Statistica Sinica Preprint No: SS-2025-0155	
Title	Subgroup Testing in Change-Plane Models and Its Applications to Medical Data
Manuscript ID	SS-2025-0155
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0155
Complete List of Authors	Xu Liu, Jian Huang, Yong Zhou, Feipeng Zhang and Panpan Ren
Corresponding Authors	Panpan Ren
E-mails	panpanren@stu.sufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Subgroup testing in change-plane models and its applications to medical data

Xu Liu[1], Jian Huang[2], Yong Zhou[3], Feipeng Zhang[4] and Panpan Ren[1,*]

[1]*Shanghai University of Finance and Economics,* [2]*The Hong Kong Polytechnic University*

[3]East China Normal University and [4]Xi'an Jiaotong University

Abstract: Considered here is a hypothesis test for coefficients in change-plane models to detect the existence of a change plane, which in practice can guide personalized treatment recommendations. The considered test is from a class of problems where some parameters are not identifiable under the null hypothesis. Classic exponential average tests do not work well in practice. To overcome this, a novel test statistic is proposed by taking the weighted average of the squared score test statistic (WAST) over the grouping parameter's space, which has a closed form with an appropriate weight. The WAST significantly improves power in practice. Asymptotic distributions of the WAST are derived under the null and alternative hypotheses. A bootstrap method for approximating critical values is investigated and theoretically guaranteed. Moreover, the method is extended to the generalized estimating equation (GEE) framework and multiple change planes. The WAST performs well in simulations, and its performance is shown further by applying it to three medical datasets.

*Corresponding author. Email address: panpanren@stu.sufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Key words and phrases: Nonstandard tests, Precision medicine, Quantile regression, Subgroup detection, U-statistics.

1. Introduction

Subgroup analysis is a powerful tool in diverse fields covering clinical medicine and physical and social sciences. In clinical medicine, treatment effects are usually heterogeneous, meaning that the same treatment can have different effects in different subpopulations. Ignoring this heterogeneity in classical models may cause biased estimates, and identifying subgroups is the core problem in precision medicine (Fan et al., 2017; Huang et al., 2021). Methods for estimating and identifying subgroups have been studied extensively using various models (Assmann et al., 2000; Zhao et al., 2013; Wei and Kosorok, 2018; Yu and Fan, 2020; Hu et al., 2021; Wang et al., 2019; Liu et al., 2025). Before identifying subgroups, an appropriate statistic is required to test their existence. However, this test problem has received limited attention, especially for covariates that can divide the population into subgroups with different associations between predictors and responses. Existing literature on subgroup testing primarily focuses on change-plane models. Therefore, we address the testing problem in change-plane models, which belongs to the class of tests where some parameters are not identifi-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

able under the null hypothesis.

It is well known that breast cancer is the most common cancer among women worldwide. Many studies have shown that women with a *PALB2* mutation face a 14% risk of breast cancer by age 50 and 35% by age 70, therefore it is important to detect whether there is a subpopulation with the breast invasive carcinoma (BRCA) tumour stage based on gene mutations including important genes *BRCA1*, *BRCA2*, *PALB2*, among others (Antoniou et al., 2014). Also, skin cutaneous melanoma (SKCM) is an aggressive cancer with rising incidence, mortality, and disease burden (Siegel et al., 2021; Xiong et al., 2019). It is believed that *CREG1*, *TMEM201*, and *CCL8* are skin-cancer susceptibility genes (Hu et al., 2020; Zhang et al., 2021; Wan et al., 2020; Liao et al., 2020), so it is important to identify the subgroup of individuals with SKCM tumour status based on the gene mutations including those sensitive genes. Li et al. (2021) considered the multi-threshold change-plane model, which divides individuals into multiple subgroups with different covariate effects. As shown in Section 5, considering Study 175 of the AIDS Clinical Trials Group by modelling it with a similar model, the proposed test shows the existence of multiple change planes, and there is strong evidence for stratifying subjects into two or three groups based on covariates haemophilia, CD4 count at baseline, and race.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Herein, both discrete (e.g., BRCA data) and continuous (e.g., CD4 counts at 20 ± 5 weeks post-baseline in Study 175) outcomes are considered as responses, enabling the development of logistic and linear change-plane models, respectively. For the different types of data and goals, there are many change-plane models considered in the literature. This paper addresses the subgroup testing problem in change-plane models, where some parameters are non-identifiable under the null hypothesis.

In the ordinary testing problem with no identifiability issue under either the null or alternative hypothesis, the classic Wald- and score-type tests perform well. When nuisance parameters are present, Andrews and Ploberger (1994, 1995) studied tests using the weighted average power criterion, which was originally introduced by Wald (1943) when studying the likelihood ratio test under regularity conditions. Davies (1977, 2002) investigated the supremum of the squared score test statistic for mixture models. Kosorok and Song (2007); Song et al. (2009); Wu et al. (2016), and Kang et al. (2017) considered nonstandard tests for semiparametric models with applications of censoring data. Su et al. (2020) studied a score-type test to detect the existence of social network dependence with time-to-event data.

All the aforementioned methods require taking the supremum of a classic Wald- or score-type test statistic, which may not perform well due to

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

the large feasible region of calculating the supremum. To overcome these drawbacks, this paper introduces a novel test statistic based on the weighted average of the squared score test statistic (WAST) over the nuisance parameter's space. Our main contributions are threefold. First, based on generalized estimating equation (GEE), we propose a novel U-like test statistic, called *WAST*, which admits a closed-form expression when an appropriate weight is chosen. Second, we derive its asymptotic distributions under both the null and alternative hypothesis, based on the theory of degenerate U-statistics and representation of bivariate function in Hilbert-Schmidt class via spectral decomposition. The critical values can be easily obtained by a bootstrap method, whose validity is also proved theoretically. Third, we investigate several useful models under the GEE framework, such as quantile regression, generalized linear models and semiparametric models, and further extend the method to multiple change-plane detection. Extensive simulation studies show that the WAST controls type-I error well and achieves consistently higher power than competitors.

The rest of this paper is organized as follows. In Section 2, we introduce the models, and the novel test statistic, WAST, and establish its limiting distributions under the null and alternative hypotheses. In Section 3, we investigate the choice of the weight function based on several

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

useful examples that are all special cases of the considered GEEs, with detailed computations included for each example. In Section 4, we report the results of simulation studies conducted to evaluate the finite-sample performance of the proposed test statistic and to compare its power with competitors. In Section 5, we apply the proposed test method to a data from an AIDS study. Finally, in Section 6, we conclude with remarks and further extensions. The Assumptions are provided in the Appendix, the proofs, additional simulation studies and the real data analysis of the data set from BRCA are included in the Supplementary Material and in Supplementary S3.2, we further extend the proposed test procedure to the GEEs with multiple change planes. An R package named "wast" is available at https://github.com/xliusufe/wast.

2. Methodology

2.1 Preliminaries

In this section, we first present several change-plane models to motivate the proposed test statistic for subgroup identification. These examples—widely applied in fields such as finance and medicine—demonstrate the method's broad applicability. Building on these illustrations, we then formally define the hypothesis testing framework and derive the test statistic.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Example 1. Quantile regression. Zhang et al. (2022) and Lee et al. (2018, 2011) considered single-index thresholding in quantile regression:

$$Q_{Y_i}(\tau|\tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i) = \tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha}(\tau) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\tau) \geq 0), \tag{2.1}$$

where $Q_Y(\tau|\cdot)$ denotes the τth conditional quantile of the response Y given covariates $\tilde{\boldsymbol{X}}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$, $\tau \in (0,1)$ is the quantile level of interest, and $\boldsymbol{X}_i$ is a subset of $\tilde{\boldsymbol{X}}_i$. Let $\{\boldsymbol{V}_i = (Y_i, \tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i), i = 1, \cdots, n\}$. For fixed $\boldsymbol{\theta}$, the score function with respect to $\boldsymbol{\beta}$ under the null hypothesis is

$$\sum_{i=1}^{n} \psi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\tau) \geq 0)\psi_0(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}})$$

where $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \{\mathbf{1}(Y_i - \tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha} - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0) \leq 0) - \tau\}\boldsymbol{X}_i$, $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}(\tau)$ is the estimate of $\boldsymbol{\alpha}(\tau)$ under the null hypothesis through

$$\sum_{i=1}^{n} \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \{\mathbf{1}(Y_i - \tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha}(\tau) \leq 0) - \tau\}\tilde{\boldsymbol{X}}_i = \mathbf{0}.$$

Example 2. Generalized linear model (GLM). Huang et al. (2021) studied the logistic change-plane model, a type of GLMs with the likelihood

$$\exp\{Y_i\mu_i - c(\mu_i)\}\, a(Y_i), \text{ with } \mu_i = \tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0), \tag{2.2}$$

where μ_i is the canonical parameter, and $a(\cdot)$ and $c(\cdot)$ are known functions. For fixed $\boldsymbol{\theta}$, the score function with respect to $\boldsymbol{\beta}$ under the null is

$$\sum_{i=1}^{n} \psi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

where $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \{Y_i - \partial c(\mu_i)/\partial \mu_i\}\boldsymbol{X}_i$, $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\alpha}}$ is the estimate of $\boldsymbol{\alpha}$ under the null hypothesis according to

$$\sum_{i=1}^{n} \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \{Y_i - \partial c(\tilde{\boldsymbol{X}}_i^{\mathrm{T}} \boldsymbol{\alpha})/\partial \mu_i\}\tilde{\boldsymbol{X}}_i = \mathbf{0}.$$

