Statistica Sinica Preprint No: SS-2023-0231	
Title	Goodness of Fit Checking for Stukel Generalized Logistic Regression Models
Manuscript ID	SS-2023-0231
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0231
Complete List of Authors	Yu Xia, Xinmin Li, Feifei Chen and Hua Liang
Corresponding Authors	Hua Liang
E-mails	hliang@gwu.edu
Notice: Accepted version subject to English editing.	

Goodness of Test checking for Stukel generalized logistic regression models

Yu Xia[1], Xinmin Li[2], Feifei Chen[3], and Hua Liang[1]

[1] *George Washington University*

[2] *Qingdao University, Shandong 266071, China*

[3] *Beijing Normal University, Zhuhai 519087, China*

Abstract: As a useful generalization, Stukel generalized logistic models can fit binary response variable more flexibly and conveniently comparing to other generalizations. In this paper, we propose a projection-based test to check Stukel models. The test is shown to be consistent and can detect root-n local alternative hypotheses, and can be used to check the standard logistic models as well. We establish the asymptotic distribution of the proposed test under the null hypothesis and analyze asymptotic properties under the local and global alternatives. We evaluate the finite-sample performance via simulation studies and apply the proposed method to analyze a real dataset as an illustration. *Key words and phrases:* Consistent test; Generalized logistic models; Marked empirical process; Projection.

1. Introduction

The logistic regression model is often used when a binary response is to be regressed upon one or more explanatory variables. For example, this may occur when the response represents the survival or death of a patient, and the explanatory variable might be the characteristics of the individual or various treatment methods. The relationship between the response probability and dosage is often modeled using a logistic distribution function. Estimation and inference based on the maximum likelihood estimator (MLE) in logistic regression have been extensively studied in theory and widely used in social science, finance industry, and medical science (Hosmer & Lemeshow 1989, Lindsey 1997). See comprehensive introductions and techniques for fitting the logistic models in McCullagh & Nelder (1989), Nelder & Wedderburn (1972).

The standard logistic model assumes the response y with mean μ, which depends on the p explanatory variables $X = (x_1, \ldots, x_p)^\top$ in the form $\mu(\eta) = \exp(\eta)/\{1 + \exp(\eta)\}$, where $\eta = X^\top \boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is a $p-$vector of unknown parameters in $\mathbf{R}^p$. The model still has its limitations, though its popularity is irreproachable. As Stukel (1988) pointed out that the form of $\mu(\eta)$ restricts a symmetric pattern about $\eta = 0$. When data are not symmetric, or even symmetric but have a steeper or gentler incline in the central probability region, the logistic model may not fit the data well. Furthermore, the maximum likelihood estimation procedure weights each observation according to its estimated variance $\mu(\eta)\{1 - \mu(\eta)\}$ so that points in the central probability region near $\mu = 1/2$ have the strongest influence on the fit.

To release these limitations, several authors have generalized the standard logistic model to a more general parametric form. Prentice (1976) modeled the expected probability curve with the cumulative distribution function of the logarithm of a $F-$distribution. Pregibon (1980) defined a family of link functions as a solution of some implicit equation, including the logit link as a special case. Aranda-Ordaz (1981) introduced two separate one-parameter models for symmetric and asymmetric departures, respectively, from the logistic model. Guerrero & Johnson (1982) suggested a one-parameter Box-Cox transformation of the odds, $\tau^{-1}[\{\mu/(1-\mu)\}^{\tau}-1]=\eta$, which reduces to the logit transform in the limit when τ is 0. The parameters (except $\boldsymbol{\beta}$) for all of these models are shape parameters that generally influence the symmetry and heaviness of tails of the fitted curve $\mu(\eta)$. Except for specific values, they usually do not correspond to known distributions, although they all include the logistic model as a special case. All of the models require nonlinear estimation procedures to calculate the maximum likelihood estimates.

Stukel (1988) introduced a generalization of the standard logistic model, a class of models indexed by two shape parameters (α_1, α_2), to extend the scope of the standard logistic model to asymmetric probability curves and improve the fit in the noncentral probability regions; i.e., the logit(μ) was generalized to h_α defined in (2.1) and (2.2) below. The added parameters allow the tails of the logistic regression model to be heavier/longer or lighter/shorter than the standard logistic regression model. This two-parameter model presents a unified structure, to account simultaneously for symmetric and asymmetric departures from the standard logistic model. Besides containing provisions for lack of symmetry, the Stukel model is useful particularly when improvement in fit in the tail regions is important. If adding two parameters to the logistic model leads to overfitting, the model can be collapsed in a simple fashion to give one symmetric and three asymmetric one-parameter families, one of which may provide a more useful alternative and also additional information as to the source of lack of fit. Thus this model provides an extensive but flexible generalization of the standard logistic model. Among the aforementioned generalized logistic models, Stukel model has the broadest scope, followed by the Prentice model and then other one-parameter models. Hosmer et al. (1997) concluded that a comparison of the Prentice model to the Stukel model showed that both offer the same level of flexibility in terms of generating alternative models but that Stukel model is analytically easier to use since it does not require the integration needed with the Prentice model.

As far as we know, there is little work on developing goodness of fit to check Stukel generalized logistic model, though Stukel score test has widely been applied to check the standard logistic models. In this paper, we propose a marked empirical process based test for checking goodness of fit (GoF) in the Stukel generalized logistic model. The similar idea has been used by Xia (2006), Stute & Zhu (2002), Escanciano (2006).

This paper is organized as follows. Section 2 states the proposed statistic, the statistical properties of the proposed test under the null and alternative hypotheses, and a model based bootstrap procedure for calculating the critical value. Section 3 presents simulation studies to evaluate numerical performance of the proposed procedure. In the first simulation study, we specifically focus on the performance of several GoF tests for checking the standard logistic models. In the second simulation study, we evaluate the performance of the proposed procedure to check the Stukel models. Sections 4 analyzes a real dataset to illustrate the utilization of the proposed procedure. Section 5 discusses potential further work. Technical details are deferred to the Appendix.

2. Stukel Model and Proposed Test Statistic

2.1 Stukel Model

The form of the Stukel generalized logistic model (Stukel 1988) is:

$$P(Y=1|X) = \mu_\alpha(\eta) = \exp\left(h_\alpha(\eta)\right) / \left\{1 + \exp\left(h_\alpha(\eta)\right)\right\}$$

where $\eta = X^\top \boldsymbol{\beta}$, and $h_\alpha(\eta)$ is a strictly increasing nonlinear function of η indexed by two shape parameters, α_1 and α_2, defined as follows. For $\eta \geqslant 0$(i.e., $\mu \geqslant 1/2$),

$$h_\alpha = \begin{cases} \alpha_1^{-1}\left\{\exp\left(\alpha_1|\eta|\right) - 1\right\}, & \alpha_1 > 0 \\ \eta, & \alpha_1 = 0 \\ -\alpha_1^{-1}\log\left(1 - \alpha_1|\eta|\right), & \alpha_1 < 0 \end{cases} \tag{2.1}$$

and for $\eta \leqslant 0$(i.e., $\mu \leqslant 1/2$),

$$h_\alpha = \begin{cases} -\alpha_2^{-1}\left\{\exp\left(\alpha_2|\eta|\right)-1\right\}, & \alpha_2 > 0 \\ \eta, & \alpha_2 = 0 \\ \alpha_2^{-1}\log\left(1-\alpha_2|\eta|\right), & \alpha_2 < 0 \end{cases} . \tag{2.2}$$

The Stukel model indexed by (α_1, α_2) was called h family model (Stukel 1988). The subclass with $\alpha_1 = 0$ was called the $\alpha_2 h$ family model, and the subclass with $\alpha_2 = 0$ called the $\alpha_1 h$ family model. The one-parameter subclasses can be used to examine symmetric or asymmetric deviations from the standard logistic model. See Figure 1 for the patterns of $h_\alpha(\eta)$ versus η (left) and $\mu_\alpha(\eta)$ versus η with different (α_1, α_2) values.

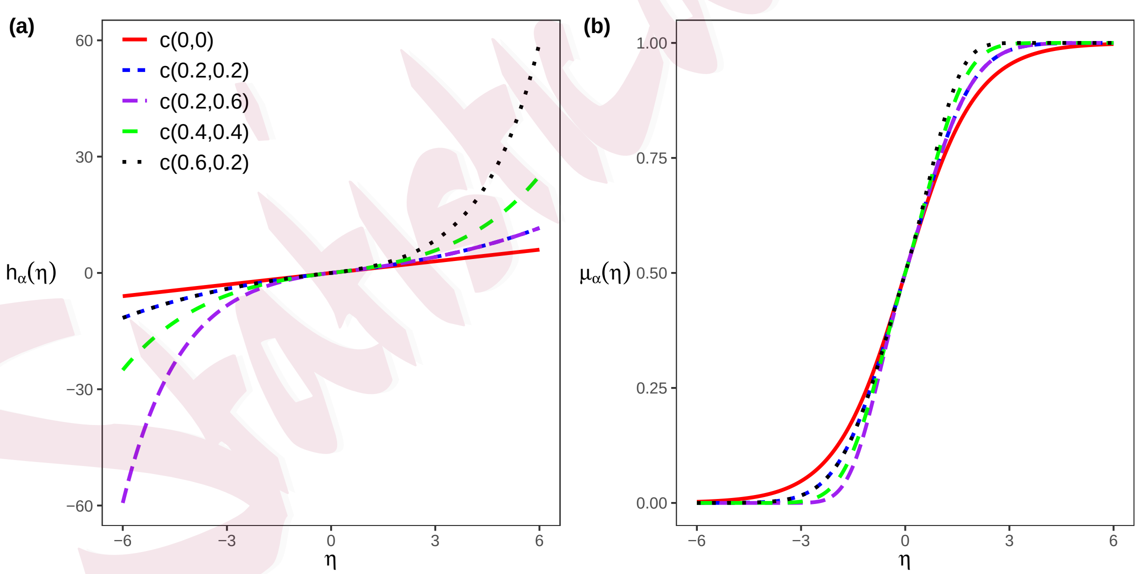

Figure 1: Plots of $h_\alpha(\eta)$ (left) and $\mu_\alpha(\eta)$ (right) against η for different (α_1, α_2).

