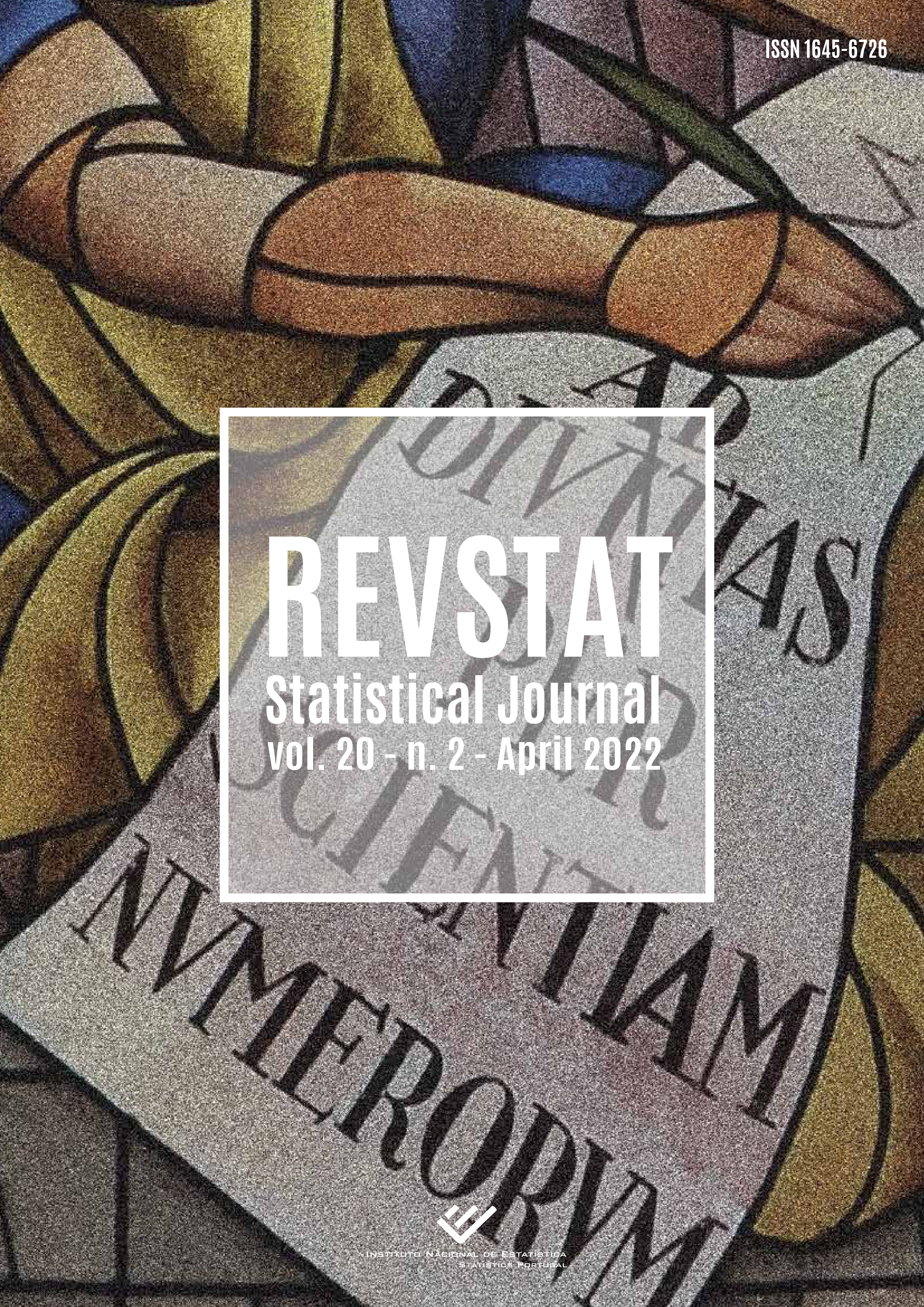


ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 20 - n. 2 - April 2022

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT-Statistical Journal, vol.20, n. 2 (April 2022)

vol.1, 2003- . - Lisbon : Statistics Portugal, 2003- .

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Publisher – *Statistics Portugal*

Layout-Graphic Design – *Carlos Perpétuo* | **Cover Design*** – *Helena Nogueira*

Edition - 140 copies | **Legal Deposit Registration** - 191915/03 | **Price** [VAT included] - € 9,00

**image*: stain glass window by Abel Manta (1888-1982)

Editorial

In memoriam - Ross Leadbetter [1931-2022]

It was very sad to hear the news on the death of Ross Leadbetter, Doctor Honoris Causa of ULisboa since 2013 and Professor Emeritus at the University of North Carolina at Chapel Hill, last February 26, 2022, at the age of 90. We are indeed missing Ross dearly. Ross Leadbetter was a true friend of his friends, among whom we are. It is our hope that present and future generations of statisticians will foster Ross' scientific legacy. Ross had a huge soul and was (is) one of the giants in the field of Extremes, having been interviewed about his career in the prestigious magazine *Extremes* (An interview with Ross Leadbetter | SpringerLink). The book he co-authored with Georg Lindgren and Holger Rootzén, entitled '*Extremes and Related Properties of Random Sequences and Processes*', is a landmark in the field, with over 5000 citations. Indeed, we cannot fail to mention that Ross Leadbetter was one of the guests at the 'NATO *Advanced Statistical Institute* (ASI), entitled *'Statistical Extremes and Applications*', which took place in Vimeiro, almost 40 years ago, in the summer of 1983, an event currently recognized as a milestone in the affirmation of this area and a milestone in the launch of the *'School of Extremes in Portugal*', which Ross Leadbetter, with his sense of humour, used to call the *'Portuguese Gang*', and which has more recently been referred to as PORTSEA, from the English '*Portuguese School of Extremes and Applications'*.

I also would like to mention that Ross Leadbetter honored ULisboa by accepting, at the request of DEIO, the distinction of *Doctor Honoris Causa* of ULisboa, because without a doubt, and just as can be read in Extremistas da minha terra | Faculdade de Ciências da Universidade de Lisboa (ulisboa.pt), '... when the university honors researchers of this importance it is also honoring itself.'

And Ross Leadbetter was also a very active participant in another of PORTSEA's major milestones, the 'EVT 2013—*Extremes in Vimeiro Today'*, organized by our colleagues and great friends, Antónia Amaral Turkman, Isabel Fraga Alves and Manuela Neves, to commemorate the 30 years of the Vimeiro meeting in 1983. Similar memories, written in Portuguese, can be found at the recent *Bulletin of the Portuguese Statistical Society* and at Info-Ciências Digital/14-03-2022.

*Participants in EVT 2013, who had also participated in the Vimeiro Meeting in 1983: *Top row:* Clive Anderson, Georg Lindgren, Jef Teugels, Ivette Gomes; *Bottom row*: Rolf Dieter-Reiss, Richard Davis, Anthony Davison, Ishay Weissman, Barry Arnold, Ross Leadbetter, Antónia Amaral Turkman, Kamil Feridun Turkman, Dinis Pestana, Juerg Huesler, Helena Iglésias, Isabel Barão, Manuela Neves

At last, Ross was deeply involved with *REVSTAT-Statistical Journal*, both as a Guest Editor of the *Special Issue on Statistics of Extremes and Related Fields* and as a reviewer of some articles submitted to this journal.

Ross is undoubtedly greatly missed, but his work will always be present and available for present and future generations!

March 31, 2022

Ivette Gomes (*Former Editor-in-Chief*)
Feridun Turkman (*Former Associate Editor*)

INDEX

Abstracted/indexed in: *Current Index to Statistics, Google Scholar, Mathematical Reviews, Science Citation Index Expanded®, Zentralblatt für Mathematic, Scimago Journal & Country Rank* and *SCOPUS.*

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 135–149

OPTIMAL REINSURANCE OF DEPENDENT RISKS

Authors: A. Bugalho de Moura
– ISEG – School of Economics and Management, Universidade de Lisboa; REM – Research in Economics and Mathematics; CEMAPRE, Lisbon, Portugal
amoura@iseg.ulisboa.pt

M.L. Centeno
– ISEG – School of Economics and Management, Universidade de Lisboa; REM – Research in Economics and Mathematics; CEMAPRE, Lisbon, Portugal
lcenteno@iseg.ulisboa.pt

Received: April 2019 Revised: December 2019 Accepted: January 2020

Abstract:

- We analyse the problem of finding the optimal combination of quota-share and stop loss treaties, maximizing the expected utility or the adjustment coefficient of the cedent, for each of two risks dependent through a copula structure. By risk we mean a line of business or a portfolio of policies. Results are obtained numerically, using the software *Mathematica*. Sensitivity of the optimal reinsurance strategy to several factors are investigated, including:
  i) the dependence level, by means of the Kendall's tau and the dependence parameter;
  ii) the type of dependence, using different copulas describing different tail behaviour;
  iii) the reinsurance calculation principles, where expected value, variance and standard deviation principles are considered.

  Results show that different dependence structures, yield significantly different optimal solutions. The optimal treaty is also very sensible to the reinsurance premium calculation principle. Namely, for variance related premiums the optimal solution is not the pure stop loss. In general, the maximum adjustment coefficient decreases when dependence increases.

Keywords:

- *reinsurance; dependent risks; copulas; premium calculation principles; expected utility; adjustment coefficient.*

AMS Subject Classification:

- 62P05.

1. INTRODUCTION

From the vast range of literature intended for the financial and insurance community, it is widely accepted that dependencies play a determinant role in risk assessment and management. Namely, reinsurance is a risk mitigating tool, constituting an important instrument in the management of risk of an insurance company where dependencies should be taken into account. When transferring risk, the cedent seeks a trade-off between profit and safety, which is dependent on the nature of the insured underlying risk and on the reinsurance premium calculation principle. This optimization problem has been largely studied in the literature, however only recently dependencies among risks have been considered. The goal is always to find the reinsurance strategy, which is usually defined by the forms of reinsurance to be considered and the specific retention levels, that minimizes a given measure of the underlying risk.

In [19] (for the aggregate claim model) and [13] (for the individual claim model) the authors obtain analytically the optimal reinsurance strategy maximizing the adjustment coefficient or the expected utility assuming independence. The premium calculation principle used is a convex functional, including the expected value, standard deviation and variance premium principles as special cases. In the case of "variance related" premium calculation principles, the optimal reinsurance contract is a specific, implicitly defined, non-linear function of the retained risk such that the tail of the underlying risk is shared by both the insurer and the reinsurer. If the expected value calculation principle is considered, the pure stop loss treaty is optimal. In fact, the pure stop loss, which appears as the optimal form of reinsurance in an innumerable amount of cases where the expected value premium principle is used, is not realistic in practice. It means all the risk in the tail is ceded to the reinsurer which will not accept it but at a very high premium loading, in which case the stop loss is probably not optimal anymore (as shown in [19, 13]). Other works considering convex premium principles include [21, 22] and [17], where convex risk measures (e.g. the variance or semi-variance of the retained risk) are used as optimality criteria. In all these works, independence is assumed. Indeed, while a large quantity of analytical studies can be found regarding optimal reinsurance, only a few number consider dependence. Notwithstanding, the interest in studying optimal reinsurance strategies under dependencies is increasing, driven by the need for real, robust and reliable quantitative risk models.

Article [12] is one of the first works including the effects of dependence when investigating analytical optimal forms or risk transfer. The optimal retention limit for the excess-loss (XL) reinsurance is studied considering two classes of insurance businesses, dependent through the number of claims by means of a bivariate Poisson, when the cedent intends to maximize the expected utility or the adjustment coefficient, using the expected value premium principle. Other authors have considered the optimal reinsurance problem under dependence between claim numbers, such as [28] and [5]. In [26] the impact of dependencies from year to year reinsurance payoffs are investigated using copulas and simulation, however optimal reinsurance is not directly addressed. In [6] positive dependencies in the individual risk are considered by means of the stochastic ordering. By considering a fixed reinsurance premium, calculated through the expected value principle, the authors demonstrate that in this case the optimal form of reinsurance is the XL treaty, when the optimality criterion is the maximization of the expectation of a convex function of the retained risk, including the expected utility for the exponential functional. In that paper, the authors refer to the non-proportional reinsurance as excess of loss (XL), assuming the risks are individual claims and then considering their sum. Accounting for dependence have protruded the use of numeral techniques, such as Dynamical Financial Analysis (DFA), Linear Programming, see e.g. [2], dynamic control problems, see e.g. [4], or simulation, see e.g. [26], which is often based on Monte Carlo simulation. Very recently, in [3], it has been advocated that when constraints on dependencies and economic and solvency factors are included in the optimal reinsurance problem,"the optimal contract can only be found numerically". Hence, they propose a numerical framework, based on the Second-Order Canonical Problem for numerical optimization. Other works regarding the application of numerical techniques to solve optimal reinsurance problems consider numerical methods for stochastic control theory (see for instance [29]). Most of these numerical works deal with real data.

In this work, we aim at studying the sensitiveness of the optimal reinsurance strategy, in presence of dependencies, to different factors such as premium calculation principles and dependence structures and levels. We account not only for the expected value principle, but also for the standard deviation and the variance principles. We consider two underlying risks and by risk we mean the aggregate claims of a line of business, a portfolio of policies or a policy. Dependence between the two risks is modelled through copulas, allowing to easily change the dependence structure and strength. We construct the optimal problem as finding the optimal combination of quota share (QS) and stop loss treaties, for each risk, that maximizes the expected utility or the adjustment coefficient of the total wealth of the first insurer. The analytical results in [6] for the expected value principle are not straightforwardly extendable to the variance related premiums, thus, we use numerical methods. To properly study the sensitivity of the optimal reinsurance strategy to several dependence structures and levels, and to a variety of reinsurance premium calculation principles, the problem setting is kept as simple as possible and no real data is used. The distributions of the underlying risks are assumed to be known and different distributions are considered. This controlled environment allows for a systematically analysis of the optimal reinsurance and its sensitivity to the several factors considered.

The layout of this paper is as follows. In Section 2 we set the optimization problem to be solved, introducing the copulas that will be used, the premium calculation principles and optimality criteria. In Section 3 we present the numerical results and their discussion. Finally, conclusions and future perspectives are drawn in Section 4.

2. SETTING THE OPTIMIZATION PROBLEM

We consider two risks, X_1 and X_2, with distribution functions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$, respectively. By risk we mean a line of business, a portfolio of policies or a policy. We assume that the two risks are dependent through a copula, denoted by C_α, $\alpha > 0$, such that the joint distribution is given by $C_\alpha(x_1, x_2) = C_\alpha(F_{X_1}(x_1), F_{X_2}(x_2))$, and the joint density function is given by $f_{X_1,X_2}(x_1, x_2)$. We use the notation $dC_\alpha(x_1, x_2) = f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2$. We assume that the insurer reinsures each risk by means of a QS contract topped by a stop loss treaty. Thus, the retained risks are $Y_i = Y_i(a_i, M_i) = \min(a_i X_i, M_i)$, $i = 1, 2$, where $0 \leqslant a_i \leqslant 1$, represents the QS retained level of risk i, $i = 1, 2$, and $M_i \geqslant 0$, denotes the stop loss retention limit, above which all the risk is ceded to the reinsurer, for risk i, $i = 1, 2$.

Therefore, the total wealth of the insurer after reinsurance is given by

$$\begin{aligned} W(a_1, M_1, a_2, M_2) &= W(a_1, M_1) + W(a_2, M_2) \\ &= (1-e_1)\, P_1 - P_{R1} - Y_1 + (1-e_2)\, P_2 - P_{R2} - Y_2\,, \end{aligned} \tag{2.1}$$

where $P_i > 0$, represents the premium received by the insurer for each risk i, $i = 1, 2$, and $e_i > 0$, $i = 1, 2$ are the corresponding insurer expenses; $P_{Ri} = P_{Ri}(a_i, M_i) > 0$ denotes the premium charged by the reinsurer for each risk i, $i = 1, 2$.

2.1. The dependence structure

When two risks are assumed not to be independent, an infinite range of possible dependencies between them can be at stake. The first question is, if they are dependent, what is the best model to explain the existing dependencies. Copulas constitute a convenient and elegant way of describing dependencies between two or more random variables. Also, using copulas, measures of non-linear dependence can be explored, such as the Kendall's rank correlation coefficient, which is a measure of concordance [14].

Our underlying risks[1], X_1 and X_2, are continuous random variables and the joint density function is given by $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, c\big(F_{X_1}(x_1), F_{X_2}(x_2)\big)$, where $c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2)$, $(u_1, u_2) \in [0,1]^2$ is the so-called copula density. In our case, the retained risk after the combination of QS and stop loss, $Y_i = \min(a_i X_i, M_i)$, $i = 1, 2$, $0 \leqslant a_i \leqslant 1$, $M_i \geqslant 0$, is non-decreasing function of X_i, hence the dependence structure is maintained for the retained risks (see [14]). That is, if the joint distribution of (X_1, X_2) is described by copula C, $F_{X_1,X_2}(x_1, x_2) = C\big(F_{X_1}(x_1), F_{X_2}(x_2)\big)$, then the joint distribution of (Y_1, Y_2) is also described by copula C, $F_{Y_1,Y_2}(y_1, y_2) = C\big(F_{Y_1}(y_1), F_{Y_2}(y_2)\big)$.

In this work, we consider Clayton's and Frank's copulas, which belong to the Archimedean family of copulas and have lower and no tail dependency, respectively. In these cases the Kendall's tau rank coefficient can be easily described by $\tau_\alpha = \frac{\alpha}{\alpha+2}$, for Clayton's copula, and $\tau_\alpha = 1 - 4\,\frac{1-D_1(\alpha)}{\alpha}$, with $D_1(\alpha) = \frac{1}{\alpha}\int_0^\alpha \frac{t}{e^t-1}\, dt$, for Frank's copula (see [14]). We also consider the Pareto's copula which can be derived as the "natural" bivariate distribution of two Pareto distributions with the same shape parameter α and it is as heavy right tail copula. Indeed, the Pareto's copula is the survival Clayton's copula with dependence parameter $1/\alpha$. Thus, the Pareto's copula Kendall's tau is $\tau_\alpha = \frac{1}{1+2\alpha}$.

2.2. The reinsurance premium

We analyse optimal reinsurance strategies for the expected value calculation principle, where the loading is proportional to the expected value of the risk, and also for the variance and standard deviation calculation principles. The later belong to the so-called (see [19]) variance related premium principles, as the premium loading is an increasing function of the

[1] In this work the random variables of interest are the two risks considered. Hence, often the underlying random variables are designated by risks.

variance of the covered risk. Noticing that the amount of risk ceded to the reinsurer, per risk i, $i = 1, 2$, is $X_i - Y_i$, with $Y_i = \min(a_i X_i, M_i)$, we can compute the reinsurance premium on each total ceded risk.

Expected value principle: $P_{Ri} = E(X_i - Y_i) + \delta_i E(X_i - Y_i) = (1+\delta_i) E(X_i - Y_i)$.

Variance principle: $P_{Ri} = E(X_i - Y_i) + \delta_i \operatorname{Var}(X_i - Y_i)$.

Standard deviation principle: $P_{Ri} = E(X_i - Y_i) + \delta_i \sqrt{\operatorname{Var}(X_i - Y_i)}$.

Here $\delta_i > 0$, $i = 1, 2$, is the loading coefficient. This is how the authors in [6], using the expected value principle, as well as in [19, 13], for variance related principles, compute the reinsurance premium. However, when a combination of QS and stop loss is taken into account, the QS and stop loss premiums can be considered separately. This is the procedure followed for instance in [8, 9], and it corresponds to many practical cases, where the stop-loss contract is independent of the QS treaty, coming on top of the QS. In fact, the QS premium is usually proportional to the ceded risk minus a commission. In this case, the QS premium is the proportion of the premium received by the insurer P_i correspondent to the ceded risk, $(1 - a_i)P_i$, subtracting the commission, $c_i > 0$: $P_{QSi} = (1 - a_i)(1 - c_i)P_i$. The stop loss premium will be computed on the ceded risk after QS: $Z_i = \max(a_i X_i - M_i, 0)$, $i = 1, 2$. Thereby, the total reinsurance premium turns out as follows.

Expected value principle: $P_{Ri} = P_{QSi} + (1 + \delta_i) E(Z_i)$.

Variance principle: $P_{Ri} = P_{QSi} + E(Z_i) + \delta_i \operatorname{Var}(Z_i)$.

Standard deviation principle: $P_{Ri} = P_{QSi} + E(Z_i) + \delta_i \sqrt{\operatorname{Var}(Z_i)}$.

Here, we will study and compare optimal reinsurance strategies in both cases where the premium is computed on the total ceded risk or separately for QS and stop loss.

2.3. The expected utility and the adjustment coefficient

Several authors have considered to use the expected utility of wealth as optimality criteria when ascertaining the optimal reinsurance strategy, e.g. [8, 9, 12, 19, 13, 23]. The adjustment coefficient can be regarded as a given coefficient of aversion of the exponential utility function. On the other hand, the adjustment coefficient is connected to the ultimate probability of ruin. From the well known Lundberg Inequality, the larger the adjustment coefficient is, the smaller the upper bound of the probability of ultimate ruin is. Thus, maximizing the adjustment coefficient R instead of minimizing the probability of ruin $\Phi(u)$ is reasonable. Because of this, many authors have considered maximizing the adjustment coefficient as optimality criteria for reinsurance, e.g. [7, 11, 12, 19, 13, 10, 28]. In [18] reinsurance strategies minimizing directly the probability of the insurer's ruin are studied. There, the authors consider that the reinsurance premium is an increasing function of the expected value of the transferred risk. They show that in this case the stop loss, or the truncated stop loss if there are reinsurance premium budget restrictions, is the optimal strategy.

In [27] the same problem, also considering the expected value premium principle, is analyzed in the presence of background risk. Other works can be found, where strategies minimizing directly the probability of ruin are obtained, such as in [20, 25, 1, 24]. However, in such works the framework is usually a dynamical setting, with a diffusion setup and a continuous time adaptation of the contract, which is not the case of the present paper.

Notice that the adjustment coefficient is independent from the initial capital, u, of the insurer. Thus, the optimal strategy that maximizes the adjustment coefficient is also independent of u. In [15] an upper bound for the probability of ruin, dependent on the initial capital, is provided. In [16] this inequality is further refined and used to approximate the probability of ruin in regime-switching Markovian models. This upper bound represents an improvement to the Lundberg bound, specially for the cases where the initial capital is small. Hence, it is expected that using such upper bound as optimality criteria will lead to different optimal retention levels, specially for small values of the initial capital. However, it requires the distribution of losses to be new worse than used (NWU) and represents a significantly more complex bound from the computational point of view, when compared to the Lundberg bound, as it includes the need to solve an extra minimization problem.

In this work we will consider maximizing the expected exponential utility and the adjustment coefficient. Interesting future works include the minimization of the improved upper bound for the probability of ruin provided in [15] as optimality criteria, and to compare it with the results here presented.

2.3.1. Maximizing the expected utility

The goal is to determine the optimal reinsurance contract for a risk-averse insurer which purpose is to maximize the expected utility of its wealth. We consider the exponential utility function, for risk averse investors, defined through $U(x) = \frac{1-e^{-\beta x}}{\beta}$, where $\beta = -U''(x)/U'(x) > 0$ is the coefficient of risk aversion. In this case, the expected utility of the wealth for a given (fixed) coefficient of aversion β is:

$$E\Big[U\Big(W\big(a_1, M_1, a_2, M_2\big)\Big)\Big] = \frac{1}{\beta}\Big(1 - E\Big[e^{-\beta W(a_1,M_1,a_2,M_2)}\Big]\Big). \tag{2.2}$$

Maximizing the expected utility (2.2) corresponds to find the reinsurance strategy, (a_1, M_1, a_2, M_2), that maximizes $E[U(W)]$ for a given (fixed) coefficient of risk aversion β. Recalling (2.1), this is equivalent to minimize the following functional:

$$\begin{aligned} E\Big[e^{-\beta W(a_1,M_1,a_2,M_2)}\Big] &:= G\big(\beta, a_1, M_1, a_2, M_2\big) = \\ &= e^{-\beta\big((1-e_1)P_1+(1-e_2)P_2\big)}\, e^{\beta\big(P_{R1}(a_1,M_1)+P_{R2}(a_2,M_2)\big)} \times \\ &\quad \times \int_0^{+\infty}\!\!\int_0^{+\infty} e^{\beta\big(Y_1(a_1,M_1)+Y_2(a_2,M_2)\big)}\, dC_\alpha(x_1,x_2) \end{aligned} \tag{2.3}$$

for a given (fixed) β.

2.3.2. Maximizing the adjustment coefficient

The adjustment coefficient, R, of the retained risk after reinsurance is defined as the unique positive root, if it exists, of $G(R, a_1, M_1, a_2, M_2) = 1$, where G is given by (2.3). The coefficient of adjustment is related to the coefficient of risk aversion of the exponential utility, as it corresponds to the value of the risk aversion coefficient for which the expected utility (2.2) is zero, see [19]. In [19] it is demonstrated that, under general regularity assumptions on the functional G verified in our case, a reinsurance policy maximizes the adjustment coefficient, $\widehat{R}$, if and only if:

- **i)** The expected utility, with coefficient of risk aversion $\widehat{R}$, is maximum for that policy, and
- **ii)** $G\big(\widehat{R}, a_1, M_1, a_2, M_2\big) = 1$.

Thus, as suggested in [19], the problem of maximizing the adjustment coefficient can be split in two sub problems:

1. For each $\beta > 0$, find the reinsurance strategy, (a_1, M_1, a_2, M_2) that minimizes G.
2. Solve $G\big(\beta, a_1, M_1, a_2, M_2\big) = 1$ with respect to the single variable β.

Whence, given the algorithm to find the optimal reinsurance maximizing the expected utility it is straightforward to obtain the reinsurance strategy maximizing the adjustment coefficient. However, maximizing the adjustment coefficient requires the solution of several expected utility maximization problems, until the desired root is found.

3. NUMERICAL RESULTS AND DISCUSSION

The numerical implementation was performed the *Mathematica*. All the double and single integrals involved in the evaluation of $G(\beta, a_1, M_1, a_2, M_2)$ are solved using *Mathematica* numerical integration, which applies global adaptive Gauss–Kronrod quadrature rules. The resolution of the minimization problems were carried out using numerical algorithms for non-linear constrained global optimization already implemented in *Mathematica*, namely the Nelder Mead and Differential Evolution algorithms. Strictly speaking, the Nelder Mead algorithm is not a global optimization method, but it tends to work quite well if the objective function does not have many local minima, which is the case here. The numerical procedure, namely the numerical optimization problem, is amenable for improvement as no particular features of the functional to minimize were taken into consideration and general global optimization was applied. The existence of *plateaux* regions in the functional to minimize, specially regarding the stop loss retention values, made the convergence to the optimal solution slower in some cases. Nevertheless, results were achieved and analysis of the sensitiveness to the several factors, such as premium calculation principles and dependence structures and levels, of the optimal reinsurance for two dependent risks were performed.

In the following, the premium received by the insurer is computed by means of the expected value principle with a loading coefficient of $\gamma_i = 0.2,\ i = 1, 2$. For the underlying risks, X_1 and X_2, we will consider different distributions, but in such way that the expected value is always 1. Hence, the premium loading charged by the insurer is $\gamma_i\, E(X_i) = \gamma_i$, $i = 1, 2$. We assume expenses are 5% of the premium, $e_i = 0.05$, $i = 1, 2$. Whenever the QS premium is computed on a proportional basis, separately from the stop-loss premium, the commission is $c_i = 0.03$, $i = 1, 2$. Indeed, the QS reinsurance commission should be lower than the insurer expenses $c_i < e_i$, meaning it is impossible to reinsure the whole risk through QS with a certain profit. This implies that the QS premium loading is $E(X_i)\left[(1 - c_i)(1 + \gamma_i) - 1\right] = 0.164\, E(X_i)$. When maximizing the expected utility, we consider a coefficient of risk aversion $\beta = 0.1$.

In Table 1 are presented the premium loadings. With these values, the premium loading when all the risk is transferred by means of a pure stop loss contract, *i.e* when $a_i = 1$ and $M_i = 0$, is the same for all three premium principles. Indeed, in this case the moments involved in computing the reinsurance premiums, either for QS and stop loss together or separately, correspond to the moments of the underlying risk. However, if QS and stop loss are considered separately that is true only when $a_i = 1$ (and $M_i = 0$), whereas if the premium is computed for the QS and stop loss together that is true no matter the value of a_i (as long as $M_i = 0$).

Table 1: Loading coefficients for the three premium principles considered, where δ is the loading coefficient for the expected value principle and X is the underlying risk.

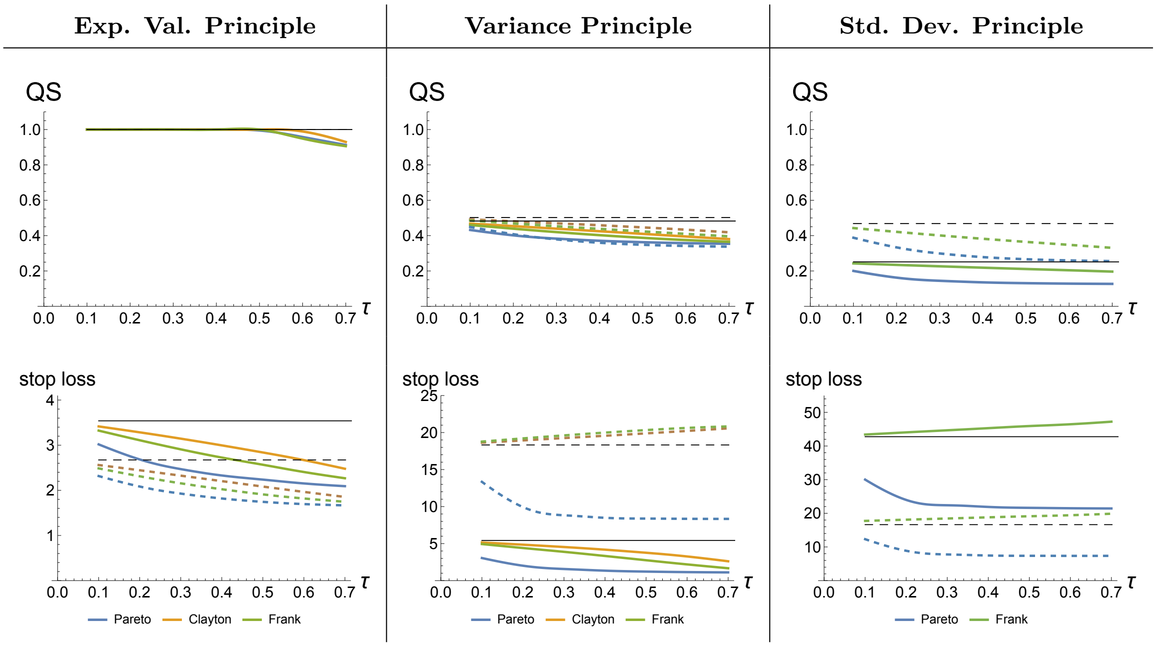
premium principle	loading coefficient
expected value	δ
variance	$\delta\, E(X)/\,\mathrm{Var}(X)$
standard deviation	$\delta\, E(X)/\sqrt{\mathrm{Var}(X)}$

We first consider two Pareto distributions with expected value 1 and shape parameter 3. The loading coefficients in this case are shown in Table 2. In this case, independently of the premium calculation principle, the optimal retention levels of QS and stop loss contracts are the same for both risks, as they are equal.

Table 2: Loading coefficients, for the three premium principles, considering two Pareto risks with expected value 1 and shape parameter 3.

premium principle	QS and stop loss separately	QS and stop loss together
expected value	0.3	0.2
variance	0.1	0.0666667
standard deviation	0.173205	0.11547

Results for the optimal reinsurance, as function of Kendall's tau coefficient, maximizing the expected utility with coefficient of risk aversion $\beta = 0.1$ and the loading coefficients in Table 2 are presented in Figure 1.



horizontal lines: independent case;
dashed lines: premiums computed on the total ceded risk;
solid lines: premiums computed on the ceded risk after QS, with the QS premium computed on original terms;

Figure 1: Optimal reinsurance maximizing the expected utility with $\beta = 0.1$.

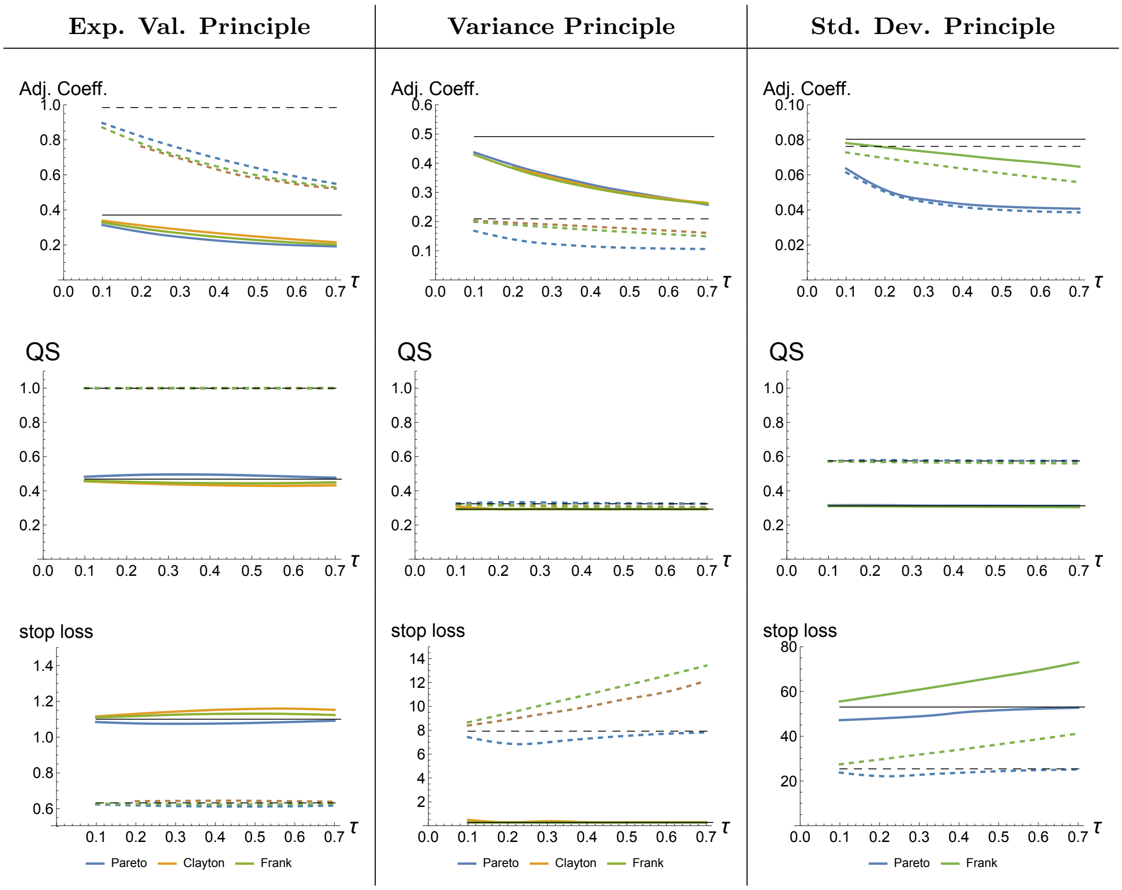
From the results, we can see that when the expected value principle is computed on the total ceded risk, the optimal reinsurance contract is always the pure stop loss, independently of the dependence structure and strength. This is expected, from the results in [6]. If the expected value principle is computed only on the ceded risk through stop-loss, after QS, the pure stop loss is no longer the optimal contract. In this case, for larger values of the Kendall's tau correlation, the optimal QS levels decrease below the independence optimal QS level. For larger values of the Kendall's tau, it compensates to cede part of the risk trough QS and to cede trough stop loss on top of that, independently of the dependence structure. This is related with the QS premium loading in this case, that for strong dependence compensates the stop-loss premium loading. This is not verified when independence is assumed, for this loading coefficients. Thus, the results suggest that affects the type of optimal contract even when the expected value principle is considered, if QS and stop-loss premiums are computed separately. We also observe that, no matter what the optimal contract is, the optimal stop loss limits for the expected value principle, computed together or separately for QS and stop loss, decrease as dependence strength increases.

Regarding the variance and standard deviation principles, we can see that the pure stop loss is never the optimal treaty, not even in the independent case. That is in accordance with the results in [19]. In these cases, the optimal QS levels always decrease as the dependence strength increases, independently of computing the reinsurance premium together, for the whole ceded risk, or separately for QS and stop-loss. However, the optimal stop-loss limit does not always decrease as dependence strength increases. For the variance principle, if the premium is computed together for QS and stop-loss, the optimal limit of stop loss decreases, as dependence increases, in the case of Pareto's copula, but it increases, as dependence increases (and the QS optimal level decreases) for Clayton's and Frank's copulas. For the standard deviation principle, this behaviour of Clayton's and Frank's copulas is observed not only for the premium computed on the whole ceded risk, but also when the premiums of QS and stop-loss are computed separately. These results show that dependence impacts the optimal levels of retention in non-intuitive ways, especially in the cases of variance related premium principles.

Next we consider the maximization of the adjustment coefficient as optimality criteria. As described in Chapter 2, to obtain the optimal reinsurance treaty maximizing the adjustment coefficient, it is enough to find the optimal solution for the expected utility problem with the coefficient of risk aversion $\beta > 0$ such that the expected utility value in (2.2) is equal to 0. In order to solve equation $G(R, a_1, M_1, a_2, M_2) = 1$, for (a_1, M_1, a_2, M_2) minimizing $G(R, a_1, M_1, a_2, M_2)$, a bisection method was applied. Amongst the root finding numerical methods, bisection is the simplest. Although its convergence is not very fast when compared with Newton-type methods, it has the advantage of not requiring the computation of derivatives of the functional. Also, convergence to a tolerance of 10^{-6} was reached within an average of 10 iterations, as the initial points were easily chosen close enough to the solution. Situations where convergence was more difficult regard instances where converge of the constraint global optimization algorithm to the minimum of functional G was slow. This was the case of Clayton's copula, when using the standard deviation principle computing QS and stop loss premiums separately. Results are depicted in Figure 2.

Again, as expected from the results in [6] and [19], the optimal treaty when the expected value principle is applied to QS and stop loss together is the pure stop loss, for all three copulas and all values of the dependence parameter. That is not the case for the variance related premium principles, even in the independent case. The optimal retention levels vary with the dependence parameter, although not so significantly as for the case where the risk aversion coefficient was fixed. Instead, the impact of dependence is very relevant in the value of the maximum adjustment coefficient of the optimal contract. It can be observed, for all copulas and premium principles considered, that the adjustment coefficient decreases as dependence strength increases. This means that the higher the dependence, the higher the upper bound of the ultimate probability of ruin.

The remarks made for the case of a fixed coefficient of aversion apply here, although now the differences in the standard deviation and variance principles, computing QS and stop loss premiums together, are more accentuated. The adjustment coefficient always decreases when dependence increases. It can be observed that the maximum adjustment coefficient using the standard deviation principle is always below those using the expected value and variance principles. If QS and stop loss premiums are computed together, then the maximum adjustment coefficient using the expected value principle is higher. If premiums are computed separately, then the maximum adjustment coefficient using the variance principle is higher.



horizontal lines: independent case;
dashed lines: premiums computed on the total ceded risk;
solid lines: premiums computed on the ceded risk after QS, with the QS premium computed on original terms;

Figure 2: Optimal reinsurance maximizing the adjustment coefficient, and corresponding optimal values of the adjustment coefficient.

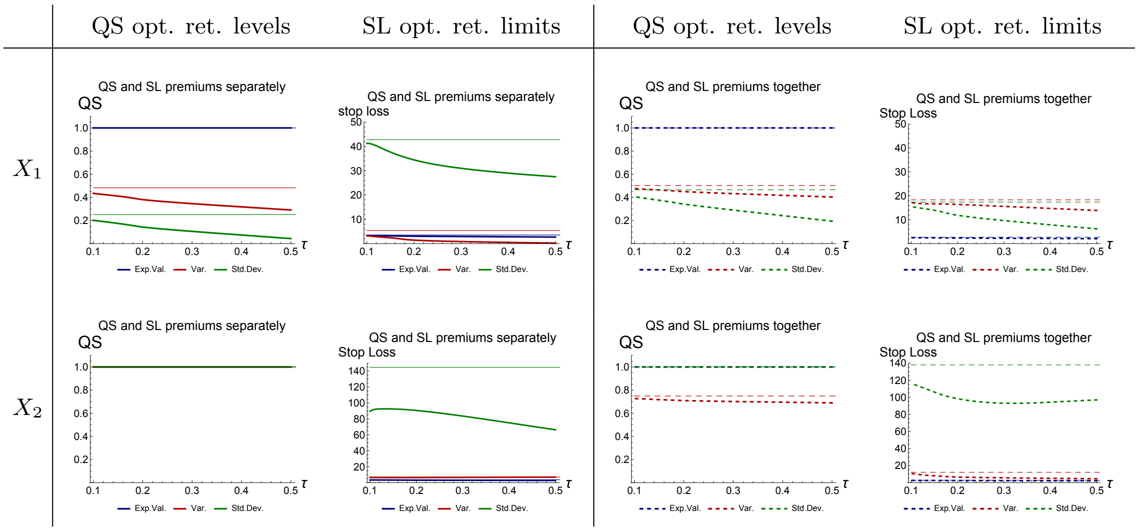
This is verified for all three copulas. When the standard deviation principle is considered, the maximum adjustment coefficient of the optimal contract is similar computing the premium together on the whole ceded risk, or separately for QS and stop-loss. It is worth noticing the differences in the optimal reinsurance for the different copulas. Differences are particularly significant between the Pareto's copula and Clayton's and Frank's copulas. This is because Pareto's copula has right tail dependence, while Clayton's and Frank's copulas do not.

Afterwards we have considered two risks with different tail heaviness: two Pareto distributions with expected value 1 and shape parameters 3 and 12, respectively. In this case, the variances are 3 and 1.2, respectively. For this case, we have considered dependence by means of the Pareto's copula, where dependence is stronger on the right tail, and we aim at maximizing the expected utility with coefficient of risk aversion $\beta = 0.1$. Regarding the loading coefficients, we apply the same reasoning as before, which is described in Table 1 and leads to the loading coefficients presented in Table 3. Results are shown in Figure 3.

Table 3: Loading coefficients, for the three premium principles, considering two Pareto risks with expected value 1 and shape parameter 3 (X_1) and 12 (X_2).

premium principle	QS and stop loss separately		QS and stop loss together	
	X_1	X_2	X_1	X_2
expected value	0.3	0.3	0.2	0.2
variance	0.1	0.25	0.0666667	0.166667
standard deviation	0.173205	0.273861	0.11547	0.182574

Whenever the expected value principle is applied (blue lines of Figure 3), either on the total ceded risk or just on the stop loss contract, the pure stop loss treaty is optimal for both risks. The optimal stop loss retention limits are similar for both risks and decrease as dependence increases.



horizontal lines: independent case;
dashed lines: premiums computed on the total ceded risk;
solid lines: premiums computed on the ceded risk after QS, with the QS premium computed on original terms;

Figure 3: Optimal reinsurance maximizing the expected utility with $\beta = 0.1$ for two dependent risks with Pareto distributions with mean 1 and shape parameters 3 (X_1) and 12 (X_2).

Regarding the standard deviation principle (green lines in Figure 3), the pure stop loss contract is optimal for the second (lighter tailed) risk, when computing the reinsurance premium both on the total ceded risk or separately for QS and stop loss. The optimal stop loss retention values for this pure stop loss contract on the second risk are significantly high, compared to the first risk or with the other, expected value and variance, premium calculation principles, and decrease as dependence strength increases. For the first (heavier tailed) risk, the optimal reinsurance contract is not the pure stop loss anymore and the optimal stop loss retention limits are much lower than those of the second risk, though still higher compared to the expected value premium principle. The optimal QS levels are quite low and both QS and stop loss optimal retention limits decrease as dependence increases. With the standard deviation premium principle, much of the first risk is transferred, while much of the second risk is kept.

For what concerns the variance principle (red lines in Figure 3), the pure stop loss contract is optimal for the second risk only when the QS premium is computed on a proportional basis. Again, the optimal QS and stop loss retention values decrease with dependence. The optimal stop loss retention limits of the first risk are significantly different when QS and stop loss premiums are computed together or separately. This difference is less accentuated for the second risk, where the stop loss contract is optimal when computing QS and stop loss premiums separately.

In general, for all three premium principles and for both risks, the optimal QS and stop loss retained levels decrease as dependence increases. In most cases the pure stop loss contract is optimal for the second (lighter tailed) risk. Thus, in most cases only the tail of the second risk is transferred. On the contrary, for the first (heavier tailed) risk, the pure stop loss contract is optimal only for the expected value principle, meaning that for the standard deviation and variance principles it is optimal to transfer more of the first (heavier tailed) risk.

4. CONCLUSIONS

Clearly dependencies alter the optimal treaty, as compared with the independent case, and the impact of these dependences on the optimal treaty may be non-intuitive. Different dependence structures, yield significantly different optimal solutions. As expected, the optimal treaty is also highly sensitive to the premium calculation principle and relevant differences are encountered between premiums calculated on the total ceded risk or separately for QS and stop loss. In some cases, this behaviour is accentuated in the presence of dependencies. The results here presented can be useful in bringing insight on the impact of dependence on the optimal reinsurance strategy. Such insight can be helpful in the design of more general theoretical results on optimal reinsurance of dependent risks. It can also be beneficial when analysing *real world* case studies of applications.

ACKNOWLEDGMENTS

The authors were partially supported by the Project CEMAPRE–UID/MULTI/00491/2019 financed by FCT/MCTES through national funds.

REFERENCES

[1] Albrecher, H. and Haas, S. (2011). Ruin theory with excess of loss reinsurance and reinstatements, *Applied Mathematics and Computation*, **217**(8031):8043.

[2] Asimit, A.; Bignozzi, V.; Cheung, K.C.; Hu, J. and Kim, E.S. (2017). Robust and Pareto optimality of insurance contracts. *European Journal of Operational Research.*

[3] Asimit, A.V.; Gao, T.; Hu, J. and Kim, E.S. (2017). *Optimal risk transfer: a numerical optimisation approach*, Technical report, City University of London.

[4] Bai, L.; Cai, J. and Zhou, M. (2013). Optimal reinsurance policies for an insurer with a bivariate reserve risk process in a dynamic setting, *Insurance: Mathematics and Economics.*

[5] Bi, J.; Liand, Z. and Xu, F. (2016). Optimal mean variance investment and reinsurance problems for the risk model with common shock dependence, *Insurance: Mathematics and Economics.*

[6] Cai, J. and Wei, W. (2012). Optimal reinsurance with positively dependent risks, *Insurance: Mathematics and Economics*, **50**, 57–63.

[7] Centeno, M.L. (1986). Measuring the effects of insurance by the adjustment coefficient, *Insurance: Mathematics and Economics*, **5**, 169–182.

[8] Centeno, M.L. (1986). Some mathematical aspects of combining proportional and non-proportional reinsurance, *Insurance and Risk Theory*, 247–266.

[9] Centeno, M.L. (1988). The expected utility applied to reinsurance, *Risk, Decision and Rationality*, 679–689.

[10] Centeno, M.L. (1997). Excess of loss reinsurance and the probability of ruin in finite horizon, *ASTIN Bulletin*, **27**(1), 59–70.

[11] Centeno, M.L. (2002). Measuring the effects of reinsurance by the adjustment coefficient in the Sparre Anderson model, *Insurance: Mathematics and Economics*, **30**, 37–49.

[12] Centeno, M.L. (2005). Dependent risks and excess of loss reinsurance, *Insurance: Mathematics and Economics*, **37**, 229–238.

[13] Centeno, M.L. and Guerra, M. (2010). The optimal reinsurance strategy – the individual case, *Insurance: Mathematics and Economics*, **46**, 450–460.

[14] Embrechts, P.; Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In "Handbook of Heavy Tailed Distributions in Finance" (S. Rachev, Ed.), pp. 329–384, Elsevier, Amsterdam.

[15] Gajek, L. (2005). On the deficit distribution when ruin occurs – discrete time model, *Insurance: Mathematics and Economics*, **36**, 13–24.

[16] Gajek, L. and Rudz, M. (2018). Banach contraction principle and ruin probabilities in the regime-switching models, *Insurance: Mathematics and Economics*, **80**, 45–53.

[17] Gajek, L. and Zagrodny, D. (2000). Insurer's optimal reinsurance strategies, *Insurance: Mathematics and Economics*, **27**, 105–112.

[18] Gajek, L. and Zagrodny, D. (2004). Reinsurance arrangements maximizing insurer's survival probability, *The Journal of Risk and Insurance*, **71**(3), 421–435.

[19] Guerra, M. and Centeno, M.L. (2008). Optimal reinsurance policy: the adjustment coefficient and the expected utility criteria, *Insurance: Mathematics and Economics*, **42**, 529–539.

[20] Hipp, C. and Vogt, M. (2003). Optimal dynamic XL reinsurance, *Astin Bulletin*, **33**(2), 193–207.

[21] Kaluszka, M. (2004). An extension of Arrow's result on optimality of a stop-loss contract, *Insurance: Mathematics and Economics*, **35**, 527–536.

[22] Kaluszka, M. (2005). Optimal reinsurance under convex principles of premium calculations, *Insurance: Mathematics and Economics*, **36**, 375–398.

[23] Karageyik, B.B. and Dickson, D.C.M. (2015). Optimal reinsurance under multiple attribute decision making, *Annals of Actuarial Science*, **10**(1), 65–86.

[24] Liang, X. and Young, V.R. (2018). Minimizing the probability of ruin: optimal per-loss reinsurance, *Insurance: Mathematics and Economics*, **82**, 181–190.

[25] Schmidli, H. (2010). Optimal proportional reinsurance policies in a dynamic setting, *Scandinavian Actuarial Journal*, **2001**(1), 55–68.

[26] Wang, P. (2013). Risk modeling and multi-year, multi-line reinsurance using copulas, *Journal of Insurance Issues*, **36**(1), 58–81.

[27] Weng, C. and Zhuang, S.C. (2017). CDF formulation for solving an optimal reinsurance problem, *Scandinavian Actuarial Journal*, **2017**(5), 395–418.

[28] Zhang, L.; Hu, X. and Duan, B. (2015). Optimal reinsurance under adjustment coefficient measure in a discrete risk model based on Poisson MA(1) process, *Scandinavian Actuarial Journal*, **2015**(5), 455–467.

[29] Zhang, N.; Jin, Z.; Li, S. and Chen, P. (2016). Optimal reinsurance under dynamic VaR constraint, *Insurance: Mathematics and Economics*.

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 151–177

A COMPUTATIONAL APPROACH TEST FOR THE EQUALITY OF TWO MULTIVARIATE NORMAL MEAN VECTORS UNDER HETEROGENEITY OF COVARIANCE MATRICES

Authors: Esra Gökpınar
– Department of Statistics, Gazi University, Turkey eyigit@gazi.edu.tr

Sezen Karanfil
– Department of Statistics, Gazi University, Turkey sznkaranfil06@gmail.com

Meral Ebegil
– Department of Statistics, Gazi University, Turkey mdemirel@gazi.edu.tr

Yaprak Arzu Ozdemir
– Department of Statistics, Gazi University, Turkey yaprak@gazi.edu.tr

Fikri Gökpınar
– Department of Statistics, Gazi University, Turkey fikri@gazi.edu.tr

Received: November 2018 Revised: December 2018 Accepted: January 2020

Abstract:

- In this paper, a computational approach test (CAT) was proposed to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrices. The proposed test was compared with the other popular tests as well as their CAT versions in terms of estimated type I error rate and power. Simulation study shows that the proposed test and CAT versions of tests can be used as a good alternative test to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrices.

Keywords:

- *computational approach test; parametric bootstrap approach; simulation study.*

AMS Subject Classification:

- 62F03, 65C05, 65C60.

1. INTRODUCTION

In statistical analysis, the comparison of the means of two or more groups is a very common problem. However, in many real world problems, there can be more than one variable which are related with each other. In this case, it is not appropriate to use univariate statistical methods. Therefore, it is required to utilize some multivariate statistical methods. One of the most common methods is the Hotelling T^2 statistic to compare the mean vectors of two independent groups under multivariate normality. It is known that Hotelling T^2 statistic requires the assumption of equality of two covariance matrices. However, this assumption is not valid in many statistical application areas. The violation of this assumption is called multivariate Behrens–Fisher problem in statistical analysis. In the case of multivariate Behrens–Fisher problem, the type I error rates of Hotelling T^2 statistic are not close to the nominal level and this also affects the power of the test negatively. Ito and Schull [12] indicated that when variances of two groups are not equal, the type I error rate for Hotelling's T^2 is approximately equal to the nominal level rate only when the sample sizes of groups are large and equal. Therefore, many solutions can be found for this problem in the literature.

Bennett [2] is one of the pioneers who presented the exact solution to the multivariate Behrens–Fisher problem. Since the Bennett's test depends on the order of the observations, it is not useful for larger sample sizes. In addition, James [13] improved the simple chi-square approximation by the Cornish–Fisher expansion until the third order term. Yao [27] suggested the approximate degrees-of-freedom solution and indicated that type I error rate of this test is lower than that of James' test in almost all cases. Subrahmaniam and Subrahmaniam [21, 22] compared the tests of Bennett, James, and Yao according to type I error rates and powers. Johansen [14] also studied the Behrens–Fisher problem in the context of general linear models. Christensen and Rencher [6] compared the seven tests given by Bennett [2], James [13], Yao [27], Johansen [14], Nel and Van der Merwe [18], Hwang and Paulson [11] and Kim [15] for the multivariate Behrens–Fisher problem in terms of type I error rates and powers. Algina, Oshima and Tang [1] also compared the tests given by Yao [27] , James [13] and Johansen [14] under various conditions of heteroscedasticity and non-normality. They showed that the type I error rate of Johansen's test is roughly equivalent to that of Yao's solution. In addition, the type I error rates of Johansen's test improve as number of variables increases. Kim [15] showed that the type I error rate of the test is more conservative than that of Yao's test in almost every situation. However, Kim's test has higher power than Yao's test when the smaller sample size is associated with the large variance [15].

De la Rey and Nel's [7] compared the tests given by Bennett [2], James [13], Yao [27] and Nel and Van der Merwe [18] and showed that Nel and Van der Merwe [18] and Yao [27] gave better solutions. According to the results of the comparative papers mentioned above, apparently there is no definitive solution that shows good performance in all circumstances. Finally, Krishnamoorthy and Yu [17] modified Nel and Van Der Merwe's [18] procedure by providing an invariant statistic. Recently, bootstrap-based methods for multivariate hypothesis testing were proposed. For example, Smaga [20] developed bootstrap methods of some test statistics based on different weight matrices for testing the mean vector of a multivariate distribution. Konietschke *et al.* [16] developed parametric and nonparametric bootstrap methods of Wald-type test for multi-factor multivariate data which also includes multivariate Behren Fisher problem. They compared these tests via simulation study under both normality and non-normality models.

The purpose of this paper is to test the equality of two normal mean vectors under heterogeneity of covariance matrices by using computational approach test (CAT). This method which was firstly introduced by Pal *et al.* [19] is used in situations where traditional approaches do not provide useful solutions. The CAT is a special case of parametric bootstrap and based on restricted maximum likelihood estimation under null hypothesis. One of the most important advantages of this procedure is that it does not require the knowledge of any sampling distribution. Pal *et al.* [19] showed the application of the CAT to Gamma and Weibull distributions for hypothesis testing and interval estimations. CAT was also applied by Chang and Pal [3] for testing the equality of two normal population means under heteroscedastisity. Chang *et al.* [4, 5] suggested test procedures based on CAT for hypotheses testing of the Poisson and Gamma models. Gökpınar and Gökpınar [8] applied CAT to test the equality of several normal population means when the variances are unknown and arbitrary and Gökpınar *et al.* [9] proposed CAT for the equality of several inverse Gaussian means under heterogeneity of scale parameters. Moreover, Gökpınar and Gökpınar [10] proposed CAT for the equality of coefficient of variations in k populations. In these studies, it was shown that the CAT procedure is a good alternative for other testing procedures for various statistical problems.

For this reason, in this study, the CAT method to the equality of two normal mean vectors under heterogeneity of covariance matrices was applied and this method was also compared with Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] tests in terms of their type I error rates and powers under various situations.

The rest of this study was organized as follows. In Section 2, the method was described to obtain the maximum likelihood estimates (MLEs) over unrestricted parameter space and over a restricted parameter space. Simple fixed point iteration was proposed to compute the MLEs under restricted parameters space. In Section 3, the tests given by Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] and Konietschke *et al.* [16] were presented. In Section 4, the concept of CAT procedure and its application to the equality of two normal mean vectors under heterogeneity of covariance matrices were given. In Section 5, simulation studies were presented to assess the performance of the proposed test in terms of the type I error rates and powers under multivariate normal distribution with different parameter combinations. Furthermore, to see robustness of all tests under nonnormal distribution, the estimated type I error rates and powers of all tests were calculated. Finally, concluding remarks were summarized in Section 6.

2. THE MAXIMUM LIKELIHOOD ESTIMATES

Let $\boldsymbol{Y_{i1}}, \boldsymbol{Y_{i2}}, ..., \boldsymbol{Y_{in_i}}$, have p-variate normal distribution with mean vector $\boldsymbol{\mu_i} = (\mu_{i1}, \mu_{i2}, ..., \mu_{ip})^\top$ and covariance matrix $\boldsymbol{\Sigma_i}$, $i = 1, 2$. Assume that sample units are independent from each other. n_i $(i = 1, 2)$ is the sample size of i-th group.