Example 3. Semiparametric model. Fan et al. (2017) considered the semiparametric change-plane model

$$Y_i = \gamma(\boldsymbol{Z}_i) + \beta X_i \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\theta} \geq 0) + \epsilon_i, \tag{2.3}$$

where $\gamma(\boldsymbol{Z})$ is an unknown baseline mean function for patients in treatment $X = 0$, $X_i \in \{0, 1\}$, and $E(\epsilon_i | X_i, \boldsymbol{Z}_i)$=0. Let $\gamma(\boldsymbol{Z}, \boldsymbol{\alpha}_2)$ denote the baseline mean function, and $\pi(\boldsymbol{Z}, \boldsymbol{\alpha}_1)$ the propensity model. For fixed $\boldsymbol{\theta}$, the score function with respect to β under the null is

$$\sum_{i=1}^{n} \psi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}),$$

where $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \beta, \boldsymbol{\theta}) = \{X_i - \pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1)\}\{Y_i - \gamma(\boldsymbol{Z}_i, \boldsymbol{\alpha}_2) - \beta X_i \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\theta} \geq 0)\}$, $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, 0, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\alpha}} = \left(\hat{\boldsymbol{\alpha}}_1^{\mathrm{T}}, \hat{\boldsymbol{\alpha}}_2^{\mathrm{T}}\right)^{\mathrm{T}}$ is estimate of $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}})^{\mathrm{T}}$ under the null hypothesis according to

$$\sum_{i=1}^{n} \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \sum_{i=1}^{n} (D_{\boldsymbol{\alpha}_1}(\boldsymbol{Z}_i)\{X_i - \pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1)\}^{\mathrm{T}}, D_{\boldsymbol{\alpha}_2}(\boldsymbol{Z}_i)\{Y_i - \gamma(\boldsymbol{Z}_i, \boldsymbol{\alpha}_2)\}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{0}$$

with $D_{\boldsymbol{\alpha}_1}(\boldsymbol{Z}_i) = \{\pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1)(1 - \pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1))\}^{-1} \partial \pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1)/\partial \boldsymbol{\alpha}_1$ and $D_{\boldsymbol{\alpha}_2}(\boldsymbol{Z}_i) = \partial \gamma(\boldsymbol{Z}_i, \boldsymbol{\alpha}_2)/\partial \boldsymbol{\alpha}_2$. Following Fan et al. (2017), we emphasize that either the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

baseline mean function $\gamma(\boldsymbol{Z}, \boldsymbol{\alpha}_2)$ or the propensity model $\pi(\boldsymbol{Z}_i, \boldsymbol{\alpha}_1)$ is specified correctly, but not necessarily both.

Motivated by Examples 1–3, the subgroup testing problem considered herein is to test whether $\boldsymbol{\beta}$ is $\mathbf{0}$ or not, i.e.,

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}, \tag{2.4}$$

where the nuisance parameter $\boldsymbol{\theta}$ is only identifiable under the alternative hypothesis. For convenience of expression, we call $\boldsymbol{Z}$ the grouping variable, $\boldsymbol{\theta}$ the grouping parameter, $\boldsymbol{X}$ the grouping difference variable, and $\boldsymbol{\beta}$ the grouping difference parameter. The grouping difference parameter measures the effect difference between two subgroups. $\tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha}$ is the baseline model, $\tilde{\boldsymbol{X}}$ is the baseline variable, and $\boldsymbol{\alpha}$ is the effect of the baseline group.

From examples 1–3, it is natural to extend these classical models to the generalized estimating equation (GEE) framework. Let the observed data consist of $\{\boldsymbol{V}_i = (Y_i, \tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i), i = 1, \cdots, n\}$, which are n independent and identically distributed copies of $\boldsymbol{V} = (Y, \tilde{\boldsymbol{X}}, \boldsymbol{X}, \boldsymbol{Z})$. Consider the GEEs $\mathrm{E}[\mathbf{1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})] = \mathbf{0}$ and $\mathrm{E}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}) = \mathbf{0}$ and the corresponding empirical estimating functions

$$\begin{aligned} \Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}), \\ \Psi_{1n}(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}), \end{aligned} \tag{2.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

where $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha})$ are specified for different models, please see Examples 1–3 for more details. And $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_r)^{\mathrm{T}} \in \Theta_\alpha \subseteq \mathbb{R}^r$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{\mathrm{T}} \in \Theta_\beta \subseteq \mathbb{R}^p$, and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)^{\mathrm{T}} \in \Theta_\theta \subseteq \mathbb{R}^q$ are unknown parameters. Here, $\mathbf{1}(\cdot)$ denotes the indicator function, and $\psi_0(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})$ is independent of both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. For ease of expression, we denote $\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\psi_0(\boldsymbol{V}, \boldsymbol{\alpha}) = \psi_0(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})$. In many hypothesis testing problems, $\boldsymbol{\theta}$ is a nuisance parameter, such as in the examples above. Functions (2.5) for considering problem (2.4) is flexible, with many existing models included as special cases, such as examples 1–3.

We are interested in the testing problem (2.4). For any known $\boldsymbol{\theta} \in \Theta_\theta$, we consider a score test statistic (SST) for testing $\boldsymbol{\beta} = \mathbf{0}$, i.e.,

$$\tilde{T}_n(\boldsymbol{\theta}) = n^{-1}\|\Psi_n(\hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta})\|^2_{\tilde{V}(\boldsymbol{\theta})^{-1}}, \tag{2.6}$$

where $\hat{\boldsymbol{\alpha}}$ is an estimate of $\boldsymbol{\alpha}$ based on the empirical GEE $\Psi_{1n}(\boldsymbol{\alpha}) = 0$ and $\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\Psi_{1n}(\boldsymbol{\alpha})$ are defined in equation (2.5). The matrix $\tilde{V}(\boldsymbol{\theta})$ is given by $\tilde{V}(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n\{\psi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) - \hat{K}(\boldsymbol{\theta})\hat{J}\psi_1(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}})\}^{\otimes 2}$, where $\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0)\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha})$ and $\psi_0(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ are also defined in equation (2.5). Here, $\hat{K}(\boldsymbol{\theta})$ and $\hat{J}$ are consistent estimators of $K(\boldsymbol{\theta}) = \partial \mathrm{E}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})/\partial\boldsymbol{\alpha}^{\mathrm{T}} \in \mathbb{R}^{p\times r}$ and $J = \left\{\partial \mathrm{E}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)/\partial\boldsymbol{\alpha}^{\mathrm{T}}\right\}^{-1} \in \mathbb{R}^{r\times r}$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Lemma 1. *If Assumptions (A1)–(A5) in the Appendix hold, then for any fixed* $\boldsymbol{\theta} \in \Theta_\theta$, $\tilde{T}_n(\boldsymbol{\theta})$ *converges in distribution to a* χ^2 *random variable with* p *degrees of freedom under* H_0 *as* $n \to \infty$.

Remark 1. Because $\tilde{T}_n(\boldsymbol{\theta})$ includes an unknown parameter $\boldsymbol{\theta}$, Lemma 1 cannot be used directly in practice. However, it shows that the asymptotic distributions are the same for any nuisance parameter $\boldsymbol{\theta} \in \Theta_\theta$. Thus, the supremum and the weighted average of $\tilde{T}_n(\boldsymbol{\theta})$ achieve correct type-I errors. We construct the proposed test statistic by making use of this point.

Because equations (2.5) does not depend on $\boldsymbol{\theta}$ when $\boldsymbol{\beta} = 0$, the parameter $\boldsymbol{\theta}$ is identifiable only under the alternative hypothesis. Lemma 1 suggests that $\tilde{T}_n(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta_\theta$ is a test statistic that can control type-I errors. However, $\tilde{T}_n(\boldsymbol{\theta})$ may suffer a loss of power for some $\boldsymbol{\theta}$, and it is difficult to determine $\boldsymbol{\theta}$ in practice. To pursue the largest power theoretically, a good choice is to take the supremum of the squared score test statistic (SST) over the parametric space Θ_θ, i.e.,

$$\tilde{T}_n = \sup_{\boldsymbol{\theta} \in \Theta_\theta} \left\{ n^{-1} \| \Psi_n(\hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) \|^2_{\tilde{V}(\boldsymbol{\theta})^{-1}} \right\}. \tag{2.7}$$

The test statistic $\tilde{T}_n$ has been widely studied in the literature, see Andrews (1993); Andrews and Ploberger (1994); Davies (1977, 2002); Song et al. (2009); Fan et al. (2017) and Shen and Qu (2020). We summarize their

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

results under the null and alternative hypotheses in the GEE framework in Lemmas S4 and S5 in the Supplementary Material.

2.2 Weighted average of squared score test statistic

The SST considered in Section 2.1 may sacrifice power in practice for pursuing theoretically the largest power, especially for large q and p. To overcome this drawback and take advantage of Lemma 1, it is natural to average $\tilde{T}_n(\boldsymbol{\theta})$ over Θ_θ. In practice, we can take a weighted average over Θ_θ.

Given above arguments, we propose WAST based on $\tilde{T}_n(\boldsymbol{\theta})$ by taking integral after removing the inverse of covariance and cross-interaction terms:

$$T_n = \frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij} \tag{2.8}$$

with

$$\rho_{ij} = \psi_0(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \psi_0(\boldsymbol{V}_j, \hat{\boldsymbol{\alpha}}) \int_{\boldsymbol{\theta} \in \Theta_\theta} \mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\theta} \geq 0) \mathbf{1}(\boldsymbol{Z}_j^{\mathrm{T}} \boldsymbol{\theta} \geq 0) w(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

where $w(\boldsymbol{\theta})$ is a non-negative weight for all $\boldsymbol{\theta} \in \Theta_\theta$ and $\int_{\boldsymbol{\theta} \in \Theta_\theta} w(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$.

Another explanation for $w(\boldsymbol{\theta})$ comes from a Bayesian perspective, where the weight $w(\boldsymbol{\theta})$ can be treated as the prior density of the grouping parameter. Herein, we focus mainly on testing the existence of subgroups instead of estimating the grouping parameter, so there is no requirement to calculate the posterior distribution, simplifying the implementation.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

In general, there is no closed-form expression for ρ_{ij}, but it usually has a simple formulation if we choose a specific weight $w(\boldsymbol{\theta})$. The choice of weight can affect the computation of the test statistic since the numerical integration over $\mathbb{R}^q$ is required. Fortunately, there is a closed-form expression for an appropriate weight from the perspective of the conjugate priors, which is quite attractive in practice. Fu and Hong (2019) and Fu et al. (2023) have discussed a similar weight that facilitates the derivation of the closed-form expression. We discuss the calculation of ρ_{ij} in detail in Section 3.