Note that Stukel model can also be expressed as

$$\eta = h_\alpha^{-1}[\text{logit}\{P(Y=1|X)\}] = X^\top \boldsymbol{\beta}. \tag{2.3}$$

Write $\mu_\alpha(\eta) = P(Y=1|X) = g(h_\alpha(\eta)) = \{1+\exp(-h_\alpha(\eta))\}^{-1}$ and $\varepsilon(X,\boldsymbol{\beta}) = Y - g(h_\alpha(X^\top\boldsymbol{\beta}))$. Whether the mean function $g(h_\alpha(X^\top\boldsymbol{\beta}))$ can properly describe a generalized logistic regression relationship is equivalent to checking whether $\mathrm{E}\{\varepsilon(X,\boldsymbol{\beta})|X\} = 0$, which holds if and only if $\mathrm{E}\{\varepsilon(X,\boldsymbol{\beta})|X^\top W\} = 0$ for all unit vectors $W \in \mathbb{R}^p$. We are motivated by this fact to define our test statistic as follows.

Consider

$$\mathcal{H}_0 : \Pr\left(\mathrm{E}\{\varepsilon(X,\boldsymbol{\beta})|X\} = 0\right) = 1 \text{ for some } \boldsymbol{\beta} \tag{2.4}$$

against the alternative hypothesis:

$$\mathcal{H}_1 : \Pr\left(\mathrm{E}\{\varepsilon(X,\boldsymbol{\beta})|X\} = 0\right) < 1 \quad \text{for any } \boldsymbol{\beta} \in \mathbb{R}^p.$$

Let $\{(Y_i, X_i), i = 1, \cdots, n\}$ be a sample from (Y, X), then the log-likelihood function is

$$L(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^n \left[Y_i \log\left\{g(h_\alpha({X_i}^\top\boldsymbol{\beta}))\right\} + (1-Y_i)\log\left\{1 - g(h_\alpha({X_i}^\top\boldsymbol{\beta}))\right\}\right]. \tag{2.5}$$

The maximum likelihood estimators (MLE) $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\alpha}}_n$ are the solution to

$$L_n(\boldsymbol{\beta}) = \frac{\partial L(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n X_i^\top \{Y_i - g(h_\alpha({X_i}^\top\boldsymbol{\beta}))\}\frac{\partial h_\alpha(\eta_i)}{\partial \eta_i} = 0 \tag{2.6}$$

$$L_n(\boldsymbol{\alpha}) = \frac{\partial L(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^{n} \{Y_i - g(h_\alpha(X_i{}^\top \boldsymbol{\beta}))\} \frac{\partial h_\alpha(\eta_i)}{\partial \alpha} = 0 \tag{2.7}$$

Let $\widehat{\varepsilon}(X_i, \widehat{\beta}_n) = Y_i - g(h_{\widehat{\alpha}_n}(X_i, \widehat{\beta}_n))$. The sample version of $\mathrm{E}\{\varepsilon(X, \beta) I(X^\top W \le u)\}$ is $1/n \sum_{i=1}^n \widehat{\varepsilon}(X_i, \widehat{\boldsymbol{\beta}}_n) I(X_i^\top W \le u)$. Define

$$M_n(u, W) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\varepsilon}(X_i, \widehat{\boldsymbol{\beta}}_n) I(X_i^\top W \le u) \tag{2.8}$$

for $u \in \mathbb{R}^1$, where W is uniformly distributed on the unit sphere in $\mathbb{R}^p : \mathbb{S}^p = \{w \in \mathbb{R}^p : \|w\| = 1\}$.

Our test statistic is

$$T_n = \int_{-\infty}^{\infty} \int_{\mathbb{S}^p} \{M_n(u, W)\}^2 \, dw F_{nw}(du), \tag{2.9}$$

where $F_{nw}(u) = 1/n \sum_{i=1}^n I(X_i^\top W \le u)$ and w has been integrated out. When the test statistic is sufficiently large, the null hypothesis can be rejected. The estimated empirical process $M_n(u, W)$ is actually the cumulative sum of the estimated model error, and T_n is a Crámer-von Mises (CvM) statistic.

Note that the test statistic T_n can be formulated as a summation.

$$T_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \hat{\varepsilon}(X_i, \hat{\beta}_n) \hat{\varepsilon}(X_j, \hat{\beta}_n) A_{ijl}, \tag{2.10}$$

where $A_{ijl} = \int_{\mathbb{S}^p} I(X_i^\top w \le X_l^\top w) I(X_j^\top w \le X_l^\top w) dw$, which can be calculated as (Escanciano 2006):

$$A_{ijl} = C_q \left| \pi - \arccos \left\{ \frac{(X_i - X_l)^\top (X_j - X_l)}{\|X_i - X_l\| \cdot \|X_j - X_l\|} \right\} \right|,$$

here $C_q = \pi^{(p/2)-1}/\Gamma(p/2+1)$, $\Gamma(\cdot)$ is the gamma function. Therefore, T_n is equal to a summation multiplying a constant. This equivalence avoids multiple integrations and makes the computations much easier.

2.2 Statistical Properties

We now examine the statistical properties of the test statistic under the null hypothesis and the power performance under the alternative hypotheses.

Let $H(\cdot)$ be the derivative of $g(\cdot)$ to $\boldsymbol{\beta}$. For any function η, denote $\eta_i(\boldsymbol{\beta}) = \eta(X_i^\top \boldsymbol{\beta})$ and $\eta_i = \eta(X_i^\top \boldsymbol{\beta}_0)$ with $\boldsymbol{\beta}_0$ being the true value, for instance, $H_i(\boldsymbol{\beta}) = H(X_i^\top \boldsymbol{\beta})$, $g_i = g_i(\boldsymbol{\beta}_0)$. Denote $h'(\eta) = \frac{\partial h_\alpha(\eta)}{\partial \eta}$ and $h''(\eta) = \frac{\partial^2 h_\alpha(\eta)}{\partial \eta^2}$. Let $G_n(\boldsymbol{\beta}) = \sum_{i=1}^n X_i h_i' H_i(\boldsymbol{\beta}) h_i' X_i^\top$ and $S_n = \sum_{i=1}^n X_i X_i^\top$. Let $\mathbf{H} = \mathrm{diag}(h'(\eta_1) H_1 h'(\eta_1), \ldots, h'(\eta_n) H_n h'(\eta_n))$ and $\mathbb{X} = (X_1, \ldots, X_n)^\top$.

In what follows, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues for a symmetric matrix A respectively. $A_1 \geq A_2$ means $A_1 - A_2$ is semi-positive definite for two matrices A_1 and A_2. C is a generic constant with different values in different places.

The following assumptions are needed for the main result.

(a) There exist two positive constants b_{min} and B_{max} such that $b_{min} \leq h'(\eta_i) \leq B_{max}$.

(b) With probability 1, $\max_{i,j} |x_{ij}| < \infty$ and there exist two positive constants $c_{\min}$ and $C_{\max}$ such that $c_{\min} \leq \lambda_{\min}(S_n/n) \leq \lambda_{\max}(S_n/n) \leq C_{\max}$.

(c) There is a positive constant c_{00} such that

$$\min_{1\le i\le n} g(h_\alpha(X_i^\top \boldsymbol{\beta}_0))\{1-g(h_\alpha(X_i^\top \boldsymbol{\beta}_0))\} \ge c_{00}. \tag{2.11}$$

(2.11) indicates that, for any p-vector $\mathbf{v}$,

$$c_{00}b_{min}^2\mathbf{v}^\top S_n\mathbf{v} \le \mathbf{v}^\top \sum_{i=1}^n X_i h^{'}(\eta_i)H_i(\boldsymbol{\beta})h^{'}(\eta_i)X_i^\top \mathbf{v} \le 1/4B_{max}^2\mathbf{v}^\top S_n\mathbf{v}.$$

Assumption (a) ensures the probabilities lie between 0 and 1. The rest of the assumption bounds the eigenvalues of S_n in probability. This is a stability assumption to ensure S_n/n is not ill-conditioned.

Write $\Sigma_\beta = \mathrm{E}(\mathbb{X}^\top \mathbf{H}\mathbb{X})$, $\Gamma(u) = \mathrm{E}\{g'(X^\top \boldsymbol{\beta}_0)X^\top \Sigma_\beta^{-1} I(X^\top W \le u)\}$, $\widetilde{A}_n(\alpha_1,\alpha_2) = \int_0^1 A_n(\alpha_1+s(\alpha_2-\alpha_1))ds$, $A_n(\alpha) = -\partial L_n(\boldsymbol{\alpha})/\partial\alpha$, $\Theta(u) = E\{\Sigma_\alpha^{-1} I(X^\top W \le u)g'(\alpha_0)\}$, and $\Psi_u(X,\varepsilon) = \{I(X^\top W \le u) - \Gamma(u)Xh^{'} - \Theta(u)\frac{\partial h_\alpha}{\partial \alpha}\}\varepsilon$. We have the following result.

