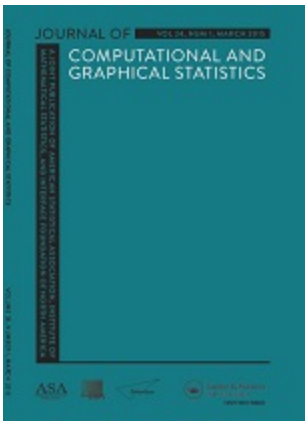

Surrogate Residuals for Discrete Choice Models

Journal:	*Journal of Computational and Graphical Statistics*
Manuscript ID	JCGS-19-067.R1
Manuscript Type:	Original Article
Keywords:	Categorical outcome, Model diagnostics, Multinominal logistic regression, Residual analysis

SCHOLARONE™
Manuscripts

Surrogate Residuals for Discrete Choice Models

Chao Cheng
Department of Mathematical Sciences, Tsinghua University
Rui Wang
Department of Public Economics, School of Economics, Xiamen University
and
Heping Zhang*
Department of Biostatistics, School of Public Health, Yale University

August 27, 2019

For Peer Review Only

Abstract

Discrete Choice Models (DCMs) are a class of models for modelling response variables that take values from a set of alternatives. Examples include logistic regression, probit regression, and multinomial logistic regression. These models are also referred together as generalized linear models. Although there exist methods for the goodness of fit of DCMs, defining intuitive residuals for such models has been difficult due to the fact that the responses are categorical values instead of continuous numbers. In this article, we propose the surrogate residual for DCMs based on the surrogate approach [Liu and Zhang, *J. Am. Stat. Assoc.*, **113**, 522 (2018) 845–854], which deals with an ordinal response. We consider categorical responses that may or may not be ordered. We shall show that our residual can be used to diagnose misspecification in the aspects of mean structure, individual-specific coefficients, and interaction effects.

Keywords: Categorical outcome; model diagnostics; multinominal logistic regression; residual analysis.

*Address for correspondence: Heping Zhang, Department of Biostatistics, School of Public Health, Yale University, New Haven, CT, USA. Email: heping.zhang@yale.edu

1 Introduction

Discrete Choice Models (DCMs) (McFadden, 1978), a class of models designed for modelling the choice from a finite set of alternatives, are widely applied in medical research (Chan, 2005), transportation demand forecasting (Ben-Akiva and Lerman, 1985), environmental analysis (Train, 1998), and evacuation modelling (Lovreglio et al., 2014). Specific forms of DCMs consist of logistic regression, probit regression, multinominal logistic regression, nested logit regression (Koppelman and Bhat, 2006) and so on. Plenty of diagnostic tests (Goeman and le Cessie, 2006; Pagan and Vella, 1989; Hausman and McFadden, 1981; Nagel and Hatzinger, 1992) were developed to check the validity of model assumptions for DCMs, whereas there are few useful tools for residual diagnosis for such models. Those diagnostic tests can provide useful measures for the probability to reject the null hopothesis, while the residual diagnosis can be more intuitive and revealing if appropriately developed. For instance, through the residual-by-covariate plot in the ordinary linear regression, we can straightforwardly review which parts of observed data are poorly fitted, thus producing potentially useful hints on the model improvement.

The difficulty of defining residuals for DCMs arises from the fact that the response variable is just a label of alternatives instead of a continuous value. Nagel and Hatzinger (1992) proposed a residual based on a quadratic-form statistic and applied the residual-by-leverage plot to help diagnose the model. However, this method results in an always-positive residual which may be inconvenient to interpret.

An intuitive approach is to assign numeric values for the discrete alternatives, and then applying the Pearson's residual, deviance residual or some other generalized residuals for above assigned numeric values (Dunn and Smyth, 1996; Gupta et al., 2008). However, this approach may be also problematic due to the following two reasons. First, how to assign the numeric values is arbitrary and subjective, lacking a sound rationale. Second, noting that the choice set is finite, the assigned numbers must be discrete; and the residuals based on discrete space could lead to unusual residual patterns even if the model is fitted correctly (Dunn and Smyth, 1996; Liu and Zhang, 2018). The following example illustrates the limitations of this approach.

