm{E}(\lambda \boldsymbol{v} \cdot \pi\epsilon)^2 = \frac{1}{\mathrm{n} - 1} \|\boldsymbol{v}\|^2 \|\epsilon - \bar{\epsilon}\|^2 = \frac{1}{\mathrm{n}} \|\boldsymbol{v}\|^2 \mathrm{a}^2.$$

□

Figure 1.5 shows the conditional distribution of $\boldsymbol{v} \cdot \pi\epsilon$ and its incomplete permutation bootstrap approximation for the case of four–point regression with single

outlier $\epsilon = (1, 0, 0, \ldots, 0)'$. We consider a vector $v = x$ and assume that v consists of two blocks

$$[\![v]\!]^1 = \begin{pmatrix} -10 \\ -10 \\ \vdots \\ 10 \\ 10 \end{pmatrix} \qquad \text{and} \qquad [\![v]\!]^2 = \begin{pmatrix} -3 \\ -3 \\ \vdots \\ 3 \\ 3 \end{pmatrix}$$

with dimensions m_1 and m_2, respectively.

The unconditional distribution of $v \cdot \pi\epsilon$ assigns probabilities $\frac{m_1}{2n}$ to the points -10 and 10 and probability $\frac{m_2}{2n}$ to the points -3 and 3.

The conditional distributions of $v \cdot \pi\epsilon$ given that in $\pi\epsilon$ the error value 1 aligns with $[\![v]\!]^1$, that is $v \cdot \pi\epsilon \mid \pi(1) \in A_1$, consists of the points -10 and 10, each with probability 1/2. Similarly, $v \cdot \pi\epsilon \mid \pi(1) \in A_2$ is uniform over the points -3 and 3.

The full permutation bootstrap of the residuals approximates the unconditional distribution of $v \cdot \pi\epsilon$ while the independent block permutation bootstrap approximates the conditional distributions. The approximations shown in Figure 1.5 were obtained via Monte-Carlo using 5000 random permutations. The dimensions of the blocks were chosen $m_1 = 100$ and $m_2 = 200$.

Figure 1.6 shows the distributions of $v \cdot \pi\epsilon$ for simple linear regression $y = \alpha + \beta x + \pi\epsilon$ with pulse error $\epsilon = (1, 0, 0, \ldots, 0)'$ and

$$x = \begin{pmatrix} -10 \\ -9.8 \\ -9.6 \\ \vdots \\ 9.8 \\ 10 \end{pmatrix}. \tag{1.40}$$

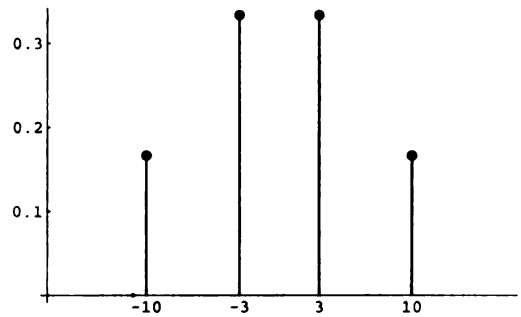
The contrast vector v was chosen equal to x with the first block $[\![v]\!]^1$ consisting of

the 7 largest and 7 smallest components of $\boldsymbol{x}$.

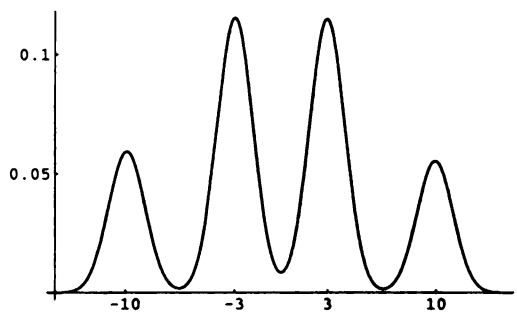
For $i = 1, 2$ the conditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$, given the order statistics of errors in each block, is uniform over the components of $[\![\boldsymbol{v}]\!]^i$ and, as Figure 1.5 shows, is well approximated by the independent block permutation of bootstrap of the residuals. The approximations shown were obtained using Monte-Carlo with 5000 replicas of $\boldsymbol{v} \cdot \lambda(\pi\boldsymbol{\epsilon})^{\perp}|Y$.

Figure 1.9 illustrate conditional distributions of $\boldsymbol{v} \cdot \boldsymbol{\epsilon}$ in the same regression model with errors $\boldsymbol{\epsilon}$ being i.i.d. Gaussian with one outlier. In Figure 1.10, $\epsilon_i = \delta_i U_i^{-2}$ where $\delta_1, \ldots, \delta_n$ are i.i.d. symmetrical random signs and $U_1, \ldots, U_n$ are i.i.d. uniform random variables independent of the random signs.

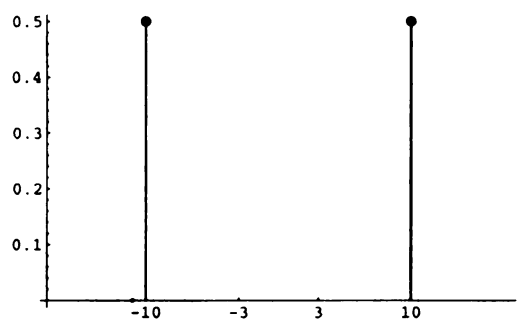
Figure 1.5: Four point linear regression with a single outlier. The conditional and unconditional distributions of the errors $v \cdot \pi\epsilon$ and their corresponding independent block and full permutation bootstrap approximations. The approximations were obtained via Monte–Carlo with 5000 replicas of the corresponding random permutations.



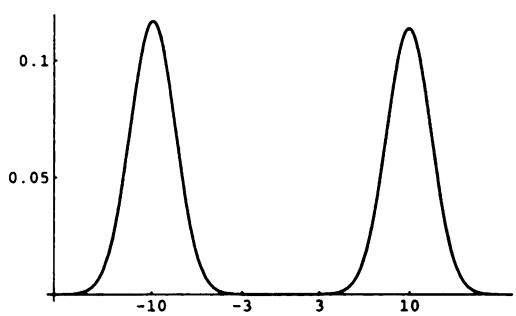
Unconditional distribution of $v \cdot \pi\epsilon$



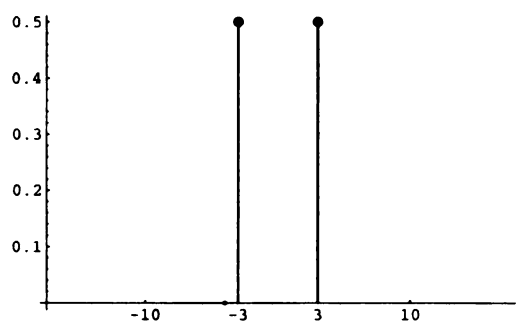
Full permutation bootstrap approximation of $v \cdot \pi\epsilon$



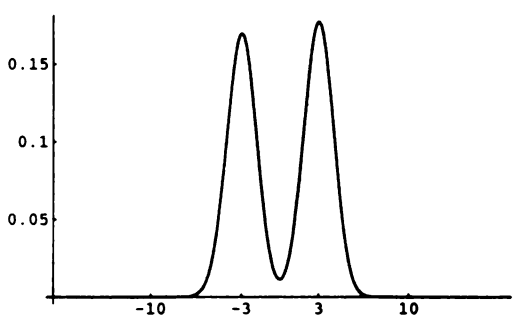
Distribution of $v \cdot \pi\epsilon \mid \pi(1) \in A_1$



Incomplete permutation bootstrap approximation of $v \cdot \pi\epsilon \mid \pi(1) \in A_1$

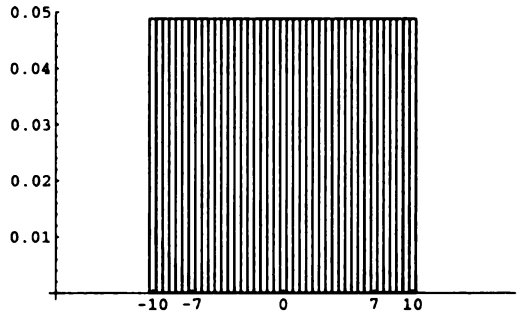


Distribution of $v \cdot \pi\epsilon \mid \pi(1) \in A_2$

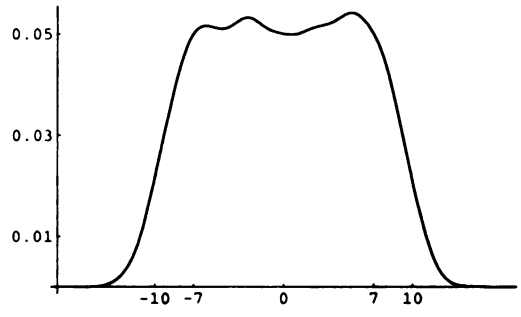


Incomplete permutation bootstrap approximation of $v \cdot \pi\epsilon \mid \pi(1) \in A_2$

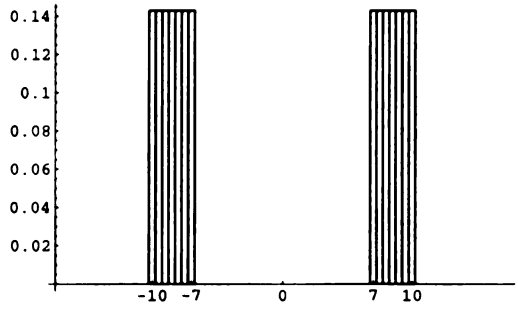
Figure 1.6: The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. The corresponding independent block and full permutation bootstrap approximations were obtained using Monte–Carlo with 5000 replicas of the corresponding random permutations. (Same as Figure 1.3)