Theorem 2.1. *Suppose that Assumptions (a)–(c) hold. Under the null hypothesis (2.4), the estimated empirical process $M_n(u,W)$ converges in distribution to $M(u)$, $-\infty < u < \infty$, in the Skorohod space $S[-\infty,\infty]$, where $M(u)$ is a centered Gaussian process with covariance functions*

$$\mathrm{cov}\{M(u_1), M(u_2)\} = \mathrm{E}\{\Psi_{u_1}(X,\varepsilon)\Psi_{u_2}(X,\varepsilon)\}.$$

For the test statistic T_n, we have

$$T_n \xrightarrow{L} \int \{M(u)\}^2 F(du), \tag{2.12}$$

where F is the distribution of $X^\top W$.

It is shown that the estimated empirical process $M_n(u, W)$, $-\infty < u < \infty$, converges to a centered Gaussian process, and T_n converges to an integrated squared Gaussian process. The convergence theoretically demonstrates that the proposed test T_n has attractive asymptotic properties.

In the following, the power behavior of the statistic under the local and global alternatives are investigated. Consider the local alternative with a deviation in $h_\alpha(X^\top\beta)$ from the null hypothesis; i.e.,

$$\mathcal{H}_{1n} : \Pr\left\{E(Y|X) = g(h_\alpha(X^\top\beta + r_n D(X)))\right\} = 1 \tag{2.13}$$

where $D(X)$ is not a function of $X^\top\beta$ for any $\boldsymbol{\beta}$ and is a measurable function of X satisfying with $0 < \mathrm{E}D^2(X) < \infty$, r_n is a constant depending on n.

It is easily proved that, under the alternatives (2.13), $T_n \longrightarrow \infty$ if $r_n n^{1/2} \to \infty$, and is different from $\int\{M(u)\}^2 F_X(du)$ if $r_n n^{1/2}$ approaches a positive but finite constant. Consequently, the proposed statistic can detect a local alternative that approaches the null hypothetical model at the parametric rate, and the statistic T_n diverges to infinity under the global alternative hypothesis (2.13). Hence, it has asymptotic power 1 and is consistent.

2.3 A bootstrap option for critical value calculation

Though Theorem 2.1 gives the asymptotic distribution of the statistic T_n under the null hypothesis, this distribution may be case-dependent, which makes the calculation of the critical value inconvenient and sometimes difficult. To overcome this potential difficulty, a bootstrap procedure is used to determine the critical value.

Step 1. Calculate the test statistic T_n.

Step 2. Generate a sample $\{Y_i^*, i = 1, \ldots, n\}$ of independent Bernoulli random variables, where Y_i^* has the probability of success given by $\{g(h_{\hat{\alpha}_n}(X_i^\top \hat{\beta}_n)), i = 1, \ldots, n\}$.

Step 3. Calculate the bootstrap test statistic, denoted by T_n^*, based on the bootstrap sample $\{(Y_i^*, X_i), i = 1, \ldots, n\}$

Step 4. Repeat Steps 2 and 3 B times and obtain $T_{n1}^*, \ldots, T_{nB}^*$. Then calculate the $1 - \alpha$ empirical quantile of the bootstrap test statistics $\{T_{n1}^*, \ldots, T_{nB}^*\}$ as the α-level critical value.

This resampling strategy is proposed by Dikta et al. (2006). The numerical experience shows that the results based on the current resampling strategy are better than those based on the existing resampling strategy. For space concerns, we did not present these comparisons. As argued by Dikta et al. (2006) this model-based bootstrap strategy integrates the model assumption and is therefore expected to be more efficient than the wild bootstrap resampling method, which is confirmed by our numerical studies. When the outcome is binary, the residuals also have binary distributions, so the wild bootstrap's ability to accommodate an arbitrary residual distribution is not needed and, apparently, entails some cost.

Note that for the bootstrap procedure, it is not necessary to estimate any new quantities such as the influence function. Also, the testing procedure is data-driven; given only the sample, $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$, the proposed testing procedure using the bootstrap-generated critical value can determine whether the Stukel model fits the data adequately without any other information on the data generation process.

For the bootstrap testing statistic T_n^*, we have the following result.

Theorem 2.2. *Under the null hypothesis (2.4) or alternative hypothesis (2.13), if Assumption (a)–(c) are satisfied, the conditional distribution of* T_n^* *converges in distribution to the limiting null distribution of* T_n *given the sample* $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$.

This assertion indicates that T_n^* is not case-dependent, and gains the convenience of calculating the critical value. By generating i.i.d. random variable series $\{Y_i^*\}_{i=1}^n$ repeatedly, we can get a series of bootstrap test statistics, which can be looked upon as a sample coming from the population T_n. Then we can calculate the empirical quantile of the distribution of T_n. The critical value determined by this method approximates the theoretical one regardless of whether the data are from the null hypothetical model (2.4) or the alternative hypothetical model (2.13).

3. Simulation Studies

In this section, we assess the numerical performance of the proposed test. Note that when $\alpha_1 = \alpha_2 = 0$, Stukel model reduces to a standard logistic regression model. It means that the proposed test is applicable to GoF checking for logistic models. We conduct two simulation studies for the assessment purpose. In the first example, we focus on GoF checking for logistic models. In the second example, we evaluate the numerical performance of the proposed method for checking Stukel generalized logistic models in terms of type I error and power performance. Let $\mathcal{N}$ be the standard normal variable.

Example 1. The data were generated from the simple logistic model

$$P(Y = 1|X) = \{1 + \exp(-X\beta + cX^2)\}^{-1} \tag{3.1}$$

where $X \sim N(0, 1)$, $\beta = 2$, c ranges from 0 to 0.8 with increment 0.1. This is a simple standard logistic model with an additional term cX^2 in the linear predictor, and when $c = 0$ it reduces to the simple logistic model. Three scenarios with clusters $m = 3, 7$, and 10 were taken into account in this example. Within each group, 140 to 170 independent samples were generated. At each configuration, 1000 independent datasets were generated. One thousand bootstrap samples were generated from each of the 1000 datasets for calculating the empirical levels. The nominal levels of 0.05 and 0.10 were used. For the comparisons, the Hosmer-Lemeshow test T_n^{HL}, Osius and Rojek test T_n^{OR} (Osius & Rojek 1992), Stukel test T_n^{Stukel} (Stukel 1988), and the proposed test are evaluated T_n.

We calculated the proportions of times the null hypothesis was rejected among the 1000 replicates. They are the empirical sizes under the null hypothesis (i.e., $c = 0$) and the empirical powers under the alternative hypothesis (i.e., $c \neq 0$).

Intuitively, increasing the deviance cX^2 creates a larger distance away from the null hypothesis, resulting in an increased chance to reject the null hypothesis. We now report the rejection proportions of the tests in Figure 2. The performance of the proposed test is very promising in the sense that the empirical sizes are close to the nominal levels and the empirical powers increase with the distance away from the null hypothesis. The Osius and Rojek test is the most conservative one and the prerequisite of the test is the dataset should be large enough to maintain Type 1 error. The proposed test is comparable to Stukel's test. When sample sizes are comparatively low, for example, in the scenario with $m = 3$, the proposed test is significantly better than Stukel's test. Both the proposed test and Stukel's test work better than the Hosmer-Lemeshow test. When the deviance is small, the proposed test works better in distinguishing the difference from the null hypothesis than the Hosmer-Lemeshow test.

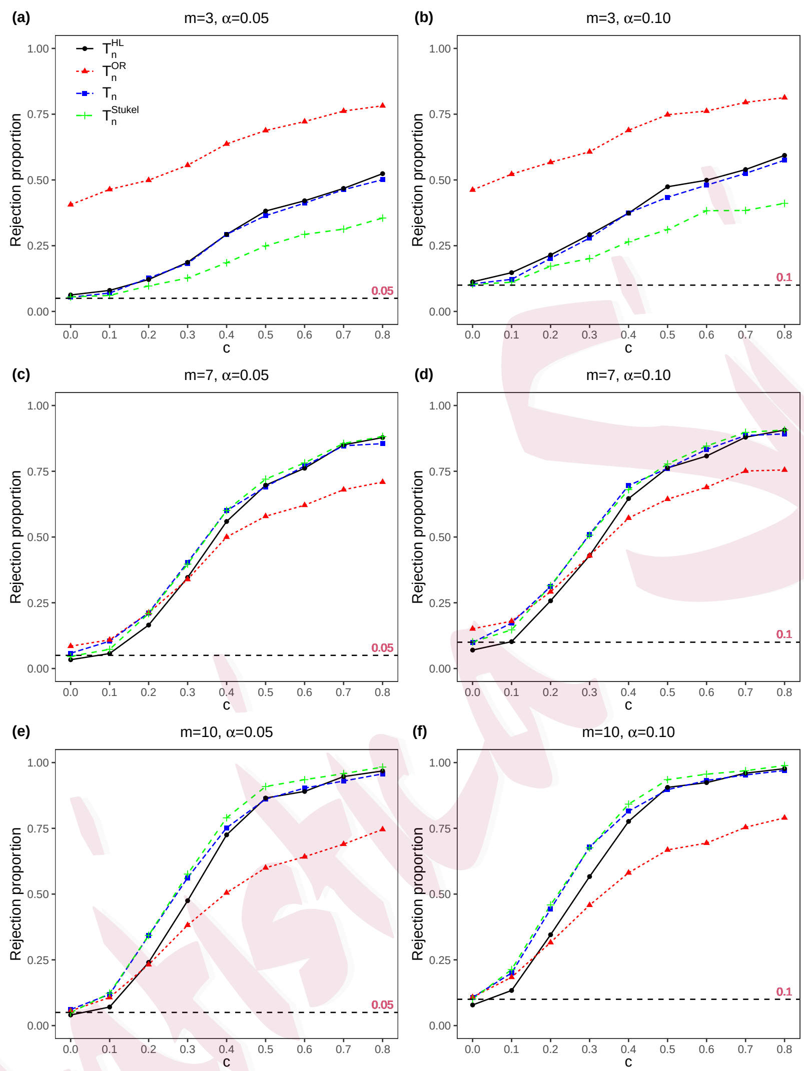

Figure 2: Simulation results for Example 1: Rejection proportion of the four tests: Hosmer-Lemeshow test (T_n^{HL}), Osius and Rojek test (T_n^{OR}), Stukel test (T_n^{Stukel}), and the proposed test (T_n), with the nominal levels 0.05 and 0.10 for different cluster cases.