For Peer Review Only

Example 1. (Correct specification of the model). Suppose 2000 observations are generated from the following logistic regression

$$Pr(Y = 1|X) = logit(1 + 2X),$$

where $Y = \{1, 2\}$ is the response representing alternatives 1 and 2, $X \sim Unif(-5, 5)$, and $logit(x) = \frac{e^x}{1+e^x}$. Noting Y is a categorical variable, we may use a numeric value $\tilde{Y} = \{1, 0\}$ to represent the alternatives 1 and 2 respectively. Now, we use the true model to fit the simulated data, and obtain the Pearson's and deviance residuals, which are defined by $r^{(P)} = (\tilde{y} - \hat{p})/\sqrt{\hat{p}(1-\hat{p})}$ and $r^{(D)} = \mathbb{I}(\tilde{y} - \hat{p})\sqrt{2[(1-\tilde{y})\ln(1-\hat{p}) - \tilde{y}\ln(\hat{p})]}$, respectively. Here, $\hat{p} = \hat{P}(\tilde{Y} = 1|X = x)$, and $\mathbb{I}(c) = 1$ if $c > 0$ and 0 otherwise.

Figure 1 (panels a - d) presents the residual-by-covariate plots (r-by-x) and quantile-quantile (QQ) plots (the empirical distribution of r versus the standard normal distribution) for both Pearson's and deviance residuals. It is difficult to observe anything useful from Pearson's and deviance residuals even if the logistic model is correctly fitted. In fact, how these residuals are defined is problematic. First, the distribution of the residual variable conditional on covariate, i.e $R|X$, varies across the values of X (see Figure 1 (a) and (c)). Second, the null distribution of R may depend on the distribution of X (see Figure 1 (b) and (d)). These issues limit the use of these residuals for model diagnostic.

Liu and Zhang (2018) provided the surrogate residual for ordinal regression. Specifically, they constructed a continuous variable, S, named surrogate variable, which is given conditionally on the ordinal outcome variable Y, and then defined the surrogate residual as the difference of S and its expectation under the null hypothesis. Due to the continuity of S, the surrogate residual avoids the problems encountered in example 1. This advantage motivated us to develop a residual based on a continuous surrogate variable rather than the discrete outcomes.