$\boldsymbol{\bar{Y}_i}$ is the sample mean vector of i-th group and $\boldsymbol{S_i}$ is the maximum likelihood estimation of covariance matrix of i-th group. $\boldsymbol{S_{(i)}}$ is the unbiased estimation of covariance matrix of

i-th group. Thus,

$$\bar{\mathbf{Y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{Y}_{ij}\,, \qquad i=1,2\,, \tag{2.1}$$

$$\mathbf{S}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_i)\,(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_i)^\top\,, \qquad i=1,2\,. \tag{2.2}$$

The log-likelihood function under the unrestricted parameter space is given by

(2.3)

$$\begin{aligned} lnL = {} & \frac{-p(n_1+n_2)}{2}\,ln(2\pi) - \frac{n_1}{2}\,ln(|\boldsymbol{\Sigma_1}|) - \frac{n_2}{2}\,ln(|\boldsymbol{\Sigma_2}|) - \frac{1}{2}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\boldsymbol{\mu_1}]^\top\,\boldsymbol{\Sigma}_1^{-1}[\mathbf{Y_{1j}}-\boldsymbol{\mu_1}] \\ & - \frac{1}{2}\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\boldsymbol{\mu_2}]^\top\,\boldsymbol{\Sigma}_2^{-1}[\mathbf{Y_{2j}}-\boldsymbol{\mu_2}]\,. \end{aligned}$$

To find the unrestricted MLEs, the partial derivatives of Eq. (2.3) with respect to $\boldsymbol{\Sigma_1}$, $\boldsymbol{\Sigma_2}$, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ yield the following equations:

$$\hat{\boldsymbol{\mu}}_{\mathbf{1}} = \bar{\mathbf{Y}}_1\,, \qquad \hat{\boldsymbol{\mu}}_{\mathbf{2}} = \bar{\mathbf{Y}}_2\,,$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{1}} = \frac{\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\bar{\mathbf{Y}}_{\mathbf{1}}]^\top\,[\mathbf{Y_{1j}}-\bar{\mathbf{Y}}_{\mathbf{1}}]}{n_1} = \mathbf{S_1}\,, \qquad \hat{\boldsymbol{\Sigma}}_{\mathbf{2}} = \frac{\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\bar{\mathbf{Y}}_{\mathbf{2}}]^\top\,[\mathbf{Y_{2j}}-\bar{\mathbf{Y}}_{\mathbf{2}}]}{n_2} = \mathbf{S_2}\,.$$

To find restricted MLE (RMLE), under $H_0\colon \boldsymbol{\mu_1}=\boldsymbol{\mu_2}=\boldsymbol{\mu}$, the log-likelihood function can be expressed as

(2.4)

$$\begin{aligned} lnL = {} & \frac{-p(n_1+n_2)}{2}\,ln(2\pi) - \frac{n_1}{2}\,ln(|\boldsymbol{\Sigma_1}|) - \frac{n_2}{2}\,ln(|\boldsymbol{\Sigma_2}|) - \frac{1}{2}\,tr\left(\boldsymbol{\Sigma}_1^{-1}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\boldsymbol{\mu}]^\top\,[\mathbf{Y_{1j}}-\boldsymbol{\mu}]\right) \\ & - \frac{1}{2}\,tr\left(\boldsymbol{\Sigma}_2^{-1}\sum_{j=1}^{n_2}[\mathbf{Y_{2j}}-\boldsymbol{\mu}]^\top\,[\mathbf{Y_{2j}}-\boldsymbol{\mu}]\right), \end{aligned}$$

where $\boldsymbol{\mu}$ denotes the unknown common mean under H_0. To find the restricted MLEs (RMLEs), by using the following equations as $\frac{\partial tr(\boldsymbol{X^{-1}}A)}{\partial \boldsymbol{X}} = -\boldsymbol{X^{-1}}A\boldsymbol{X^{-1}}$ or $\frac{\partial tr(\boldsymbol{X^{-1}}A)}{\partial \boldsymbol{X}} = -(\boldsymbol{X^{-1}})^\top A^\top(\boldsymbol{X^{-1}})^\top$ the partial derivatives of Eq. (2.4) with respect to $\boldsymbol{\Sigma_1}$, $\boldsymbol{\Sigma_2}$ and $\boldsymbol{\mu}$ yield the following equations:

$$\frac{\partial lnL}{\partial \boldsymbol{\Sigma}_1} = -\frac{n_1}{2}\,\boldsymbol{\Sigma}_1^{-1} + \frac{1}{2}\,\boldsymbol{\Sigma}_1^{-1}\sum_{j=1}^{n_1}[\mathbf{Y_{1j}}-\boldsymbol{\mu}]^\top\,[\mathbf{Y_{1j}}-\boldsymbol{\mu}]\boldsymbol{\Sigma}_1^{-1}\,,$$

$$\frac{\partial lnL}{\partial \boldsymbol{\Sigma}_2} = -\frac{n_2}{2}\,\boldsymbol{\Sigma}_2^{-1} + \frac{1}{2}\,\boldsymbol{\Sigma}_2^{-1}\sum_{j=1}^{n_2}\left[\mathbf{Y}_{\mathbf{2j}} - \boldsymbol{\mu}\right]^\top \left[\mathbf{Y}_{\mathbf{2j}} - \boldsymbol{\mu}\right]\boldsymbol{\Sigma}_2^{-1}\,,$$

$$\frac{\partial lnL}{\partial \boldsymbol{\mu}} = n_1 \boldsymbol{\Sigma}_1^{-1}\bar{\mathbf{Y}}_1 - \frac{1}{2}\Big(2n_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}^\top\Big) + n_2\,\boldsymbol{\Sigma}_2^{-1}\bar{\mathbf{Y}}_2 - \frac{1}{2}\Big(2n_2\,\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}^\top\Big)\,.$$

The RMLEs are given by

$$\hat{\boldsymbol{\Sigma}}_{1(\mathrm{RML})} = \frac{\sum\limits_{j=1}^{n_1}\left[\mathbf{Y}_{\mathbf{1j}} - \hat{\boldsymbol{\mu}}_{(\mathrm{RML})}\right]^\top \left[\mathbf{Y}_{\mathbf{1j}} - \hat{\boldsymbol{\mu}}_{(\mathrm{RML})}\right]}{n_1}\,,$$

$$\hat{\boldsymbol{\Sigma}}_{2(\mathrm{RML})} = \frac{\sum\limits_{j=1}^{n_2}\left[\mathbf{Y}_{\mathbf{2j}} - \hat{\boldsymbol{\mu}}_{(\mathrm{RML})}\right]^\top \left[\mathbf{Y}_{\mathbf{2j}} - \hat{\boldsymbol{\mu}}_{(\mathrm{RML})}\right]}{n_2}\,,$$

$$\hat{\boldsymbol{\mu}}^\top_{(\mathrm{RML})} = \Big(n_1\hat{\boldsymbol{\Sigma}}^{-1}_{1(\mathrm{RML})} + n_2\,\hat{\boldsymbol{\Sigma}}^{-1}_{2(\mathrm{RML})}\Big)^{-1}\Big(n_1\hat{\boldsymbol{\Sigma}}^{-1}_{1(\mathrm{RML})}\bar{\mathbf{Y}}_1 + n_2\,\hat{\boldsymbol{\Sigma}}^{-1}_{2(\mathrm{RML})}\bar{\mathbf{Y}}_2\Big)\,. \tag{2.5}$$

Since there are no close forms of these equations, these estimators can be obtained iteratively as follows: updating the estimates from l-step estimates ($\boldsymbol{\Sigma}_1^{(l)}, \boldsymbol{\Sigma}_2^{(l)}$ and $\boldsymbol{\mu}^{(l)}$) by

$$\boldsymbol{\Sigma}_1^{(l+1)} = \frac{\sum\limits_{j=1}^{n_1}\left[\mathbf{Y}_{\mathbf{1j}} - \boldsymbol{\mu}^{(l+1)}\right]^\top \left[\mathbf{Y}_{\mathbf{1j}} - \boldsymbol{\mu}^{(l+1)}\right]}{n_1}\,,$$

$$\boldsymbol{\Sigma}_2^{(l+1)} = \frac{\sum\limits_{j=1}^{n_2}\left[\mathbf{Y}_{\mathbf{2j}} - \boldsymbol{\mu}^{(l+1)}\right]^\top \left[\mathbf{Y}_{\mathbf{2j}} - \boldsymbol{\mu}^{(l+1)}\right]}{n_2}\,,$$

$$\boldsymbol{\mu}^{(l+1)} = \Big(n_1\big(\boldsymbol{\Sigma}_1^{(l+1)}\big)^{-1} + n_2\big(\boldsymbol{\Sigma}_2^{(l+1)}\big)^{-1}\Big)^{-1}\Big(n_1\big(\boldsymbol{\Sigma}_1^{(l+1)}\big)^{-1}\bar{\mathbf{Y}}_1 + n_2\big(\boldsymbol{\Sigma}_2^{(l+1)}\big)^{-1}\bar{\mathbf{Y}}_2\Big)\,.$$

where initial value $\boldsymbol{\mu}^{(0)}$ could set as $\boldsymbol{\mu}^{(0)} = \big(n_1\mathbf{S}_1^{-1} + n_2\mathbf{S}_2^{-1}\big)^{-1}\big(n_1\mathbf{S}_1^{-1}\bar{\mathbf{Y}}_1 + n_2\mathbf{S}_2^{-1}\bar{\mathbf{Y}}_2\big)$. $\boldsymbol{\Sigma}_1^{(l)}$, $\boldsymbol{\Sigma}_2^{(l)}$ and $\boldsymbol{\mu}^{(l)}$ converge to the RMLEs under H_0 denoted as $\hat{\boldsymbol{\Sigma}}_{i(\mathrm{RML})}$ and $\hat{\boldsymbol{\mu}}_{(\mathrm{RML})}$. For example, let

$$p = 3\,, \qquad n_1 = n_2 = 5\,, \qquad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = [1 \quad 1 \quad 1]\,,$$

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix},$$

and

$$\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}.$$

The Monte Carlo estimates of the expected value of $\hat{\mathbf{\Sigma}}_{i(\mathrm{RML})}$ and $\hat{\boldsymbol{\mu}}_{(\mathrm{RML})}$ are

$$\tilde{E}(\hat{\boldsymbol{\mu}}_{\mathrm{RML}}) = [1.005 \quad 1.006 \quad 1.003]\,,$$

$$\tilde{E}(\hat{\mathbf{\Sigma}}_{1(\mathrm{RML})}) = \begin{bmatrix} 0.969 & 0.190 & 0.199 \\ 0.190 & 0.981 & 0.197 \\ 0.199 & 0.197 & 0.969 \end{bmatrix},$$

and

$$\tilde{E}(\hat{\mathbf{\Sigma}}_{2(\mathrm{RML})}) = \begin{bmatrix} 0.999 & 0.795 & 0.798 \\ 0.795 & 0.994 & 0.794 \\ 0.798 & 0.794 & 0.992 \end{bmatrix}.$$

As seen from above, the obtained result is well.

3. TEST STATISTICS

Let $\boldsymbol{Y_{i1}}, \boldsymbol{Y_{i2}}, ..., \boldsymbol{Y_{in_i}}$, have p-variate normal distribution with mean vector $\boldsymbol{\mu_i}$ and covariance matrix $\mathbf{\Sigma_i}$, $i = 1, 2$. Let us denote

$$\mathbf{S}_{(i)} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)\,(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)^\top\,, \qquad i = 1, 2\,,$$

$$\tilde{\mathbf{\Sigma}}_i = \frac{1}{n_i}\mathbf{\Sigma}_i \quad \text{and} \quad \tilde{\mathbf{S}}_i = \frac{1}{n_i}\mathbf{S}_i\,, \qquad i = 1, 2\,.$$

Since the sample mean vector of i-th group, $\bar{\boldsymbol{Y}}_i$ and the maximum likelihood estimation of covariance matrix of i-th group, $\boldsymbol{S}_i$'s are independent from each other, the following equations can be written as follows:

$$\bar{\mathbf{Y}}_i \sim N_p\Big(\boldsymbol{\mu}_i, \frac{1}{n_i}\mathbf{\Sigma}_i\Big) \quad \text{and} \quad \tilde{\mathbf{S}}_i \sim W_p\Big(n_i - 1, \frac{1}{n_i - 1}\tilde{\mathbf{\Sigma}}_i\Big), \qquad i = 1, 2\,.$$

Here $W_p(r, \mathbf{\Sigma})$ is p-variate Wishart distribution with degrees of freedom r. This distribution is also known as generalized chi-square distribution which is obtained by Wishart [25].

The null and alternative hypotheses for testing the equality of two multivariate normal mean vectors are as follows:

$$H_0\colon\ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2\,, \qquad H_1\colon\ \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2\,. \tag{3.1}$$

For this problem, a natural statistic, the multivariate version of the statistic considered by Welch [24], is given as follows:

$$T = (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}})^\top (\mathbf{S}_e)^{-1} (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{Y}_2)\,, \tag{3.2}$$

where $\boldsymbol{S}_e = \frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2}$.

T statistic is asymptotically distributed as chi-square with degrees of freedom p when n_1 and n_2 approach to infinite. This approach is not valid for the small values of n_1 and n_2. Under H_0 and the assumption of the homogeneity of covariance matrices ($\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$), $(n-p-1)T/(p(n-2))$ is distributed as F with degrees of freedom p and $n-p-1$, where $n = n_1 + n_2$.

In the rest of this section, the tests given by Bennett [2], Johansen [14], Nel and Van Der Merwe [18], Krishnamoorthy and Yu [17] were introduced briefly.

3.1. Bennett test

Bennett [2] proposed a test for the equality of two mean vectors for $n_2 \geq n_1$. This test statistic can be given as follows:

$$T_B = n_1 \bar{\mathbf{z}}^\top \mathbf{S}_{\mathbf{z}}^{\mathbf{-1}} \bar{\mathbf{z}}\,, \tag{3.3}$$

where $\boldsymbol{z_j} = \mathbf{Y_{1\boldsymbol{j}}} - \sqrt{\frac{n_1}{n_2}}\,\mathbf{Y_{2\boldsymbol{j}}} + \frac{1}{\sqrt{n_1 n_2}} \sum\limits_{k=1}^{n_1} \mathbf{Y_{1\boldsymbol{k}}} - \frac{1}{n_2} \sum\limits_{k=1}^{n_2} \mathbf{Y_{2\boldsymbol{k}}},\ \ j = 1, ..., n_1,$

and also $\bar{\boldsymbol{z}}$ and $\boldsymbol{S_z}$ are the mean and variance-covariance matrix of $\boldsymbol{z_j}$, $j = 1, .., n_1$, respectively. By using the following transformation, the distribution of test statistic can be obtained as follows:

$$F = \frac{n_1 - p}{p(n_1 - 1)} T_B \sim F_{p, n_1 - p}\,.$$

3.2. Johansen test

Johansen [14] obtained a test given below:

$$T_J = \frac{T}{C}\,. \tag{3.4}$$

Here, T is given in Eq. (2.6),

$$C = p - 2D - 6D/\left[p(p-1)+2\right],$$

$$D = \sum_{i=1}^{2} \frac{1}{2(n_i-1)} \left\{ tr(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i)^2 + \left[tr(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i)\right]^2 \right\},$$

where $\boldsymbol{V}_i = (\boldsymbol{S_i}/n_i)^{-1}$, $i = 1, 2$ and $\boldsymbol{V} = \boldsymbol{V}_1 + \boldsymbol{V}_2$.

This test statistic is distributed as F with degrees of freedom p and $f = p(p+2)/(3D)$.

3.3. Nel and Van der Merwe test

Nel and Van der Merwe [18] modified the test statistic given in Eq. (3.2) as follows:

$$T_{\mathrm{NV}} = \frac{v-p+1}{p\,v}\, T \sim F_{p,v-p+1}, \tag{3.5}$$

where $v = \dfrac{tr(\boldsymbol{S_e})^2 + [tr(\boldsymbol{S_e})]^2}{\frac{1}{n_1-1}\left\{tr\left(\frac{\boldsymbol{S_1}}{n_1}\right)^2 + \left[tr\left(\frac{\boldsymbol{S_1}}{n_1}\right)\right]^2\right\} + \frac{1}{n_2-1}\left\{tr\left(\frac{\boldsymbol{S_2}}{n_2}\right)^2 + \left[tr\left(\frac{\boldsymbol{S_2}}{n_2}\right)\right]^2\right\}}$.

3.4. Krishnamoorthy and Yu test

Krishnamoorthy and Yu [17] obtained a test statistic by modifying the test statistic given by Nel and Van der Merwe [18]. The test statistic is as follows:

$$T_{\mathrm{M}} = \frac{(\hat{v}_{\mathrm{M}} - p + 1)\, T}{p\, \hat{v}_{\mathrm{M}}}, \tag{3.6}$$

where

$$\hat{v}_{\mathrm{M}} = \frac{p\,(p+1)\,(n-2)}{\hat{\varphi}_1 + \hat{\varphi}_2},$$

$$\hat{\varphi}_1 = \frac{n_2^2\,(n-2)}{n^2\,(n_1-1)} \left\{tr(\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}})\right\}^2 + \frac{n_1^2\,(n-2)}{n^2\,(n_2-1)} \left\{tr(\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}})\right\}^2,$$

$$\hat{\varphi}_2 = \frac{n_2^2\,(n-2)}{n^2\,(n_1-1)}\, tr(\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}}\mathbf{S}_1\bar{\mathbf{S}}^{-\mathbf{1}}) + \frac{n_1^2\,(n-2)}{n^2\,(n_2-1)}\, tr(\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}}\mathbf{S}_2\bar{\mathbf{S}}^{-\mathbf{1}})$$

and

$$\bar{\mathbf{S}} = \frac{n_2}{n}\mathbf{S}_1 + \frac{n_1}{n}\mathbf{S}_2 .$$

T_{M} is distributed as F with degrees of freedom p and $\hat{v}_{\mathrm{M}} - p + 1$ [26].

3.5. Yao test

Yao [27] proposed a test which is an extension of the Welch test provided by Tukey [23]. This test statistic T_Y based on T in Eq. (3.2) can be given as:

$$T_Y = \frac{m-p+1}{pm} T^2 \sim F_{p,m-p+1}, \tag{3.7}$$

where $\frac{1}{m} = \frac{1}{(T)^2} \sum_{i=1}^{2} \frac{1}{n_i - 1} \left[(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2)^\top (\boldsymbol{S_e})^{-1} \frac{\boldsymbol{S}_i}{n_i} (\boldsymbol{S_e})^{-1} (\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2) \right]^2$.

3.6. Wald test and its bootstrap approach

Konietschke *et al.* [16] developed parametric bootstrap methods for the multivariate Behrens–Fisher problem. According to this, for two multivariate normal mean vectors , H_0 stated in Eq. (3.1) is equivalent to testing $H_0^\top$: $\boldsymbol{H}\boldsymbol{\mu}^* = \mathbf{0}$, where $\boldsymbol{\mu}^* = (\boldsymbol{\mu_1}^\top, \boldsymbol{\mu_2}^\top)^\top$ and contrast matrix is given by $\boldsymbol{H} = \boldsymbol{P} \otimes \boldsymbol{I}_p$.

Here,

$$\mathbf{P} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and $\boldsymbol{I}_p$ is the p-dimensional unit matrix.

The Wald test statistic for testing $H_0^\top$ is

$$Q_n(H) = n\bar{\mathbf{Y}}^\top \mathbf{H}(\mathbf{H}\hat{\mathbf{V}}_\mathbf{n}\mathbf{H})^+ \boldsymbol{H}\bar{\boldsymbol{Y}}, \tag{3.8}$$

where $(\cdot)^+$ denotes the Moore-Penrose inverse and $\hat{\boldsymbol{V}}_{\boldsymbol{n}} = \mathrm{diag}\big(\frac{n}{n_i}\boldsymbol{S}_{\boldsymbol{(i)}}; 1 \le i \le 2\big)$. $Q_n(H)$ is asymptotically distributed as χ^2 with degrees of freedom rank$(\boldsymbol{H})$.

Konietschke *et al.* [16] applied the nonparametric and parametric bootstrap of the Wald test. The algorithm of these tests, respectively, are given below:

The nonparametric bootstrap of the Wald test:

1. For given data, calculate value of test statistic given in Eq. (3.8).
2. Generate nonparametric bootstrap sample, $\boldsymbol{Y}_{i1}^\bullet, \boldsymbol{Y}_{i2}^\bullet, ..., \boldsymbol{Y}_{in_i}^\bullet$, which are drawn with replacement from the pooled observation vectors, $\mathbf{Y}_{11}, ..., \boldsymbol{Y}_{2n_2}$.
3. Compute value of test statistic given in Eq. (3.8) from the nonparametric bootstrap sample and denote it by $Q_n^{\bullet(l)}(H)$.
4. Repeat the steps 2 and 3 for a large number of times (say, L times).
5. Compute the Monte Carlo estimates of the p-values as $\hat{p} = \sum_{l=1}^{L} I\Big(Q_n^{\bullet(l)}(H) > Q_n(H)\Big)/L$, where I is the indicator function.
6. If $\hat{p} < \alpha$, H_0 is rejected.

We refer to the nonparametric bootstrap of the Wald test as WB in rest of study.

The parametric bootstrap of the Wald test:

1. For given data, calculate value of test statistic given in Eq. (3.8).
2. Generate parametric bootstrap variables as

$$\boldsymbol{Y}_{i1}^*, \boldsymbol{Y}_{i2}^*, ..., \boldsymbol{Y}_{in_i}^* \sim N(0, \boldsymbol{S}_{(i)}), \qquad i = 1, 2.$$

3. Compute value of test statistic given in Eq. (3.8) from the parametric bootstrap vectors, and denote it by $Q_n^{*(l)}(H)$.
4. Repeat the steps 2 and 3 for a large number of times (say, L times).
5. Compute the Monte Carlo estimates of the p-values as $\hat{p} = \sum_{l=1}^{L} I\Big(Q_n^{*(l)}(H) > Q_n(H)\Big)/L$, where I is the indicator function.
6. If $\hat{p} < \alpha$, H_0 is rejected.

We refer to the parametric bootstrap of the Wald test as WPB in rest of study.

4. COMPUTATIONAL APPROACH TEST

Initially, before applying CAT for testing the null hypothesis given in Eq. (3.1), the general technique in CAT was first given.

Let $X_1, X_2, ..., X_n$ be a random sample having a probability density function as $f(x|\theta)$, where the functional form of f is assumed to be known. Let $\theta = (\theta^{(1)}, \theta^{(2)})$ be the parameter vector and our primary interest lies in $\theta^{(1)}$, i.e., $\theta^{(2)}$ is the nuisance parameter. Our goal is to test $H_0^\top\colon \theta^{(1)} = \theta_0^{(1)}$ versus a suitable alternative. To test $H_0^\top\colon \theta^{(1)} = \theta_0^{(1)}$ against $H_1^\top$, $H_0^\top$ was first expressed as $H_0^{\top *}\colon \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against $H_1^{\top *}$, where η is a scalar valued function. The general methodology of the CAT for testing $H_0^{\top *}\colon \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against a suitable alternative at a desired level α was given through the following steps [19]:

1. Obtain the MLEs of the parameters, $\theta^{(1)}$ and $\theta^{(2)}$. Obtain a suitable $\eta\big(\theta^{(1)}, \theta_0^{(1)}\big)$ and the MLE of η, $\hat{\eta} = \hat{\eta}(\hat{\theta}^{(1)}, \theta_0^{(1)})$ can be used as a test statistic.
2. Under H_0, find the RMLEs of $\theta^{(2)}$ parameter, which is denoted by $\tilde{\theta}^{(2)}$. Generate artificial sample $Y_1, Y_2, ..., Y_n$ from $f\big(y|\theta_0^{(1)}, \tilde{\theta}^{(2)}\big)$ large number of times, say L times.
3. For each of these replicated samples, recalculate the MLE of η, $\hat{\eta}^{*(l)}$, $l = 1, ..., L$.
4. Estimate the p-value as, $\hat{p} = \sum_{l=1}^{L} \big(\hat{\eta}^{*(l)} > \hat{\eta}\big)/L$. In the case of $\hat{p} < \alpha$, H_0 is rejected.

CAT is based on restricted maximum likelihood estimations (RMLEs) under null hypothesis. There is no need to obtain theoretical distribution of test statistic and the p-value can be calculated directly; therefore, this method is quite easy to apply. Then, the testing procedure based on CAT for the equality of two multivariate normal mean vectors under heterogeneity of covariance matrix can be given as follows:

The observed value of the test statistic based on random sample is calculated as follows:

Step 1. The observed value of the test statistic based on random sample is calculated as follows:

$$\hat{\eta}_{\mathrm{ML}} = (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}})^{\top} \left(\frac{\hat{\Sigma}_1}{n_1} + \frac{\hat{\Sigma}_2}{n_2} \right)^{-1} (\bar{\mathbf{Y}}_{\mathbf{1}} - \bar{\mathbf{Y}}_{\mathbf{2}}) \,. \tag{4.1}$$

Step 2. Under H_0, the RMLEs of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ are obtained as $\hat{\boldsymbol{\mu}}_{(\mathrm{RML})}$ and $\hat{\boldsymbol{\Sigma}}_{i(\mathrm{RML})}$ in Eq. (2.5), iteratively.

Step 3. A large number is generated, say L, of artificial sample from $N_p(\hat{\boldsymbol{\mu}}_{(\mathrm{RML})}, \hat{\boldsymbol{\Sigma}}_{i(\mathrm{RML})})$, $i = 1, 2$. For every artificial sample $\hat{\eta}_{\mathrm{ML}}^{(l)}$, $l = 1, ..., L$ are calculated.

Step 4. p value is calculated as $\hat{p} = \sum_{l=1}^{L} \frac{(\hat{\eta}_{\mathrm{ML}}^{(l)} > \hat{\eta}_{\mathrm{ML}})}{L}$. H_0 is rejected when $\hat{p} < \alpha$.

By using these steps, a simulation study was carried out.

First, we carry out the CAT for testing the H_0 using the test statistic $\hat{\eta}_{\mathrm{ML}}$ in Eq. (4.1) and refer to this test statistics as CAT. Beside, we carry out the CAT using the test statistics T_B in Eq. (3.3), T_{M} in Eq. (3.6), T_J in Eq. (3.4), T_Y in Eq. (3.7) and T_{NV} in Eq. (3.5), we refer to these test statistics as B-CAT, M-CAT, J-CAT, Y-CAT, NV-CAT, respectively.

5. SIMULATION STUDY

In this section, all tests were compared with respect to their estimated type-I error rates and powers for multivariate normality and non-normality. For this purpose, the cases of p= 2, 3, 4 with different combinations of equal and unequal sample sizes were considered. To estimate type-I error rates and the powers of all tests under multivariate normality assumption, 2000 random numbers with a sample size n_i (i=1,2) from the multivariate normal distribution were generated. Mean vectors were used as $\boldsymbol{\mu}_1 = (0, 0, ..., 0)_{1\times p}$ and $\boldsymbol{\mu}_2 = (\Delta, \Delta, ..., \Delta)_{1\times p}$.

Following Konietschke *et al.* [16], we considered six covariance structure as:

Setting 1: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p) = \boldsymbol{\Sigma}_2$,

Setting 2: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|} = \boldsymbol{\Sigma}_2$,

Setting 3: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p)$ and $\boldsymbol{\Sigma}_2 = 3I_p + 0.5(J_p - I_p)$,

Setting 4: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|}$ and $\boldsymbol{\Sigma}_2 = (0.6)^{|r-s|} + 2I_p$,

Setting 5: $\boldsymbol{\Sigma}_1 = I_p + 0.5(J_p - I_p)$ and $\boldsymbol{\Sigma}_2 = 9I_p + 0.5(J_p - I_p)$,

Setting 6: $\boldsymbol{\Sigma}_1 = [\sigma_{rs}] = (0.6)^{|r-s|}$ and $\boldsymbol{\Sigma}_2 = (0.6)^{|r-s|} + 8I_p$,

where I_p is an identity vector with dimension p and J_p is the $p \times p$ matrix of 1's. While setting 1 represents a scenario with homoscedastic compound symmetric, settings 3 and 5 represent the scenarios with moderate and severe heteroscedastic versions of this structure, respectively.

While setting 2 represents a scenario with homoscedastic autoregressive covariance structure, setting 4 and 6 represent the scenarios with moderate and severe heteroscedastic versions of this structure, respectively.

To calculate the $\hat{p}$-values of the CAT and CAT versions of the tests, m was taken as 2000. The simulation study was conducted in MATLAB. Initially, the estimated type-I error rates of all tests under the null hypothesis were calculated. The simulation results were provided in Tables 1–3 under the multivariate normal model at the nominal level 0.05.

Table 1: The estimated type I error rates of all tests under multivariate normal distribution for $p = 2$. In the table, the estimated type I error rates exceeding 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,038	0,045	0,040	0,040	0,039	0,041	0,045	0,039	0,042	0,042	0,038	**0,101**	0,045	0,034
	15, 15	0,049	0,052	0,050	0,050	0,049	0,050	0,051	0,050	0,050	0,049	0,050	**0,079**	0,051	0,051
	30, 30	0,054	0,054	0,055	0,054	0,055	0,055	0,054	0,054	0,054	0,054	0,053	**0,067**	0,054	0,050
	7, 15	0,055	0,050	0,055	0,055	0,057	0,052	0,050	0,055	0,058	0,059	0,051	**0,101**	0,053	0,050
	15, 30	0,044	0,043	0,045	0,045	0,045	0,050	0,043	0,045	0,045	0,046	0,045	**0,069**	0,043	0,057
2	7, 7	0,040	0,052	0,041	0,041	0,041	0,039	0,051	0,039	0,043	0,043	0,038	**0,105**	0,045	0,055
	15, 15	0,053	0,056	0,053	0,053	0,052	0,044	0,055	0,052	0,052	0,053	0,051	**0,093**	0,054	0,053
	30, 30	0,051	0,055	0,052	0,052	0,052	0,044	0,056	0,050	0,050	0,051	0,050	**0,061**	0,050	0,047
	7, 15	0,045	0,050	0,045	0,045	0,045	0,048	0,051	0,047	0,047	0,052	0,042	**0,098**	0,041	0,054
	15, 30	0,041	0,048	0,041	0,041	0,041	0,052	0,047	0,041	0,041	0,042	0,042	0,058	0,040	0,047
3	7, 7	0,049	0,051	0,051	0,051	0,051	0,042	0,057	0,051	0,055	0,059	0,046	**0,121**	**0,069**	0,051
	15, 15	0,056	0,057	0,056	0,056	0,056	0,049	0,058	0,059	0,059	0,060	0,059	**0,085**	**0,060**	0,059
	30, 30	0,055	0,052	0,055	0,055	0,055	0,048	0,054	0,055	0,055	0,055	0,055	**0,071**	0,059	0,056
	7, 15	0,049	0,056	0,049	0,049	0,049	0,054	0,054	0,049	0,049	0,050	0,048	**0,082**	0,037	0,049
	15, 30	0,047	0,043	0,047	0,047	0,047	0,056	0,045	0,047	0,047	0,047	0,047	0,058	0,044	0,053
4	7, 7	0,040	0,042	0,042	0,042	0,041	0,045	0,042	0,040	0,045	0,048	0,040	**0,113**	0,057	0,059
	15, 15	0,053	0,054	0,053	0,053	0,053	0,055	0,055	0,054	0,054	0,055	0,054	**0,077**	0,057	0,054
	30, 30	0,056	0,057	0,056	0,056	0,056	0,052	0,056	0,056	0,055	0,057	0,055	**0,070**	0,058	0,053
	7, 15	0,051	0,054	0,051	0,051	0,051	0,054	0,052	0,050	0,050	0,052	0,048	**0,084**	0,037	0,052
	15, 30	0,053	0,045	0,053	0,053	0,053	0,051	0,046	0,053	0,053	0,054	0,053	**0,069**	0,051	0,047
5	7, 7	0,050	0,048	0,050	0,050	0,049	0,049	0,048	0,051	0,057	0,061	0,048	**0,132**	**0,083**	0,057
	15, 15	0,053	0,054	0,053	0,053	0,053	0,056	0,052	0,053	0,053	0,054	0,053	**0,090**	**0,071**	0,049
	30, 30	0,045	0,044	0,045	0,045	0,045	0,054	0,045	0,045	0,045	0,046	0,045	**0,063**	0,054	0,050
	7, 15	0,052	0,045	0,052	0,052	0,052	0,046	0,046	0,051	0,051	0,051	0,050	**0,079**	0,039	0,052
	15, 30	0,047	0,050	0,047	0,047	0,047	0,059	0,050	0,050	0,050	0,050	0,050	**0,068**	0,045	0,045
6	7, 7	0,052	0,047	0,051	0,051	0,052	0,056	0,047	0,053	0,056	0,065	0,049	**0,143**	**0,092**	0,057
	15, 15	0,054	0,048	0,055	0,055	0,055	0,047	0,050	0,055	0,056	0,058	0,054	**0,088**	**0,070**	0,046
	30, 30	0,047	0,047	0,042	0,046	0,046	0,049	0,045	0,043	0,045	0,045	0,045	0,059	0,051	0,058
	7, 15	0,052	0,050	0,053	0,053	0,053	0,054	0,050	0,052	0,053	0,052	0,052	**0,085**	0,040	0,047
	15, 30	0,048	0,054	0,048	0,048	0,048	0,057	0,054	0,047	0,047	0,048	0,048	**0,067**	0,045	0,049

It can be seen in Table 1 that for $p = 2$, in cases of homoscedastic and moderate heteroscedastic structures (setting 1, 2, 3 and 4), all tests except Wald test have the estimated type I error rates close to the nominal level 0.05. As heterogeneity is getting severe (setting 5 and 6), in cases of small sample sizes, the estimated type I error rates of the WB and T_Y tests as well as the Wald test exceed the nominal level 0.05, that is, these tests tend to be liberal.

It can be seen in Table 2 that in case of homoscedastic structure, the results of the tests for $p = 3$ show similar pattern as those for the cases of $p = 2$. However, in case of heterogeneity even when moderate heterogeneity, the T_Y, T_J, Wald, WB tests tend to be liberal for small sample sizes. Furthermore, as heterogeneity is getting severe, these tests tend to be highly liberal for small sample size. A remarkable consequence is that when sample sizes are different, the estimated type I error rates of the WB test are considerably lower than the nominal level 0.05.

Table 2: The estimated type I error rates of all tests under multivariate normal distribution for $p = 3$. In the table, the estimated type I error rates exceeding 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,037	0,046	0,040	0,039	0,040	0,041	0,046	0,039	0,046	0,045	0,036	**0,150**	0,046	0,041
	15, 15	0,046	0,062	0,048	0,048	0,047	0,049	0,061	0,047	0,048	0,048	0,047	**0,087**	0,049	0,049
	30, 30	0,042	0,045	0,042	0,042	0,041	0,048	0,046	0,042	0,042	0,042	0,042	0,059	0,042	0,052
	7, 15	0,045	0,047	0,045	0,045	0,046	0,049	0,049	0,044	0,049	**0,061**	0,035	**0,140**	0,039	**0,066**
	15, 30	0,055	0,050	0,055	0,055	0,056	0,036	0,050	0,057	0,058	0,059	0,054	**0,088**	0,056	0,059
2	7, 7	0,045	0,056	0,047	0,046	0,045	0,040	0,058	0,046	0,056	0,050	0,046	**0,156**	0,059	0,041
	15, 15	0,049	0,051	0,050	0,050	0,049	0,057	0,050	0,047	0,048	0,050	0,047	**0,088**	0,048	0,056
	30, 30	0,050	0,046	0,050	0,050	0,050	0,050	0,046	0,048	0,047	0,048	0,048	**0,065**	0,048	0,057
	7, 15	0,049	0,041	0,048	0,049	0,052	0,050	0,042	0,048	0,053	**0,064**	0,036	**0,143**	0,046	0,056
	15, 30	0,049	0,056	0,048	0,048	0,049	0,052	0,056	0,050	0,050	0,054	0,050	**0,085**	0,048	0,047
3	7, 7	0,052	0,048	0,054	0,052	0,056	0,048	0,049	0,055	**0,068**	**0,074**	0,050	**0,200**	**0,083**	0,048
	15, 15	0,052	0,049	0,052	0,052	0,052	0,049	0,049	0,052	0,053	0,056	0,050	**0,099**	**0,065**	0,045
	30, 30	0,049	0,057	0,049	0,049	0,049	0,049	0,059	0,048	0,048	0,048	0,051	**0,071**	0,057	0,047
	7, 15	0,053	0,048	0,053	0,053	0,054	0,048	0,050	0,051	0,054	0,056	0,051	**0,117**	0,037	0,042
	15, 30	0,043	0,047	0,044	0,044	0,044	0,058	0,048	0,044	0,044	0,043	0,043	**0,066**	0,036	0,056
4	7, 7	0,045	0,043	0,046	0,045	0,048	0,058	0,044	0,044	0,058	0,059	0,039	**0,183**	**0,070**	0,051
	15, 15	0,053	0,046	0,053	0,053	0,053	0,051	0,048	0,052	0,053	0,058	0,052	**0,093**	**0,063**	0,050
	30, 30	0,054	0,058	0,054	0,054	0,054	0,050	0,056	0,055	0,055	0,056	0,055	**0,076**	**0,062**	0,047
	7, 15	0,045	0,052	0,047	0,045	0,047	0,047	0,053	0,047	0,050	0,048	0,040	**0,107**	0,027	0,046
	15, 30	0,043	0,044	0,043	0,043	0,043	0,047	0,042	0,045	0,045	0,045	0,044	**0,064**	0,038	0,053
5	7, 7	0,056	**0,064**	0,055	0,056	0,058	0,056	**0,061**	0,058	**0,075**	**0,089**	0,044	**0,236**	**0,121**	0,059
	15, 15	0,054	**0,060**	0,052	0,053	0,053	0,049	0,059	0,054	0,057	**0,060**	0,052	**0,120**	**0,085**	0,048
	30, 30	0,038	0,043	0,039	0,039	0,040	0,051	0,044	0,040	0,040	0,041	0,039	**0,067**	0,054	0,054
	7, 15	0,043	0,047	0,044	0,044	0,042	0,055	0,048	0,045	0,045	0,047	0,045	**0,112**	0,029	0,051
	15, 30	0,041	0,050	0,041	0,041	0,041	0,056	0,050	0,043	0,043	0,044	0,043	**0,069**	0,040	0,051
6	7, 7	0,054	0,055	0,053	0,053	0,057	0,056	0,054	0,054	**0,075**	**0,090**	0,044	**0,237**	**0,127**	**0,062**
	15, 15	0,044	0,043	0,044	0,044	0,045	0,051	0,041	0,045	0,045	0,048	0,043	**0,099**	**0,069**	0,049
	30, 30	0,057	0,056	0,057	0,056	0,057	0,051	0,055	0,054	0,054	0,057	0,054	**0,091**	**0,069**	0,050
	7, 15	0,045	0,052	0,045	0,045	0,046	0,050	0,054	0,044	0,046	0,047	0,042	**0,105**	0,032	0,057
	15, 30	0,053	0,054	0,053	0,053	0,053	0,054	0,055	0,052	0,052	0,053	0,053	**0,085**	0,046	0,048

It can be seen in Table 3 that for $p = 4$, in case of homoscedastic structure and different sample sizes, the T_Y and T_J, tests tend to be liberal. Furthermore, when small and equal sample sizes, these tests tend to be highly liberal in case of heterogeneity. As sample sizes increase, these tests have the estimated type I error rates close to the nominal level 0.05. Besides, in case of heteroscedastic structure, the estimated type I error rates of the WB test exhibit similar pattern as those for the cases of $p = 2$ and $p = 3$. However, as p increases, this test tend to be highly liberal. It can be seen from all tables that as p increase, the T_Y, T_J, and WB tests tend to be highly liberal. Furthermore, as heterogeneity is getting severe, these tests also tend to be highly liberal. The CAT versions of the $T_Y$and T_J, tests, J-CAT and Y-CAT, greatly improved these tests' behavior. In this cases, these tests have the estimated type I error rates close to the nominal level 0.05 in most cases. As for the WPB test, in homoscedastic structure, the estimated type I error rates of this test is slightly higher than the nominal level 0.05 when sample size is small and unequal. Furthermore, as heterogeneity is getting severe, the estimated type I error rates of this test is getting quite higher than the nominal level 0.05.

Table 3: The estimated type I error rates of all tests under multivariate normal distribution for $p = 4$. In the table, the estimated type I error rates exceeding 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,040	0,051	0,042	0,042	0,042	0,050	0,050	0,040	0,059	0,049	0,036	**0,215**	0,054	0,039
	15, 15	0,053	0,052	0,055	0,055	0,055	0,050	0,055	0,052	0,056	0,057	0,051	**0,118**	0,054	0,043
	30, 30	0,051	0,046	0,051	0,051	0,051	0,048	0,046	0,051	0,051	0,050	0,051	**0,076**	0,051	0,051
	7, 15	0,055	0,053	0,051	0,053	0,056	0,057	0,052	0,050	0,065	0,082	0,034	**0,206**	0,043	0,059
	15, 30	0,050	0,052	0,049	0,049	0,050	0,054	0,054	0,050	0,050	0,059	0,048	**0,108**	0,050	0,048
2	7, 7	0,039	0,060	0,043	0,041	0,041	0,044	0,059	0,038	**0,061**	0,055	0,039	**0,211**	0,054	0,040
	15, 15	0,057	0,049	0,058	0,057	0,058	0,048	0,050	0,057	0,059	0,057	0,056	**0,116**	0,058	0,048
	30, 30	0,048	0,056	0,048	0,048	0,048	0,052	0,056	0,048	0,048	0,049	0,048	**0,080**	0,051	0,050
	7, 15	0,054	0,056	0,054	0,054	0,059	0,059	0,056	0,051	**0,065**	**0,072**	0,037	**0,197**	0,049	**0,066**
	15, 30	0,048	0,053	0,048	0,048	0,048	0,053	0,052	0,047	0,048	0,052	0,043	**0,106**	0,047	0,055
3	7, 7	0,051	0,049	0,052	0,051	0,056	0,047	0,048	0,048	**0,079**	**0,083**	0,041	**0,267**	**0,091**	0,059
	15, 15	0,046	0,053	0,046	0,046	0,048	0,044	0,055	0,046	0,049	0,055	0,046	**0,133**	**0,066**	0,052
	30, 30	0,048	0,052	0,047	0,047	0,048	0,051	0,053	0,049	0,049	0,050	0,049	**0,073**	0,056	0,053
	7, 15	0,045	0,055	0,045	0,045	0,047	0,053	0,056	0,047	0,050	0,052	0,043	**0,138**	0,029	0,037
	15, 30	0,044	0,044	0,044	0,044	0,044	0,048	0,045	0,043	0,043	0,043	0,042	**0,075**	0,029	0,047
4	7, 7	0,053	0,054	0,053	0,053	0,055	0,050	0,056	0,052	**0,081**	**0,083**	0,042	**0,264**	**0,093**	0,052
	15, 15	0,049	0,049	0,049	0,049	0,050	0,056	0,050	0,049	0,053	0,057	0,050	**0,135**	**0,070**	0,057
	30, 30	0,054	0,051	0,054	0,054	0,054	0,042	0,052	0,052	0,052	0,053	0,052	**0,083**	**0,061**	0,050
	7, 15	0,044	0,062	0,045	0,045	0,045	0,052	0,061	0,043	0,048	0,051	0,038	**0,143**	0,027	0,041
	15, 30	0,058	0,056	0,058	0,058	0,058	0,054	0,059	0,057	0,057	0,057	0,055	**0,090**	0,044	0,058
5	7, 7	0,057	**0,060**	0,055	0,056	**0,062**	**0,062**	**0,060**	0,055	**0,100**	**0,121**	0,029	**0,343**	**0,155**	**0,083**
	15, 15	0,047	0,047	0,045	0,046	0,046	0,054	0,047	0,048	0,053	0,057	0,044	**0,152**	**0,089**	0,054
	30, 30	0,048	0,048	0,044	0,046	0,046	0,053	0,045	0,045	0,047	0,050	0,046	**0,104**	**0,072**	0,054
	7, 15	0,053	0,047	0,053	0,053	0,054	0,050	0,045	0,053	0,058	0,056	0,052	**0,144**	0,037	0,043
	15, 30	0,046	0,055	0,046	0,046	0,046	0,040	0,054	0,048	0,048	0,050	0,048	**0,080**	0,042	0,044
6	7, 7	**0,064**	0,050	**0,064**	**0,064**	**0,066**	**0,063**	0,048	0,059	**0,107**	**0,123**	0,034	**0,360**	**0,162**	**0,067**
	15, 15	0,057	0,050	0,056	0,056	0,055	0,059	0,048	0,057	**0,062**	**0,068**	0,053	**0,158**	**0,097**	0,051
	30, 30	0,053	0,053	0,053	0,053	0,053	0,050	0,053	0,052	0,053	0,056	0,052	**0,100**	**0,074**	0,060
	7, 15	0,046	0,055	0,046	0,046	0,047	0,050	0,057	0,045	0,050	0,050	0,044	**0,144**	0,030	0,048
	15, 30	0,054	0,055	0,054	0,054	0,053	0,057	0,055	0,053	0,053	0,055	0,054	**0,093**	0,047	0,051

In summary, according to the results obtained from all tables, the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_{M} and WPB tests have the estimated type I error rates close to the nominal level 0.05 in most cases. Furthermore, as p increase and heterogeneity is getting severe, the estimated type I error rates of all tests are affected negatively from this case, that is, these tests behaviors depend on the p and degree of heterogeneity.

The simulated powers of all tests were provided in Tables 4–6 under the normal model. The tests attaining nominal level closely can be compared meaningfully in terms of power. Since the estimated type I error rates of the Wald test exceed 6% in all considered cases, the Wald test was ignored and excluded from tables. While the powers of the tests were interpreted, the tests which had greater the estimated type I error rates than 6% given in Tables 1–3 were disregarded. Thus, the estimated powers of these tests were denoted by $*$.

Table 4: The estimated powers of all tests under the multivariate normal distribution for $p = 2$. The estimated powers of the tests exceeding 6% in terms of the estimated type I error rates were indicated by '*' in the tables.

		Setting1		Setting2		Setting3		Setting4		Setting5		Setting6	
n	**Tests**	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75
(7, 7)	CAT	0,116	0,222	0,097	0,208	0,076	0,131	0,075	0,144	0,059	0,089	0,066	0,090
	B-CAT	0,103	0,193	0,093	0,170	0,074	0,117	0,079	0,130	0,064	0,087	0,057	0,083
	M-CAT	0,119	0,224	0,100	0,211	0,075	0,131	0,076	0,145	0,059	0,090	0,065	0,091
	J-CAT	0,117	0,222	0,099	0,210	0,076	0,131	0,075	0,144	0,059	0,090	0,065	0,090
	Y-CAT	0,121	0,223	0,099	0,210	0,075	0,132	0,077	0,145	0,057	0,090	0,067	0,092
	NV_CAT	0,115	0,231	0,113	0,196	0,076	0,137	0,076	0,136	0,062	0,085	0,072	0,085
	T_B	0,104	0,195	0,094	0,171	0,075	0,117	0,081	0,133	0,064	0,089	0,059	0,081
	T_M	0,116	0,224	0,102	0,209	0,078	0,132	0,078	0,148	0,062	0,092	0,068	0,091
	T_J	0,122	0,233	0,106	0,218	0,084	0,139	0,081	0,153	0,067	0,102	0,072	0,098
	T_Y	0,122	0,231	0,110	0,224	0,091	0,143	0,085	0,162	0,071	0,112	0,082	0,109
	T_{NV}	0,116	0,222	0,101	0,208	0,076	0,126	0,077	0,145	0,058	0,087	0,061	0,084
	WB	0,129	0,234	0,119	0,233	0,104	0,169	0,102	0,185	*	*	*	*
	WPB	0,099	0,224	0,116	0,211	0,084	0,142	0,078	0,141	0,071	0,107	0,067	0,088
(15, 15)	CAT	0,244	0,480	0,232	0,463	0,154	0,298	0,145	0,296	0,086	0,161	0,090	0,161
	B-CAT	0,222	0,440	0,206	0,423	0,152	0,282	0,140	0,281	0,088	0,151	0,086	0,155
	M-CAT	0,246	0,481	0,233	0,463	0,154	0,299	0,143	0,296	0,086	0,161	0,090	0,161
	J-CAT	0,246	0,481	0,233	0,463	0,154	0,298	0,143	0,296	0,086	0,161	0,090	0,161
	Y-CAT	0,245	0,481	0,234	0,463	0,156	0,299	0,144	0,296	0,086	0,161	0,090	0,162
	NV_CAT	0,246	0,503	0,248	0,466	0,146	0,312	0,162	0,293	0,098	0,162	0,087	0,156
	T_B	0,221	0,438	0,207	0,420	0,155	0,285	0,137	0,282	0,086	0,152	0,089	0,157
	T_M	0,250	0,481	0,228	0,464	0,159	0,300	0,142	0,295	0,088	0,161	0,092	0,162
	T_J	0,251	0,482	0,229	0,464	0,160	0,301	0,142	0,295	0,089	0,162	0,093	0,164
	T_Y	0,252	0,486	0,229	0,465	0,166	0,307	0,148	0,303	0,092	0,170	0,093	0,167
	T_{NV}	0,250	0,483	0,229	0,465	0,160	0,300	0,143	0,297	0,086	0,161	0,089	0,162
	WB	0,254	0,486	0,234	0,469	0,176	0,322	0,159	0,319	*	*	*	*
	WPB	0,242	0,500	0,236	0,482	0,147	0,320	0,157	0,311	0,088	0,154	0,087	0,155
(30, 30)	CAT	0,461	0,841	0,461	0,822	0,312	0,613	0,280	0,580	0,153	0,305	0,150	0,289
	B-CAT	0,439	0,813	0,440	0,804	0,294	0,596	0,274	0,566	0,152	0,292	0,146	0,286
	M-CAT	0,462	0,841	0,461	0,823	0,312	0,613	0,280	0,580	0,153	0,305	0,150	0,289
	J-CAT	0,462	0,841	0,461	0,823	0,312	0,613	0,280	0,580	0,153	0,305	0,150	0,289
	Y-CAT	0,462	0,841	0,461	0,823	0,313	0,614	0,282	0,582	0,153	0,305	0,151	0,290
	NV_CAT	0,493	0,845	0,455	0,815	0,280	0,600	0,292	0,585	0,167	0,312	0,153	0,297
	T_B	0,441	0,815	0,438	0,805	0,298	0,599	0,270	0,568	0,151	0,295	0,147	0,284
	T_M	0,465	0,841	0,462	0,822	0,311	0,613	0,282	0,582	0,155	0,301	0,149	0,290
	T_J	0,465	0,839	0,462	0,822	0,311	0,613	0,281	0,582	0,155	0,301	0,149	0,289
	T_Y	0,466	0,840	0,462	0,823	0,313	0,617	0,283	0,586	0,157	0,306	0,150	0,298
	T_{NV}	0,466	0,840	0,462	0,823	0,312	0,614	0,282	0,583	0,155	0,301	0,149	0,290
	WB	0,464	0,838	0,461	0,821	0,318	0,625	0,289	0,598	0,172	0,328	0,165	0,315
	WPB	0,474	0,837	0,449	0,789	0,296	0,594	0,301	0,589	0,162	0,298	0,166	0,305
(7, 15)	CAT	0,157	0,307	0,146	0,281	0,112	0,216	0,125	0,219	0,083	0,146	0,078	0,145
	B-CAT	0,122	0,228	0,114	0,220	0,092	0,163	0,101	0,174	0,076	0,114	0,079	0,118
	M-CAT	0,156	0,308	0,145	0,283	0,112	0,217	0,125	0,221	0,084	0,147	0,078	0,146
	J-CAT	0,157	0,309	0,145	0,282	0,112	0,217	0,125	0,220	0,084	0,146	0,078	0,146
	Y-CAT	0,158	0,305	0,146	0,282	0,112	0,217	0,126	0,220	0,085	0,147	0,079	0,146
	NV_CAT	0,154	0,315	0,143	0,288	0,119	0,231	0,119	0,227	0,087	0,130	0,086	0,144
	T_B	0,124	0,227	0,115	0,220	0,093	0,163	0,100	0,174	0,080	0,115	0,078	0,120
	T_M	0,155	0,308	0,147	0,286	0,117	0,223	0,128	0,226	0,087	0,143	0,077	0,150
	T_J	0,158	0,313	0,149	0,294	0,118	0,225	0,129	0,228	0,088	0,144	0,079	0,151
	T_Y	0,171	0,321	0,155	0,312	0,116	0,223	0,126	0,223	0,090	0,146	0,079	0,151
	T_{NV}	0,153	0,297	0,139	0,279	0,116	0,217	0,123	0,220	0,087	0,143	0,079	0,150
	WB	0,151	0,302	0,141	0,286	0,093	0,179	0,097	0,183	0,068	0,120	0,062	0,125
	WPB	0,155	0,306	0,168	0,310	0,125	0,215	0,138	0,230	0,097	0,144	0,087	0,136
(15, 30)	CAT	0,313	0,645	0,301	0,618	0,237	0,467	0,218	0,468	0,146	0,270	0,138	0,269
	B-CAT	0,289	0,593	0,261	0,548	0,210	0,413	0,197	0,414	0,119	0,247	0,137	0,246
	M-CAT	0,313	0,644	0,300	0,617	0,237	0,467	0,218	0,468	0,146	0,271	0,138	0,269
	J-CAT	0,313	0,645	0,300	0,617	0,237	0,467	0,218	0,468	0,147	0,271	0,138	0,269
	Y-CAT	0,313	0,645	0,299	0,615	0,238	0,468	0,217	0,469	0,147	0,271	0,139	0,270
	NV_CAT	0,343	0,642	0,289	0,596	0,243	0,469	0,227	0,469	0,141	0,273	0,134	0,265
	T_B	0,292	0,593	0,261	0,551	0,209	0,416	0,197	0,416	0,123	0,243	0,141	0,246
	T_M	0,315	0,647	0,297	0,618	0,234	0,473	0,218	0,466	0,148	0,273	0,139	0,268
	T_J	0,315	0,647	0,297	0,618	0,234	0,472	0,216	0,466	0,148	0,272	0,138	0,268
	T_Y	0,318	0,649	0,298	0,618	0,231	0,469	0,217	0,463	0,149	0,276	0,142	0,270
	T_{NV}	0,312	0,644	0,295	0,615	0,233	0,471	0,216	0,465	0,149	0,273	0,139	0,267
	WB	0,310	0,642	0,293	0,615	0,218	0,444	0,205	0,446	0,138	0,257	0,129	0,252
	WPB	0,330	0,650	0,308	0,589	0,244	0,455	0,227	0,475	0,156	0,276	0,130	0,283

Table 5: The estimated powers of all tests under the multivariate normal distribution for $p = 3$. The estimated powers of the tests exceeding 6% in terms of the estimated type I error rates were indicated by '*' in the tables.

		Setting1		Setting2		Setting3		Setting4		Setting5		Setting6	
n	Tests	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75
(7, 7)	CAT	0,088	0,177	0,096	0,182	0,082	0,132	0,077	0,120	0,066	0,095	0,072	0,090
	B-CAT	0,078	0,125	0,089	0,139	0,075	0,111	0,069	0,110	*	*	0,062	0,073
	M-CAT	0,091	0,180	0,098	0,184	0,082	0,135	0,076	0,122	0,066	0,095	0,071	0,091
	J-CAT	0,089	0,179	0,097	0,183	0,082	0,134	0,075	0,122	0,066	0,095	0,072	0,090
	Y-CAT	0,089	0,178	0,099	0,182	0,084	0,138	0,081	0,126	0,069	0,098	0,076	0,094
	NV_CAT	0,088	0,185	0,112	0,187	0,081	0,114	0,084	0,120	0,071	0,093	0,073	0,098
	T_B	0,079	0,130	0,091	0,137	0,077	0,113	0,070	0,109	0,053	0,083	0,061	0,073
	T_M	0,088	0,174	0,096	0,183	0,084	0,136	0,076	0,121	0,066	0,093	0,073	0,092
	T_J	0,109	0,201	0,111	0,205	0,102	0,162	0,095	0,145	*	*	*	*
	T_Y	0,104	0,194	0,112	0,200	0,106	0,169	0,100	0,155	*	*	*	*
	T_{NV}	0,089	0,176	0,097	0,177	0,075	0,120	0,069	0,113	0,051	0,071	0,061	0,072
	WB	0,110	0,209	0,115	0,214	*	*	*	*	*	*	*	*
	WPB	0,084	0,180	0,108	0,182	0,081	0,120	0,089	0,127	0,082	0,103	*	*
(15, 15)	CAT	0,213	0,478	0,229	0,468	0,179	0,294	0,155	0,314	0,107	0,168	0,090	0,170
	B-CAT	0,185	0,420	0,214	0,410	0,165	0,281	0,149	0,282	*	*	0,090	0,154
	M-CAT	0,213	0,481	0,228	0,471	0,178	0,293	0,154	0,315	0,106	0,167	0,091	0,167
	J-CAT	0,212	0,480	0,228	0,469	0,179	0,293	0,154	0,315	0,107	0,167	0,091	0,167
	Y-CAT	0,211	0,481	0,229	0,472	0,180	0,297	0,157	0,317	0,106	0,170	0,090	0,173
	NV_CAT	0,227	0,493	0,220	0,471	0,163	0,302	0,153	0,320	0,100	0,167	0,097	0,181
	T_B	0,185	0,418	0,217	0,410	0,167	0,280	0,146	0,284	0,099	0,157	0,088	0,158
	T_M	0,211	0,479	0,229	0,467	0,179	0,298	0,153	0,318	0,106	0,167	0,090	0,169
	T_J	0,214	0,480	0,229	0,469	0,182	0,302	0,155	0,325	0,109	0,174	0,096	0,177
	T_Y	0,214	0,483	0,230	0,470	0,193	0,321	0,168	0,335	*	*	0,102	0,191
	T_{NV}	0,210	0,478	0,228	0,467	0,179	0,300	0,154	0,321	0,102	0,163	0,088	0,164
	WB	0,214	0,485	0,232	0,475	*	*	*	*	*	*	*	*
	WPB	0,225	0,490	0,218	0,475	0,163	0,306	0,154	0,321	0,100	0,171	0,099	0,186
(30, 30)	CAT	0,460	0,829	0,472	0,846	0,297	0,620	0,298	0,618	0,164	0,336	0,166	0,337
	B-CAT	0,438	0,796	0,440	0,808	0,280	0,592	0,285	0,594	0,164	0,329	0,162	0,333
	M-CAT	0,460	0,830	0,472	0,846	0,296	0,619	0,298	0,619	0,164	0,336	0,166	0,335
	J-CAT	0,460	0,830	0,472	0,846	0,297	0,619	0,297	0,619	0,164	0,335	0,166	0,336
	Y-CAT	0,461	0,830	0,472	0,846	0,297	0,623	0,299	0,620	0,165	0,337	0,166	0,338
	NV_CAT	0,456	0,827	0,451	0,844	0,296	0,627	0,297	0,631	0,167	0,333	0,163	0,331
	T_B	0,436	0,794	0,440	0,806	0,282	0,594	0,286	0,596	0,164	0,326	0,160	0,335
	T_M	0,462	0,828	0,475	0,846	0,297	0,618	0,302	0,618	0,165	0,337	0,166	0,338
	T_J	0,462	0,827	0,475	0,845	0,296	0,617	0,301	0,616	0,165	0,337	0,166	0,339
	T_Y	0,463	0,828	0,473	0,845	0,308	0,629	0,310	0,627	0,171	0,347	0,172	0,347
	T_{NV}	0,462	0,828	0,476	0,845	0,300	0,618	0,304	0,619	0,164	0,338	0,165	0,339
	WB	0,458	0,830	0,474	0,847	0,319	0,637	*	*	0,192	0,380	*	*
	WPB	0,457	0,825	0,448	0,842	0,300	0,627	0,295	0,626	0,168	0,334	0,159	0,328
(7, 15)	CAT	0,140	0,272	0,154	0,275	0,122	0,214	0,121	0,211	0,089	0,137	0,084	0,134
	B-CAT	0,100	0,164	0,110	0,168	0,086	0,139	0,092	0,134	0,071	0,093	0,062	0,097
	M-CAT	0,138	0,269	0,151	0,270	0,124	0,216	0,122	0,213	0,090	0,138	0,084	0,136
	J-CAT	0,138	0,270	0,151	0,271	0,124	0,215	0,121	0,213	0,089	0,138	0,084	0,135
	Y-CAT	0,141	0,268	0,157	0,274	0,123	0,218	0,119	0,212	0,091	0,138	0,085	0,137
	NV-CAT	0,160	0,289	0,142	0,292	0,123	0,226	0,116	0,207	0,080	0,142	0,085	0,135
	T_B	0,101	0,165	0,111	0,168	0,089	0,142	0,092	0,136	0,073	0,095	0,062	0,099
	T_M	0,136	0,267	0,151	0,276	0,125	0,215	0,120	0,212	0,092	0,141	0,083	0,137
	T_J	0,148	0,289	0,163	0,296	0,128	0,219	0,125	0,221	0,096	0,144	0,087	0,141
	T_Y	*	*	*	*	0,124	0,214	0,121	0,213	0,097	0,146	0,092	0,142
	T_{NV}	0,119	0,239	0,132	0,245	0,121	0,205	0,119	0,201	0,093	0,139	0,083	0,137
	WB	0,131	0,260	0,138	0,259	0,089	0,159	0,090	0,152	0,063	0,106	0,059	0,104
	WPB	*	*	0,148	0,298	0,123	0,226	0,113	0,205	0,083	0,141	0,088	0,134
(15, 30)	CAT	0,311	0,583	0,291	0,618	0,234	0,496	0,239	0,498	0,132	0,294	0,148	0,282
	B-CAT	0,278	0,509	0,243	0,537	0,196	0,405	0,213	0,418	0,120	0,247	0,125	0,252
	M-CAT	0,310	0,583	0,292	0,617	0,236	0,496	0,240	0,499	0,132	0,294	0,148	0,282
	J-CAT	0,310	0,583	0,291	0,618	0,236	0,496	0,240	0,499	0,133	0,294	0,148	0,282
	Y-CAT	0,310	0,578	0,287	0,618	0,235	0,494	0,239	0,498	0,133	0,295	0,148	0,286
	NV_CAT	0,310	0,614	0,305	0,618	0,250	0,526	0,227	0,495	0,161	0,279	0,148	0,287
	T_B	0,279	0,512	0,239	0,532	0,198	0,409	0,208	0,416	0,119	0,253	0,125	0,252
	T_M	0,312	0,588	0,291	0,616	0,237	0,498	0,238	0,500	0,134	0,299	0,149	0,289
	T_J	0,315	0,588	0,292	0,618	0,237	0,496	0,237	0,499	0,133	0,298	0,149	0,289
	T_Y	0,320	0,588	0,296	0,623	0,237	0,494	0,238	0,495	0,138	0,303	0,150	0,295
	T_{NV}	0,308	0,579	0,281	0,611	0,236	0,492	0,237	0,495	0,134	0,298	0,149	0,290
	WB	0,305	0,582	0,280	0,611	0,219	0,458	0,213	0,465	0,119	0,273	0,133	0,263
	WPB	0,317	0,617	0,305	0,616	0,240	0,522	0,226	0,491	0,159	0,282	0,149	0,293

Table 6: The estimated powers of all tests under the multivariate normal distribution for $p = 4$. The estimated powers of the tests exceeding 6% in terms of the estimated type I error rates were indicated by '*' in the tables.