Example 2. We extend the application to the generalized logistic model

$$
\begin{aligned}
P(Y=1|X) &= \mu_\alpha(X\beta - cX^2) \\
&= \exp\left(h_\alpha(X\beta - cX^2)\right) / \left(1+\exp\left(h_\alpha(X\beta - cX^2)\right)\right)
\end{aligned}
\tag{3.2}
$$

There are $m = 20$ different clusters contained in each dataset, where Xs are 20 equally spaced numbers from -1.5 to 1.5, $\beta = 2$. This is a generalized logistic model with an additional term $-cX^2$ in the linear predictor, and when $c = 0$ it reduces to Stukel's generalized logistic model. The nominal levels of 0.05 and 0.10 were used.

Several scenarios were explored in this example. We conducted the simulations among the h family model with four different sets of parameters $(\alpha_1, \alpha_2) = c(0.2, 0.2)$, $c(0.2, 0.6)$, $c(0.4, 0.4)$, and $c(0.6, 0.2)$; $\alpha_2 h$ family model with four different parameters: $\alpha_2 = 0$, 0.2, 0.4 , and 0.6; and h family model with given, fixed parameters $(\alpha_1, \alpha_2) = c(0.2, 0.2)$, $c(0.2, 0.6)$, $c(0.4, 0.4)$, and $c(0.6, 0.2)$. Within each group, 140 to 2000 independent samples were generated. At each configuration, 200 independent datasets ($N = 200$) were generated. 200 bootstrap ($B = 200$) samples were generated from each of the 200 datasets for calculating the empirical levels. c ranges from 0 to 1.4 with increment 0.2.

In Figure 3, on the left panel, we estimated the additional parameters (α_1, α_2) first, and the rejection proportion gets to increase to 1 when $c = 0.4$. The testing performance works better with the setting $c(0.6, 0.2)$, especially, when compared to the curve with $c(0.2, 0.6)$, which can be explained by Figure 1. Overall, this pattern

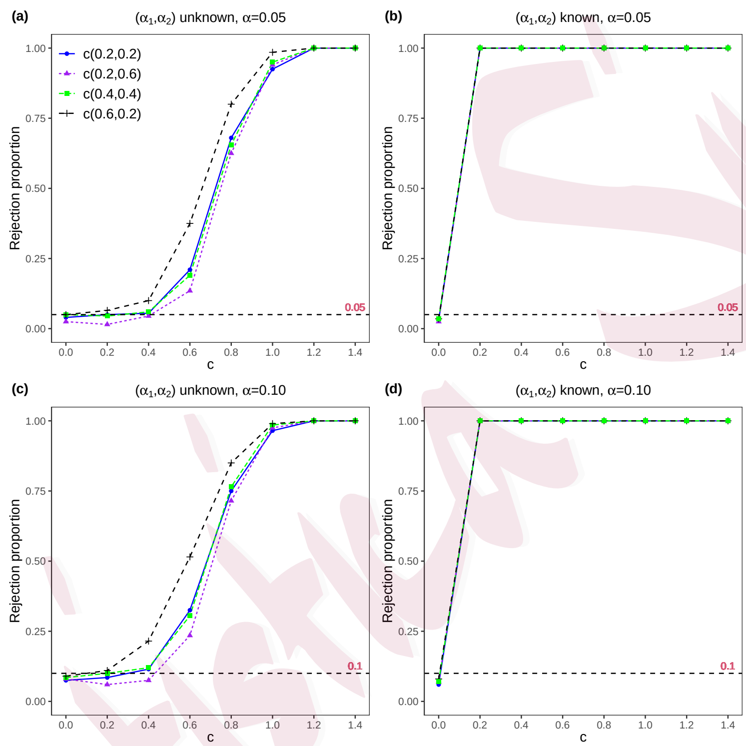

Figure 3: Simulation results for Example 2: Rejection proportion of the proposed test, left panel with (α_1, α_2) unknown, and right panel with (α_1, α_2) known.

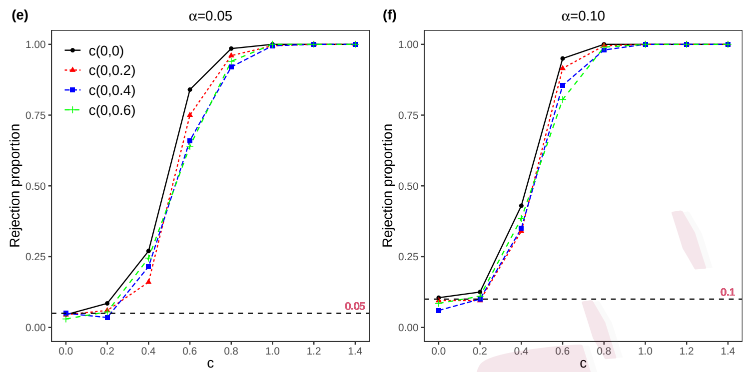

Figure 4: Simulation results for Example 2: Rejection proportion of the proposed test for the $\alpha_2 h$ family, with the nominal levels 0.05 and 0.10.

confirmed that the proposed test maintains the type 1 error at null hypothesis well and achieves good power performance within a reasonable deviation. The right panel is similar to what we observed in Example 1. However, instead of taking $c(0,0)$, i.e., the standard logistic regression, we treated the parameters (α_1, α_2) as given information. As a result, the rejection proportion approaches 100% at $c = 0.2$. Figure 4 depicts the performance of the proposed test on $\alpha_2 h$ family. With only one additional shape parameter to be estimated, the rejection proportion gets to increase to 1 when $c = 0.2$.

The second scenario we explored in this example is to compare the difference in performance between balanced data and unbalanced data. We considered two settings: the balanced data with 200 independent samples within each group, and the unbalance data with 50 to 400 samples within each group. 20 clusters in total, $N = 200$, and $B = 500$. From Figure 5, no significant difference between these two scenarios, which indicates that the loss of balance won't affect the performance of the proposed test. However, the rejection proportion curves for the balanced data are smoother than the crooked curves for unbalanced data. We expect to enlarge the bootstrap times B and replication cases N to achieve a smoother and more reliable curve with unbalanced data.

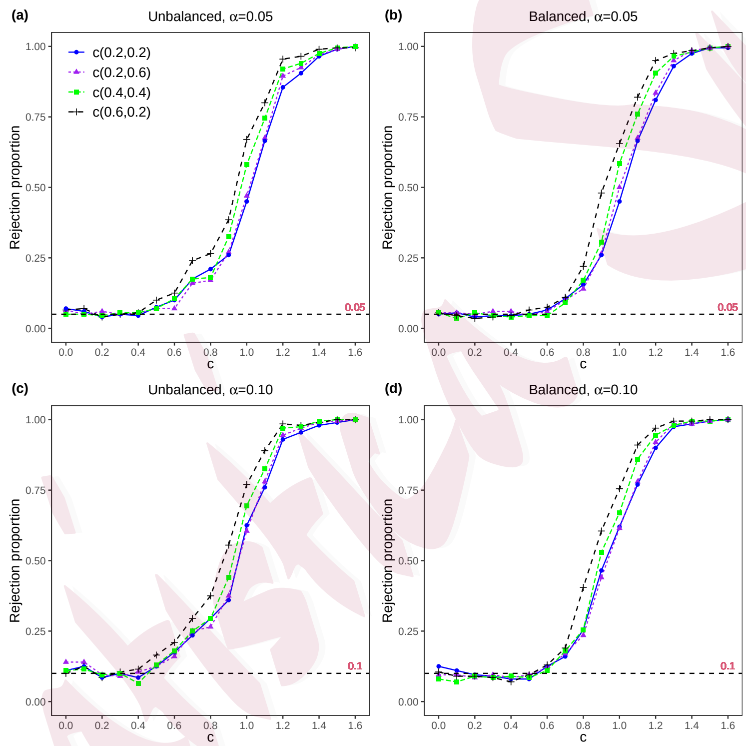

Figure 5: Simulation results for Example 2: Rejection proportion of the proposed test for balanced data and unbalanced data, with the nominal levels 0.05 and 0.10.

4. Real Data Example

In this example we analyzed a dataset determining the age of menarche in a sample of $3,918$ Warsaw girls with more details in Milicer & Szczotka (1966), and analyzed by Stukel (1988) for model checking, and referred to as the Warsaw study. This dataset is available in R package `MASS` and contains 25 different age groups. By fitting a standard logistic model, we obtained the intercept $\hat{\beta}_0 = -21.226$ and the slope of age $\hat{\beta}_1 = 1.632$ with residual deviance 26.7. An examination of the residuals shows that possible improvements to the fit could be made on the tails (Stukel 1988). We first applied the Stukel test the following three hypotheses.