To establish the asymptotic distribution of our statistic, we introduce additional notation. Let the kernel of a U-statistic under the null be

$$h(\boldsymbol{V}_i, \boldsymbol{V}_j) = \int_{\boldsymbol{\theta}\in\Theta_\theta} \psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_j, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta} + \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\mathrm{T}} K_j + K_i^{\mathrm{T}} \psi_1(\boldsymbol{V}_j, \boldsymbol{\alpha}_0) + \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\mathrm{T}} H \psi_1(\boldsymbol{V}_j, \boldsymbol{\alpha}_0), \tag{2.9}$$

where

$$H = \int_{\boldsymbol{\theta}\in\Theta_\theta} J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} K(\boldsymbol{\theta}) J w(\boldsymbol{\theta}) d\boldsymbol{\theta} \text{ and } K_i = \int_{\boldsymbol{\theta}\in\Theta_\theta} J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Theorem 1. *If Assumptions (A1)–(A5) in the Appendix hold, then under the null hypothesis, we have*

$$nT_n - \mu_0 \xrightarrow{\mathcal{L}} \zeta_0,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

where $\xrightarrow{\mathcal{L}}$ *denotes convergence in distribution,* $\mu_0 = 2\mathrm{E}[\psi_1(\boldsymbol{V}_1, \boldsymbol{\alpha}_0)^{\mathrm{T}} K_1] + \mathrm{E}[\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)^{\mathrm{T}} H \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)]$, ζ_0 *is a random variable of the form* $\zeta_0 = \sum_{j=1}^{\infty} \lambda_j (\chi^2_{1j} - 1)$, *and* $\chi^2_{11}, \chi^2_{12}, \cdots$ *are independent* χ^2_1 *variables, i.e.,* ζ_0 *has the characteristic function*

$$\mathrm{E}\left[e^{it\zeta_0}\right] = \prod_{j=1}^{\infty} (1 - 2it\lambda_j)^{-\frac{1}{2}} e^{-it\lambda_j}.$$

Here, $i = \sqrt{-1}$ *standards for the imaginary unit, and* $\{\lambda_j\}$ *are the eigenvalues of the kernel* $h(\boldsymbol{v}_1, \boldsymbol{v}_2)$ *under* $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$, *i.e., they are the solutions of* $\lambda_j g_j(\boldsymbol{v}_2) = \int_0^{\infty} h(\boldsymbol{v}_1, \boldsymbol{v}_2) g_j(\boldsymbol{v}_1) f(\boldsymbol{v}_1, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0) d\boldsymbol{v}_1$ *for nonzero* g_j, *where* $f(\boldsymbol{v}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ *is the density of* $\boldsymbol{V}$ *with parameters* $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, *and* $\boldsymbol{\theta}$.

Next, we investigate the power performance of the proposed test statistic under the local alternative.

Theorem 2. *If Assumptions (A1)–(A6) in the Appendix hold, then under the local alternative hypothesis* H_1, *that is,* $\boldsymbol{\beta} = n^{-1/2}\boldsymbol{\xi}$ *with a fixed vector* $\boldsymbol{\xi} \in \Theta_\beta$, *we have*

$$nT_n - \mu_0 \xrightarrow{\mathcal{L}} \zeta_1,$$

where μ_0 *is defined in Theorem 1,* ζ_1 *is a random variable of the form* $\zeta_1 = \sum_{j=1}^{\infty} \lambda_j \{\chi^2_{1j}(\mu_{aj}) - 1\}$, *and* $\chi^2_{11}(\mu_{a1}), \chi^2_{12}(\mu_{a2}), \cdots$ *are independent*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

noncentral χ_1^2 *variables, i.e.,* ζ_1 *has the characteristic function*

$$\mathrm{E}\left[e^{it\zeta_1}\right] = \prod_{j=1}^{\infty}(1-2it\lambda_j)^{-1/2}\exp\left\{(-it\lambda_j + it\lambda_j\mu_{aj})/(1-2it\lambda_j)\right\}.$$

Here, $\{\lambda_j\}$ *are the eigenvalues of the kernel* $h(\boldsymbol{v}_1, \boldsymbol{v}_2)$ *defined in* (2.9) *under* $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$*, and each noncentrality parameter of* $\chi^2_{1j}(\mu_{aj})$ *is*

$$\mu_{aj} = \mathrm{E}\left[\phi_j(\boldsymbol{V}_0)\boldsymbol{\xi}^{\mathrm{T}}\partial\log(f(\boldsymbol{V}_0, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0))/\partial\boldsymbol{\beta}\right], \quad j = 1, 2, \cdots,$$

where $\{\phi_j(\boldsymbol{v})\}$ *denotes orthonormal eigenfunctions corresponding to the eigenvalues* $\{\lambda_j\}$*, and* $\boldsymbol{V}_0$ *is generated from the null distribution* $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$*.*

Theorem 2 implies that the power function of $nT_n - \mu_0$ can be theoretically approximated by F_{ζ_1}, where F_{ζ_1} is the cumulative distribution function of ζ_1. The proof of Theorem 2 shows that $0 < \mathrm{E}[nh(\boldsymbol{V}_1, \boldsymbol{V}_2)] = \sum_{j=1}^{\infty}\lambda_j\mu_{aj} + o(1)$ under the local alternative hypothesis. The additional mean $\sum_{j=1}^{\infty}\lambda_j\mu_{aj}$ under H_1 can be viewed as a measure of the difference between H_0 and H_1. Because the asymptotic distribution is not common, it is difficult to use in practice; instead, in Section 2.3, we provide a bootstrap method for calculating the critical value or p-value.

Theorems 1–2 show that the bias μ_0 arises from estimating $\boldsymbol{\alpha}$, which means that μ_0 is zero if $\boldsymbol{\alpha}$ is known. In fact, if $\boldsymbol{\alpha}$ is known, then T_n is exactly a U-statistic with zero mean under the null hypothesis.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

2.3 Computation of critical value

In Section 2.2, we deduced the asymptotic distributions for the WAST T_n under the null and alternative hypotheses. However, because the distributions are not as common as normal, χ^2, and F distributions, it is difficult to calculate critical values directly. Instead, many related articles recommend the strategy of approximating the limiting null distribution of the statistic. Fan et al. (2017) proposed a perturbed test statistic for the SST

$$\tilde{T}_n^* = \sup_{\boldsymbol{\theta}\in\Theta_\theta} \frac{1}{n} \left\| \sum_{i=1}^{n} \left[\nu_i \{ \psi(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) + \hat{K}(\boldsymbol{\theta}) \hat{J} \psi_1(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}) \} \right] \right\|^2_{\tilde{V}(\boldsymbol{\theta})^{-1}},$$

where $\tilde{V}(\boldsymbol{\theta})$ is defined in (2.6), and $\{\nu_i, i = 1, \cdots, n\}$ are n independent and identically distributed samples from $\mathcal{N}(0, 1)$, independent of $\boldsymbol{V}$.

The critical value C_α can be calculated by the empirical distribution of $\tilde{T}_n^*$ according to replicates, say $\{\tilde{T}_n^{*,j}, j = 1, \cdots, N\}$, where C_α is an upper quantile of the empirical distribution of $\tilde{T}_n^*$ and N is a large integer. The null hypothesis is rejected for an α-level test when $\tilde{T}_n > C_\alpha$, and the p-value is approximated by $N^{-1} \sum_{j=1}^{N} \mathbf{1}(\tilde{T}_n^{*,j} > \tilde{T}_n)$.

For the WAST, we recommend the parametric bootstrap method to approximate the critical value or p-value. For each bootstrap sample $\{\boldsymbol{V}_i^{*,b}, i =$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

$1, \cdots, n\}$, $b = 1, \cdots, B$, we calculate $T_n^{*,b}$ by

$$T_n^{*,b} = \frac{1}{n(n-1)} \sum_{i \neq j} \int_{\boldsymbol{\theta} \in \Theta_\theta} \psi(\boldsymbol{V}_i^{*,b}, \hat{\boldsymbol{\alpha}}^{*,b}, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_j^{*,b}, \hat{\boldsymbol{\alpha}}^{*,b}, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.10}$$

where B is a large integer and $\hat{\boldsymbol{\alpha}}^{*,b}$ is the estimate of $\boldsymbol{\alpha}$ under the null. The critical value can be approximated by the upper quantile of $\{T_n^{*,b}, b = 1, \cdots, B\}$. The p-value is approximated by $B^{-1} \sum_{b=1}^{B} \mathbf{1}(T_n^{*,b} > T_n)$. In Section 3, we discuss the bootstrap method using several useful examples.

If the density $f(\boldsymbol{v}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ of $\boldsymbol{V}$ is given except for the unknown parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$, and bootstrap samples $\{\boldsymbol{V}_i^*, i = 1, \cdots, n\}$ are generated from the distribution $f(\boldsymbol{v}; \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}_0)$, then Theorem 3 provides the asymptotic consistency of the bootstrap distribution of the test statistic, where $\hat{\boldsymbol{\alpha}}$ is the estimate of $\boldsymbol{\alpha}$ under the null hypothesis.

Theorem 3. *If Assumptions (A1)–(A5) and (A7) in the Appendix hold, then under the null hypothesis, we have*

$$\sup_{x \in \mathbb{R}} |\mathbf{P}^*(nT_n^* \leq x) - \mathbf{P}(nT_n \leq x)| \xrightarrow{p} 0,$$

where $\xrightarrow{p}$ *denotes convergence in probability and* $\mathbf{P}^*(\cdot) = \mathbf{P}(\cdot | \{\boldsymbol{V}_i, i = 1, \cdots, n\})$ *denote the probability conditional on the observed data, that is, the probability measure induced by the bootstrap resampling procedure given the original sample* $\{\boldsymbol{V}_i, i = 1, \cdots, n\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

3. Choice of the Weight

In Section 2, we have established the asymptotic distributions of T_n under the null and alternative hypotheses in the GEE framework. Here, we discuss the closed-form of T_n by taking an appropriate weight. We also consider the bootstrap method for each example.