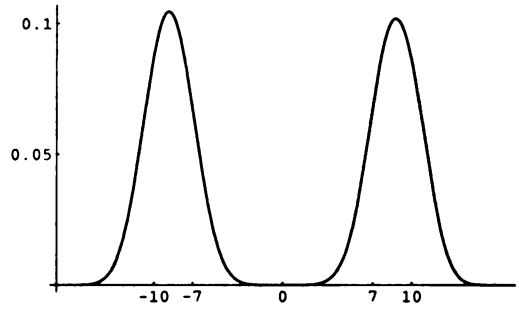
Unconditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



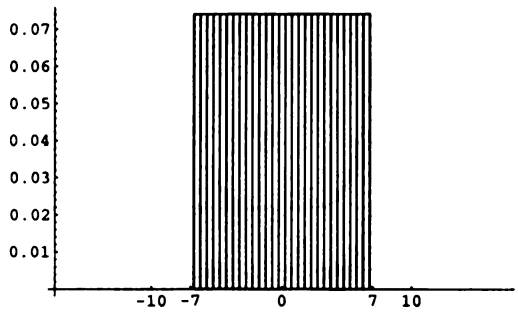
Full permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



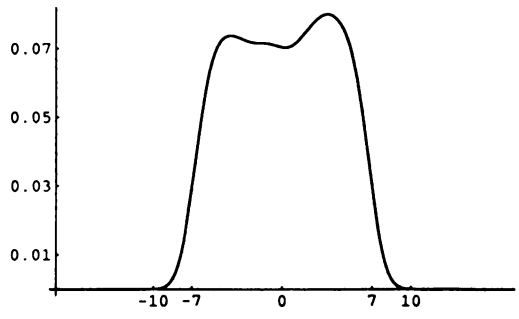
Distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$



Incomplete permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$

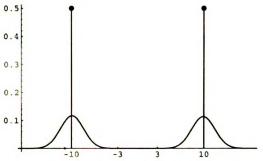


Distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$

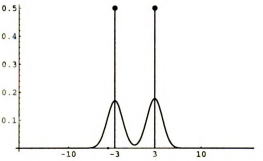


Incomplete permutation bootstrap approximation of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$

Figure 1.7: Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a four point linear regression with a single outlier. (Data from Figure 1.5.)

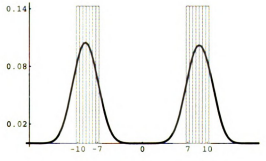


Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$ and its approximation

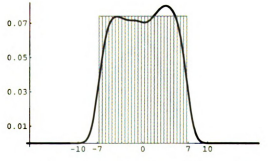


Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$ and its approximation

Figure 1.8: Comparison of the exact and approximated conditional distributions of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with a single outlier. (Same as Figure 1.4, see data in Figure 1.6.)

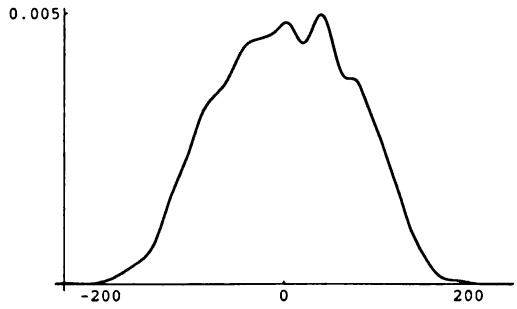


Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_1$ and its approximation

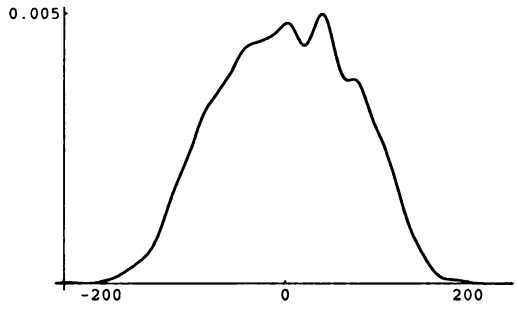


Comparison of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi(1) \in A_2$ and its approximation

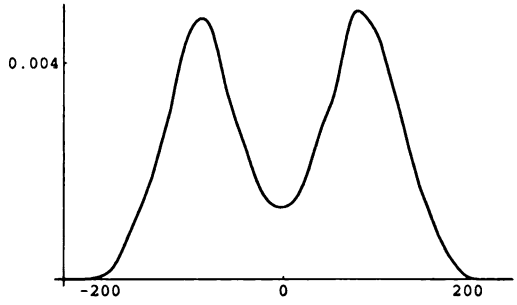
Figure 1.9: The conditional and unconditional distribution of the errors $v \cdot \pi\epsilon$ in a simple linear regression with i.i.d. Gaussian errors. The corresponding independent block and full permutation bootstrap approximations were obtained using Monte–Carlo with 2000 replicas of the corresponding random permutations.



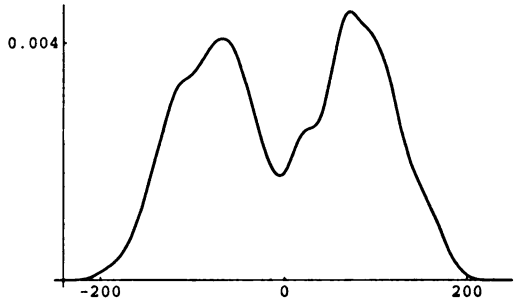
Unconditional distribution of $v \cdot \pi\epsilon$



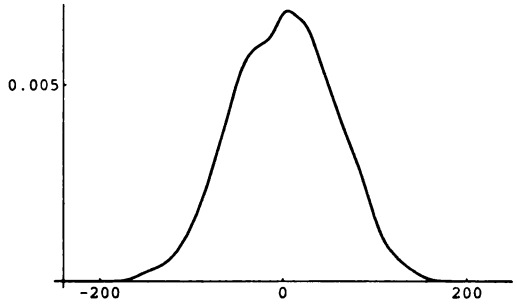
Full permutation bootstrap approximation of the distribution of $v \cdot \pi\epsilon$



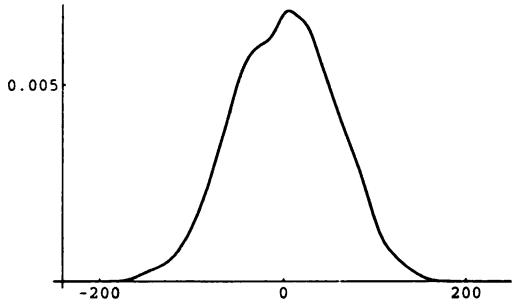
Conditional distribution $\mathcal{L}_1$ of $v \cdot \pi\epsilon$ given that extreme values of $\pi\epsilon$ align with v^1



Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_1$

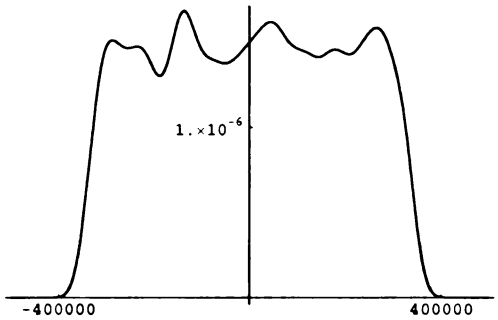


Conditional distribution $\mathcal{L}_2$ of $v \cdot \pi\epsilon$ given that extreme values of $\pi\epsilon$ align with v^2

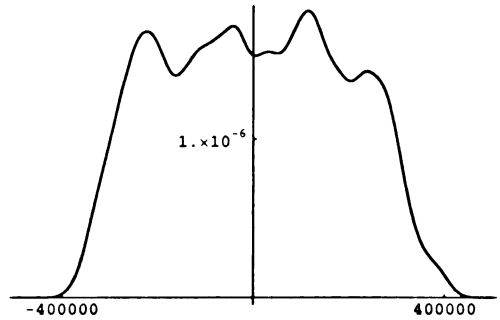


Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_2$

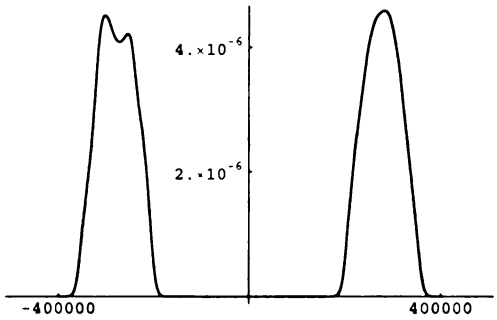
Figure 1.10: The conditional and unconditional distribution of the errors $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ in a simple linear regression with i.i.d. errors from the domain of attraction of a symmetric stable law with $\alpha = 0.5$ (in particular, $\epsilon_i = \delta_i U_i^{-2}$ are reciprocals of squared i.i.d. uniform variables with independent symmetrical random signs). The estimates of the distributions were obtained via Monte–Carlo using 2000 replicas of the corresponding random permutations. (Same as Figure 1.1)



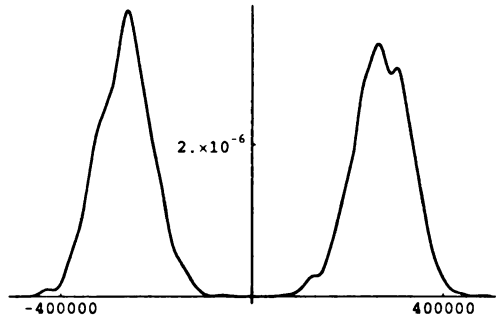
Unconditional distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



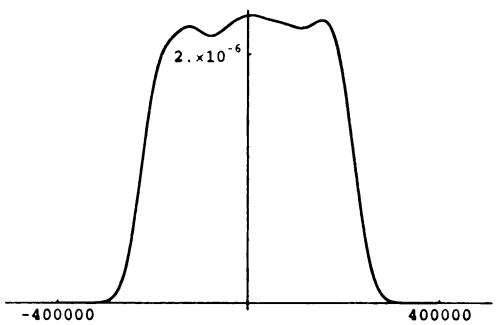
Full permutation bootstrap approximation of the distribution of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$