		Setting1		Setting2		Setting3		Setting4		Setting5		Setting6	
n	**Tests**	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75	Δ=.5	Δ=.75
	CAT	0,068	0,133	0,079	0,165	0,068	0,113	0,072	0,106	0,083	0,080	0,065	0,097
	B-CAT	0,077	0,104	0,084	0,109	0,068	0,092	0,059	0,087	0,062	0,069	0,052	0,080
	M-CAT	0,073	0,136	0,081	0,169	0,067	0,114	0,073	0,105	0,078	0,077	0,061	0,093
	J-CAT	0,071	0,136	0,080	0,166	0,067	0,113	0,072	0,106	0,080	0,078	0,064	0,095
	Y-CAT	0,069	0,136	0,079	0,165	0,075	0,117	0,074	0,107	*	*	0,068	0,102
	NV_CAT	0,094	0,165	0,102	0,178	0,073	0,139	0,080	0,124	*	*	0,070	0,085
(7, 7)	T_B	0,076	0,106	0,083	0,109	0,066	0,094	0,062	0,086	0,060	0,067	0,058	0,080
	T_M	0,068	0,136	0,080	0,161	0,067	0,109	0,069	0,102	0,080	0,078	0,065	0,090
	T_J	0,105	0,179	*	*	*	*	*	*	*	*	*	*
	T_Y	0,091	0,158	0,102	0,198	*	*	*	*	*	*	*	*
	T_{NV}	0,064	0,127	0,077	0,157	0,052	0,079	0,056	0,082	0,040	0,033	0,033	0,050
	WB	0,097	0,168	0,105	0,203	*	*	*	*	*	*	*	*
	WPB	0,082	0,151	0,093	0,160	0,087	0,124	0,084	0,135	*	*	*	*
	CAT	0,211	0,425	0,233	0,474	0,157	0,300	0,158	0,305	0,099	0,157	0,097	0,172
	B-CAT	0,168	0,364	0,201	0,402	0,135	0,261	0,147	0,266	0,097	0,149	0,093	0,165
	M-CAT	0,212	0,426	0,234	0,479	0,155	0,299	0,157	0,305	0,098	0,153	0,097	0,171
	J-CAT	0,212	0,426	0,234	0,478	0,157	0,299	0,157	0,305	0,099	0,155	0,097	0,172
	Y-CAT	0,213	0,428	0,233	0,476	0,160	0,304	0,161	0,310	0,101	0,157	0,101	0,176
	NV_CAT	0,191	0,416	0,220	0,495	0,141	0,295	0,168	0,322	0,102	0,157	0,100	0,162
(15, 15)	T_B	0,172	0,366	0,206	0,407	0,133	0,265	0,149	0,270	0,096	0,151	0,093	0,168
	T_M	0,213	0,427	0,235	0,476	0,156	0,300	0,158	0,305	0,100	0,154	0,098	0,172
	T_J	0,219	0,431	0,239	0,484	0,162	0,313	0,168	0,313	0,111	0,167	*	*
	T_Y	0,218	0,435	0,238	0,480	0,176	0,330	0,180	0,335	0,124	0,190	*	*
	T_{NV}	0,211	0,424	0,232	0,475	0,153	0,301	0,160	0,307	0,096	0,149	0,094	0,166
	WB	0,215	0,434	0,241	0,486	*	*	*	*	*	*	*	*
	WPB	0,209	0,420	0,224	0,474	0,150	0,317	0,163	0,293	0,124	0,165	0,105	0,175
	CAT	0,414	0,807	0,452	0,846	0,309	0,637	0,313	0,665	0,167	0,343	0,173	0,365
	B-CAT	0,390	0,771	0,426	0,815	0,290	0,605	0,290	0,631	0,171	0,327	0,169	0,356
	M-CAT	0,416	0,807	0,452	0,847	0,308	0,637	0,312	0,665	0,167	0,343	0,172	0,366
	J-CAT	0,416	0,807	0,452	0,846	0,309	0,637	0,312	0,665	0,168	0,343	0,172	0,365
	Y-CAT	0,413	0,807	0,453	0,847	0,311	0,640	0,315	0,668	0,171	0,345	0,175	0,370
	NV_CAT	0,454	0,819	0,476	0,846	0,310	0,626	0,316	0,676	0,172	0,324	0,177	0,372
(30, 30)	T_B	0,389	0,772	0,426	0,817	0,290	0,607	0,290	0,634	0,170	0,329	0,169	0,358
	T_M	0,415	0,807	0,454	0,850	0,305	0,640	0,307	0,663	0,167	0,342	0,174	0,367
	T_J	0,414	0,807	0,453	0,850	0,305	0,640	0,308	0,663	0,170	0,344	0,176	0,370
	T_Y	0,417	0,810	0,456	0,852	0,323	0,657	0,322	0,674	0,183	0,363	0,185	0,388
	T_{NV}	0,415	0,808	0,454	0,850	0,311	0,643	0,312	0,667	0,168	0,342	0,172	0,367
	WB	0,415	0,808	0,455	0,850	0,336	0,670	*	*	*	*	*	*
	WPB	0,423	0,825	0,466	0,862	0,301	0,623	0,290	0,647	0,177	0,346	0,170	0,381
	CAT	0,135	0,251	0,132	0,254	0,111	0,207	0,116	0,219	0,087	0,133	0,081	0,137
	B-CAT	0,093	0,134	0,083	0,142	0,075	0,109	0,084	0,109	0,067	0,086	0,056	0,095
	M-CAT	0,130	0,243	0,130	0,246	0,112	0,209	0,118	0,221	0,087	0,135	0,082	0,137
	J-CAT	0,132	0,246	0,131	0,250	0,113	0,209	0,116	0,221	0,087	0,134	0,082	0,137
	Y-CAT	0,141	0,248	0,136	0,255	0,112	0,205	0,117	0,215	0,089	0,136	0,084	0,141
	NV_CAT	0,131	0,257	0,146	0,290	0,113	0,224	0,114	0,239	0,082	0,130	0,080	0,140
(7, 15)	T_B	0,094	0,135	0,084	0,142	0,078	0,113	0,083	0,110	0,067	0,084	0,051	0,099
	T_M	0,128	0,239	0,130	0,246	0,110	0,209	0,115	0,219	0,089	0,136	0,081	0,139
	T_J	0,156	0,276	*	*	0,125	0,225	0,128	0,235	0,100	0,143	0,090	0,150
	T_Y	0,176	0,300	*	*	0,119	0,212	0,125	0,219	0,098	0,145	0,091	0,150
	T_{NV}	0,100	0,188	0,091	0,195	0,103	0,193	0,111	0,205	0,087	0,133	0,081	0,137
	WB	0,121	0,233	0,122	0,234	0,075	0,148	0,073	0,150	0,065	0,100	0,055	0,103
	WPB	0,143	0,275	*	*	0,103	0,204	0,112	0,239	0,088	0,134	0,089	0,147
	CAT	0,268	0,565	0,297	0,633	0,229	0,491	0,237	0,522	0,147	0,304	0,135	0,298
	B-CAT	0,213	0,452	0,243	0,514	0,183	0,379	0,195	0,412	0,121	0,242	0,109	0,228
	M-CAT	0,265	0,561	0,294	0,628	0,229	0,492	0,238	0,526	0,147	0,305	0,135	0,299
	J-CAT	0,266	0,561	0,295	0,630	0,229	0,492	0,238	0,525	0,147	0,305	0,135	0,298
	Y-CAT	0,265	0,558	0,294	0,627	0,227	0,487	0,236	0,520	0,151	0,310	0,137	0,303
(15, 30)	NV_CAT	0,280	0,602	0,298	0,615	0,228	0,491	0,232	0,499	0,139	0,316	0,154	0,320
	T_B	0,213	0,456	0,240	0,517	0,180	0,379	0,199	0,409	0,122	0,241	0,110	0,229
	T_M	0,265	0,561	0,301	0,622	0,229	0,491	0,242	0,527	0,149	0,302	0,139	0,298
	T_J	0,268	0,567	0,303	0,630	0,230	0,491	0,242	0,528	0,149	0,305	0,140	0,299
	T_Y	0,279	0,576	0,308	0,636	0,227	0,483	0,241	0,525	0,157	0,317	0,142	0,309
	T_{NV}	0,253	0,545	0,287	0,612	0,223	0,484	0,235	0,522	0,149	0,303	0,139	0,299
	WB	0,257	0,554	0,292	0,624	0,192	0,440	0,210	0,488	0,134	0,279	0,125	0,273
	WPB	0,281	0,580	0,317	0,618	0,247	0,479	0,238	0,512	0,147	0,284	0,151	0,311

When all tests were compared in terms of their powers in Table 4–6, the T_B test and the CAT version of this test, B-CAT, has a smaller power than the other tests in most cases. In cases where the estimated type I error rates of the T_J, T_Y and WB tests are close the nominal level, the powers of these tests are close to the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_{M} and WPB tests, even sometimes the powers of these tests are slightly higher than those of these tests. However, as p and the degree of heterogeneity increase, and in case of small sample sizes, the T_J, T_Y and WB tests tend to be highly liberal, which is a disadvantage for them.

When the CAT, K-CAT, J-CAT, Y-CAT, NV-CAT, T_{M} and WPB tests were compared in terms of their powers, these tests have powers close to each other. Besides, while the CAT has a bit higher power than the other tests in some cases, the NV-CAT has a bit higher power than the other tests in some cases. Since the CAT has simple form than the other CAT versions of the tests, it can be preferred instead of the others where they have similar powers.

To get idea about robustness of the above tests against multivariate non-normality we conduct simulation study under multivariate non-normal models. Following Konietschke *et al.* [16], we generated data as

$$\mathbf{Y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\varepsilon}_{ij}\,, \qquad i = 1, ..., n_i; \quad j = 1, 2\,,$$

using the Cholesky decomposition $\boldsymbol{\Sigma}_i^{1/2}$ of a given covariance matrix $\boldsymbol{\Sigma}_i$. The independent and identically distributed random error vectors $\varepsilon_{ij} = (\varepsilon_{ij}^{(1)}, ..., \varepsilon_{ij}^{(p)})^\top$ were generated from different standardized symmetric or skewed distributions by

$$\varepsilon_{ij}^{(s)} = \frac{W_{ij}^{(s)} - E\big(W_{ij}^{(s)}\big)}{\sqrt{\mathrm{Var}\big(W_{ij}^{(s)}\big)}}\,.$$

Here $W_{ij}^{(s)}$ are double exponential distribution (DE), t-distribution with degrees of freedom 7 (t_7), χ^2 distribution with degrees of freedom 15 (χ^2_{15}) and χ^2 distribution with degrees of freedom 20 (χ^2_{20}). We refer to these distributions as distribution 1, 2, 3 and 4 in tables, respectively. We estimated the type-I error rates of all tests under these distributions, respectively. The simulated results were provided in Tables 7–18.

The results under the double exponential model are almost similar to those under the normal model. However, unlike normal distribution, it can be seen that the estimated type I error rates of all tests are smaller than the nominal level 0.05 especially in small sample size. Besides, while the estimated type I error rates of the WB test are close to the nominal level in cases of homogeneity structure, those of this test are higher than the nominal level as the degree of heterogeneity.

The results under t_7-model are quite similar to those under the normal model. While the results under the χ^2_{15}-model are somewhat similar to those under the normal model, it can be seen that the estimated type I error rates of tests increase significantly. Because of skewed distribution, the T_J, T_Y and WB tests tend to be highly liberal when heterogeneity is severe and sample size is small. A remarkable consequence is that as p increase, especially in cases of small sample size, the estimated type I error rates of all tests are higher than the nominal level. However, the B-CAT test has the estimated type I error rate close to the nominal level. Also, note that as the degree of freedom increase, since χ^2_{20} distribution is more close to a symmetric distribution than χ^2_{15} distribution, the results under this model are more similar to those under the normal model.

In cases of normal distributed and symmetric distributed models, it can be seen that the CAT, the CAT versions of tests, T_{M} and WPB tests have the estimated type I error rates close to the nominal level 0.05. The T_J, T_Y and WB tests tend to be highly liberal in cases of small sample size and heterogeneity. In cases of model with skewed distribution, that is, the χ^2_{15}-model, the estimated type I error rates of many tests significantly exceed the nominal level 0.05. However, the B-CAT performs well than other tests in terms of type I error rate under this model.

As Konietschke *et al.* [16] noted and our simulation study can be seen, the WB test's behavior depends on the p, degree of heterogeneity and the amount of skewness. Furthermore, T_J and T_Y tests' behavior also depend on the p, degree of heterogeneity and the amount of skewness. Thus, as seen from simulation study, CAT method can be uses as a good alternative for the equality of two multivariate normal mean vectors under heterogeneity of covariance.

Table 7: The estimated type I error rates of all tests under distribution 1 for $p = 2$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,034	0,040	0,033	0,033	0,033	0,032	0,041	0,032	0,036	0,038	0,032	**0,100**	0,044	0,040
	15, 15	0,049	0,052	0,049	0,049	0,050	0,047	0,051	0,049	0,049	0,048	0,048	**0,075**	0,053	0,051
	30, 30	0,050	0,047	0,050	0,050	0,049	0,041	0,049	0,050	0,050	0,050	0,050	**0,060**	0,051	0,043
	7, 15	0,046	0,048	0,045	0,046	0,046	0,036	0,049	0,043	0,043	0,050	0,040	**0,093**	0,046	0,046
	15, 30	0,043	0,043	0,043	0,043	0,042	0,050	0,044	0,043	0,043	0,043	0,041	**0,063**	0,045	0,051
2	7, 7	0,038	0,045	0,038	0,038	0,039	0,035	0,047	0,040	0,041	0,042	0,038	**0,105**	0,049	0,038
	15, 15	0,046	0,042	0,046	0,046	0,046	0,044	0,043	0,048	0,048	0,047	0,046	**0,075**	0,050	0,053
	30, 30	0,035	0,041	0,035	0,035	0,035	0,051	0,039	0,035	0,035	0,035	0,035	0,049	0,037	0,050
	7, 15	0,037	0,047	0,038	0,038	0,038	0,042	0,046	0,038	0,039	0,041	0,034	**0,085**	0,036	0,040
	15, 30	0,050	0,054	0,049	0,049	0,049	0,047	0,054	0,047	0,047	0,047	0,047	**0,071**	0,048	0,043
3	7, 7	0,038	0,040	0,038	0,038	0,039	0,039	0,041	0,039	0,041	0,044	0,035	**0,122**	0,059	0,040
	15, 15	0,046	0,045	0,046	0,046	0,046	0,046	0,044	0,049	0,049	0,049	0,048	**0,072**	0,054	0,042
	30, 30	0,055	0,057	0,055	0,055	0,055	0,046	0,057	0,056	0,056	0,057	0,056	**0,069**	0,060	0,051
	7, 15	0,039	0,048	0,040	0,039	0,039	0,050	0,048	0,041	0,041	0,042	0,039	**0,073**	0,035	0,045
	15, 30	0,043	0,058	0,043	0,043	0,043	0,050	0,058	0,043	0,043	0,043	0,043	**0,060**	0,041	0,052
4	7, 7	0,037	0,044	0,037	0,037	0,037	0,039	0,046	0,037	0,039	0,043	0,034	**0,107**	0,053	0,039
	15, 15	0,046	0,049	0,046	0,046	0,046	0,043	0,049	0,046	0,046	0,046	0,046	**0,076**	0,055	0,042
	30, 30	0,052	0,054	0,052	0,052	0,052	0,046	0,052	0,052	0,052	0,052	0,052	**0,063**	0,058	0,053
	7, 15	0,044	0,052	0,044	0,044	0,045	0,043	0,048	0,044	0,044	0,043	0,042	**0,071**	0,038	0,035
	15, 30	0,050	0,054	0,050	0,050	0,049	0,044	0,052	0,050	0,050	0,049	0,049	**0,070**	0,045	0,045
5	7, 7	0,040	0,040	0,044	0,040	0,040	0,048	0,042	0,042	0,047	0,048	0,036	**0,131**	**0,078**	0,040
	15, 15	0,041	0,048	0,041	0,041	0,042	0,048	0,047	0,042	0,043	0,045	0,041	**0,082**	**0,065**	0,047
	30, 30	0,038	0,060	0,038	0,038	0,038	0,045	0,061	0,039	0,040	0,040	0,040	**0,074**	0,033	0,042
	7, 15	0,040	0,052	0,040	0,040	0,040	0,045	0,052	0,039	0,039	0,040	0,039	0,058	0,037	0,044
	15, 30	0,051	0,048	0,051	0,051	0,051	0,055	0,047	0,052	0,052	0,052	0,051	**0,067**	**0,061**	0,050
6	7, 7	0,040	0,040	0,042	0,040	0,040	0,038	0,042	0,043	0,045	0,050	0,037	**0,126**	**0,077**	0,038
	15, 15	0,037	0,039	0,037	0,037	0,038	0,046	0,040	0,038	0,039	0,043	0,038	**0,079**	**0,061**	0,046
	30, 30	0,043	0,061	0,043	0,043	0,042	0,047	0,059	0,043	0,044	0,045	0,042	**0,089**	0,034	0,049
	7, 15	0,043	0,055	0,043	0,043	0,043	0,046	0,054	0,041	0,041	0,043	0,042	**0,064**	0,040	0,042
	15, 30	0,041	0,040	0,041	0,041	0,041	0,048	0,042	0,042	0,042	0,043	0,042	**0,062**	0,052	0,045

Table 8: The estimated type I error rates of all tests under distribution 1 for $p = 3$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,028	0,043	0,029	0,029	0,029	0,035	0,041	0,030	0,037	0,038	0,026	**0,140**	0,041	0,035
	15, 15	0,047	0,038	0,047	0,047	0,047	0,041	0,040	0,047	0,047	0,048	0,046	**0,087**	0,054	0,047
	30, 30	0,048	0,046	0,048	0,048	0,048	0,048	0,046	0,048	0,047	0,049	0,048	**0,061**	0,051	0,041
	7, 15	0,044	0,047	0,044	0,044	0,044	0,041	0,046	0,040	0,044	0,051	0,028	**0,128**	0,042	0,044
	15, 30	0,046	0,052	0,046	0,046	0,047	0,053	0,053	0,046	0,046	0,050	0,043	**0,080**	0,047	0,043
2	7, 7	0,038	0,037	0,041	0,040	0,040	0,032	0,038	0,039	0,049	0,050	0,036	**0,150**	0,056	0,041
	15, 15	0,042	0,042	0,043	0,043	0,044	0,048	0,043	0,044	0,044	0,044	0,043	**0,088**	0,050	0,046
	30, 30	0,042	0,044	0,043	0,043	0,043	0,041	0,044	0,042	0,041	0,043	0,041	0,058	0,044	0,048
	7, 15	0,046	0,048	0,046	0,046	0,045	0,042	0,049	0,044	0,050	0,055	0,035	**0,136**	0,047	0,044
	15, 30	0,045	0,045	0,044	0,045	0,044	0,053	0,046	0,043	0,043	0,046	0,042	**0,071**	0,045	0,044
3	7, 7	0,030	0,039	0,031	0,031	0,031	0,036	0,038	0,032	0,041	0,043	0,026	**0,172**	**0,061**	0,041
	15, 15	0,040	0,043	0,040	0,040	0,039	0,048	0,043	0,041	0,041	0,043	0,041	**0,094**	0,054	0,042
	30, 30	0,047	0,054	0,047	0,047	0,047	0,045	0,053	0,048	0,048	0,049	0,048	**0,069**	0,055	0,054
	7, 15	0,042	0,057	0,042	0,042	0,043	0,040	0,056	0,042	0,046	0,048	0,035	**0,110**	0,026	0,035
	15, 30	0,047	0,059	0,048	0,048	0,048	0,052	0,059	0,047	0,046	0,046	0,046	**0,073**	0,042	0,051
4	7, 7	0,038	0,043	0,037	0,037	0,039	0,039	0,044	0,040	0,051	0,058	0,033	**0,180**	**0,070**	0,046
	15, 15	0,040	0,050	0,040	0,040	0,041	0,045	0,047	0,039	0,040	0,043	0,037	**0,101**	0,055	0,043
	30, 30	0,047	0,046	0,047	0,047	0,047	0,049	0,045	0,048	0,048	0,048	0,048	**0,074**	0,053	0,047
	7, 15	0,035	0,050	0,037	0,036	0,038	0,050	0,050	0,036	0,038	0,039	0,033	**0,096**	0,025	0,038
	15, 30	0,048	0,050	0,048	0,048	0,050	0,045	0,051	0,048	0,048	0,048	0,047	**0,069**	0,042	0,040
5	7, 7	0,033	0,031	0,031	0,032	0,033	0,046	0,035	0,033	0,052	**0,060**	0,021	**0,198**	**0,096**	0,049
	15, 15	0,049	0,047	0,048	0,048	0,049	0,045	0,047	0,049	0,052	0,054	0,044	**0,113**	**0,078**	0,048
	30, 30	0,045	0,045	0,046	0,045	0,045	0,046	0,046	0,046	0,046	0,048	0,046	**0,071**	0,056	0,050
	7, 15	0,045	0,064	0,045	0,046	0,044	0,046	0,062	0,044	0,047	0,047	0,044	**0,107**	0,031	0,045
	15, 30	0,049	0,062	0,050	0,050	0,049	0,042	0,062	0,045	0,045	0,047	0,045	**0,078**	0,044	0,042
6	7, 7	0,040	0,043	0,038	0,039	0,039	0,035	0,041	0,040	0,055	**0,063**	0,027	**0,207**	**0,100**	0,051
	15, 15	0,050	0,048	0,050	0,050	0,051	0,047	0,046	0,048	0,051	0,055	0,047	**0,115**	**0,082**	0,047
	30, 30	0,055	0,050	0,055	0,055	0,054	0,048	0,050	0,054	0,054	0,055	0,054	**0,084**	**0,069**	0,048
	7, 15	0,047	0,059	0,048	0,048	0,047	0,048	0,059	0,048	0,050	0,052	0,048	**0,119**	0,037	0,039
	15, 30	0,056	0,058	0,056	0,056	0,056	0,044	0,058	0,057	0,057	0,055	0,055	**0,087**	0,049	0,056

Table 9: The estimated type I error rates of all tests under distribution 1 for $p = 4$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,033	0,047	0,034	0,033	0,034	0,046	0,048	0,032	0,053	0,047	0,025	**0,220**	0,054	0,036
	15, 15	0,040	0,045	0,040	0,040	0,039	0,041	0,044	0,040	0,042	0,043	0,039	**0,107**	0,047	0,047
	30, 30	0,045	0,046	0,045	0,045	0,046	0,046	0,046	0,046	0,046	0,047	0,046	**0,065**	0,049	0,043
	7, 15	0,042	0,053	0,040	0,041	0,043	0,045	0,051	0,040	0,052	**0,062**	0,026	**0,185**	0,045	0,051
	15, 30	0,044	0,053	0,043	0,043	0,044	0,042	0,052	0,042	0,044	0,049	0,038	**0,098**	0,041	0,047
2	7, 7	0,033	0,037	0,035	0,034	0,037	0,031	0,039	0,033	0,049	0,045	0,031	**0,209**	0,049	0,031
	15, 15	0,048	0,049	0,049	0,048	0,047	0,044	0,048	0,046	0,048	0,047	0,045	**0,111**	0,055	0,052
	30, 30	0,046	0,050	0,047	0,046	0,046	0,054	0,050	0,045	0,045	0,046	0,046	**0,077**	0,052	0,052
	7, 15	0,043	0,046	0,042	0,043	0,047	0,043	0,044	0,041	0,054	**0,065**	0,026	**0,172**	0,037	0,043
	15, 30	0,042	0,051	0,042	0,042	0,041	0,044	0,050	0,041	0,043	0,046	0,040	**0,094**	0,043	0,041
3	7, 7	0,035	0,051	0,035	0,034	0,035	0,043	0,049	0,036	0,057	**0,060**	0,025	**0,261**	**0,075**	0,041
	15, 15	0,046	0,041	0,046	0,046	0,047	0,049	0,043	0,047	0,051	0,055	0,047	**0,130**	**0,069**	0,039
	30, 30	0,043	0,042	0,042	0,042	0,045	0,056	0,041	0,043	0,043	0,044	0,043	**0,070**	0,051	0,050
	7, 15	0,034	0,051	0,036	0,036	0,035	0,039	0,050	0,034	0,041	0,041	0,030	**0,133**	0,020	0,047
	15, 30	0,045	0,051	0,046	0,045	0,044	0,039	0,051	0,044	0,045	0,046	0,044	**0,079**	0,038	0,040
4	7, 7	0,033	0,050	0,033	0,033	0,036	0,046	0,050	0,033	0,057	0,059	0,027	**0,262**	**0,069**	0,047
	15, 15	0,051	0,049	0,050	0,050	0,052	0,047	0,049	0,049	0,053	0,056	0,050	**0,132**	**0,072**	0,043
	30, 30	0,044	0,052	0,044	0,044	0,043	0,051	0,051	0,045	0,045	0,048	0,047	**0,082**	**0,060**	0,050
	7, 15	0,035	0,050	0,036	0,036	0,036	0,041	0,048	0,036	0,041	0,041	0,030	**0,137**	0,025	0,041
	15, 30	0,043	0,053	0,044	0,044	0,045	0,046	0,054	0,045	0,046	0,047	0,043	**0,079**	0,037	0,049
5	7, 7	0,041	0,041	0,038	0,040	0,041	0,045	0,040	0,038	**0,077**	**0,086**	0,017	**0,309**	**0,122**	0,053
	15, 15	0,040	0,041	0,040	0,040	0,039	0,048	0,040	0,041	0,047	0,051	0,036	**0,148**	**0,086**	0,044
	30, 30	0,043	0,048	0,041	0,042	0,043	0,050	0,048	0,045	0,045	0,048	0,044	**0,087**	**0,067**	0,044
	7, 15	0,040	0,056	0,041	0,040	0,041	0,042	0,057	0,040	0,045	0,044	0,039	**0,128**	0,026	0,042
	15, 30	0,047	**0,060**	0,046	0,047	0,048	0,046	0,058	0,045	0,046	0,049	0,045	**0,084**	0,042	0,042
6	7, 7	0,048	0,051	0,044	0,047	0,048	0,048	0,050	0,044	**0,081**	**0,090**	0,019	**0,324**	**0,128**	0,054
	15, 15	0,045	0,043	0,045	0,044	0,046	0,042	0,042	0,046	0,051	0,056	0,040	**0,135**	**0,084**	0,046
	30, 30	0,045	0,045	0,044	0,044	0,045	0,049	0,043	0,046	0,046	0,049	0,045	**0,088**	**0,069**	0,045
	7, 15	0,036	**0,064**	0,037	0,036	0,038	0,040	**0,065**	0,037	0,047	0,045	0,035	**0,144**	0,026	0,039
	15, 30	0,052	**0,062**	0,052	0,052	0,052	0,051	**0,064**	0,050	0,050	0,052	0,049	**0,094**	0,045	0,051

Table 10: The estimated type I error rates of all tests under distribution 2 for $p = 2$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,046	0,055	0,047	0,046	0,046	0,048	0,054	0,047	0,049	0,048	0,048	**0,103**	0,056	0,044
	15, 15	0,047	0,045	0,047	0,047	0,047	0,051	0,043	0,047	0,047	0,048	0,047	**0,067**	0,048	0,050
	30, 30	0,042	0,042	0,042	0,042	0,042	0,053	0,043	0,044	0,043	0,043	0,044	0,056	0,045	0,041
	7, 15	0,052	0,052	0,052	0,052	0,052	0,041	0,051	0,050	0,052	**0,061**	0,048	**0,108**	0,053	0,057
	15, 30	0,045	0,040	0,045	0,045	0,045	0,044	0,041	0,045	0,045	0,045	0,044	**0,065**	0,043	0,049
2	7, 7	0,041	0,051	0,042	0,042	0,041	0,045	0,049	0,042	0,045	0,045	0,042	**0,095**	0,051	0,046
	15, 15	0,043	0,041	0,044	0,044	0,043	0,052	0,042	0,044	0,044	0,044	0,043	**0,068**	0,046	0,046
	30, 30	0,056	0,056	0,056	0,056	0,056	0,047	0,055	0,055	0,055	0,055	0,055	**0,070**	0,055	0,054
	7, 15	0,044	0,049	0,043	0,043	0,042	0,042	0,048	0,045	0,047	0,048	0,040	**0,094**	0,041	0,051
	15, 30	0,042	0,054	0,042	0,042	0,042	0,049	0,055	0,043	0,043	0,044	0,043	**0,071**	0,040	0,055
3	7, 7	0,039	0,040	0,040	0,040	0,041	0,051	0,039	0,038	0,042	0,043	0,034	**0,104**	0,054	0,046
	15, 15	0,050	0,052	0,050	0,050	0,050	0,043	0,050	0,050	0,050	0,050	0,050	**0,085**	0,059	0,050
	30, 30	0,050	0,046	0,051	0,051	0,050	0,047	0,045	0,050	0,050	0,050	0,050	**0,061**	0,051	0,050
	7, 15	0,045	0,047	0,045	0,045	0,045	0,045	0,049	0,043	0,044	0,045	0,043	**0,084**	0,032	0,047
	15, 30	0,054	0,053	0,054	0,054	0,054	0,049	0,052	0,052	0,052	0,053	0,052	**0,068**	0,048	0,051
4	7, 7	0,050	0,049	0,050	0,050	0,050	0,047	0,048	0,054	0,058	**0,064**	0,050	**0,119**	**0,072**	0,044
	15, 15	0,056	0,053	0,056	0,056	0,056	0,051	0,055	0,057	0,057	0,057	0,058	**0,086**	**0,064**	0,044
	30, 30	0,047	0,048	0,047	0,047	0,047	0,057	0,046	0,047	0,047	0,048	0,048	**0,062**	0,053	0,053
	7, 15	0,045	0,051	0,046	0,046	0,046	0,054	0,052	0,048	0,048	0,047	0,045	**0,083**	0,036	0,046
	15, 30	0,049	0,051	0,050	0,050	0,050	0,047	0,052	0,050	0,050	0,050	0,050	**0,064**	0,044	0,048
5	7, 7	0,048	0,046	0,049	0,049	0,049	0,044	0,046	0,047	0,053	0,059	0,044	**0,144**	**0,083**	0,051
	15, 15	0,050	0,050	0,051	0,051	0,052	0,048	0,049	0,052	0,052	0,053	0,050	**0,081**	**0,064**	0,054
	30, 30	0,050	0,050	0,050	0,050	0,050	0,044	0,050	0,050	0,050	0,050	0,050	**0,064**	0,057	0,049
	7, 15	0,043	0,051	0,044	0,043	0,044	0,047	0,052	0,045	0,045	0,046	0,044	**0,081**	0,036	0,040
	15, 30	0,049	0,050	0,049	0,049	0,048	0,047	0,050	0,048	0,048	0,049	0,049	**0,070**	0,045	0,047
6	7, 7	0,047	0,051	0,047	0,047	0,047	0,044	0,051	0,049	0,057	**0,062**	0,047	**0,143**	**0,087**	0,051
	15, 15	0,044	0,044	0,043	0,043	0,044	0,046	0,042	0,043	0,044	0,046	0,042	**0,076**	**0,061**	0,044
	30, 30	0,045	0,042	0,045	0,045	0,045	0,049	0,041	0,044	0,044	0,044	0,045	**0,064**	0,055	0,046
	7, 15	0,045	0,052	0,045	0,045	0,045	0,043	0,051	0,043	0,044	0,044	0,043	**0,081**	0,034	0,048
	15, 30	0,049	0,056	0,049	0,049	0,049	0,048	0,057	0,050	0,050	0,050	0,050	**0,066**	0,046	0,051

Table 11: The estimated type I error rates of all tests under distribution 2 for $p = 3$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,041	0,048	0,042	0,041	0,044	0,040	0,049	0,040	0,048	0,048	0,038	**0,155**	0,053	0,049
	15, 15	0,049	0,050	0,049	0,049	0,048	0,047	0,049	0,049	0,049	0,050	0,048	**0,088**	0,051	0,042
	30, 30	0,052	0,053	0,053	0,052	0,053	0,052	0,053	0,050	0,050	0,050	0,050	**0,070**	0,052	0,050
	7, 15	0,052	0,054	0,051	0,052	0,052	0,058	0,052	0,052	0,058	**0,065**	0,038	**0,137**	0,050	0,046
	15, 30	0,047	0,045	0,046	0,046	0,048	0,050	0,046	0,046	0,047	0,050	0,046	**0,085**	0,047	0,047
2	7, 7	0,036	0,049	0,036	0,036	0,034	0,038	0,050	0,034	0,041	0,040	0,031	**0,131**	0,042	0,036
	15, 15	0,049	0,050	0,050	0,050	0,049	0,050	0,050	0,049	0,049	0,049	0,049	**0,089**	0,053	0,040
	30, 30	0,055	0,058	0,055	0,055	0,054	0,047	0,058	0,056	0,056	0,055	0,056	**0,076**	0,055	0,041
	7, 15	0,053	0,048	0,052	0,052	0,055	0,056	0,050	0,052	0,059	**0,068**	0,043	**0,141**	0,051	0,056
	15, 30	0,058	0,055	0,058	0,058	0,058	0,049	0,055	0,057	0,058	**0,062**	0,053	**0,096**	0,056	0,054
3	7, 7	0,043	0,048	0,043	0,043	0,043	0,040	0,045	0,044	0,055	0,056	0,037	**0,172**	**0,071**	0,047
	15, 15	0,052	0,051	0,053	0,053	0,052	0,055	0,053	0,052	0,053	0,057	0,052	**0,101**	**0,065**	0,048
	30, 30	0,048	0,049	0,048	0,048	0,047	0,050	0,049	0,048	0,048	0,049	0,049	**0,072**	0,056	0,053
	7, 15	0,044	0,055	0,045	0,045	0,046	0,044	0,055	0,043	0,045	0,048	0,039	**0,102**	0,031	0,038
	15, 30	0,047	0,059	0,047	0,047	0,047	0,059	0,060	0,045	0,045	0,046	0,044	**0,078**	0,041	0,045
4	7, 7	0,041	0,042	0,042	0,042	0,041	0,046	0,041	0,041	0,054	0,058	0,032	**0,185**	**0,073**	0,046
	15, 15	0,049	0,046	0,049	0,049	0,050	0,041	0,046	0,054	0,056	0,058	0,053	**0,106**	**0,066**	0,048
	30, 30	0,043	0,047	0,043	0,043	0,043	0,049	0,048	0,043	0,043	0,045	0,044	**0,069**	0,050	0,046
	7, 15	0,048	0,050	0,048	0,048	0,047	0,039	0,050	0,048	0,051	0,050	0,044	**0,106**	0,031	0,038
	15, 30	0,045	0,050	0,045	0,045	0,044	0,044	0,049	0,042	0,042	0,043	0,041	**0,070**	0,036	0,051
5	7, 7	0,052	0,046	0,051	0,051	0,053	0,046	0,046	0,053	**0,071**	**0,081**	0,038	**0,234**	**0,120**	0,054
	15, 15	0,042	0,049	0,042	0,042	0,043	0,047	0,048	0,044	0,046	0,048	0,041	**0,108**	**0,072**	0,049
	30, 30	0,055	0,047	0,055	0,055	0,055	0,041	0,048	0,054	0,054	0,057	0,054	**0,083**	**0,068**	0,050
	7, 15	0,038	0,050	0,039	0,038	0,038	0,046	0,052	0,038	0,041	0,042	0,038	**0,096**	0,027	0,038
	15, 30	0,049	0,049	0,049	0,049	0,049	0,056	0,050	0,046	0,046	0,047	0,046	**0,071**	0,044	0,041
6	7, 7	0,050	0,042	0,048	0,049	0,048	0,051	0,043	0,049	**0,062**	**0,068**	0,035	**0,199**	**0,097**	0,047
	15, 15	0,053	0,052	0,052	0,052	0,054	0,053	0,052	0,053	0,056	0,058	0,050	**0,122**	**0,084**	0,049
	30, 30	0,057	0,058	0,057	0,057	0,057	0,052	0,056	0,055	0,055	0,059	0,055	**0,084**	**0,072**	0,051
	7, 15	0,047	0,049	0,048	0,048	0,047	0,042	0,049	0,048	0,049	0,049	0,047	**0,104**	0,036	0,040
	15, 30	0,056	0,056	0,056	0,056	0,055	0,048	0,057	0,054	0,054	0,056	0,054	**0,090**	0,042	0,047

Table 12: The estimated type I error rates of all tests under distribution 2 for $p = 4$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,034	0,055	0,035	0,035	0,035	0,034	0,056	0,032	0,051	0,046	0,029	**0,201**	0,048	0,037
	15, 15	0,048	0,054	0,049	0,049	0,048	0,045	0,052	0,048	0,051	0,051	0,046	**0,113**	0,053	0,047
	30, 30	0,043	0,045	0,043	0,043	0,042	0,048	0,046	0,044	0,044	0,044	0,044	**0,067**	0,044	0,048
	7, 15	0,050	0,045	0,049	0,050	0,055	0,046	0,044	0,048	**0,062**	**0,073**	0,029	**0,188**	0,043	0,059
	15, 30	0,048	0,046	0,048	0,048	0,049	0,049	0,047	0,049	0,049	0,051	0,046	**0,102**	0,046	0,048
2	7, 7	0,043	0,044	0,045	0,044	0,045	0,048	0,041	0,042	0,061	0,057	0,041	**0,207**	0,057	0,041
	15, 15	0,049	0,055	0,050	0,050	0,051	0,043	0,057	0,050	0,053	0,056	0,049	**0,113**	0,055	0,048
	30, 30	0,052	0,048	0,052	0,052	0,053	0,046	0,047	0,051	0,050	0,052	0,051	**0,079**	0,054	0,052
	7, 15	0,058	0,059	0,059	0,058	0,060	0,055	0,059	0,057	**0,070**	**0,086**	0,037	**0,200**	0,055	0,054
	15, 30	0,051	0,055	0,051	0,051	0,050	0,051	0,056	0,049	0,050	0,057	0,044	**0,112**	0,048	0,049
3	7, 7	0,040	0,049	0,040	0,040	0,043	0,040	0,050	0,040	**0,068**	**0,070**	0,032	**0,258**	**0,076**	0,049
	15, 15	0,051	0,044	0,051	0,051	0,051	0,054	0,044	0,051	0,052	0,056	0,051	**0,116**	**0,067**	0,046
	30, 30	0,047	0,045	0,047	0,047	0,046	0,049	0,043	0,048	0,048	0,050	0,048	**0,078**	0,055	0,048
	7, 15	0,045	0,052	0,045	0,045	0,045	0,040	0,055	0,044	0,047	0,048	0,040	**0,140**	0,030	0,031
	15, 30	0,046	0,055	0,046	0,046	0,046	0,053	0,055	0,046	0,046	0,045	0,043	**0,082**	0,034	0,046
4	7, 7	0,033	0,044	0,034	0,034	0,036	0,051	0,045	0,033	0,053	0,056	0,025	**0,260**	**0,069**	0,048
	15, 15	0,053	0,052	0,053	0,052	0,054	0,045	0,051	0,052	0,055	0,059	0,052	**0,132**	**0,074**	0,047
	30, 30	0,051	0,052	0,051	0,051	0,052	0,055	0,053	0,052	0,052	0,053	0,053	**0,083**	0,058	0,048
	7, 15	0,044	0,048	0,045	0,045	0,047	0,040	0,047	0,045	0,052	0,050	0,039	**0,143**	0,025	0,040
	15, 30	0,042	0,052	0,042	0,042	0,042	0,052	0,053	0,044	0,044	0,044	0,043	**0,081**	0,034	0,046
5	7, 7	0,051	0,042	0,049	0,050	0,051	0,052	0,044	0,046	**0,098**	**0,108**	0,027	**0,332**	**0,140**	**0,063**
	15, 15	0,047	0,045	0,047	0,047	0,048	0,050	0,045	0,050	0,055	0,059	0,047	**0,155**	**0,101**	0,055
	30, 30	0,052	0,050	0,052	0,052	0,052	0,047	0,049	0,052	0,053	0,055	0,051	**0,089**	**0,068**	0,049
	7, 15	0,051	0,048	0,051	0,051	0,051	0,042	0,047	0,051	0,057	0,057	0,051	**0,142**	0,036	0,045
	15, 30	0,046	0,050	0,046	0,046	0,046	0,054	0,049	0,047	0,047	0,047	0,046	**0,077**	0,041	0,048
6	7, 7	0,055	0,045	0,052	0,052	0,059	0,058	0,046	0,052	**0,098**	**0,108**	0,030	**0,341**	**0,149**	0,057
	15, 15	0,051	0,049	0,050	0,050	0,053	0,050	0,049	0,050	0,054	**0,064**	0,045	**0,150**	**0,089**	0,045
	30, 30	0,055	0,055	0,055	0,055	0,055	0,043	0,058	0,055	0,056	0,059	0,055	**0,105**	**0,079**	0,052
	7, 15	0,045	0,052	0,045	0,045	0,046	0,055	0,051	0,045	0,052	0,053	0,042	**0,138**	0,032	0,040
	15, 30	0,052	0,059	0,052	0,052	0,052	0,051	0,061	0,052	0,053	0,053	0,052	**0,086**	0,047	0,041

Table 13: The estimated type I error rates of all tests under distribution 3 for $p = 2$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,035	0,058	0,038	0,035	0,037	0,043	0,058	0,033	0,052	0,048	0,031	**0,207**	0,048	0,034
	15, 15	0,051	0,057	0,052	0,051	0,052	0,044	0,058	0,053	0,054	0,054	0,052	**0,106**	0,053	0,041
	30, 30	0,050	0,050	0,050	0,050	0,050	0,050	0,049	0,051	0,049	0,050	0,050	**0,061**	0,050	0,048
	7, 15	0,057	0,056	0,058	0,058	0,058	0,056	0,055	0,058	0,059	0,066	0,055	**0,108**	0,054	**0,061**
	15, 30	0,043	0,046	0,043	0,043	0,043	0,050	0,046	0,042	0,042	0,044	0,041	**0,065**	0,038	0,056
2	7, 7	0,037	0,061	0,040	0,038	0,037	0,043	0,058	0,037	0,055	0,044	0,031	**0,209**	0,051	0,043
	15, 15	0,044	0,048	0,044	0,044	0,044	0,050	0,046	0,044	0,045	0,046	0,044	**0,103**	0,046	0,050
	30, 30	0,045	0,044	0,045	0,045	0,045	0,047	0,043	0,045	0,044	0,044	0,045	0,057	0,045	0,051
	7, 15	0,058	0,059	0,058	0,058	0,058	0,053	0,059	0,056	0,056	**0,061**	0,051	**0,095**	0,050	0,055
	15, 30	0,050	0,055	0,050	0,050	0,049	0,050	0,054	0,050	0,050	0,051	0,050	**0,073**	0,047	0,054
3	7, 7	0,058	0,046	0,058	0,059	0,062	0,051	0,049	0,055	**0,086**	**0,087**	0,040	**0,281**	**0,094**	0,058
	15, 15	0,056	0,062	0,056	0,056	0,060	0,053	0,059	0,059	**0,062**	**0,067**	0,058	**0,141**	**0,076**	0,051
	30, 30	0,041	0,041	0,041	0,041	0,041	0,053	0,041	0,041	0,040	0,041	0,041	0,057	0,042	0,050
	7, 15	0,043	0,053	0,043	0,043	0,043	0,049	0,053	0,043	0,043	0,043	0,041	**0,077**	0,029	0,049
	15, 30	0,052	0,045	0,052	0,052	0,052	0,048	0,045	0,052	0,052	0,052	0,052	**0,065**	0,047	0,048
4	7, 7	0,050	0,053	0,049	0,050	0,049	0,051	0,052	0,047	**0,072**	**0,076**	0,036	**0,270**	**0,087**	0,054
	15, 15	0,056	0,056	0,055	0,055	0,057	0,052	0,057	0,056	0,057	**0,063**	0,054	**0,125**	**0,072**	0,054
	30, 30	0,048	0,051	0,048	0,048	0,048	0,056	0,054	0,050	0,050	0,049	0,050	**0,066**	0,056	0,046
	7, 15	0,048	0,051	0,048	0,048	0,048	0,051	0,051	0,047	0,048	0,048	0,045	**0,087**	0,034	0,058
	15, 30	0,050	0,052	0,050	0,050	0,050	0,043	0,050	0,051	0,051	0,052	0,052	**0,066**	0,045	0,057
5	7, 7	0,057	0,059	0,056	0,057	0,057	**0,066**	0,058	**0,062**	**0,067**	**0,076**	0,057	**0,165**	**0,102**	**0,065**
	15, 15	**0,063**	**0,062**	**0,062**	**0,062**	**0,063**	0,054	**0,062**	**0,064**	**0,064**	**0,064**	**0,061**	**0,105**	**0,079**	0,057
	30, 30	0,059	0,054	0,059	0,059	0,059	0,052	0,055	**0,061**	**0,061**	**0,060**	**0,061**	**0,076**	**0,066**	0,046
	7, 15	0,043	0,051	0,043	0,043	0,042	0,054	0,052	0,044	0,044	0,045	0,043	**0,076**	0,029	0,058
	15, 30	0,050	0,053	0,050	0,050	0,050	0,057	0,053	0,048	0,048	0,048	0,048	**0,062**	0,044	0,055
6	7, 7	0,059	0,052	0,059	0,059	0,060	0,059	0,053	0,063	**0,070**	**0,073**	0,056	**0,158**	**0,098**	**0,067**
	15, 15	0,058	0,052	0,057	0,057	0,058	0,055	0,053	0,058	0,058	**0,060**	0,056	**0,095**	**0,075**	0,053
	30, 30	0,055	0,057	0,055	0,055	0,056	0,051	0,056	0,055	0,055	0,057	0,055	**0,073**	**0,066**	0,058
	7, 15	0,052	0,057	0,053	0,052	0,052	0,050	0,057	0,054	0,055	0,054	0,054	**0,089**	0,040	0,049
	15, 30	0,052	0,059	0,052	0,052	0,052	0,046	0,062	0,051	0,051	0,052	0,051	**0,066**	0,046	0,053

Table 14: The estimated type I error rates of all tests under distribution 3 for $p = 3$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,041	0,048	0,042	0,041	0,042	0,037	0,047	0,039	0,047	0,046	0,040	**0,149**	0,050	0,041
	15, 15	0,043	0,043	0,043	0,043	0,043	0,056	0,041	0,043	0,044	0,043	0,043	**0,080**	0,045	0,042
	30, 30	0,047	0,043	0,047	0,047	0,047	0,055	0,043	0,046	0,046	0,046	0,046	0,059	0,045	0,048
	7, 15	0,050	0,055	0,049	0,050	0,052	0,054	0,056	0,050	0,057	**0,067**	0,041	**0,148**	0,045	**0,063**
	15, 30	0,046	0,047	0,046	0,046	0,046	0,054	0,048	0,047	0,047	0,050	0,044	**0,073**	0,042	0,053
2	7, 7	0,040	0,046	0,041	0,040	0,040	0,038	0,047	0,040	0,048	0,045	0,037	**0,161**	0,054	0,042
	15, 15	0,044	0,050	0,045	0,045	0,044	0,039	0,050	0,045	0,046	0,045	0,046	**0,092**	0,048	0,042
	30, 30	0,044	0,050	0,044	0,044	0,043	0,042	0,047	0,044	0,044	0,044	0,045	**0,068**	0,045	0,046
	7, 15	0,056	0,063	0,053	0,054	0,055	0,050	0,062	0,052	0,058	**0,066**	0,042	**0,145**	0,048	0,050
	15, 30	0,061	0,057	0,061	0,061	0,061	0,046	0,057	0,059	0,059	**0,064**	0,058	**0,099**	0,059	0,049
3	7, 7	0,048	0,048	0,048	0,048	0,048	0,056	0,049	0,048	0,057	0,059	0,041	**0,187**	**0,073**	0,053
	15, 15	0,047	0,050	0,047	0,047	0,048	0,051	0,050	0,047	0,050	0,052	0,047	**0,100**	0,056	0,057
	30, 30	0,053	0,054	0,053	0,053	0,054	0,055	0,055	0,056	0,055	0,057	0,057	**0,082**	0,064	0,051
	7, 15	0,050	0,054	0,052	0,051	0,052	0,044	0,053	0,053	0,056	0,057	0,049	**0,119**	0,032	0,043
	15, 30	0,048	0,054	0,048	0,048	0,048	0,050	0,053	0,048	0,048	0,048	0,047	**0,077**	0,037	0,044
4	7, 7	0,059	0,052	0,059	**0,060**	**0,063**	0,056	0,051	0,058	**0,071**	**0,072**	0,052	**0,219**	**0,084**	0,053
	15, 15	0,054	0,056	0,054	0,054	0,053	0,049	0,056	0,054	0,055	0,057	0,056	**0,102**	**0,065**	0,055
	30, 30	0,045	0,048	0,045	0,045	0,045	0,049	0,048	0,044	0,044	0,045	0,045	**0,067**	0,052	0,051
	7, 15	0,051	0,055	0,052	0,052	0,052	0,043	0,054	0,050	0,054	0,055	0,047	**0,120**	0,032	0,036
	15, 30	0,052	0,052	0,052	0,052	0,052	0,057	0,050	0,052	0,052	0,052	0,051	**0,073**	0,040	0,047
5	7, 7	**0,067**	0,059	**0,066**	**0,067**	**0,069**	**0,060**	**0,060**	**0,068**	**0,088**	**0,093**	0,049	**0,246**	**0,126**	**0,075**
	15, 15	0,054	0,053	0,053	0,053	0,053	0,054	0,054	0,053	0,055	0,059	0,048	**0,115**	**0,078**	**0,064**
	30, 30	0,056	0,056	0,059	0,056	0,056	0,053	0,059	0,054	0,054	0,054	0,054	**0,088**	**0,068**	0,046
	7, 15	0,045	0,055	0,046	0,045	0,046	0,049	0,054	0,047	0,047	0,048	0,047	**0,107**	0,033	0,054
	15, 30	0,043	0,049	0,043	0,043	0,043	0,051	0,051	0,043	0,043	0,043	0,043	**0,073**	0,038	0,055
6	7, 7	**0,067**	0,059	**0,065**	**0,066**	**0,069**	0,054	0,057	**0,067**	**0,096**	**0,108**	0,052	**0,257**	**0,139**	**0,075**
	15, 15	0,051	0,053	0,051	0,051	0,052	0,059	0,053	0,052	0,054	0,056	0,050	**0,114**	**0,081**	0,049
	30, 30	0,058	0,058	0,054	0,057	0,058	0,050	0,054	0,056	0,055	0,058	0,059	**0,088**	**0,075**	0,054
	7, 15	0,050	0,058	0,051	0,051	0,051	0,053	0,058	0,051	0,054	0,055	0,051	**0,124**	0,037	**0,065**
	15, 30	0,055	0,051	0,055	0,055	0,054	0,053	0,050	0,055	0,055	0,055	0,055	**0,082**	0,047	0,059

Table 15: The estimated type I error rates of all tests under distribution 3 for $p = 4$. In the table, the estimated type I error rates of 6% were given in bold characters.

		Tests													
Setting	**n**	**CAT**	**B-CAT**	**M-CAT**	**J-CAT**	**Y-CAT**	**NV-CAT**	T_B	T_M	T_J	T_Y	T_{NV}	**Wald**	**WB**	**WPB**
1	7, 7	0,027	0,033	0,030	0,028	0,032	0,042	0,034	0,029	0,048	0,042	0,029	**0,206**	0,045	0,038
	15, 15	0,048	0,051	0,050	0,049	0,048	0,049	0,049	0,047	0,050	0,048	0,045	**0,104**	0,050	0,045
	30, 30	0,052	0,054	0,053	0,053	0,052	0,055	0,055	0,052	0,051	0,052	0,052	**0,073**	0,052	0,050
	7, 15	**0,063**	0,056	**0,060**	**0,062**	**0,068**	0,057	0,056	0,056	**0,073**	**0,081**	0,037	**0,203**	0,048	0,057
	15, 30	0,050	0,057	0,050	0,050	0,049	0,045	0,055	0,049	0,050	0,052	0,047	**0,095**	0,049	0,045
2	7, 7	0,042	0,049	0,043	0,043	0,043	0,040	0,049	0,042	0,055	0,053	0,039	**0,211**	0,053	0,046
	15, 15	0,041	0,043	0,043	0,042	0,041	0,053	0,044	0,044	0,045	0,044	0,043	**0,097**	0,048	0,042
	30, 30	0,049	0,050	0,049	0,049	0,048	0,048	0,049	0,048	0,047	0,047	0,047	**0,076**	0,047	0,043
	7, 15	0,053	0,048	0,051	0,052	0,058	0,056	0,051	0,049	**0,063**	**0,078**	0,030	**0,190**	0,047	0,051
	15, 30	0,056	0,055	0,056	0,056	0,057	0,046	0,055	0,057	0,058	**0,060**	0,049	**0,113**	0,050	0,052
3	7, 7	0,049	0,045	0,049	0,048	0,053	0,053	0,046	0,047	**0,075**	**0,079**	0,035	**0,259**	**0,085**	0,058
	15, 15	0,053	0,052	0,053	0,053	0,054	0,055	0,052	0,054	0,058	**0,063**	0,054	**0,135**	**0,078**	0,054
	30, 30	0,053	0,054	0,053	0,053	0,054	0,044	0,052	0,051	0,050	0,051	0,052	**0,091**	**0,062**	0,054
	7, 15	0,035	0,048	0,037	0,036	0,037	0,053	0,048	0,036	0,043	0,043	0,031	**0,126**	0,018	0,054
	15, 30	0,041	0,043	0,042	0,042	0,042	0,053	0,046	0,042	0,042	0,042	0,040	**0,078**	0,032	0,044
4	7, 7	0,051	0,046	0,051	0,052	0,054	0,053	0,045	0,050	**0,078**	**0,081**	0,039	**0,271**	**0,096**	0,056
	15, 15	0,047	0,051	0,047	0,047	0,047	0,054	0,049	0,046	0,049	0,055	0,046	**0,129**	**0,067**	0,057
	30, 30	0,057	0,059	0,057	0,057	0,056	0,052	0,059	0,057	0,057	0,059	0,059	**0,095**	**0,069**	0,056
	7, 15	0,049	0,050	0,050	0,050	0,050	0,046	0,049	0,049	0,055	0,056	0,043	**0,154**	0,027	0,047
	15, 30	0,054	0,053	0,054	0,054	0,054	0,054	0,052	0,054	0,055	0,055	0,054	**0,091**	0,041	0,055
5	7, 7	**0,074**	**0,060**	**0,071**	**0,073**	**0,079**	**0,072**	**0,060**	**0,073**	**0,119**	**0,128**	0,040	**0,360**	**0,162**	**0,074**
	15, 15	**0,063**	0,058	**0,063**	**0,063**	**0,065**	0,057	0,057	**0,064**	**0,069**	**0,075**	**0,061**	**0,156**	**0,105**	**0,062**
	30, 30	0,058	0,059	0,057	0,057	0,059	0,048	0,058	0,056	0,057	**0,060**	0,056	**0,110**	**0,079**	0,057
	7, 15	0,046	0,054	0,049	0,049	0,049	0,057	0,055	0,050	0,055	0,056	0,050	**0,161**	0,030	0,055
	15, 30	0,053	0,052	0,053	0,053	0,053	0,052	0,052	0,053	0,053	0,055	0,053	**0,091**	0,045	0,056
6	7, 7	**0,073**	**0,066**	**0,071**	**0,071**	**0,079**	**0,069**	**0,066**	**0,072**	**0,112**	**0,124**	0,041	**0,352**	**0,153**	**0,080**
	15, 15	**0,064**	**0,062**	**0,063**	**0,063**	**0,065**	0,059	**0,061**	**0,064**	**0,071**	**0,083**	**0,061**	**0,173**	**0,114**	0,057
	30, 30	**0,062**	0,059	**0,062**	**0,062**	**0,064**	0,054	0,057	**0,061**	**0,062**	**0,065**	**0,061**	**0,113**	**0,090**	0,058
	7, 15	0,047	0,050	0,048	0,047	0,050	0,056	0,051	0,047	0,056	0,056	0,046	**0,156**	0,032	0,054
	15, 30	0,050	0,049	0,050	0,050	0,050	0,056	0,048	0,049	0,049	0,050	0,049	**0,086**	0,039	0,052

Table 16: The estimated type I error rates of all tests under distribution 4 for $p = 2$. In the table, the estimated type I error rates of 6% were given in bold characters.

Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,044	0,055	0,044	0,044	0,045	0,048	0,055	0,044	0,048	0,048	0,043	**0,108**	0,050	0,051
	15, 15	0,044	0,046	0,044	0,044	0,044	0,049	0,045	0,043	0,043	0,043	0,043	**0,065**	0,044	0,049
	30, 30	0,049	0,049	0,049	0,049	0,049	0,056	0,048	0,050	0,049	0,050	0,050	**0,060**	0,050	0,049
	7, 15	0,055	0,055	0,056	0,056	0,055	0,050	0,054	0,057	**0,060**	**0,062**	0,052	**0,111**	0,050	0,051
	15, 30	0,043	0,046	0,042	0,043	0,042	0,054	0,046	0,045	0,045	0,045	0,043	**0,067**	0,044	0,043
2	7, 7	0,042	0,049	0,042	0,042	0,041	0,042	0,049	0,043	0,046	0,045	0,046	**0,114**	0,050	0,034
	15, 15	0,047	0,050	0,047	0,047	0,047	0,043	0,048	0,048	0,048	0,046	0,047	**0,073**	0,049	0,040
	30, 30	0,041	0,043	0,041	0,041	0,042	0,047	0,044	0,044	0,044	0,044	0,044	0,052	0,043	0,055
	7, 15	0,051	0,052	0,051	0,051	0,052	0,054	0,052	0,052	0,053	0,058	0,047	**0,103**	0,045	0,058
	15, 30	0,052	0,050	0,052	0,052	0,052	0,054	0,049	0,050	0,050	0,052	0,049	**0,073**	0,049	0,044
3	7, 7	0,052	0,059	0,053	0,053	0,053	0,053	0,057	0,051	0,054	0,057	0,051	**0,127**	**0,069**	0,057
	15, 15	0,054	0,057	0,054	0,054	0,054	0,053	0,059	0,057	0,057	0,058	0,057	**0,085**	**0,063**	0,056
	30, 30	0,041	0,044	0,041	0,041	0,041	0,041	0,041	0,042	0,042	0,042	0,042	0,054	0,046	0,045
	7, 15	0,053	0,051	0,052	0,052	0,052	0,046	0,051	0,051	0,052	0,052	0,048	**0,090**	0,036	0,051
	15, 30	0,048	0,058	0,048	0,048	0,048	0,054	0,058	0,048	0,048	0,048	0,048	**0,061**	0,043	0,053
4	7, 7	0,041	0,049	0,041	0,041	0,042	0,051	0,048	0,042	0,044	0,049	0,040	**0,114**	**0,061**	0,057
	15, 15	0,053	0,056	0,052	0,052	0,053	0,054	0,059	0,054	0,055	0,056	0,054	**0,093**	**0,063**	0,056
	30, 30	0,044	0,046	0,044	0,044	0,044	0,054	0,047	0,043	0,043	0,044	0,045	0,059	0,048	0,054
	7, 15	0,046	0,052	0,047	0,047	0,046	0,046	0,053	0,046	0,046	0,045	0,045	**0,086**	0,034	0,047
	15, 30	0,049	0,044	0,049	0,049	0,049	0,048	0,044	0,048	0,048	0,047	0,048	**0,066**	0,044	0,059
5	7, 7	0,053	0,049	0,054	0,053	0,055	0,058	0,047	0,058	**0,060**	**0,068**	0,054	**0,147**	**0,090**	0,059
	15, 15	0,053	0,047	0,054	0,054	0,053	0,052	0,049	0,055	0,055	0,057	0,054	**0,095**	**0,072**	0,065
	30, 30	0,053	0,053	0,054	0,054	0,054	0,049	0,054	0,055	0,055	0,055	0,054	**0,072**	**0,063**	0,055
	7, 15	0,051	0,051	0,052	0,051	0,052	0,044	0,051	0,050	0,051	0,050	0,050	**0,087**	0,036	**0,066**
	15, 30	0,056	0,058	0,056	0,056	0,056	0,050	0,061	0,056	0,056	0,056	0,056	**0,076**	0,052	0,054
6	7, 7	**0,065**	**0,062**	**0,065**	**0,065**	**0,065**	**0,064**	**0,063**	**0,065**	**0,069**	**0,076**	**0,061**	**0,162**	**0,104**	**0,067**
	15, 15	0,052	0,054	0,052	0,053	0,051	**0,061**	0,052	0,052	0,052	0,055	0,050	**0,093**	**0,071**	0,058
	30, 30	0,056	0,056	0,058	0,055	0,055	0,056	0,059	0,056	0,056	0,057	0,056	**0,072**	**0,063**	0,051
	7, 15	0,047	0,050	0,048	0,048	0,048	0,045	0,051	0,045	0,046	0,047	0,044	**0,083**	0,038	0,055
	15, 30	0,055	0,062	0,055	0,055	0,055	0,055	0,061	0,053	0,053	0,053	0,055	**0,070**	0,049	0,050

Table 17: The estimated type I error rates of all tests under distribution 4 for $p = 3$. In the table, the estimated type I error rates of 6% were given in bold characters.

Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,039	0,049	0,041	0,040	0,038	0,046	0,049	0,038	0,045	0,043	0,037	**0,153**	0,048	0,036
	15, 15	0,052	0,058	0,052	0,052	0,053	0,046	0,058	0,054	0,055	0,054	0,053	**0,094**	0,055	0,046
	30, 30	0,053	0,051	0,053	0,053	0,053	0,053	0,054	0,053	0,053	0,053	0,053	**0,064**	0,052	0,045
	7, 15	0,055	0,053	0,054	0,055	0,054	0,048	0,056	0,054	0,058	0,069	0,047	**0,145**	0,048	0,057
	15, 30	0,050	0,051	0,050	0,050	0,052	0,049	0,049	0,049	0,050	0,054	0,047	**0,084**	0,050	0,051
2	7, 7	0,043	0,054	0,045	0,044	0,043	0,045	0,053	0,044	0,051	0,048	0,038	**0,158**	0,054	0,042
	15, 15	0,051	0,054	0,052	0,052	0,051	0,052	0,054	0,052	0,052	0,052	0,052	**0,091**	0,053	0,050
	30, 30	0,056	0,053	0,057	0,057	0,057	0,059	0,051	0,055	0,054	0,055	0,054	**0,071**	0,055	0,047
	7, 15	0,052	0,049	0,052	0,052	0,055	0,054	0,050	0,051	0,057	**0,070**	0,041	**0,144**	0,045	0,053
	15, 30	0,051	0,047	0,051	0,051	0,051	0,046	0,047	0,053	0,053	0,056	0,051	**0,082**	0,050	0,055
3	7, 7	0,041	0,048	0,042	0,042	0,044	0,053	0,049	0,042	0,056	**0,062**	0,037	**0,182**	**0,068**	0,051
	15, 15	0,052	0,055	0,052	0,052	0,053	0,054	0,053	0,051	0,053	0,056	0,052	**0,101**	**0,060**	0,053
	30, 30	0,058	0,055	0,059	0,059	0,059	0,054	0,057	0,059	0,059	0,060	0,059	**0,082**	**0,065**	0,055
	7, 15	0,051	0,052	0,052	0,051	0,051	0,051	0,051	0,050	0,053	0,052	0,043	**0,100**	0,029	0,053
	15, 30	0,045	0,045	0,044	0,044	0,044	0,055	0,045	0,043	0,043	0,044	0,042	**0,072**	0,037	0,051
4	7, 7	0,050	0,054	0,051	0,051	0,052	0,050	0,055	0,050	**0,064**	**0,068**	0,046	**0,187**	**0,083**	0,059
	15, 15	0,050	0,047	0,050	0,050	0,051	0,052	0,046	0,048	0,050	0,054	0,050	**0,113**	**0,062**	0,055
	30, 30	0,052	0,052	0,052	0,052	0,051	0,055	0,054	0,052	0,051	0,054	0,053	**0,077**	**0,060**	0,059
	7, 15	0,049	0,051	0,050	0,049	0,049	0,046	0,051	0,047	0,051	0,050	0,043	**0,111**	0,029	0,044
	15, 30	0,055	0,059	0,055	0,055	0,055	0,050	0,059	0,055	0,055	0,055	0,054	**0,080**	0,044	0,052
5	7, 7	0,058	0,055	0,057	0,059	**0,060**	**0,070**	0,053	0,057	**0,078**	**0,089**	0,043	**0,228**	**0,118**	**0,067**
	15, 15	0,051	0,049	0,051	0,051	0,053	0,047	0,048	0,051	0,054	0,058	0,050	**0,120**	**0,076**	0,058
	30, 30	0,059	0,058	0,059	0,059	0,059	0,055	0,058	0,060	0,060	**0,061**	0,060	**0,096**	**0,076**	0,050
	7, 15	0,053	0,051	0,053	0,053	0,054	0,045	0,051	0,053	0,058	**0,060**	0,052	**0,123**	0,034	0,055
	15, 30	0,052	0,047	0,052	0,052	0,052	0,049	0,048	0,051	0,051	0,052	0,051	**0,079**	0,046	0,054
6	7, 7	0,053	0,054	0,053	0,053	0,055	**0,062**	0,056	0,056	**0,081**	**0,093**	0,039	**0,229**	**0,122**	**0,065**
	15, 15	0,047	0,051	0,045	0,046	0,045	0,056	0,050	0,045	0,048	0,052	0,043	**0,114**	**0,075**	**0,065**
	30, 30	0,053	0,054	0,053	0,053	0,053	0,051	0,054	0,051	0,051	0,053	0,052	**0,084**	**0,066**	0,058
	7, 15	0,047	0,042	0,048	0,047	0,047	0,051	0,042	0,049	0,051	0,052	0,047	**0,109**	0,031	0,051
	15, 30	0,048	0,048	0,048	0,048	0,048	0,057	0,047	0,049	0,049	0,050	0,050	**0,073**	0,040	0,053

Table 18: The estimated type I error rates of all tests under distribution 4 for $p = 4$. In the table, the estimated type I error rates of 6% were given in bold characters.

Setting	n	CAT	B-CAT	M-CAT	J-CAT	Y-CAT	NV-CAT	T_B	T_M	T_J	T_Y	T_{NV}	Wald	WB	WPB
1	7, 7	0,039	0,048	0,041	0,040	0,042	0,035	0,047	0,039	**0,061**	0,053	0,039	**0,213**	0,054	0,042
	15, 15	0,055	0,056	0,055	0,055	0,055	0,047	0,057	0,055	0,057	0,056	0,053	**0,120**	0,056	0,048
	30, 30	0,046	0,051	0,047	0,047	0,046	0,040	0,047	0,049	0,048	0,048	0,049	**0,072**	0,050	0,042
	7, 15	0,068	0,054	0,066	0,065	0,068	0,058	0,054	0,066	0,074	**0,088**	0,043	**0,213**	0,056	**0,060**
	15, 30	0,058	0,058	0,056	0,057	0,057	0,048	0,058	0,056	0,058	**0,062**	0,052	**0,111**	0,053	0,050
2	7, 7	0,038	0,052	0,038	0,038	0,041	0,039	0,054	0,037	0,056	0,048	0,034	**0,204**	0,052	0,048
	15, 15	0,047	0,051	0,048	0,047	0,048	0,046	0,051	0,045	0,049	0,049	0,046	**0,112**	0,052	0,046
	30, 30	0,057	0,055	0,057	0,057	0,056	0,047	0,057	0,057	0,057	0,057	0,057	**0,086**	0,056	0,048
	7, 15	0,047	0,043	0,044	0,046	0,050	0,053	0,043	0,044	0,055	**0,064**	0,028	**0,179**	0,039	0,053
	15, 30	0,049	0,049	0,049	0,049	0,050	0,053	0,050	0,048	0,051	0,056	0,047	**0,103**	0,042	0,049
3	7, 7	0,055	0,049	0,055	0,055	0,057	0,046	0,051	0,051	**0,083**	**0,089**	0,040	**0,293**	**0,095**	**0,072**
	15, 15	0,055	0,052	0,055	0,055	0,056	0,048	0,055	0,055	**0,061**	**0,063**	0,054	**0,137**	**0,074**	0,048
	30, 30	0,058	0,054	0,057	0,058	0,057	0,047	0,056	0,057	0,058	0,059	0,058	**0,101**	**0,069**	0,049
	7, 15	0,047	0,053	0,049	0,048	0,048	0,039	0,052	0,047	0,055	0,053	0,041	**0,140**	0,027	0,052
	15, 30	0,054	0,059	0,053	0,053	0,053	0,057	0,058	0,051	0,052	0,053	0,051	**0,085**	0,041	0,050
4	7, 7	0,049	0,057	0,050	0,050	0,054	0,057	0,058	0,047	**0,079**	**0,089**	0,034	**0,286**	**0,098**	0,058
	15, 15	0,049	0,048	0,049	0,049	0,047	0,047	0,048	0,049	0,052	0,055	0,048	**0,131**	**0,063**	0,052
	30, 30	0,048	0,051	0,048	0,048	0,048	0,050	0,050	0,049	0,049	0,049	0,049	**0,081**	0,056	0,050
	7, 15	0,043	0,063	0,045	0,044	0,048	0,053	**0,062**	0,043	0,049	0,049	0,039	**0,141**	0,025	0,047
	15, 30	0,055	0,051	0,054	0,054	0,054	0,048	0,050	0,054	0,054	0,054	0,053	**0,087**	0,044	0,046
5	7, 7	**0,065**	0,052	0,059	**0,062**	**0,068**	**0,069**	0,053	**0,061**	**0,108**	**0,130**	0,032	**0,358**	**0,165**	**0,075**
	15, 15	0,049	0,046	0,048	0,048	0,047	0,049	0,046	0,045	0,051	0,056	0,042	**0,147**	**0,083**	**0,063**
	30, 30	0,055	0,057	0,055	0,055	0,056	0,058	0,054	0,055	0,055	0,057	0,054	**0,099**	**0,077**	0,053
	7, 15	0,051	0,054	0,051	0,051	0,052	0,059	0,055	0,052	0,057	**0,060**	0,051	**0,146**	0,034	0,050
	15, 30	0,047	0,056	0,047	0,047	0,047	0,054	0,056	0,047	0,048	0,049	0,048	**0,095**	0,040	0,044
6	7, 7	**0,070**	**0,061**	**0,068**	**0,070**	**0,073**	**0,075**	**0,061**	**0,067**	**0,111**	**0,127**	0,034	**0,358**	**0,169**	**0,078**
	15, 15	0,059	0,057	0,057	0,058	0,059	**0,061**	0,057	0,058	**0,064**	**0,067**	0,052	**0,162**	**0,103**	0,054
	30, 30	0,051	0,052	0,050	0,050	0,051	**0,063**	0,051	0,052	0,052	0,053	0,052	**0,098**	**0,070**	0,057
	7, 15	0,056	0,056	0,057	0,056	0,056	0,050	0,057	0,054	**0,065**	**0,064**	0,054	**0,159**	0,035	0,046
	15, 30	0,049	0,053	0,049	0,049	0,049	0,059	0,052	0,046	0,046	0,047	0,047	**0,094**	0,037	0,053

6. CONCLUSIONS

In this study, the CAT were proposed and compared it against the other popular tests (T_B, T_{M}, T_{NV}, T_J, T_Y, WB, WPB) as well as their CAT versions (B-CAT, M-CAT, J-CAT, Y-CAT, NV-CAT) to test the equality of two multivariate normal mean vectors under heterogeneity of covariance matrix. The results of Monte Carlo simulations that were conducted to compare the estimated type I error rates and powers of these tests were presented. The simulation study shows that the CAT, M-CAT, J-CAT, Y-CAT, NV-CAT, T_{M} and WPB tests performed better than the others in terms of both the estimated type I error rates and power, even the CAT and NV-CAT had a bit higher power than the other tests in some cases. This method can be adapted to the heterogeneity MANOVA models.

ACKNOWLEDGMENTS

The author is thankful to the referee and the editor for their constructive comments that led to a significant improvement of the paper.

REFERENCES

[1] Algina, J.; Oshima, T.C. and Tang, K.L. (1991). Robustness of Yao's, James', and Johansen's tests under variance-covariance heteroscedasticity and nonnormality, *Journal of Educational and Behavioral Statistics*, **16**(2), 125–139.

[2] Bennett, B.M. (1951). Note on a solution of the generalized Behrens–Fisher problem, *Annals of the Institute of Statistical Mathematics*, **2**, 87–90.

[3] Chang, C.H. and Pal, N. (2008). A revisit to the Behrens–Fisher problem: comparison of five test methods, *Communications in Statistics – Simulation and Computation*, **37**(6), 1064–1085.

[4] Chang, C.H.; Pal, N.; Lim, W.K. and Lin, J.J. (2010). Comparing several population means: a parametric bootstrap method, and its comparison with usual ANOVA F test as well as ANOM, *Computational Statistics*, **25**(1), 71–95.

[5] Chang, C.H.; Lin, J.J. and Pal, N. (2011). Testing the equality of several gamma means: a parametric bootstrap method with applications, *Computational Statistics*, **26**(1), 55–76.

[6] Christensen, W.F. and Rencher, A.C. (1997). A comparison of Type I error rates and power levels for seven solutions to the multivariate Behrens–Fisher problem, *Communications in Statistics-Simulation and Computation*, **26**(4), 1251–1273.

[7] De la Rey, N. and Nel, D.G. (1993). A comparison of the significance levels and power functions of several solutions to the multivariate Behrens–Fisher problem, *South African Statistical Journal*, **27**(2), 129–148.

[8] Gökpinar, E.Y. and Gökpinar, F. (2012). A test based on the computational approach for equality of means under the unequal variance assumption, *Hacettepe Journal of Mathematics and Statistics*, **41**(4), 1–9.

[9] Gökpinar, E.Y.; Polat, E.; Gökpinar, F. and Günay, S. (2013). A new computational approach for testing equality of inverse Gaussian means under heterogeneity, *Hacettepe Journal of Mathematics and Statistics*, **42**(5), 581–590.

[10] Gökpinar, F. and Gökpinar, E. (2015). A computational approach for testing equality of coefficients of variation in k normal populations, *Hacettepe Journal of Mathematics and Statistics*, **44**(5), 1197–1213.

[11] Hwang, H.L. and Paulson, A.S. (1986). *Some methods for the multivariate two-sample location problem*, unpublished manuscript, Oak Ridge National Laboratory, Oak Ridge, TN.

[12] Ito, K. and Schull, W.J. (1964). On the robustness of the To2 test in multivariate analysis of variance when variance-covariance matrices are not equal, *Biometrika*, **51**(1–2), 71–82.

[13] James, G.S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown, *Biometrika*, **41**, 19–43.

[14] Johansen, S. (1980). The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression, *Biometrika*, **67**, 85–92.

[15] Kim, S.J. (1992) A practical solution to the multivariate Behrens–Fisher problem, *Biometrika*, **79**(1), 171–176.

[16] Konietschke, F.; Bathke, A.C.; Harrar, S.W. and Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA, *Journal of Multivariate Analysis*, **140**, 291–301.

[17] Krishnamoorthy, K. and Yu, J. (2004). Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem, *Statistics & Probability Letters*, **66**(2), 161–169.

[18] Nel, D.G. and Van der Merwe, C.A. (1986). A solution to the multivariate Behrens–Fisher problem, *Communications in Statistics – Theory and Methods*, **15**(12), 3719–3735.

[19] Pal, N.; Lim, W.K. and Ling, C.H. (2007). A computational approach to statistical inferences, *Journal of Applied Probability and Statistics*, **2**(1), 13–35.

[20] Smaga, Ł. (2017). Bootstrap methods for multivariate hypothesis testing, *Communications in Statistics – Simulation and Computation*, **46**, 7654–7667.

[21] Subrahmaniam, K. and Subrahmaniam, K. (1973). On the multivariate Behrens–Fisher problem, *Biometrika*, **60**(1), 107–11.

[22] Subrahmaniam, K. and Subrahmaniam, K. (1975). On the confidence region comparison of some solutions for the multivariate Behrens–Fisher problem, *Communications in Statistics – Theory and Methods*, **4**(1), 57–67.

[23] Tukey, J.W. (1959). A quick compact two sample test to Duckworth's specifications, *Technometrics*, **1**(1), 31–48.

[24] Welch, B.L. (1938). The significance of the difference between two means when the population variances are unequal, *Biometrika*, **29**(3/4), 350–362.

[25] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A**(1/2), 32–52.

[26] Yanagihara, H. and Yuan, K.H. (2005). Three approximate solutions to the multivariate Behrens–Fisher problem, *SCommunications in Statistics – Simulation and Computation*, **34**(4), 975–988.

[27] Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem, *Biometrika*, **52**(1/2), 139–147.

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 179–198

ASYMPTOTICS OF THE ADAPTIVE ELASTIC NET ESTIMATION FOR CONDITIONAL HETEROSCEDASTIC TIME SERIES MODELS

Authors: Yuanyuan Liao
– Department of Mathematics, Nanjing University, Nanjing, PR China

Lihong Wang
– Department of Mathematics, Nanjing University, Nanjing, PR China
lhwang@nju.edu.cn

Received: January 2019 Revised: September 2019 Accepted: January 2020

Abstract:

- In this paper we propose an iteratively reweighted adaptive elastic net estimation method for conditional heteroscedastic time series models. The sign consistency and the asymptotic normality of the estimator are investigated. Compared with the Lasso method, the elastic net is more efficient for autoregressive time series models, because it benefits not only from the selection of the Lasso but also from the grouping effect inherited from the ridge penalty. The Monte Carlo simulation studies based on an AR-ARCH model are reported to assess the finite-sample performance of the proposed elastic net method.

Keywords:

- *adaptive elastic net; AR-ARCH models; asymptotic normality; iteratively reweighted algorithm; sign consistency.*

AMS Subject Classification:

- 62J07, 62M10.

1. INTRODUCTION

The Lasso introduced in [10] is a shrinkage and selection method for linear regression models. As variable selection is of increasing importance in big data analysis, the lasso is much more appealing owing to its sparse representation. However, the literature about the penalization techniques mainly deals with homoscedastic linear regression models, see, e.g., [2], [5], [9], [15], [16], and [18], among others. The investigation of the Lasso type estimator for heteroscedastic models started relatively late. Recently, [11] and [12] analysed the weighted lasso type estimators in a linear heteroscedastic regression model setting. [17] derived an iteratively reweighted adaptive lasso algorithm for time series models under conditional heteroscedasticity, and proved that the resulting estimator has sign consistency and asymptotic normality. The proposed method can be applied to various AR-ARCH type processes.

In this paper, we generalize the results of [17] to the adaptive elastic net method. That is, we consider the model similar to the one used by [17], but suggest the use of an iteratively reweighted adaptive elastic net algorithm. The elastic net introduced by [19] is a convex combination of the Lasso and ridge penalty. The ridge part of the penalty shrinks the estimated coefficients of all the variables and induces coefficients of correlated variables to be close to one another. The Lasso part of the penalty shrinks and selects the coefficients of the variables. As discussed in [4], the elastic net benefits from the selection of the Lasso, as well as from the finite-sample grouping effect inherited from the ridge penalty. This makes the elastic net particularly useful for estimating the autoregressive time series models, since this estimation procedure leaves out irrelevant variables but does not exclude correlated variables that may be relevant as part of a group.

In the next section, we introduce the iteratively reweighted adaptive elastic net algorithm for high-dimensional sparse linear regression models under conditional heteroscedasticity. The sign consistency and the asymptotic normality of the weighted adaptive elastic net estimators of the parameters are also addressed. Section 3 gives the Monte Carlo simulations based on a specific AR-ARCH model, evaluating and comparing the performance of the proposed adaptive elastic net algorithm and the adaptive Lasso method. The proof of the theorem is given in Appendix.

Throughout the paper, all limits are taken as $n \to \infty$, unless specified otherwise. The symbol C denotes an absolute positive constant whose value may vary at each occurrence. $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, $\xrightarrow{\mathcal{P}}$ denotes convergence in probability, Z stands for a standard normal random variable. For any two real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ means that there are constants $c > 0$ and $C < \infty$ such that $c \leq a_n/b_n \leq C$ for all sufficiently large n.

2. THE ITERATIVELY REWEIGHTED ADAPTIVE ELASTIC NET ALGORITHM

We now introduce the model and the basic ideas of the algorithm. The model discussed here is similar to the one used by [14] and [17]. We consider a stationary random process $Y_t \in \mathbb{R}$ and a possibly infinite vector of covariates of stationary processes, $X_{t,\infty} = (X_{t,1}, X_{t,2}, ...)'$, $t \in \mathbb{Z}$, $\mathbb{Z} := \{0, \pm 1, \pm 2, ...\}$, obeying the model

$$Y_t = X'_{t,\infty}\beta^0_\infty + \varepsilon_t\,, \qquad t \in \mathbb{Z}\,, \tag{2.1}$$

where $\beta^0_\infty = (\beta^0_1, \beta^0_2, ...)'$ satisfying $\sum_{i=1}^\infty |\beta^0_i|^2 < +\infty$, ε_t is zero mean and independent of the covariates $X_{t,\infty}$, and

$$\varepsilon_t = \sigma_t Z_t\,, \qquad \sigma_t = g(\alpha^0_\infty; L_{\infty,t})\,, \quad t \in \mathbb{Z}\,,$$

where Z_t, $t \in \mathbb{Z}$, are i.i.d. standardized r.v.'s, g is a positive function, $L_{\infty,t} = (L_{1,t}, L_{2,t}, ...)$ is a possibly infinite vector of covariates of stationary processes $L_{i,t}$, $t \in \mathbb{Z}$, and $\alpha^0_\infty = (\alpha^0_1, \alpha^0_2, ...)'$ is a parameter vector. Here the covariates $X_{t,\infty}$ and $L_{\infty,t}$ can contain lagged versions of Y_t and $(\varepsilon_t, \sigma_t)$, respectively, which allows flexible modelling of autoregressive processes and a class of conditional variance models such as GARCH type models.

The observed data consists of $(\mathbf{X}_n, \mathbf{Y}_n)$, where

$$\mathbf{Y}_n = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X}_n = \begin{pmatrix} X_{1,1} & \cdots & X_{1,p_n} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \cdots & X_{n,p_n} \end{pmatrix}, \qquad \boldsymbol{\beta}^0_n = \begin{pmatrix} \beta^0_1 \\ \vdots \\ \beta^0_{p_n} \end{pmatrix}, \qquad \boldsymbol{\varepsilon}^0_n = \mathbf{Y}_n - \mathbf{X}_n \boldsymbol{\beta}^0_n\,,$$

where p_n is the number of possible parameters which increases with sample size n, $\boldsymbol{\beta}^0_n$ is the restriction of β^0_∞ to its first p_n coordinates, $\boldsymbol{\varepsilon}^0_n = (\varepsilon^0_1, ..., \varepsilon^0_n)'$.

The fact $\sum_{i=1}^\infty |\beta^0_i|^2 < +\infty$ implies that there is a positive sequence a_n decreasing to zero such that $\lim_{n\to\infty} P(\max_{1\le t\le n} |\varepsilon^0_t - \varepsilon_t| < a_n) \to 1$ holds. Thus, for a sufficiently large n we can approximately write

$$\varepsilon^0_t = \sigma_t Z_t\,, \qquad \sigma_t = g_n(\boldsymbol{\alpha}^0_n; \mathbf{L}^0_{n,t})\,, \quad 1 \le t \le n\,, \tag{2.2}$$

here $\boldsymbol{\alpha}^0_n$ and $\mathbf{L}^0_{n,t}$ are the restrictions of α^0_∞ and $L_{\infty,t}$ to their first p_n coordinates, respectively, and g_n is the restriction of g that corresponds to $\boldsymbol{\alpha}^0_n$ and $\mathbf{L}^0_{n,t}$. Without loss of generality we assume that only q_n of the p_n parameters are non-zero. That is, $\boldsymbol{\beta}^0_n = (\beta^0_1, ..., \beta^0_{q_n}, 0, ..., 0)' = (\boldsymbol{\beta}^0_n(1)', \mathbf{0}')'$. In a similar manner, $\mathbf{X}_n = (\mathbf{X}_n(1), \mathbf{X}_n(2))$ and $\mathbf{X}_{t,n} = (\mathbf{X}_{t,n}(1)', \mathbf{X}_{t,n}(2)')'$, where $\mathbf{X}_{t,n}$ is the t-th row of $\mathbf{X}_n$.

We now introduce the adaptive elastic net algorithm based on an iteratively reweighted technique which is similar to the approaches in [7], [8], and [17]. Rewrite Model (2.1) as

$$\widetilde{Y}_t = \widetilde{\mathbf{X}}'_{t,n}\boldsymbol{\beta}^0_n + Z_t\,, \qquad 1 \le t \le n\,, \tag{2.3}$$

where $\widetilde{Y}_t = \frac{1}{\sigma_t}Y_t$, $\widetilde{\mathbf{X}}_{t,n} = \frac{1}{\sigma_t}\mathbf{X}_{t,n}$. It is obvious that the error Z_t is homoscedastic.

Since we have no a priori information about the conditional standard deviation σ_t, at first step we assume homoscedasticity. Then we use a weighted adaptive elastic net algorithm to estimate $\boldsymbol{\beta}_n^0$ in each iteration step. That is,

$$\begin{aligned}(2.4)\qquad &\boldsymbol{\beta}_{n,\text{elastic}}(\lambda_n, \gamma_n, w_n)\\ &\qquad = \arg\min_{\boldsymbol{\beta}} (\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta})' W_n^2 (\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}) + \lambda_n \|\Sigma_1 \boldsymbol{\beta}\|_1 + \gamma_n \|\Sigma_2^{1/2}\boldsymbol{\beta}\|_2^2,\end{aligned}$$

where $\lambda_n \geq 0$, $\gamma_n \geq 0$, $\Sigma_1 = \text{diag}(v_n)$, $v_n = (v_{n,1}, ..., v_{n,p_n}) = |\boldsymbol{\beta}_{n,\text{init1}}|^{-\tau_1}$, $\Sigma_2 = \text{diag}(u_n)$, $u_n = (u_{n,1}, ..., u_{n,p_n}) = |\boldsymbol{\beta}_{n,\text{init2}}|^{-\tau_2}$, $\boldsymbol{\beta}_{n,\text{init1}}$ and $\boldsymbol{\beta}_{n,\text{init2}}$ are two initial estimators of $\boldsymbol{\beta}_n^0$ for some $\tau_1 \geq 0$ and $\tau_2 \geq 0$, and $W_n = \text{diag}(w_n)$, $w_n = (w_{n,1}, ..., w_{n,n}) = (\hat{\sigma}_{n,1}^{-1}, ..., \hat{\sigma}_{n,n}^{-1})$, $\hat{\sigma}_{n,t}$ is a suitable estimator of σ_t. Moreover, let $\hat{\boldsymbol{\alpha}}_n\ (\boldsymbol{\beta}_{n,\text{elastic}}; \mathbf{X}_n, \mathbf{Y}_n)$ and $\hat{\mathbf{L}}_{n,t}(\boldsymbol{\beta}_{n,\text{elastic}}; \mathbf{X}_n, \mathbf{Y}_n)$ be the suitable known plug-in estimators for $\boldsymbol{\alpha}_n^0$ and $\mathbf{L}_{n,t}^0$, respectively. For relevant literature on estimation methods for the conditional variance part, see e.g. [7], [8], [17], and the references therein. For example, if the error process is an ARCH(p) model as in the simulation studies of Section 3, the usual maximum likelihood methods can be applied to estimate the unknown parameters of the conditional variance part based on the residuals from step 2 of the following algorithm.

The iteratively reweighted adaptive elastic net algorithm:

1. Let $k = 1$, $w_n^{[0]} = \mathbf{1}$. Determine the initial values of v_n, u_n, λ_n and γ_n.
2. Calculate the estimator $\boldsymbol{\beta}_n^{[k]} = \boldsymbol{\beta}_{n,\text{elastic}}(\lambda_n, \gamma_n, w_n^{[k-1]})$ of $\boldsymbol{\beta}_n^0$ for Model (2.3) using the weighted adaptive elastic net algorithm (2.4), compute the residuals $\boldsymbol{\varepsilon}_n^{[k]} = \mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}_n^{[k]}$.
3. Estimate the conditional variances $\sigma_{n,t}^{[k]} = g_n(\boldsymbol{\alpha}_n^{[k]}; \mathbf{L}_{n,t}^{[k]})$, $1 \leq t \leq n$, where $\boldsymbol{\alpha}_n^{[k]} = \hat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}_n^{[k]}; \mathbf{X}_n, \mathbf{Y}_n)$, $\mathbf{L}_{n,t}^{[k]} = \hat{\mathbf{L}}_{n,t}(\boldsymbol{\beta}_n^{[k]}; \mathbf{X}_n, \mathbf{Y}_n)$ based on Model (2.2) and the residuals from step 2.
4. Calculate new weights $w_{n,t}^{[k]} = g_n(\boldsymbol{\alpha}_n^{[k]}; \mathbf{L}_{n,t}^{[k]})^{-1}$. Let $w_n^{[k]} = (w_{n,1}^{[k]}, ..., w_{n,n}^{[k]})$.
5. Let $k = k + 1$ and back to step 2 until a specified stopping criterion is satisfied. Return estimate $\boldsymbol{\beta}_n^{[k]}$.

As stated in [17], a plausible stopping criterion should measure the convergence of $\sigma_n^{[k]}$, where $\sigma_n^{[k]} = (\sigma_{n,1}^{[k]}, ..., \sigma_{n,n}^{[k]})'$. One can stop the algorithm if $\|\sigma_n^{[k]} - \sigma_n^{[k-1]}\|_2 < \zeta$ for some small $\zeta > 0$. It is suggested that, under certain conditions, $k = 2$ is sufficient to get an optimal estimator if n is large.

For the two initial estimators $\boldsymbol{\beta}_{n,\text{init1}}$ and $\boldsymbol{\beta}_{n,\text{init2}}$, as stated in [17], there are several options available. When $p_n < n$, one can simply choose the OLS estimator. Alternatively, one can select the lasso estimator as $\boldsymbol{\beta}_{n,\text{init1}}$, the ridge regression estimator as $\boldsymbol{\beta}_{n,\text{init2}}$, or set both $\boldsymbol{\beta}_{n,\text{init1}}$ and $\boldsymbol{\beta}_{n,\text{init2}}$ equal the elastic net estimator.

Next we show the sign consistency and asymptotic normality of the non-vanishing components of $\boldsymbol{\beta}_n^{[k]}$. Let $b_n = \min\{|\boldsymbol{\beta}_n^0(1)|\}$, $W_n^{[k]} = \text{diag}(w_n^{[k]})$, $\widetilde{\mathbf{X}}_n^{[k]} = W_n^{[k-1]}\mathbf{X}_n$, $\widetilde{\mathbf{Y}}_n^{[k]} = W_n^{[k-1]}\mathbf{Y}_n$, $\widetilde{\Gamma}_n^{[k]} = \frac{1}{n}(\widetilde{\mathbf{X}}_n^{[k]})'\widetilde{\mathbf{X}}_n^{[k]}$, $\Gamma_n = \widetilde{\Gamma}_n^{[1]} = \frac{1}{n}\mathbf{X}_n'\mathbf{X}_n$. Let W_n^0 and $\widetilde{\Gamma}_n^0$ be the true matrices, and the submatrices to $\boldsymbol{\beta}_n^0(1)$ are denoted as $\widetilde{\Gamma}_n^{[k]}(1)$, $\widetilde{\Gamma}_n^0(1)$, $\Gamma_n(1)$, $\Sigma_1(1)$ and $\Sigma_2(1)$. Similarly to [12] and [17], we require the following assumptions.

Assumption (A):

(**A1**) $\{Y_t, X_{t,1}, ..., X_{t,m}, \sigma_t\}_{t\in\mathbb{Z}}$ is weakly stationary for all $m \geq 1$, $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. standardized random sequence and $E(Z_t^4) < \infty$, Z_t is independent of $X_{t,\infty}$ for any $t \in \mathbb{Z}$, and $E(\sigma_t^4) < \infty$.

(**A2**) $E(X_{t,i}^2) = 1$ for any $i \geq 1$ and $t \in \mathbb{Z}$.

(**A3**) There is a positive sequence $\{\upsilon_n\}$ such that $\max_{1\leq t\leq n} \|\mathbf{X}_{t,n}(1)\|_2 = \mathcal{O}_p(\upsilon_n\sqrt{q_n})$.

(**A4**) There are constants $a_1 > 0$ and $a_2 > 0$ such that

$$\lim_{n\to\infty} P\Big(a_1 \min\big\{|\boldsymbol{\beta}_{n,\text{init1}}(1)|^{\tau_1}\big\} < b_n\Big) = 0\,,$$
$$\lim_{n\to\infty} P\Big(a_2 \min\big\{|\boldsymbol{\beta}_{n,\text{init2}}(1)|^{\tau_2}\big\} < b_n\Big) = 0\,.$$

(**A5**) There exists a positive sequence $\{r_n\}$ with $r_n \to \infty$ such that

$$\lim_{n\to\infty} P\Big(\max\big\{|\beta_{n,\text{init1}}(2)|^{\tau_1}\big\} \geq {r_n}^{-1}\Big) = 0\,.$$

(**A6**) There are positive constants $\lambda_{0,\min} < \lambda_{0,\max}$ and $\lambda_{1,\min}$ such that the eigenvalues satisfy

$$\lim_{n\to\infty} P\Big(\lambda_{0,\min} < \lambda_{\min}(\Gamma_n(1)) \leq \lambda_{\max}(\Gamma_n(1)) < \lambda_{0,\max}\Big) = 1\,,$$

and

$$\lim_{n\to\infty} P\Big(\lambda_{1,\min} < \lambda_{\min}(\widetilde{\Gamma}_n^0(1)) \leq \lambda_{\max}(\widetilde{\Gamma}_n^0(1))\Big) = 1\,.$$

(**A7**) There are constants $0 < \lambda_{2,\min}$ and $\lambda_{3,\min} > 0$ such that the eigenvalues satisfy

$$\lim_{n\to\infty} P\Big(\lambda_{2,\min} < \lambda_{\min}(D_n) \leq \lambda_{\max}(D_n)\Big) = 1\,,$$

and

$$\lim_{n\to\infty} P\Big(\lambda_{3,\min} < \lambda_{\min}(E_n) \leq \lambda_{\max}(E_n)\Big) = 1\,,$$

where

$$D_n \;=\; \Big(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\,\widetilde{\Gamma}_n^0(1)\,\Big(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1},$$

$$E_n \;=\; \Big(\Gamma_n(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\,\Gamma_n(1)\,\Big(\Gamma_n(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}.$$

(**A8**) There is a positive constant $\sigma_{\min}$ such that

$$0 < \sigma_{\min} < g_n\Big(\hat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}_n;\mathbf{X}_n,\mathbf{Y}_n),\, \hat{\mathbf{L}}_{n,t}(\boldsymbol{\beta}_n;\mathbf{X}_n,\mathbf{Y}_n)\Big)\,, \qquad 1 \leq t \leq n\,,$$

for all large enough n and $\boldsymbol{\beta}_n$ in an open neighbourhood of $\boldsymbol{\beta}_n^0$.

(**A9**) For all n and any $1 \leq t \leq n$, the estimators $\hat{\boldsymbol{\alpha}}_n$ and $\hat{\mathbf{L}}_{n,t}$ are consistent for $\boldsymbol{\alpha}_n^0$ and $\mathbf{L}_{n,t}^0$, and there is a sequence $\{h_n\}$ with $h_n n^{-1/2} \to 0$ such that

$$\max_{1\leq t\leq n}\left|g\big(\boldsymbol{\alpha}_\infty^0;\mathbf{L}_{\infty,t}\big)^{-2} - g_n\Big(\hat{\boldsymbol{\alpha}}_n(\boldsymbol{\beta}_n^0;\mathbf{X}_n,\mathbf{Y}_n),\, \hat{\mathbf{L}}_{n,t}(\boldsymbol{\beta}_n^0;\mathbf{X}_n,\mathbf{Y}_n)\Big)^{-2}\right| = \mathcal{O}_p\big(h_n/\sqrt{n}\big).$$

(**A10**) There are positive constants C_1, C_2 and d with $1 \le d \le 2$ such that, for any $t \in Z$,

$$P\big(|\varepsilon_t| > x\big) \le C_1 \exp\big(-C_2 x^d\big)\,.$$

(**A11**)

① $\frac{(\log n)^{I\{d=1\}}(\log(1+q_n))^{1/d}}{\sqrt{n}b_n} \to 0$, ② $\frac{h_n}{\sqrt{n}b_n} \to 0$,

③ $\frac{\lambda_n\sqrt{q_n}}{\sqrt{n}b_n} \to 0$, ④ $\frac{\sqrt{n}(\log n)^{I\{d=1\}}(\log(1+p_n-q_n))^{1/d}}{\lambda_n r_n} \to 0$,

⑤ $\frac{h_n\sqrt{n}}{\lambda_n r_n} \to 0$, ⑥ $\frac{\sqrt{q_n}}{b_n r_n} \to 0$,

⑦ $\frac{\upsilon_n\sqrt{q_n}}{\sqrt{n}} \to 0$, ⑧ $\frac{h_n\sqrt{q_n}}{\sqrt{n}} \to 0$,

⑨ $\frac{\gamma_n\left((\log n)^{I\{d=1\}}+h_n\right)}{\sqrt{n}b_n} \to 0$.

Similar assumptions as in (A1)–(A11) are also imposed in [17] to study the asymptotic behaviour of the iteratively reweighted adaptive lasso algorithm. Assumption (A1) is standard for variable selection in a time series setting. Assumption (A2) is the usual scale standardization required in a lasso setting without loss of generality (see e.g. [6]), because $\{X_{t,i}\}$ is stationary and hence its mean and variance are constants. Assumption (A3) characterises the structure of regressors. For instance, if $\{X_{t,n}(1)\}$ is stationary and $\boldsymbol{\beta}_n^0$ contains a finite number of non-zero components, then we can choose $\upsilon_n = O_P(1)$ for Assumption (A3) to hold. Assumptions (A4) and (A5) actually assume that the weights υ_n and u_n are not too large for $\beta_j^0 \neq 0$ and not too small for $\beta_j^0 = 0$. They also mean that the initial estimators can distinguish between zero and non-zero components of the parameter vector well. For the Lasso initial estimators, Assumptions (A4) and (A5) can be derived from sharp thresholds and sign consistency of the Lasso estimate under some additional mild assumptions (see, e.g., [13] and [16]). Assumption (A6) is needed to address heteroscedasticity in high-dimensional regression models (see, for example, [3]). Since we deal with the weighted adaptive elastic net algorithm, additional similar assumptions such as (A7) are also needed here. It is worth mentioning that, under certain conditions, $D_n - \widetilde{\Gamma}_n^0(1) \to 0$ and $E_n - \Gamma_n(1) \to 0$ as $n \to \infty$. Assumptions (A8) and (A9) are standard in heteroscedastic regression and Assumption (A10) excludes heavy-tailed errors.

Assumption (A11) postulate properties required for deriving the asymptotics of the proposed estimator. As a simple example, to better understand Assumption (A11) assume b_n to be fixed and $d = 1$, ① and ② permit $h_n \sim 1$ and $q_n \sim n^{1/2+\delta}$ for any $0 < \delta < 1/4$. With these choices we can choose $\lambda_n \sim n^{1/4-\delta}$, $r_n \sim n^{1/2+\delta}$, $\upsilon_n \sim 1$ and $\gamma_n \sim n^{1/4-\delta}$ by ③, ④, ⑦ and ⑨, and p_n can grow with every polynomial order. Obviously these selections satisfy Assumption (A11), and also Assumptions (A3)–(A5) and (A9). Moreover, by ④ and ⑨, we obtain $\frac{\gamma_n}{b_n\lambda_n r_n} \to 0$ as $n \to \infty$.

The following theorem shows the sign consistency and asymptotic normality of the estimator. The proof will be given in the Appendix. The sign consistency introduced by [16] is stronger than the usual selection consistency which only requires the zeros to be matched, but not the signs. The reason for using sign consistency is to avoid dealing with situations where a model is estimated with matching zeros but reversed signs.

Theorem 2.1. *Under Assumption (A), it holds for all $k \geq 1$ that:*

(**1**) *(Sign consistency)*

$$\lim_{n\to\infty} P\Big(\operatorname{sign}(\boldsymbol{\beta}_n^{[k]}) = \operatorname{sign}(\boldsymbol{\beta}_n^0)\Big) = 1\,,$$

where $\operatorname{sign}(\cdot)$ *maps positive entry to 1, negative entry to -1 and zero to zero, that is, $\boldsymbol{\beta}_n^{[k]}$ asymptotically matches the zeros and signs of $\boldsymbol{\beta}_n^0$ with probability one.*

(**2**) *(Asymptotic normality)*

$$\sqrt{n}\,(s_n(k))^{-1}\,\boldsymbol{\xi}_n'\Big(\boldsymbol{\beta}_n^{[k]}(1) - \boldsymbol{\beta}_n^0(1)\Big) \xrightarrow{\mathcal{D}} Z\,,$$

where $\boldsymbol{\xi}_n \in \mathbb{R}^{q_n}$ with $\|\boldsymbol{\xi}_n\|_2 = 1$, $s_n^2(1) = \boldsymbol{\xi}_n' E_n \boldsymbol{\xi}_n$ and $s_n^2(k) = \boldsymbol{\xi}_n' D_n \boldsymbol{\xi}_n$ for $k \geq 2$.

3. SIMULATION STUDIES

In this section, we provide simulation studies to check the finite sample performance of the iteratively reweighted adaptive elastic net algorithm (IRAEN) for an AR-ARCH model. The comparison with the iteratively reweighted adaptive Lasso algorithm (IRAL) introduced in [17] is also considered.

We consider the following AR-ARCH model

$$Y_t = \textstyle\sum_{i\in I} \phi_i Y_{t-i} + \varepsilon_t\,,$$

and

$$\varepsilon_t = \sigma_t Z_t\,, \qquad \sigma_t = \sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2}\,,$$

where the true values of the parameters are $\alpha_0 = 0.02$ and $\alpha_1 = \alpha_2 = 0.49$, $Z_t \sim N(0,1)$, $\phi_i = 0.95(\phi^{-1}-1)\phi^{\sqrt{i}}$, $\phi = 0.85$, $I = \{1,4,9,16,...\}$. It is easy to see that $\sum_{i\in I}|\phi_i| = 0.95 < 1$ and $\sum_{i\in I}\phi_i^2 < \infty$ which imply the stationarity of Y_t. Note that, by the properties of the AR-ARCH model, $EY_t^2 = E\sigma_t^2 = \alpha_0/(1-\alpha_1-\alpha_2) = 1$. This implies that Assumption (A2) is satisfied.

Let $p_n = [2\sqrt{n}]$ and $q_n = [\sqrt{p_n}]$, where n is the sample size. For example, when $n = 500$, $p_n = 44$, $q_n = 6$ and $I = \{1,4,9,16,25,36\}$. If $n = 1000$, then $p_n = 63$, $q_n = 7$ and $I = \{1,4,9,16,25,36,49\}$.

After generating data from the above AR-ARCH model with sample size $n = 500$ and $n = 1000$, respectively, we use two methods, IRAEN and IRAL, to estimate the parameters ϕ_i and to check the sign consistency of the estimators. In the simulations, we use the C_p criterion to choose the appropriate λ_n and γ_n. The two initial estimators $\boldsymbol{\beta}_{n,\mathrm{init1}}$ and $\boldsymbol{\beta}_{n,\mathrm{init2}}$ are chosen to be the OLS estimator.

3.1. The iteratively reweighted adaptive elastic net algorithm

To apply the proposed iteratively reweighted elastic net algorithm, we consider two cases: the homoscedastic case ($k = 1$) and the heteroscedastic case with one additional replication ($k = 2$).

For the $k = 1$ case, Table 1 reports the estimation results for two sample sizes $n = 500$ and $n = 1000$ based on 1000 replications. We hope that the covariates with non-zero coefficients (relevant parameters) can be selected from the estimation procedure, but the covariates with zero coefficients (irrelevant parameters) shouldn't be included. Table 1 shows the proportions of both the relevant and irrelevant included parameters of all estimated parameters for the homoscedastic case. Proportion 1 (the accuracy rate) denotes the proportion of the relevant included parameters and Proportion 2 (the error rate) is the proportion of the irrelevant included parameters. The number of times each parameter has been selected during 1000 simulations are also reported.

Table 1: Proportions of relevant included parameters and irrelevant included parameters for the case of $k = 1$ using IRAEN.

sample size	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	...	Proportion 1	Proportion 2
$n = 500$	866	691	622	834	568	506	454	482	864	...	81.25%	33.61%
$n = 1000$	932	721	668	902	575	542	502	502	927	...	89.41%	29.66%

It is seen from Table 1 that the accuracy rate increases with larger sample size n, while the error rate decreases in n. This is consistent with the theoretical results in Theorem 2.1.

In a similar way, we apply the proposed iteratively reweighted elastic net algorithm with $k = 2$. Proportions of both the relevant and irrelevant included parameters of all estimated parameters and the number of times each parameter has been selected during 1000 simulations for the heteroscedastic case are given in Table 2. Inspection of Table 2 reveals that, as in the $k = 1$ case, the accuracy rate increases with larger sample size n, while the error rate decreases in n. Comparing two tables, we conclude that the heteroscedastic case with $k = 2$ has better selection properties than the homoscedastic case $k = 1$ for the conditional heteroscedastic models.

Table 2: Proportions of relevant included parameters and irrelevant included parameters for the case of $k = 2$ using IRAEN.

sample size	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	...	Proportion 1	Proportion 2
$n = 500$	938	415	395	969	383	363	322	345	979	...	94.92%	29.54%
$n = 1000$	992	407	371	999	377	287	316	294	999	...	98.77%	21.98%

Moreover, the plots in Figure 1 show the selection results for both the $k = 1$ and $k = 2$ cases from one simulation with $n = 500$. For each plot, the vertical axis represents the values of the estimated coefficients, the horizontal axis (bottom) represents the values of $\ln \lambda_n$, and the top shows the numbers of the non-zero coefficients selected for different values of $\ln \lambda_n$. The 44 curves illustrate the change of the values of 44 estimated coefficients with $\ln \lambda_n$ changing. Note that there are only six non-zero positive coefficients in the true model.

It can be seen that, when $k = 2$, these six coefficients tend to zero from positive side, while when $k = 1$, there exist some coefficients tending to zero from negative side, which means that no matter what value $\ln \lambda_n$ takes, the sign consistency may not be satisfied. This is consistent with the conclusions drawn from the comparison of Tables 1 and 2. Figure 1 again visually displays that the heteroscedastic algorithm with $k = 2$ outperforms its homoscedastic counterpart.

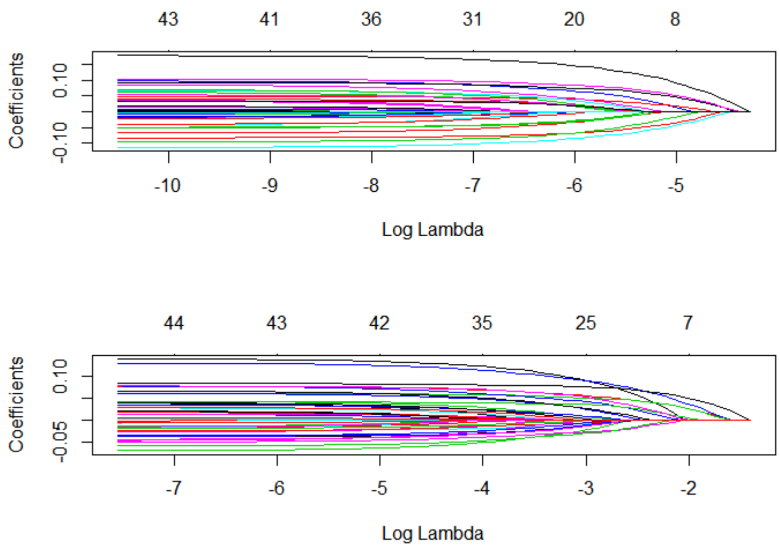


Figure 1: Estimated coefficients for different λ_n values with $n = 500$ and $k = 1$ (upper) or $k = 2$ (lower) using IRAEN.

3.2. The iteratively reweighted adaptive Lasso algorithm

Next we report the estimation results using the iteratively reweighted adaptive Lasso algorithm. Proportions of both the relevant and irrelevant included parameters and the number of times each parameter has been selected during 1000 simulations for the homoscedastic case and the heteroscedastic case with $n = 500$ and $n = 1000$ are given in Tables 3 and 4, respectively. Figure 2 shows the selection results for both the $k = 1$ and $k = 2$ cases from one simulation with $n = 500$. Similarly to the IRAEN algorithm, Tables 3–4 and Figure 2 indicate that the heteroscedastic algorithm with $k = 2$ outperforms its homoscedastic counterpart.

Comparing Tables 1 and 2 with Tables 3 and 4, it is clear that the IRAEN algorithm proposed in this paper uniformly improves the accuracy rate as compared to the IRAL method, while the error rate is increased as a price to pay for using IRAEN algorithm. This implies that the IRAL method excludes irrelevant variables more thoroughly. It is also consistent with the conclusions of [19]. That is, if the covariates have grouping effect (a group of variables among which the pairwise correlations are very high), then the IRAL algorithm tends to arbitrarily select only one variable from the group, while the IRAEN algorithm has the capacity of selecting groups of correlated variables. Generally speaking, the IRAEN algorithm produces a sparse model with good estimation accuracy, while encouraging a grouping effect.

This makes the IRAEN algorithm particularly useful for estimating the models containing several correlated variables such as the AR-ARCH type processes.

Table 3: Proportions of relevant included parameters and irrelevant included parameters for the case of $k = 1$ using IRAL.

sample size	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	...	Proportion 1	Proportion 2
$n = 500$	871	682	619	863	515	501	481	452	827	...	80.20%	31.95%
$n = 1000$	927	710	612	893	555	519	506	471	931	...	88.37%	27.80%

Table 4: Proportions of relevant included parameters and irrelevant included parameters for the case of $k = 2$ using IRAL.

sample size	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	...	Proportion 1	Proportion 2
$n = 500$	930	399	358	971	338	296	327	313	974	...	93.40%	26.35%
$n = 1000$	995	349	325	997	305	277	253	252	1000	...	98.66%	19.98%

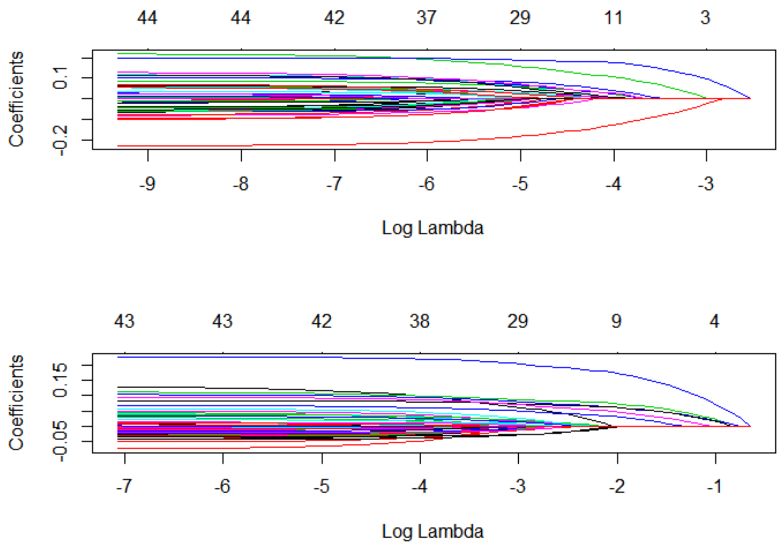


Figure 2: Estimated coefficients for different λ_n values with $n = 500$ and $k = 1$ (upper) or $k = 2$ (lower) using IRAL.

A. APPENDIX

Proof of Theorem 2.1: The basic ideas of the proof are mainly from [12] and [17]. Since we are dealing with the elastic net algorithm, we need some extra steps to achieve our goal.

Let $\|X\|_{\psi_d} = \inf\left\{C > 0 \,|\, E[\psi_d(|X|/C)] \le 1\right\}$ be the Orlicz norm of a random variable X, where $\psi_d(x) = \exp(x^d) - 1$, $1 \le d \le 2$. Denote $e_{n,j}$ the j-th unit vector in $\mathbb{R}^{q_n}$. For any vector a and b, $a =_{\mathrm{s}} b$ means that $\mathrm{sign}(a) = \mathrm{sign}(b)$. Let $k \ge 2$, the case $k = 1$ can be proved in a similar way.

(I) The sign consistency

The Karush–Kuhn–Tucker (KKT) conditions yield that $(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta})'(W_n^{[k-1]})^2\,(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}) + \lambda_n\|\Sigma_1\boldsymbol{\beta}\|_1 + \gamma_n\|\Sigma_2^{1/2}\boldsymbol{\beta}\|_2^2$ is minimised by $\boldsymbol{\beta} = (\boldsymbol{\beta}(1)', \mathbf{0}')'$ if and only if

$$\mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}) - \gamma_n u_{n,j}\,\beta_j = \frac{\lambda_n}{2}\,v_{n,j}\,\mathrm{sign}(\beta_j)\,, \qquad \text{if} \quad \beta_j \ne 0\,, \tag{A.1}$$

$$\left|\mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta})\right| < \frac{\lambda_n}{2}\,v_{n,j}\,, \qquad \text{if} \quad \beta_j = 0\,, \tag{A.2}$$

where $\mathbf{X}_j^0$ is the j-th column of $\mathbf{X}_n$. Let

$$\begin{aligned} \delta_n^{[k]}(1) \;=\; & \boldsymbol{\beta}_n^0(1) + \frac{1}{n}\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'\,(W_n^{[k-1]})^2\,\boldsymbol{\varepsilon}_n^0 \\ & - \frac{\lambda_n}{2n}\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1} s_n^0(1) \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\beta}_n^{[k]}(1) \;=\; & \Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\,\widetilde{\Gamma}_n^{[k]}(1)\,\boldsymbol{\beta}_n^0(1) \\ & + \frac{1}{n}\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'\,(W_n^{[k-1]})^2\,\boldsymbol{\varepsilon}_n^0 \\ & - \frac{\lambda_n}{2n}\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1} s_n^0(1)\,, \end{aligned} \tag{A.3}$$

where $s_n^0(1) = \Sigma_1(1)\,\mathrm{sign}(\boldsymbol{\beta}_n^0(1))$. In addition, let $\delta_n^{[k]} = (\delta_n^{[k]}(1)', \mathbf{0}')'$ and $\boldsymbol{\beta}_n^{[k]} = (\boldsymbol{\beta}_n^{[k]}(1)', \mathbf{0}')'$.

First we show

$$\lim_{n\to\infty} P\Big(\boldsymbol{\beta}_n^0 \ne_{\mathrm{s}} \delta_n^{[k]}\Big) = 0\,. \tag{A.4}$$

Let $\eta_{1,j} = e_{n,j}'\big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\big)^{-1}\mathbf{X}_n(1)'(W_n^{[k-1]})^2\boldsymbol{\varepsilon}_n^0$, $\eta_{2,j} = e_{n,j}'\big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\big)^{-1}s_n^0(1)$, and let $A_1 = \big\{\frac{1}{n}|\eta_{1,j}| \ge \frac{1}{2}|\beta_j^0|$, for some $j \le q_n\big\}$ and $A_2 = \big\{\frac{\lambda_n}{n}|\eta_{2,j}| \ge |\beta_j^0|$, for some $j \le q_n\big\}$. Thus, to prove (A.4), it is enough to show that $P(A_j) \to 0$ as $n \to \infty$ for $j = 1, 2$.

For $P(A_1)$, we obtain

$$\begin{aligned} P(A_1) &\le P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta_{1,j}|\ge\frac{b_n}{2}\bigg)\\ &\le P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta^{0,\infty}_{1,j}|\ge\frac{b_n}{4}\bigg)+P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta_{1,j}-\eta^0_{1,j}|\ge\frac{b_n}{8}\bigg)\\ &\quad+P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta^0_{1,j}-\eta^{0,\infty}_{1,j}|\ge\frac{b_n}{8}\bigg), \end{aligned}\tag{A.5}$$

where $\eta^0_{1,j}=e'_{n,j}(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W^0_n)^2\boldsymbol{\varepsilon}^0_n$ and $\eta^{0,\infty}_{1,j}=e'_{n,j}(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'\cdot$ $\cdot\,(W^0_n)^2\,\boldsymbol{\varepsilon}^0_{n,\infty}$, and $\boldsymbol{\varepsilon}^0_{n,\infty}$ is the restriction of the true error $\boldsymbol{\varepsilon}^0_\infty=(\varepsilon^0_1,\varepsilon^0_2,\ldots)'$ in Model (2.1).

Regarding the first term of (A.5), Assumptions (A6), (A8) and (A9) imply that $\|W^0_n\|_2\le\sigma^{-1}_{\min}$ and $\|\Gamma_n(1)\|_2\le\lambda_{0,\max}$. Note that$\lambda_{1,\min}<\lambda(\widetilde{\Gamma}^0_n(1))$ and $0\le\lambda(\frac{\gamma_n}{n}\Sigma_2(1))$, then $\lambda_{1,\min}<\lambda(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1))$. That is, $\lambda((\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1})\le\lambda^{-1}_{1,\min}$ and hence $\|(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\|\le\lambda^{-1}_{1,\min}$. Thus we arrive at

$$\begin{aligned} &\bigg\|\frac{1}{\sqrt{n}}\,e'_{n,j}\Big(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'(W^0_n)^2\bigg\|_2\\ &\le\bigg\|\Big(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\bigg\|_2\,\bigg\|\frac{1}{\sqrt{n}}\mathbf{X}_n(1)\bigg\|_2\,\|(W^0_n)^2\|_2\;\le\;\lambda^{-1}_{1,\min}\sqrt{\lambda_{0,\max}}\,\sigma^{-2}_{\min}\,. \end{aligned}\tag{A.6}$$

This implies that, as $n\to\infty$,

$$P\Bigg(\bigg\|\frac{1}{\sqrt{n}}\,e'_{n,j}\Big(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}X_n(1)'(W^0_n)^2\bigg\|_2\le\lambda^{-1}_{1,\min}\sqrt{\lambda_{0,\max}}\,\sigma^{-2}_{\min}\Bigg)\to 1\,.$$

This, together with Lemma 1(i) of [6] and Assumption (A10), yields that

$$\bigg\|\frac{1}{\sqrt{n}}\,\eta^{0,\infty}_{1,j}\bigg\|_{\psi_d}=\bigg\|\frac{1}{\sqrt{n}}\,e'_{n,j}\Big(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}X_n(1)'(W^0_n)^2\,\varepsilon^0_{n,\infty}\bigg\|_{\psi_d}\le C\,(\log n)^{I\{d=1\}}\,.\tag{A.7}$$

Combining this with Equation (16) of [17], we obtain

$$P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta^{0,\infty}_{1,j}|\ge\frac{b_n}{4}\bigg)\le\psi^{-1}_d\Bigg(\frac{b_n\sqrt{n}}{4C\,\big(\log(1+q_n)\big)^{1/d}\,(\log n)^{I\{d=1\}}}\Bigg).\tag{A.8}$$

Now it follows from Assumption (A11) that

$$P\bigg(\frac{1}{n}\max_{1\le j\le q_n}|\eta^{0,\infty}_{1,j}|\ge\frac{b_n}{4}\bigg)\to 0\,.$$

For the second term of (A.5), Assumptions (A8) and (A9) ensure that $\|W^{[k-1]}_n\|_2=\mathcal{O}_p(1)$ and $\|(W^0_n)^2-(W^{[k-1]}_n)^2\|_2=\mathcal{O}_p(\frac{h_n}{\sqrt{n}})$. Furthermore, we notice that $\|\boldsymbol{\varepsilon}^0_n\|_2\le\|\boldsymbol{\varepsilon}^0_n-\boldsymbol{\varepsilon}^0_{n,\infty}\|_2+\|\boldsymbol{\varepsilon}^0_{n,\infty}\|_2$, while $\|\boldsymbol{\varepsilon}^0_n-\boldsymbol{\varepsilon}^0_{n,\infty}\|_2\xrightarrow{\mathcal{P}}0$, and Assumption (A1) and the weak law of large numbers yield that $\|\boldsymbol{\varepsilon}^0_{n,\infty}\|_2=\mathcal{O}_p(\sqrt{n})$. This bound implies that $\|\boldsymbol{\varepsilon}^0_n\|_2=\mathcal{O}_p(\sqrt{n})$.