Example 1 (Continued). We first illustrate the choice of the weight function using the quantile regression model in Example 1 of Section 2.1. The same weighting scheme applies to the GLM and semiparametric models in Examples 2 and 3, with the corresponding expressions for ρ_{ij} given in Equations (3.4) and (3.5). By Lemma S1 in the Supplementary Material and taking the weight $w(\boldsymbol{\theta})$ as the multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance Σ, ρ_{ij} in the WAST (2.8) reduces to

$$\rho_{ij} = \omega_{ij} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_j \{\mathbf{1}(Y_i - \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \hat{\boldsymbol{\alpha}}(\tau) \le 0) - \tau\} \{\mathbf{1}(Y_j - \tilde{\boldsymbol{X}}_j^{\mathrm{T}} \hat{\boldsymbol{\alpha}}(\tau) \le 0) - \tau\}, \tag{3.1}$$

where for $i \neq j$, $\omega_{ij} = \int_{-\infty}^{\frac{\boldsymbol{Z}_j^{\mathrm{T}} \boldsymbol{\mu}}{\|\boldsymbol{Z}_j\|_\Sigma}} \phi(z) \Phi\left(\frac{1}{\sqrt{1-\varrho_{ij}^2}} \left(\frac{\boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\mu}}{\|\boldsymbol{Z}_i\|_\Sigma} - \varrho_{ij} z \right) \right) dz$, and $\varrho_{ij} = \boldsymbol{Z}_i^{\mathrm{T}} \Sigma \boldsymbol{Z}_j (\|\boldsymbol{Z}_i\|_\Sigma \|\boldsymbol{Z}_j\|_\Sigma)^{-1}$.

From a Bayesian perspective, $\boldsymbol{\mu}$ and Σ can be treated as priors and chosen empirically. Although there is no closed-form expression of ω_{ij} for $i \neq j$ in general, we can approximate ω_{ij} by the Monte Carlo method. More

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

specifically, we approximate ω_{ij} by

$$\omega_{ij} \approx \frac{1}{N}\sum_{k=1}^{N}\left\{\Phi\left(\frac{1}{\sqrt{1-\varrho_{ij}^2}}\left(\frac{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\mu}}{\|\boldsymbol{Z}_i\|_\Sigma}-\varrho_{ij}z_k\right)\right)\boldsymbol{I}\left(z_k \le \|\boldsymbol{Z}_j\|_\Sigma^{-1}\boldsymbol{Z}_j^{\mathrm{T}}\boldsymbol{\mu}\right)\right\}, \tag{3.2}$$

where z_k follows $\mathcal{N}(0,1)$ and N is a large integer. In practice, we can simply choose $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \boldsymbol{I}$, whereby Lemma S1 in the Supplementary Material implies that ω_{ij} reduces to

$$\omega_{ij} = \frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{\varrho_{ij}}{\sqrt{1-\varrho_{ij}^2}}\right) \quad \text{if } i \neq j. \tag{3.3}$$

If $\boldsymbol{\theta} \in [0,1]$, then we can take $w(\boldsymbol{\theta})$ as being the density of the beta distribution Beta(λ_1, λ_2) with shape parameters $\lambda_1 > 0$ and λ_2. In this case, we find that

$$\omega_{ij} = B(\min\{Z_i, Z_j\}; \lambda_1, \lambda_2),$$

where $B(\cdot; \lambda_1, \lambda_2)$ is the cumulative distribution function of Beta(λ_1, λ_2). If $\boldsymbol{\theta} \in \mathbb{R}$, then we can take $w(\boldsymbol{\theta}) = \phi(\boldsymbol{\theta}; \mu, \sigma^2)$ as the density of the Gaussian distribution with mean μ and variance σ^2, which implies that

$$\omega_{ij} = \Phi(\min\{Z_i, Z_j\}; \mu, \sigma^2),$$

where $\Phi(\cdot; \mu, \sigma^2)$ is the cumulative distribution function of $\mathcal{N}(\mu, \sigma^2)$. Alternatively, we can write $\check{\boldsymbol{\theta}} = (-\boldsymbol{\theta}, 1)^{\mathrm{T}}$ and $\check{\boldsymbol{Z}} = (1, Z)^{\mathrm{T}}$, which imply that

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

$\mathbf{1}(Z \geq \boldsymbol{\theta})$ is equal to $\mathbf{1}(\check{\boldsymbol{Z}}^{\mathrm{T}}\check{\boldsymbol{\theta}} \geq 0)$. This allows us to use the bivariate normal density as the weight, which results in ω_{ij} in (3.3).

We conduct numerical studies based on GLMs, quantile and semiparametric models to investigate the sensitivity of the weight's choice by comparing ω_{ij} with the closed-form in (3.3) and the approximated ω_{ij} in (3.2), see details in Supplementary S2.2. From the results, compared with the approximated ω_{ij} in (3.2), the test statistic with ω_{ij} in (3.3) has higher power uniformly, and only take 10% time computationally when $N = 10000$. This strongly recommends one to use the WAST with the closed-form in (3.3).

For the bootstrap procedure, as with the wild bootstrap of Feng et al. (2011) and Cheng et al. (2022), we independently generate a random variable ν satisfying the following conditions: (i) there are two positive constants c_1 and c_2 satisfying $\sup\{\nu \in \mathcal{G} : \nu \leq 0\} = -c_1$ and $\inf\{\nu \in \mathcal{G} : \nu \geq 0\} = c_2$, where $\mathcal{G}$ is the support of ν; (ii) the distribution G of ν satisfies $\int_0^\infty \nu^{-1} g(\nu) d\nu = \int_{-\infty}^0 \nu^{-1} g(\nu) d\nu = 1/2$ and $\mathrm{E}[\nu] < \infty$, where $g(\nu)$ is the density of ν; (iii) the τth quantile of ν is zero. We obtain the bootstrapped samples $\{(Y_i^*, \tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i), i = 1, \cdots, n\}$ and calculate T_n^* given in (2.10), where $Y_i^* = \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\alpha}}(\tau) + \nu_i |\tilde{r}_i|$ and $\tilde{r}_i = Y_i - \tilde{\boldsymbol{X}}_i \hat{\boldsymbol{\alpha}}(\tau)$, with $\hat{\boldsymbol{\alpha}}(\tau)$ being the estimate of $\boldsymbol{\alpha}(\tau)$ based on $\Psi_n(\boldsymbol{\alpha}) = \mathbf{0}$ under the null hypothesis. In practice, one can generate ν following a discrete distribution with

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

$\mathbf{P}(\nu = 2(1-\tau)) = 1-\tau$ and $\mathbf{P}(\nu = -2\tau) = \tau$.

Example 2 (Continued). ρ_{ij} in the test statistic becomes

$$\rho_{ij} = \omega_{ij} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{X}_j \{Y_i - \partial c(\tilde{\boldsymbol{X}}_i^{\mathrm{T}} \hat{\boldsymbol{\alpha}})/\partial \mu_i\}\{Y_j - \partial c(\tilde{\boldsymbol{X}}_j^{\mathrm{T}} \hat{\boldsymbol{\alpha}})/\partial \mu_j\}, \tag{3.4}$$

where the weight $w(\boldsymbol{\theta})$ can be chosen as in the case of quantile regression, yielding the form of ω_{ij} given in Equation (3.3). We generate the bootstrap samples from GLM with the density $f(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}_0)$, which is the density under the null hypothesis by replacing $\boldsymbol{\alpha}$ with $\hat{\boldsymbol{\alpha}}$. We obtain the bootstrap samples $\{(Y_i^*, \tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i), i = 1, \cdots, n\}$ and calculate T_n^* given in (2.10).

Example 3 (Continued). We have

$$\rho_{ij} = \omega_{ij}\{X_i - \pi(\boldsymbol{Z}_i, \hat{\boldsymbol{\alpha}}_1)\}\{Y_i - \gamma(\boldsymbol{Z}_i, \hat{\boldsymbol{\alpha}}_2)\}\{X_j - \pi(\boldsymbol{Z}_j, \hat{\boldsymbol{\alpha}}_1)\}\{Y_j - \gamma(\boldsymbol{Z}_j, \hat{\boldsymbol{\alpha}}_2)\}, \tag{3.5}$$

where ω_{ij} is the same as the term defined in the quantile regression of Example 1 (3.3). To calculate the critical value or p-value of T_n, we can use the wild bootstrap strategy to generate bootstrap samples (Wu, 1986; Liu, 1988; Mammen, 1993). More specifically, we obtain the bootstrap samples as $\{(Y_i^*, X_i, \boldsymbol{Z}_i), i = 1, \cdots, n\}$ with $Y_i^* = \gamma(\boldsymbol{Z}_i, \hat{\boldsymbol{\alpha}}_2) + \nu_i \tilde{r}_i$ and calculate T_n^* given in (2.10), where $\tilde{r}_i = Y_i - \gamma(\boldsymbol{Z}_i, \hat{\boldsymbol{\alpha}}_2)$ and ν_i is independent of $(Y_i, X_i, \boldsymbol{Z}_i)$ and normally distributed with zero mean and unit variance.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

4. Simulation Studies

In this section, we demonstrate the finite-sample performance of the proposed WAST. We report simulation results from various scenarios involving different numbers of $\boldsymbol{Z}$ and distributions of grouping variable $\boldsymbol{Z}$ for GLMs, quantile regression, and semiparametric models. Due to limit of space, we give simulation results of quantile regression and semiparametric models, as well as multiple change planes, in Appendices S2 and S3 of the Supplementary Material. As mentioned in Section 3, we choose Gaussian density with zero mean and covariance $\boldsymbol{I}$ as the weight in all simulations.

For fair comparison, we use same settings as considered by Huang et al. (2021). Consider the GLM in model (2.2) with the canonical parameter

$$\mu = \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} \geq 0),$$

where $\boldsymbol{\beta} = \kappa\mathbf{1}_p$ and $\alpha_2 = \cdots = \alpha_r = \log(1.4)$. For Gaussian and Poisson families, we set $\alpha_1 = 0.5$, and for binomial family α_1 is chosen so that the proportion of cases in the data set is $1/3$ on average under the null hypothesis. $\theta_2, \cdots, \theta_q$ are equally spaced numbers from -1 to 1, and θ_1 is chosen as the negative of the 0.65 percentile of $Z_2\theta_2 + \cdots + Z_q\theta_q$, which means that $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta}$ divides the population into two groups with 0.35 and 0.65 observations, respectively. For the predictor, we generate $(v_2, \cdots, v_{\max\{r,p\}}, Z_2, \cdots, Z_q)^{\mathrm{T}}$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\Sigma = (\tilde{\rho}_{ij})$, where $\tilde{\rho}_{ij} = 1$ if $i = j$ and $\tilde{\rho}_{ij} = \rho$ otherwise. Here, we consider both $\rho = 0$ and $\rho = 0.3$. We set $\widetilde{X}_j = \mathbf{1}(v_j > 0)$ with $j = 2, \cdots, r$, $X_k = \mathbf{1}(v_k > 0)$ with $k = 2, \cdots, p$, and $\widetilde{X}_1 = 1$, $X_1 = 1$, and $Z_1 = 1$.