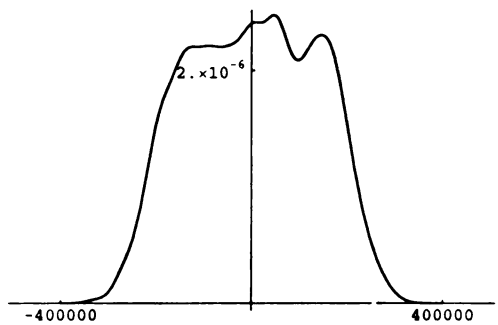
Conditional distribution $\mathcal{L}_1$ of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ given that extreme values of $\pi\boldsymbol{\epsilon}$ align with $\boldsymbol{v}^1$



Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_1$



Conditional distribution $\mathcal{L}_2$ of $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$ given that extreme values of $\pi\boldsymbol{\epsilon}$ align with $\boldsymbol{v}^2$



Incomplete permutation bootstrap approximation of the distribution $\mathcal{L}_2$

1.7 Asymptotics for Independent Block Permutations

Our results in the previous sections show that the conditional distribution $\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \Lambda\pi$, i.e. the distribution $\lambda\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} \mid \pi$ or $\boldsymbol{v} \cdot \lambda\boldsymbol{\epsilon} \mid \boldsymbol{\epsilon}$, can under some conditions on the design be well approximated by incomplete permutation bootstrap of the observed residuals $(\pi\boldsymbol{\epsilon})^{\perp}$ with proper compensation for the conditional bias of the estimator $\boldsymbol{v} \cdot \boldsymbol{y}$.

To support the use of conditional confidence intervals based on incomplete permutations bootstrap and to clarify the relationship of the conditional distribution to the unconditional one we establish a strong invariance principle (as n increases to infinity) for the conditional distributions

$$\boldsymbol{v}_n \cdot \lambda_n \boldsymbol{\epsilon}_n \mid \boldsymbol{\epsilon}_n \tag{1.41}$$

for a sequence of particularly chosen block-wise contrast vectors $\boldsymbol{v}_n \in I\!R^n$ in the special case when for each n the errors $\boldsymbol{\epsilon}_n = (\epsilon_{n,i})_{i=1}^n$ are i.i.d. random variables with distribution attracted to a fixed α–stable law of index $0 < \alpha < 2$, and the random permutation λ_n consists of K independent uniform block permutations.

Theorem 1.7.1 (Limit Theorem) *Let $\mathcal{A} = \{A_k\}_{k=1}^K$ be a partition of the interval $[0,1]$ into K finite unions of intervals, and for every n let λ_n be a random permutation consisting of K independent block permutations induced by $\mathcal{A}$ as described in (1.43).*

For every distribution F in the domain of attraction of α–stable law having index $\alpha < 2$ there exists a triangular array such that for every n the errors in the $n^{-\text{th}}$ row of the array $\boldsymbol{\epsilon}_n = (\epsilon_{n,i})_{i=1}^n$ are i.i.d. F and if v is a continuous function

on $[0,1]$ *with finite variation, satisfying* $\int_{A_k} v = 0$ *for all* $k = 1, \ldots, K$ *then there exist cádlág versions* $L^1, \ldots, L^K$ *of* K *independent* α*-stable Lévy processes on* $[0,1]$ *such that*

$$\frac{1}{a_n} \boldsymbol{v}_n \cdot \lambda_n \epsilon_n \mid \epsilon_n \rightarrow \sum_{k=1}^{K} \mu(A_k)^{-\alpha} \int_0^1 v(f_k) dL^k \mid \mathcal{F}^k, \quad (a.s.) \tag{1.42}$$

as $n \rightarrow \infty$, *where* μ *is the Lebesgue measure, the vectors* $\boldsymbol{v}_n$ *are defined using* v *as specified in (1.44), the scaling constants* a_n *are defined in (1.45), and for every* $k = 1, \ldots, K$ *the* σ*-algebra* $\mathcal{F}^k$ *is generated by the jumps of the process* L^k *while* f_k *is an appropriate monotone, piece-wise linear scaling of [0,1] onto* A_k.

Those familiar with convergence to stable laws may note the absence of centering in (1.42), but this is taken care of by centering $\boldsymbol{v}_n$ (see (1.46)).

Proof: We will adopt a modification of the approach used by authors in [17] who obtained similar results for the complete permutation bootstrap case. This method is based on a series representation of a cádlág version of a Lévy process which is described in great detail in [18].

Assume that for every n the permutation λ_n is generated by K independent block permutations, i.e. that for a fixed integer K there is a partition $\mathcal{A} = \{A_k\}_{k=1}^K$ of the interval $[0,1]$ with A_k being a finite union of intervals for every $k = 1, \ldots, K$, and that for every n the random permutation λ_n applied to a vector $\boldsymbol{u} \in \mathbb{R}^n$ satisfies

$$[\![\lambda_n \boldsymbol{u}]\!]^k = \lambda_n^k [\![\boldsymbol{u}]\!]^k, \quad k = 1, \ldots, K, \tag{1.43}$$

where $\{\lambda_n^k\}_{k=1}^K$ are independent random permutations, each distributed uniformly over $\Sigma_{m_n^k}$. Here m_n^k denotes the dimension of $[\![\boldsymbol{u}]\!]^k = (u_i)_{i \in nA_k}$ which, similarly as before, represents a vector consisting of the $k^{-\text{th}}$ block of components of the vector

$\boldsymbol{u}$. Notice that, in contrast to our previous definition, the set of indices $\{1,\ldots,n\}$ is partitioned into K blocks by the sets in $n\mathcal{A}$.

We will generalize the notation of block vectors to functions on $[0,1]$ as follows. For a positive integer n and a function $v:[0,1]\to\mathbb{R}$ define $[\![v]\!]_n$ as the n–dimensional vector with components equal to values of v sampled at the successive points i/n, $i=1,\ldots,n$. For every $k=1,\ldots,K$ let $[\![v]\!]_n^k$ denote the $k^{-\text{th}}$–block of $[\![v]\!]_n$.

This notation will be used to specify the special form of the sequence of contrast vectors $\{\boldsymbol{v}_n\}_{n\geq 1}$. For what follows we will assume that $v\colon[0,1]\to\mathbb{R}$ is a continuous function with finite variation satisfying

$$\int_{A_k} v(x)dx = 0, \quad k=1,\ldots,K,$$

and for every n we will define a vector $\boldsymbol{v}_n\in\mathbb{R}^n$ by requiring

$$[\![\boldsymbol{v}_n]\!]^k = [\![v]\!]_n^k - \overline{[\![v]\!]_n^k}\,, \quad k=1,\ldots,K. \tag{1.44}$$

The specific choice of the vectors $\boldsymbol{v}_n$ guarantees that all the vectors $\boldsymbol{v}_n$ are blockwise contrasts, in other words,that all the corresponding blocks $[\![\boldsymbol{v}_n]\!]^k$ are contrast vectors. As noted earlier, this condition is equivalent to the condition that $E\lambda_n\boldsymbol{v}_n = 0$ for all $n\geq 1$.

Let us now state the results for the full permutation case as developed in [18]. For a random variable X with distribution F define $G_+(x)=P(X\geq x\mid X\geq 0)$, $G_-(x)=P(-X>x\mid -X>0)$, and $G(x)=P(|X|\geq 0)$. The inverse of a function F will be defined as $F^{-1}(x)=\inf\{y\geq 0\colon F(y)\leq x\}$. For every n let

$$a_n = G^{-1}(1/n) \tag{1.45}$$

If F belongs to the domain of attraction of an α–stable law with $\alpha<2$ then

the following limits exist (see [18])

$$p = \lim_{x\to\infty} \frac{P(X \geq x)}{G(x)}, \qquad q = \lim_{x\to\infty} \frac{P(-X > x)}{G(x)}.$$

For a sequence $u = (u_i)_{i=1}^{\infty}$ of numbers from $[0,1]$ and $0 \leq t \leq 1$ define (as in [15], reprinted in [16])

$$\begin{aligned} I_1^n(u,t) &= 1_{\{s:\, u_1 n \leq [sn]\}}(t) \text{ and recursively} \\ I_i^n(u,t) &= 1_{\{s:\, u_i(n+1-i) \leq [sn] - \sum_{j=1}^{i-1} I_j^n(u,s)\}}(t), \end{aligned}$$

where $i = 1, \ldots, n$.

If $\Gamma = (\Gamma_i)_{i=1}^{\infty}$ is a sequence of the arrival times of a standard Poisson process, $U = (U_i)_{i=1}^{\infty}$ is a sequence of i.i.d. random variables with uniform distribution over the interval $[0,1]$, and if $\delta = (\delta_i)_{i=1}^{\infty}$ is a sequence of random, possibly non-symmetric signs such that Γ, U, and δ are mutually independent then according to [18, Lemma 3 and (11)] the random vector

$$\epsilon_n = (\epsilon_{n,i})_{i=1}^n = \pi_n^U \left(\delta_i G_{\delta_i}^{-1} \left(\frac{\Gamma_i}{\Gamma_{n+1}} \right) \right)_{i=1}^n$$

consists of i.i.d. F random variables. Here π_n^U represents a uniform random permutation of $\{1, \ldots, n\}$ chosen by U via the positions $(J_i^n)_{i=1}^n$ of jumps of the functions $(I_i^n(U,t))_{i=1}^n$.