On the other hand, since

$$\begin{aligned} \bigg\|\Big(\widetilde{\Gamma}^0_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)-\Big(\widetilde{\Gamma}^{[k]}_n(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)\bigg\|_2&=\Big\|\widetilde{\Gamma}^0_n(1)-\widetilde{\Gamma}^{[k]}_n(1)\Big\|_2\\ &=\big\|\Gamma_n(1)\big\|_2\,\Big\|(W^0_n)^2-(W^{[k-1]}_n)^2\Big\|_2=\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big), \end{aligned}$$

we obtain

$$\begin{aligned}\left\|A^{-1}-(A+B)^{-1}\right\|_2 &\le \left\|A^{-1}-(A+B)^{-1}+A^{-1}BA^{-1}\right\|_2+\left\|A^{-1}BA^{-1}\right\|_2\\ &\le O_p(\|B\|_2)+\left\|A^{-1}\right\|_2^2\,\|B\|_2 \;=\; \mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big),\end{aligned}$$

where $A=\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)$ and $B=\big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\big)-\big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\big)$. That is,

$$\left\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}-\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\| \;=\; \mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big). \tag{A.9}$$

We conclude that, for all $1\le j\le q_n$,

$$\begin{aligned}\left|\eta_{1,j}-\eta_{1,j}^0\right| &= \bigg|e_{n,j}'\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'\left((W_n^0)^2-(W_n^{[k-1]})^2\right)\boldsymbol{\varepsilon}_n^0\\ &\quad+e_{n,j}'\bigg(\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)-\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)\bigg)\mathbf{X}_n(1)'(W_n^{[k-1]})^2\,\boldsymbol{\varepsilon}_n^0\bigg|\\ &\le \left\|n\Gamma_n(1)\right\|_2^{1/2}\left\|\boldsymbol{\varepsilon}_n^0\right\|_2\bigg\{\left\|(W_n^0)^2-(W_n^{[k-1]})^2\right\|_2\left\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2\\ &\quad+\left\|(W_n^{[k-1]})^2\right\|_2\left\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}-\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2\bigg\}\\ &= \mathcal{O}_p\big(\sqrt{n}\big)\,\mathcal{O}_p\big(\sqrt{n}\big)\,\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,\mathcal{O}_p(1)\\ &= \mathcal{O}_p\big(h_n\sqrt{n}\big)\,.\end{aligned}$$

Thus it follows from Assumption (A11) that $P\big(\frac{1}{n}\max_{1\le j\le q_n}|\eta_{1,j}-\eta_{1,j}^0|\ge\frac{b_n}{8}\big)\le P\big(\frac{h_n}{\sqrt{n}b_n}\ge C\big)$ $\to 0$ as $n\to\infty$.

We proceed to deal with the third term of (A.5). By (A.6),

$$\begin{aligned}\frac{1}{\sqrt{n}}\left|\eta_{1,j}^0-\eta_{1,j}^{0,\infty}\right| &= \frac{1}{\sqrt{n}}\left|e_{n,j}'\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'(W_n^0)^2\left(\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\right)\right|\\ &\le \lambda_{1,\min}^{-1}\sqrt{\lambda_{0,\max}}\,\sigma_{\min}^{-2}\left\|\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2 \;\xrightarrow{\mathcal{P}}\; 0\,.\end{aligned}$$

Hence, by Assumption (A11), we have

$$P\bigg(\frac{1}{n}\max_{1\le j\le q_n}\left|\eta_{1,j}^0-\eta_{1,j}^{0,\infty}\right|\ge\frac{b_n}{8}\bigg)\;\le\;P\bigg(\frac{1}{\sqrt{n}\,b_n}\max_{1\le j\le q_n}\frac{1}{\sqrt{n}}\left|\eta_{1,j}^0-\eta_{1,j}^{0,\infty}\right|\ge\frac{1}{8}\bigg)\;\to\;0\,.$$

Then (A.5) implies that $P(A_1)\to 0$ as $n\to\infty$. In order to prove $P(A_2)\to 0$ as $n\to\infty$, we examine the bound of $\|(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\|_2$. By (A.9) and Weyl's perturbation theorem for eigenvalues of the matrices, for all $1\le j\le q_n$,

$$\begin{aligned}&\left|\lambda_j\bigg(\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\bigg)-\lambda_j\bigg(\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\bigg)\right|\\ &\qquad\le\left\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}-\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2=\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,.\end{aligned}$$

Therefore $\big\|\big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\big)^{-1}\big\|_2 \le \lambda_{1,\min}^{-1} + C$ with probability arbitrarily close to 1 for sufficiently large n. It follows from Assumptions (A4), (A6) and (A11) that

$$\begin{aligned} P(A_2) \;&\le\; P\left(\frac{\lambda_n}{n}\max_{1\le j\le q_n}|\eta_{2,j}| \ge b_n\right)\\ &\le\; P\left(\frac{\lambda_n}{n}\left\|\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1}\right\|_2 \left\|s_n^0(1)\right\|_2 \ge b_n\right)\\ &\le\; P\left(\frac{\lambda_n\sqrt{q_n}}{n\,b_n^2} \ge C\right) \;\to\; 0 \end{aligned}$$

due to the fact that $\|s_n^0(1)\| \le \|\Sigma_1(1)\|_2\,\|\operatorname{sign}(\boldsymbol{\beta}_n^0(1))\|_2 \le \frac{b_1\sqrt{q_n}}{b_n} = \mathcal{O}_p\big(\frac{\sqrt{q_n}}{b_n}\big)$.

This completes the proof of (A.4). We now turn to show that

$$\lim_{n\to\infty} P\left(\delta_n^{[k]} \ne_{\mathrm{s}} \boldsymbol{\beta}_n^{[k]}\right) \;=\; 0\,. \tag{A.10}$$

Observe that

$$\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1} = \left(\widetilde{\Gamma}_n^{[k]}(1)\right)^{-1} - \left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1}\frac{\gamma_n}{n}\Sigma_2(1)\left(\widetilde{\Gamma}_n^{[k]}(1)\right)^{-1}. \tag{A.11}$$

Then, by Assumptions (A4) and (A11),

$$\begin{aligned} \left\|\boldsymbol{\beta}_n^{[k]} - \delta_n^{[k]}\right\|_2 \;&=\; \left\|\left[\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1}\widetilde{\Gamma}_n^{[k]}(1) - I_{q_n}\right]\boldsymbol{\beta}_n^0(1)\right\|_2\\ &=\; \left\|-\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1}\frac{\gamma_n}{n}\Sigma_2(1)\,\boldsymbol{\beta}_n^0(1)\right\|_2\\ &\le\; \frac{\gamma_n}{n}\left\|\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\right)^{-1}\right\|_2 \left\|\Sigma_2(1)\right\|_2\left\|\boldsymbol{\beta}_n^0(1)\right\|_2\\ &=\; \mathcal{O}_p\left(\frac{\gamma_n}{n\,b_n}\right) \;\xrightarrow{\mathcal{P}}\; 0\,. \end{aligned}$$

This implies (A.10). Combining (A.4) and (A.10) leads to

$$\lim_{n\to\infty} P\left(\boldsymbol{\beta}_n^0 \ne_{\mathrm{s}} \boldsymbol{\beta}_n^{[k]}\right) \;=\; 0\,. \tag{A.12}$$

Hence, to prove the sign consistency of the iteratively reweighted adaptive elastic net estimator, it suffices to show that, as $n\to\infty$, $\boldsymbol{\beta}_n^{[k]}$ satisfies the KKT conditions (A.1) and (A.2), so that $\boldsymbol{\beta}_n^{[k]}$ is indeed the solution of (2.4).

The above arguments for proving (A.4) and (A.10) imply that

$$\left\|\boldsymbol{\beta}_n^{[k]} - \boldsymbol{\beta}_n^0(1)\right\|_2 \;=\; \mathcal{O}_p\left(\frac{\gamma_n}{n\,b_n} + \frac{(\log n)^{I\{d=1\}} + h_n}{\sqrt{n}} + \frac{\lambda_n\sqrt{q_n}}{n\,b_n}\right). \tag{A.13}$$

From (A.3), (A.11)–(A.13) and Assumption (A11), for $1 \le j \le q_n$,

$$\begin{aligned}
&\mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\left(\mathbf{Y}_n - \mathbf{X}_n\,\boldsymbol{\beta}_n^{[k]}(1)\right) - \gamma_n\,u_{n,j}\,\beta_{n,j}^{[k]}\\
&\quad= \mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,\boldsymbol{\varepsilon}_n^0 + \mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,\mathbf{X}_n(1)\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\right)^{-1}\frac{\gamma_n}{n}\,\Sigma_2(1)\,\boldsymbol{\beta}_n^0(1)\\
&\quad\quad - \mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,\mathbf{X}_n(1)\,\frac{1}{n}\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\right)^{-1}\mathbf{X}_n(1)'\,(W_n^{[k-1]})^2\,\boldsymbol{\varepsilon}_n^0\\
&\quad\quad + \mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\,\mathbf{X}_n(1)\,\frac{\lambda_n}{2n}\left(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\right)^{-1} s_n^0(1) - \gamma_n\,u_{n,j}\,\beta_{n,j}^{[k]}\\
&\quad= \gamma_n\,u_{n,j}\,\beta_j + \frac{\lambda_n}{2}\,v_{n,j}\,\mathrm{sign}(\beta_j) - \gamma_n\,u_{n,j}\,\beta_{n,j}^{[k]} + \mathcal{O}_p\left(\frac{\gamma_n}{\sqrt{n}\,b_n}\right)\\
&\quad= \frac{\lambda_n}{2}\,v_{n,j}\,\mathrm{sign}\big(\beta_{n,j}^{[k]}\big) + \mathcal{O}_p\left(\frac{\gamma_n}{\sqrt{n}\,b_n} + \frac{\gamma_n\big((\log n)^{I\{d=1\}} + h_n\big)}{\sqrt{n}\,b_n} + \frac{\gamma_n\lambda_n\sqrt{q_n}}{n\,{b_n}^2}\right).
\end{aligned}$$

This means that $\boldsymbol{\beta}_n^{[k]}$ satisfies the first KKT condition (A.1) as $n \to \infty$.

Let $\eta_{3,j} = \mathbf{X}_j^{0\prime}(W_n^{[k-1]})^2\big[I_n - \frac{1}{n}\mathbf{X}_n(1)(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W_n^{[k-1]})^2\big]\boldsymbol{\varepsilon}_n^0$ and $\eta_{4,j} = \frac{\lambda_n}{2n}\mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\mathbf{X}_n(1)(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}s_n^0(1) + \mathbf{X}_j^{0\prime}\,(W_n^{[k-1]})^2\mathbf{X}_n(1)\,\big[(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}\,\widetilde{\Gamma}_n^{[k]}(1) - I_{q_n}\big]\boldsymbol{\beta}_n^0(1)$. Denote $A_3 = \big\{|\eta_{3,j}| \ge \frac{\lambda_n}{4}v_{n,j},\ \text{for some } j > q_n\big\}$ and $A_4 = \big\{|\eta_{4,j}| \ge \frac{\lambda_n}{4}v_{n,j},$ for some $j > q_n\big\}$.

Then, to show that $\boldsymbol{\beta}_n^{[k]}$ satisfies the second KKT condition (A.2), we only need to prove that $P\big(|\eta_{3,j} - \eta_{4,j}| < \frac{\lambda_n}{2}\,v_{n,j}\big) \to 0$ as $n \to \infty$ for any $q_n < j \le p_n$. So it is enough to show that $P(A_j) \to 0$ as $n \to \infty$ for $j = 3, 4$.

Let $\eta_{3,j}^0 = \mathbf{X}_j^{0\prime}(W_n^0)^2\big[I_n - \frac{1}{n}\mathbf{X}_n(1)(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W_n^0)^2\big]\boldsymbol{\varepsilon}_n^0$ and $\eta_{3,j}^{0,\infty} = \mathbf{X}_j^{0\prime}(W_n^0)^2\,\big[I_n - \frac{1}{n}\mathbf{X}_n(1)(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W_n^0)^2\big]\boldsymbol{\varepsilon}_{n,\infty}^0$. Then

$$\begin{aligned}
P(A_3) \;\le\;& P\left(\max_{q_n<j\le p_n}\big|\eta_{3,j}^{0,\infty}\big| \ge \frac{\lambda_n r_n}{8}\right) + P\left(\max_{q_n<j\le p_n}\big|\eta_{3,j} - \eta_{3,j}^0\big| \ge \frac{\lambda_n r_n}{16}\right)\\
&+ P\left(\max_{q_n<j\le p_n}\big|\eta_{3,j}^0 - \eta_{3,j}^{0,\infty}\big| \ge \frac{\lambda_n r_n}{16}\right)\\
&+ P\left(\max_{q_n<j\le p_n}\big|\beta_{j,\mathrm{init1}}\big|^{\tau_1} \ge \frac{1}{r_n}\right),
\end{aligned}\tag{A.14}$$

where $\beta_{j,\mathrm{init1}}$ is the j-th element of $\boldsymbol{\beta}_{n,\mathrm{init1}}$.

For estimating the first term of (A.14), let $H_{n,j}^0 = \mathbf{X}_j^{0\prime}(W_n^0)^2\big[I_n - \frac{1}{n}\mathbf{X}_n(1)(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1))^{-1}\,\mathbf{X}_n(1)'(W_n^0)^2\big]$. Thus we have $\eta_{3,j}^{0,\infty} = H_{n,j}^0\,\boldsymbol{\varepsilon}_{n,\infty}^0$. Note that

$$\begin{aligned}
\|H_{n,j}^0\|_2 \;\le\;& \|\mathbf{X}_j^0\|_2\,\|(W_n^0)^2\|_2\left[1 + \left\|\frac{1}{n}\,\mathbf{X}_n(1)\left(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\,\Sigma_2(1)\right)^{-1}\mathbf{X}_n(1)'\right\|_2\,\|(W_n^0)^2\|_2\right]\\
=\;& \mathcal{O}_p\big(\sqrt{n}\big)\,.
\end{aligned}$$

In the same way as in (A.7) and (A.8), by Assumption (11), we obtain

$$P\left(\max_{q_n<j\le p_n}\big|\eta_{3,j}^{0,\infty}\big| \ge \frac{\lambda_n r_n}{8}\right) \le \psi_d^{-1}\left(\frac{\lambda_n r_n}{c_7\sqrt{n}\big(\log\,(1+p_n-q_n)\big)^{1/d}(\log n)^{I\{d=1\}}}\right) \longrightarrow 0\,. \tag{A.15}$$

Since

$$\begin{aligned}\left|\eta_{3,j}-\eta_{3,j}^{0}\right| \;=\; &\Bigg|\mathbf{X}_j^{0\prime}\Bigg\{(W_n^0)^2\Big[I_n-\frac{1}{n}\,\mathbf{X}_n(1)\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'(W_n^0)^2\Big]\\ &-\big(W_n^{[k-1]}\big)^2\Big[I_n-\frac{1}{n}\,\mathbf{X}_n(1)\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\mathbf{X}_n(1)'\big(W_n^{[k-1]}\big)^2\Big]\Bigg\}\,\boldsymbol{\varepsilon}_n^0\Bigg|\\ \le\; &\|\mathbf{X}_j^0\|_2\,\Big\|(W_n^0)^2-\big(W_n^{[k-1]}\big)^2\Big\|_2\,\|\boldsymbol{\varepsilon}_n^0\|_2+\|\mathbf{X}_j^0\|_2\,\|G_n\|_2\,\|\boldsymbol{\varepsilon}_n^0\|_2\,,\end{aligned}$$

where $G_n=\frac{1}{n}(W_n^0)^2\mathbf{X}_n(1)(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W_n^0)^2-\frac{1}{n}(W_n^{[k-1]})^2\mathbf{X}_n(1)\,(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_n(1)'(W_n^{[k-1]})^2$, and

$$\begin{aligned}\|G_n\|_2\;\le\;&\Big\|(W_n^0)^2-\big(W_n^{[k-1]}\big)^2\Big\|_2\,\|\Gamma_n(1)\|_2^2\,\Big\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\Big\|_2\,\|(W_n^0)^2\|_2\\ &+\|(W_n^0)^2\|_2\,\|\Gamma_n(1)\|_2^2\,\Big\|\Big(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}-\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\Big\|_2\,\|\big(W_n^{[k-1]}\big)^2\|_2\\ &+\Big\|(W_n^0)^2-\big(W_n^{[k-1]}\big)^2\Big\|_2\,\|\Gamma_n(1)\|_2^2\,\Big\|\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\Big\|_2\,\|\big(W_n^{[k-1]}\big)^2\|_2\\ =\;&\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,\mathcal{O}_p(1)\,\mathcal{O}_p(1)\,\mathcal{O}_p(1)\;=\;\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,,\end{aligned}$$

then we have

$$\left|\eta_{3,j}-\eta_{3,j}^{0}\right|\;=\;\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,\mathcal{O}_p(\sqrt{n})+\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,\mathcal{O}_p(\sqrt{n})\;=\;\mathcal{O}_p(h_n\sqrt{n})\,.$$

This, together with Assumption (A11), yields that

$$P\Big(\max_{q_n<j\le p_n}\big|\eta_{3,j}^0-\eta_{3,j}\big|\ge\frac{\lambda_n r_n}{16}\Big)\;\le\;P\Big(\frac{h_n\sqrt{n}}{\lambda_n r_n}\ge C\Big)\to 0\,. \tag{A.16}$$

Moreover, since $\frac{1}{\sqrt{n}}|\eta_{3,j}^0-\eta_{3,j}^{0,\infty}|\le\frac{1}{\sqrt{n}}\|H_{n,j}^0\|_2\|\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\|_2=\mathcal{O}_p(1)$, it follows from Assumption (A11) that

$$P\Big(\max_{q_n<j\le p_n}\big|\eta_{3,j}^0-\eta_{3,j}^{0,\infty}\big|\ge\frac{\lambda_n r_n}{16}\Big)\;\le\;P\Big(\frac{\sqrt{n}}{\lambda_n r_n}\ge C\Big)\to 0\,. \tag{A.17}$$

By (A.14)–(A.17) and Assumption (A5), we arrive at $P(A_3)\to 0$ as $n\to\infty$.

For A_4, notice that

$$\begin{aligned}|\eta_{4,j}|\;\le\;&\frac{\lambda_n}{2n}\,\|\mathbf{X}_j^0\|_2\,\|\big(W_n^{[k-1]}\big)^2\|_2\,\|\mathbf{X}_n(1)\|_2\,\Big\|\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\Big\|_2\,\|s_n^0(1)\|_2\\ &+\|\mathbf{X}_j^0\|_2\,\|\big(W_n^{[k-1]}\big)^2\|_2\,\|\mathbf{X}_n(1)\|_2\,\Big\|\Big(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\widetilde{\Gamma}_n^{[k]}(1)-I_{q_n}\Big\|_2\,\|\boldsymbol{\beta}_n^0(1)\|_2\\ =\;&\frac{\lambda_n}{n}\,\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p(1)\,\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p\Big(\frac{\sqrt{q_n}}{b_n}\Big)+\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p(1)\,\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p\Big(\frac{\gamma_n}{n\,b_n}\Big)\,\mathcal{O}_p(1)\\ =\;&\mathcal{O}_p\Big(\frac{\lambda_n\sqrt{q_n}}{b_n}\Big)+\mathcal{O}_p\Big(\frac{\gamma_n}{b_n}\Big)\,.\end{aligned}$$

Then Assumption (A11) implies that

$$P\Big(\max_{q_n<j\le p_n}|\eta_{4,j}|\ge\frac{\lambda_n r_n}{4}\Big)\;\le\;P\Big(\frac{\sqrt{q_n}}{b_n r_n}\ge C\Big)+P\Big(\frac{\gamma_n}{b_n\lambda_n r_n}\ge C\Big)\to 0\,.$$

Assumption (A5) yields that

$$P(A_4) \;\le\; P\biggl(\max_{q_n<j\le p_n}\bigl|\eta_{4,j}\bigr| \ge \frac{\lambda_n r_n}{4}\biggr) + P\biggl(\max_{q_n<j\le p_n}\bigl|\beta_{j,\mathrm{init1}}\bigr|^{\tau_1} \ge \frac{1}{r_n}\biggr) \to 0$$

as $n\to\infty$.

This concludes the proof of the sign consistency of the estimator $\boldsymbol{\beta}_n^{[k]}$. Next we proceed to show the asymptotic normality of $\boldsymbol{\beta}_n^{[k]}$.

(II) The asymptotic normality

From (A.3), we have

$$\begin{aligned}
\frac{\sqrt{n}}{s_n(k)}\,\boldsymbol{\xi}_n'\Bigl(\boldsymbol{\beta}_n^{[k]}(1)-\boldsymbol{\beta}_n^{0}(1)\Bigr) \;=\;& \frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'\Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\mathbf{X}_n(1)'\,\bigl(W_n^{[k-1]}\bigr)^2\boldsymbol{\varepsilon}_n^0\\
&-\frac{\lambda_n}{2\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'\Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}s_n^0(1)\\
&+\frac{\sqrt{n}}{s_n(k)}\,\boldsymbol{\xi}_n'\biggl[\Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\widetilde{\Gamma}_n^{[k]}(1)-I_{q_n}\biggr]\boldsymbol{\beta}_n^0(1)\,. \qquad \text{(A.18)}
\end{aligned}$$

For the first term of (A.18), similarly to the proof of part (I), we have the decomposition

$$\Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\mathbf{X}_n(1)'\,\bigl(W_n^{[k-1]}\bigr)^2 \;=\; B_1+B_2+B_3\,,$$

where

$$\begin{aligned}
B_1 &= \Bigl(\widetilde{\Gamma}_n^{0}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\mathbf{X}_n(1)'\,\bigl(W_n^{0}\bigr)^2,\\
B_2 &= \biggl[\Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}-\Bigl(\widetilde{\Gamma}_n^{0}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\biggr]\mathbf{X}_n(1)'\,\bigl(W_n^{0}\bigr)^2,\\
B_3 &= \Bigl(\widetilde{\Gamma}_n^{[k]}(1)+\frac{\gamma_n}{n}\Sigma_2(1)\Bigr)^{-1}\mathbf{X}_n(1)'\Bigl(\bigl(W_n^{[k-1]}\bigr)^2-\bigl(W_n^{0}\bigr)^2\Bigr)\,.
\end{aligned}$$

Note that

$$\frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'B_1\boldsymbol{\varepsilon}_n^0 \;=\; \frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'B_1\boldsymbol{\varepsilon}_{n,\infty}^0+\frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'B_1\bigl(\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\bigr)\,,$$

while

$$\frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'B_1\boldsymbol{\varepsilon}_{n,\infty}^0 \;=\; \sum_{t=1}^{n}a_tZ_t\,,$$

with $a_t=\frac{1}{\sqrt{n}\,s_n(k)\,\sigma_t}\,\boldsymbol{\xi}_n'(\widetilde{\Gamma}_n^0(1)+\frac{\gamma_n}{n}\Sigma_2(1))^{-1}\mathbf{X}_{t,n}(1)$. It is easy to see that $\{a_tZ_t,\mathcal{F}_{n,t},\,1\le t\le n\}$ is a martingale difference array, where $\mathcal{F}_{n,t}=\sigma\{Z_{t_1},X_{t_2,\infty},\,1\le t_2\le n,\,1\le t_1\le t\}$ is the σ-field. Moreover, $E(\sum_{t=1}^n a_tZ_t)=0$ and $E(\sum_{t=1}^n a_tZ_t)^2=E(Z_t^2)\,E(\sum_{t=1}^n a_t^2)=1$.

In addition, Assumption (A7) implies that $1/s_n(k)\le 1/\sqrt{\lambda_{2,\min}}$. Then it follows from Assumptions (A3), (A8), (A9) and (A11) that

$$\begin{aligned}
\max_{1\le t\le n}|a_t| \;&\le\; \frac{1}{\sqrt{n}\,s_n(k)}\,\bigl\|\boldsymbol{\xi}_n\bigr\|_2\biggl\|\Bigl(\widetilde{\Gamma}_n^{0}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Bigr)^{-1}\biggr\|_2\max_{1\le t\le n}\biggl\|\frac{1}{\sigma_t}\,\mathbf{X}_{t,n}(1)\biggr\|_2\\
&\le\; \frac{C}{\sqrt{n}}\,\bigl\|\mathbf{X}_{t,n}(1)\bigr\|_2 \;=\; \mathcal{O}_p\biggl(\frac{\sqrt{q_n}\,v_n}{\sqrt{n}}\biggr)\xrightarrow{\;\mathcal{P}\;}0\,.
\end{aligned}$$

So the conditional Lindeberg condition is satisfied and the martingale central limit theorem (see, e.g. Theorem 2 of [1]) yields that

$$\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_1 \boldsymbol{\varepsilon}_{n,\infty}^0 \xrightarrow{\mathcal{D}} Z\,. \tag{A.19}$$

On the other hand,

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_1 \big(\boldsymbol{\varepsilon}_n^0 - \boldsymbol{\varepsilon}_{n,\infty}^0\big)\right| \\
&\qquad \le \frac{1}{\sqrt{n}\, s_n(k)}\, \|\boldsymbol{\xi}_n\|_2 \left\|\Big(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2 \|\mathbf{X}_n(1)\|_2 \,\big\|(W_n^0)^2\big\|_2 \,\|\boldsymbol{\varepsilon}_n^0 - \boldsymbol{\varepsilon}_{n,\infty}^0\|_2 \\
&\qquad \le C\,\|\boldsymbol{\varepsilon}_n^0 - \boldsymbol{\varepsilon}_{n,\infty}^0\|_2 \xrightarrow{\mathcal{P}} 0\,.
\end{aligned}$$

By Slutsky's Theorem,

$$\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_1 \boldsymbol{\varepsilon}_{n,\infty}^0 \xrightarrow{\mathcal{D}} Z\,. \tag{A.20}$$

For B_2, we know that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_2 \boldsymbol{\varepsilon}_n^0\right| \\
&\quad \le \frac{1}{\sqrt{n}\, s_n(k)} \left\|\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1} - \Big(\widetilde{\Gamma}_n^0(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2 \|\boldsymbol{\xi}_n\|_2 \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_n^0\right\|_2,
\end{aligned}$$

and

$$\begin{aligned}
\left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_n^0\right\|_2 &\le \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 (\boldsymbol{\varepsilon}_n^0 - \boldsymbol{\varepsilon}_{n,\infty}^0)\right\|_2 + \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2 \\
&\le \|\mathbf{X}_n(1)\|_2 \,\big\|(W_n^0)^2\big\|_2 \,\|\boldsymbol{\varepsilon}_n^0 - \boldsymbol{\varepsilon}_{n,\infty}^0\|_2 + \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2 \\
&\le \mathcal{O}_p\big(\sqrt{n}\big)\, \mathcal{O}_p(1)\, \mathcal{O}_p(1) + \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2 \\
&= \mathcal{O}_p\big(\sqrt{n}\big) + \left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2 .
\end{aligned}$$

Markov's inequality and Assumptions (A1), (A2) and (A8) give

$$\begin{aligned}
P\bigg(\frac{1}{q_n n}\left\|\mathbf{X}_n(1)'\,(W_n^0)^2 \boldsymbol{\varepsilon}_{n,\infty}^0\right\|_2^2 > C\bigg) &\le \frac{1}{C\, q_n n} \sum_{i=1}^{q_n} E\bigg(\sum_{t=1}^{n} X_{t,i}\, \frac{Z_t}{\sigma_t}\bigg)^2 \\
&\le \frac{1}{C\, q_n n} \sum_{i=1}^{q_n} E\bigg(\sum_{t=1}^{n} X_{t,i}\, \frac{Z_t}{\sigma_{\min}}\bigg)^2 \le \frac{1}{C\,\sigma_{\min}^2}\,.
\end{aligned}$$

This means that $\|\mathbf{X}_n(1)'(W_n^0)^2\boldsymbol{\varepsilon}_{n,\infty}^0\|_2 = \mathcal{O}_p(\sqrt{q_n n})$. These bounds together with Assumption (A11) imply that

$$\left|\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_2 \boldsymbol{\varepsilon}_n^0\right| = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)\, \mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\, \mathcal{O}_p\big(\sqrt{q_n\, n}\big) \xrightarrow{\mathcal{P}} 0\,. \tag{A.21}$$

Along similar lines for B_2, we obtain

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}\, s_n(k)}\, \boldsymbol{\xi}_n' B_3 \boldsymbol{\varepsilon}_n^0\right| \\
&\qquad \le \frac{1}{\sqrt{n}\, s_n(k)} \left\|\Big(\widetilde{\Gamma}_n^{[k]}(1) + \frac{\gamma_n}{n}\Sigma_2(1)\Big)^{-1}\right\|_2 \|\boldsymbol{\xi}_n\|_2 \left\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2 - \big(W_n^{[k-1]}\big)^2\Big)\,\boldsymbol{\varepsilon}_n^0\right\|_2
\end{aligned}$$

and

$$\begin{aligned}&\Big\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big)\,\boldsymbol{\varepsilon}_n^0\Big\|_2\\&\quad\le\Big\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big)\,\big(\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\big)\Big\|_2+\Big\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big)\,\boldsymbol{\varepsilon}_{n,\infty}^0\Big\|_2.\end{aligned}$$

Moreover,

$$\begin{aligned}&\Big\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big)\,\big(\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\big)\Big\|_2\\&\quad\le\big\|\mathbf{X}_n(1)\big\|_2\,\Big\|\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big\|_2\,\big\|\boldsymbol{\varepsilon}_n^0-\boldsymbol{\varepsilon}_{n,\infty}^0\big\|_2\le\mathcal{O}_p(\sqrt{n})\,\mathcal{O}_p\Big(\frac{h_n}{\sqrt{n}}\Big)\,\mathcal{O}_p(1)\,=\,\mathcal{O}_p(h_n).\end{aligned}$$

From Markov's inequality and Assumptions (A1), (A2) and (A9):

$$\begin{aligned}&P\bigg(\frac{1}{q_n h_n^2}\,\Big\|\mathbf{X}_n(1)'\Big(\big(W_n^0\big)^2-\big(W_n^{[k-1]}\big)^2\Big)\,\boldsymbol{\varepsilon}_{n,\infty}^0\Big\|_2^2>C\bigg)\\&\qquad\le\frac{1}{C\,q_n h_n^2}\sum_{i=1}^{q_n}E\bigg(\sum_{t=1}^{n}X_{t,i}\Big(\frac{1}{\sigma_t^2}-\frac{1}{\hat\sigma_t^{[k-1]}}\Big)\,\varepsilon_t\bigg)^2\le\frac{1}{C\,q_n h_n^2}\,\frac{h_n^2}{n}\sum_{i=1}^{q_n}E\bigg(\sum_{t=1}^{n}X_{t,i}\,\varepsilon_t\bigg)^2.\end{aligned}$$

That is $\|\mathbf{X}_n(1)'((W_n^0)^2-(W_n^{[k-1]})^2)\boldsymbol{\varepsilon}_{n,\infty}^0\|_2=\mathcal{O}_p(\sqrt{q_n}\,h_n)$. Therefore we have

$$\text{(A.22)}\qquad\bigg|\frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'\,B_3\,\boldsymbol{\varepsilon}_n^0\bigg|=\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)\,\mathcal{O}_p(1)\,\mathcal{O}_p\big(\sqrt{q_n}\,h_n\big)\xrightarrow{\mathcal{P}}0.$$

By (A.20)–(A.22) and Slutsky's Theorem,

$$\text{(A.23)}\qquad\frac{1}{\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'\,(B_1+B_2+B_3)\,\boldsymbol{\varepsilon}_n^0\xrightarrow{\mathcal{D}}Z.$$

Now it suffices to show that the last two terms of (A.18) converge to zero in probability. By Assumption (A11),

$$\begin{aligned}&\bigg|\frac{\lambda_n}{2\sqrt{n}\,s_n(k)}\,\boldsymbol{\xi}_n'\Big(\widetilde\Gamma_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}s_n^0(1)\bigg|\\&\qquad\le\frac{\lambda_n}{2\sqrt{n}\,s_n(k)}\,\|\boldsymbol{\xi}_n\|_2\,\bigg\|\Big(\widetilde\Gamma_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\bigg\|_2\,\big\|s_n^0(1)\big\|_2\\&\qquad=\mathcal{O}_p\Big(\frac{\lambda_n\sqrt{q_n}}{\sqrt{n}\,b_n}\Big)\xrightarrow{\mathcal{P}}0.\end{aligned}$$

For the last term of (A.18), by (A.11), we obtain

$$\bigg|\frac{\sqrt{n}}{s_n(k)}\,\boldsymbol{\xi}_n'\bigg[\Big(\widetilde\Gamma_n^{[k]}(1)+\frac{\gamma_n}{n}\,\Sigma_2(1)\Big)^{-1}\widetilde\Gamma_n^{[k]}(1)-I_{q_n}\bigg]\boldsymbol{\beta}_n^0(1)\bigg|=\mathcal{O}_p\Big(\frac{\gamma_n}{\sqrt{n}\,b_n}\Big)\xrightarrow{\mathcal{P}}0.$$

This completes the proof of Theorem 2.1. □

ACKNOWLEDGMENTS

This work has been supported by National Natural Science Foundation of China (NSFC) Grants 11671194 and 11501287.

REFERENCES

[1] Brown, B.M. (1971). Martingale central limit theorems, *The Annals of Mathematical Statistics*, **42**, 59–66.

[2] Chen, K. and Chan, K. (2011). Subset ARMA selection via the adaptive Lasso, *Statistics and Its Interface*, **4**, 197–205.

[3] Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters, *The Annals of Statistics*, **32**, 928–961.

[4] Furman, Y. (2014). *VAR estimation with the adaptive elastic net*, Available at SSRN: https://ssrn.com/abstract=2456510 or http://dx.doi.org/10.2139/ssrn.2456510

[5] Hsu, N.; Hung, H. and Chang, Y. (2008). Subset selection for vector autoregressive processes using Lasso, *Computational Statistics and Data Analysis*, **52**, 3645–3657.

[6] Huang, J.; Ma, S. and Zhang, C.H. (2008). Adaptive lasso for sparse high-dimensional regression, *Statistica Sinica*, **18**, 1603–1618.

[7] Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models, *Journal of Econometrics*, **140**, 849–873.

[8] Mak, T.; Wong, H. and Li, W. (1997). Estimation of nonlinear time series with conditional heteroskedastic variance by iteratively weighted least squares, *Computational Statistics and Data Analysis*, **24**, 169–178.

[9] Schomaker, M. (2012). Shrinkage averaging estimation, *Statistical Papers*, **53**, 1015–1034.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, **58**, 267–288.

[11] Wagener, J. and Dette, H. (2012). Bridge estimators and the adaptive lasso under heteroscedasticity, *Mathematical Methods of Statistics*, **21**, 109–126.

[12] Wagener, J. and Dette, H. (2013). The adaptive lasso in high-dimensional sparse heteroscedastic models, *Mathematical Methods of Statistics*, **22**, 137–154.

[13] Wainwright, M.J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using constrained quadratic programming (Lasso), *IEEE Transactions on Information Theory*, **55**, 2183–2202.

[14] Yoon, Y.J.; Park, C. and Lee, T. (2013). Penalized regression models with autoregressive error terms, *Journal of Statistical Computation and Simulation*, **83**, 1756–1772.

[15] Yüzbaşi, B.; Ahmed, S.E. and Güngör, M. (2017). Improved penalty strategies in linear regression models, *REVSTAT – Statistical Journal*, **15**(2), 251–276.

[16] Zhao, P. and Yu, B. (2006). On model selection consistency of lasso, *The Journal of Machine Learning Research*, **7**, 2541–2563.

[17] Ziel, F. (2016). Iteratively reweighted adaptive lasso for conditional heteroscedastic time series with applications to AR-ARCH type processes, *Computational Statistics and Data Analysis*, **100**, 773–793.

[18] Zou, H. (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429.

[19] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **67**, 301–320.

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 199–220

WORD DISTINCTIVITY — QUANTIFYING IMPROVEMENT OF TOPIC MODELING RESULTS FROM N-GRAMMING *

Author: Christine P. Chai
– Microsoft Corporation
United States
cpchai21@gmail.com

Received: February 2019 Revised: June 2019 Accepted: January 2020

Abstract:

- Text data cleaning is an important but often overlooked step in text mining because it is difficult to quantify the contribution. Therefore, we propose the word distinctivity to measure the improvement of topic modeling results from n-gramming, which preserves special phrases in a corpus. The word distinctivity evaluates the signal strength of a word's topic assignments, and a high distinctivity means a high posterior probability for the word to come from a certain topic. We implemented the latent Dirichlet allocation for topic modeling, and discovered that some special phrases show an increase in word distinctivity, reducing uncertainty in topic identification.

Keywords:

- *latent Dirichlet allocation; text mining; topic modeling; n-gramming; data cleaning; quantification.*

AMS Subject Classification:

- 62-07, 62C10, 68U15.

*The research was conducted while the author was a PhD student in statistical science at Duke University. The opinions and views expressed in this manuscript are those of the author and do not necessarily state or reflect those of Microsoft.

1. INTRODUCTION

Data cleaning is an essential step to prepare the data for analysis [34, 54], because most datasets in real life contain much noise and missing values [22]. For example, a table consisting of a variable "years of education" may encode a missing value as "99". It is almost impossible for a person to have 99 years of education, and we need to remove this value before running a regression on the data. Otherwise, the regression results would be distorted.

Nevertheless, data cleaning is often overlooked for various reasons. One reason is that early-career professionals focus on the analysis due to the over-emphasis of statistical modeling in graduate school programs [41]. Another reason is that data cleaning is often viewed as a tedious and time-consuming task [43]. According to a report done by CrowdFlower in 2016 [14], most data scientists spend more than half of their work time cleaning and organizing data.

1.1. Text data cleaning

The under-appreciation of text data cleaning is a more severe problem because people tend to care less about text data than numerical data [48, 44]. This is unfortunate, although understandable because text data are unstructured and more difficult to analyze than the numerical counterparts [39]. Many characteristics of numerical data are not transferable to text data, such as mean and standard deviation.

Recently, due to the emerging need of text data mining [17], more and more resources are available for text data processing [18, 3, 40]. However, most of them describe text data cleaning as an important step before the analysis, without providing concrete evidence of why this step is crucial. If we can quantify the text data cleaning results, the importance of preprocessing the data is clearly demonstrated. Quantifiable results are published in many different fields to show new research findings, and text data cleaning results should not be an exception.

A unique issue with text data is that many statistical models perform random permuation of words and do not account for word order [55], resulting in confusion and loss of semantic information. For example, the two sentences "the department chair couches offers" and "the chair department offers couches" comprise the exact same words, but with completely different meanings. The first sentence probably came from a university administration report, and the second sentence may be written by a retail store [50].

1.2. N-gramming and word distinctivity

Text data cleaning needs to preserve the word order to mitgate the problem, and one common solution is n-gramming [47, 26], i.e. retaining special phrases in the corpus, such as "white house" and "new york". The phrases are called n-grams [5]. The n-gramming process helps recover semantic information because words within a preserved special phrase are regarded as a single token in text analysis. For example, the two sentences "The white person lives in the house." and "The person lives in the White House." contain the exact same words but have completely different meaning. If the term "White House" is separated as two words, the original meaning is lost.

N-gramming has been widely used in natural language processing, such as machine translation [53], speech recognition [52], and information retrieval [25, 15]. Discussion and comparison of various n-gram sizes (i.e., how many words each n-gram contains) are also extensive in the literature [29, 32].

Nevertheless, few previous studies [50] evaluate the information gain from n-gramming, not to mention quantifying it. Even though many researchers in the text mining field regard n-gramming as a necessity in text data processing, it can be challenging to explain to business stakeholders why n-gramming is worth the time spent. Most business stakeholders would like to see quantifiable results, such as "a 20% increase in model accuracy."

To quantify the improvement of text classification results from n-gramming, we propose the "word distinctivity" as a metric. Word distinctivity refers to how "distinctive" a word is, or in plain language, how likely the word is assigned to a certain topic.

In mathematical terms, word distinctivity is defined as

$$\max_i P\big(\text{topic } i \mid \text{word } j,\, \text{data}\big),$$

i.e., taking the maximum probability of how likely a topic i is, given a specific word j and the data.

Here, text classification is also known as topic modeling, the classification process that assigns text documents into various topics.

1.3. Overview of topic modeling

Topic modeling is actually an automated process of classifying text documents into different topics, which is a part of natural language processing. A common approach for topic modeling is the latent Dirichlet allocation (LDA) [6], and the algorithm outputs a topic assignment vector for each document, as well as the "top" words for each topic. The number of topics is a preset constant, while the contents of each topic are to be determined by the text corpus. LDA seems to be the standard algorithm of topic modeling because many researchers extend LDA to more advanced topic models [35, 38, 27].

Nevertheless, to the best of my knowledge, few have questioned the fundamental criteria of how LDA selects the "top" words in each topic — the selection is based on the posterior probability $P(\text{word } j \mid \text{topic } i, \text{data})$. This answers the question "Given topic i and the data, which words would the model generate?" But more often than not, we are given a document with words, and would like to know which topic(s) the words belong to. Hence a better selection criteria is $P(\text{topic } i \mid \text{word } j, \text{data})$, which is the word distinctivity before we take the maximum probability across each topic.

Since topic models assign each word in the corpus to one or more topics, we use the word distinctivity to measure the signal strength generated by each word. If a word (or a retained phrase) has a high probability to be assigned to a particular topic, the word is considered highly distinctive. Upon seeing this word, we know that it is highly likely that the word came from the particular topic. On the contrary, if a word is equally likely to be assigned to all topics, the word has low distinctivity.

To compare and quantify the topic modeling results, we created two versions of the same text dataset — before and after n-gramming and implemented the LDA algorithm with the same number of preset topics. To show the improvement from n-gramming, we look at the word distinctivity of a retained phrase and the word distinctivity of each word in that phrase. The former is expected to be much higher than any of the latter.

2. DATA DESCRIPTION

Our text dataset originates from the collection of 109,055 blog posts from the top 467 US political blogs, as ranked in 2012 by Technorati (now Synacor) [24]. The blog posts were written in English. The blog post collection was obtained from MaxPoint Interactive (now Valassis Digital), where the computer scientists web-scraped the text and stemmed the words using a modified version of Snowball [30] developed in-house. Therefore, the words in the corpus are actually tokens, but we use the terms "word" and "token" interchangeably.

"Stemming" a word removes its suffix and keeps only its root, and the output is a "token". In this way, words of the same root are consolidated into the same token. For example, according to the `wordStem` function in the R package `SnowballC` [7], "worry" and its present principle form "worrying" are both assigned to the same token "worri". When we summarize the word counts in the corpus, we may see that "worry" appears 50 times and "worrying" appears 30 times. After stemming the corpus, we would see that "worri" appears a combination of 80 times. The stemming process not only reduces the size of vocabulary, but also increases the readability of the results.

We focus on the articles relevant to the fatal shooting of Trayvon Martin by George Zimmerman on February 26, 2012[1], which triggered a heated debate on the media and many political blogs. Among the 109,055 blog posts, 450 contains the keyword "Trayvon", and the 450 blog posts form the actual corpus for analysis. We call the text corpus the "Trayvon Martin dataset", and we generated two versions of this dataset.

2.1. First version: stop words removed (before n-gramming)

In the first version, we removed predefined stop words (approximately 300) with little semantic meaning (e.g. "to", "for") from the Trayvon Martin Dataset. Note that negation terms, such as "no", "not", and "don't", are excluded from the stop word list, because they can reverse the meaning of the next word. For example, "not a good idea" means "a bad idea".

[1] https://en.wikipedia.org/wiki/Shooting_of_Trayvon_Martin

Since many topic models are bag-of-words models and do not preserve word order [42], one solution [11] is to replace words and a preceding negation term with its corresponding antonym, e.g. "not good" becomes "bad". However, this is outside the paper's scope because we would like to focus on the improvement of topic modeling results from n-gramming, instead of adding another variation to the data.

2.2. Second version: special phrases retained (after n-gramming)

In the second version of the dataset, we identified and preserved the special phrases. The process is called n-gramming, whose goal is to keep sets of words with high probability of co-occurrence. If a special phrase contains n words, it is called an n-gram. In particular, a word can be called a uni-gram; a two-word phrase retained this way is a bi-gram, and a three-word phrase of this kind is a tri-gram. Section 3.1 explains the n-gramming methodology in detail.

3. METHODS

The methods section describes the n-gramming process, the latent Dirichlet allocation (LDA) algorithm, and the word distinctivity measure. The n-gramming process is used to keep certain words together, so their order would not be affected by the bag-of-words models, which assume an orderless document representation. The LDA is used for topic modeling, and it determines which document contains which topic(s) in a probabilistic way. The word distinctivity measure determines which word(s) have a strong signal in topic identification, and this measure can be computed from the LDA topic assignment vectors.

3.1. N-gramming

The objective of n-gramming is to retain phrases with high probability of occurrence. One obvious solution is to select phrases that appears many times in the corpus, but this is likely to include many common expressions with little semantic meaning.

A major question we also try to answer is, "Given a particular word, how likely is this word going to follow it?" The Turbo Topics [5] software demonstrated an example: Given the word "new" in their corpus, the word "york" follows it 60% of the time. Therefore, we can infer that "new york" is a bi-gram.

To identify the n-grams, we start by searching for all n-word phrases (a.k.a. n-gram candidates) and filter them in terms of raw frequency and conditional probability. Next, we set certain thresholds to determine which are the actual n-grams, and finally provide the implementation results.

3.1.1. Search for n-gram candidates

We retrieve all n-gram candidates in the corpus by "shingling" at the word level [45, 8], a standard approach of slicing down a long sentence into phrases with n words each [19, 9]. In comparison, "shingling" at the character level creates each n-gram candidate as a string of n characters, which is not of interest here [49, 10].

If a sentence contains n words, then there are n uni-grams, $n-1$ bi-gram candidates, and $n-2$ tri-gram candidates. For example, "heard gun shot outside apartment yesterday" contains 6 words, 5 two-word phrases, and 4 three-word phrases, listed as below:

- Text: "heard gun shot outside apartment yesterday";
- Two-word phrases: "heard gun", "gun shot", "shot outside", "outside apartment", and "apartment yesterday";
- Three-word phrases: "heard gun shot", "gun shot outside", "shot outside apartment", and "outside apartment yesterday".

Note that "n-gram candidates" are different from "n-grams". The former term refers to the n-word phrases which can potentially be n-grams. The latter term "n-grams" refers to only the selected ones n-word phrases, typically with both practical and statistical significance. How we select actual n-grams from the candidates is described next.

3.1.2. Raw frequency: practical significance

For an n-word phrase to "qualify" as an n-gram candidate, the phrase must occur at least a certain number of times in the data, so setting a minimum cutoff frequency is essential [51, 16]. The raw frequency threshold corresponds to the practical significance of n-grams and removes rare words, which have little semantic meaning in the corpus. One example is the name of the police officer who arrested George Zimmerman.

A default minimum frequency of 5 is recommended by the Microsoft Azure Machine Learning Studio [31], but this is too low for the Trayvon Martin dataset. The results include lots of unmeaningful phrases, such as "countri better" and "claim obama".

Instead, the cutoff for phrase counts should be determined by corpus size, and we empirically set it to 100 for the bi-grams in the Trayvon Martin dataset. That is, to be considered a bi-gram candidate, any two-word phrase has to appear in the corpus at least 100 times.

3.1.3. Conditional probability: statistical significance

The conditional probability is also a widely used approach to filter out n-gram candidates [9, 13], and this measures the statistical significance of each n-word phrase.

The conditional probability in n-gramming is denoted as

$$P\big(\text{word } n \mid \text{words } 1, ..., n-1\big),$$

i.e., the probability of getting the n-th word given the first $n-1$ words.

In mathematical terms, the marginal probability $P(\text{word})$ is the frequency of the word, divided by the total number of words in the corpus. Similarly, $P(\text{words } 1, ..., m)$ is the frequency of the m-word phrase, divided by the total number of words in the corpus.

Hence the conditional probability of n-grams is written as

$$\begin{aligned} P\big(\text{word } n \mid \text{words } 1, ..., n-1\big) &= \frac{P(\text{words } 1, ..., n)}{P(\text{words } 1, ..., n-1)} \\ &= \frac{\text{Frequency of words } 1, ..., n}{\text{Frequency of words } 1, ..., n-1}. \end{aligned}$$

Particularly, the conditional probability for bi-grams is $P(\text{word } 2 \mid \text{word } 1)$. If the answer is "yes" to the question "Is Word 2 more likely to follow Word 1?", then the two words should form a bi-gram.

The hypotheses are:

- $H_0: \ P(\text{word } 2 \mid \text{word } 1) \leq P(\text{word } 2)$;
- $H_1: \ P(\text{word } 2 \mid \text{word } 1) > P(\text{word } 2)$.

The p-value cutoff is set to 0.05 by default, and this removes most words which just happened to appear together. For example, "said obama" is a common phrase but not a meaningful bi-gram, and the high frequency is due to the high marginal probability of "said".

Note that the raw frequency cutoff is also crucial, since rare phrases can distort the conditional probability and produce undesirable results. As an extreme example, if the first word appears only once in the data, the second word following the first word has conditional probability of 100%.

3.2. Latent Dirichlet allocation (LDA)

LDA is a Bayesian data generative process that performs topic modeling, i.e., classifies documents and words into topics. The LDA algorithm first draws each topic from a Dirichlet distribution as the prior, then updates the probabilities by using the words in the documents. Finally, the algorithm outputs the `top.topic.words` for each topic, based on the posterior probability $P(\text{word } j \mid \text{topic } i, \text{data})$ for each combination of topic i and word j.

For each document, LDA outputs the topic proportions — the probabilistic topic assignment vector. The number of components of this vector is equal to the preset number of topics. The probabilities for topics are defined using word counts, i.e., the number of relevant words in the document. For example, (0.5, 0.3, 0.2) means the document has topic proportions 50% in Topic 1, 30% in Topic 2, and 20% in Topic 3. In this document, 50% of the words belong to Topic 1, 30% of the words belong to Topic 2, and 20% of the words belong to Topic 3.

LDA also produces topic assignments at the word level, which is the main usage in this paper. For example, a word has a probabilistic topic assignment vector (0.5, 0.3, 0.2), and we implemented 100 simulations. This means the word is assigned to Topic 1 for 50 times, assigned to Topic 2 for 30 times, and assigned to Topic 3 for 20 times.

3.2.1. Setup

The LDA algorithm assumes the corpus $\mathbf{D}$ to be a fixed set of M documents and the words from a finite vocaublary set $\mathbf{W}$. The LDA also requires a predefined number of topics K, and the setup is specified as below:

- Fixed set of M documents: $\mathbf{D} = \{D_1, ..., D_M\}$;
- Words within a document D_d: $\mathbf{W}_d = \{w_{d,1}, ..., w_{d,N_d}\}$
  - the document D_d contains N_d words;
- Finite vocabulary set: $\mathbf{W} = \mathbf{W}_1 \cup \cdots \cup \mathbf{W}_M$, with size N
  - $\mathbf{W}$ is the union of all sets $\mathbf{W}_d$, where $d = 1, ..., M$,
  - the vocabulary set of $\mathbf{D}$ contains N words in total;
- Predefined number of topics K;
- Fixed vectors $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_K)$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_K)$.

For the parameters, we set the number of topics to $K = 5$ and $\alpha_i = 0.1, \beta_i = 0.1$ for all $i = 1, ..., K$, same as previous researchers did on the Trayvon Martin dataset [42, 1].

3.2.2. Algorithm description

The data generative process of LDA is defined as below, and the plate diagram is illustrated in Figure 1. The word proportion vector ϕ_k determines the relative "weights" of each word in topic k, and the topic proportion vector θ_d determines how the document D_d is composed from each of the K topics.

- For each topic k
  - Draw a word proportion vector: $\phi_k|\boldsymbol{\alpha} \sim \text{Dirichlet}_N(\boldsymbol{\alpha})$;
- For each document D_d
  - draw a topic proportion vector: $\theta_d|\boldsymbol{\beta} \sim \text{Dirichlet}_K(\boldsymbol{\beta})$,
  - for each word $w_{d,i}$ in document D_d
    * draw a topic assignment $z_{d,i}|\theta_d \sim \text{Multinomial}(\theta_d)$,
    * draw a word from the topic $w_{d,i}|\phi_{z_{d,i}} \sim \text{Multinomial}(\phi_{z_{d,i}})$.

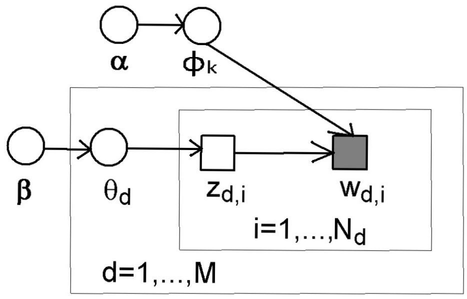


Figure 1: Plate diagram for the LDA process.

The full posterior model specification of LDA is:

$$p\big(\theta_{1:M}, z_{1:M,1:N}, \phi_{1:K} \,|\, w_{1:M,1:N}, \boldsymbol{\alpha}, \boldsymbol{\beta}, K\big) =$$

$$= \frac{p\big(\theta_{1:M}, z_{1:M,1:N}, \phi_{1:K} \,|\, \boldsymbol{\alpha}, \boldsymbol{\beta}, K\big) \times p\big(w_{1:M,1:N} \,|\, \theta_{1:M}, z_{1:M,1:N}, \phi_{1:K}, \boldsymbol{\alpha}, \boldsymbol{\beta}, K\big)}{\int_{\phi_{1:K}} \int_{\theta_{1:M}} \sum\limits_{z_{1:M,1:N}} p\big(\theta_{1:M}, z_{1:M,1:N}, \phi_{1:K}, w_{1:M,1:N} \,|\, \boldsymbol{\alpha}, \boldsymbol{\beta}, K\big)} .$$

Since a document D_d has only N_d words, $w_{d,n} = 0$ for all $n > N_d$, i.e., a non-existent word. Similarly, $z_{d,n} = 0$ for all $n > N_d$, i.e., a non-existent topic assignment.

The prior $p(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\phi} \,|\, \boldsymbol{\alpha}, \boldsymbol{\beta}, K)$ can also be written as

$$p(\boldsymbol{\phi} \,|\, \boldsymbol{\alpha}, K)\ p(\boldsymbol{\theta} \,|\, \boldsymbol{\beta}, K)\ p(\mathbf{z} | \boldsymbol{\theta}) ,$$

which is derived from the data generative process.

On the other hand, the likelihood in the denominator is intractable, and this requires Markov Chain Monte Carlo or variational inference methods to compute. Existing solutions include a variational Bayes approximation approach [6] and a collapsed Gibbs sampler that integrates out both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ [20].

3.3. Word distinctivity

Word distinctivity measures how "distinctive" a word is in terms of topic classification, and this can be regarded as an add-on to the LDA model. Since LDA returns words of the highest (posterior) probability given each topic and the data, the output compares many words and decides which ones should be assigned to the particular topic. In contrast, word distinctivity compares the assignment probabilities across topics for a particular word, so when we see the word, we know how likely it is from a certain topic.

Word distinctivity is defined as the highest posterior probability of a word to be assigned to a particular topic. For example, if a word w_1 has a topic assignment vector of (0.1, 0.6, 0.3), the word distinctivity of w_1 is 0.6. To put it differently, word distinctivity is the maximum signal level observed from the word.

For another example, assume the words w_2, w_3 have topic assignment vectors (0.33, 0.34, 0.33) and (0.80, 0.10, 0.10), respectively. Then w_2 has low distinctivity because its topic assignment vector nearly corresponds to a discrete uniform distribution, which has the largest entropy. On the other hand, w_3 has high distinctivity, because given w_3 and the data, we are 80% sure that the word w_3 came from the first topic.

Using the Bayes' theorem, we can convert the direct output of LDA $P(\text{word}\, j \mid \text{topic}\, i, \text{data})$ into the word distinctivity candidates for word j:

$$P\big(\text{topic}\, i \mid \text{word}\, j, \text{data}\big) = \frac{P\big(\text{word}\, j \mid \text{topic}\, i, \text{data}\big)\, P(\text{topic}\, i)}{\sum\limits_{\text{topic}\, k} P\big(\text{word}\, j \mid \text{topic}\, k, \text{data}\big)\, P(\text{topic}\, k)}.$$

Then we take the maximum probability across all topics, and the value is the word distinctivity of word j:

$$\max_i P\big(\text{topic}\, i \mid \text{word}\, j, \text{data}\big).$$

4. IMPLEMENTATION RESULTS

We demonstrated the statistical methods on the Trayvon Martin dataset (described in Section 2). First, we identified and created 22 bi-grams from n-gramming. Next, we implemented LDA and compared the topic modeling results before and after n-gramming. Then we converted the LDA posterior probabilities into word distinctivity, and we compared the new results again to show the information gain from bi-grams — the increase in word distinctivity. Finally, we compared the selected words for each topic under different versions of LDA implementations, showing that word distinctivity also improves the quality of topic modeling results.

4.1. N-gramming output

In the Trayvon Martin dataset, n-gramming was implemented in R using code from an existing GitHub repository [23]. The function `textToBigrams` generates bi-grams from the corpus; the cutoff frequency is set to 100; the level of statistical significance is set to 0.05.

The 22 bi-grams generated from the Trayvon Martin dataset are:

```
barack_obama, black_panther, civil_right, comment_dave
dave_surl, dont_know, dont_think, fox_news
georg_zimmerman, gregori_william, look_like, mitt_romney
neighborhood_watch, new_york, presid_obama, right_wing
self_defens, stand_ground, trayvon_martin, unit_state
white_hous, year_old
```

The bi-grams fall into three categories: people names, special phrases, and common expressions. For most bi-grams, the original form can be clearly determined, e.g. "unit_state" is originally "United States". The special phrases, such as "neighborhood_watch" and "self_defens" are the most interesting because they may not be easily identified when they were two separate words. We expect an increase in semantic information when the special phrases are regarded as single tokens.

There are few tri-grams of interest because more than 93% of the three-word phrases appear only once in the Trayvon Martin dataset. On the other hand, we found an extremely long n-gram — one blogger includes the Second Amendment to the United States Constitution[2] in every post, just like a signature.

4.2. Topic modeling results

We implemented LDA for topic modeling using the R package `lda` [12], with the main function `lda.collapsed.gibbs.sampler` [20, 28]. Then we list the top 10 words for each LDA-generated topic from the function `top.topic.words`. Table 1 shows the results before n-gramming. Table 2 shows the results after n-gramming, and four bi-grams are present — "fox_news", "trayvon_martin", "georg_zimmerman", and "dave_surl".

The five topic names are manually assigned from the vocabulary, and the topics are also aligned for easy cross-table comparison. The first four topics are the same for both tables and are explained as below:

- Topic 1 is "General" because the words are used in everyday language, such as "like", "get", "know", and "think" — obviously not a distinctive topic.
- Topic 2 is "Election", mainly due to the words "obama", "presid", and "romney", which normally appear in the 2012 US presidential election.
- Topic 3 is "Incident" due to the key words "martin" (or "trayvon_martin") and "zimmerman" (or "georg_zimmerman") for the fatal shooting incident of Trayvon Martin.
- Topic 4 is "News Coverage", because of the words "anonym" (anonymous), "fox" (Fox News), "malkin" (Michelle Malkin, an American political commentator), and "msnbc" (an American television network).

The last topic differs in the results before and after n-gramming. In Table 1, Topic 5 is "Gun Laws" because of the words "law", "gun", "ground", and "stand" (stand your ground law), although it is a little difficult to tell. In contrast, Topic 5 in Table 2 is "Zimmerman Trial" due to the words "perjuri", "bond", and "free". Both "Gun Laws" and "Zimmerman Trial" are meaningful topics, but we can reveal one, but not both, from a single iteration of the LDA topic model.

The function `top.topic.words` in the R package `lda` [12] selects words for each topic based on the posterior probability $P(\text{word}\, j \,|\, \text{topic}\, i, \text{data})$, i.e. the probability of getting word j given topic i and the data. This seems reasonable because the function returns words that are most likely to appear in each topic.

[2] "A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed."

However, some words (or bi-grams) in the topics are not "distinctive" enough — these words are common across the corpus and are (unfortunately) not stop words. For example, the bi-grams "trayvon_martin" and "georg_zimmerman" belong to Topic 3 (Incident) in Table 2 because the incident is about George Zimmerman shooting Trayvon Martin. But the opposite does not hold: Upon seeing the bi-gram "georg_zimmerman", we do not know whether it came from Topic 3 (Incident), Topic 4 (News Coverage), or Topic 5 (Zimmerman Trial).