(i) H_0: a standard logistic model $\Longleftrightarrow$ H_a: h family

(ii) H_0: a standard logistic model $\Longleftrightarrow$ H_a: $\alpha_1 h$ family

(iii) H_0: a standard logistic model $\Longleftrightarrow$ H_a: $\alpha_2 h$ family

The results for the tests of shape of parameters are presented in Table 1.

For hypothesis (i), i.e., whether the two shape parameters equal zero, the statistic values based on the Stukel likelihood-ratio and score tests are 12.117 and 7.938 with p-values 0.002 and 0.019, respectively. For hypothesis (ii), i.e., whether the shape parameter α_2 equals zero, the statistic values based on the Stukel likelihood-ratio and score tests are 0.824 and 0.706 with p-values 0.364 and 0.401, respectively. For hypothesis (iii), i.e., whether the shape parameter α_1 equals zero, the statistic values

Table 1: Tests for Shape Parameters on Warsaw Age at Menarche Data: Likelihood Ratio Test and Score Test.

Hypothesis	Stukel Likelihood-Ratio Test		Stukel Score Test	
	Test Statistic	p-value	Test Statistic	p-value
(i)	12.117	0.002	7.938	0.019
(ii)	0.824	0.364	0.706	0.401
(iii)	10.338	0.001	7.232	0.007

based on the Stukel likelihood-ratio and score tests are 10.338 and 7.232 with p-values 0.001 and 0.007, respectively. These results indicate that we do not have enough evidence to reject $\alpha_2 = 0$, but have strong evidence to reject $\alpha_1 = 0$.

We further applied the proposed GoF procedure for the Stukel model to check h family and $\alpha_2 h$ family. That is, we consider the following two hull hypotheses.

(Gi) H_0: h family

(Gii) H_0: $\alpha_2 h$ family

One thousand bootstrap samples were generated to calculate the empirical levels. As shown in Table 2, the proposed test with $B = 1000$ shows that 73.8% of the test statistics T_n^* are higher than the observed test statistic $T_n = 0.015$ for the hypothesis (Gi); 43.1% of the test statistics T_n^* are higher than the observed test statistic $T_n =$

Table 2: Goodness-of-Fit on Warsaw Age at Menarche Data.

Hypothesis	Model	Goodness-of-Fit	
		Test Statistics	p-value
(Gi)	h family	0.015	0.738
(Gii)	$\alpha_2 h$ family	0.049	0.431

0.049 for the hypothesis (Gii). These results indicate that we have no evidence to reject h family and $\alpha_2 h$ family. Since $\alpha_2 h$ family is a subfamily of h family, we choose the former as our final model. The parameter estimates and variance matrix based on the $\alpha_2 h$ family model are:

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} -18.824 \\ 1.457 \\ 0.219 \end{pmatrix}, \mathbf{V}_{\hat{\beta},\hat{\alpha}_2} = \begin{pmatrix} 0.979 & -0.073 & 0.063 \\ -0.073 & 0.005 & -0.005 \\ 0.063 & -0.005 & 0.007 \end{pmatrix}.$$

It is equivalent to the two-parameter model in terms of fitted values, except that the standard errors were much smaller. This result is consistent with the finding in Milicer & Szczotka (1966) that the age of menarche distribution was slightly positively skewed.

As a further illustration, we replaced the linear term in the standard logistic model by a nonparametric term $m(age)$, where $m(age)$ is an unknown but smooth function. This flexible replacement can avoid potential misspecifications. $m(age)$ was estimated by a data-driven strategy and obtained using the R function `gam`. We refit this particular semiparametric logistic regression and display the estimated $m(age)$ (red broken line) and the associated 95% confidence band (shaded) in Figure 6. Intuitively speaking, if a parametric model is close to the truth, the corresponding parametric curve should close to $m(age)$, i.e., covered by the reference band. If the parametric curve is far away from the reference band, we can guarantee that the candidate model is inappropriate. We also insert the estimated straight line (black solid line) based on the standard logistic model and the $\alpha_2 h$ fam-

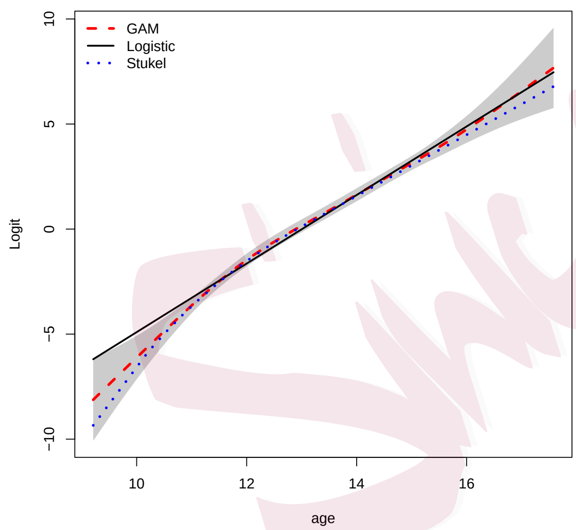

Figure 6: The results for the Warsaw Example. The estimated curve of $m(age)$ (red broken line) and the associated pointwise confidence band based on the semiparametric logistic model; The logit function based on the standard logistic model (straight solid line), and logit function based on the $\alpha_2 h$ family (blue dotted line).

ily curve (blue dotted line) based on the Stukel $\alpha_2 h$ family. It is interesting to notice that the $\alpha_2 h$ curve is closer to $m(age)$ in the left tail and is encapsuled in the confidence band, while the logit (straight) line is closer to $m(age)$ in the right tail, yet the straight line is almost but not completely covered by the confidence band. This may also illustrate why we reject the standard logistic model with a weak evidence, yet accept the $\alpha_2 h$ family. These analyses also illustrate the superiority of the proposed test.

5. Discussion

This is the first attempt to develop a test for assessing the Stukel generalized logistic model. The test is shown to be consistent and can detect parametric rate local alternative hypotheses. The test is easy to implement, theoretically reliable, and has good finite-sample performance. In implementing the proposed procedure, calculating the critical value based the bootstrap procedure is the most computationally intensive.

We focus on the Stukel's generalized logistic models, but the proposed method can also be used for other generalized logistic model such as the one-parameter generalization of the logistic model mentioned in Hosmer & Hjort (2002):

$$\Pr(Y = 1 \mid X) = \left\{ \frac{\exp(X^\top \boldsymbol{\beta})}{1 + \exp(X^\top \boldsymbol{\beta})} \right\}^{1+\gamma} \tag{5.1}$$

The model in equation (5.1) is stochastically smaller than the logistic model when $\gamma > 0$ and is stochastically larger when $\gamma < 0$. To our knowledge this transformation has never been used in logistic regression to assess overall model adequacy.

Per the simulation experience, the proposed method works well for the case of univariate x. But the precision of the estimation gets worse when introducing more covariates into the model. As a result, the proposed method may also become worse. Additional efforts are needed for improving estimation of the parameters. In this paper, we have discussed the performance of the proposed test with balanced and unbalanced observations. If most of the observations gather at the tails, the estimation of α may be out of a reasonable range and impact the reliability of the test in the end. Generally speak, if estimation of the parameters works well, the proposed procedure should perform satisfactorily.

Appendix

We first present several lemmas and their proofs if needed. These lemmas establish bounds for several terms, which will appear in the proofs of the main results, and a representation for $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$, which we need later. The proofs of Lemmas A.3-A.5 are similar to but more sophisticated than the proofs in Yin et al. (2006), Liang & Du (2012), Li et al. (2023). Below C denotes a generic constant which may varies but is independent of n.

A.1 Several Lemmas

Lemma A.1. *(Heuser 1981)* If F is continuously differentiable on $\mathbb{R}$, then

$$F(t_2) - F(t_1) = (t_2 - t_1) \int_0^1 \frac{dF}{ds}|_{s=t_1+u(t_2-t_1)} du,$$

where $t_1, t_2 \in \mathbb{R}$. □

Lemma A.2. *(Chen et al. 1999)* Let Υ be a smooth injection from $\mathbb{R}^p$ to $\mathbb{R}^p$ with $\Upsilon(\mathbf{x}_0) = \mathbf{y}_0$ and $\inf_{\|\mathbf{x}-\mathbf{x}_0\|=\delta} \|\Upsilon(\mathbf{x}) - \mathbf{y}_0\| \geq R$. Then for any $\mathbf{y}$ with $\|\mathbf{y} - \mathbf{y}_0\| \leq R$, there is an $\mathbf{x}$ with $\|\mathbf{x} - \mathbf{x}_0\| \leq \delta$ such that $\Upsilon(\mathbf{x}) = \mathbf{y}$. □

Let $N_n(\delta) = \{\boldsymbol{\beta} : \|G_n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \leq \delta\}$, and $\mathbf{v}$ below be a unit p-vector. The following lemma establishes bounds for several terms, which will appear in the proof of the main results. The bounds or inequalities hold in the sense of "in probability" unless specified otherwise.

Lemma A.3. Under the conditions of Theorem 2.1, we have

$$\max_{1\leq i\leq n} \|G_n^{-1/2} X_i\|^2 = O(n^{-1}), \tag{A.1}$$

$$\max_{1\leq i\leq n} |X_i^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| = O(n^{-1/2}\delta). \tag{A.2}$$

$$\sum_{i=1}^n |\mathbf{v}^\top G_n^{-1/2} X_i|^2 \leq C \tag{A.3}$$

for $\boldsymbol{\beta} \in N_n(\delta)$.