According to Theorem 2 in [18]: Let $\alpha \in (0,2)$ and let v be a continuous real function with bounded variation with $\int_0^1 v = 0$. Then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{i=1}^n v(J_i^n) \delta_i \, G_{\delta_i}^{-1} \left(\frac{\Gamma_i}{\Gamma_{n+1}} \right) = \sum_{i=1}^{\infty} \frac{z_i \, v(U_i)}{\Gamma_i^{1/\alpha}}, \quad \text{(a.s.)}, \tag{1.46}$$

where

$$z_i = \left(\frac{p}{P(X \geq 0)} \right)^{1/\alpha} \frac{\delta_i + 1}{2} + \left(\frac{q}{P(-X > 0)} \right)^{1/\alpha} \frac{\delta_i - 1}{2}.$$

Note that for every n the sum under the limit in (1.46) can be written as $\frac{1}{a_n}[\![v]\!]_n \cdot \pi_n^U \epsilon$. The right side of (1.46) can be according to [18] viewed as $\int v dL$ where L is a cádlág version of an α-stable Lévy motion on $[0,1]$.

Theorem 2 in [18] also states that if $b_n = E(z_i \int_{1/n}^n G_{\delta_i}(x)dx)n/a_n$ then the scaled and centered sums $1/a_n \sum_{i=1}^n \epsilon_{n,i} - b_n$ converge (a.s.) to a finite random variable as in (1.48) below with $A = [0,1]$. Note that $|b_n| \leq B\, n/a_n$ for some $B > 0$, therefore if $\{c_n\}$ is a sequence of real numbers with $nc_n/a_n \to 0$ then

$$\lim_{n\to\infty} \frac{c_n}{a_n} \sum_{i=1}^n \epsilon_{n,i} = 0 \quad \text{(a.s.)}. \tag{1.47}$$

If A is a finite union of intervals in $[0,1]$ we can use the same construction to obtain an (a.s.) limit for the sums of $[n\mu(A)]$ random variables

$$\lim_{n\to\infty} \frac{1}{a_{n\mu(A)}} \sum_{i\in nA} v(f(J_i^n))\delta_i\, G_{\delta_i}^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right) = \sum_{i=1}^{\infty} \frac{z_i\, v(f(U_i))}{\Gamma_i^{1/\alpha}}, \quad \text{(a.s.)}, \tag{1.48}$$

where f is a scaling of $[0,1]$ onto A.

We can therefore construct K independent (a.s.) limits for $\frac{1}{a_n}[\![v]\!]_n^k \cdot \lambda_{n,k}^{U_k}\epsilon_{n,k}$ as $\mu(A_k)^\alpha \int v(f_k)dL^k$. Here we use that if F is α–stable law then for every $x > 0$ it holds that $\lim_{n\to\infty} a_{nx}/a_n = x^\alpha$.

The facts that v is continuous and $\int_{A_k} v d\mu = 0$ guarantee that $\overline{[\![v]\!]_n^k} \to 0$. Replacing c_n in (1.47) by $\overline{[\![v]\!]_n^k}$ we obtain that $\frac{1}{a_n}\overline{[\![v]\!]_n^k} \cdot \epsilon_{n,k}$ converges to 0 (a.s.), for every k as $a_n \to \infty$, hence we obtain that for every $k = 1, \ldots, K$

$$\lim_{n\to\infty} \frac{1}{a_n} [\![v_n]\!]^k \cdot \lambda_{n,k}^{U_k}\epsilon_{n,k} = \mu(A_k)^\alpha \int v(f_k)dL^k, \quad \text{(a.s.)}.$$

Similarly as in Proposition 1 of [18], the conditional convergence result is obtained via conditioning on (Γ_i^k, δ_i^k) which is equivalent to conditioning on the jumps of the corresponding processes.

□

1.8 Exact Confidence Intervals Based on Incomplete Permutation Tests

Quade (1973) ([20]; see also [19] and [9]) has found a pivot which can be used to develop an exact confidence interval for β in the *simple* regression model

$$\boldsymbol{y} = \alpha\mathbf{1} + \boldsymbol{x}\beta + \pi\boldsymbol{\epsilon} \tag{1.49}$$

where $\boldsymbol{x}$ is a known vector in $\mathbb{R}^n$ and α and β are two unknown parameters. For simplicity we will assume that $\boldsymbol{x}$ is a contrast, i.e. $\boldsymbol{x} \cdot \mathbf{1} = 0$.

The model (1.49) is a special case of the model (1.1) $\boldsymbol{y} = X \cdot \boldsymbol{\beta} + \pi\boldsymbol{\epsilon}$ with the matrix X having two columns $\mathbf{1}$ and $\boldsymbol{x}$ and $\boldsymbol{\beta} = (\alpha, \beta)'$. The vector $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is, similarly as in (1.1), considered to be non-random but unknown, and the random permutation π is assumed to be distributed uniformly over the group Σ_n of all permutations of components in $\mathbb{R}^n$. Lastly, we assume that the random permutation λ is distributed uniformly over a subgroup Λ of Σ_n.

Let $\boldsymbol{v}$ denote the second row of the matrix $V = (X'X)^{-1}X'$, in other words let $\boldsymbol{v} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|^2}$, and recall that $\boldsymbol{v}$ is a contrast. Denote by $\hat{\beta} = \boldsymbol{v} \cdot \boldsymbol{y}$ the least squares estimator of β.

Under the hypothesis H_0: $\beta = \beta_0$ we can observe that the vector of marginal errors $\boldsymbol{\epsilon}^0 = \boldsymbol{y} - \bar{\boldsymbol{y}} - \boldsymbol{x}\beta_0$ is equal to $\pi\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}$. Therefore under the hypothesis H_0 we are able to recover the centered errors without postulating the value of α.

A permutation test for testing H_0 can be based on the test statistic

$$T_\pi(\beta_0) = \boldsymbol{v} \cdot \boldsymbol{\epsilon}^0$$

which under H_0 has $\Lambda\pi$-conditional distribution that is uniformly distributed over the indices of the values of $\boldsymbol{v} \cdot \lambda\boldsymbol{\epsilon}^0$. Note that if H_0 holds then $T_\pi(\beta_0)$ is equal to

$\boldsymbol{v} \cdot \pi\boldsymbol{\epsilon} = \hat{\beta} - \beta_0$.

Denote by T^* the random variable $\boldsymbol{v} \cdot \lambda\boldsymbol{\epsilon}^0$ which is $\Lambda\pi$-conditionally independent of $T_\pi(\beta_0)$, but under H_0: $\beta = \beta_0$ has the distribution $T_\pi(\beta_0) \mid \Lambda\pi$. Then the p-value for the observed value t_0 of the test statistic $T_\pi(\beta_0)$ can be expressed as $2\min(P^+, P^-, \frac{1}{2})$ where $P^+ = P(T^* > t_0)$ and $P^- = P(T^* < t_0)$.

Lemma 1.8.1 (Extension of Quade's result in [20]) *Let $T_\pi(\beta_0)$, T^*, and λ be as before. Assume in addition that there are no ties among the values of T^*. Then an α-level confidence region for β can be formed as*

$$B = \{\beta_0 \colon \tfrac{\alpha}{2} < P^\pi(T^* > T_\pi(\beta_0)) \leq 1 - \tfrac{\alpha}{2}\}, \tag{1.50}$$

where α is of the form $\frac{2j}{|\Lambda|}$ for some integer $0 < j < 2|\Lambda|$.

Proof: Denote by $P_{\beta_0}^{\Lambda\pi}$ the probability measure under the assumption that β_0 is the true value of the parameter β, conditional on $\Lambda\pi$. Let the π-conditional distribution function of T^* be denoted by F_{β_0}. To examine the probability that β_0 is not covered by the confidence region B we will write

$$P_{\beta_0}^{\Lambda\pi}(\beta_0 \notin B) = P_{\beta_0}^{\Lambda\pi}(P^\pi(T^* \leq T_\pi(\beta_0)) \leq \tfrac{\alpha}{2}) + P_{\beta_0}^{\Lambda\pi}(P^\pi(T^* \leq T_\pi(\beta_0)) > 1 - \tfrac{\alpha}{2}) \tag{1.51}$$

and observe that $P_{\beta_0}^{\Lambda\pi}(P^\pi(T^* \leq T_\pi(\beta_0)) \leq \frac{\alpha}{2}) = P_{\beta_0}^{\Lambda\pi}(F_{\beta_0}(T_\pi(\beta_0)) \leq \frac{\alpha}{2}) = \frac{\alpha}{2}$ since $F_{\beta_0}(T_\pi(\beta_0))$ is distributed uniformly over $\{\frac{k}{|\Lambda|} \colon k = 0, \ldots, |\Lambda|\}$ and α is of the form $\frac{2j}{|\Lambda|}$, for some j. Similarly, the second probability on the right hand side in (1.51) is equal to $\frac{\alpha}{2}$. Hence $P_{\beta_0}^{\Lambda\pi}(\beta_0 \notin B) = \alpha$.

□

The confidence region in (1.50) has the desired confidence level, but is hard to construct since for each value of β_0 the probability $P^\pi(T^* > T_\pi(\beta_0))$ has to be

computed individually. To simplify the construction of the confidence region B we will rewrite its definition in a more useful form using a pivotal random variable which does not depend on the value of β_0. A special case of this method, for the permutation group Σ_n, was first introduced by Quade in 1973.