Our explanation is that the two bi-grams have low distinctivity in terms of topic selection, i.e., upon seeing them, we do not know which topic they belong to. Another example is the words in Topic 1 (General) — the words in this particular category also often appear in other topics. This presents the need of "word distinctivity", that is, given the word, how likely it is going to be in a certain topic.

Table 1: Before n-gramming: Top 10 words for each LDA-generated topic.

Topic 1 General	Topic 2 Election	Topic 3 Incident	Topic 4 News Coverage	Topic 5 Gun Laws
like	obama	zimmerman	anonym	law
peopl	presid	martin	fox	gun
dont	year	said	news	ground
comment	romney	trayvon	liber	stand
get	american	polic	tommi	forc
know	democrat	georg	malkin	alec
think	said	call	msnbc	mar
right	nation	black	show	defend
make	women	prosecutor	conserv	reason
white	govern	charg	gregori	shoot

Table 2: After n-gramming: Top 10 words for each LDA-generated topic.

Topic 1 General	Topic 2 Election	Topic 3 Incident	Topic 4 News Coverage	Topic 5 Zimmerman Trial
like	obama	zimmerman	anonym	comment
peopl	presid	martin	liber	spokesmancom
get	year	case	fox	perjuri
make	american	polic	tommi	**dave_surl**
know	romney	law	show	bond
think	republican	**trayvon_martin**	conserv	free
right	govern	said	news	access
say	law	**georg_zimmerman**	**fox_news**	view
dont	women	trayvon	msnbc	surl
see	countri	shoot	hanniti	account

Remarks

We had to manually align the topics in Tables 1 and 2 from the original LDA output, because which topic is labeled as "Topic 1" is arbitrary in the LDA output. This is a common labeling issue in finite mixture models, where the label's index has no meaning to the model itself [37]. LDA does not "know" which topic the words actually belong to; instead, LDA simply determines which words belong to the same topic.

In addition, the meaning of the bi-gram "dave_surl" is difficult to determine because we do not have the original, non-stemmed version of the Trayvon Martin corpus. We used the R package `SnowballC` [7] and found the stemmed token of the word "surveillance" to be "surveil", not "surl".

4.3. Conversion to word distinctivity

For word distinctivity, we used the R package `topicmodels` [21] because it is much easier to obtain the posterior probability values than from the R package `lda`. In the R package `topicmodels`, the `posterior` function returns the posterior probabilities $P(\text{word}\, j \mid \text{topic}\, i, \text{data})$ and $P(\text{topic}\, i \mid \text{document}\, d, \text{data})$ for each combination of topic i, word j, and document d.

In Section 4.2, we demonstrated using the R package `lda` on purpose because the function `top.topic.words` returns the words with the highest posterior probability for each topic. This is an easy and straightforward way to obtain the LDA topic modeling results. Actually, the results from R package `topicmodels` based on posterior probability are not ideal — the same word or bi-gram can appear in more than one topic.

Tables 3 and 4 list the top 10 **distinctive** words for each topic from LDA. The former shows the results before n-gramming, while the latter shows the results after n-gramming and contains nine bi-grams. The two tables share the same five topics (in alphabetical order):

- Topic 1 is "Election" — the 2012 US presidential campaign between Barack Obama and Mitt Romney.
- Topic 2 is "Gun Laws" due to the words "gun", "moral", "legisl", "individu", "weapon", and "violenc". This topic is about whether people are allowed to have their own guns.
- Topic 3 is "News Coverage". The words "malkin" and "gregori" refer to the American political commentators Michelle Malkin and Dick Gregory, respectively.
- Topic 4 is "Racism" mainly because of the word "sharpton". Al Sharpton is known for his engagement in civil right cases involving racism. Moreover, the word "hoodi" refers to a hoodie because Trayvon Martin was wearing a hooded sweatshirt at the time of the shooting incident.
- Topic 5 is "Zimmerman Trial" due to the words "perjuri" and "prosecutor".

In Table 4, some bi-grams make it easier to identify the topics. For instance, the bi-grams "stand_ground" (stand your ground law) and "self_defens" make it clear that the Topic 2 is about gun laws. In Topic 4, the bi-gram "black_panther" (Black Panther Party) shows discussion about racism.

The key bi-grams to the Trayvon Martin dataset, "trayvon_martin" and "georg_zimmerman", are not present in the top 10 distinctive words. The whole corpus is related to the two terms, but given these bi-grams, it is difficult to know which sub-topic they come from. As a result, they are not "distinctive" enough within the Trayvon Martin dataset.

Table 3: Before n-gramming: Top 10 **distinctive** words for each topic from LDA.

Topic 1 Election	Topic 2 Gun Laws	Topic 3 News Coverage	Topic 4 Racism	Topic 5 Zimmerman Trial
barack	alec	tommi	hoodi	spokesmancom
mitt	gun	anonym	dispatch	surl
presid	moral	tue	mar	perjuri
obama	legisl	malkin	sharpton	corey
administr	group	idiot	minut	dave
candid	individu	stupid	polic	bond
tax	weapon	gregori	martin	expert
health	ground	palin	trayvon	access
congress	violenc	rich	walk	prosecutor
romney	violat	liber	unarm	comment

Table 4: After n-gramming: Top 10 **distinctive** words for each topic from LDA.

Topic 1 Election	Topic 2 Gun Laws	Topic 3 News Coverage	Topic 4 Racism	Topic 5 Zimmerman Trial
mitt_romney	spokesmancom	malkin	**black_panther**	**comment_dave**
barack_obama	york	palin	sharpton	**dave_surl**
alec	retreat	tue	panther	perjuri
congress	**stand_ground**	fox	young	surl
administr	access	hanniti	white	dave
econom	**neighborhood_watch**	tommi	race	bond
tax	hoodi	msnbc	black	expert
economi	sanford	dog	trayvon	comment
senat	**self_defens**	anonym	racism	voic
unit_state	florida	mitt	drug	corey

4.4. Quantitative comparison: increase in word distinctivity

This section compares the results in Tables 3 and 4. Since "People don't ask how; they ask how much", we need to quantify the improvement of topic modeling results from n-gramming. Hence we define the "change" of word distinctivity for words as the difference between the word distinctivity before and after n-gramming. The definition of the "change" is slightly different for words (uni-grams) and for bi-grams.

4.4.1. Words (uni-grams)

In mathematical terms, the **change** of word distinctivity is written as

$$\max_i\Big[P\big(\text{topic } i \mid \text{word } j, \text{ data after}\big)\Big] - \max_i\Big[P\big(\text{topic } i \mid \text{word } j, \text{ data before}\big)\Big],$$

where "data before" refers to the data before n-gramming, and "data after" refers to the data after n-gramming.

For example, if the topic assignment vector for a word w changes from $(0.8, 0.1, 0.1)$ to $(0.05, 0.9, 0.05)$, the change of word distinctivity for w is $0.9 - 0.8 = 0.1$. Since n-gramming increases the word distinctivity of w, this is evidence of n-gramming improving the topic modeling results.

Note that word distinctivity is defined as the maximum value of the components in the vector, so the order of the components does not matter. The word distinctivity values 0.8 and 0.9 do not have to be in the same position in the topic assignment vector.

4.4.2. Bi-grams

Similarly, the **change** of word distinctivity for a bi-gram is written as

$$\max_i\Big[P\big(\text{topic } i \mid \text{bi-gram } b, \text{ data after}\big)\Big] - \max_i\Big[P\big(\text{topic } i \mid \text{bi-gram } b, \text{ data before}\big)\Big],$$

where the first component refers to the word distinctivity of the bi-gram b **after** n-gramming.

We need to explicitly define the second component, since the bi-gram was not formed before the n-gramming step. A bi-gram contains two words, so the word distinctivity of the bi-gram b **before** n-gramming should be the higher of the two words' distinctivity values. That is, the "baseline" of a bi-gram's word distinctivity is the highest distinctivity of the two words.

In mathematical terms, the word distinctivity of a bi-gram before n-gramming is defined as

$$\max_i\Big[P\big(\text{topic } i \mid \text{bi-gram } b, \text{ data before}\big)\Big] =$$
$$= \max\Big\{\max_i\Big[P\big(\text{topic } i \mid \text{word } 1, \text{ data before}\big)\Big], \max_i\Big[P\big(\text{topic } i \mid \text{word } 2, \text{ data before}\big)\Big]\Big\}.$$

The bi-grams of interest would have this feature: The two words have low distinctivity, but the formed bi-gram has high distinctivity.

4.4.3. Increase in word distinctivity

Table 5 presents the eight bi-grams in the Trayvon Martin dataset whose word distinctivity increased at least 0.15 after n-gramming. They are listed in descending order of the increase in distinctivity. These bi-grams are of interest because they start with a low distinctivity of each word, but the bi-gram has a high distinctivity. In this way, the bi-gram almost always appears in a certain topic, so the uncertainty in topic identification decreases.

For example, the bi-gram "black_panther" is highly distinctive (90.1%) because it refers to the Black Panther Party. But if we look at the words "black" and "panther" separately, the meaning is not as clear. "Black" may refer to the color or the race, and "panther" may refer to the animal or the movie *Panther*[3].

For another example, the bi-gram "neighborhood_watch" is also highly distinctive (85.2%) because it refers to a group whose goal is to prevent crime within a neighborhood. If we break down the bi-gram, "neighborhood" and "watch" are common words and are often used in everyday English.

Table 5: The increase of word distinctivity in bi-grams.

Bi-gram	Distinctivity of Bi-gram	Distinctivity of Word 1	Distinctivity of Word 2	Increase in Distinctivity
black_panther	0.901	0.484	0.525	0.376
unit_state	0.848	0.479	0.475	0.369
self_defens	0.798	0.370	0.476	0.322
neighborhood_watch	0.852	0.584	0.463	0.269
white_hous	0.793	0.354	0.528	0.265
stand_ground	0.910	0.632	0.687	0.222
year_old	0.574	0.385	0.391	0.183
new_york	0.491	0.339	0.314	0.152

4.5. Qualitative comparison: improvement of topic modeling results

Last but not least, we also performed a qualitative comparison of the topic modeling results, and we examined the selected words for each topic under different versions of LDA implementations. The versions are determined by whether the input data were before or after n-gramming, and whether the selection criteria was the traditional posterior (Section 4.2) or the word distinctivity (Section 4.3).

The two LDA versions with word distinctivity both identified the topic "Racism" (Tables 3 and 4), while the versions using the traditional posterior did not (Tables 1 and 2). Next, we would like to present interesting results for the topics "Election" and "Gun Laws".

[3] https://en.wikipedia.org/wiki/Panther_(film)

Table 6 compares the selected words for the topic "Election". The word "obama" appears in all four versions because Barack Obama ran for the 2012 US presidential election. Political party names such as "democrat" and "republican" appear only in the traditional posterior versions, and they are replaced with nonpartisan government words in the word distinctivity versions, such as "congress", "tax", and "administr".

Although we performed n-gramming to the corpus, the bi-grams show up only when we applied the word distinctivity criteria to choose words (tokens, to be exact) for each topic. In the fourth column of Table 6, the two presidential candidate names "mitt_romney" and "barack_obama" are on top of the list. This is much more informative than the other three versions, showing that the combination of n-gramming and word distinctivity works better.

Table 6: Comparison of selected words for the topic "Election".

Before N-gramming Traditional Posterior	After N-gramming Traditional Posterior	Before N-gramming Word Distinctivity	After N-gramming Word Distinctivity
obama	obama	barack	**mitt_romney**
presid	presid	mitt	**barack_obama**
year	year	presid	alec
romney	american	obama	congress
american	romney	administr	administr
democrat	republican	candid	econom
said	govern	tax	tax
nation	law	health	economi
women	women	congress	senat
govern	countri	romney	**unit_state**

Table 7 also attempts to compare the selected words for the topic "Gun Laws", but this topic does not exist in the version of after n-gramming and traditional posterior. That is, when using the traditional posterior probability as the selection criteria, we could reveal the topic "Gun Laws" before n-gramming, but could not do so after n-gramming.

Table 7: Comparison of selected words for the topic "Gun Laws".

Before N-gramming Traditional Posterior	After N-gramming Traditional Posterior	Before N-gramming Word Distinctivity	After N-gramming Word Distinctivity
law	N/A	alec	spokesmancom
gun		gun	york
ground		moral	retreat
stand		legisl	**stand_ground**
forc		group	access
alec		individu	**neighborhood_watch**
mar		weapon	hoodi
defend		ground	sanford
reason		violenc	**self_defens**
shoot		violet	florida

Performing n-gramming is expected to result in information gain of the topic modeling results, but we did not see the desired outcome.

On the contrary, the combination of n-gramming and word distinctivity works well in Table 7. Three informative bi-grams appear: "stand_ground" (stand your ground law), "neighborhood_watch", and "self_defens". It is much easier to infer the context from the bi-gram "stand_ground" than from the two separate words "stand" and "ground".

In short, n-gramming improves the topic modeling results, but it is difficult to show the improvement without using word distinctivity as the selection criteria of `top.topic.words`.

4.6. Limitations

We can safely assume that topics with more highly distinctive tokens are better defined, but we are unable to prove or disprove whether n-gramming increases our ability to correctly guess the document's topic proportions, given that the Trayvon Martin dataset does not contain the ground truth. Existing literature [33] also shows that evaluating an unsupervised model is difficult. However, it is possible to create a synthetic dataset with pre-defined topic proportions from Wikipedia articles [11], then we can use the new dataset to test the hypothesis.

5. DISCUSSION

Adequate text data preprocessing helps in topic modeling, and the improvement of results from n-gramming can be quantified by word distinctivity. After we identify and combine the words with special meaning into bi-grams, more semantic information is retained, leading to a stronger signal in topic classification.

In this way, the text data cleaning quality can be measured in terms of topic modeling results at the word level. By retaining special phrases (i.e., bi-grams), n-gramming increases the word distinctivity, and word distinctivity improves the quality of the LDA-identified topics. Some bi-grams have a higher distinctivity than either of the two word components, so the signal of topic assignment is stronger after the bi-gram is formed.

On the other hand, the effect of n-gramming at the corpus level is still unclear, since bi-grams account for only a small part of a text database [2]. We attempted to measure the prediction power from the corpus after n-gramming, and we used "perplexity" as a single number to summarize how well the topic model predicts the remaining words, given a part of the document [4]. The perplexity is the effective number of equally likely words based on the model, so the perplexity is inversely proportional to the precision of the predictive model output. Nevertheless, the t-test results were inconclusive [11]. Therefore, more research is needed to evaluate the overall improvement of topic modeling results after n-gramming.

6. FUTURE DIRECTIONS

This research is a start of quantifying topic model performance, and we hope to further improve the text data cleaning process and statistically evaluate the results. We quantified the increase of word distinctivity from n-gramming in Section 4.4, and we are still looking for a metric to numerically measure the quality of selected topic words. Section 4.5 gives a preliminary and qualitative comparison to show evidence that the word distinctivity is a better selection criteria than the traditional posterior.

Moreover, a potential solution to the LDA labeling issue (Section 4.2) is seeded topic models [46]. The seeded topic model preassigns each topic with a word, then the model "grows" each topic from the preassigned word. An example is to start the first topic (Election) with "barack_obama", the second topic (Gun Law) with "self_defens", and the third topic (Racism) with "sharpton" (Al Sharpton). Other deterministic relabeling strategies are described in [37] and the R package `label.switching` [36].

A new possible direction is to compare various existing methods and determine which method is most appropriate for which type of text corpus. We used a list of predefined stop words to remove the words with little semantic meaning in the corpus, but there may exist better ways to perform text data cleaning.

Another possible extension is to combine the n-gram construction and the LDA into a single Bayesian hierarchical model. By introducing additional latent variables to determine the posterior probability of a phrase to be a multiword expression, we can go beyond the frequentist approach of testing for n-grams.

ACKNOWLEDGMENTS

The author would like to thank her PhD advisor at Duke University, Dr. David Banks, for his support on this research project. The author would also like to thank the people who also worked on this dataset, Derek Owens-Oas and Teague Rhine Henry, for the discussion. The author is also grateful for the comments from Andrew Raim (U.S. Census Bureau) and the anonymous reviewers.

REFERENCES

[1] Au, T.C. (2014). *Topics in Computational Advertising*, PhD dissertation, Duke University.

[2] Bekkerman, R. and Allan, J. (2004). *Using bigrams in text categorization*, Technical report, IR-408, Center of Intelligent Information Retrieval, University of Massachusetts Amherst.

[3] Bird, S.; Klein, E. and Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Inc..

[4] Blei, D. and Lafferty, J. (2006). Correlated topic models, *Advances in Neural Information Processing Systems*, **18**, 147.

[5] Blei, D.M. and Lafferty, J.D. (2009). Visualizing topics with multi-word expressions, *arXiv preprint*, arXiv:0907.1013.

[6] Blei, D.M.; Ng, A.Y. and Jordan, M.I. (2003). Latent Dirichlet allocation, *Journal of Machine Learning Research*, **3**, 993–1022.

[7] Bouchet-Valat, M. (2014). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library.* R package version 0.5.1.

[8] Broder, A.Z.; Glassman, S.C.; Manasse, M.S. and Zweig, G. (1997). Syntactic clustering of the web, *Computer Networks and ISDN Systems*, **29**(8-13), 1157–1166. ISDN stands for Integrated Services Digital Network.

[9] Brown, P.F.; Desouza, P.V.; Mercer, R.L.; Della Pietra, V.J. and Lai, J.C. (1992). Class-based n-gram models of natural language, *Computational Linguistics*, **18**(4), 467–479.

[10] Cavnar, W.B. and Trenkle, J.M. (1994). *N-gram-based text categorization.* In "Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval", vol. 161175, Citeseer.

[11] Chai, C.P. (2017). *Statistical Issues in Quantifying Text Mining Performance*, PhD dissertation, Duke University.

[12] Chang, J. (2015). *lda: Collapsed Gibbs sampling methods for topic models.* R package version 1.4.2.

[13] Chater, N. and Manning, C.D. (2006). Probabilistic models of language processing and acquisition, *Trends in Cognitive Sciences*, **10**(7), 335–344.

[14] CrowdFlower (2016). *2016 data science report.* http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

[15] Doraisamy, S. and Rüger, S. (2003). Robust polyphonic music retrieval with n-grams, *Journal of Intelligent Information Systems*, **21**(1), 53–70.

[16] Eck, M.; Vogel, S. and Waibel, A. (2005). *Low cost portability for statistical machine translation based on n-gram frequency and tf-idf.* In "International Workshop on Spoken Language Translation (IWSLT) 2005".

[17] Feldman, R. and Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press.

[18] Francis, L. and Flynn, M. (2010). *Text mining handbook*, Casualty Actuarial Society E-Forum. http://www.casact.org/pubs/forum/10spforum/completes10.pdf

[19] Fürnkranz, J. (1998). A study using n-gram features for text categorization, *Austrian Research Institute for Artificial Intelligence*, **3**(1998), 1–10.

[20] Griffiths, T.L. and Steyvers, M. (2004). Finding scientific topics, *Proceedings of the National Academy of Sciences*, **101**, 5228–5235.

[21] Grün, B. and Hornik, K. (2011). Topicmodels: An R package for fitting topic models, *Journal of Statistical Software*, **40**(13), 1–30.

[22] Han, J.; Pei, J. and Kamber, M. (2011). *Data Mining: Concepts and Techniques*, Elsevier.

[23] Henry, T. (2016). *Quick ngram processing script.* https://github.com/trh3/NGramProcessing

[24] Henry, T.R.; Banks, D.; Owens-Oas, D. and Chai, C. (2019). Modeling community structure and topics in dynamic text networks, *Journal of Classification*, **36**(2), 322–349.

[25] Hollink, V.; Kamps, J.; Monz, C. and De Rijke, M. (2004). Monolingual document retrieval for European languages, *Information retrieval*, **7**(1-2), 33–52.

[26] Kamps, J.; Monz, C.; De Rijke, M. and Sigurbjörnsson, B. (2003). *Language-dependent and language-independent approaches to cross-lingual text retrieval.* In "Workshop of the Cross-Language Evaluation Forum for European Languages", pp. 152–165, Springer.

[27] Lin, C. and He, Y. (2009). *Joint sentiment/topic model for sentiment analysis.* In "Proceedings of the 18th ACM Conference on Information and Knowledge Management", pp. 375–384, ACM.

[28] Liu, J.S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem, *Journal of the American Statistical Association*, **89**(427), 958–966.

[29] Marino, J.B.; Banchs, R.E.; Crego, J.M.; de Gispert, A.; Lambert, P.; Fonollosa, J.A.R. and Costa-Jussà, M.R. (2006). N-gram-based machine translation, *Computational Linguistics*, **32**(4), 527–549.

[30] McNamee, P. and Mayfield, J. (2003). *Jhu/apl experiments in tokenization and non-word translation.* In "Comparative Evaluation of Multilingual Information Access Systems", pp. 85–97, Springer.

[31] Microsoft (2019). *Extract n-gram features from text.* https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-text

[32] Niesler, T.R. and Woodland, P.C. (1996). *A variable-length category-based n-gram language model.* In "1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings", vol. 1, pp. 164–167, IEEE.

[33] Nogueira, B.M.; Moura, M.F.; Conrado, M.S.; Rossi, R.G.; Marcacini, R.M. and Rezende, S.O. (2008). *Winning some of the document preprocessing challenges in a text mining process.* In "Embrapa Informática Agropecuária – Artigo em Anais de Congresso (ALICE)". In "Simpósio Brasileiro de Banco de Dados", 23.

[34] Osborne, J.W. (2012). *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data*, Sage Publications.

[35] Owens-Oas, D. (2018). *Probabilistic Models for Text in Social Networks*, PhD dissertation, Duke University.

[36] Papastamoulis, P. (2015). label.switching: An R package for dealing with the label switching problem in mcmc outputs, *arXiv preprint*, arXiv:1503.02271.

[37] Rodriguez, C.E. and Walker, S.G. (2014). Label switching in Bayesian mixture models: Deterministic relabeling strategies, *Journal of Computational and Graphical Statistics*, **23**(1), 25–45.

[38] Rosen-Zvi, M.; Griffiths, T.; Steyvers, M. and Smyth, P. (2004). *The author-topic model for authors and documents.* In "Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence", pp. 487–494, AUAI Press.

[39] Schuman, H. and Presser, S. (1996). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, Sage.

[40] Silge, J. and Robinson, D. (2017). *Text Mining with R: A Tidy Approach*, O'Reilly Media, Inc..

[41] Simply Statistics (2012). *On the relative importance of mathematical abstraction in graduate statistical education.* https://simplystatistics.org/2012/08/08/on-the-relative-importance-of-mathematical-abstraction/

[42] Soriano, J.; Au, T. and Banks, D. (2013). Text mining in computational advertising, *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **6**(4), 273–285.

[43] Squire, M. (2015). *Clean Data*, Packt Publishing Ltd.

[44] Statistical Services Centre (2001). *Approaches to the analysis of survey data*, Technical report, The University of Reading, United Kingdom.

[45] Steorts, R.C.; Ventura, S.L.; Sadinle, M. and Fienberg, S.E. (2014). *A comparison of blocking methods for record linkage*. In "International Conference on Privacy in Statistical Databases", pp. 253–268, Springer.

[46] Stephens, M. (2000). Dealing with label switching in mixture models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(4), 795–809.

[47] Suen, C.Y. (1979). N-gram statistics for natural language understanding and text processing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**(2), 164–172.

[48] SurveyMonkey (2018). *How to analyze survey data*. https://www.surveymonkey.com/mp/how-to-analyze-survey-data/ (n.d., accessed in 2018).

[49] Vatanen, T.; Väyrynen, J.J. and Virpioja, S. (2010). *Language identification of short text segments with n-gram models*. In "Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)".

[50] Wallach, H.M. (2006). *Topic modeling: Beyond bag-of-words*. In "Proceedings of the 23rd International Conference on Machine Learning", pp. 977–984, ACM.

[51] Wei, Z.; Miao, D.; Chauchat, J.-H.; Zhao, R. and Li, W. (2009). N-grams based feature selection and text representation for Chinese text classification, *International Journal of Computational Intelligence Systems*, **2**(4), 365–374.

[52] Xiong, W.; Wu, L.; Alleva, F.; Droppo, J.; Huang, X. and Stolcke, A. (2018). *The Microsoft 2017 conversational speech recognition system*. In "2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", pp. 5934–5938, IEEE.

[53] Zens, R. and Ney, H. (2006). *N-gram posterior probabilities for statistical machine translation*. In "Proceedings on the Workshop on Statistical Machine Translation", pp. 72–77.

[54] Zhang, S.; Zhang, C. and Yang, Q. (2003). Data preparation for data mining, *Applied Artificial Intelligence*, **17**(5-6), 375–381.

[55] Zhang, Y.; Jin, R. and Zhou, Z.-H. (2010). Understanding bag-of-words model: A statistical framework, *International Journal of Machine Learning and Cybernetics*, **1**(1-4), 43–52.

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 221–249

STRONG UNIFORM CONSISTENCY RATES OF CONDITIONAL DENSITY ESTIMATION IN THE SINGLE FUNCTIONAL INDEX MODEL FOR FUNCTIONAL DATA UNDER RANDOM CENSORSHIP

Authors: Nadia Kadiri
– University Djillali LIABES of Sidi Bel Abbes,
Algeria
nad.kad06@yahoo.com

Mustapha Meghnafi
– Departement of Mathematics, Faculty of Exact Sciences, University of Béchar,
Algeria
megnafi3000@yahoo.fr

Abbes Rabhi
– Laboratory of Mathematics, University of Sidi Bel Abbes,
Algeria
rabhi_abbes@yahoo.fr

Received: July 2018 Revised: November 2019 Accepted: February 2020

Abstract:

- The main objective of this paper is to investigate the estimation of conditional density function based on the single-index model in the censorship model when the sample is considered as an independent and identically distributed (i.i.d.) random variables. First of all, a kernel type estimator for the conditional density function (*cond-df*) is introduced. Afterwards, the asymptotic properties are stated when the observations are linked with a single-index structure. The pointwise almost complete convergence and the uniform almost complete convergence (with rate) of the kernel estimate of this model are established. As an application the conditional mode in functional single-index model is presented. Finally, a simulation study is carried out to evaluate the performance of this estimate.

Keywords:

- *conditional density; functional single-index process; functional random variable; nonparametric estimation; small ball probability.*

AMS Subject Classification:

- 62G05, 62G99, 62M10.

1. INTRODUCTION

Over the last two decades, functional data modeling was highly considered in the statistical literature. The new generation of electronic devices is now allowing practitioners to have access to data continuously (over time and/or space). This change in the data structure raised several challenging statistical problems in analyzing curve-type data. In practice, one can observe functional data in several fields such as climatology, stock market analysis, signal processing, satellite images analysis, etc. For an overview of the functional data analysis, the readers can refer to Ramsay and Silverman [30, 31], Masry [26], Ferraty and Vieu [16], Cuevas [8], Hsing and Eubank [21], Goia and Vieu [18] and the references therein.

Modeling of the relationship between two concomitant variables is one of the most relevant tasks in functional data analysis. In this paper, we are interested in using semi-parametric approach to model the conditional density of a real-valued response variable given an infinite dimensional (functional) covariate. A dimension reduction approach, based on single index model, is used in this paper to estimate the conditional mode whenever the response variable is affected by a right censorship phenomenon.

The problem of estimating the conditional density function has taken considerable attention in the past for both independent and dependent data. Conditional density estimation of a scalar response given a scalar/multivariate covariate has been widely used to estimate some characteristic features of a data set, such as the conditional mode, and gained considerable interest in the statistical literature. For completely observed data, several nonparametric approaches have been proposed. Samanta and Thavaneswaran [32] showed that, under some regularity conditions, the kernel estimator of the conditional mode function was consistent and asymptotically normally distributed. Mehra *et al.* [27] established the law of iterated logarithm (LIL). Under random censoring, Ould-Saïd and Cai [29] established the uniform strong consistency of a nonparametric estimator of the censored conditional mode function, in the i.i.d. case using a step function for the interest random variable. For their part, Khardani *et al.* [22] obtained the strong consistency with rate and asymptotic normality. Ould-Saïd [28] constructed a kernel estimator of the conditional quantile under an i.i.d. censorship model and established its strong uniform convergence rate. For the censored dependent case, Khardani *et al.* [23] obtained the strong consistency with rate for the α-mixing framework. The asymptotic normality of the conditional mode estimator for the censored dependent case was proved by Khardani *et al.* [24].

Many authors are interested in the estimation of the conditional mode of a scalar response given a functional covariate. The kernel-type estimators of some characteristics of the conditional cumulative distribution function and the successive derivatives of the conditional density were introduced by Ferraty *et al.* [13]. Some asymptotic properties were established with a particular application to the conditional mode and conditional quantiles. An application to a chemometrical data set coming from food industry is also presented. The uniform strong consistency with rates and the asymptotic normality for the kernel conditional mode estimator were obtained by Ezzahrioui and Ould-Saïd [10] in the i.i.d. case. The asymptotic normality, under α-mixing conditions, of the kernel conditional quantile estimator, was established by Ezzahrioui and Ould-Saïd [11].

In multivariate statistics, where the vector of covariates belongs to a high dimensional but finite space, single index model represents one of the well-known semi-parametric models which allows to reduce the dimensionality of the covariate space and, at the same time, gives flexibility in describing the relationship between the response and the covariate through an unknown link function. Indeed, single index model reduces the curse of dimensionality effect known in pure nonparametric estimation methods and it is always seen as a reasonable compromise between nonparametric and parametric models. Consequently, reducing the dimensionality can be of great interest in practice. For instance, it allows to increase the prediction accuracy and to improve the interpretability of the relationship between a response variable with a vector of covariates. For more details about the advantages of single index models in finite dimensional space setting, the reader can be referred to [19], [20], [33] and the references therein. In our infinite dimensional purpose, we use the terminology *functional nonparametric*, where the word *functional* referees to the infinite dimensionality of the data and where the word *nonparametric* referees to the infinite dimensionality of the model. Such *functional nonparametric* statistics is also called *doubly infinite dimensional* (see Ferraty and Vieu [15], for more details).

The extension of the single index model to the functional data framework was introduced first in Ferraty *et al.* [14] to estimate semi-parametrically the regression operator where the response variable is real-valued and the covariate is a functional random variable. The single functional index model (SFIM) assumes that a functional explanatory variable acts on a scalar response only through its projection on one functional direction. The SFIM was intensively extended to estimate several statistical parameters describing the shape of the conditional distribution. For instance, Aït-Saidi *et al.* [1] were interested in using SFIM to estimate the regression operator and suggest to use a cross-validation procedure allowing the estimated the unknown link function as well as the unknown functional index. Attaoui [3] and Attaoui and Ling [6] studied, respectively, the estimation of the conditional density and the conditional cumulative distribution function based on a SFIM and assuming that the data satisfy a strong mixing condition. Bouchentouf *et al.* [7] were interested in the semi-parametric estimation of the hazard function. Goia and Vieu [17] presented a methodology allowing to approximate in a semi-parametric way the unknown regression operator through a single index approach and by taking possible structural changes into account. Furthermore, Ling *et al.* [25] obtained the asymptotic normality of the conditional density estimator and the conditional mode estimator for the α-mixing dependence functional time series data.

The main contribution of this work, is to establish the pointwise almost complete convergence and the uniform almost complete convergence (with rate) of the conditional density estimator in the single functional index model in i.i.d. case under random censorship, this result will be applied to obtain the convergence rates of the conditional mode estimator. Moreover, we prove the asymptotic normality of the estimators of conditional density function and conditional mode. The layout of the paper is as follows: Section 1 presents the functional nonparametric framework. In Section 2 we treat the almost complete convergence, while in Section 3 the uniform version is studied. The asymptotic normality is given in Section 4, and a simulation study is provided in Section 5. Finally, all the proofs of the theoretical results are given in Section 6.

1.1. The functional nonparametric framework

Consider a random pair (X,T) where T is valued in $\mathbb{R}$ and X is valued in some infinite dimensional Hilbertian space $\mathcal{H}$ with scalar product $\langle\cdot,\cdot\rangle$. Let $(X_i,T_i)_{i=1,\ldots,n}$ be the statistical sample of pairs which are identically distributed like (X,T), but not necessarily independent. X is called functional random variable *f.r.v.*

As example, in the classical regression case, the important parameter whose one assumed existence is the regression function of Y knowing the covariate X, denoted $r(x) = \mathbb{E}(Y\,|\,X=x)$, $X,Y\in\mathbb{R}^d\times\mathbb{R}$. For this model, the non-parametric method considers only regularity assumptions on the function r. Obviously, this method has some drawbacks. One can cite the problem of curse of dimensionality. This problem appears when the number of regressors d increases, the rate of convergence of the nonparametric estimator r which is supposed k times differentiable is $\mathcal{O}(n^{-k/2k+d})$ deteriorate. The second drawback is the lack of means to quantify the effect of each explanatory variable. To alleviate in these drawbacks, an alternative approach is naturally provided by the semi-parametric model which supposes the introduction of a parameter on the regressors. Assume that the conditional expectation of T given X is done through a fixed functional index θ in $\mathcal{H}$, such that by writing than the regression function is of the form

$$\mathbb{E}_\theta\big(T\,|\,X\big) \;=\; \mathbb{E}\big(T\,|\,\langle X,\theta\rangle = x\big)\,.$$

This model was introduced by Ferraty *et al.* [14] and we can refer to Attaoui *et al.* [5] for details. From this model, let $f(\theta,\cdot,x)$ be the conditional density of Y given $\langle X,\theta\rangle = \langle x,\theta\rangle$ for $x\in\mathcal{H}$, which also shows the relationship between X and Y but it often unknown.

Let $(T_i)_{i\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables, and assume that they form a strictly stationary sequence of lifetimes. Suppose that there exists a sample of i.i.d. censoring random variable (r.v) $(C_i)_{i\geq 1}$ with common unknown continuous distribution function (df).

In the censored framework, the observed random variables are the triplets (Y_i,δ_i,X_i) with

$$Y_i = \min\big\{T_i, C_i\big\} \qquad \text{and} \qquad \delta_i = \mathbf{1}_{T_i\leq C_i}\,,\; 1\leq i\leq n\,,$$

where both of T_i and C_i are expected to exhibit some kind of dependence which ensures the identifiability of the model.

In biomedical case studies, it is assumed that C_i and (T_i,X_i) are independent, this condition is plausible whenever the censoring is independent of the patient's modality.

The Kernel estimator $f_n(\theta,\cdot,x)$ of $f(\theta,\cdot,x)$ is defined by:

$$f_n(\theta,t,x) = \frac{h_H^{-1}\displaystyle\sum_{i=1}^{n} K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)\,H\big(h_H^{-1}(t-T_i)\big)}{\displaystyle\sum_{i=1}^{n} K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)}\,, \tag{1.1}$$

where the functions K and H are kernels and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) a sequence of positive real numbers.

The Kernel type estimator of the conditional density $f(\theta,\cdot,x)$ adapted for censorship model, can be reformulated from the expression (1.1) as follows:

$$\widetilde{f}(\theta,t,x) \;=\; \frac{h_H^{-1}\displaystyle\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}(Y_i)}\,K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)\,H\big(h_H^{-1}(t-Y_i)\big)}{\displaystyle\sum_{i=1}^{n}K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)}\,. \tag{1.2}$$

In practice $\bar{G}(\cdot) = 1-G(\cdot)$ is unknown, then using Kaplan and Meier (1958) estimator, $\bar{G}_n(\cdot)$ will be given as

$$\bar{G}_n(t) \;=\; 1-G_n(t) \;=\; \begin{cases} \displaystyle\prod_{i=1}^{n}\bigg(1-\frac{1-\delta_{(i)}}{n-i+1}\bigg)^{\mathbf{1}\{Y_{(i)}\le t\}}, & \text{if } t\le Y_{(n)}\,,\\ 0\,, & \text{if } t> Y_{(n)}\,, \end{cases} \tag{1.3}$$

where $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ are the order statistics of Y_i and $\delta_{(i)}$ is the non-censoring indicator corresponding to $Y_{(i)}$.

Therefore, estimator of the conditional density function $f(\theta,\cdot,x)$ is given by

$$\widehat{f}(\theta,t,x) \;=\; \frac{h_H^{-1}\displaystyle\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}_n(Y_i)}\,K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)\,H\big(h_H^{-1}(t-Y_i)\big)}{\displaystyle\sum_{i=1}^{n}K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)}\,. \tag{1.4}$$

2. ASYMPTOTIC STUDY

2.1. Pointwise almost complete rate of convergence

In the following, for any $x\in\mathcal{H}$, let N_x be a fixed neighborhood of x and $\mathcal{S}_{\mathbb{R}}$ is a fixed compact of $\mathbb{R}^+$. Denote $B_\theta(x,h) = \{f\in\mathcal{H}:\ 0<|\langle x-f,\theta\rangle|<h\}$ the ball of center x and radius h. Assume that $(C_i)_{i\ge1}$ are independent and $\tau_G<\infty$ where $\tau_G := \sup\{t : G(t)<1\}$ and let τ be a positive real number such that $\tau<\tau_G$.

In order to establish the almost complete (a.co.) convergence of our estimator, we need some regular hypotheses as follows:

- **(H1)** $\forall h>0,\ \mathbb{P}(X\in B_\theta(x,h)) = \phi_{\theta,x}(h) > 0$;
- **(H2)** The conditional density $f(\theta,t,x)$ satisfies the Hölder condition, i.e., $\forall(x_1,x_2)\in N_x\times N_x,\ \forall(t_1,t_2)\in\mathcal{S}_{\mathbb{R}}^2$,

$$\Big|f(\theta,t_1,x_1)-f(\theta,t_2,x_2)\Big| \le C_{\theta,x}\Big(\|x_1-x_2\|^{b_1}+|t_1-t_2|^{b_2}\Big)\,,\qquad b_1>0\,,\ b_2>0\,;$$

(**H3**) H is a bounded function, such that $\forall (t_1,t_2)\in\mathbb{R}^2,\ |H(t_1)-H(t_2)|\le C|t_1-t_2|$ and $\displaystyle\int |t|^{b_2} H(t)\,dt<\infty$;

(**H4**) K is a positive bounded function with support $[-1,1]$, such that $\forall u\in(0,1)$, $0<K(u)$;

(**H5**) The bandwidths h_K and h_H satisfy

$$\lim_{n\longrightarrow\infty} h_K=0\,,\qquad \frac{\log n}{n\,h_H\,\phi_{\theta,x}(h_K)}\underset{n\to\infty}{\longrightarrow}0\,.$$

- **Comments on the hypotheses**

Our hypotheses are very standard for the conditional density estimation in single functional index model, which have been adopted by Attatoui *et al.* [5]. Hypotheses (H3) and (H5) are technical conditions and are similar to those done in Ferraty and Vieu [16].

Proposition 2.1. *Under conditions (H1)–(H5), we have as n goes to infinity*

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}(\theta,t,x)-f(\theta,t,x)\right| = \mathcal{O}\Big(h_K^{b_1}+h_H^{b_2}\Big)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,x}(h_K)}}\right). \tag{2.1}$$

Proof of Proposition 2.1: Consider now, for $i=1,...,n$, in what follows, let's denote:

$$K_i(\theta,x)=K\Big(h_K^{-1}\big(\langle x-X_i,\theta\rangle\big)\Big)\,,\qquad H_i(t)=H\big(h_H^{-1}(t-Y_i)\big)\,,\qquad \bar{G}_i=\bar{G}(Y_i)\,,$$

$$\begin{aligned}
\widehat{f}_N(\theta,t,x) &= \frac{1}{n\,h_H\,\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}_n(Y_i)}\,K_i(\theta,x)\,H_i(t)\,,\\
\widetilde{f}_N(\theta,t,x) &= \frac{1}{n\,h_H\,\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}(Y_i)}\,K_i(\theta,x)\,H_i(t)\,,\\
\widehat{F}_D(\theta,x) &= \frac{1}{n\,\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}K_i(\theta,x)\,.
\end{aligned}$$

The proof is based on the following decomposition, valid for any $t\in\mathcal{S}_{\mathbb{R}}$:

$$\begin{aligned}
\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}(\theta,t,x)-f(\theta,t,x)\right| &\le \frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left\{\left|\widehat{f}_N(\theta,t,x)-\widetilde{f}_N(\theta,t,x)\right|\right\}\\
&\quad+\frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left\{\left|\widetilde{f}_N(\theta,t,x)-\mathbb{E}\widetilde{f}_N(\theta,t,x)\right|\right\}\\
&\quad+\frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left\{\left|\mathbb{E}\widetilde{f}_N(\theta,t,x)-f(\theta,t,x)\right|\right\}\\
&\quad+\frac{f(\theta,t,x)}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|1-\widehat{F}_D(\theta,x)\right|\,.
\end{aligned}\tag{2.2}$$

Finally, the proof of this proposition is a direct consequence of the following intermediate results. $\square$

Lemma 2.1. *Under hypotheses (H1)–(H4), and if*

$$n\, h_H\, \phi_{\theta,x}(h_K) \longrightarrow \infty\,, \qquad \frac{\log n}{n h_H\, \phi_{\theta,x}(h_K)} \underset{n\to\infty}{\longrightarrow} 0\,,$$

we have

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}} \left\{ \left| \widehat{f}_N(\theta,t,x) - \widetilde{f}_N(\theta,t,x) \right| \right\} \;=\; \mathcal{O}_{a.s.}\left(\frac{\log\log n}{n} \right).$$

The following lemma shows the asymptotic bias term of $\widetilde{f}_N(\theta,t,x)$ and $\widehat{f}_D(\theta,x)$ as n tends to infinity.

Lemma 2.2. *Under hypotheses (H1)–(H3), we have as $n \to \infty$*

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}} \left| \mathbb{E}\left[\widetilde{f}_N(\theta,t,x) \right] - f(\theta,t,x) \right| = \mathcal{O}\left(h_K^{b_1} + h_H^{b_2} \right). \tag{2.3}$$

The following result deals with the variance term of the right-hand side of (2.2) which is expressed by: $\sup_{t\in\mathcal{S}_{\mathbb{R}}} \left\{ \left| \widetilde{f}_N(\theta,t,x) - \mathbb{E}\, \widetilde{f}_N(\theta,t,x) \right| \right\}$. For $\widehat{F}_D(\theta,x) - \mathbb{E}\left[\widehat{F}_D(\theta,x) \right]$ the same arguments will be used with a slight difference.

Lemma 2.3. *Under hypotheses (H1), (H4)–(H5), as n goes to infinity, we have*

$$\widehat{F}_D(\theta,x) - \mathbb{E}\, \widehat{F}_D(\theta,x) \;=\; \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\, \phi_{\theta,x}(h_K)}} \right),$$

furthermore, we have

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \widehat{F}_D(\theta,x) \right| \le 1/2 \right) < \infty\,.$$

Lemma 2.4. *Under the conditions of Proposition 2.1, we have as $n \to \infty$*

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}} \left\{ \left| \widetilde{f}_N(\theta,t,x) - \mathbb{E}\, \widetilde{f}_N(\theta,t,x) \right| \right\} \;=\; \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\, h_H\, \phi_{\theta,x}(h_K)}} \right).$$

We conclude the proof of Proposition 2.1 by making use of the inequality (2.2), in conjunction with Lemmas 2.1, 2.2, 2.3 and 2.4.

2.2. Pointwise almost complete rate of convergence

In this section, we will consider the problem of the estimation of the conditional mode in the functional single-index model, denoted by $M_\theta(x)$. For this, we assume that $M_\theta(x)$ satisfies the following uniqueness property:

(H6) $\forall \varepsilon_0 > 0$, $\exists \eta > 0$, $\forall \varphi$:

$$\left| M_\theta(x) - \varphi(x) \right| \ge \varepsilon_0 \implies \left| f\big(\theta, \varphi(x), x\big) - f\big(\theta, M_\theta(x), x\big) \right| \ge \eta\,.$$

We estimate the conditional mode $M_\theta(x)$ with a random variable $\widehat{M_\theta}(x)$ such that

$$\widehat{M_\theta}(x) = \arg\sup_{t\in\mathcal{S}_\mathbb{R}} \hat{f}(\theta,t,x)\,.$$

The difficulty of the problem is naturally linked with the flatness of the function $f(\theta,t,x)$ around the mode M_θ. This flatness can be controlled by the number of vanishing derivatives at point M_θ, and this parameter will also have a great influence on the asymptotic rates of our estimates. More precisely, we introduce the following additional smoothness condition:

(**H7**) There exists some integer $j>1$ such that $\forall x\in\mathcal{S}_\mathcal{H}$, the function $f(\theta,\cdot,x)$ is j times continuously differentiable w.r.t. t on $\mathcal{S}_\mathbb{R}$ with

$$f^{(l)}\big(\theta,M_\theta(x),x\big)=0\,,\qquad \text{if } \ 1\le l<j\,,$$

and $f^{(j)}(\theta,\cdot,x)$ is uniformly continuous on $\mathcal{S}_\mathbb{R}$ such that

$$f^{(j)}\big(\theta,M_\theta(x),x\big)\neq 0\,,$$

where $f^{(j)}(\theta,\cdot,x)$ is the j^{th} order derivative of the conditional density $f(\theta,\cdot,x)$.

Theorem 2.1. *Under hypotheses of Proposition 2.1 and if the conditional density $f(\theta,\cdot,x)$ satisfies (H6) and (H7), then we get*

$$\left|\widehat{M_\theta}(x)-M_\theta(x)\right| = \mathcal{O}\left(h_K^{\frac{b_1}{j}}+h_H^{\frac{b_2}{j}}\right)+\mathcal{O}_{a.co.}\left(\left(\frac{\log n}{n\,h_H\,\phi_{\theta,x}(h_K)}\right)^{\frac{1}{2j}}\right). \tag{2.4}$$

Proof of Theorem 2.1: By the Taylor expansion of $f(\theta,t,x)$ in neighborhood of $M_\theta(x)$, we get

$$\widehat{f}\big(\theta,\widehat{M_\theta}(x),x\big) = f\big(\theta,M_\theta(x),x\big)+\frac{f^{(j)}\big(\theta,M_\theta^*(x),x\big)}{j!}\Big(\widehat{M_\theta}(x)-M_\theta(x)\Big)^j\,, \tag{2.5}$$

where $M_\theta^*(x)$ is between $M_\theta(x)$ and $\widehat{M_\theta}(x)$.

Combining the last equality with the fact that

$$\left|\widehat{f}\big(\theta,\widehat{M_\theta}(x),x\big)-f\big(\theta,M_\theta(x),x\big)\right| \le 2\sup_{t\in\mathcal{S}_\mathbb{R}}\left|\widehat{f}(\theta,t,x)-f(\theta,t,x)\right|,$$

allow to write:

$$\left|\widehat{M_\theta}(x)-M_\theta(x)\right|^j \le \frac{j!}{f^{(j)}(\theta,M_\theta^*,x)}\sup_{t\in\mathcal{S}_\mathbb{R}}\left|\widehat{f}(\theta,t,x)-f(\theta,t,x)\right|.$$

Using the second part of (H7) we obtain that

$$\exists\, c>0\,,\qquad \sum_{n=1}^{\infty}\mathbb{P}\Big(f^{(j)}(\theta,M_\theta^*,x)<c\Big)<\infty\,.$$

So, we would have

$$\left|\widehat{M_\theta}(x) - M_\theta(x)\right|^j = \mathcal{O}_{a.co.}\left(\sup_{t\in\mathcal{S}_\mathbb{R}}\left|\widehat{f}(\theta,t,x) - f(\theta,t,x)\right|\right). \tag{2.6}$$

Finally, Theorem 2.1 can be deduced from Proposition 2.1. □

Theorem 2.2. *Under the hypotheses of Proposition 2.1, thus we have*

$$\widehat{M_\theta}(x) - M_\theta(x) \underset{n\to\infty}{\longrightarrow} 0\,, \qquad \text{a.co.} \tag{2.7}$$

Proof of Theorem 2.2: Because the continuity of the function $f(\theta,t,x)$, we have, for all $\varepsilon > 0$, $\exists\,\eta(\varepsilon) > 0$ such that

$$\left|f(\theta,t,x) - f\big(\theta, M_\theta(x), x\big)\right| \le \eta(\varepsilon) \implies \left|t - M_\theta(x)\right| \le \varepsilon\,.$$

Therefore, for $t = \widehat{M_\theta}(x)$,

$$\mathbb{P}\Big(\big|\widehat{M_\theta}(x) - M_\theta(x)\big| > \varepsilon\Big) \le \mathbb{P}\Big(\big|f\big(\theta, \widehat{M_\theta}(x), x\big) - f\big(\theta, M_\theta(x), x\big)\big| > \eta(\varepsilon)\Big)\,.$$

Then, according to theorem, $\widehat{M_\theta} - M_\theta$ go almost completely to 0, as n goes to infinity. □

3. UNIFORM ALMOST COMPLETE CONVERGENCE AND RATE OF CONVERGENCE

In this section, we devote the result of the uniform version of Proposition 2.1. The study of the uniform consistency is a crucial tool for studying the asymptotic properties of all estimates of the functional index if is unknown. In the multivariate case, the uniform consistency is a standard extension of the pointwise one, nevertheless, in the studied case, it requires some additional tools and topological conditions (see Ferraty *et al.* [12]). Consequently, coupled with the conditions introduced antecedently, we need the following ones. Firstly, consider

$$\mathcal{S}_\mathcal{H} \subset \bigcup_{k=1}^{d_n^{\mathcal{S}_\mathcal{H}}} B_\theta(x_k, r_n) \qquad \text{and} \qquad \Theta_\mathcal{H} \subset \bigcup_{q=1}^{d_n^{\Theta_\mathcal{H}}} B_\theta(\theta_q, r_n)\,, \tag{3.1}$$

with x_k (resp. θ_q) $\in \mathcal{H}$ and r_n, $d_n^{\mathcal{S}_\mathcal{H}}$, $d_n^{\Theta_\mathcal{H}}$ are sequences of positive real numbers which tend to infinity as n goes to infinity and suppose that $d_n^{\mathcal{S}_\mathcal{H}}$, $d_n^{\Theta_\mathcal{H}}$ are the minimal numbers of open balls with radius r_n in $\mathcal{H}$, which are required to cover $\mathcal{S}_\mathcal{H}$ and $\Theta_\mathcal{H}$. Moreover, the following assumptions are also satisfied:

(**A1**) There exists a differentiable function $\phi(\cdot)$ such that, $\forall x \in \mathcal{S}_\mathcal{H}$ and $\forall \theta \in \Theta_\mathcal{H}$,

$$0 < C\phi(h) \le \phi_{\theta,x}(h) \le C'\phi(h) < \infty \qquad \text{and} \qquad \exists\,\eta_0 > 0,\ \forall \eta < \eta_0,\ \phi'(\eta) < C\,;$$

(**A2**) The kernel K satisfy (H4) and Lipschitz's condition holds

$$\Big|K(x)-K(y)\Big| \leq C\,\|x-y\|\,;$$

(**A3**) The conditional density $f(\theta,t,x)$ satisfies the uniform Hölder condition, i.e, $\forall(t_1,t_2)\in\mathcal{S}_{\mathbb{R}}\times\mathcal{S}_{\mathbb{R}}$, $\forall(x_1,x_2)\in\mathcal{S}_{\mathcal{H}}\times\mathcal{S}_{\mathcal{H}}$ and $\forall\theta\in\Theta_{\mathcal{H}}$,

$$\Big|f(\theta,t_1,x_1)-f(\theta,t_2,x_2)\Big| \leq C\Big(\|x_1-x_2\|^{b_1}+|t_1-t_2|^{b_2}\Big)\,;$$

(**A4**) For some $\nu\in(0,1)$, $\lim\limits_{n\to\infty} n^{\nu}h_H=\infty$, and for $r_n=\mathcal{O}\big(\frac{\log n}{n}\big)$, the sequences $d_n^{\mathcal{S}_{\mathcal{H}}}$ and $d_n^{\Theta_{\mathcal{H}}}$ satisfy:

$$\frac{(\log n)^2}{n\,h_H\,\phi(h_K)} < \log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}} < \frac{n\,h_H\,\phi(h_K)}{\log n}\,,$$

and

$$\sum_{n=1}^{\infty} n^{(3\gamma+1)/2}\big(d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\big)^{1-\beta}<\infty\,,\qquad \text{for some}\quad \beta>1\,.$$

In what follows, denote

$$\begin{aligned}
\Lambda_i(x,\theta) &= \frac{1}{h_K\,\phi(h_K)}\;\mathbf{1}_{B_\theta(x,h)\cup B_\theta(x_{k(x)},h)}(X_i)\,,\\
\Omega_i(x,\theta) &= \frac{1}{h_K\,\phi(h_K)}\;\mathbf{1}_{B_\theta(x_{k(x)},h)\cup B_{\theta_{q(\theta)}}(x_{k(x)},h)}(X_i)\,,\\
\Delta_i\big(x_{k(x)},\theta_{q(\theta)}\big) &= \frac{K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}{\mathbb{E}\,K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_i\big(x_{k(x)},v_{k_t},\theta_{q(\theta)}\big) &= \frac{1}{h_H}\,\frac{K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}{\mathbb{E}\,K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}\;H\big(h_H^{-1}(v_{k_t}-Y_i)\big)\\
&\quad-\frac{1}{h_H}\,\mathbb{E}\left(\frac{K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}{\mathbb{E}\,K\Big(h_K^{-1}\big\langle x_{k(x)}-X_i,\,\theta_{q(\theta)}\big\rangle\Big)}\;H\big(h_H^{-1}(v_{k_t}-Y_i)\big)\right).
\end{aligned}$$

Remark 3.1. Note that Assumptions (A1) and (A3) are, respectively, the uniform version of (H1) and (H2). Assumptions (A1) and (A4) are linked with the topological structure of the functional variable, see Ferraty *et al.* [12].

Theorem 3.1. *Under Assumptions (A1)–(A4), we have, as n goes to infinity*

(3.2)

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\widehat{f}(\theta,t,x)-f(\theta,t,x)\Big| = \mathcal{O}\big(h_K^{b_1}\big)+\mathcal{O}\big(h_H^{b_2}\big)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right).$$

Corollary 3.1. *Under the assumptions of Theorem 3.1 and hypotheses (H6)–(H7), we have*

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{M}_\theta(x)-M_\theta(x)\right|^j = \mathcal{O}\big(h_K^{b_1}\big)+\mathcal{O}\big(h_H^{b_2}\big)+\mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right). \tag{3.3}$$

The proof of Theorem 3.1 and Corollary 3.1 can be completed by the following lemmas.

Lemma 3.1. *Under assumptions (A1), (A3) and (A4), we have as $n\longrightarrow\infty$*

$$\sup_{\theta\in\Theta_{\mathcal{F}}}\sup_{x\in\mathcal{S}_{\mathcal{F}}}\left|\widehat{F}_D(\theta,x)-\mathbb{E}\,\widehat{F}_D(\theta,x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{n\,\phi(h_K)}}\right).$$

Corollary 3.2. *Under assumptions of Lemma 3.1, we have*

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\inf_{\theta\in\Theta_{\mathcal{F}}}\inf_{x\in\mathcal{S}_{\mathcal{F}}}\widehat{F}_D(\theta,x)<\frac{1}{2}\right)<\infty\,.$$

Lemma 3.2. *Under assumptions (A1), (A3) and (H3), we have as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|f(\theta,t,x)-\mathbb{E}\,\widehat{f}_N(\theta,t,x)\right| = \mathcal{O}\big(h_K^{b_1}\big)+O\big(h_H^{b_2}\big)\,. \tag{3.4}$$

Lemma 3.3. *Under the assumptions of Theorem 3.1, we have as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widetilde{f}_N(\theta,t,x)-\mathbb{E}\,\widetilde{f}_N(\theta,t,x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{n\,h_H\,\phi(h_K)}}\right).$$

4. ASYMPTOTIC NORMALITY

In this section, the asymptotic normality of the conditional density and the conditional mode are established. Therefore, further assumptions are required. Assume that:

(N1) There exists a function $\xi^{\theta,x}$, such that

$$\forall u\in[0,1]\,,\qquad \lim_{h\longrightarrow 0}\frac{\phi_{\theta,x}(uh)}{\phi_{\theta,x}(h)}=\lim_{h\longrightarrow 0}\xi_h^{\theta,x}(u)=\xi_0^{\theta,x}(u)\,;$$

(N2) The bandwidth h_H satisfies

$$n\,h_H^3\,\phi_{\theta,x}^3(h_K)\longrightarrow 0,\qquad \text{as}\quad n\longrightarrow\infty\,;$$

(N3) The *df* of the censored random variable, G has a bounded first derivative G';

(**N4**) The conditional density function $f(\theta,t,x)$ satisfies: $\exists\,\beta_0>0, \forall (t_1,t_2)\in \mathcal{S}_{\mathbb{R}}\times\mathcal{S}_{\mathbb{R}}$,

$$\left|f^{(q)}(\theta,t_1,x)-f^{(q)}(\theta,t_2,x)\right| \le C\left(|t_1-t_2|^{\beta_0}\right), \qquad \forall q=1,2\,;$$

(**N5**) $H^{'}$ and $H^{''}$ are bounded respectively with

$$\int \left(H^{'}(t)\right)^2 dt < \infty\,, \qquad \int |t|^{\beta_0} H(t)\,dt < \infty\,.$$

Theorem 4.1. *Under assumptions (H1)–(H5) and (N1)–(N3) for all $x\in\mathcal{H}$, we have*

$$\sqrt{\frac{n\,h_H\,\phi_{\theta,x}(h_K)}{\sigma^2(\theta,t,x)}}\left(\widehat{f}(\theta,t,x)-f(\theta,t,x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \qquad \text{as} \;\; n\longrightarrow\infty\,,$$

where

$$\sigma^2(\theta,t,x) \;=\; \frac{a_2(\theta,x)\,f(\theta,t,x)}{(a_1(\theta,x))^2\,\bar{G}(t)}\int_{\mathbb{R}} H^2(u)\,du\,,$$

with

$$a_l(\theta,x) \;=\; K^l(1)-\int_0^1 (K^l)^{'}(u)\;\xi_0^{\theta,x}(u)\;du\,, \qquad l=1,2\,.$$

"$\xrightarrow{\mathcal{D}}$" *means the convergence in distribution.*

Proof: In order to establish the asymptotic normality of $\widehat{f}(\theta,t,x)$, we need further notations and definitions. First we consider the following decomposition:

$$\begin{aligned}
\widehat{f}(\theta,t,x)-f(\theta,t,x) \;&=\; \frac{\widehat{f}_N(\theta,t,x)}{\widehat{F}_D(\theta,x)}-\frac{a_1(\theta,x)\,f(\theta,t,x)}{a_1(\theta,x)}\\
&=\; \frac{1}{\widehat{F}_D(\theta,x)}\left(\widehat{f}_N(\theta,t,x)-\mathbb{E}\,\widehat{f}_N(\theta,t,x)\right)\\
&\quad -\frac{1}{\widehat{F}_D(\theta,x)}\left(a_1(\theta,x)\,f(\theta,t,x)-\mathbb{E}\,\widehat{f}_N(\theta,t,x)\right)\\
&\quad +\frac{f(\theta,t,x)}{\widehat{F}_D(\theta,x)}\left(a_1(\theta,x)-\mathbb{E}\,\widehat{F}_D(\theta,x)\right)\\
&\quad -\frac{f(\theta,t,x)}{\widehat{F}_D(\theta,x)}\left(\widehat{F}_D(\theta,x)-\mathbb{E}\,\widehat{F}_D(\theta,x)\right)\\
&=\; \frac{1}{\widehat{F}_D(\theta,x)}\left(A_n(\theta,t,x)+B_n(\theta,t,x)\right).
\end{aligned}$$

□

Where:

$$\begin{aligned}
A_n(\theta,t,x) \;&=\; \frac{1}{n\,h_H\,\mathbb{E}\,K_1(\theta,x)}\sum_{i=1}^n\left\{\left(\frac{\delta_i}{\bar{G}(Y_i)}\,H_i(t)-h_H\,f(\theta,t,x)\right)K_i(\theta,x)\right.\\
&\qquad \left. -\;\mathbb{E}\left[\left(\frac{\delta_i}{\bar{G}(Y_i)}\,H_i(t)-h_H\,f(\theta,t,x)\right)K_i(\theta,x)\right]\right\}\\
&=\; \frac{1}{n\,h_H\,\mathbb{E}\,K_1(\theta,x)}\sum_{i=1}^n N_i(\theta,t,x)\,.
\end{aligned}$$

It follows that

$$\begin{aligned} n\, h_H\, \phi_{\theta,x}(h_K)\, \mathrm{Var}\big(A_n(\theta,t,x)\big) \;&=\; \frac{\phi_{\theta,x}(h_K)}{h_H\big(\mathbb{E}\, K_1(\theta,x)\big)^2}\, \mathrm{Var}\big(N_1(\theta,t,x)\big) \\ &=\; V_n(\theta,t,x) \end{aligned}$$

and

$$B_n(\theta,t,x) \;=\; a_1(\theta,x)\, f(\theta,t,x) - \mathbb{E}\, \widehat{f}_N(\theta,t,x) + f(\theta,t,x)\, \Big(a_1(\theta,x) - \mathbb{E}\, \widehat{F}_D(\theta,x)\Big)\,.$$

Then, the proof of Theorem 4.1 can be deduced from the following Lemmas.