Proof. Note that

$$\max_{1\leq i\leq n} \|G_n^{-1/2} X_i\|^2 = \max_{1\leq i\leq n} X_i^\top G_n^{-1} X_i$$

$$
\begin{aligned}
&= \max_{1\le i\le n} X_i^\top (\mathbb{X}^\top \mathbf{H}\mathbb{X})^{-1} X_i \\
&\le \max_{1\le i\le n} X_i^\top \lambda_{\min}^{-1}(\mathbf{H})(\mathbb{X}^\top \mathbb{X})^{-1} X_i \\
&\le c_{00}^{-1} b_{min}^{-2} c_{\min}^{-1} n^{-1} \max_{1\le i\le n} X_i^\top X_i \\
&= O(n^{-1}),
\end{aligned}
$$

and

$$
\begin{aligned}
\max_{1\le i\le n} |X_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)| &\le \max_{1\le i\le n} \|G_n^{-1/2} X_i\| \|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \\
&= O(n^{-1/2}\delta).
\end{aligned}
$$

In addition,

$$
\begin{aligned}
\sum_{i=1}^n |\mathbf{v}^\top G_n^{-1/2} X_i|^2 &= \sum_{i=1}^n \mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v} \\
&= \mathbf{v}^\top G_n^{-1/2} S_n G_n^{-1/2}\mathbf{v} \\
&= \mathbf{v}^\top (\mathbb{X}^\top \mathbf{H}\mathbb{X})^{-1/2} S_n (\mathbb{X}^\top \mathbf{H}\mathbb{X})^{-1/2}\mathbf{v} \\
&\le c_{00}^{-1} b_{min}^{-2} \mathbf{v}^\top (\mathbb{X}^\top \mathbb{X})^{-1/2} S_n (\mathbb{X}^\top \mathbb{X})^{-1/2}\mathbf{v} = c_{00}^{-1} b_{min}^{-2}.
\end{aligned}
$$

□

Lemma A.4. Under the conditions of Theorem 2.1, we have

$$
\sup_{\boldsymbol{\beta}\in N_n(\delta)} |\mathbf{v}^\top G_n^{-1/2} Q_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{v} - 1| \to 0, \tag{A.4}
$$

where $Q_n(\boldsymbol{\beta}) = -\partial L_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^\top$.

Proof. Recall $\varepsilon_i = Y_i - g_i = Y_i - g(h_\alpha(X_i^\top \boldsymbol{\beta}_0))$ and $\sigma_i^2 = \mathrm{var}(\varepsilon_i)$. A direct calculation yields

$$\mathbf{v}^\top G_n^{-1/2} Q_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{v} - 1 = A_n(\boldsymbol{\beta}) - B_n - C_n(\boldsymbol{\beta}) - D_n(\boldsymbol{\beta}), \tag{A.5}$$

where

$$A_n(\boldsymbol{\beta}) = \mathbf{v}^\top G_n^{-1/2} G_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{v} - 1,$$
$$B_n = \sum_{i=1}^n \mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v} h_i^{''}\varepsilon_i,$$
$$C_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v}\{h_i^{''}(\boldsymbol{\beta}) - h_i^{''}\}\varepsilon_i,$$
$$D_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v} h_i^{''}(\boldsymbol{\beta})\{g_i - g_i(\boldsymbol{\beta})\}.$$

We will show that each of these four terms approaches to zero on $N_n(\delta)$.

Note that

$$\begin{aligned}
|A_n(\boldsymbol{\beta})| &= \mathbf{v}^\top G_n^{-1/2}\{G_n(\boldsymbol{\beta}) - G_n(\boldsymbol{\beta}_0)\} G_n^{-1/2}\mathbf{v} \\
&= \mathbf{v}^\top G_n^{-1/2} \sum_{i=1}^n X_i H_i^{(1)}(\widetilde{\boldsymbol{\beta}}) X_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0) X_i^\top G_n^{-1/2}\mathbf{v} \\
&\le \mathbf{v}^\top G_n^{-1/2} \sum_{i=1}^n X_i C_1 |X_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)| X_i^\top G_n^{-1/2}\mathbf{v} \\
&\le C_1 \max_{1\le i\le n} |X_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \sum_{i=1}^n |\mathbf{v}^\top G_n^{-1/2} X_i|^2.
\end{aligned}$$

Then (A.2) and (A.3) indicate that $\sup_{\boldsymbol{\beta}\in N_n(\delta)} \|A_n(\boldsymbol{\beta})\| \to 0$.

A direct calculation yields

$$\mathrm{var}(B_n) = \sum_{i=1}^n (\mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v})^2 (h_i^{''})^2 \sigma_i^2$$

$$\leq \sum_{i=1}^{n} |\mathbf{v}^\top G_n^{-1/2} X_i|^2 |X_i^\top G_n^{-1/2}\mathbf{v}|^2 C_2$$

$$\leq \max_{1\leq i\leq n} |\mathbf{v}^\top G_n^{-1/2} X_i|^2 \sum_{i=1}^{n} |X_i^\top G_n^{-1/2}\mathbf{v}|^2 C_2 \to 0. \tag{A.6}$$

This along with the fact that $E(B_n) = 0$ indicates $|B_n| \to 0$.

Noting that $\sup \mathrm{E}\,|\varepsilon_i| < \infty$, with the analogous arguments as in $A_n(\boldsymbol{\beta})$, $\sup_{\boldsymbol{\beta}\in N_n(\delta)} \|h_i^{''}(\boldsymbol{\beta}) - h_i^{''}\| \to 0$, and

$$\mathrm{E} \max_{\boldsymbol{\beta}\in N_n(\delta)} \|C_n(\boldsymbol{\beta})\| \leqslant \sup_{\boldsymbol{\beta}\in N_n(\delta)} \left\|h_i^{''}(\boldsymbol{\beta}) - h_i^{''}\right\| \sup \mathrm{E}|\varepsilon_i| \sum_{i=1}^{n} \left|\mathbf{v}^\top G_n^{-1/2} X_i\right|^2 \to 0,$$

which implies that

$$\max_{\boldsymbol{\beta}\in N_n(\delta)} \|C_n(\boldsymbol{\beta})\| \xrightarrow{p} 0.$$

Similarly, noting that $\sup_{\boldsymbol{\beta}\in N_n(\delta)} \left\|h_i^{''}(\boldsymbol{\beta})\right\| < \infty$, we obtain

$$\sup_{\boldsymbol{\beta}\in N_n(\delta)} \|D_n(\boldsymbol{\beta})\| \leq C_3 \max_{1\leq i\leq n} |X_i^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)| \sum_{i=1}^{n} |\mathbf{v}^\top G_n^{-1/2} X_i|^2 \to 0.$$

□

If $\alpha_1 = \alpha_2 = 0$, the right-hand side of A.5 simplifies to $A_n(\boldsymbol{\beta})$.

Lemma A.5. Suppose Assumptions (a)-(b) hold. Then there exist a sequence of random variables $\widehat{\boldsymbol{\beta}}_n$ such that

$$P\{L_n(\widehat{\boldsymbol{\beta}}_n) = 0\} \to 1 \tag{A.7}$$

and

$$\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 = \{\widetilde{Q}_n(\widehat{\boldsymbol{\beta}}_n)\}^{-1} L_n(\boldsymbol{\beta}_0), \tag{A.8}$$

where $\widetilde{Q}_n(\boldsymbol{\beta}) = \int_0^1 Q_n(\boldsymbol{\beta}_0 + s(\boldsymbol{\beta} - \boldsymbol{\beta}_0))ds$.

(A.7) indicates that with probability tending to 1, there exists a solution of the equation $L_n(\boldsymbol{\beta}) = 0$, while (A.8) gives a representative of $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$.

Proof. We first show

$$\mathbf{v}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0,1). \tag{A.9}$$

Let $\xi_i = \mathbf{v}^\top G_n^{-1/2} X_i h_i' \varepsilon_i$. It is easy to verify that $E(\xi_i) = 0$. It now suffices to prove that (Lindeberg's condition), for any $\zeta > 0$, as $n \to \infty$,

$$g_n(\zeta) = \sum_{i=1}^n E\{|\xi_i|^2 I_{(|\xi_i|>\zeta)}\} \to 0. \tag{A.10}$$

Let $a_{ni} = \mathbf{v}^\top G_n^{-1/2} X_i h_i'$. Similar to (A.3), we can show that

$$\max_{1\le i\le n} \|a_{ni}\|^2 \le \max_{1\le i\le n} B_{max}^2 \mathbf{v}^\top G_n^{-1/2} X_i X_i^\top G_n^{-1/2}\mathbf{v} = \max_{1\le i\le n} B_{max}^2 \|\mathbf{v}^\top G_n^{-1/2} X_i\|^2 \to 0.$$

Also, (A.3) showed that $\sum_{i=1}^n \|a_{ni}\|^2$ is bounded. Combining these with the Cauchy-Schwartz inequality, and (2.11) ensures (A.10). The central limiting theorem then yields (A.9).