Proposition 1.8.1 (Extends Quade [20]) *Let* $\alpha = \frac{2j}{|\Lambda|}$ *for some integer* $0 < j < 2|\Lambda|$. *Then the* α*-level confidence region (1.50) can be obtained as*

$$B = \{\beta_0 \colon \tfrac{\alpha}{2} < P^{\pi}(\hat{\beta} - \tfrac{\boldsymbol{v}\cdot\lambda(\pi\epsilon)^{\perp}}{(1-(\boldsymbol{v}\cdot\lambda\boldsymbol{x}))} < \beta_0) < 1 - \tfrac{\alpha}{2}\}. \tag{1.52}$$

Proof: Recall first that if H_0 holds then $T_{\pi}(\beta_0) = \hat{\beta} - \beta_0$. Thus (1.50) can be written as

$$B = \{\beta_0 \colon \tfrac{\alpha}{2} < P^{\pi}(\boldsymbol{v}\cdot\lambda\boldsymbol{\epsilon}^0 > \hat{\beta} - \beta_0) < 1 - \tfrac{\alpha}{2}\}.$$

Notice further that under H_0

$$(\pi\boldsymbol{\epsilon})^{\perp} = \boldsymbol{y} - \bar{\boldsymbol{y}} - \boldsymbol{x}\hat{\beta} = \boldsymbol{\epsilon}^0 - \boldsymbol{x}(\hat{\beta} - \beta_0)$$

which implies that $\boldsymbol{v}\cdot\lambda\boldsymbol{\epsilon}^0 = \boldsymbol{v}\cdot\lambda(\pi\boldsymbol{\epsilon})^{\perp} + (\hat{\beta} - \beta_0)(\boldsymbol{v}\cdot\lambda\boldsymbol{x})$. Hence the probability $P^{\pi}(\boldsymbol{v}\cdot\lambda\boldsymbol{\epsilon}^0 > \hat{\beta} - \beta_0)$ becomes

$$P^{\pi}(\boldsymbol{v}\cdot\lambda(\pi\boldsymbol{\epsilon})^{\perp} > (\hat{\beta} - \beta_0)(1 - (\boldsymbol{v}\cdot\lambda\boldsymbol{x})))$$

and the α-level confidence region (1.50) can be obtained as

$$B = \{\beta_0 \colon \tfrac{\alpha}{2} < P^{\pi}(\hat{\beta} - \tfrac{\boldsymbol{v}\cdot\lambda(\pi\epsilon)^{\perp}}{(1-(\boldsymbol{v}\cdot\lambda\boldsymbol{x}))} < \beta_0) < 1 - \tfrac{\alpha}{2}\}$$

which completes the proof.

□

The pivot $\hat{\beta} - \frac{\boldsymbol{v}\cdot\lambda(\pi\boldsymbol{\epsilon})^{\perp}}{(1-(\boldsymbol{v}\cdot\lambda\boldsymbol{x}))}$ does not depend on the hypothetical value β_0 of the parameter β, therefore the confidence region B can be constructed for example by resampling from the distribution of λ.

Chapter 2

Wavelet Expansion Model

2.1 Introduction

Consider a problem of estimating a function f that has been observed with errors at N points

$$y_i = f(t_i) + \epsilon_i, i = 1, \ldots, N, \tag{2.1}$$

where $(t_i)_{i=1}^N$ is an equidistant division of the interval $[0, 1]$ and the random vector of errors $(\epsilon_i)_{i=1}^N$ has a distribution with exchangeable components. Authors in [5] use wavelet shrinkage for model (2.1) with i.i.d. normal errors.

Here we consider a modified discrete wavelet expansion of the function f as a means of estimating the function f at the points $(t_i)_{i=1}^N$. Our modification of the wavelet expansion will allow us to use permutation resampling of the residuals to find conditional confidence regions for the wavelet coefficients of the function f.

The only assumption about the distribution of the vector of errors is exchangeability of the components.

2.2 Modified Discrete Wavelet Transform

Consider a discrete wavelet transform based on a finite collection $\mathcal{W}_0$ of wavelets and a corresponding matrix W_0 with columns representing discrete versions of the wavelets, evaluated at the points of interest $(t_i)_{i=1}^N$.

In many situations all the columns of the matrix W_0 are orthogonal to the vector $\mathbf{1}$, therefore in such cases the results of this dissertation can be used directly for the model (1.1) with $\mathbf{1} \perp X$.

Also in cases when the column space of W_0 contains the vector $\mathbf{1}$ we can use our results directly. For example the Haar wavelet basis on a finite interval contains the constant function 1 hence our results can be utilized.

The modification of the original concept of discrete wavelet transforms is concerned with the case when neither of the above situations applies, that is when $\mathbf{1}$ is neither orthogonal to nor contained in the column space of W_0. Is such a case, instead of the wavelet collection $\mathcal{W}_0$ we use a collection of functions

$$\mathcal{W} = \{g_1\} \cup \mathcal{W}_0 = \{g_1, g_2, \ldots, g_K\}, \tag{2.2}$$

where g_1 is the constant function 1 on the interval $[0, 1]$. In place of the matrix W_0 we use a matrix W with columns $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K\}$ corresponding to the functions in $\mathcal{W}$.

This modification allows us to assume that the first column $\boldsymbol{w}_1$ of the matrix W is the vector $\mathbf{1}$ with all components equal to 1, which is required to use our results.

For simplicity we will also use $\mathcal{W}$ and W in place of $\mathcal{W}_0$ and W_0, respectively, when the modification of the wavelet transform is not required, that is when $\mathbf{1} \perp W$

or when $\mathbf{1}$ is spanned by the columns of W_0.

2.2.1 Linear Regression Model

Let f^* denote the L_2 projection of the function f on the $L_2([0,1])$ subspace spanned by the collection of functions in $\mathcal{W}$. We can write $f^* = \sum_{k=1}^{K} g_k \theta_k$ (a.s.) for some $\theta_1, \ldots, \theta_K$.

Consider now a problem of estimating f^* using the same observations and errors as in (2.1). This problem can be described by the following linear regression model

$$y_i = \sum_{k=1}^{K} (\boldsymbol{w}_k \theta_k) + \epsilon_i, i = 1, \ldots, N, \tag{2.3}$$

which in matrix form becomes

$$\boldsymbol{y} = W \cdot \boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{2.4}$$

where $\boldsymbol{y}$ is the vector of N observations $\{y_1, \ldots, y_N\}$, W is an $N \times K$ matrix, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_K\}$ is the vector of unknown parameters, and $\boldsymbol{\epsilon}$ is an N–dimensional random vector with exchangeable distribution.

The conditional version of model (2.4), given the order statistics of $\boldsymbol{\epsilon}$, is a special case of model (1.1) with $\boldsymbol{\theta}$ playing the role of $\boldsymbol{\beta}$ and W in place of the matrix X.

The assumption that $\mathbf{1} \perp X$ or that $\mathbf{1}$ is the first column of X will in light of our modification of the wavelet transform be always satisfied. The matrix W_0, and ergo also the matrix W, usually has full rank, therefore all the assumptions of the model (1.1) are satisfied.

2.2.2 Asymptotics for the Modified Wavelet Transform

The modification of the wavelet transform described above is essential in that it allows us to use our method of incomplete permutation bootstrap to approximate

the conditional distribution of the estimation errors. The effect of the modification on the original wavelet expansion becomes less important as the number K of wavelets used in the expansion increases.

Proposition 2.2.1 *The difference between the modified and traditional wavelet transforms vanishes with increasing K in the L_2 sense as described by (2.5), (2.7), and (2.8).*

Let us consider the case when $\mathcal{W}$ corresponds to the modified wavelet transform as described in (2.2). Let us define $f^* = f/_{\mathcal{W}}$ and $f^\perp = f/_{\mathcal{W}^\perp}$. We can then rewrite the function f in a few useful ways as follows:

$$f = f/_{\mathcal{W}} + f/_{\mathcal{W}^\perp} = f^* + f^\perp \tag{2.5}$$

$$f = f/_{\mathcal{W}_0} + f/_{\mathcal{W}_0^\perp} \tag{2.6}$$

$$f = f/_{\mathcal{W}_0} + f/_{\mathcal{W}_1} + f/_{\mathcal{W}^\perp}, \tag{2.7}$$

where $\mathcal{W}_1 = \{g_1/_{\mathcal{W}_0^\perp}\}$. Recall that g_1 represents the constant function 1 on $[0,1]$ and note that in the representation (2.7) all three sets $\mathcal{W}_0$, $\mathcal{W}_1$, and $\mathcal{W}_1^\perp$ are mutually orthogonal.

According to the theory of discrete wavelet transforms, $f/_{\mathcal{W}_0^\perp} \xrightarrow{L_2} 0$ as $K \to \infty$. Therefore, with $K \to \infty$ also $f/_{\mathcal{W}^\perp} \xrightarrow{L_2} 0$ using $\mathcal{W}_0 \subseteq \mathcal{W}$ in (2.5) and (2.6). As an easy consequence of (2.7) we then obtain that also

$$f/_{\mathcal{W}_1} \xrightarrow{L_2} 0. \tag{2.8}$$

2.3 Illustration

We will illustrate the potential benefits of incomplete permutation bootstrap in the area of wavelet expansion in a simulation study by recovering certain conditional

Figure 2.1: Mexican hat mother wavelet $\psi(t) = (1 - t^2)e^{-t^2/2}$.