Lemma 4.1. *Under conditions of Theorem 4.1, we have*

$$\sqrt{n\, h_H\, \phi_{\theta,x}(h_K)}\, A_n(\theta,t,x) \xrightarrow{\;\mathcal{D}\;} \mathcal{N}\big(0, \sigma^2(\theta,t,x)\big)\,,$$

where $\sigma^2_{\theta,x}$ is given in Theorem 4.1.

Lemma 4.2. *Under assumptions (H1)–(H5) and (N1)–(N2), we have as $n \longrightarrow \infty$,*

$$\sqrt{n\, h_H\, \phi_{\theta,x}(h_K)}\, B_n(\theta,t,x) \longrightarrow 0 \qquad \text{in probability}\,.$$

Corollary 4.1. *If the assumptions (H1)–(H7) as well as (N1)–(N5) hold, then, we have:*

$$\sqrt{\frac{n\, h_H^3\, \phi_{\theta,x}(h_K)}{\sigma_1^2(\theta,x)}}\, \Big(\widehat{M}_\theta(x) - M_\theta(x)\Big) \xrightarrow{\;\mathcal{D}\;} \mathcal{N}(0,1)\,, \qquad \text{as} \;\; n \longrightarrow \infty\,, \tag{4.1}$$

where

$$\sigma_1^2(\theta,t,x) \;=\; \frac{a_2(\theta,x)\, f\big(\theta, M_\theta(x), x\big)}{\Big(a_1(\theta,x)\, f^{(2)}\big(\theta, M_\theta(x), x\big)\Big)^2\, \bar{G}(t)} \int_{\mathbb{R}} H'^2(u)\, du\,.$$

5. SIMULATION STUDY

This section aims at illustrating our study which the forecast via the conditional mode. More precisely, we will compare our model CFSIM (1.4) (censored functional single index model) with CNPFDA (5.1) (censored nonparametric functional data analysis) in censored data:

$$\widehat{f}_n(t,x) \;=\; \frac{h_H^{-1} \sum_{i=1}^{n} \dfrac{\delta_i}{\bar{G}_n(Y_i)}\, K\big(h_K^{-1} d(x,X_i)\big)\, H\big(h_H^{-1}(t-Y_i)\big)}{\sum_{i=1}^{n} K\big(h_K^{-1}\, d(x,X_i)\big)}\,. \tag{5.1}$$

Note that all the routines for functional data used in this implementation (developed in R/S-Plus software) are available on the website https://www.math.univ-toulouse.fr/staph/npfda/.

We consider a diffusion process on the interval $[0,1]$:

$$X_i(t) = \cos(\pi b_i t) + a_i t^2 , \qquad i = 1, ..., 200; \quad t \in [0,1] , \tag{5.2}$$

where a_i are uniformly distributed on $[0,1]$ ($a \sim \mathcal{U}(0,1)$) and b_i are standard normal distribution ($b \sim \mathcal{N}(0,1)$). We carry out the simulation with a 200 sample of the curves X(t) (see Figure 1).

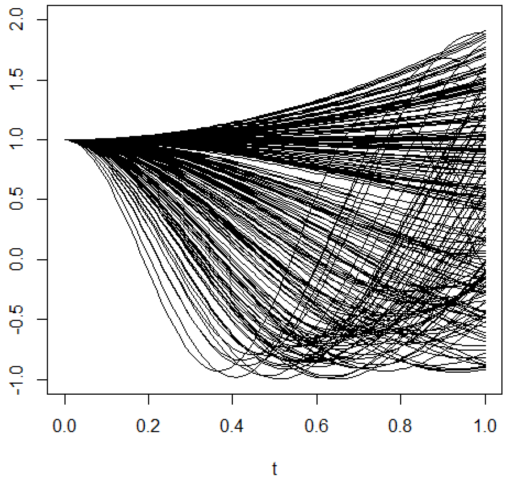


Figure 1: The curves $X_{i=1,...,200}(t)$, $t \in [0,1]$.

5.1. Estimating the single index in practice

The single index θ is unknown and has to be estimated. In practice this parameter can be selected by cross-validation approach (see Aït Saidi *et al.* [2]). To simulate the single functional index model as follows, first, we choose the functional parameter θ.

So for $\mathcal{L} = \{1, ..., 200\}$, the best approximation of θ is to estimate the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}(X'))\langle X', \cdot\rangle_{\mathcal{H}}]$ by its empirical covariance $\frac{1}{\mathcal{L}} \sum_{i \in \mathcal{L}} (X_i' - \mathbb{E}(X'))^t (X_i' - \mathbb{E}(X'))$ [4]. Figures 2, 3 and 4 show the discretization of the two first eigenfunction, twenty and all the eigenfunctions $\theta_i(t)$, respectively.

Taking θ^* the first eigenfunction corresponding to the first higher eigenvalue, and compute the inner product $\langle \theta^*, X_1\rangle, ..., \langle \theta^*, X_{200}\rangle$, then simulate the response variables $T_i = r(\langle \theta^*, X_i\rangle) + \epsilon$, where $r(\langle \theta^*, X_i\rangle) = \exp(10(\langle \theta^*, X_i\rangle - 0.05))$ and ϵ generate independently from a centered gaussian of variance equal to 0.05 times the empirical variance of $r(\langle \theta^*, X_i\rangle)$.

We simulate n i.i.d. rv C_i, $i = 1, ..., n$ with the exponential distribution $\mathcal{E}(1,5)$. Noting that the computation of those estimators are based on the observed data $(X_i, Y_i, \delta_i)_{i=1,...,n}$, where $Y_i = \min(T_i, C_i)$ and $\delta_i = \mathbf{1}_{\{T_i \le C_i\}}$. On the other hand, we choose the quadratic kernels defined by:

$$K(u) = \frac{3}{2} (1 - u^2) \, \mathbf{1}_{(0,1)}(u)$$

and

$$H(t) = \frac{3}{4} (1 - t^2) \, \mathbf{1}_{(-1,1)}(t) .$$

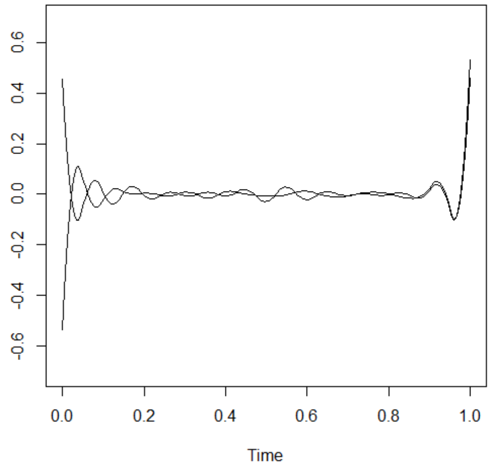


Figure 2: The curves $\theta_{i=1,2}(t)$, $t \in [0,1]$.

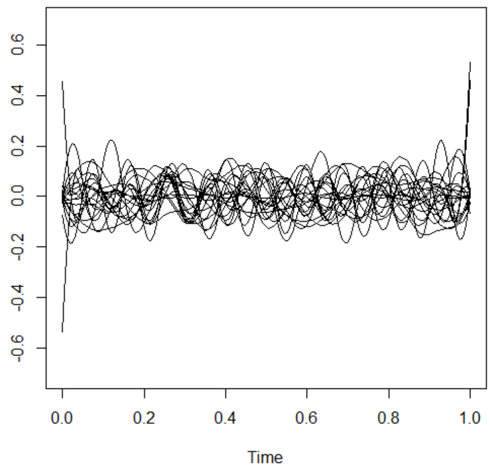


Figure 3: The curves $\theta_{i=1,\dots,20}(t)$, $t \in [0,1]$.

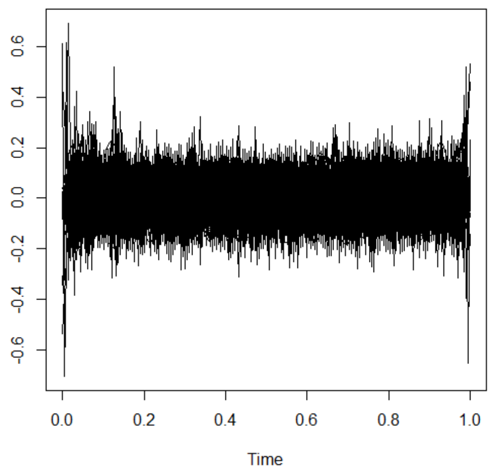


Figure 4: The curves $\theta_{i=1,\dots,200}(t)$, $t \in [0,1]$.

Then, taking into account the smoothness of the curves $X_i(t)$, we choose for the CNPFDA model the semi-metric in $\mathcal{H}$:

$$d(x_i, x_j) = \sqrt{\int_0^1 \Big(x_i'(t) - x_j'(t)\Big)^2 dt}\,, \qquad x_i, x_j \in \mathcal{H}\,.$$

For the bandwidths $h_H \sim h_K =: h$ is automatically selected by the procedure of the cross-validation method on the k-nearest neighbors ([16]).

In our simulation, sample sizes are $n = 200$, we take it into two parts, one is a learning sample of 150 observations and the others 50 observations are a test sample. Then using the learning sample to compute the estimator of $\widehat{Y}_i = \widehat{M}_{\theta^*}(X_i)$ and $\widehat{Y}_{ni} = \widehat{M}(X_i)$ for $i = \{151, ..., 200\}$.

Finally we show the results by plotting the true values versus the predicted values for the MSE under censored data for both estimators (1.4) and (5.1) which are respectively defined as:

$$\text{CFSIM.MSE} = \frac{1}{50}\sum_{i=151}^{200} (Y_i - \widehat{Y}_i)^2\,,$$

$$\text{CNPFDA.MSE} = \frac{1}{50}\sum_{i=151}^{200} (Y_i - \widehat{Y}_{ni})^2\,.$$

By Figures 5 and 6, we can say that both estimators on weak censored rates of 3.5% works almost as well as if we had the complete data-set. To show how the different censored rates (CRs) affects the prediction results, we present some CRs and their corresponding MSE for CFSIM and CNPFDA. Two sample sizes are considered, $n = 200$ and 300, and for each sample size different censoring rates are taken: CR = 5%, 16%, 27% and 54%. We carried 100 independent replications of the experiment and then we computed the average of mean squared error. These quantities are presented in Table 1.

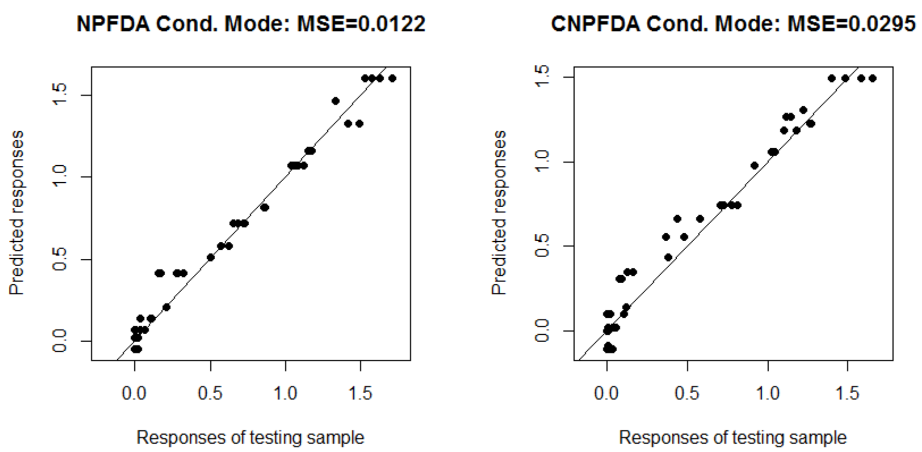


Figure 5: Prediction via the conditional mode by NPFDA for complete data (MSE = 0.0122) and censored data with CR ∼ 3.5% (MSE = 0.0295).

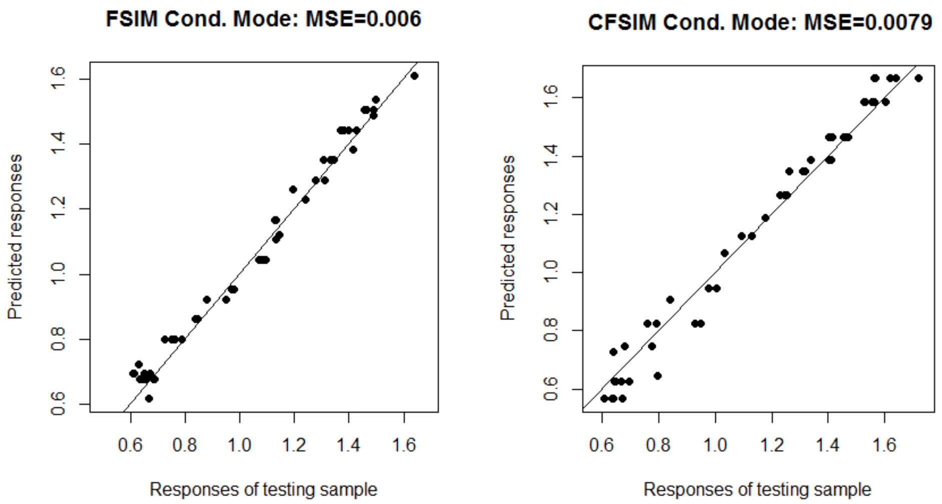


Figure 6: Prediction via the conditional mode by FSIM for complete data (MSE = 0.006) and censored data with CR ∼ 3.5% (MSE = 0.0079).

Table 1: MSE comparison for FSIM and NPFDA.

n	CR	NPFDA	CNPFDA	FSIM	CFSIM
200	5%	0.0122	0.0343	0.0065	0.0168
	16%		0.0844		0.0554
	27%		0.1274		0.1084
	54%		0.3349		0.3234
300	5%	0.0115	0.0303	0.0032	0.0063
	16%		0.0779		0.0473
	27%		0.1245		0.0952
	54%		0.3165		0.1600

One can observe that both estimators have a reasonable performance for lower censored rates. However, they are strongly affected when the percentage of censored rate is high, but the FSIM estimator stays more accurate than the NPFDA one in all cases. And on the other hand, when the sample size increases, the preciseness of forecast also increases.

5.2. Real data example: peak electricity demand

We evaluate and compare the finite sample performance between a nonparametric functional model with our estimator (the functional single index model). To this end, we apply our method to the data constituting hourly electricity demand for the Rocky Mountain region (WACM) of the United States. The data are daily electricity demands divided into 24 grids, where each hour of the day corresponds to a grid, from July 2015 to November 2018. The updated version of the data can be found on the site http://www.eia.gov/.

We construct our variables as follows: the observations of our covariate X are the daily electricity demands from 2016 to 2018, $X_i = (x_{i1}, ..., x_{i24})$, our sample consists of $n = 1037$ observations. The observations of our response variable Y are $Y_i = \min(\max(X_i), 1408)$, $i = 1, ..., n$, where 1408 is the maximum peak of electricity demands in 2015.

In this part, we use Kaplan–Meier's estimator $\overline{G}_n(\cdot)$ as an estimator of $\overline{G}(\cdot)$ to construct our conditional distribution estimator, by taking the variables $(C_i)_i$ as deterministic (all equal to 1408, which is the maximum of the peak observed in 2015).

Since we are performing analysis on a time series spread over 4 years, considering the year 2015 as a base year, and in the simulation we are interested only in the years 2016–2018, we can consider 1408 as a maximum amplitude, that is, any value (or hourly observation) greater than 1408 can be considered as aberrant data. So, on this basis, we built our response variable.

Concerning the estimation of our parameters, we chose $deriv_1$ (the semi-metric based on the first derivatives of the curves) as semi-metric, the kernel $K(\cdot)$ and the cumulative *df* $H(u)$ are defined in the Subsection 5.1. Then, as previously discussed, the optimal bandwidth $h_n = h_{K,n} = h_{H,n}$ are chosen using the cross-validation method on the k-nearest neighbors. Finally, we replace θ by the first eigenfunction corresponding to the first higher eigenvalue of the empirical covariance operator. The curves of the data are represented in Figure 7.

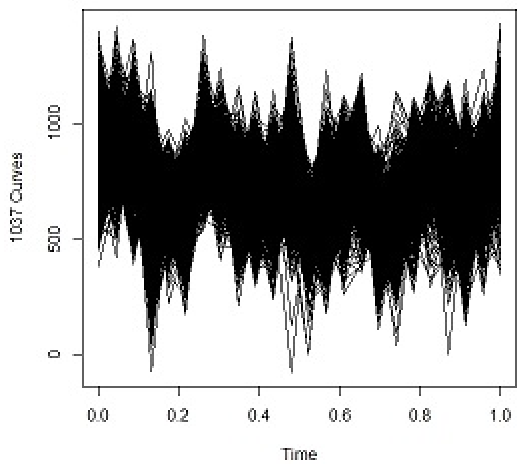


Figure 7: Sample curves $\{X_i(t),\, t \in [0,1]\}_{i=1,...,1037}$.

To assess the in-sample estimation accuracy and out-of sample prediction accuracy of the models, we split the original 1037 samples into two samples. The first one (learning set), from 1 to 960, used for the estimation, while the second sample (testing set), from 961 to 1037, is served for the prediction. To measure the estimation and prediction accuracies, we evaluate and compare the forecast accuracy using the testing sample, from which we predict responses in the testing sample. To measure the performance of each functional prediction method, we consider the mean square errors (MSE).

After performing the calculations, we find MSE = 0.0025 for our estimator, and an MSE = 0.0197 for that of NPFDA (see Figure 8). We can therefore conclude that there is an improvement in estimation and prediction accuracies for our model in comparison to the nonparametric functional model.

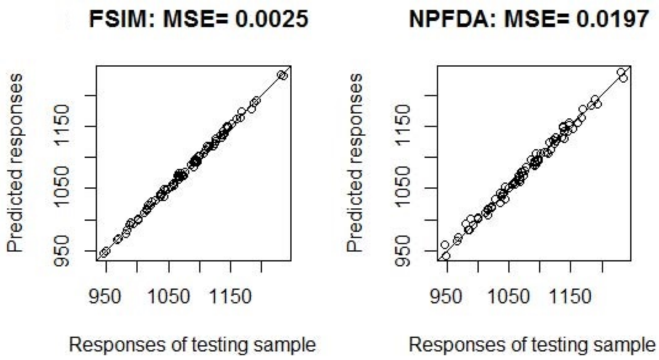


Figure 8: Prediction via the conditional mode by FSIM with error MSE = 0.0025 against NPFDA with error MSE = 0.0197.

6. PROOFS OF TECHNICAL LEMMAS

Proof of Lemma 2.1: The proof is similar to that of Lemma 5.2 in [22]. From Equations (1.2) and (1.4), we have

$$\begin{aligned}\left|\widehat{f}_N(\theta,t,x)-\widetilde{f}_N(\theta,t,x)\right| &\le \frac{h_H^{-1}}{n\mathbb{E}K_1(\theta,x)}\sum_{i=1}^n\left|\frac{\delta_i}{\bar{G}_n(Y_i)}K_i(\theta,x)H_i(t)-\frac{\delta_i}{\bar{G}(Y_i)}K_i(\theta,x)H_i(t)\right|\\ &\le \frac{h_H^{-1}}{n\mathbb{E}K_1(\theta,x)}\sum_{i=1}^n\left|\delta_iK_i(\theta,x)H_i(t)\right|\left|\frac{1}{\bar{G}_n(Y_i)}-\frac{1}{\bar{G}(Y_i)}\right|\\ &\le \frac{h_H^{-1}}{\phi_{\theta,x}(h_K)}\,\frac{C}{\bar{G}_n(\tau_G)\bar{G}(\tau_G)}\,\sup_{t\le\tau_G}{}_{\mathbb{R}}\left|\bar{G}_n(t)-\bar{G}(t)\right|\,\frac{1}{n}\sum_{i=1}^n\left|K_i(\theta,x)H_i(t)\right|.\end{aligned}$$

Since $\bar{G}(t_G)>0$, together with the SLLN and the LIL on the censoring law (see formula (4.28) in Deheuvels and Einmahl [9]), we obtain

$$\sup_{t\le\tau_G}\left|\bar{G}_n(t)-\bar{G}(t)\right| = O_{a.s.}\left(\frac{\log\log n}{n}\right).$$

We achieve the proof by considering the conditions (H3) and (H4). □

Proof of Lemma 2.2: We have

$$\begin{aligned}\mathbb{E}\widetilde{f}_N(\theta,t,x)-f(\theta,t,x) &= \frac{1}{h_H\,\mathbb{E}K_1(x,\theta)}\,\mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)}K_i(x,\theta)H_i(t)\right)-f(\theta,t,x)\\ &= \frac{1}{h_h\,\mathbb{E}K_1(x,\theta)}\,\mathbb{E}\left(K_i(x,\theta)\left[\mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)}H_i(t)\,|\,\langle X_1,\theta\rangle\right)-h_H\,f(\theta,t,x)\right]\right).\end{aligned} \tag{6.1}$$

Using the fact that H is a *cdf* and the use a double conditioning with respect to T_1, we can easily get

$$
\begin{aligned}
I &= \mathbb{E}\Big(\frac{\delta_i}{\bar{G}(Y_i)}H_i(t)\,|\,\langle X_1,\theta\rangle\Big)\\
&= \mathbb{E}\Big(\mathbb{E}\Big[\frac{\mathbf{1}_{T_1\le C_1}}{\bar{G}(T_1)}H\Big(\frac{t-T_1}{h_H}\Big)\,|\,\langle X_1,\theta\rangle, T_1\Big]\Big)\\
&= \mathbb{E}\Big(\frac{1}{\bar{G}(T_1)}H\Big(\frac{t-T_1}{h_H}\Big)\mathbb{E}\Big[\mathbf{1}_{T_1\le C_1}\,|\,T_1\Big]\,|\,\langle X_1,\theta\rangle\Big)\\
&= \mathbb{E}\Big[H\Big(\frac{t-T_1}{h_H}\Big)\,|\,\langle X_1,\theta\rangle\Big]\\
&= \int_{\mathbb{R}} H\Big(\frac{t-u}{h_H}\Big)\, f(\theta,u,X_1)\,du\\
&= h_H\int_{\mathbb{R}} H(v)\, f(\theta,t-vh_H,X_1)\,dv\\
&= h_H\int_{\mathbb{R}} H(v)\,\Big(f(\theta,t-vh_H,X_1)-f(\theta,t,x)\Big)\,dv \,+\, h_H\, f(\theta,t,x)\int_{\mathbb{R}} H(v)\,dv\,,
\end{aligned}
$$

we can write, because of (H2) and (H3):

$$
\begin{aligned}
I &\le h_H\, C_{x,\theta}\int_{\mathbb{R}} H(v)\Big(h_K^{b_1}+|v|^{b_2}h_H^{b_2}\Big)\,dv \,+\, h_H\, f(\theta,t,x)\\
&= \mathcal{O}\big(h_K^{b_1}+h_H^{b_2}\big)+h_H\, f(\theta,t,x)\,.
\end{aligned}
$$

Combining this last result with (6.1) allows us to achieve the proof. $\square$

Proof of Lemma 2.4: Using the compactness of $\mathcal{S}_{\mathbb{R}}$, we can write that:

$\mathcal{S}_{\mathbb{R}}\subset\bigcup_{k=1}^{\tau_n}(z_k-l_n,z_k+l_n)$, where l_n and τ_n can be chosen such that $l_n=C\tau_n^{-1}\sim Cn^{-\varsigma-1/2}$.

Taking $k_t = \underset{\{z_1,\ldots,z_{\tau_n}\}}{\arg\min}\,|t-z_k|$.

Thus, we have the following decomposition:

$$
\begin{aligned}
\frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\widetilde{f}_N(\theta,t,x)-\mathbb{E}\widetilde{f}_N(\theta,t,x)\Big| &\le \frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\widetilde{f}_N(\theta,t,x)-\widehat{f}_N(\theta,t_k,x)\Big|\\
&\quad+\frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widehat{f}_N(\theta,t_k,x)\Big|\\
&\quad+\frac{1}{\widehat{F}_D(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\mathbb{E}\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widetilde{f}_N(\theta,t,x)\Big|\\
&\le T_1+T_2+T_3\,.
\end{aligned}
$$

• On one hand, as the first and the third terms can be treated in the same manner, we deal only with first term. Making use of (H3) we get

$$\begin{aligned}
\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widetilde{f}_N(\theta,t,x)-\widehat{f}_N(\theta,t_k,x)\right| &\leq \\
&\leq \frac{1}{n\,h_H\,\mathbb{E}K_1(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\sum_{i=1}^{n}\left|\frac{\delta_i}{\bar{G}(Y_i)}H_i(t)-\frac{\delta_i}{\bar{G}_n(Y_i)}H_i(t_k)\right||K_i(\theta,x)| \\
&\leq \frac{C}{n\,h_H\,\mathbb{E}K_1(\theta,x)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\frac{|t-t_k|}{h_H}\left(\sum_{i=1}^{n}K_i(\theta,x)\left(\frac{1}{\bar{G}(Y_i)}-\frac{1}{\bar{G}_n(Y_i)}\right)\right) \\
&\leq \frac{Cl_n}{h_H^2\bar{G}_n(\tau_G)\bar{G}(\tau_G)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|G_n(t)-G(t)|\,\widehat{F}_D(\theta,x)\,.
\end{aligned}$$

Using $l_n = n^{-\varsigma-1/2}$ we obtain

$$T_1 \leq \frac{Cn^{-\varsigma-1/2}}{h_H^2\bar{G}_n(\tau_G)\bar{G}(\tau_G)}\left(\frac{\log n\log n}{n}\right)^{1/2},$$

and note that, because of (H2)-(i), we have

$$\frac{l_n}{h_H^2} = o\left(\sqrt{\frac{\log n}{n\,h_h\,\phi_{\theta,x}(h_K)}}\right).$$

Thus, for n large enough, we have

$$T_1 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,x}(h_K)}}\right).$$

Following similar arguments, we can write

$$T_3 \leq T_1\,.$$

• Concerning T_2, let us consider $\varepsilon = \epsilon_0\sqrt{\dfrac{\log n}{n\,h_H\,\phi_{\theta,x}(h_K)}}$. Since for all $\epsilon_0 > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widehat{f}_N(\theta,t_k,x)\right|>\varepsilon\right) &\leq \mathbb{P}\left(\max_{k\in\{1,\ldots,\tau_n\}}\left|\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widehat{f}_N(\theta,t_k,x)\right|>\varepsilon\right) \\
&\leq \tau_n\max_{k\in\{1,\ldots,\tau_n\}}\mathbb{P}\left(\left|\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widehat{f}_N(\theta,t_k,x)\right|>\varepsilon\right).
\end{aligned}$$

Applying Berstain's exponential inequality to:

$$\Psi_i = \frac{1}{h_H\,\mathbb{E}K_1(x,\theta)}\left[\frac{\delta_i}{\bar{G}(Y_i)}K_i(x,\theta)H_i(t_k)-\mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)}K_i(x,\theta)H_i(t_k)\right)\right].$$

Firstly, it follows from the fact that the Kernels K and H are bounded, we get

$$\mathbb{P}\left(\left|\widehat{f}_N(\theta,t_k,x)-\mathbb{E}\widehat{f}_N(\theta,t_k,x)\right|>\varepsilon\right) \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Psi_i\right|>\varepsilon\right) \leq 2\,n^{-c\varepsilon_0^2}\,.$$

Finally, by choosing ε_0 large enough, the proof can be concluded by the use of the Borel–Cantelli lemma. the result can be easily deduced. □

Proof of Lemma 3.3: For all $x \in \mathcal{S}_{\mathcal{H}}$ and $\forall \theta \in \Theta_{\mathcal{H}}$, we set

$$k(x) = \underset{k\in\{1,\dots,d_n^{\mathcal{S}_{\mathcal{H}}}\}}{\arg\min} \|x - x_k\| \qquad \text{and} \qquad q(\theta) = \underset{m\in\{1,\dots,d_n^{\Theta_{\mathcal{H}}}\}}{\arg\min} \|\theta - \theta_q\|,$$

and by the compact property of $\mathcal{S}_{\mathbb{R}} \subset \mathbb{R}$ we have $\mathcal{S}_{\mathbb{R}} \subset \bigcup_{k=1}^{\tau_n}(v_k - l_n, v_k + l_n)$ where l_n and τ_n can be selected such as $l_n = \mathcal{O}\big(\tau_n^{-1}\big) = \mathcal{O}\big(n^{-(3\varsigma+1)/2}\big)$. Taking $k_t = \underset{\{v_1,\dots,v_{\tau_n}\}}{\arg\min} |t - v_k|$.

Let us consider the following decomposition:

$$\begin{aligned}
\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} \Big|\widetilde{f}_N(\theta,t,x) - \mathbb{E}\Big(\widetilde{f}_N(\theta,t,x)\Big)\Big| &\leq \\
\leq \sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} \Big\{ &\Big|\widetilde{f}_N(\theta,t,x) - \widetilde{f}_N(\theta,t,x_{k(x)})\Big| \\
&+ \Big|\widetilde{f}_N(\theta,t,x_{k(x)}) - \widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)})\Big| \\
&+ \Big|\widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)}) - \widetilde{f}_N(\theta_{q(\theta)},v_{k_t},x_{k(x)})\Big| \\
&+ \Big|\widetilde{f}_N(\theta_{q(\theta)},v_{k_t},x_{k(x)}) - \mathbb{E}\Big(\widetilde{f}_N(\theta_{q(\theta)},v_{k_t},x_{k(x)})\Big)\Big| \\
&+ \Big|\mathbb{E}\Big(\widetilde{f}_N(\theta_{q(\theta)},v_{k_t},x_{k(x)})\Big) - \mathbb{E}\Big(\widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)})\Big)\Big| \\
&+ \Big|\mathbb{E}\Big(\widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)})\Big) - \mathbb{E}\Big(\widehat{f}_N(\theta,t,x_{k(x)})\Big)\Big| \\
&+ \Big|\mathbb{E}\Big(\widetilde{f}_N(\theta,t,x_{k(x)})\Big) - \mathbb{E}\Big(\widetilde{f}_N(\theta,t,x)\Big)\Big|\Big\} \\
\leq \Psi_1 + \Psi_2 + \Psi_3 + \Psi_4 + \Psi_5 + \Psi_6 + \Psi_7 .
\end{aligned}$$

- Concerning Ψ_3 and Ψ_5; by conditions (H3) and (A4), boundness of K, we obtain

$$\begin{aligned}
&\Big|\widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)}) - \widetilde{f}_N(\theta_{q(\theta)},v_{k_t},x_{k(x)})\Big| \leq \\
&\leq \frac{1}{nh_H\,\mathbb{E}K_1(\theta,x)} \sup_{t\in\mathcal{S}_{\mathbb{R}}} \sum_{i=1}^{n} \left|\frac{\delta_i}{\bar{G}(Y_i)} K_i(\theta_{q(\theta)},x_{k(x)})\right| \left|H\left(\frac{t-Y_i}{h_H}\right) H\left(\frac{v_{k_t}-Y_i}{h_H}\right)\right| \\
&\leq \sup_{t\in\mathcal{S}_{\mathbb{R}}} C\frac{|t-v_{k_t}|}{h_H^2}\left(\frac{1}{n\mathbb{E}(K_1(\theta_{q(\theta)},x_{k(x)}))}\sum_{i=1}^{n}\left|K_i(\theta_{q(\theta)},x_{k(x)})\frac{1}{\bar{G}(Y_i)}\right|\right) \\
&\leq \frac{Cl_n}{\phi(h_K)h_H^2} = \mathcal{O}\left(\frac{l_n}{h_H^2\phi(h_K)}\right).
\end{aligned}$$

Now, the fact that $\lim_{n\to\infty} n^{\nu}h_H = \infty$, and choosing $l_n = n^{-(3\nu+1)/2}$ and using the second part of (A4), imply that

$$\frac{l_n}{h_H^2\phi(h_K)} = o\left(\sqrt{\frac{logn}{nh_H\phi(h_K)}}\right)$$

as $n \to \infty$, therefore, it follows

$$\Psi_5 \leq \Psi_3 = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right).$$

- Concerning Ψ_4, let us consider $\varepsilon = \varepsilon_0 \sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{n h_H \phi(h_K)}}$. For all $\varepsilon_0 > 0$, we have

$$\begin{aligned}
\mathbb{P}(\Psi_4 > \varepsilon) &= \mathbb{P}\left(\max_{q\in\{1,\ldots,d_n^{\Theta_{\mathcal{H}}}\}} \max_{k\in\{1,\ldots,d_n^{\mathcal{S}_{\mathcal{H}}}\}} \max_{k_t\in\{1,2,\ldots,\tau_n\}} |\Gamma_i - \mathbb{E}\Gamma_i| > \varepsilon \right) \\
&\leq \tau_n\, d_n^{\mathcal{S}_{\mathcal{H}}}\, d_n^{\Theta_{\mathcal{H}}}\, \mathbb{P}\Big(|\Gamma_i - \mathbb{E}\Gamma_i| > \varepsilon \Big). \qquad (6.2)
\end{aligned}$$

Applying Bernstein's exponential inequality, under (H4), to get $\forall q \leq d_n^{\Theta_{\mathcal{H}}}$, $\forall k \leq d_n^{\mathcal{S}_{\mathcal{H}}}$ and $\forall k_t \leq \tau_n$,

$$\mathbb{P}\Big(|\Gamma_i - \mathbb{E}\Gamma_i| > \varepsilon \Big) \leq 2(d_n^{\Theta_{\mathcal{H}}} d_n^{\mathcal{S}_{\mathcal{H}}})^{-C\varepsilon_0^2}.$$

Choosing $\tau_n \leq C n^{(3\varsigma+1)/2}$, we get

$$\mathbb{P}\big(\Psi_4 > \varepsilon\big) \leq C'\, \tau_n (d_n^{\Theta_{\mathcal{H}}} d_n^{\mathcal{S}_{\mathcal{H}}})^{1-C\varepsilon_0^2}.$$

Putting $C\varepsilon_0^2 = \beta$ and using (A4), to get

$$\Psi_4 = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{n\, h_H\, \phi(h_K)}} \right). \qquad (6.3)$$

- Concerning Ψ_1 and Ψ_2, by assumption (A1), it follows

$$\begin{aligned}
\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} \left| \widetilde{f}_N(\theta,t,x) - \widetilde{f}_N(\theta,t,x_{k(x)}) \right| &\leq \\
&\leq \frac{1}{n h_H \mathbb{E} K_1(\theta,x)} \\
&\quad \cdot \sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\bar{G}(Y_i)} \right| |H_i(t)|\, |(K_i(\theta,x) - K_i(\theta,x_k))| \\
&\leq \frac{1}{n h_H \phi(h_K)} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{\theta\in\Theta_{\mathcal{H}}} \sum_{i=1}^n \left| \Delta_i(x,\theta) - \Delta_i\big(x_{k(x)},\theta\big) \right| \\
&\leq \frac{1}{h_H \phi(h_K)} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{\theta\in\Theta_{\mathcal{H}}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{B_\theta(x,h)\cup B_\theta(x_{k(x)},h)}(X_i) \\
&\leq \frac{C}{h_H} \sup_{x\in\mathcal{S}_{\mathcal{H}}} \sup_{\theta\in\Theta_{\mathcal{H}}} \frac{1}{n} \sum_{i=1}^n \Lambda_i(x,\theta).
\end{aligned}$$

Therefore, similar to the arguments for (6.3), we can get that

$$\Psi_1 = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{n h_H \phi(h_K)}} \right),$$

$$\begin{aligned}
&\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\Big|\widetilde{f}_N(\theta,t,x)-\widetilde{f}_N(\theta_{q(\theta)},t,x_{k(x)})\Big| \leq \\
&\qquad\leq \frac{h_H^{-1}}{n\mathbb{E}K_1(\theta,x)} \\
&\qquad\quad\cdot \sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}\sum_{i=1}^{n}\left|\frac{\delta_i}{\bar{G}(Y_i)}\right| |H_i(t)|\,\big|(K_i(\theta,x_{(k)})-K_i(\theta_{q(\theta)},x_{(k)}))\big| \\
&\qquad\leq \frac{Ch_H^{-1}}{n\phi(h_K)}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{\theta\in\Theta_{\mathcal{H}}}\sum_{i=1}^{n}\big|\Delta_i(\theta,x_k)-\Delta_i\big(\theta_{q(\theta)},x_{k(x)}\big)\big| \\
&\qquad\leq \frac{Ch_H^{-1}}{\phi(h_K)}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{\theta\in\Theta_{\mathcal{H}}}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{B_\theta(x_{(k)},h)\cup B_{q(\theta)}(x_{k(x)},h)}(X_i) \\
&\qquad\leq \frac{C}{h_H}\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{\theta\in\Theta_{\mathcal{H}}}\frac{1}{n}\sum_{i=1}^{n}\Omega_i(x,\theta)\,.
\end{aligned}$$

Similar to the deduce of (6.3), it yields

$$\Psi_2 \;=\; \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi(h_K)}}\right).$$

On the other hand, since $\Psi_7\leq\Psi_1$ and $\Psi_6\leq\Psi_2$, it also leads to

$$\Psi_6 \;=\; \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi(h_K)}}\right)$$

and

$$\Psi_7 \;=\; \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}}+\log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi(h_K)}}\right).$$

Then the proof of Lemma 3.3 can be completed. □

Proof of Lemma 4.1:

$$\begin{aligned}
V_n(\theta,t,x) &= \frac{\phi_{\theta,x}(h_K)}{h_H(\mathbb{E}K_1(\theta,x))^2}\,\mathbb{E}\left[K_1^2(\theta,x)\left(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)-h_hf(\theta,t,x)\right)^2\right] \\
&= \frac{\phi_{\theta,x}(h_K)}{h_H(\mathbb{E}K_1(\theta,x))^2}\,\mathbb{E}\left[K_1^2(\theta,x)\;\mathbb{E}\left(\left(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)-h_Hf(\theta,t,x)\right)^2\Big/\langle\theta,X_1\rangle\right)\right]. \qquad (6.4)
\end{aligned}$$

Using the definition of conditional variance, we have

$$\mathbb{E}\left[\left(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)-h_Hf(\theta,t,x)\right)^2\,\Big|\,\langle\theta,X_1\rangle\right] \;=\; J_{1n}+J_{2n}\,,$$

where

$$\begin{aligned}
J_{1n} &= \operatorname{Var}\left(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)\,\Big|\,\langle\theta,X_1\rangle\right), \\
J_{2n} &= \left[\mathbb{E}\left(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)\,\Big|\,\langle\theta,X_1\rangle\right)-h_Hf(\theta,t,x)\right]^2.
\end{aligned}$$

- Concerning J_{1n}:

$$J_{1n} \;=\; \mathbb{E}\bigg(\frac{\delta_1}{\bar{G}^2(Y_1)}H_1^2(t)\,\big|\,\langle\theta,X_1\rangle\bigg) - \mathbb{E}\bigg(\frac{\delta_1}{\bar{G}(Y_1)}H_1(t)\,\big|\,\langle\theta,X_1\rangle\bigg)^2 \;=\; J_1 + J_2\,.$$

As for J_1, by the property of double conditional expectation and by changing variables, we get that

$$\begin{aligned}
J_1 \;&=\; \mathbb{E}\bigg[\mathbb{E}\bigg(\frac{\delta_1}{\bar{G}^2(Y_1)}H_1^2\Big(\frac{t-Y_1}{h_H}\Big)\,|\,\langle\theta,X_1\rangle,T_1\bigg)\bigg]\\
&=\; \mathbb{E}\bigg(\frac{1}{\bar{G}^2(T_1)}H_1^2\Big(\frac{t-T_1}{h_H}\Big)\,\mathbb{E}\Big[\mathbf{1}_{T_1\le C_1}\,\big|\,T_1\Big]\,|\,\langle\theta,X_1\rangle\bigg)\\
&=\; \mathbb{E}\bigg(\frac{1}{\bar{G}(T_1)}H_1^2\Big(\frac{t-T_1}{h_H}\Big)\,|\,\langle\theta,X_1\rangle\bigg)\\
&=\; \int_{\mathbb{R}}\frac{1}{\bar{G}(v)}H_1^2\Big(\frac{t-v}{h_H}\Big)f(\theta,v,X_1)\,dv\\
&=\; \int_{\mathbb{R}}\frac{1}{\bar{G}(t-uh_H)}H_1^2(u)\,dF(\theta,t-uh_H,X_1)\,.
\end{aligned}$$

By the first order Taylor's expansion of the function $\bar{G}^{-1}(\cdot)$ around zero, one gets

$$J_1 = \int_{\mathbb{R}}\frac{1}{\bar{G}(t)}H_1^2(u)\,dF(\theta,t-uh_H,X_1) + \frac{h_H^2}{\bar{G}(t)^2}\int_{\mathbb{R}}uH_1^2\bar{G}^{(1)}(t^*)\,f(\theta,t-uh_H,X_1)\,du \;+\; o(1)\,,$$

where t^* is between t and $t-uh_H$.

Under assumptions (N3) and using hypothesis (H2), we get

$$\frac{h_H^2}{\bar{G}^2(t)}\int_{\mathbb{R}}uH_1^2\bar{G}^{(1)}(t^*)\,f(\theta,t-uh_H,X_1)\,du \;=\; o(h_H^2)\,.$$

Indeed,

$$\frac{h_H^2}{\bar{G}^2(t)}\int_{\mathbb{R}}uH_1^2\bar{G}^{(1)}(t^*)\,f(\theta,t-uh_H,X_1)\,du \;\le\; h_H^2\Big(\sup_{u\in\mathbb{R}}|G^{'}(u)||\bar{G}^2(t)\Big)\int_{\mathbb{R}}uf(\theta,t-yh_H,x)\,du\,.$$

On the other hand, by applying (H2) and (H3), we have

$$\begin{aligned}
&\int_{\mathbb{R}}\frac{1}{\bar{G}(t)}H_1^2(u)\,dF(\theta,t-uh_H,X_1) \;=\\
&\qquad=\; h_H\int_{\mathbb{R}}\frac{1}{\bar{G}(t)}H_1^2(u)\,f(\theta,t-uh_H,X_1)\,du\\
&\qquad\le\; \frac{h_H}{\bar{G}(t)}\int_{\mathbb{R}}H_1^2(u)\,\Big(f(\theta,t-uh_H,X_1)-f(\theta,t,x)\Big)\,du\\
&\qquad\quad+\frac{h_H}{\bar{G}(t)}\int_{\mathbb{R}}H_1^2(u)\,f(\theta,t,x)\,du\\
&\qquad\le\; \frac{h_H}{\bar{G}(t)}\bigg(C_{x,\theta}\int_{\mathbb{R}}H^2(u)\Big(h_K^{b_1}+|v|^{b_2}h_H^{b_2}\Big)\,du \;+\; f(\theta,t,x)\int_{\mathbb{R}}H^2(u)\,du\bigg)\\
&\qquad=\; \mathcal{O}\Big(h_k^{b_1}+h_H^{b_2}\Big) + \frac{h_H}{\bar{G}(t)}\,f(\theta,t,x)\int_{\mathbb{R}}H^2(u)\,du\,.
\end{aligned}\tag{6.5}$$

As for J_2,

$$\begin{aligned}
J_2' &= \mathbb{E}\bigg(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) \,\Big|\, \langle \theta, X_1\rangle\bigg)\\
&= \mathbb{E}\bigg[\mathbb{E}\bigg(\frac{\delta_1}{\bar{G}(Y_1)} H_1\Big(\frac{t-Y_1}{h_H}\Big) \,\Big|\, \langle \theta, X_1\rangle, T_1\bigg)\bigg]\\
&= \mathbb{E}\bigg(\frac{1}{\bar{G}(T_1)} H_1\Big(\frac{t-T_1}{h_H}\Big)\,\mathbb{E}\Big[\mathbf{1}_{T_1\le C_1} \,\Big|\, T_1\Big] \,\Big|\, \langle \theta, X_1\rangle\bigg)\\
&= \mathbb{E}\bigg(H_1\Big(\frac{t-T_1}{h_H}\Big) \,\Big|\, \langle \theta, X_1\rangle\bigg)\\
&= \int_{\mathbb{R}} H^{(1)}\Big(\frac{t-v}{h_H}\Big) f(\theta, t, X_1)\, dv\,.
\end{aligned}$$

Moreover, we have by changing variables:

$$J_2' = h_H \int_{\mathbb{R}} H(u)\Big(f(\theta, t-uh_H, X_1 - f(\theta,t,x)\Big)\, du \,+\, h_H f(\theta,t,x)\int_{\mathbb{R}} H(u)\, du\,.$$

The last equality is due to the fact that H is a probability density.

Thus, we have:

$$J_2' = \mathcal{O}\big(h_k^{b_1} + h_H^{b_2}\big) + h_H f(\theta,t,x)\,.$$

Finally we get $J_2 \underset{n\to\infty}{\longrightarrow} 0$.

As for J_{2n}, by (H1)–(H3), we obtain that

$$J_{2n} \longrightarrow 0, \qquad \text{as} \quad n \longrightarrow \infty\,. \tag{6.6}$$

Meanwhile, by (H1)–(H3) and (N3), it follows that

$$\frac{\phi_{\theta,x}(h_K)\,\mathbb{E}K_1^2(\theta,x)}{\mathbb{E}^2 K_1(\theta,x)} \underset{n\longrightarrow\infty}{\longrightarrow} \frac{a_2(\theta,x)}{(a_1(\theta,x))^2}\,,$$

which, combining equations (6.4) and (6.5), leads to

$$V_n(\theta,t,x) \longrightarrow \frac{a_2(\theta,x)}{(a_1(\theta,x))^2}\,\frac{f(\theta,t,x)}{\bar{G}(t)}\int_{\mathbb{R}} H^2(u)\, du\,. \tag{6.7}$$

□

Proof of Lemma 4.2: We have

$$\begin{aligned}
\sqrt{nh_H\phi_{\theta,x}(h_K)}\, B_n(\theta,t,x) \;=\; \frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}}{\widehat{F}_D(\theta,x)}\,\Big\{&\mathbb{E}\widehat{f}_N(\theta,t,x) - a_1(\theta,x) f(\theta,t,x)\\
&+ f(\theta,t,x)\Big(a_1(\theta,x) - \mathbb{E}\widehat{F}_D(\theta,x)\Big)\Big\}.
\end{aligned}$$

Firstly, observe that the results below

$$\text{(6.8)} \qquad \frac{1}{\phi_{\theta,x}(h_K)}\mathbb{E}\left[K^l\left(\frac{\langle x-X_i,\theta\rangle}{h_K}\right)\right] \longrightarrow a_l(\theta,x), \qquad \text{as} \quad n\longrightarrow\infty, \quad \text{for} \quad l=1,2\,,$$

$$\text{(6.9)} \qquad \mathbb{E}\left[\widehat{F}_D(\theta,x)\right] \longrightarrow a_1(\theta,x), \qquad \text{as} \quad n\longrightarrow\infty\,,$$

and

$$\text{(6.10)} \qquad \mathbb{E}\left[\widehat{f}_n(\theta,t,x)\right] \longrightarrow a_1(\theta,x)\, f(\theta,t,x), \qquad \text{as} \quad n\longrightarrow\infty\,,$$

can be proved in the same way as in Ezzahrioui and Ould Saïd [10] corresponding to their Lemmas 5.1 and 5.2, and then their proofs omitted.

Secondly, on one hand, making use of (6.8), (6.9) and (6.10) we have

$$\left\{\mathbb{E}\widehat{f}_N(\theta,t,x) - a_1(\theta,x)f(\theta,t,x) + f(\theta,t,x)\left(a_1(\theta,x) - \mathbb{E}\widehat{F}_D(\theta,x)\right)\right\} \underset{n\longrightarrow\infty}{\longrightarrow} 0\,.$$

On the other hand,

$$\text{(6.11)} \qquad \frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}}{\widehat{F}_D(\theta,x)} = \frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}\,\widetilde{f}(\theta,t,x)}{\widehat{F}_D(\theta,x)\widetilde{f}(\theta,t,x)} = \frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}\,\widetilde{f}(\theta,t,x)}{\widetilde{f}_N(\theta,t,x)}\,.$$

Then, using Proposition 2.1, it suffices to show that $\frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}}{\widetilde{f}_N(\theta,t,x)}$ tend to zero as n goes to infinity.

Indeed

$$\widetilde{f}_N(\theta,t,x) = \frac{1}{nh_H\mathbb{E}K_1(\theta,x)}\sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)}\,K\left(\frac{\langle x-X_i,\theta\rangle}{h_K}\right)H\left(\frac{t-Y_i}{h_H}\right).$$

Because $K(\cdot)\,H(\cdot)$ is continuous with support on [0,1], then by (H3) and (H4) $\exists\, m = \inf_{[0,1]} K(t)H(t)$ if follows that

$$\widetilde{f}_N(\theta,t,x) \geq \frac{m}{h_H\phi_{\theta,x}(h_K)}\,,$$

which gives

$$\frac{\sqrt{nh_H\phi_{\theta,x}(h_K)}}{\widetilde{f}_N(\theta,t,x)} \leq \frac{\sqrt{nh_H^3\phi_{\theta,x}^3(h_K)}}{m}\,.$$

Finally, using (N2), completes the proof of Lemma 4.2. □

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the referees and chief editor for their valuable suggestions which have considerably improved the earlier version of the paper.

REFERENCES

[1] Aït-Saidi, A.; Ferraty, F. and Kassa, R. (2005). Single functional index model for a time series, *Revue Roumaine de Mathématique Pures et Appliquées*, **50**, 321–330.

[2] Aït-Saidi, A.; Ferraty, F.; Kassa, R. and Vieu, P. (2008). Cross-validated estimation in the single functional index model, *Statistics*, **24**, 475–494.

[3] Attaoui, S. (2014). On the nonparametric conditional density and mode estimates in the single functional index model with strongly mixing data, *Sankhyã: The Indian Journal of Statistics*, **76**(A), 356–378.

[4] Attaoui, S. (2014). Strong uniform consistency rates and asymptotic normality of conditional density estimator in the single functional index modeling for time series data, *J. AStA.*, **98**, 257–286.

[5] Attaoui, S.; Laksaci, A. and Ould-Saïd, E. (2011). A note on the conditional density estimate in the single functional index model, *Statist. Probab. Lett.*, **81**, 45–53.

[6] Attaoui, S. and Ling, N. (2016). Asymptotic results of a nonparametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications, *Metrika*, **79**, 485–511.

[7] Bouchentouf, A.A.; Djebbouri, T.; Rabhi, A. and Sabri, K. (2014). Strong uniform consistency rates of some characteristics of the conditional distribution estimator in the functional single-index model, *Appl. Math. (Warsaw)*, **41**(4), 301–322.

[8] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data, *J. Statist. Plann. Inference.*, **147**, 1–23.

[9] Deheuvels, P. and Einmahl, J.H.J. (2000). Functional limit laws for the increments of Kaplan–Meier product-limit processes and applications, *Ann. Probab.*, **28**(3), 1301–1335.

[10] Ezzahrioui, M. and Ould Saïd, E. (2008). Asymptotic results of a nonparametric conditional quantile estimator for functional time series, *Comm. Statist. Theory and Methods*, **37**(16–17), 2735–2759.

[11] Ezzahrioui, M. and Ould Saïd, E. (2010). Some asymptotic results of a nonparametric conditional mode estimator for functional time series data, *Statist. Neerlandica*, **64**, 171–201.

[12] Ferraty, F.; Laksaci, A.; Tadj, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, *J. Statist. Plann. and Inf.*, **140**, 335–352.

[13] Ferraty, F.; Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models, *Statist. Inference Stoch. Process*, **9**, 47–76.

[14] Ferraty, F.; Peuch, A. and Vieu, P. (2003). Modèle à indice fonctionnel simple, *C. R. Acad. Sci., Paris*, **336**, 1025–1028.

[15] Ferraty, F. and Vieu, P. (2003). *Functional nonparametric statistics: a double infinite dimensional framework*. In "Recent Advances and Trends in Nonparametric Statistics" (M. Akritas and D. Politis, Ed.), Elsevier.

[16] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, New York.

[17] Goia, A. and Vieu, P. (2015). A partitioned single functional index model, *Computational Statistics*, **30**(3), 673–692.

[18] Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics, *J. Mutiv. Anal.*, **146**, 1–6.

[19] Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation, *Ann. Statist.*, **13**, 1465–1481.

[20] HRISTACHE, M.; JUDITSKY, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficient in the single-index model, *Ann. Statist.*, **29**, 595–623.

[21] HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*, Wiley, Chichester.

[22] KHARDANI, S.; LEMDANI, M. and OULD SAÏD, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship, *Journal of the Korean Statistical Society*, **39**, 455–469.

[23] KHARDANI, S.; LEMDANI, M. and OULD SAÏD, E. (2011). Uniform rate of strong consistency for a smooth Kernel estimator of the conditional mode under random censorship, *J. Statist. Plann. and Inf.*, **141**, 3426–3436.

[24] KHARDANI, S.; LEMDANI, M. and OULD SAÏD, E. (2014). On the Central Limit Theorem for a conditional mode estimator of a randomly censored time series, *Journal of Statistical Theory and Practice*, **8**, 722–742.

[25] LING, N. and XU, Q. (2012). Asymptotic normality of conditional density estimation in the single index model for functional time series data, *Statistics and Probability Letters*, **82**, 2235–2243.

[26] MASRY, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stochastic Process. Appl.*, **115**, 155–177.

[27] MEHRA, K.L.; RAMAKRISHNAIAH, Y.S. and SASHIKALA, P. (2000). Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode, *Ann. Inst. Math. Stat.*, **52**, 630–645.

[28] OULD-SAÏD, E. (2006). A strong uniform convergence rate of Kernel conditional quantile estimator under random censorship, *Statistics and Probability Letters*, **76**, 579–586.

[29] OULD-SAÏD, E. and CAI, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function, *Nonparametric Statistics*, **17**(7), 797–806.

[30] RAMSAY, J. and SILVERMAN, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.

[31] RAMSAY, J. and SILVERMAN, B. (2005). *Functional Data Analysis* (2nd ed.), Springer, New York.

[32] SAMANTA, M. and THAVANESWARAN, A. (1990). Nonparametric estimation of the conditional mode, *Commun. Stat. Theory Methods*, **16**, 4515–4524.

[33] XIA, Y. and HÄRDLE, W. (2006). semi-parametric estimation of partially linear single-index models, *J. Multivar. Anal.*, **97**, 1162–1184.

REVSTAT – Statistical Journal
Volume 20, Number 2, April 2022, 251–276

CONSTRUCTION OF T_m-TYPE AND T_m-ASSISTED PBIB DESIGNS

Authors: Parneet Kaur *
– Research Scholar, Department of Statistics, Punjabi University, Patiala, India
parneet.nonu93@gmail.com

Davinder Kumar Garg
– Professor, Department of Statistics, Punjabi University, Patiala, India
dkgarg_stat@yahoo.co.in

Received: November 2018 Revised: January 2020 Accepted: February 2020

Abstract:

- Two-associate class triangular designs have been explored greatly but the T_m-type PBIB designs ($m \geq 3$) largely remains unexplored. The paper is written with an objective to construct a new series of T_m-type PBIB designs and to derive some more series of PBIB designs based on these PBIB designs, which we have called as T_m-assisted PBIB designs. For this, we begin by first constructing a series of T_m-type PBIB designs and then based on these designs, three series of T_m-assisted PBIB designs have been constructed. The association schemes of T_m-type and T_m-assisted PBIB designs have been discussed in their complete generalized form.

Keywords:

- *block designs; triangular association scheme; PBIB designs.*

AMS Subject Classification:

- 05B05, 05E30, 51E05, 62K10.

*Corresponding author.

1. INTRODUCTION

The first step before performing an experiment is to devise the way various treatments are allotted to different experimental units. Experimental designs assist us in performing this task keeping in mind various constraints like constraints on experimental material, constraints on the cost of the experimental setup, etc.

In the class of block designs, incomplete block designs are used whenever the constraints on the experimental materials do not allow us to use complete blocks or when the heterogeneity increases as a result of formation of complete blocks. Partially Balanced Incomplete Block (PBIB) designs, which fall under the category of incomplete block designs, are among the popular incomplete block designs which help in making treatment comparisons by utilizing lesser experimental material.

Since their introduction, the two-associate class PBIB designs have been studied to a greater extent by many authors. Notable among them are Bose and Nair ([1]), Bose and Shimamoto ([2]), Raghavarao ([12]), Clatworthy *et al.* ([5]), etc. Among the two associate class association scheme, triangular association scheme is interesting due to its specific arrangement of symbols in the form of a matrix. Ogaswara ([11]) generalized the triangular association scheme and introduced $\mathrm{T_m}$ – association scheme with m – associate classes. Later, John ([7]), Saha ([15]), Sinha ([22]), Sinha ([23]), Cheng *et al.* ([4]), Meitei ([8]), Sinha *et al.* ([24]), Singh ([19]), etc., studied the $\mathrm{T_m}$-association scheme and constructed some $\mathrm{T_m}$-type PBIB designs. Recently, Ruj and Roy (2007) have shown the applications of PBIB designs in key predistribution by using triangular association scheme.

Even today, the higher associate class $\mathrm{T_m}$-type PBIB designs ($\mathrm{m} \geq 3$) largely remains unexplored and because the higher associate class PBIB designs are needed as they provide us new and more efficient PBIB designs as discussed by Raghavarao ([12]), Sharma and Garg ([17]), etc., the present paper has been written to make some contributions towards the construction of higher-associate class $\mathrm{T_m}$-type PBIB designs. For this, we will begin by discussing $\mathrm{T_m}$-type association scheme and then construct a new series of $\mathrm{T_m}$-type PBIB designs for which we have discussed a methodology to directly obtain the incidence matrix of the designs. After this, we will employ some easy and interesting techniques which utilize triangular type PBIB designs to construct some more four-associate class PBIB designs with different parametric combinations. We have called these designs as $\mathrm{T_m}$-assisted PBIB designs. Their corresponding association schemes have also been discussed. We have also prepared a table containing PBIB designs constructed by using different methodologies discussed in this paper. This table compares the listed designs with designs having same parameters listed in Clatworthy *et al.* ([5]). The R programming codes for the computerized construction of all the series have been given in the Appendices.