For Peer Review Only

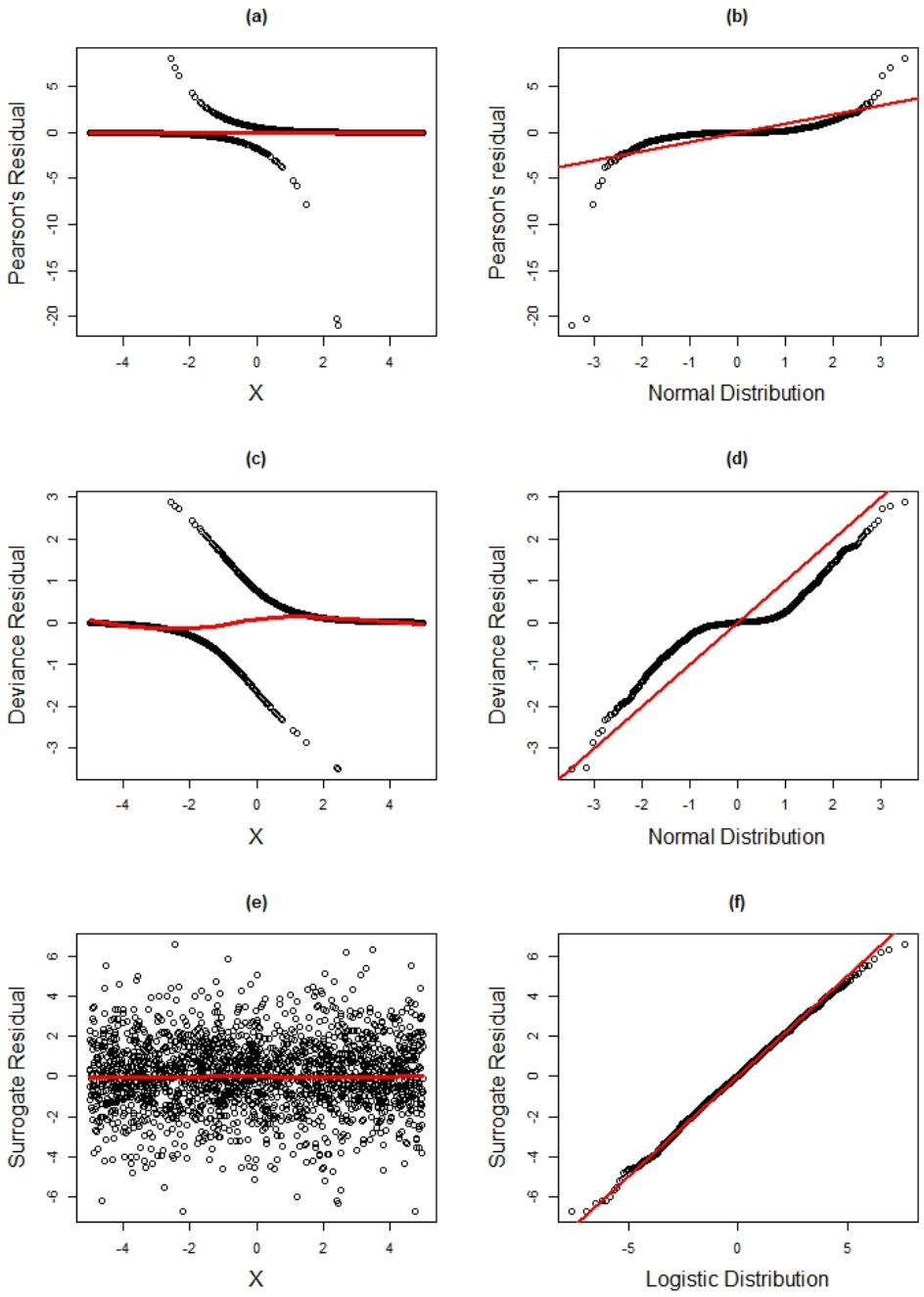

Figure 1: Model diagnostics using the Pearson's residual (upper row), deviance residual (middle row) and surrogate residual (lower row) in the case that the logistic regression is correctly fitted. Figures (a), (c) and (e) present the plots of the residual-by-covariate (r-by-x), and loess curves (red solid) are added as references. Figures (b), (d) and (f) are the corresponding QQ-plots (quantile-quantile plots) of above mentioned residuals.

In this article, we are concerned with DCMs for which the response variable takes discrete choices without orders. This response variable is different from the ordinal response investigated by Liu and Zhang (2018). However, we extend the use of the surrogate approach to define the residual for DCMs. Because discrete choices have no order, each alternative has its unique utility function. This is the major difference between our work and that of Liu and Zhang (2018). The multiple utility functions create additional challenges in defining the residual. First, an one-dimensional residual is no longer appropriate because it cannot represent the multi-dimensional utilities. Second, the residual needs to reflect the relationship among different alternatives. In this work, the surrogate residual $\boldsymbol{R}$ for DCMs is a vector for which each dimension corresponding to one particular alternative, and has a reference distribution with mean $\mathbf{0}$ and is independent of the covariate $\boldsymbol{X}$ if the assumed model holds. Based on theoretical and graphical analyses, we shall show that, by comparing the surrogate residual with its reference distribution and synthesizing the information across different residual dimensions, our residual is able to detect various misspecifications, including detecting the misspecified mean structure, checking individual-specific coefficients, and detecting interaction effects.

The framework of this paper is as follows: Discrete choice models are briefly introduced in Section 2. In Section 3, we define the surrogate residual for DCMs, and present its theoretical and graphical properties. We generalize the surrogate residual for general statistical models with unordered categorical responses in Section 4. In Section 5, we re-analyze a publicly available dataset (Hoogendoorn-Lanser et al., 2015) to demonstrate the use of the surrogate residual in real data analysis. We conclude this paper with a brief discussion in Section 6.