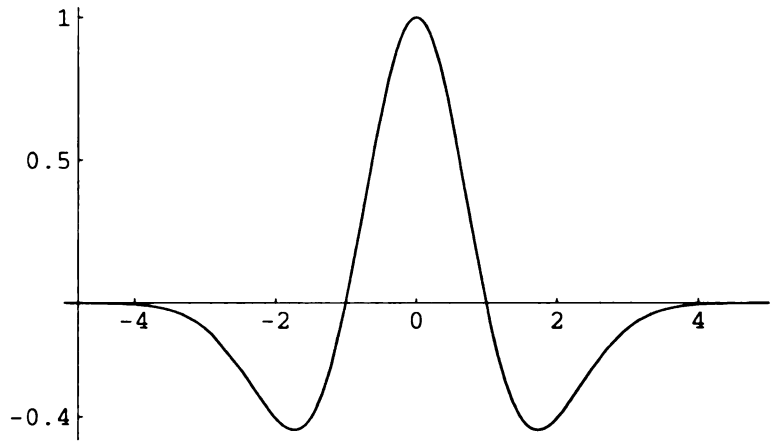
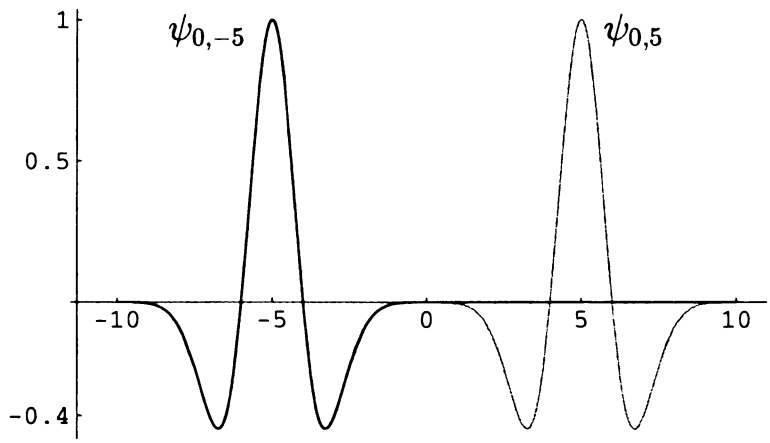


Figure 2.2: Wavelets $\psi_{0,-5}$ and $\psi_{0,5}$ based on Mexican hat mother wavelet.



distributions of the estimation errors for one wavelet coefficient in a simplified discrete wavelet expansion by two selected wavelets.

For the sake of simplicity we will consider wavelets based on the Mexican hat mother wavelet

$$\psi(t) = (1 - t^2)e^{-t^2/2}$$

which is shown in Figure 2.1. For any two integers m, n (possibly negative) a wavelet $\psi_{m,n}$ will be defined using a shifted and scaled or dilated version of the mother wavelet ψ

$$\psi_{m,n} = 2^{-m/2}\psi(2^{-m}x - n).$$

(See [7], [4] or [12] for details on orthonormal wavelet bases and multi–resolution analysis.)

Let us consider a special case of the regression model (2.4) with the matrix W consisting of three columns $\boldsymbol{w}_1, \boldsymbol{w}_2$ and $\boldsymbol{w}_3$ with $\boldsymbol{w}_1 = \mathbf{1}$, and $\boldsymbol{w}_2$ and $\boldsymbol{w}_3$ equal to n–dimensional discrete versions of wavelets $\psi_{0,-5}$ and $\psi_{0,5}$, respectively. The wavelets $\psi_{0,-5}$ and $\psi_{0,5}$ (which are based on the Mexican hat mother wavelet ψ) are shown if Figure 2.2.

In the simulation study we have considered independent Gaussian errors with standard deviation of 0.1, and discrete versions of the wavelets with $n = 400$. All distributions were approximated using Monte–Carlo based on 2000 replicas from the distribution of the corresponding random permutations.

Figure 2.3 compares the distribution of the estimation errors for the coefficient θ_2 of $\psi_{0,-5}$ (conditional on the order statistics of $\boldsymbol{\epsilon}$) to its approximation based on full permutation bootstrap of the observed residuals.

To illustrate the approximation of blockwise conditional distributions of the

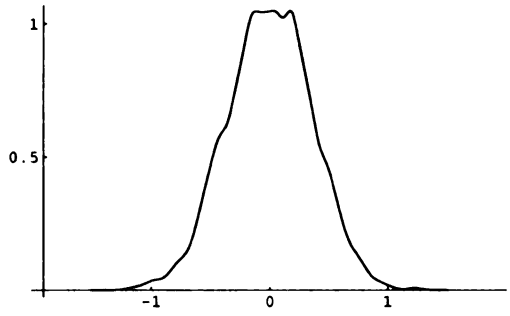
estimation errors we consider distributions conditional on which of two blocks of indices extreme values of the errors ϵ align with. These two blocks of indices are in this particular case hinted at by the form of the vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$.

Assume that n is even and consider a random permutation λ consisting of two independent block permutations λ_1 and λ_2 which uniformly permute the first $n/2$, respectively the last $n/2$ components of the vector to which λ is being applied. In the case of the vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, λ thus permutes independently their close-to zero and non-zero values among themselves.

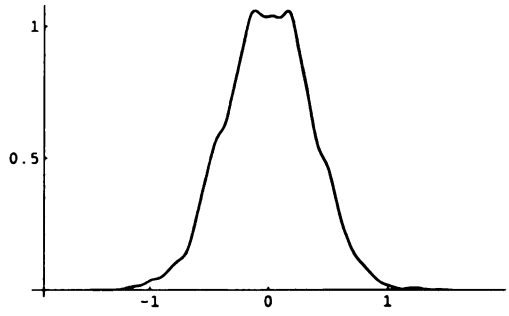
The next three figures illustrate the approximation of the conditional distribution of the estimation errors for θ_2. The approximation is based on incomplete permutation bootstrap of the observed residuals using Monte-Carlo from the distribution of the random permutation λ.

Figures 2.4 and 2.5 show approximation of the distribution conditional on whether the first $n/2$ components of ϵ contain errors with small or large absolute values, while Figure 2.6 illustrates approximation of the block-wise conditional distribution of the estimation errors when the components of ϵ are not in a particular order. A comparison of the distributions and their permutation bootstrap approximations are shown in Figure 2.7.

Figure 2.3: The conditional distribution of the estimation errors for the coefficient θ_2 of $\psi_{0,-5}$, given the order statistics of the errors ϵ(left), and its estimate based on full permutation bootstrap of the observed residuals (right).

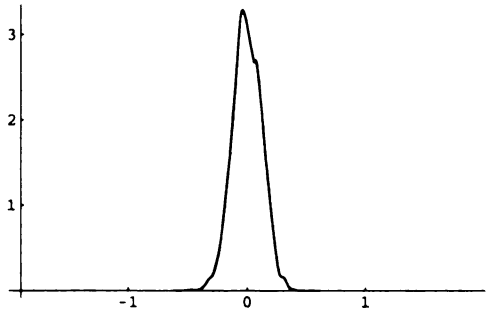


True distribution of the errors

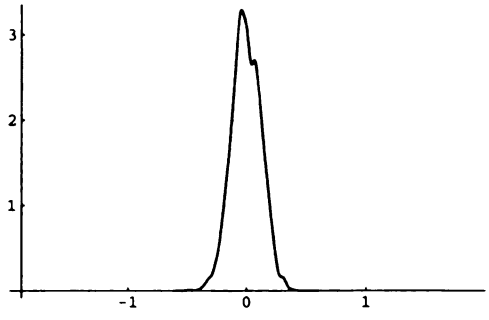


Full permutation bootstrap approximation

Figure 2.4: The incomplete permutation bootstrap approximation of the distribution of the estimation errors for the coefficient θ_2, conditional on the fact that the errors with large absolute values align with the last $n/2$ components of ϵ. In the simulation study the components of ϵ have been sorted to satisfy this condition.

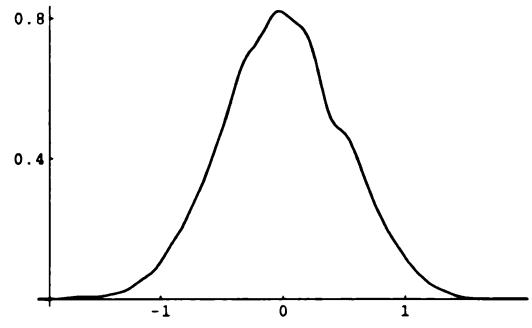


True conditional distribution of the errors

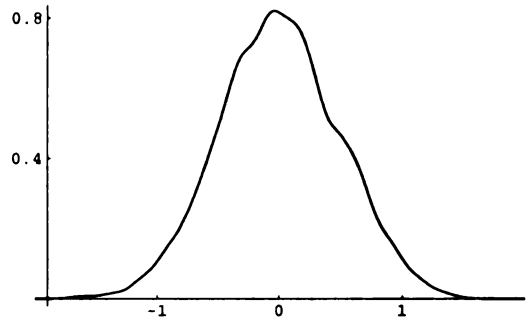


Incomplete permutation bootstrap approximation

Figure 2.5: The approximation of the conditional distribution of the estimation errors for the coefficient θ_2, given that the errors with large absolute values lie among the first $n/2$ components of ϵ. In the simulation study the components of ϵ have been sorted to satisfy this condition.

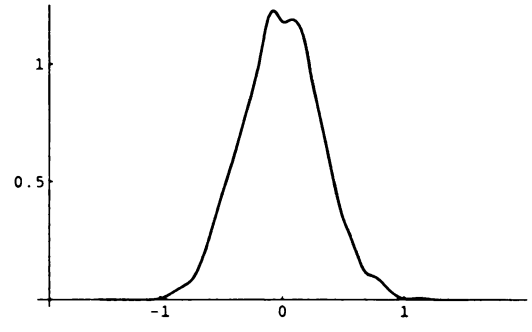


True conditional
distribution of the errors

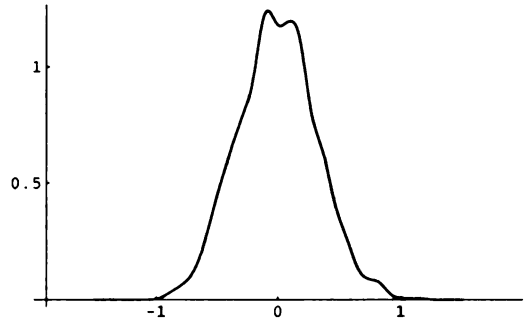


Incomplete permutation
bootstrap approximation

Figure 2.6: For a vector of errors with randomly arranged components, the block-wise conditional distribution of the estimation errors for the coefficient θ_2, given the order statistics of the errors in each of the two blocks are well recovered by the incomplete permutation bootstrap of the observed residuals. In this particular case the true block-wise conditional distribution appears similar to the conditional distribution given the order statistics of ϵ (see Figures 2.3 and 2.7 for comparison).