By Lemma A.1, we have

$$L_n(\boldsymbol{\beta}) - L_n(\boldsymbol{\beta}_0) = -\widetilde{Q}_n(\boldsymbol{\beta})(\boldsymbol{\beta} - \boldsymbol{\beta}_0). \tag{A.11}$$

Furthermore, it follows from Lemma A.4 that

$$\sup_{\boldsymbol{\beta}\in N_n(\delta)} |\mathbf{v}^\top G_n^{-1/2} \widetilde{Q}_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{v} - 1| \to 0 \tag{A.12}$$

and

$$\sup_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\in N_n(\delta)} |\mathbf{v}^\top G_n^{-1/2}\widetilde{Q}_n(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)G_n^{-1/2}\mathbf{v}-1| \to 0, \tag{A.13}$$

where $\widetilde{Q}_n(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\int_0^1 Q_n(\boldsymbol{\beta}_1+s(\boldsymbol{\beta}_2-\boldsymbol{\beta}_1))ds$. Next, we prove that for any $\zeta>0$ there is a $\delta>0$ such that when n is large enough

$$P\{\text{there is } \widehat{\boldsymbol{\beta}}_n\in N_n(\delta) \text{ such that } L_n(\widehat{\boldsymbol{\beta}}_n)=0\}>1-\zeta. \tag{A.14}$$

Write $\partial N_n(\delta)=\{\boldsymbol{\beta}:\|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|=\delta\}$. Note that $\|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|/\delta=1$ for $\boldsymbol{\beta}\in\partial N_n(\delta)$. By the Cauchy-Schwartz inequality, we have that for any $\delta>0$,

$$\begin{aligned}\inf_{\boldsymbol{\beta}\in\partial N_n(\delta)}&(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\widetilde{Q}_n(\boldsymbol{\beta})^\top G_n^{-1}\widetilde{Q}_n(\boldsymbol{\beta})(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\ &\geq \inf_{\boldsymbol{\beta}\in\partial N_n(\delta)}\delta^2\{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top\widetilde{Q}_n(\boldsymbol{\beta})^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)/\delta^2\}^2.\end{aligned} \tag{A.15}$$

It follows from (A.12) that, for any $\epsilon>0$ and $\delta>0$, there is a $c_0\in(0,1)$ independent of δ, such that

$$P\left\{\inf_{\|\mathbf{e}\|=1,\boldsymbol{\beta}\in\partial N_n(\delta)}\mathbf{e}^\top G_n^{-1/2}\widetilde{Q}_n(\boldsymbol{\beta})^\top G_n^{-1/2}\mathbf{e}\geq c_0\right\}>1-\frac{\epsilon}{4}. \tag{A.16}$$

(A.11), (A.15), and (A.16) indicate that, for any $\delta>0$ such that

$$P\left\{\inf_{\boldsymbol{\beta}\in\partial N_n(\delta)}|\mathbf{v}^\top G_n^{-1/2}\{L_n(\boldsymbol{\beta})-L_n(\boldsymbol{\beta}_0)\}|\geq c_0\delta\right\}>1-\frac{\epsilon}{4}. \tag{A.17}$$

Taking $\delta=(4/\epsilon)^{1/2}/c_0$ and using the Markov inequality and (A.9) yield

$$P\{|\mathbf{v}^\top G_n^{1/2}L_n(\boldsymbol{\beta}_0)|\leq c_0\delta\}\geq 1-E|\mathbf{v}^\top G_n^{-1/2}L_n(\boldsymbol{\beta}_0)|^2/(c_0\delta)^2$$

$$\geq 1 - 1/(c_0\delta)^2 = 1 - \frac{\epsilon}{4}. \tag{A.18}$$

Write $E_n = \left\{|\mathbf{v}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0)| \leq \inf_{\boldsymbol{\beta}\in\partial N_n(\delta)} |\mathbf{v}^\top G_n^{-1/2}\{L_n(\boldsymbol{\beta}) - L_n(\boldsymbol{\beta}_0)\}|\right\}$. (A.17) and (A.18) imply that

$$P(E_n) > 1 - \frac{\epsilon}{2}. \tag{A.19}$$

Write $\widetilde{E}_n = \left\{\det\{\widetilde{Q}_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\} \neq 0 \text{ for all } \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in N_n(\delta)\right\}$. Then (A.13) indicates that

$$P(\widetilde{E}_n) > 1 - \frac{\epsilon}{2}. \tag{A.20}$$

Lemma A.1 indicates that the map: $\boldsymbol{\beta} \to \mathbf{v}^\top G_n^{-1/2} L_n(\boldsymbol{\beta})$ is an injection for $\boldsymbol{\beta} \in N_n(\delta)$ on the set $\widetilde{E}_n$. Using Lemma A.2 we know that, on $E_n \cap \widetilde{E}_n$, there is a $\widehat{\boldsymbol{\beta}}_n$ such that

$$\widehat{\boldsymbol{\beta}}_n \in N_n(\delta) \quad \text{and } L_n(\widehat{\boldsymbol{\beta}}_n) = 0. \tag{A.21}$$

(A.14) follows from (A.19)-(A.21). Then (A.7) holds. □

A.2 Proof of Theorem 2.1

By the definition of $M_n(u, W)$, we have

$$\begin{aligned} M_n(u, W) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i - g(h_{\hat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n))\} I(X_i^\top W \leq u) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i - g(h_{\alpha_0}(X_i^\top \boldsymbol{\beta}_0))\} I(X_i^\top W \leq u) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(h_{\hat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)) - g(h_{\hat{\alpha}}(X_i^\top \boldsymbol{\beta}_0))\} I(X_i^\top W \leq u) \end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{g(h_{\hat{\alpha}}(X_i^\top \boldsymbol{\beta}_0)) - g(h_{\alpha_0}(X_i^\top \boldsymbol{\beta}_0))\} I(X_i^\top W \le u) \\
:= \; & B_{n1}(u,W) - B_{n2}(u,W) - B_{n3}(u,W). \qquad (A.22)
\end{aligned}
$$

It follows from model (2.4) that

$$
B_{n1}(u,W) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i I(X_i^\top W \le u). \qquad (A.23)
$$

Using Lemma A.1, the second term can be expressed as

$$
B_{n2}(u,W) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i^\top(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) I(X_i^\top W \le u)\int_0^1 g'(X_i^\top \boldsymbol{\beta}_0 + s(X_i^\top(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)))ds.
$$

Substituting $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$ given in (A.8) and applying the mean-value theorem yield

$$
\begin{aligned}
B_{n2}(u,W) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i^\top\{\widetilde{Q}_n(\widehat{\boldsymbol{\beta}}_n)\}^{-1}\sum_{j=1}^{n} X_j h_j' \varepsilon_j I(X_i^\top W \le u) g'(X_i^\top \widetilde{\boldsymbol{\beta}}_n) \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\sum_{i=1}^{n} X_i^\top\{\widetilde{Q}_n(\widehat{\boldsymbol{\beta}}_n)\}^{-1} I(X_i^\top W \le u) g'(X_i^\top \widetilde{\boldsymbol{\beta}}_n)\right] X_j h_j' \varepsilon_j,
\end{aligned}
$$

where $\widetilde{\boldsymbol{\beta}}_n$ is a vector such that $X_i^\top \widetilde{\boldsymbol{\beta}}_n$ lies between $X_i^\top \widehat{\boldsymbol{\beta}}_n$ and $X_i^\top \boldsymbol{\beta}_0$. Applying the result of Lemma A.3 with additional simplifications yields

$$
\begin{aligned}
\max_j \Bigg| \frac{1}{n}\sum_{i=1}^{n} \Big[& X_i^\top\{\widetilde{Q}_n(\widehat{\boldsymbol{\beta}}_n)\}^{-1} I(X_i^\top W \le u) g'(X_i^\top \widetilde{\boldsymbol{\beta}}_n) \\
& - E\{X^\top \Sigma_\beta^{-1} I(X^\top W \le u) g'(X^\top \boldsymbol{\beta}_0)\}\Big] X_j h_j' \Bigg| \to 0. \qquad (A.24)
\end{aligned}
$$

As a result, we have

$$
B_{n2}(u,W) = \frac{1}{\sqrt{n}}\Gamma(u)\sum_{i=1}^{n}\varepsilon_i X_i h_i' + o_p(1). \qquad (A.25)
$$

Similarly, the third term can be expressed as

$$B_{n3}(u,W) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{\alpha}_n-\alpha_0)I(X_i^\top W\le u)\int_0^1 g'(\alpha_0+s(\widehat{\alpha}_n-\alpha_0))ds.$$

For simplicity, we consider the case with one additional shape parameter. Substituting $\widehat{\alpha}_n-\alpha_0$ using Lemma A.1 and applying the mean-value theorem yield

$$\begin{aligned} B_{n3}(u,W) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widetilde{A}_n(\widehat{\alpha})\}^{-1}\sum_{j=1}^{n}X_jh_j^{'}\varepsilon_jI(X_i^\top W\le u)g'(\widetilde{\alpha}) \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\{\widetilde{A}_n(\widehat{\alpha})\}^{-1}I(X_i^\top W\le u)g'(\widetilde{\alpha})\right]X_jh_j^{'}\varepsilon_j, \end{aligned}$$

where $\widetilde{\alpha}$ is a value lies between $\widehat{\alpha}$ and α_0, $\widetilde{A}_n(\alpha_1,\alpha_2)=\int_0^1 A_n(\alpha_1+s(\alpha_2-\alpha_1))ds$, $A_n(\alpha)=-\partial L_n(\alpha)/\partial\alpha$.