We evaluate the power under various alternative models indexed by κ, i.e., $H_1^\kappa : \boldsymbol{\beta} = \kappa \mathbf{1}_p$ with κ equally spaced in the range $(0, \kappa_{\max}]$. We set the sample size $n = (300, 600)$, and we set 1000 repetitions and 1000 bootstrap samples. For comparison, we consider three methods here: (i) the proposed WAST, (ii) the SST, and (iii) the approximated supremum of likelihood-ratio test statistic (SLRT) given by Huang et al. (2021). We calculate both SLRT and SST over $\{\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}} : k = 1, \cdots, K\}$. Let $\boldsymbol{\theta}_{-1}^{(k)} = \tilde{\boldsymbol{\theta}}_{-1}^{(k)} / \|\tilde{\boldsymbol{\theta}}_{-1}^{(k)}\|$, where $\boldsymbol{\theta}_{-1}^{(k)} = (\theta_2^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)} = (\tilde{\theta}_2^{(k)}, \cdots, \tilde{\theta}_q^{(k)})^{\mathrm{T}}$, and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)}$ is drawn independently from the multivariate normal distribution $\mathcal{N}(\mathbf{0}_{q-1}, \mathbf{1}_{q-1})$. For each $\theta_1^{(k)}$, $k = 1, \cdots, K$, we select it by an equal grid search in the range from the lower tenth percentile to the upper tenth percentile of the data points of $\{\theta_2^{(k)} Z_{i2} + \cdots + \theta_q^{(k)} Z_{iq}\}_{i=1}^n$, which is the same as that in Huang et al. (2021). Here, we set $K = 1000$. And the critical value of the test statistic SST is calculated with the resampling method used in Fan et al. (2017). The results for each scenario are summarized based on 1000 repetitions.

Table 1 lists the type-I errors ($\kappa = 0$) with the significance level of 0.05

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Table 1: Type-I errors of the WAST, SST, and SLRT for Gaussian and binomial families. The nominal significant level is 0.05.

Family	(r,p,q)	$\rho=0$						$\rho=0.3$					
		$n=300$			$n=600$			$n=300$			$n=600$		
		WAST	SST	SLRT	WAST	SST	SLRT	WAST	SST	SLRT	WAST	SST	SLRT
Gaussian	$(2,2,3)$	0.045	0.033	0.048	0.052	0.036	0.049	0.046	0.032	0.049	0.048	0.044	0.046
	$(6,6,3)$	0.049	0.013	0.040	0.050	0.030	0.042	0.059	0.016	0.042	0.049	0.028	0.043
	$(2,2,11)$	0.044	0.030	0.058	0.060	0.041	0.059	0.049	0.020	0.051	0.052	0.031	0.047
	$(6,6,11)$	0.050	0.016	0.048	0.052	0.026	0.044	0.050	0.008	0.048	0.047	0.023	0.048
	$(2,51,11)$	0.045	0.032	0.044	0.048	0.000	0.059	0.044	0.034	0.050	0.048	0.000	0.045
	$(6,51,11)$	0.046	0.026	0.039	0.054	0.000	0.050	0.056	0.023	0.043	0.043	0.000	0.046
Binomial	$(2,2,3)$	0.053	0.109	0.052	0.050	0.082	0.054	0.055	0.100	0.048	0.051	0.072	0.059
	$(6,6,3)$	0.056	0.081	0.055	0.045	0.110	0.067	0.059	0.076	0.048	0.046	0.117	0.059
	$(2,2,11)$	0.046	0.124	0.053	0.057	0.089	0.065	0.052	0.106	0.048	0.050	0.099	0.057
	$(6,6,11)$	0.052	0.069	0.049	0.036	0.099	0.041	0.043	0.057	0.068	0.043	0.129	0.044
	$(2,51,11)$	0.051	0.035	0.054	0.046	0.019	0.010	0.043	0.046	0.044	0.049	0.011	0.008
	$(6,51,11)$	0.042	0.046	0.034	0.046	0.014	0.015	0.051	0.036	0.044	0.053	0.013	0.019

for Gaussian and binomial families. The results for Poisson family can be found in Supplementary S3.1. As can be seen, the sizes of the proposed WAST and the SLRT method are close to 0.05. For the SST, its sizes are far from 0.05 in most of the scenarios, and it performs the worst of these three methods. Table 1 reveals that the proposed WAST has a good control over the type-I error rate.

Figures 1–2 show powers for the scenarios considered in Table 1. As can be seen from Figures 1–2 and Figure S20, the powers for GLMs increase

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

as the sample size n increases, which verifies the asymptotic theory as expected. For the Gaussian, binomial and Poisson families, the power of the proposed WAST increases faster than those of the SST and SLRT, and this superiority is particularly evident when n is small or p and q increase. In summary, the proposed test statistic shows very competitive performance.

For large numbers of $\boldsymbol{Z}$, we analyze performance of the WAST in Supplementary S2.1, from which the WAST outperforms the competitors as scenarios with small numbers of grouping variables. Additionally, the simulation results for quantile regression and semiparametric models in Supplementary S3.3 further show the excellent performance of the WAST method.

The WAST achieves much cheaper computational cost by compared with two competitors, the SST and SLRT. On a computer with an Intel(R) Core(TM) i7-6700 CPU @3.40 GHz, when $(r, p, q) = (2, 2, 11)$ and $n = 600$, the WAST takes less than 0.045 seconds on average for 1000 repetitions with 1000 bootstrap samples, whereas the SLRT and SST take around 4.5 seconds on average for 1000 repetitions with 1000 bootstrap samples and 1000 threshold values ($K = 1000$). This indicates the computational time of the SST or SLRT is on average around 100 times that of the WAST.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

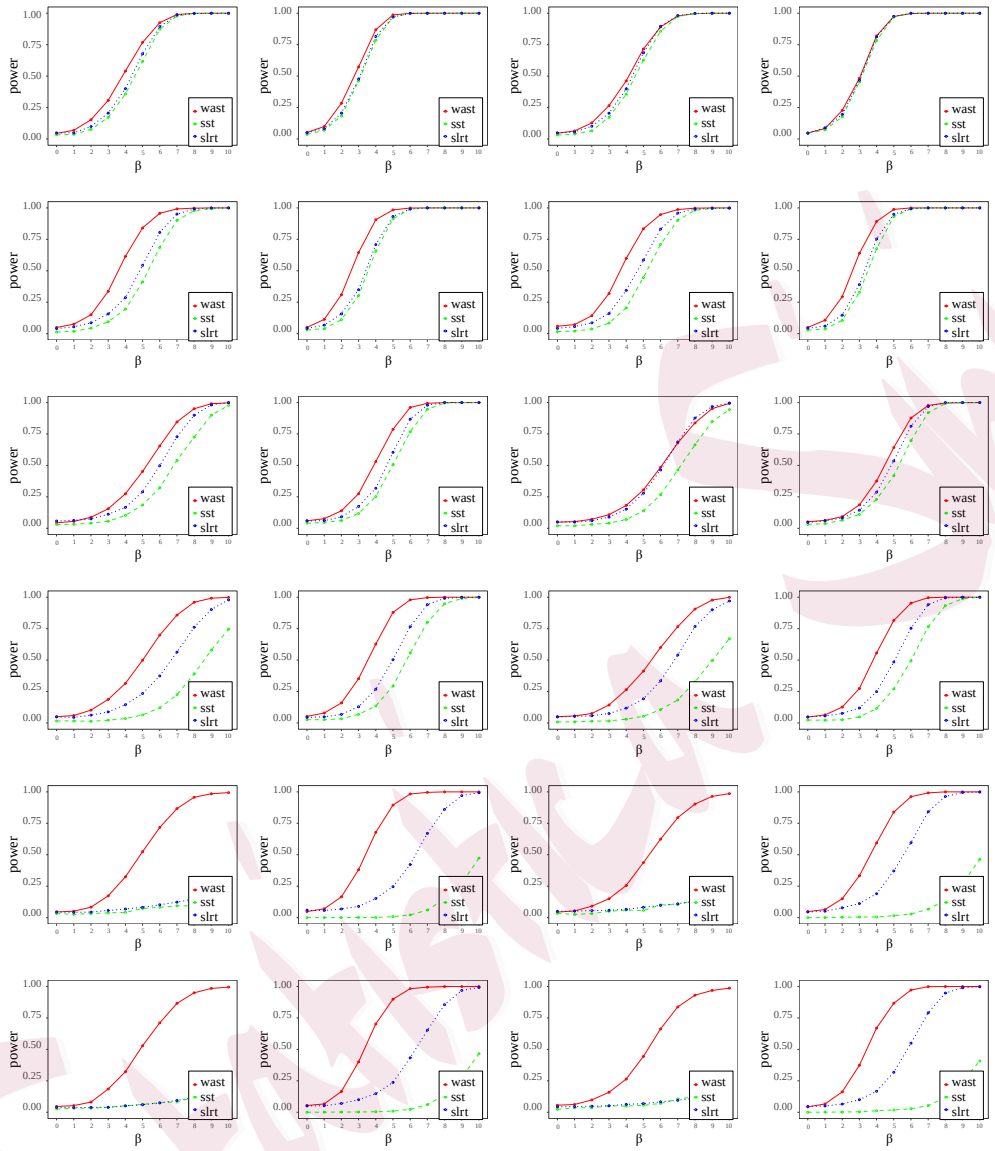

Figure 1: Powers of testing linear model with Gaussian error by WAST (solid line), SST (dashed line), and SLRT (dotted line). From top to bottom, each row depicts the powers for $(r, p, q) = (2, 2, 3)$, $(6, 6, 3)$, $(2, 2, 11)$, $(6, 6, 11)$, $(2, 51, 11)$, and $(6, 51, 11)$. From left to right, each column represents different combinations of sample size and correlation: $(n, \rho) = (300, 0)$, $(600, 0)$, $(300, 0.3)$, and $(600, 0.3)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