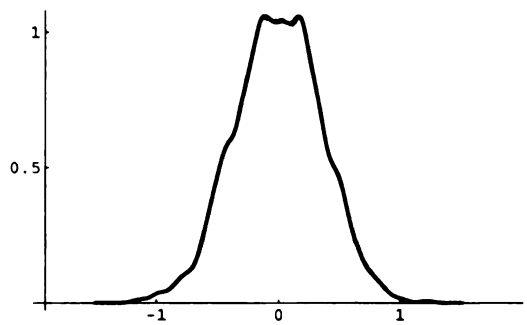


True conditional
distribution of the errors

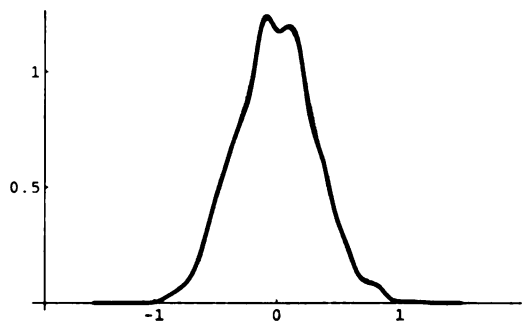


Incomplete permutation
bootstrap approximation

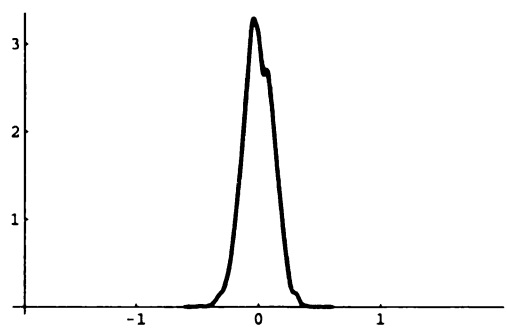
Figure 2.7: Comparison of conditional distributions of the estimation errors for the coefficient θ_2 and the corresponding incomplete permutation bootstrap recoveries. Full permutation bootstrap was used for the distribution conditional on the order statistics of ϵ.



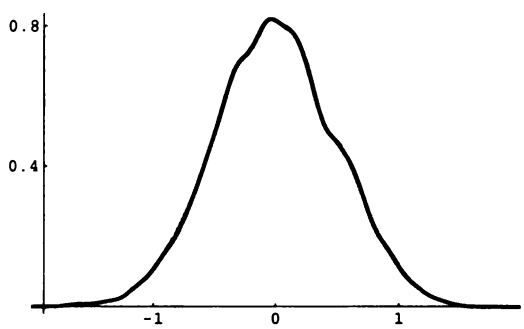
Distribution conditional on the order statistics of ϵ



Distribution conditional on the order statistics of the errors in each block.



Conditional distribution given large absolute errors fall into the second block



Conditional distribution given large absolute errors fall into the first block

2.4 Conclusions - Incomplete Permutations of Residuals

As shown above, the problem of estimating the coefficients of discrete wavelet transforms of functions in L_2 with additive exchangeable noise is a special case of model (1.1). The results of this thesis are in principle applicable to wavelet shrinking in a context of the model (2.1) with incomplete permutation bootstrap estimates of the conditional sampling error used as our alternative to the normal errors of [5].

The methods described in [5] consider model (2.1) with Gaussian white noise and use a wavelet transform and the method of thresholding of the resulting wavelet coefficients to remove the noise. In models with strongly non-Gaussian noise the wavelet transform may not yield i.i.d. wavelet coefficients. For example in the case of Cauchy errors (see [6]) the transformed errors are neither independent nor identically distributed.

Therefore our method which requires only that the errors be exchangeable is of possible interest as a solution to noise removal in non-Gaussian cases.

Appendices

Appendix A

Matrix formulation of Basic Results

All the main results obtained in the previous sections can be re-stated in terms of random permutations of the elements of the matrix X and the projection matrix corresponding to the projection on the column space of X, rather than in terms of random permutations of the contrast vector $\boldsymbol{v}$. Therefore all the necessary calculations involving random permutations can be performed without prior knowledge of the contrast vector $\boldsymbol{v}$ which is advantageous in the case of multiple contrast vectors $\boldsymbol{v}_i$. In addition, the matrix form of the formulas allows us to find an upper bounds for the relative mean square bias and discrepancy.

Before stating these modified results we need to introduce notation for permutations applied to matrices.

Definition A.0.1 *Assume that* $\pi \in \Sigma_n$ *and* $\rho \in \Sigma_d$ *are two permutations of* $\{1,\ldots,n\}$ *and* $\{1,\ldots,d\}$, *respectively. Let* $A = (a_{i,j})_{i,j=1}^{n,d}$ *be an* $n \times d$ *matrix with real components.*

Then the row *and* column permutations of A *are defined by* $A_\pi = (a_{\pi(i),j})_{i,j=1}^{n,d}$ *and* $A^\rho = (a_{i,\rho(j)})_{i,j=1}^{n,d}$. *If* A *is a square matrix then the* π-permutation of A *will be*

defined as $\pi A = A_{\pi}^{\pi}$.

The definition of permutations of a vector introduced earlier is a slight extension of Definition A.0.1 for matrices with a single column. For a vector $\boldsymbol{v} \in I\!R^n$ it is natural to write $\pi \boldsymbol{v} = \boldsymbol{v}_{\pi}$.

In the proofs below we will use the following simple properties of matrix permutations. For any two matrices $A_{n\times d}$ and $B_{d\times n}$ and a permutation $\pi \in \Sigma_n$ it holds that

$$A^{\pi}B = AB_{\pi^{-1}} \text{ and} \tag{A.1}$$

$$A_{\pi}B = (AB)_{\pi}. \tag{A.2}$$

Assume that Λ is a subgroup of Σ_n and that λ is a random permutations distributed uniformly over Λ. For a matrix $A_{n\times d}$ the expectation EA_{λ} will be denoted by A_{Λ}. The symbols $A^{\Lambda} = EA^{\lambda}$ and $\Lambda A = E\lambda A$ will be used similarly whenever appropriate. Note that the meaning of the latter symbol differs from A_{Λ}^{Λ} which represents $E(EA_{\lambda})^{\lambda}$.

We will now turn our attention to restating our previous results. We will naturally use the same assumptions for the conditional version of model (1.1) as before, namely we will henceforward assume that Λ is a subgroup of Σ_n and that λ and π are two independent random permutations, distributed uniformly over Λ and Σ_n, respectively.

In addition to our previous notation we will denote by M the projection matrix of the projection to the column space of the matrix X from model (1.1), while the projection to $X^{\perp}$ will be represented by $M^{\perp} = I\!I - M$. Here $I\!I$ is used to denote the unit matrix of the appropriate dimension, which in this case is $n \times n$.

The results describing the basic properties of our proposed bias adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$, in particular the representation of its remaining conditional bias and error, are restated in the following two lemmas.

Corollary A.0.1 (of Lemma 1.3.1 and 1.3.2) *The remaining $\Lambda\pi$-conditional bias of the bias adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$, and the mean square of the conditional bias satisfy*

$$\begin{aligned} \mathrm{B}_{\text{adj}}(\Lambda, \pi) &= M_{\Lambda}^{\Lambda} \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}, \text{ and} \\ E(\mathrm{B}_{\text{adj}}(\Lambda, \pi))^2 &= \frac{1}{n-1} \left\| M_{\Lambda}^{\Lambda} \boldsymbol{v} \right\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 \\ &\leq \frac{1}{n-1} \left\| M_{\Lambda}^{\Lambda} \right\|^2 \|\boldsymbol{v}\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 . \end{aligned}$$

Proof: According to Lemma 1.3.1 the remaining conditional bias is $\mathrm{B}_{\text{adj}}(\Lambda, \pi) = \mathrm{E}(\lambda \left[(\mathrm{E}\lambda \boldsymbol{v})/_{\mathrm{X}} \right]) \cdot \pi\boldsymbol{\epsilon}$. Using the rules (A.1) and (A.2) we can write

$$E(\lambda \left[(E\lambda \boldsymbol{v})/_{X} \right]) = E(\lambda(ME\boldsymbol{v}_{\lambda})) = E(M^{\Lambda}\boldsymbol{v})_{\lambda} = M_{\Lambda}^{\Lambda}\boldsymbol{v}$$

to obtain $\mathrm{B}_{\text{adj}}(\Lambda, \pi) = \mathrm{M}_{\Lambda}^{\Lambda} \boldsymbol{v} \cdot \pi\boldsymbol{\epsilon}$.

The second part of the lemma concerning the mean square bias is a direct consequence of Lemma 1.3.2 with $M_{\Lambda}^{\Lambda}\boldsymbol{v}$ substituted for $E(\lambda \left[(E\lambda \boldsymbol{v})/_{X} \right])$. The inequality is a basic inequality which follows from the common definition of the norm of a matrix $\|M\| = \sup_{\|\boldsymbol{v}\|=1} \|M\boldsymbol{v}\|$.

□

Corollary A.0.2 (of Lemma 1.3.3) *For the mean squared error of the bias adjusted estimator $(\boldsymbol{v} \cdot \boldsymbol{y})_{\text{adj}}$ it holds that*

$$\begin{aligned} \mathrm{MSE}_{\text{adj}} &= \frac{1}{n-1} \left\| (I\!I - M^{\perp^{\Lambda}}) \boldsymbol{v} \right\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 \\ &\leq \frac{1}{n-1} \left\| (I\!I - M^{\perp^{\Lambda}}) \right\|^2 \|\boldsymbol{v}\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 . \end{aligned}$$

Proof: Let us substitute $(E\lambda \boldsymbol{v})^{\perp} = M^{\perp}E\boldsymbol{v}_{\lambda} = M^{\perp_{\Lambda}}\boldsymbol{v}$ in formula (1.18) of Lemma 1.3.3 to obtain

$$\mathrm{MSE}_{\mathrm{adj}} = \frac{1}{\mathrm{n}-1}\left\|\boldsymbol{v} - (\mathrm{E}\lambda\boldsymbol{v})^{\perp}\right\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 = \frac{1}{\mathrm{n}-1}\left\|\boldsymbol{v} - \mathrm{M}^{\perp_{\Lambda}}\boldsymbol{v}\right\|^2 \|\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}\|^2 .$$

Similarly as in the proof of Lemma A.0.1, the inequality is a basic result for the norm of matrices.