Denote $\Sigma_\alpha = E\{\sum_{i=1}^n g(h_\alpha(\eta))(1-g(h_\alpha(\eta)))\frac{\partial h(\alpha)}{\partial\alpha}\frac{\partial h(\alpha)}{\partial\alpha}\}$. Applying the result of Lemma A.3 with additional simplifications yields

$$\begin{aligned} \max_j\Bigg|\frac{1}{n}\sum_{i=1}^{n}\Big[&\{\widetilde{A}_n(\widehat{\alpha})\}^{-1}I(X_i^\top W\le u)g'(\widetilde{\alpha}) \\ &-E\{\Sigma_\alpha^{-1}I(X^\top W\le u)g'(\alpha_0)\}\Big]\frac{\partial h_\alpha(\eta_j)}{\partial\alpha}\Bigg|\to 0. \end{aligned} \quad (A.26)$$

Denote $\Theta(u)=E\{\Sigma_\alpha^{-1}I(X^\top W\le u)g'(\alpha_0)\}$. As a result, we have

$$B_{n3}(u,W)=\frac{1}{\sqrt{n}}\Theta(u)\sum_{i=1}^{n}\varepsilon_i\frac{\partial h_\alpha}{\partial\alpha}+o_p(1). \quad (A.27)$$

So we have the following expression for $M_n(u,W)$.

$$M_n(u,W)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{I(X_i^\top W\le u)-\Gamma(u)X_ih_i^{'}-\Theta(u)\frac{\partial h_\alpha}{\partial\alpha}\right\}\varepsilon_i+o_p(1). \quad (A.28)$$

It is easy to see that $I(X^\top W \leq u)$ is monotone with respect to u. By Lemma 9.10 of Kosorok (2008), the function class $\{I(X^\top W \leq u) : u \in \mathbb{R}^1\}$ is a VC-class. Similarly the function classes $\{\Gamma(u) : u \in \mathbb{R}^1\}$ and $\{\Theta(u) : u \in \mathbb{R}^1\}$ are VC-class as well. By Theorem 2.6.8 of van der Vaart & Wellner (1996), the function classes $\{\varepsilon I(X^\top W \leq u) : u \in \mathbb{R}^1\}$, the class $\{\Gamma(u)\varepsilon X h' : u \in \mathbb{R}^1\}$ and the class $\{\Theta(u)\varepsilon \frac{\partial h_\alpha}{\partial \alpha} : u \in \mathbb{R}^1\}$ are all VC-class. Then by Lemma 9.17 of Kosorok (2008), the function class $\{\Psi_u(u, y, \varepsilon, w) : u \in \mathbb{R}^1\}$ is a VC-class. We can take the envelope function as $|\varepsilon| + \mathrm{E}(\| X \|)|\varepsilon||\Sigma_\beta^{-1}|X||h'| + |\varepsilon|\Sigma_\alpha^{-1}|\frac{\partial h_\alpha}{\partial \alpha}|$. By Theorems 2.6.7 and 2.5.2 of van der Vaart & Wellner (1996), we can prove that the estimated empirical process $M_n(u, W)$ converges weakly to $M(u)$ in the Skorohod space $S[\Pi]$. By the continuous mapping theorem, we can prove the result for T_n. $\square$

A.3 Proof of Theorem 2.2

Let $\widehat{\boldsymbol{\beta}}_n^*$ be the MLE of $\boldsymbol{\beta}$ based on the bootstrap samples (X_i, Y_i^*) for $i = 1, \cdots, n$. Analogously to establish (A.8), we can verify that

$$\widetilde{Q}_n(\widehat{\boldsymbol{\beta}}_n^*)(\widehat{\boldsymbol{\beta}}_n^* - \widehat{\boldsymbol{\beta}}_n) = L_n^*(\widehat{\boldsymbol{\beta}}_n) = \sum_{i=1}^n X_i\{Y_i^* - g(h_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n))\}h_i'(\widehat{\boldsymbol{\beta}}_n). \tag{A.29}$$

Write the bootstrap version of $M_n(u, W)$ as

$$M_n^*(u, W) = 1/\sqrt{n}\sum_{i=1}^n \{Y_i^* - g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n^*)\}I(X_i^\top W \leq u).$$

Note that $Y_i^* - g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n^*) = \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} - \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n^*) - g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} - \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n) - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\}$. Then, we have

$$\begin{aligned} M_n^*(u, W) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} I(X_i^\top W \le u) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n^*) - g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} I(X_i^\top W \le u) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n) - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} I(X_i^\top W \le u) \\ &:= M_{n1}^*(u, W) - M_{n2}^*(u, W) - M_{n3}^*(u, W). \end{aligned} \tag{A.30}$$

Applying (A.29) along with the similar proof to that for (A.25) yields that

$$\begin{aligned} M_{n2}^*(u, W) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n^*) - g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} I(X_i^\top W \le u) \\ &= \frac{1}{\sqrt{n}} \Gamma(u) \sum_{i=1}^n X_i \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} h_i'(\widehat{\boldsymbol{\beta}}_n) + o_p(1). \end{aligned} \tag{A.31}$$

Applying (A.29) along with the similar proof to that for (A.27) yields that

$$\begin{aligned} M_{n3}^*(u, W) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_{\widehat{\alpha}^*}(X_i^\top \widehat{\boldsymbol{\beta}}_n) - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} I(X_i^\top W \le u) \\ &= \frac{1}{\sqrt{n}} \Theta(u) \sum_{i=1}^n \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} \frac{\partial h_{\widehat{\alpha}}}{\partial \widehat{\alpha}} + o_p(1). \end{aligned} \tag{A.32}$$

It follows from (A.30)-(A.32) that

$$\begin{aligned} M_n^*(u, W) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} [I(X_i^\top W \le u) \\ &\quad - \frac{1}{\sqrt{n}} \Gamma(u) \sum_{i=1}^n X_i \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} h_i'(\widehat{\boldsymbol{\beta}}_n) \\ &\quad - \frac{1}{\sqrt{n}} \Theta(u) \sum_{i=1}^n \{Y_i^* - g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n)\} \frac{\partial h_{\widehat{\alpha}}}{\partial \widehat{\alpha}} + o_p(1). \end{aligned} \tag{A.33}$$

Recall that $P(Y_i^* = 1|\text{data}) = g_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n) = g(h_{\widehat{\alpha}}(X_i^\top \widehat{\boldsymbol{\beta}}_n))$ and $g(v) = \{1+\exp(-v)\}$, the similar arguments to the proof of Theorem 2.1 along the line with the proof of Theorem 2 in Dikta et al. (2006) can prove that the conditional distribution of T_n^* converges in distribution to the limiting null distribution of T_n.

Note that the validity of (A.29) is independent of $D(x) = 0$ or not. We can similarly prove that the conditional distribution of T_n^* converges in distribution to the limiting alternative distribution of T_n. Theorem 2.2 follows. $\square$

References

Aranda-Ordaz, F. J. (1981), 'On two families of transformations to additivity for binary response data', *Biometrika* **68**(2), 357–363.

Chen, K., Hu, I. & Ying, Z. (1999), 'Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs', *Annals of Statistics* **27**(4), 1155–1163.

Dikta, G., Kvesic, M. & Schmidt, C. (2006), 'Bootstrap approximations in model checks for binary data', *Journal of the American Statistical Association* **101**(474), 521–530.

Escanciano, J. C. (2006), 'A consistent diagnostic test for regression models using projections', *Econometric Theory* **22**, 1030–1051.

Guerrero, V. M. & Johnson, R. A. (1982), 'Use of the Box-Cox transformation with binary response models', *Biometrika* **69**(2), 309–314.

Heuser, H. (1981), *Lehrbuch der Analysis. Teil 2*, B. G. Teubner.

Hosmer, D. W. & Hjort, N. L. (2002), 'Goodness-of-fit processes for logistic regression: simulation results', *Statistics in Medicine* **21**(18), 2723–2738.

Hosmer, D. W., Hosmer, T., Le Cessie, S. & Lemeshow, S. (1997), 'A comparison of goodness-of-fit tests for the logistic regression model', *Statistics in Medicine* **16**(9), 965–980.

Hosmer, D. W. & Lemeshow, S. (1989), *Applied Logistic Regression*, John Wiley & Sons.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics, Springer-Verlag.

Li, X., Chen, F., Liang, H. & Ruppert, D. (2023), 'Model checking for logistic models when the number of parameters tends to infinity', *Journal of Computational and Graphical Statistics* **32**, 241–251.

Liang, H. & Du, P. (2012), 'Maximum likelihood estimation in logistic regression models with a diverging number of covariates', *Electronic Journal of Statistics* **6**, 1838–1846.

Lindsey, J. K. (1997), *Applying Generalized Linear Models*, Springer.

McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, Vol. 37, Chapman and Hall.

Milicer, H. & Szczotka, F. (1966), 'Age at menarche in Warsaw girls in 1965', *Human Biology* **38**(3), 199–203.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, Series A* **135**, 370–384.

Osius, G. & Rojek, D. (1992), 'Normal goodness-of-fit tests for multinomial models with large degrees of freedom', *Journal of the American Statistical Association* **87**(420), 1145–1152.

Pregibon, D. (1980), 'Goodness of link tests for generalized linear models', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **29**(1), 15–24.

Prentice, R. L. (1976), 'A generalization of the probit and logit methods for dose response curves', *Biometrics* pp. 761–768.

Stukel, T. A. (1988), 'Generalized logistic models', *Journal of the American Statistical Association* **83**(402), 426–431.

Stute, W. & Zhu, L.-X. (2002), 'Model checks for generalized linear models', *Scandinavian Journal of Statistics. Theory and Applications* **29**, 535–545.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

Xia, Y. (2006), 'Asymptotic distributions for two estimators of the single-index model', *Econometric Theory* **22**, 1112–1137.

Yin, C., Zhao, L. & Wei, C. (2006), 'Asymptotic normality and strong consistency of maximum quasi-likelihood estimates in generalized linear models', *Science in China. Series A. Mathematics* **49**, 145–157.