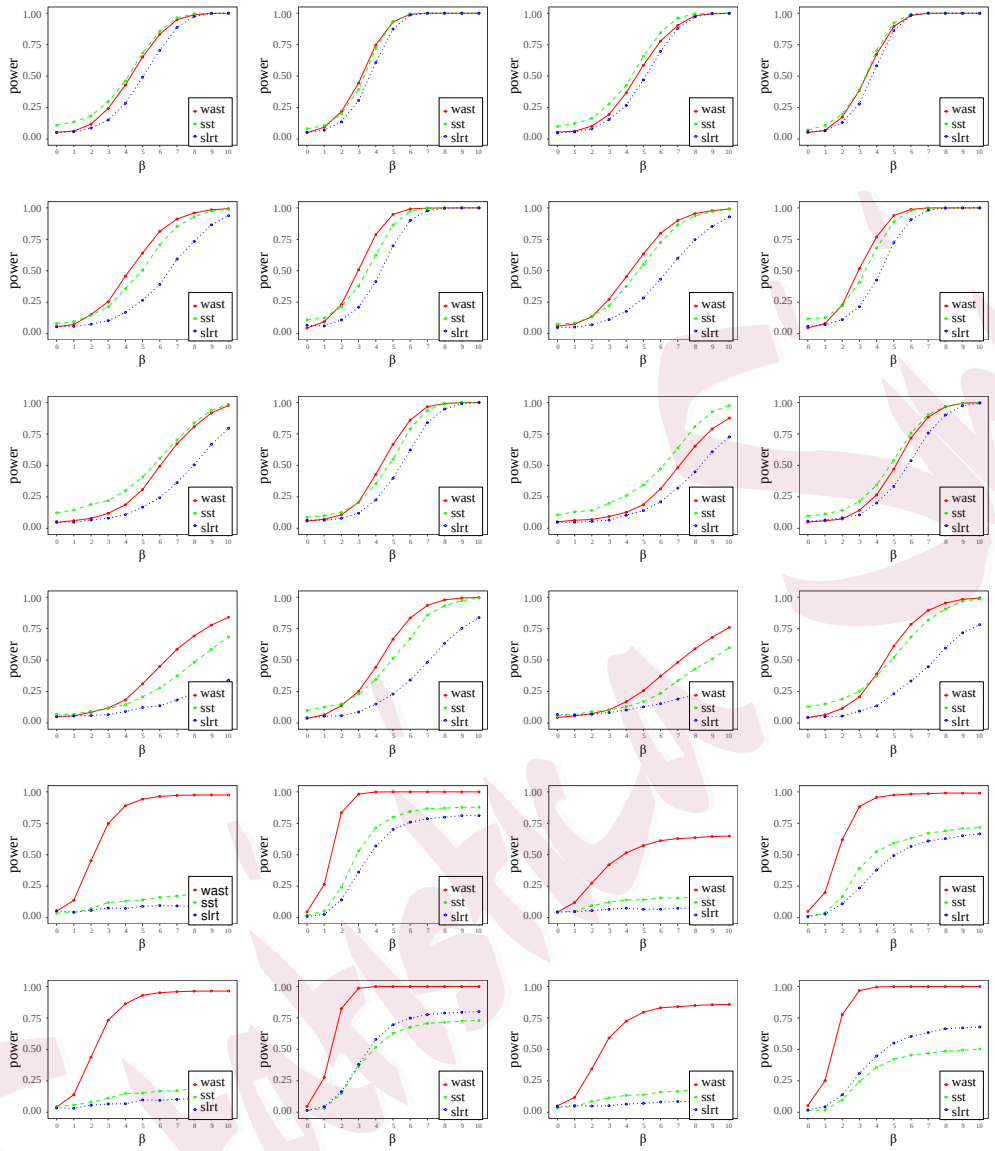

Figure 2: Powers of testing logistic regression by WAST (solid line), SST (dashed line), and SLRT (dotted line). From top to bottom, each row depicts the powers for $(r, p, q) = (2, 2, 3)$, $(6, 6, 3)$, $(2, 2, 11)$, $(6, 6, 11)$, $(2, 51, 11)$, and $(6, 51, 11)$. From left to right, each column represents different combinations of sample size and correlation: $(n, \rho) = (300, 0)$, $(600, 0)$, $(300, 0.3)$, and $(600, 0.3)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

5. Case Study

We show the proposed method in applications to prognostic data from Study 175 of the AIDS Clinical Trials Group and cancer type BRCA downloaded from the Cancer Genome Atlas at https://tcga-data.nci.nih.gov/. For the limit of space, we present analysis of the dataset from BRCA in Supplementary S4.

We apply the proposed test to the dataset from Study 175 of the AIDS Clinical Trials Group, which consists of 2139 subjects infected by the human immunodeficiency virus (HIV). All the subjects are partitioned randomly into four different treatment groups, i.e., zidovudine (ZDV) monotherapy, ZDV + didanosine (ddI), ZDV + zalcitabine (zal), and ddI monotherapy, which are coded by 0, 1, 2, and 3, respectively. Our purpose is to test for the existence of subgroups. As did in Li et al. (2021) and Fan et al. (2017), we treat the CD4 cell count (cells per cubic millimetre) at 20 ± 5 weeks post-baseline as the response variable. We consider the following ten covariates: (i) haemophilia (0, no; 1, yes), (ii) CD4 count at baseline, (iii) antiretroviral history (0, naive; 1, experienced), (iv) Karnofsky score, (v) CD8 count at baseline, (vi) race (0, white; 1, non-white) , (vii) symptomatic status (0, asymptomatic; 1, symptomatic), (viii) treatment arm (1, ZDV + ddI; 0, others), (ix) treatment arm (1, ZDV + zal; 0, others), and (x) treatment

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

arm (1, ddI; 0, others). These covariates are denoted by $X_1, \cdots, X_{10}$, respectively; covariates X_1, X_2, and X_3 are chosen to be threshold variables, and continuous covariates are normalized to have unit variance.

We consider the linear regression model with single change plane, i.e.,

$$Y_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0) + \epsilon_i, \quad i = 1, \cdots, 2139,$$

where $\boldsymbol{X} = (1, X_1, \cdots, X_{10})^{\mathrm{T}}$ and $\boldsymbol{Z} = (1, X_1, X_2, X_3)^{\mathrm{T}}$. As shown in Li et al. (2021) by fitting the linear regression model, gender, age, weight, homosexual activity, and history of intravenous drug use are not significant and so are excluded in our model.

The p-value for WAST, SST, and SLRT are 0.018, 0.0018, and < 0.001, respectively, based on 2000 candidate grouping parameters and 5000 bootstrap samples. Therefore, all three test methods give strong evidence for rejecting the null hypothesis. Table 2 lists the estimated coefficients for a single change plane, and the population is divided by $\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\theta}}$ into two groups with 227 individuals for $\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\theta}} < 0$ and 1912 for $\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \geq 0$.

We also consider the linear model with two planes as in Li et al. (2021),

$$Y_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_1\mathbf{1}(\boldsymbol{Z}_{-1,i}^{\mathrm{T}}\boldsymbol{\theta}_{-1} \geq a_1) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_2\mathbf{1}(\boldsymbol{Z}_{-1,i}^{\mathrm{T}}\boldsymbol{\theta}_{-1} \geq a_2) + \epsilon_i, \quad i = 1, \cdots, 2139,$$

where $\boldsymbol{Z}_{-1} = (X_1, X_2, X_3)^{\mathrm{T}}$ and $\boldsymbol{\theta}_{-1} = (\theta_2, \cdots, \theta_q)^{\mathrm{T}}$. The p-value of the WAST is 0.030, which indicates that there is strong evidence for rejecting

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

the null hypothesis. Table 2 lists the estimated coefficients of the proposed estimation procedure for multiple change planes, and the two threshold locations are $a_1 = -1.352$ and $a_2 = -1.321$. The population is divided by $\boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1}$ into three groups with 217 individuals for $\boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1} \leq -1.352$, 43 for $-1.517 < \boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1} \leq -1.321$, and 1879 for $\boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1} > -1.321$.

A Venn diagram is depicted in Figure 3, where g11 with 1912 subjects and g12 with 227 subjects denote two subgroups partitioned by single change plane, and g21 with 217 subjects, g22 with 43 subjects and g23 with 1879 subjects denote three subgroups partitioned by two change planes. From Figure 3, combining the partitions of a single change plane and multiple change planes, we find that there are 210 subjects who belong simultaneously in both the subgroup of a single change plane with $\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\theta}} < 0$ and the subgroup of multiple change planes with $\boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1} \leq -1.352$. 98% subjects in the subgroup of a single change plane with $\boldsymbol{Z}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \geq 0$ belong simultaneously in the subgroup of multiple change planes with $\boldsymbol{Z}_{-1}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{-1} \geq -1.321$.

6. Conclusion

Herein, we have considered a class of change-plane tests for heterogeneous estimating equations. A novel test statistic, WAST, is introduced to test whether the change plane exists. In comparison with existing test proce-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

Table 2: Estimates of parameter if null hypothesis has been rejected.

Change plane	Parameter	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Single	$\hat{\boldsymbol{\alpha}}$	0.710	-0.103	-0.126	0.285	0.147	0.054	-0.158	-0.176	0.057	0.033	0.197
	$\hat{\boldsymbol{\beta}}$	-0.630	0.041	0.023	0.464	0.061	-0.133	0.128	0.144	0.138	0.066	-0.098
	$\hat{\boldsymbol{\theta}}$	0.805	0.035	-0.025	-0.591							
Multiple	$\hat{\boldsymbol{\alpha}}$	0.861	0.093	-0.142	0.245	0.079	0.046	-0.157	-0.155	0.065	0.019	0.171
	$\hat{\boldsymbol{\beta}}_1$	1.592	-0.044	0.104	-0.866	-0.890	-0.026	0.351	0.145	0.480	0.439	0.302
	$\hat{\boldsymbol{\beta}}_2$	-2.368	-0.115	-0.070	1.351	1.040	-0.099	-0.231	-0.024	-0.361	-0.373	-0.380
	$\hat{\boldsymbol{\theta}}$		0.195	-0.082	-0.977							
	$\hat{a}_1$	-1.352										
	$\hat{a}_2$	-1.321										

Here, $X_0 = 1$ is the intercept, X_1 = haemophilia, X_2 = CD4 count at baseline, X_3 = antiretroviral history, X_4 = Karnofsky score, X_5 = CD8 count at baseline, X_6 = race, X_7 = symptomatic status, X_8 = treatment arm (1, ZDV + ddI; 0, others), X_9 = treatment arm (1, ZDV + zal; 0, others), and X_{10} = treatment arm (1, ddI; 0, others).