□

These two lemmas allow us to reformulate the result for the relative mean square bias (RMSB) of $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$ stated in Theorem 1.3.4. In addition, we find an upper bound for RMSB.

Corollary A.0.3 (of Theorem 1.3.4) *Assume that all assumptions of Theorem 1.3.4 hold. Then the relative mean squared bias of the bias-adjusted estimator* $(\boldsymbol{v}\cdot\boldsymbol{y})_{\mathrm{adj}}$ *satisfies*

$$RMSB = \frac{\left\|M_{\Lambda}^{\Lambda}\boldsymbol{v}\right\|^2}{\left\|(I\!I - M^{\perp_{\Lambda}})\boldsymbol{v}\right\|^2} \leq \frac{\left\|M_{\Lambda}^{\Lambda}\right\|^2}{\left\|(I\!I - M^{\perp_{\Lambda}})\boldsymbol{v}^{*}\right\|^2}, \tag{A.3}$$

where $\boldsymbol{v}^{*}$ *is the normalized vector* $\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.

Proof: Lemma A.0.1 and Lemma A.0.2 yield that

$$\mathrm{RMSB} = \frac{E(\mathrm{B}_{\mathrm{adj}}(\Lambda,\pi))^2}{\mathrm{MSE}_{\mathrm{adj}}} = \frac{\left\|M_{\Lambda}^{\Lambda}\boldsymbol{v}\right\|^2}{\left\|(I\!I - M^{\perp_{\Lambda}})\boldsymbol{v}\right\|^2} \leq \frac{\left\|M_{\Lambda}^{\Lambda}\right\|^2\|\boldsymbol{v}\|^2}{\left\|(I\!I - M^{\perp_{\Lambda}})\boldsymbol{v}\right\|^2}$$

which completes the proof.

□

The next theorem restates (in the projection matrix notation) and strengthens the result for RMSD of Theorem 1.4.1.

Corollary A.0.4 (of Theorem 1.4.1) *The relative mean squared discrepancy and its upper bound can be expressed as*

$$RMSD = \frac{E\left\|(M^{\lambda} - M^{\Lambda})\boldsymbol{v}\right\|^2}{\left\|(I\!I - M^{\perp^{\Lambda}})\boldsymbol{v}\right\|^2} \leq \frac{E\left\|M^{\lambda} - M^{\Lambda}\right\|^2}{\left\|(I\!I - M^{\perp^{\Lambda}})\boldsymbol{v}^*\right\|^2}, \tag{A.4}$$

where $\boldsymbol{v}^* = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.

Proof: As shown in the proof of Theorem 1.4.1 we can write

$$E(BS_{\Lambda} - True_{\Lambda})^2 = \frac{1}{n-1} E\left\|[\lambda\boldsymbol{v} - E\lambda\boldsymbol{v}]/_X\right\|^2 \|\epsilon - \bar{\epsilon}\|^2.$$

According to rules (A.1) and (A.1) we obtain

$$E\left\|[\lambda\boldsymbol{v} - E\lambda\boldsymbol{v}]/_X\right\|^2 = E\left\|M\boldsymbol{v}_{\lambda} - EM\boldsymbol{v}_{\lambda}\right\|^2 = E\left\|M^{\lambda}\boldsymbol{v} - M^{\Lambda}\boldsymbol{v}\right\|^2, \tag{A.5}$$

thus we can conclude, using also Lemma A.0.2, that

$$\text{RMSD} = \frac{E(BS_{\Lambda} - True_{\Lambda})^2}{\text{MSE}_{\text{adj}}} = \frac{E\left\|M^{\lambda}\boldsymbol{v} - M^{\Lambda}\boldsymbol{v}\right\|^2}{\left\|(I\!I - M^{\perp^{\Lambda}})\boldsymbol{v}\right\|^2}.$$

The inequality holds since $E\left\|M^{\lambda}\boldsymbol{v} - M^{\Lambda}\boldsymbol{v}\right\|^2 \leq E\left\|M^{\lambda} - M^{\Lambda}\right\|^2 \|\boldsymbol{v}\|^2$.

□

It is worthwhile to mention that the matrix $I\!I - M^{\perp^{\Lambda}}$ used in formulas (A.3) and (A.4) above can be rewritten as $I\!I - I\!I^{\Lambda} + M^{\Lambda}$. This should, for example, help us to investigate the ratio in (A.3).

Finally, the next theorem uses the formula of Theorem 1.4.3 concerning the relative performance improvement of the incomplete versus the complete permutation bootstrap and states an upper bound for the performance improvement. This upper bound is completely free of the contrast vector $\boldsymbol{v}$.

Corollary A.0.5 (of Corollary 1.4.3) *Let $\boldsymbol{v}$ be a contrast vector and BS_Λ, BS_{Σ_n}, $True_\Lambda$, and $True_{\Sigma_n}$ be as in Theorem 1.4.3. Then the ratio of the incomplete bootstrap mean square discrepancy versus the mean square discrepancy under the full permutation bootstrap model satisfies*

$$\frac{E(BS_\Lambda - True_\Lambda)^2}{E(BS_{\Sigma_n} - True_{\Sigma_n})^2} = \frac{n-1}{d-1} E \left\| (M^\lambda - M^\Lambda)\boldsymbol{v}^* \right\|^2 \leq \frac{n-1}{d-1} E \left\| M^\lambda - M^\Lambda \right\|^2, \tag{A.6}$$

where $\boldsymbol{v}^* = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$.

Proof: The equation is a direct consequence of Theorem 1.4.3 and (A.5). The inequality follows from the fact that $E \left\| (M^\lambda - M^\Lambda)\boldsymbol{v}^* \right\| \leq E \left\| M^\lambda - M^\Lambda \right\| \|\boldsymbol{v}^*\|$. □

Bibliography

Bibliography

[1] Anděl, J. (1985), "Matematická Statistika," *SNTL - Nakladatelství Technické Literatury, Praha, the Czech Republic.*

[2] Basu, D. (1980), "Randomization Analysis of Experimental Data: The Fisher Randomization Test," *Journal of the American Statistical Association*, 75, 575–582, (with discussions).

[3] Bickel, P. and Freedman, D. (1981), "Some Asymptotic Theory of the Bootstrap," *The Annals of Statistics*, Vol. 9, No. 6, 1196–1217.

[4] Daubechies, I. (1992), "Ten Lectures on Wavelets," *CBMS-NSF regional conference series in applied mathematics*, Printed by Capital City Press, Montpelier, Vermont.

[5] Donoho, D. and Johnstone, I. (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, 90, 1200–1224.

[6] Donoho, D. and Yu, T. P. Y. (1995), "Nonlinear Wavelet Transforms based on Median-Interpolation," *Technical Report, Department of Statistics, Stanford University.*

[7] Frazier, M. (1997), "An Introduction to Wavelets Through Linear Alegebra," *to Appear.*

[8] Freedman, D. A., and Lane, D. (1983), "A Nonstochastic Interpretation of Reported Significance Levels," *Journal of Bus. Econ. Statist.*, 1, 292–298.

[9] Gabriel, K. R. and Hall, W. J. (1983), "Rerandomization Inference on Regression and Shift Effects: Computationally Feasible Methods," *Journal of the American Statistical Association*, 78, 827–836.

[10] Gasela, G. (1983), "Conditional Inference from Confidence Sets," *Paper in honor of D. Basu's 65^{-th} birthday*, Cornell University 1–12.

[11] John, R. D. and Robinson, J. (1983), "Significance Levels and Confidence Intervals for Permutation Tests," *J. Statist. Comput. Simul.*, 16, 161–173.

[12] Kaiser, G. (1994), "A Friendly Guide to Wavelets," *Birkhäuser Boston.*

[13] Kolaczyk, E. (1996), "A Wavelet Shrinkage Approach to Tomographic Image Reconstruction," *Journal of the American Statistical Association*, 91, 1079–1089.

[14] Lehmann, E. L. (1959), "Testing Statistical Hypotheses," *John Willey & Sons, Inc., New York.*

[15] LePage, R. (1980), "Multidimensional Infinitely Divisible Variables and Processes, Part I: Stable Case," *Technical Report No. 292, Statistics Department, Stanford University.*

[16] LePage, R. (1987), "Conditional Moments for Coordinates of Stable Vectors," *Lecture Notes in Mathematics: Probability Theory on Vector Spaces IV*, 1391, 148–163.

[17] LePage, R. and Podgórski, K. (1996), "Resampling Permutations in Regression without Second Moments," *Journal of Multivariate Analysis*, 57, 119–141.

[18] LePage, R., Podgórski, K., Ryznar, M. (1997), "Strong and Conditional Invariance Principles for Samples Attracted to Stable Laws," *Probab. Theory Related Fields*, 108, no. 2, 281–298.

[19] Robinson, J., "Nonparametric Confidence Intervals in Regression: The Bootstrap and Randomization Methods," 243–255.

[20] Quade, D. (1973), "A Randomization Confidence Interval for a Shift," *Mimeographed lecture notes*, Dept. of Biostatistics, University of North Carolina, 1–7.

[21] Zvára, K. (1989), "Regresní analýza," *Academia, Praha, the Czech Republic.*