2. $\mathrm{T_m}$-TYPE ASSOCIATION SCHEME

$\mathrm{T_m}$-type or m-dimensional triangular association scheme was first defined by Ogasawara ([11]) and we can refer to it for its original definition. We have also defined the $\mathrm{T_m}$-type association scheme, for any arbitrary $\mathrm{m} \geq 2$, in the following way:

Let us consider a vector of size b containing only binary elements 0's and 1's such that the vector contains r unit elements and (b – r) 0's. The number of associate classes, i.e., m is given by:

$$\mathrm{m} = \min(\mathrm{r}, (\mathrm{b} - \mathrm{r})).$$

Let us say that this vector denotes our first treatment. The remaining (v – 1) treatments can be obtained by taking different combinations of the elements of this vector and in all we will get $\mathrm{v} = \binom{\mathrm{b}}{\mathrm{r}}$ treatments. From this set of treatments, a treatment is the 0^{th}-associate of its own and two treatments are mutually $\mathrm{i^{th}}$-associates if the corresponding vectors have (m – i) unit or null elements in common according to if $\mathrm{r} < (\mathrm{b} - \mathrm{r})$ or $(\mathrm{b} - \mathrm{r}) < \mathrm{r}$ respectively. The parameters of the association scheme are:

$$\mathrm{v} = \binom{\mathrm{b}}{\mathrm{r}},$$

$$\mathrm{n_i} = \binom{\mathrm{m}}{\mathrm{i}}\binom{\mathrm{b}-\mathrm{m}}{\mathrm{i}}; \quad \mathrm{i} = 0, 1, 2, ..., \mathrm{m},$$

$$\mathrm{p^i_{jk}} = \sum_{\mathrm{u}=0}^{\mathrm{m}-\mathrm{i}} \binom{\mathrm{m}-\mathrm{i}}{\mathrm{u}}\binom{\mathrm{i}}{\mathrm{m}-\mathrm{j}-\mathrm{u}}\binom{\mathrm{i}}{\mathrm{m}-\mathrm{k}-\mathrm{u}}\binom{\mathrm{b}-\mathrm{m}-\mathrm{i}}{\mathrm{j}+\mathrm{k}+\mathrm{u}-\mathrm{m}}; \quad \mathrm{i,j,k} = 0, 1, 2, ..., \mathrm{m}.$$

3. CONSTRUCTION OF $\mathrm{T_m}$-TYPE PBIB DESIGNS

Just like in association scheme, form a vector of size b. Now make all possible combinations of the elements of this vector and in all we will get $\binom{\mathrm{b}}{\mathrm{r}}$ different vectors. Now make a matrix N of order $\binom{\mathrm{b}}{\mathrm{r}} \times \mathrm{b}$ matrix using these $\binom{\mathrm{b}}{\mathrm{r}}$ vectors as its rows such that any vector may form any row of this matrix. The resulting $\binom{\mathrm{b}}{\mathrm{r}} \times \mathrm{b}$ matrix will be a matrix of 0's and 1's which will form the incidence matrix of a PBIB design following the $\mathrm{T_m}$-type association scheme defined in Section 2 with the following parameters:

$$\mathrm{v} = \binom{\mathrm{b}}{\mathrm{r}}, \quad \mathrm{b}, \quad \mathrm{r}, \quad \mathrm{k} = \binom{\mathrm{b}-1}{\mathrm{r}-1}, \qquad \lambda_{\mathrm{i}} = \mathrm{r} - \mathrm{i}; \quad \mathrm{i} = 1, 2, ..., \mathrm{m}.$$

Example 3.1. To illustrate the above construction methodology, let us consider a vector of size 6 having three unit (1) elements and three zero (0) elements. Following the above discussed methodology and taking all possible combinations of this vector, we will get 20 distinct vectors in all which will form the rows of the following incidence matrix N:

$$N = \begin{bmatrix} 1&0&0&1&0&1\\ 1&1&1&0&0&0\\ 1&1&0&0&0&1\\ 1&1&0&0&1&0\\ 1&1&0&1&0&0\\ 1&0&1&1&0&0\\ 1&0&1&0&0&1\\ 1&0&1&0&1&0\\ 1&0&0&0&1&1\\ 1&0&0&1&1&0\\ 0&1&1&1&0&0\\ 0&1&1&0&0&1\\ 0&1&1&0&1&0\\ 0&1&0&1&1&0\\ 0&1&0&0&1&1\\ 0&1&0&1&0&1\\ 0&0&1&1&1&0\\ 0&0&1&1&0&1\\ 0&0&1&0&1&1\\ 0&0&0&1&1&1 \end{bmatrix}.$$

The incidence matrix N corresponds to the following set of six blocks:

$$\begin{array}{ll}
(1, 2, 3, 4, 5, 6, 7, 8, 9, 10), & (2, 3, 4, 5, 11, 12, 13, 14, 15, 16),\\
(2, 6, 7, 8, 11, 12, 13, 17, 18, 19), & (1, 5, 6, 10, 11, 14, 16, 17, 18, 20),\\
(4, 8, 9, 10, 13, 14, 15, 17, 19, 20), & (1, 3, 7, 9, 12, 15, 16, 18, 19, 20).
\end{array}$$

The above set of blocks will constitute a T_m-type PBIB design with the following parameters:

$$\begin{array}{llll}
v = 20, & b = 6, & r = 3, & k = 10,\\
\lambda_1 = 2, & \lambda_2 = 1, & \lambda_3 = 0, & \\
n_1 = 9, & n_2 = 9, & n_3 = 1, & \\
P_1 = \begin{bmatrix} 4&4&0\\ 4&4&1\\ 0&1&0 \end{bmatrix}, & P_2 = \begin{bmatrix} 4&4&1\\ 4&4&0\\ 1&0&0 \end{bmatrix}, & P_3 = \begin{bmatrix} 0&9&0\\ 9&0&0\\ 0&0&0 \end{bmatrix}. &
\end{array}$$

4. CONSTRUCTION OF SOME T_m-ASSISTED PBIB DESIGNS

There is always a requirement for the construction of PBIB designs which are either new or more efficient than the already existing PBIB designs in the literature. For this purpose, the researchers are working towards the construction of higher associate class PBIB designs by devising new construction techniques and association schemes. In this section, we will use the T_m-type PBIB designs constructed above and apply some interesting techniques to construct some more PBIB designs. Already existing T_m-type PBIB designs can also be used. The association schemes of these derived PBIB designs have also been discussed.

4.1. Method – I

Using the incidence matrix of the $\mathrm{T_m}$-type PBIB design constructed in Section 3 and its complementary matrix, we can obtain the incidence matrix of another PBIB design through matrix augmentation. Let us say that N be the incidence matrix of the $\mathrm{T_m}$-type PBIB design as obtained in Section 3 and $\mathrm{N^C}$ be its complement. Then another matrix $\mathrm{N_1}$ can be obtained through matrix augmentation as below:

$$\mathrm{N_1} = \left[\begin{array}{c|c} \mathrm{N} & \mathrm{N^C} \\ \hline \mathrm{N^C} & \mathrm{N} \end{array}\right].$$

Here, N is the incidence matrix of the $\mathrm{T_m}$-type PBIB designs with the following parameters:

$$\mathrm{v} = \binom{\mathrm{b}}{\mathrm{r}}, \quad \mathrm{b}, \quad \mathrm{r}, \quad \mathrm{k} \quad \text{and} \quad \lambda_{\mathrm{i}};\ \mathrm{i} = 1, 2, ..., \mathrm{m},$$

and $\mathrm{n_i}$ be the number of $\mathrm{i^{th}}$-associates of any treatment. (A)
Therefore, $\mathrm{N^C}$ is the incidence matrix of the Tm-type PBIB designs with the parameters given below:

$$\mathrm{v^c} = \binom{\mathrm{b^c}}{\mathrm{r^c}}, \quad \mathrm{b^c}, \quad \mathrm{r^c}, \quad \mathrm{k^c} \quad \text{and} \quad \lambda^{\mathrm{c}}_{\mathrm{i}};\ \mathrm{i} = 1, 2, ..., \mathrm{m},$$

and $\mathrm{n_i}$ be the number of $\mathrm{i^{th}}$-associates of any treatment. (B)
Then the matrix $\mathrm{N_1}$ will form the incidence matrix of the PBIB design having (2m + 1) associate classes and with the following parameters:

$$\mathrm{v_1} = \mathrm{v} + \mathrm{v^c}, \qquad \mathrm{b_1} = \mathrm{b} + \mathrm{b^c}, \qquad \mathrm{r_1} = \mathrm{r} + \mathrm{r^c}, \qquad \mathrm{k_1} = \mathrm{k} + \mathrm{k^c}.$$

Let the difference between r and $\mathrm{r^c}$ be given by $\mathrm{d} = |\mathrm{r} - \mathrm{r^c}|$.

(i) $\mathrm{d} = 0$	(ii) $\mathrm{d} = 1$	(iii) $\mathrm{d} \geq 2$
$\lambda^1_{\mathrm{i}} = \lambda_{\mathrm{i}} + \lambda^{\mathrm{c}}_{\mathrm{i}}$; $\mathrm{i} = 1, 2, ..., \mathrm{m}$ $\lambda^1_{\mathrm{m+1}} = \mathrm{r_1}$ $\lambda^1_{\mathrm{j}} = \lambda^1_{\mathrm{j-(m+1)}}$; $\mathrm{j} = \mathrm{m}+2, \mathrm{m}+3, ..., 2\mathrm{m}+1$	$\lambda^1_{\mathrm{i}} = \mathrm{r_1} - \mathrm{i}$; $\mathrm{i} = 1, 2, ..., 2\mathrm{m}+1$	$\lambda^1_{\mathrm{i}} = \lambda_{\mathrm{i}} + \lambda^{\mathrm{c}}_{\mathrm{i}}$; $\mathrm{i} = 1, 2, ..., \mathrm{m}$ $\lambda^1_{\mathrm{m+1}} = 0$ $\lambda^1_{\mathrm{j}} = \lambda^1_{\mathrm{j-(m+1)}} - (\mathrm{d} - 2)$; $\mathrm{j} = \mathrm{m}+2, \mathrm{m}+3, ..., 2\mathrm{m}+1$

On the basis of the value of d, we have the following three association schemes. Let $\mathrm{S_i}(\alpha)$ denotes the set of $\mathrm{i^{th}}$-associates of treatment α in $\mathrm{T_m}$-association scheme (i = 1, 2, ..., m). Therefore,

$$\mathrm{p^i_{jk}} = |\mathrm{S_j}(\alpha) \cap \mathrm{S_k}(\beta)|,$$

where α and β are mutually $\mathrm{i^{th}}$-associates and $|\mathrm{S_i}(\alpha)|$ represents the size of set $\mathrm{S_i}(\alpha)$.

4.1.1. Association scheme for d = 0

We already know that the rows of the matrix N denote the corresponding treatment number. Now keeping in view, the parameters (A), (B) and incidence matrix N_1, we define the following association scheme:

i. Two treatments are said to be i^{th}-associates if the inner product of the corresponding rows in the matrix N_1 is $\lambda_i^1 = \lambda_i + \lambda_i^c$ (i = 1, 2, ..., m).

ii. Two treatments are said to be $(m+1)^{th}$-associates if the inner product of the corresponding rows in the matrix N_1 is r.

iii. Two treatments are said to be j^{th}-associates if the inner product of the corresponding rows in the matrix N_1 is $\lambda_j^1 = \lambda_{(j-(m+1))}^1$; $j = m+2, m+3, ..., 2m+1$.

The above defined association scheme has the following parameters:

$$v_1 = \binom{b}{r} + \binom{b^c}{r^c}, \qquad n_i^1 = n_i; \quad i = 1, 2, ..., m,$$

$$n_{(m+1)}^1 = 1, \qquad n_{(m+j+1)}^1 = n_j; \quad j = 1, 2, ..., m,$$

$$p_{jk}^i = |S'_j(\alpha) \cap S'_k(\beta)|,$$

where α and β are mutually i^{th}-associates (i, j, k = 1, 2, ..., 2m+1), and
$S'_i(\alpha) = S_i(\alpha); \ \ i = 1, 2, ..., m,$
$S'_{i+m}(\alpha) = \{\alpha + v\},$
where v denotes the number of treatments in T_m-type association scheme,
$S'_{j+m+1}(\alpha) = S_j(\alpha + v); \ \ j = 1, 2, ..., m.$

4.1.2. Association scheme for d = 1

Like in association scheme 4.1.1, we keep in view the parameters (A), (B) and incidence matrix N_1 to define the following association scheme as:
Two treatments are said to be i^{th}-associates if the inner product of the corresponding rows in the matrix N is $\lambda_i^1 = r_1 - i; i = 1, 2, ..., 2m+1$.
Following are the parameters of the association scheme:

Suppose X is the set of values containing n_i's (i = 1, 2, ..., m) in the increasing order of their magnitudes. Therefore X contains m values with X(i) corresponding to the i^{th} value such that X(1) has the minimum and X(m) has the maximum value. Let $K_i(\alpha)$ denotes the sets $S_j(\alpha)$ arranged in ascending order of magnitudes for i, j = 1, 2, ..., m such that $K_1(\alpha)$ is minimum among $S_j(\alpha)$ and $K_m(\alpha)$ is maximum among $S_j(\alpha)$ (i, j = 1, 2, ..., m). Now the sets of i^{th}-associates in this association scheme are:

$S'_i(\alpha) = K_i(\alpha);\ \ i = 1, 2, ..., m,$
$S'_{i+m}(\alpha) = K_{(m-i+1)}(\alpha + v);\ \ i = 1, 2, ..., m,$
$S'_{2m+1}(\alpha) = \{\alpha + v\}.$
Keeping the cyclic order of treatments in mind.
Thus we have

$$p^i_{jk} = |S'_j(\alpha) \cap S'_k(\beta)|,$$

where α and β are mutually i^{th}-associates and i, j and k = 1, 2, ..., 2m+1,

$$v_1 = \binom{b}{r} + \binom{b^c}{r^c}, \qquad n^1_i = X(i);\ \ i = 1, 2, ..., m,$$

$$n^1_{(m+j)} = X(m - (j-1));\ \ j = 1, 2, ..., m \qquad \text{and} \qquad n^1_{(2m+1)} = 1.$$

4.1.3. Association scheme for d $\geq$ 2

Similar to association schemes 4.1.1 and 4.1.2, we keep in mind the parameters (A), (B) and incidence matrix N_1 to define our third association scheme:

i. Two treatments are said to be i^{th}-associates if the inner product of the corresponding rows in the matrix N_1 is $\lambda^1_i = \lambda_i + \lambda^c_i$ $(i = 1, 2, ..., m)$.

ii. Two treatments are said to be $(m+1)^{th}$-associates if the inner product of the corresponding rows in the matrix N_1 is 0.

iii. Two treatments are said to be j^{th}-associates if the inner product of the corresponding rows in the matrix N_1 is $\lambda^1_j = \lambda^1_{(j-(m+1))} - (d-2);\ j = m+2, m+3, ..., 2m+1.$

This association scheme has the following parameters:

$$v_1 = \binom{b}{r} + \binom{b^c}{r^c}, \qquad n^1_i = n_i;\ \ i = 1, 2, ..., m,$$

$$n^1_{(m+1)} = 1 \qquad \text{and} \qquad n^1_{(m+j+1)} = n_{(m-(j-1))};\ \ j = 1, 2, ..., m,$$

$$p^i_{jk} = |S'_j(\alpha) \cap S'_k(\beta)|,$$

where α and β are mutually i^{th}-associates (i, j, k = 1, 2, ldots, 2m+1), and
$S'_i(\alpha) = S_i(\alpha);\ \ i = 1, 2, ..., m,$
$S'_{i+m}(\alpha) = \{\alpha + v\},$
where v denotes the number of treatments in T_m-type association scheme,
$S'_{j+m+1}(\alpha) = S_{(m-j+1)}(\alpha + v);\ \ j = 1, 2, ..., m.$

Example 4.1. Suppose that we have a T_m-type PBIB design, with $v = \binom{5}{2}$ treatments where m = 2, and its complementary PBIB design has $v^c = \binom{5}{3}$ treatments.

For $v = \binom{5}{2}$ treatments, let us consider the following vector of size 5 with two unit elements and three zeroes. From the methodology discussed in Section 3 by taking all possible combinations of this vector, we will get 10 distinct vectors in all which will form the rows of the following incidence matrix N:

$$N = \begin{bmatrix} 1&1&0&0&0\\ 1&0&0&0&1\\ 1&0&0&1&0\\ 1&0&1&0&0\\ 0&1&1&0&0\\ 0&1&0&0&1\\ 0&1&0&1&0\\ 0&0&1&1&0\\ 0&0&1&0&1\\ 0&0&0&1&1 \end{bmatrix}.$$

The complement of this matrix, i.e., N^C is given as:

$$N^C = \begin{bmatrix} 0&0&1&1&1\\ 0&1&1&1&0\\ 0&1&1&0&1\\ 0&1&0&1&1\\ 1&0&0&1&1\\ 1&0&1&1&0\\ 1&0&1&0&1\\ 1&1&0&0&1\\ 1&1&0&1&0\\ 1&1&1&0&0 \end{bmatrix}.$$

Therefore, according to the above discussed method, we can use the above T_m-type PBIB design and its complementary PBIB designs to construct another PBIB design with $v_1 = \binom{5}{2} + \binom{5}{3} = 20$ treatments having five-associate classes. Following is the incidence matrix of the five-associate class PBIB design constructed using the above discussed method which can be easily obtained using the incidence matrices of T_m-type PBIB design and its complementary PBIB design:

$$N_1 = \left[\begin{array}{c|c} N & N^C \\ \hline N^C & N \end{array}\right],$$

$$N_1 = \left[\begin{array}{ccccc|ccccc} 1&1&0&0&0&0&0&1&1&1\\ 1&0&0&0&1&0&1&1&1&0\\ 1&0&0&1&0&0&1&1&0&1\\ 1&0&1&0&0&0&1&0&1&1\\ 0&1&1&0&0&1&0&0&1&1\\ 0&1&0&0&1&1&0&1&1&0\\ 0&1&0&1&0&1&0&1&0&1\\ 0&0&1&1&0&1&1&0&0&1\\ 0&0&1&0&1&1&1&0&1&0\\ 0&0&0&1&1&1&1&1&0&0\\ \hline 0&0&1&1&1&1&1&0&0&0\\ 0&1&1&1&0&1&0&0&0&1\\ 0&1&1&0&1&1&0&0&1&0\\ 0&1&0&1&1&1&0&1&0&0\\ 1&0&0&1&1&0&1&1&0&0\\ 1&0&1&1&0&0&1&0&0&1\\ 1&0&1&0&1&0&1&0&1&0\\ 1&1&0&0&1&0&0&1&1&0\\ 1&1&0&1&0&0&0&1&0&1\\ 1&1&1&0&0&0&0&0&1&1 \end{array}\right].$$

The above incidence matrix corresponds to the following five-associate class PBIB design:

$$\begin{array}{lllll}
v_1 = 20, & b_1 = 10, & r_1 = 5, & k_1 = 10, & \\
\lambda_1^1 = 4, & \lambda_2^1 = 3, & \lambda_3^1 = 2, & \lambda_4^1 = 1, & \lambda_5^1 = 0, \\
n_1^1 = 3, & n_2^1 = 6, & n_3^1 = 6, & n_4^1 = 3, & n_5^1 = 1,
\end{array}$$

$$P_1^1 = \begin{bmatrix} 0&2&0&0&0\\ 2&0&4&0&0\\ 0&4&0&2&0\\ 0&0&2&0&1\\ 0&0&0&1&0 \end{bmatrix}, \quad P_2^1 = \begin{bmatrix} 1&0&2&0&0\\ 0&3&0&2&0\\ 2&0&3&0&1\\ 0&2&0&1&0\\ 0&0&1&0&0 \end{bmatrix}, \quad P_3^1 = \begin{bmatrix} 0&2&0&1&0\\ 2&0&3&0&1\\ 0&3&0&2&0\\ 1&0&2&0&0\\ 0&1&0&0&0 \end{bmatrix},$$

$$P_4^1 = \begin{bmatrix} 0&0&2&0&1\\ 0&4&0&2&0\\ 2&0&4&0&0\\ 0&2&0&0&0\\ 1&0&0&0&0 \end{bmatrix}, \quad P_5^1 = \begin{bmatrix} 0&0&0&3&0\\ 0&0&6&0&0\\ 0&6&0&0&0\\ 3&0&0&0&0\\ 0&0&0&0&0 \end{bmatrix}.$$

4.2. Method – II

4.2.1. Association scheme

Let us consider four quadrants and each of these contains exactly $n = \frac{n'(n'-1)}{2}$; $n' > 3$ treatments. These n treatments in each quadrant follow two-associate triangular association scheme with n treatments and can be arranged in a $n' \times n'$ matrix containing blank diagonal elements in such a way that all the n treatments are symmetric about the principal diagonal as shown in Figure 1. Thus, we have a total of $v = 4n$ treatments in all which follow the following four-associate class association scheme:

i. Two treatments are said to be the first associates if they belong to the same row and same column of the matrix in the same quadrant.

ii. The remaining treatments of the matrix in the same quadrant which are not the first associates of the treatment are said to be its second associates.

iii. Two treatments are said to be the third associates if they are in the adjacent quadrants.

iv. Treatments in the diagonal quadrants are the fourth associates.

Following are the parameters of the association scheme:

$$\begin{array}{ll}
v = 4n, & n = \dfrac{n'(n'-1)}{2}; n' > 3, \\
n_1 = 2(n'-2), & n_2 = \dfrac{(n'-2)(n'-3)}{2}, \\
n_3 = n'(n'-1) = 2n, & n_4 = \dfrac{n'(n'-1)}{2} = n,
\end{array}$$

$$\mathrm{P_1} = \begin{bmatrix} (\mathrm{n}'-2) & (\mathrm{n}'-3) & 0 & 0 \\ (\mathrm{n}'-3) & \frac{(\mathrm{n}'-3)(\mathrm{n}'-4)}{2} & 0 & 0 \\ 0 & 0 & \mathrm{n}'(\mathrm{n}'-1) & 0 \\ 0 & 0 & 0 & \frac{\mathrm{n}'(\mathrm{n}'-1)}{2} \end{bmatrix},$$

$$\mathrm{P_2} = \begin{bmatrix} 4 & 2\mathrm{n}'-8 & 0 & 0 \\ 2\mathrm{n}'-8 & \frac{(\mathrm{n}'-4)(\mathrm{n}'-5)}{2} & 0 & 0 \\ 0 & 0 & \mathrm{n}'(\mathrm{n}'-1) & 0 \\ 0 & 0 & 0 & \frac{\mathrm{n}'(\mathrm{n}'-1)}{2} \end{bmatrix},$$

$$\mathrm{P_3} = \begin{bmatrix} 0 & 0 & 2(\mathrm{n}'-2) & 0 \\ 0 & 0 & \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2} & 0 \\ 2(\mathrm{n}'-2) & \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2} & 0 & \frac{\mathrm{n}'(\mathrm{n}'-1)}{2} \\ 0 & 0 & \frac{\mathrm{n}'(\mathrm{n}'-1)}{2} & 0 \end{bmatrix},$$

$$\mathrm{P_4} = \begin{bmatrix} 0 & 0 & 0 & 2(\mathrm{n}'-2) \\ 0 & 0 & 0 & \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2} \\ 0 & 0 & \mathrm{n}'(\mathrm{n}'-1) & 0 \\ 2(n'-2) & \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2} & 0 & 0 \end{bmatrix}.$$

Note: It has been observed that we can consider any number of, say t ($\mathrm{t} \geq 3$), partitions instead of just four such that each partition has exactly $\mathrm{n} = \frac{\mathrm{n}'(\mathrm{n}'-1)}{2}; \mathrm{n}' > 3$ treatments following two-associate triangular association scheme arranged in $\mathrm{n}' \times \mathrm{n}'$ matrix in each partition. In such a situation, we have the following m-associate class association scheme:

i. Two treatments in the same row or same column of the matrix in the same partition are mutually first associates.

ii. Two treatments in the same partition but not in the same row or column of the matrix are second associates of each other.

iii. Treatments in the $(\mathrm{i}-2)^{\mathrm{th}}$ partition from the partition to which treatment, say α, belongs, are said to be the i^{th}-associates of treatment α (i = 3, 4, ..., m):

$$\mathrm{m} = \begin{cases} \frac{\mathrm{t}-1}{2} + 2; & \text{if t is odd,} \\ \frac{\mathrm{t}}{2} + 2; & \text{if t is even.} \end{cases}$$

The parameters of the above association scheme are:

$$\mathrm{v} = \mathrm{tn}; \quad \mathrm{t} \geq 3, \qquad \mathrm{n} = \frac{\mathrm{n}'(\mathrm{n}'-1)}{2}; \quad \mathrm{n}' > 3.$$

If t is odd

$$\mathrm{n_1} = 2(\mathrm{n}'-2), \qquad \mathrm{n_2} = \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2},$$

$$\mathrm{n_3} = \mathrm{n_4} = \mathrm{n_5} = \cdots = \mathrm{n_m} = 2\mathrm{n} = \mathrm{n}'(\mathrm{n}'-1).$$

If t is even

$$\mathrm{n_1} = 2(\mathrm{n}'-2), \qquad \mathrm{n_2} = \frac{(\mathrm{n}'-2)(\mathrm{n}'-3)}{2},$$

$$\mathrm{n_3} = \mathrm{n_4} = \mathrm{n_5} = \cdots = \mathrm{n_{(m-1)}} = 2\mathrm{n} = \mathrm{n}'(\mathrm{n}'-1), \qquad \mathrm{n_m} = \mathrm{n} = \frac{\mathrm{n}'(\mathrm{n}'-1)}{2}.$$

$*$	1	2	3	$\dots$	$n' - 1$	$*$	$n + 1$	$n + 2$	$n + 3$	$\dots$	$n + n' - 1$
1	$*$	n'	$n' + 1$	$\dots$	$2n' - 3$	$n + 1$	$*$	$n + n'$	$n + n' + 1$	$\dots$	$n + 2n' - 3$
2	n'	$*$	$2n' - 2$	$\dots$	$3n' - 6$	$n + 2$	$n + n'$	$*$	$n + 2n' - 2$	$\dots$	$n + 3n' - 6$
3	$n' + 1$	$2n' - 2$	$*$	$\vdots$	$\vdots$	$n + 3$	$n + n' + 1$	$n + 2n' - 2$	$*$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$*$	$n = \frac{n'(n'-1)}{2}$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$*$	$2n$
$n' - 1$	$2n' - 3$	$3n' - 6$	$\dots$	$n = \frac{n'(n'-1)}{2}$	$*$	$n + n' - 1$	$n + 2n' - 3$	$n + 3n' - 6$	$\dots$	$2n$	$*$
$*$	$3n + 1$	$3n + 2$	$3n + 3$	$\dots$	$3n + n' - 1$	$*$	$2n + 1$	$2n + 2$	$2n + 3$	$\dots$	$2n + n' - 1$
$3n + 1$	$*$	$3n + n'$	$3n + n' + 1$	$\dots$	$3n + 2n' - 3$	$2n + 1$	$*$	$2n + n'$	$2n + n' + 1$	$\dots$	$2n + 2n' - 3$
$3n + 2$	$3n + n'$	$*$	$3n + 2n' - 2$	$\dots$	$3n + 3n' - 6$	$2n + 2$	$2n + n'$	$*$	$2n + 2n' - 2$	$\dots$	$2n + 3n' - 6$
$3n + 3$	$3n + n' + 1$	$3n + 2n' - 2$	$*$	$\vdots$	$\vdots$	$2n + 3$	$2n + n' + 1$	$2n + 2n' - 2$	$*$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$*$	$4n$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$*$	$3n$
$3n + n' - 1$	$3n + 2n' - 3$	$3n + 3n' - 6$	$\dots$	$4n$	$*$	$2n + n' - 1$	$2n + 2n' - 3$	$2n + 3n' - 6$	$\dots$	$3n$	$*$

Figure 1: Arrangement of 4n treatments in four quadrants.

4.2.2. Series – I

Let us consider v = 4n treatments as discussed in association scheme defined in section 4.2.1 being arranged in four quadrants each containing exactly n treatments, where $n = \frac{n'(n'-1)}{2}$; $n' > 3$. We can form a set of b blocks from these v = 4n treatments such that i^{th} block has treatment i along with the treatments present in the quadrant diagonal to the quadrant containing treatment i. Thus in all we will obtain a set of b = 4n blocks and this set of b blocks constitute a four-associate class PBIB design following the association scheme 4.2.1 with the following parameters:

$$v = 4n = b, \qquad r = n + 1 = k, \quad n = \frac{n'(n'-1)}{2}; \quad n' > 3,$$

$$\lambda_1 = n = \lambda_2, \qquad \lambda_3 = 0, \qquad \lambda_4 = 2.$$

Example 4.2. Let us illustrate the above construction methodology by taking $n' = 4$. Thus, we have $n = \frac{4(4-1)}{2} = 6$ and we have a set of v = 24 treatments arranged as following:

$$\begin{array}{cccc|cccc}
* & 1 & 2 & 3 & * & 7 & 8 & 9 \\
1 & * & 4 & 5 & 7 & * & 10 & 11 \\
2 & 4 & * & 6 & 8 & 10 & * & 12 \\
3 & 5 & 6 & * & 9 & 11 & 12 & * \\
\hline
* & 19 & 20 & 21 & * & 13 & 14 & 15 \\
19 & * & 22 & 23 & 13 & * & 16 & 17 \\
20 & 22 & * & 24 & 14 & 16 & * & 18 \\
21 & 23 & 24 & * & 15 & 17 & 18 & *
\end{array}$$

Figure 2: Arrangement of 24 treatments.

From the above arrangement by taking the combinations of the i^{th} treatment with the treatments present in the diagonal quadrant, we will obtain the following set of blocks:

$$\begin{array}{lll}
(1,13,14,15,16,17,18) & (2,13,14,15,16,17,18) & (3,13,14,15,16,17,18) \\
(4,13,14,15,16,17,18) & (5,13,14,15,16,17,18) & (6,13,14,15,16,17,18) \\
(7,19,20,21,22,23,24) & (8,19,20,21,22,23,24) & (9,19,20,21,22,23,24) \\
(10,19,20,21,22,23,24) & (11,19,20,21,22,23,24) & (12,19,20,21,22,23,24) \\
(13,1,2,3,4,5,6) & (14,1,2,3,4,5,6) & (15,1,2,3,4,5,6) \\
(16,1,2,3,4,5,6) & (17,1,2,3,4,5,6) & (18,1,2,3,4,5,6) \\
(19,7,8,9,10,11,12) & (20,7,8,9,10,11,12) & (21,7,8,9,10,11,12) \\
(22,7,8,9,10,11,12) & (23,7,8,9,10,11,12) & (24,7,8,9,10,11,12)
\end{array}$$

The above set of 24 blocks will form a four-associate class PBIB design with the following parameters:

$$\mathrm{v} = 24 = \mathrm{b}, \qquad \mathrm{r} = 7 = \mathrm{k}, \qquad \lambda_1 = 6, \qquad \lambda_2 = 6, \qquad \lambda_3 = 0, \qquad \lambda_4 = 2.$$

The parameters of the association scheme are:

$$\mathrm{n}_1 = 4, \qquad \mathrm{n}_2 = 1, \qquad \mathrm{n}_3 = 12, \qquad \mathrm{n}_4 = 6,$$

$$\mathrm{P}_1 = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}, \qquad \mathrm{P}_2 = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix},$$

$$\mathrm{P}_3 = \begin{bmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & 1 & 0 \\ 4 & 1 & 0 & 6 \\ 0 & 0 & 6 & 0 \end{bmatrix}, \qquad \mathrm{P}_4 = \begin{bmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 12 & 0 \\ 4 & 1 & 0 & 0 \end{bmatrix}.$$

4.2.3. Series – II

As already stated earlier in the association scheme that n treatments, in each of the four quadrants, are arranged in $\mathrm{n}' \times \mathrm{n}'$ matrix with blank diagonal entries, we can form a set of blocks by taking a set of two rows of treatments from the adjacent quadrants and its treatments as the elements of block. In all, we will get a set of $\mathrm{b} = 4\mathrm{n}'^2$ blocks and this set of blocks will give us a four-associate class PBIB design following the association scheme 4.2.1 with the given below parameters:

$$\mathrm{v} = 4\mathrm{n}; \quad \mathrm{n} = \frac{\mathrm{n}'(\mathrm{n}' - 1)}{2}; \quad \mathrm{n}' > 3,$$

$$\mathrm{b} = 4\mathrm{n}'^2, \qquad \mathrm{r} = 4\mathrm{n}', \qquad \mathrm{k} = 2(\mathrm{n}' - 1).$$

$$\lambda_1 = 2\mathrm{n}', \qquad \lambda_2 = 0, \qquad \lambda_3 = 4, \qquad \lambda_4 = 0.$$

Example 4.3. To illustrate the above construction methodology, let us take $\mathrm{n}' = 4$. In this case, we have a set of $\mathrm{v} = 24$ treatments. Now from Figure 2, by taking the combinations of rows with rows in adjacent quadrants we obtain the following set of $\mathrm{b} = 64$ blocks:

(1, 2, 3, 7, 8, 9)	(1, 2, 3, 7, 10, 11)	(1, 2, 3, 8, 10, 12)
(1, 2, 3, 9, 11, 12)	(1, 2, 3, 19, 20, 21)	(1, 2, 3, 19, 22, 23)
(1, 2, 3, 20, 22, 24)	(1, 2, 3, 21, 23, 24)	(1, 4, 5, 7, 8, 9)
(1, 4, 5, 7, 10, 11)	(1, 4, 5, 8, 10, 12)	(1, 4, 5, 9, 11, 12)
(1, 4, 5, 19, 20, 21)	(1, 4, 5, 19, 22, 23)	(1, 4, 5, 20, 22, 24)
(1, 4, 5, 21, 23, 24)	(2, 4, 6, 7, 8, 9)	(2, 4, 6, 7, 10, 11)
(2, 4, 6, 8, 10, 12)	(2, 4, 6, 9, 11, 12)	(2, 4, 6, 19, 20, 21)
(2, 4, 6, 19, 22, 23)	(2, 4, 6, 20, 22, 24)	(2, 4, 6, 21, 23, 24)
(3, 5, 6, 7, 8, 9)	(3, 5, 6, 7, 1011)	(3, 5, 6, 8, 10, 12)
(3, 5, 6, 9, 11, 12)	(3, 5, 6, 19, 20, 21)	(3, 5, 6, 19, 22, 23)
(3, 5, 6, 20, 22, 24)	(3, 5, 6, 21, 23, 24)	(7, 8, 9, 13, 14, 15)
(7, 8, 9, 13, 16, 17)	(7, 8, 9, 14, 16, 18)	(7, 8, 9, 15, 17, 18)
(7, 10, 11, 13, 14, 15)	(7, 10, 11, 13, 16, 17)	(7, 10, 11, 14, 16, 18)
(7, 10, 11, 15, 17, 18)	(13, 14, 15, 19, 20, 21)	(13, 14, 15, 19, 22, 23)
(13, 14, 15, 20, 22, 24)	(13, 14, 15, 21, 23, 24)	(13, 16, 17, 19, 20, 21)
(13, 16, 17, 19, 22, 23)	(13, 16, 17, 20, 22, 24)	(13, 16, 17, 21, 23, 24)
(14, 16, 18, 19, 20, 21)	(14, 16, 18, 19, 22, 23)	(14, 16, 18, 20, 22, 24)
(15, 17, 18, 21, 23, 24)	(15, 17, 18, 19, 20, 21)	(15, 17, 18, 19, 22, 23)
(15, 17, 18, 20, 22, 24)	(15, 17, 18, 21, 23, 24)	

The above set of 64 blocks constitutes a four-associate class PBIB design following association scheme 4.2.1 having the following parameters:

$$v = 24, \quad b = 64, \quad r = 16, \quad k = 6, \quad \lambda_1 = 8, \quad \lambda_2 = 0, \quad \lambda_3 = 4, \quad \lambda_4 = 0.$$

The parameters of the association scheme are same as those given in Example 4.2.

4.2.4. Series – III

From the arrangement of v = 4n treatments in four quadrants as already discussed, if we form a set of blocks by taking two rows of treatments such that each of the row is placed in a quadrant diagonal to that of the other, and taking the treatment numbers as the elements of block, we will obtain a total of $b = 2n'^2$ blocks and this set of blocks will constitute a four-associate class PBIB design following the association scheme 4.2.1 with the following parameters:

$$v = 4n; \quad n = \frac{n'(n'-1)}{2}; \quad n' \geq 3,$$

$$b = 2n'^2, \qquad r = 2n', \qquad k = 2(n'-1),$$

$$\lambda_1 = n', \qquad \lambda_2 = 0, \qquad \lambda_3 = 0, \qquad \lambda_4 = 4.$$

Example 4.4. As an illustration, let us take of $n' = 4$. Now using Figure 2 and following the above discussed methodology, for the set of $v = 24$ treatments, we will get the following set of blocks:

$(1, 2, 3, 13, 14, 15)$	$(1, 2, 3, 13, 16, 17)$	$(1, 2, 3, 14, 16, 18)$
$(1, 2, 3, 15, 17, 18)$	$(3, 5, 6, 13, 14, 15)$	$(3, 5, 6, 13, 16, 17)$
$(3, 5, 6, 14, 16, 18)$	$(3, 5, 6, 15, 17, 18)$	$(1, 4, 5, 13, 14, 15)$
$(1, 4, 5, 13, 16, 17)$	$(1, 4, 5, 14, 16, 18)$	$(1, 4, 5, 15, 17, 18)$
$(7, 8, 9, 19, 20, 21)$	$(7, 8, 9, 19, 22, 23)$	$(7, 8, 9, 20, 22, 24)$
$(7, 8, 9, 21, 23, 24)$	$(2, 4, 6, 13, 14, 15)$	$(2, 4, 6, 13, 16, 17)$
$(2, 4, 6, 14, 16, 18)$	$(2, 4, 6, 15, 17, 18)$	$(7, 10, 11, 19, 20, 21)$
$(7, 10, 11, 19, 22, 23)$	$(7, 10, 11, 20, 22, 24)$	$(7, 10, 11, 21, 23, 24)$
$(8, 10, 12, 19, 20, 21)$	$(8, 10, 12, 19, 22, 23)$	$(8, 10, 12, 20, 22, 24)$
$(8, 10, 12, 21, 23, 24)$	$(9, 11, 12, 19, 20, 21)$	$(9, 11, 12, 19, 22, 23)$
$(9, 11, 12, 20, 22, 24)$	$(9, 11, 12, 21, 23, 24)$	

The above set of 32 blocks constitutes a four-associate class PBIB design based on association scheme with the following parameters:

$$v = 24, \quad b = 32, \quad r = 8, \quad k = 6, \quad \lambda_1 = 4, \quad \lambda_2 = 0, \quad \lambda_3 = 0, \quad \lambda_4 = 4.$$

The parameters of the association scheme are same as given in Example 4.2.

5. APPLICATIONS

Not only PBIB designs have found their applications in agricultural experi-mentation, these designs are being largely explored in various other fields also, for example, in group testing, cryptography, medicine, clinical trials, reliability theory, etc. Relevant work in this direction is due to authors like Smith ([25]), Hinkelmann and Kempthorne ([6]), Narain and Arya ([9]), Braun ([3]), Singh and Hinkelmann ([20]), etc. The applications of PBIB designs in sample surveys have been discussed by Raghavarao and Singh ([13]), Singh *et al.* ([21]), See *et al.* ([16]), Sharma and Garg ([18]), etc.

Apart from this, the application of triangular designs having two associate classes (T_2-type designs) in the key predistribution has been discussed by Ruj and Roy ([14]) in their landmark paper thereby increasing the scope of T_m-type PBIB designs. Let us discuss the applications of PBIB designs using the illustration of testing of car tires as discussed by Naseer and Jawad ([10]). Suppose that we want to test 10 different kinds of car tires manufactured by different companies. In this case, we have a set of $v = 10$ treatments and the blocks are in the form of different cars. Since each car can have only four tires, our blocks in this case are capable of accommodating only four treatments and complete block designs cannot be used. Therefore, we need an incomplete block design to test the different types of tires. Among the incomplete block designs, balanced incomplete block (BIB) designs are the only designs which are both variance balanced and efficiency balanced but these designs sometimes require a large number of experimental units to test a particular set of treatments.

For example, in order to test v = 10 different types of tires using blocks of size k = 4, we need a minimum of b = 15 blocks, i.e., we require at least 60 experimental units. Now, if such a large number of experimental units are not available due to various constraints, BIB design cannot be used. In such a situation, we can opt for a PBIB design. For example, in order to test the above set of v = 10 treatments, we can use method discussed in section 3 to construct a PBIB design having parameters v = 10, b = 5, r = 2 and k = 4, i.e., we need only 20 experimental units which are $\frac{1}{3}^{\text{rd}}$ of the total experimental units required in above discussed BIB design. The layout of this PBIB design is as given below:

Position of tire	Cars				
	1	2	3	4	5
Front Left	A	A	D	C	B
Front Right	B	E	E	G	F
Back Left	C	F	H	H	I
Back Right	D	G	I	J	J

6. CONCLUSION

The higher-associate class PBIB designs are important because these provide designs with new parametric combinations of v, b, r and k. Moreover, higher-associate class PBIB designs, many times, come out to be more efficient than the corresponding lower-associate class PBIB designs having the same values of the parameters v, b, r and k. This give rise to the need for the study and construction of PBIB designs with associate classes $m \geq 3$. In this direction, the present paper has been written to make contributions towards higher associate-class PBIB designs and for this we have first constructed a new series of T_m-type PBIB designs and then based on this series, we have obtained some series of T_m-assisted PBIB designs by employing some easy and interesting construction techniques. Thus, we have enhanced the literature of PBIB designs. We have also discussed the corresponding association schemes in their fully generalized forms. We have also demonstrated how these designs can further be used to construct more PBIB designs with different parameters based on different types of association schemes. It shows how unique combinatorial properties of T_m-type PBIB designs can assist in deriving more series of PBIB designs which should inspire the present and future researchers to further explore these designs.

We have also provided a table which enlists some of the PBIB designs constructed in this paper and compares them with some of the PBIB designs listed in the table of PBIB designs provided by Clatworthy *et al.* ([5]) for the same values of parameters v, b, r and k.

Table 1: PBIB designs constructed in this paper.

Sr. No.	**v**	**b**	**r**	**k**	λ_i	$\mathbf{O}_{\text{eff}}$	**Obtained From**
†1	6	4	2	3	$\lambda_1 = 1, \lambda_2 = 0$	0.6	Method 3
1 a	6	4	2	3	$\lambda_1 = 0, \lambda_2 = 1$	0.6	SR18
†2	10	5	2	4	$\lambda_1 = 1, \lambda_2 = 0$	0.5	Method 3
2 a	10	5	2	4	$\lambda_1 = 1, \lambda_2 = 0$	0.5	T28
†3	10	5	3	6	$\lambda_1 = 2, \lambda_2 = 1$	0.7778	Method 3
3 a	10	5	3	6	$\lambda_1 = 2, \lambda_2 = 1$	0.7778	T57
†4	15	6	2	5	$\lambda_1 = 1, \lambda_2 = 0$	0.4286	Method 3
4 a	15	6	2	5	$\lambda_1 = 1, \lambda_2 = 0$	0.4286	T48
*5	20	6	3	10	$\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0$	0.6842	Method 3
5 a	20	6	3	10	$\lambda_1 = 3, \lambda_2 = 1$	0.6491	S106
§6	20	10	5	10	$\lambda_1 = 4, \lambda_2 = 3, \lambda_3 = 2, \lambda_4 = 1, \lambda_5 = 0$	0.6758	Method 4.1
†7	21	7	2	6	$\lambda_1 = 1, \lambda_2 = 0$	0.375	Method 3
7 a	21	7	2	6	$\lambda_1 = 1, \lambda_2 = 0$	0.375	T65
‖8	21	7	5	15	$\lambda_1 = 4, \lambda_2 = 3$	0.9	Method 3
§9	24	24	7	7	$\lambda_1 = 6, \lambda_2 = 6, \lambda_3 = 0, \lambda_4 = 2$	0.3407	Method 4.2.2
‖10	24	64	16	6	$\lambda_1 = 8, \lambda_2 = 0, \lambda_3 = 4, \lambda_4 = 0$	0.3587	Method 4.2.3
11	24	32	8	6	$\lambda_1 = 4, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 4$	0.3261	Method 4.2.4
‖12	30	12	6	15	$\lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 0, \lambda_4 = 4, \lambda_5 = 2$	0.6973	Method 4.1
‖13	35	7	3	15	$\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0$	0.6078	Method 3
‖14	35	7	4	20	$\lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 1$	0.7794	Method 3
‖15	40	12	6	20	$\lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 0, \lambda_4 = 6, \lambda_5 = 4, \lambda_6 = 2, \lambda_7 = 0$	0.6923	Method 4.1
§16	40	50	10	8	$\lambda_1 = 5, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 4$	0.2795	Method 4.2.4
‖17	40	40	11	11	$\lambda_1 = 10, \lambda_2 = 10, \lambda_3 = 0, \lambda_4 = 2$	0.3136	Method 4.2.2
‖18	42	14	7	21	$\lambda_1 = 5, \lambda_2 = 3, \lambda_3 = 0, \lambda_4 = 4, \lambda_5 = 2$	0.7068	Method 4.1
‖19	56	8	3	21	$\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0$	0.5454	Method 3
‖20	70	14	7	35	$\lambda_1 = 6, \lambda_2 = 5, \lambda_3 = 4, \lambda_4 = 3, \lambda_5 = 2, \lambda_6 = 1, \lambda_7 = 0$	0.7057	Method 4.1

[*] Design constructed in the chapter and has greater overall efficiency factor than the corresponding design listed in Clatworthy *et al.* ([5]) for same parametric combination v, b, r and k.

[†] Design constructed in the chapter and has exactly same overall efficiency factor than the corresponding design listed in Clatworthy *et al.* ([5]) for same parametric combination v, b, r and k.

[§] Design constructed in the chapter and there is no listed design in Clatworthy *et al.* ([5]) for same parametric combination v, b, r and k.

[‖] Design constructed in the chapter with r, k > 10.

A. R PROGRAMMING CODES

The R programming code for the computerized construction of T_m-type and T_m-assisted PBIB designs discussed are as follows:

A.1. Code 1

Code for developing the PBIB designs discussed in Section 3:

```
series1<-function(x){
b<-length(x)
r<-0
for(i in 1:b){
if (x[i]==1){
r<-r+1 } }
v<-factorial(b)/(factorial(r)*factorial(b-r))
k<-v*r/b
l<-c()
m<-min(r,b-r)
for(i in 1:m){
l[i]<-r-i }
p<-1
mat<-combn(b,r)
N<-matrix(nrow=v,ncol=b)
cat("The Blocks of the Tm type PBIB design are:\n")
for(ii in 1:v){
for(jj in 1:b){
if(mat[p,ii]==jj){
N[ii,jj]<-1
p<-p+1
if(p>r){
p<-p%%r } }
else
N[ii,jj]<-0 } }
for(j in 1:b){
cat("(")
for(kk in 1:v){
if (N[kk,j]==1)
cat(kk," ") }
cat(")")
cat("\n") }
cat("The parameters of the above design are:\n")
cat("v = ",v,"\t b = ", b, "\t r = ", r, "\t k = ", k, "\n")
for(i in 1:m)
cat("lambda ",i, " = ", l[i],"\t" ) }
cat("\n") }
```

A.2. Code 2

Code for developing the PBIB designs discussed in 4.1:

```
series2<-function(x){
b1<-length(x)
r1<-0
r2<-0
for(i in 1:b1){
if (x[i]==1){
r1<-r1+1 }
else
r2<-r2+1 }
m1<-min(r1,b1-r1)
l1<-c()
l2<-c()
for(i in 1:m1){
l1[i]<-r1-i
l2[i]<-r2-i }
v1<-factorial(b1)/(factorial(r1)*factorial(b1-r1))
k1<-v1*r1/b1
p<-1
mat<-combn(b1,r1)
N<-matrix(nrow=v1,ncol=b1)
for(ii in 1:v1){
for(jj in 1:b1){
if(mat[p,ii]==jj){
N[ii,jj]<-1
p<-p+1
if(p>r1){
p<-p%%r1 } }
else
N[ii,jj]<-0 } }
Nc<-matrix(nrow = v1, ncol = b1)
for(p in 1:v1){
for (q in 1:b1) {
if(N[p,q]==1){
Nc[p,q]<-0 }
else
Nc[p,q]<-1 } }
v<-2*v1
b<-2*b1
r<-r1+r2
k<-v1
N1<-cbind(N,Nc)
N2<-cbind(Nc,N)
Nf<-rbind(N1,N2)
```

```
cat("The Blocks of the Tm assisted PBIB design are:\n")
for(j in 1:b){
cat("(")
for(kk in 1:v){
if (Nf[kk,j]==1)
cat(kk," ") }
cat(")")
cat("\n") }
D<-abs(r1-r2)
m<-2*m1+1
l<-c()
if(D==0){
for (i in 1:m1) {
l[i]<-l1[i]+l2[i] }
l[m1+1]<-r
for(i in (m1+2):m){
l[i]<-l[i-(m1+1)]
} } else if(D==1){
for (i in 1:m) {
l[i]<-r-i
} }else{
for (i in 1:m1) {
l[i]<-l1[i]+l2[i] }
l[m1+1]<-0
for (i in (m1+2):m) {
l[i]<-l[i-(m1+1)]-(D-2) } }
cat("The parameters of the above design are:\n")
cat("v = ",v,"\t b = ", b, "\t r = ", r, "\t k = ", k, "\n")
for(i in 1:m){
cat("lambda ",i, " = ", l[i],"\t" ) } }
```

A.3. Code 3

To execute the following function first install and load the R package 'Matrix'. Following is the R programming Code for developing the PBIB designs discussed in Section 4.2.2:

```
series3<-function(n1){
n<-n1*(n1-1)/2
v<-4*n
a<-1:n
A1<-matrix(nrow=n1,ncol=n1)
diag(A1)<-0
A1[lower.tri(A1)]<-1:n
B1<-forceSymmetric(A1,"L")
A2<-matrix(nrow=n1,ncol=n1)
```

```
diag(A2)<-0
A2[lower.tri(A1)]<-(n+1):(2*n)
B2<-forceSymmetric(A2,"L")
A3<-matrix(nrow=n1,ncol=n1)
diag(A3)<-0
A3[lower.tri(A3)]<-(2*n+1):(3*n)
B3<-forceSymmetric(A3,"L")
A4<-matrix(nrow=n1,ncol=n1)
diag(A4)<-0
A4[lower.tri(A4)]<-(3*n+1):(4*n)
B4<-forceSymmetric(A4,"L")
C1<-cbind(B1,B2)
C2<-cbind(B4,B3)
D<-rbind(C1,C2)
b<-v
r<-n+1
k<-r
l<-c(n,n,0,2)
cat("The Blocks of the design are:\n")
p<-1
for (i in 1:4) {
for (j in ((i-1)*n+1):(i*n)) {
if(i==1){
cat("(",j,(2*n+1):(3*n),")")
cat("\n")
} else if(i==2){
cat("(",j,(3*n+1):(4*n),")")
cat("\n")
} else if(i==3){
cat("(",j,(1):(n),")")
cat("\n")
} else{
cat("(",j,(n+1):(2*n),")")
cat("\n") } } }
cat("The parameters of the design are:\n")
cat("v = ",v,"\t","b = ",b,"\t","r = ",r,"\t","k = ",k,"\n")
for(i in 1:4){
cat("lambda_",i," = ",l[i],"\t") } }
```

A.4. Code 4

To execute the following function first install and load the R package 'Matrix'. Following is the R programming code for developing the PBIB designs discussed in Section 4.2.3:

```
series4<-function(n1){
n<-n1*(n1-1)/2
v<-4*n
a<-1:n
A1<-matrix(nrow=n1,ncol=n1)
diag(A1)<-0
A1[lower.tri(A1)]<-1:n
B1<-forceSymmetric(A1,"L")
A2<-matrix(nrow=n1,ncol=n1)
diag(A2)<-0
A2[lower.tri(A1)]<-(n+1):(2*n)
B2<-forceSymmetric(A2,"L")
A3<-matrix(nrow=n1,ncol=n1)
diag(A3)<-0
A3[lower.tri(A3)]<-(2*n+1):(3*n)
B3<-forceSymmetric(A3,"L")
A4<-matrix(nrow=n1,ncol=n1)
diag(A4)<-0
A4[lower.tri(A4)]<-(3*n+1):(4*n)
B4<-forceSymmetric(A4,"L")
C1<-cbind(B1,B2)
C2<-cbind(B4,B3)
D<-rbind(C1,C2)
b<-4*n1*n1
r<-4*n1
k<-2*(n1-1)
l<-c(2*n1,0,4,0)
p<-c()
pp<-c()
cat("The blocks of the design are:\n")
for (i in 1:n1) {
ii<-1
for (j in 1:n1) {
if(D[i,j]!=0){
p[ii]<-D[i,j]
ii<-ii+1 } }
for (jj in (1):(n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[jj,ij]!=0){
pp[iii]<-D[jj,ij]
iii<-iii+1 } }
```

```
cat(p,pp,"\n") } }
for (i in (n1+1):(2*n1)) {
ii<-1
for (j in (1):(n1)) {
if(D[i,j]!=0){
p[ii]<-D[i,j]
ii<-ii+1 } }
for (jj in (n1+1):(2*n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[jj,ij]!=0){
pp[iii]<-D[jj,ij]
iii<-iii+1 } }
cat(p,pp,"\n") } }
for (i in 1:n1) {
ii<-1
for (j in 1:n1) {
if(D[j,i]!=0){
p[ii]<-D[j,i]
ii<-ii+1 } }
for (jj in (1):(n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[ij,jj]!=0){
pp[iii]<-D[ij,jj]
iii<-iii+1 } }
cat(p,pp,"\n") } }
for (i in (n1+1):(2*n1)) {
ii<-1
for (j in (1):(n1)) {
if(D[j,i]!=0){
p[ii]<-D[j,i]
ii<-ii+1 } }
for (jj in (n1+1):(2*n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[ij,jj]!=0){
pp[iii]<-D[ij,jj]
iii<-iii+1 } }
cat(p,pp,"\n") } }
cat("The parameters of the above design are:\n")
cat("v = ",v,"\t b = ", b, "\t r = ", r, "\t k = ", k, "\n")
for(i in 1:4){
cat("lambda_",i, " = ", l[i],"\t" ) } }
```

A.5. Code 5

To execute the following function first install and load the R package 'Matrix'. Following is the R programming code for developing the PBIB designs discussed in Section 4.2.4:

```
series5<-function(n1){
n<-n1*(n1-1)/2
v<-4*n
a<-1:n
A1<-matrix(nrow=n1,ncol=n1)
diag(A1)<-0
A1[lower.tri(A1)]<-1:n
B1<-forceSymmetric(A1,"L")
A2<-matrix(nrow=n1,ncol=n1)
diag(A2)<-0
A2[lower.tri(A1)]<-(n+1):(2*n)
B2<-forceSymmetric(A2,"L")
A3<-matrix(nrow=n1,ncol=n1)
diag(A3)<-0
A3[lower.tri(A3)]<-(2*n+1):(3*n)
B3<-forceSymmetric(A3,"L")
A4<-matrix(nrow=n1,ncol=n1)
diag(A4)<-0
A4[lower.tri(A4)]<-(3*n+1):(4*n)
B4<-forceSymmetric(A4,"L")
C1<-cbind(B1,B2)
C2<-cbind(B4,B3)
D<-rbind(C1,C2)
b<-2*n1*n1
r<-2*n1
k<-2*(n1-1)
l<-c(n1,0,0,4)
p<-c()
pp<-c()
cat("The blocks of the design are:\n")
for (i in 1:n1) {
ii<-1
for (j in 1:n1) {
if(D[i,j]!=0){
p[ii]<-D[i,j]
ii<-ii+1 } }
for (jj in (n1+1):(2*n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[jj,ij]!=0){
pp[iii]<-D[jj,ij]
iii<-iii+1 } }
```

```
cat(p,pp,"\n") } }
for (i in (n1+1):(2*n1)) {
ii<-1
for (j in (1):(n1)) {
if(D[j,i]!=0){
p[ii]<-D[j,i]
ii<-ii+1 } }
for (jj in (1):(n1)) {
iii<-1
for (ij in (n1+1):(2*n1)) {
if(D[ij,jj]!=0){
pp[iii]<-D[ij,jj]
iii<-iii+1 } }
cat(p,pp,"\n") } }
cat("The parameters of the above design are:\n")
cat("v = ",v,"\t b = ", b, "\t r = ", r, "\t k = ", k, "\n")
for(i in 1:4){
cat("lambda_",i, " = ", l[i],"\t" ) } }
```

ACKNOWLEDGMENTS

The authors are highly thankful to the referees for their valuable suggestions and additions which make the paper more presentable.

REFERENCES

[1] Bose, R.C. and Nair, K.R. (1939). Partially balanced incomplete block designs, *Sankhyā: The Indian Journal of Statistics*, **4**(3), 337–372.

[2] Bose, R.C. and Shimamoto, T. (1952). Classification and analysis of partially balanced incomplete block designs with two associate classes, *Journal of the American Statistical Association*, **47**(258), 151–184.

[3] Braun, H.I. (1988). understanding scoring reliability: experiments in calibrating essay readers, *Journal of Educational Statistics*, **13**(1), 1–18.

[4] Cheng, C.S.; Constantine, G.M. and Hedayat, A.S. (1984). A unified method for constructing PBIB designs based on triangular and L2 schemes, *Journal of the Royal Statistical Society. Series B (Methodological)*, **41**(1), 31–37.

[5] Clatworthy, W.H.; Cameron, J.M. and Speckman, J.A. (1973). *Tables Of Two-Associate-Class Partially Balanced Designs*, NBS Appl. Math., Ser., 63.

[6] Hinkelmann, K. and Kempthorne, O. (1963). Two classes of group divisible partial diallel crosses, *Biometrika*, **50**(3/4), 281–291.

[7] John, P.W. (1966). An extension of the triangular association scheme to three associate classes, *Journal of the Royal Statistical Society. Series B (Methodological)*, **28**(2), 361–365.

[8] Meitei, K.K.S. (1996). A method of construction of triangular PBIB designs, *Journal of Indian Society of Agricultural Statistics*, **48**(2), 165–169.

[9] Narain, P. and Arya, A.S. (1981). Truncated triangular association scheme and related partial diallel crosses, *Sankhyā: The Indian Journal of Statistics, Series B*, **43**(1), 93–103.

[10] Nasser, M. and Jawad, B. (2008, January). *Design of experiments as effective design tools.* In "ASME 2008 International Mechanical Engineering Congress and Exposition", pp. 441–448, American Society of Mechanical Engineers Digital Collection.

[11] Ogasawara, M. (1965). *A necessary condition for the existence of regular and symmetrical PBIB designs of T_m-type*, Mimeograph Series No. 418, University of North Carolina, Institute of Statistics.

[12] Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments*, Wiley, New York.

[13] Raghavarao, D. and Singh, R. (1972). *Applications of PBIB designs in cluster sampling*, Paper presented at International Symposium on Combinatorial Mathematics and its Applications at New Delhi, India.

[14] Ruj, S. and Roy, B. (2007). *Key predistribution using partially balanced designs in wireless sensor networks.* In "International Symposium on Parallel and Distributed Processing and Applications", pp. 431–445, Springer, Berlin, Heidelberg.

[15] Saha, G.M. (1973). On construction of T m-type PBIB designs, *Annals of the Institute of Statistical Mathematics*, **25**(1), 605–616.

[16] See, K.; Song, S.Y. and Stufken, J. (1997). On a class of partially balanced incomplete block designs with applications in survey sampling, *Communications in Statistics – Theory and Methods*, **26**(1), 1–13.

[17] Sharma, K. and Garg, D.K. (2018). m-associate PBIB designs using Youden-m squares, *Communications in Statistics – Theory and Methods*, **47**(7), 1653–1662.

[18] Sharma, K. and Garg, D.K. (2019). Modified systematic sampling and the PBIB designs, *Utilitas Mathematica*, **113**, 243–263.

[19] Singh, G. (2014). Method of construction of triangular designs, *IOSR Journal of Mathematics (IOSR-JM)*, **10**(4), 24–25.

[20] Singh, M. and Hinkelmann, K. (1995). Partial diallel crosses in incomplete blocks, *Biometrics*, **51**(4), 1302–1314.

[21] Singh, R.; Raghavarao, D. and Federer, W.T. (1976). Applications of higher associate class PBIB designs in multidimensional cluster sampling, *Estadistica*, **30**, 202–209.

[22] Sinha, K. (1977). On the construction of m-associate PBIB designs I, *Journal of Statistical Planning and Inference*, **1**(2), 133–142.

[23] Sinha, K. (1981). On the construction of m-associate pbib designs II, *Statistics: A Journal of Theoretical and Applied Statistics*, **12**(4), 503–508.

[24] Sinha, K.; Kageyama, S.; Das, A. and Saha, G.M. (2004). Some series of block designs with nested rows and columns, *Australasian Journal of Combinatorics*, **29**, 337–352.

[25] Smith, K.J.C. (1968). *An application of incomplete block designs to the construction of error-correcting codes*, Institute of Statistics Memo Series No. 587, University of North Carolina.

REVSTAT-Statistical journal

Aims and Scope

The aim of REVSTAT-Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

Background

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work. At the time, the editorial board was mainly composed by Portuguese university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal. In 2001, the *Revista de Estatística* published a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name changed to REVSTAT-Statistical Journal, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

Editorial policy

REVSTAT-Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

All published works are Open Access (CC BY 4.0) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Journal prints may be ordered at expenses of the author(s), and prior to publication.

Abstract and Indexing services

REVSTAT-Statistical Journal is covered *by Journal Citation Reports - JCR (Clarivate); Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

Author guidelines

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage https://revstat.ine.pt/ based in Open Journal System (OJS). Authors intending to submit any work must ***register, login*** and follow the indications choosing ***Submissions*****.**

REVSTAT - Statistical Journal adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This theorem was proved later by AuthorB and AuthorC (1990); § This

subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998).

- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;
- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email and personal URL or ORCID number in the Comments for the Editor (submission form).

Accepted papers

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Copyright Notice

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information.

According to REVSTAT's *archiving policy*, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.