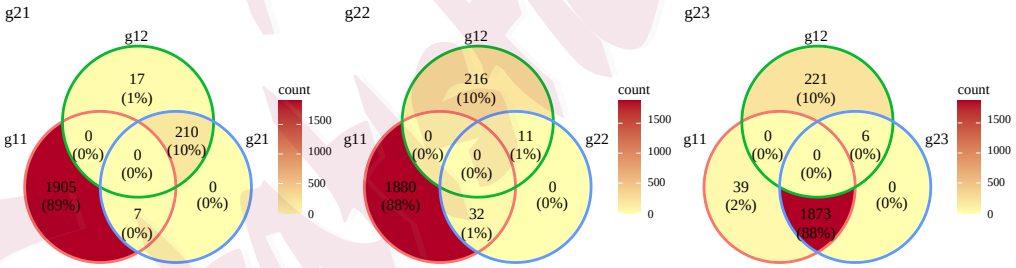

Figure 3: *Venn diagram for subgroup g2k with subgroups g11 and g12, where g2k is estimated from two change-plane model, and g1j is estimated from single change-plane model,* $k = 1, 2, 3$ *and* $j = 1, 2$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

dures, including the supremum of the squared score test from the framework of the average exponential LM test (Andrews, 1993; Andrews and Ploberger, 1994; Song et al., 2009; Fan et al., 2017), the proposed test statistic significantly improves the power in practice. As a by-product, the WAST has a closed form when an appropriate weight is used. Asymptotic distributions are derived under null and alternative hypotheses, and we extend the method to models with multiple change planes in the GEE framework, as studied by Li et al. (2021). Many classic and useful models in the GEE framework are considered herein, such as quantile regression (Zhang et al., 2022; Lee et al., 2018, 2011), GLMs (Huang et al., 2021), and semiparametric models (Fan et al., 2017). Case studies are analyzed to further demonstrate the performance of the proposed test procedure. From simulation and case studies, we conclude that compared with several competitors, the proposed WAST performs well across a wide range of settings.

We provide the hypothesis test for subgroup detection, but it is left blank how to partition subjects, particularly in the setting with large numbers of grouping variables, which needs the estimate of parameters, especially of the grouping parameter $\boldsymbol{\theta}$. It is a possible strategy for the estimation of $\boldsymbol{\theta}$ to penalize the loss function with $\|\boldsymbol{\theta}\|_1$ under sparsity assumptions. Further work could also extend the proposed method to allow

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

nonlinear change surfaces, such as those developed in Liu et al. (2025).

Supplementary material

Section S1 includes proofs of Theorems 1–3 and related Lemmas. Sections S2 and S3 present additional simulations to illustrate performance of the proposed statistic. Section S4 shows performance of the WAST by applying it to the BRCA dataset.

Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper. Liu's work is supported by the National Natural Science Foundation of China (12271329, 72331005), Guangxi Natural Science Foundation under Grant No. 2025GXNSFDA04240010, Program for Innovative Research Team of SUFE, and the Shanghai Research Center for Data Science and Decision Technology. Zhou's work is supported by the National Key R&D Program of China (2021YFA1000100, 2021YFA1000101), State Key Program of National Natural Science Foundation of China (72531003), Natural Science Foundation of Shanghai (23JS1400500), and Shanghai Municipal Education Commission (2024AI01002). Huang's work is supported

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

by the National Natural Science Foundation of China (72331005) and the research grants from The Hong Kong Polytechnic University (P0046811, P0042888, P0045417, P0045931).

Appendix: Assumptions for Establishing Asymptotic Properties

(A1) $\mathrm{E}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0) = 0$ for some $\boldsymbol{\alpha}_0 \in \Theta_\alpha$. $\mathrm{E}\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) \to \mathbf{0}$ implies that $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = o(1)$ for any sequence $\{\hat{\boldsymbol{\alpha}}\} \in \Theta_\alpha$.

(A2) The maps $\boldsymbol{\alpha} \mapsto \mathrm{E}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\boldsymbol{\alpha} \mapsto \mathrm{E}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha})$ are twice continuously differentiable for any $\boldsymbol{\beta} \in \Theta_\beta$ and $\boldsymbol{\theta} \in \Theta_\theta$. The map $\boldsymbol{\beta} \mapsto \mathrm{E}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})$ is continuously differentiable for any $\boldsymbol{\theta} \in \Theta_\theta$ and $\int_{\boldsymbol{\theta}\in\Theta_\theta} \partial \mathrm{E}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})/\partial\boldsymbol{\beta} w(\boldsymbol{\theta}) d\boldsymbol{\theta} = O(1)$.

(A3) $\max_{j=1,\cdots,p} \mathrm{E}[\psi^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta}) - \psi^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}', \mathbf{0}, \boldsymbol{\theta})]^2 \le L_n \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|^2$ and $\max_{j=1,\cdots,r} \mathrm{E}[\psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}) - \psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}')]^2 \le L_{1n} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|^2$ for any $\boldsymbol{\alpha} \in \Theta_\alpha$, $\boldsymbol{\alpha}' \in \Theta_\alpha$, and $\boldsymbol{\theta} \in \Theta_\theta$, where $\psi^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha})$ are the jth component of $\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ and $\psi_1(\boldsymbol{V}, \boldsymbol{\alpha})$, respectively.

(A4) The classes $\mathcal{F}^{(j)} = \{\boldsymbol{V} \mapsto \psi^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) : \boldsymbol{\alpha} \in \Theta_\alpha, \boldsymbol{\beta} \in \Theta_\beta, \boldsymbol{\theta} \in \Theta_\theta\}$ and $\mathcal{F}_1^{(j)} = \{\boldsymbol{V} \mapsto \psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Theta_\alpha\}$ are pointwise measurable and satisfy the uniform entropy condition $\max_{j=1,\cdots,p} \log \sup_Q N(\epsilon \|F_j\|_{Q,2}, \mathcal{F}^{(j)}, \|\cdot\|_{Q,2}) \le C_0 \log(e/\epsilon)$, where $F^{(j)} = \sup_{f\in\mathcal{F}^{(j)}} |f|$, $\epsilon \in (0, 1]$ and $C_0 \ge$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

1 is a constant. The classes $\mathcal{F}^{(j)}$ and $\mathcal{F}_1^{(j)}$ have measurable envelope functions $F_j \geq \sup_{f\in\mathcal{F}^{(j)}} |f|$ satisfying $\mathrm{E}F_j^s(\boldsymbol{V}) \leq C$ and $F_{1j} \geq \sup_{f\in\mathcal{F}_1^{(j)}} |f|$ satisfying $\mathrm{E}F_{1j}^s(\boldsymbol{V}) \leq C_1$, respectively, where $s \geq 4$.

(A5) $J = \left\{\partial \mathrm{E}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha}^{\mathrm{T}}\right\}^{-1} \in \mathbb{R}^{r\times r}$ is a finite and positive-definite deterministic matrix, and $\sup_{\boldsymbol{\theta}\in\Theta_\theta, \boldsymbol{\beta}\in\Theta_\beta} \|K(\boldsymbol{\beta}, \boldsymbol{\theta})\|_\infty \leq C_K$, $\|J\|_2 \leq C_c$, where $K(\boldsymbol{\beta}, \boldsymbol{\theta}) = \partial \mathrm{E}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})/\partial \boldsymbol{\alpha} \in \mathbb{R}^{p\times r}$ is continuous with respect to $\boldsymbol{\beta} \in \Theta_\beta$. Denote $K(\boldsymbol{\theta}) = K(\mathbf{0}, \boldsymbol{\theta})$. $\mathrm{E}\|\psi_1(\boldsymbol{V}, \boldsymbol{\alpha})\|^2$ and $\mathrm{E}\|\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})\|^2$ are bounded for any $\boldsymbol{\alpha} \in \mathcal{U}_{\alpha_0}$ and $\boldsymbol{\theta} \in \Theta_\theta$, where $\mathcal{U}_{\alpha_0}$ is a neighborhood of $\boldsymbol{\alpha}_0$.

(A6) There is a positive function $b(\boldsymbol{v}, \boldsymbol{\xi})$ of $\boldsymbol{v}$ relying on $\boldsymbol{\alpha}_0$, $\boldsymbol{\theta}_0$ such that

$$\left|\boldsymbol{\xi}^{\mathrm{T}} \frac{\partial f(\boldsymbol{v}; \boldsymbol{\alpha}_0, r_n\boldsymbol{\xi}, \boldsymbol{\theta}_0)/\partial \boldsymbol{\beta}}{f(\boldsymbol{v}; \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right| \leq b(\boldsymbol{v}, \boldsymbol{\xi}),$$

and $\mathrm{E}[b(\boldsymbol{V}, \boldsymbol{\xi})^2]$, $\lambda_{\max}\left(\mathrm{E}[b(\boldsymbol{V}, \boldsymbol{\xi})\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)^{\otimes 2}]\right)$, and $\lambda_{\max}\left(\mathrm{E}\left[b(\boldsymbol{V}, \boldsymbol{\xi})\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})^{\otimes 2}\right]\right)$ are bounded by $C_{\boldsymbol{\xi}}$, and for all k, $\mathrm{E}[\phi_k(\boldsymbol{V})^2 b(\boldsymbol{V}, \boldsymbol{\xi})]$ is bounded by $C_{\boldsymbol{\xi}}$, where $\boldsymbol{\xi} \in \Theta_\beta$, $r_n = o(1)$, $C_{\boldsymbol{\xi}} > 0$ is a constant relying on $\boldsymbol{\xi}$, $\phi_k(\cdot)$ is defined in Theorem 1, and $\boldsymbol{V}$ is generated from the null distribution with density $f(\boldsymbol{v}; \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$.

(A7) There is a positive function $b_1(\boldsymbol{v}, \boldsymbol{\xi}_1)$ of $\boldsymbol{v}$ relying on $\boldsymbol{\alpha}_0$, $\boldsymbol{\theta}_0$ such that

$$\left|\boldsymbol{\xi}_1^{\mathrm{T}} \frac{\partial f(\boldsymbol{v}; \boldsymbol{\alpha}_0 + r_n\boldsymbol{\xi}_1, \mathbf{0}, \boldsymbol{\theta}_0)/\partial \boldsymbol{\alpha}}{f(\boldsymbol{v}; \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right| \leq b_1(\boldsymbol{v}, \boldsymbol{\xi}_1),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

and $\mathrm{E}[b_1(\boldsymbol{V},\boldsymbol{\xi}_1)^2]$, $\lambda_{\max}\left(\mathrm{E}[b_1(\boldsymbol{V},\boldsymbol{\xi}_1)\psi_1(\boldsymbol{V},\boldsymbol{\alpha}_0)^{\otimes 2}]\right)$, and $\lambda_{\max}\left(\mathrm{E}\left[b_1(\boldsymbol{V},\boldsymbol{\xi})\psi(\boldsymbol{V},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta})^{\otimes 2}\right]\right)$ are bounded by $C_f(\boldsymbol{\xi}_1)$, and $\mathrm{E}[\phi_k(\boldsymbol{V})^2 b_1(\boldsymbol{V},\boldsymbol{\xi}_1)]$ and $\mathrm{E}[|\phi_k(\boldsymbol{V})\phi_j(\boldsymbol{V})|b_1(\boldsymbol{V},\boldsymbol{\xi}_1)]$ are bounded by $C_f(\boldsymbol{\xi}_1)$ for all j,k, where $\boldsymbol{\xi}_1 \in \Theta_\alpha$, $r_n = o(1)$, $C_f(\boldsymbol{\xi}_1) > 0$ is a constant relying on $\boldsymbol{\xi}_1$, $\phi_k(\cdot)$ is defined in Theorem 1, and $\boldsymbol{V}$ is generated from the null density $f(\boldsymbol{v};\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)$. Assume that, for $\boldsymbol{\xi}_2 \in \Theta_\alpha$ and any $\tilde{\boldsymbol{\alpha}} \in \Theta_\alpha$ satisfying $\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|^2 = O(n^{-1})$,

$$\mathrm{E}\left[\left\{\boldsymbol{\xi}_2^{\mathrm{T}}\frac{\partial h(\boldsymbol{V}_1,\boldsymbol{V}_2,\tilde{\boldsymbol{\alpha}})}{\partial\boldsymbol{\alpha}}\right\}^2\{1+b_1(\boldsymbol{V}_1,\boldsymbol{\xi}_1)\}\{1+b_1(\boldsymbol{V}_2,\boldsymbol{\xi}_1)\}\right] < \infty,$$

where $h(\boldsymbol{V}_1,\boldsymbol{V}_2,\boldsymbol{\alpha})$ equals $h(\boldsymbol{V}_1,\boldsymbol{V}_2)$ defined in (2.9) with $\boldsymbol{\alpha}_0$ replaced by $\boldsymbol{\alpha}$, and $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are independently generated from the null density $f(\boldsymbol{v},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)$.

Assumptions (A1)–(A5) are mild conditions for establishing the asymptotic properties of statistic $\tilde{T}_n$. Assumption (A1) is an identifiability condition for the GEE; see Theorem 2.10 of Kosorok (2008). Assumptions (A2) and (A3) are imposed for the asymptotic properties of the estimate of $\boldsymbol{\alpha}$ and the test statistic $\tilde{T}_n$. Assumption (A4) involves entropy conditions similar to those in Lee et al. (2011). Assumption (A5) is mild and easily verified in practice, and is used to derive the asymptotic distribution of the proposed statistic T_n. Assumptions (A6) and (A7) basically impose the Lipschitz condition, which is similar to the Lipschitz condition in Assumption (A2).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Andrews, D. W. K. and W. Ploberger (1995). Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative. *The Annals of Statistics 23*(5), 1609 – 1629.

Antoniou, A. C., S. Casadei, and et al. (2014). Breast-cancer risk in families with mutations in palb2. *New England Journal of Medicine 371*(6), 497–506.

Assmann, S. F., S. J. Pocock, L. E. Enos, and L. E. Kasten (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet 355*(9209), 1064–1069.

Cheng, C., X. Feng, J. Huang, and X. Liu (2022). Regularized projection score estimation of treatment effects in high-dimensional quantile regression. *Statistica Sinica 32*(1), 23–41.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: linear model case. *Biometrika 89*(2), 484–489.

Fan, A., S. Rui, and W. Lu (2017). Change-plane analysis for subgroup detection and sample

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

size calculation. *Journal of the American Statistical Association 112*(518), 769–778.

Fan, C., W. Lu, R. Song, and Y. Zhou (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology 79*(5), 1565–1582.

Feng, X., X. He, and J. Hu (2011). Wild bootstrap for quantile regression. *Biometrika 98*(4), 995–999.

Fu, Z. and Y. Hong (2019). A model-free consistent test for structural change in regression possibly with endogeneity. *Journal of Econometrics 211*(1), 206–242.

Fu, Z., Y. Hong, and X. Wang (2023). Testing for structural changes in large dimensional factor models via discrete fourier transform. *Journal of Econometrics 233*(1), 302–331.

Hu, B., Q. Wei, C. Zhou, M. Ju, L. Wang, L. Chen, Z. Li, M. Wei, M. He, and L. Zhao (2020). Analysis of immune subtypes based on immunogenomic profiling identifies prognostic signature for cutaneous melanoma. *International Immunopharmacology 89*, 107162.

Hu, X., J. Huang, L. Liu, D. Sun, and X. Zhao (2021). Subgroup analysis in the heterogeneous cox model. *Statistics in Medicine 40*(3), 739–757.

Huang, Y., J. Cho, and Y. Fong (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. *Journal of the Royal Statistical Society Series C: Applied Statistics 70*(2), 291–311.

Kang, S., W. Lu, and R. Song (2017). Subgroup detection and sample size calculation with

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

proportional hazards regression for survival data. *Statistics in Medicine 36*(29), 4646–4659.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Kosorok, M. R. and R. Song (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics 35*(3), 957 – 989.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2018). Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association 113*(523), 1184–1194.

Lee, S., M. H. Seo, and Y. Shin (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association 106*(493), 220–231.

Li, J., Y. Li, B. Jin, and M. R. Kosorok (2021). Multithreshold change plane model: Estimation theory and applications in subgroup identification. *Statistics in Medicine 40*(15), 3440–3459.

Liao, M., F. Zeng, Y. Li, Q. Gao, M. Yin, G. Deng, and X. Chen (2020, July). A novel predictive model incorporating immune-related gene signatures for overall survival in melanoma patients. *Scientific reports 10*(1), 12462.

Liu, P., Y. Li, and J. Li (2025). Change surface regression for nonlinear subgroup identification with application to warfarin pharmacogenomics data. *Biometrics 81*(1), ujae169.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

REFERENCES

Liu, R. Y. (1988). Bootstrap procedures under some non-I.I.D. models. *The Annals of Statistics 16*(4), 1696–1708.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics 21*(1), 255–285.

Shen, J. and A. Qu (2020). Subgroup analysis based on structured mixedeffects models for longitudinal data. *Journal of Biopharmaceutical Statistics 30*(4), 607–622.

Siegel, R. L., K. D. Miller, H. E. Fuchs, and A. Jemal (2021). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians 71*(1), 7–33.

Song, R., M. R. Kosorok, and J. P. Fine (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *The Annals of Statistics 37*(5A), 2409–2444.

Su, L., W. Lu, R. Song, and D. Huang (2020). Testing and estimation of social network dependence with time to event data. *Journal of the American Statistical Association 115*(530), 570–582.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society 54*(3), 426–482.

Wan, Q., J. Tang, H. Lu, L. Jin, Y. Su, S. Wang, Y. Cheng, Y. Liu, C. Li, and Z. Wang (2020). Six-gene-based prognostic model predicts overall survival in patients with uveal melanoma. *Cancer Biomark 27*(3), 343–356.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

REFERENCES

Wang, J., J. Li, Y. Li, and W. K. Wong (2019). A model-based multithreshold method for subgroup identification. *Statistics in medicine 38*(14), 2605–2631.

Wei, S. and M. R. Kosorok (2018, 10). The change-plane Cox model. *Biometrika 105*(4), 891–903.

Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics 14*(4), 1261–1295.

Wu, R.-f., M. Zheng, and W. Yu (2016). Subgroup analysis with time-to-event data under a logistic-cox mixture model. *Scandinavian Journal of Statistics 43*(3), 863–878.

Xiong, J., Z. Bing, and S. Guo (2019). Observed survival interval: A supplement to tcga pan-cancer clinical data resource. *Cancers 11*(3), 280.

Yu, P. and X. Fan (2020). Threshold regression with a threshold boundary. *Journal of Business and Economic Statistics 39*(4), 953–971.

Zhang, E., Y. Chen, S. Bao, X. Hou, J. Hu, O. Y. N. Mu, Y. Song, and L. Shan (2021). Identification of subgroups along the glycolysis-cholesterol synthesis axis and the development of an associated prognostic risk model. *Human genomics 15*(1), 1–20.

Zhang, Y., H. J. Wang, and Z. Zhu (2022). Single-index thresholding in quantile regression. *Journal of the American Statistical Association 117*(540), 2222–2237.

Zhao, L., L. Tian, T. Cai, B. Claggett, and L.-J. Wei (2013). Effectively selecting a target population for a future comparative study. *Journal of the American Statistical Associa-*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0155

tion 108(502), 527–539.

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

E-mail: liu.xu@mail.shufe.edu.cn

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong SAR, China

E-mail: j.huang@polyu.edu.hk

Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200062, China

E-mail: yzhou@amss.ac.cn

School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710049, China

E-mail: zhangfp@xjtu.edu.cn

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

E-mail: panpanren@stu.edu.cn