2 Discrete Choice Models

Suppose a categorical variable Y that has J alternatives $\{1, \ldots, J\}$. Consider there are J utility levels

$$U_j = V_j(\boldsymbol{X}_j, \boldsymbol{\beta}_j) + \epsilon_j \tag{1}$$

For Peer Review Only

corresponding to alternative $j = 1, \ldots, J$, where $V_j(\boldsymbol{X}_j, \boldsymbol{\beta}_j)$ is the deterministic part depending on covariates $\boldsymbol{X}_j = [X_{j1}, \ldots, X_{jp}]^T$ and unknown coefficients $\boldsymbol{\beta}_j$, and ϵ_j is the random part. The random parts $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_J]^T$ are assumed to follow a J-dimensional continuous distribution $G(\boldsymbol{\epsilon})$. In a DCM, $Y = y$ if and only if $\forall j \neq y$, $U_y > U_j$; i.e, the probability of selecting alternative y is

$$Pr(Y = y) = Pr(U_y > U_j, \forall j \neq y), \tag{2}$$

where $y \in \{1, 2, \ldots, J\}$. In order to ensure the uniqueness of the unknown parameters $\boldsymbol{\beta}$, some specific models select an alternative as reference and force its whole utility (both deterministic and random parts) as zero. Without loss of generality, we select the last alternative J as the reference.

The following are a few examples for how model (1) can be specified; for DCMs whose choice set contains 2 alternatives, we have:

- Logistic regression, where $U_2 \equiv 0$ and ϵ_1 follows a logistic distribution with c.d.f $F(\gamma) = e^{\gamma}/(1 + e^{\gamma})$.
- Probit regression, where $U_2 \equiv 0$ and ϵ_1 follows a standard normal distribution.

Below are multinomial choice models whose choice set contains 3 or more alternatives:

- Multinomial logistic regression, where ϵ_j's are mutually independent, and follow a standard gumbel distribution with c.d.f $F(\gamma) = \exp(-e^{-\gamma})$.
- Multinomial probit regression, where $U_J \equiv 0$ and $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_{J-1}]^T \sim N(0, \Sigma_{(J-1)\times(J-1)})$.
- Nested logit regression, where the c.d.f of $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_J]^T$ is $F(\epsilon_1, \ldots, \epsilon_J) = \exp\Big(-\sum_{s=1}^{S}\big(\sum_{j\in B_s} e^{-\epsilon_j/\lambda_s}\big)^{\lambda_s}\Big)$.

Remark 1. *Gumbel distribution will be used later, so we briefly describe its known properties here. The c.d.f of Gumbel distribution is* $F(x; \theta, \alpha) = \exp(-e^{(x-\theta)/\alpha})$, *where* θ *is the location parameter and* α *is the scale parameter. The variance is* $\frac{\pi^2}{6}\alpha^2$, *and the mean is given by* $\theta + \zeta\alpha$, *where* $\zeta \approx 0.5772$ *denotes the Euler-Mascheroni constant. Moreover, it is a positive skewed distribution with the moment coefficient of skewness 1.1423. The standard*

For Peer Review Only

Gumbel distribution is the Gumbel distribution with $\theta = 0$ and $\alpha = 1$, whose mean is ζ and variance is approximately equal to 1.6449.

3 Surrogate Residual for DCMs

3.1 Definition

In this section, we propose a residual for DCMs based on the surrogate approach proposed by Liu and Zhang (2018). We shall (1) find a hypothetical variable $\boldsymbol{Z}$ following a continuous distribution, (2) construct a joint distribution $(\boldsymbol{Z}, Y)$ to link $\boldsymbol{Z}$ and Y together, and (3) define a surrogate variable $\boldsymbol{S}$ as the conditional distribution of $\boldsymbol{Z}$ given Y. Finally, the surrogate residual $\boldsymbol{R}$ is the difference between $\boldsymbol{S}$ and the null distribution of $\boldsymbol{S}$ conditional on $\boldsymbol{X}$, i.e $E_0(\boldsymbol{S}|\boldsymbol{X})$.

In DCMs, the utility vector $\boldsymbol{U} = [U_1, \ldots, U_J]^T$ is a convenient hypothetical variable because $\boldsymbol{U}$ is a continuous variable due to the continuity of $\boldsymbol{\epsilon}$; and importantly, equation (2) induces a natural joint distribution linking $\boldsymbol{U}$ and Y. Specifically, the joint distribution can be set as

$$Y \triangleq y \iff U_y > U_j \quad \forall j \neq y.$$

Now, following the procedure of the surrogate approach, we let the surrogate variable $\boldsymbol{S}$ be a J-dimensional random variable following the conditional distribution of $\boldsymbol{U}$ given the observation Y. More specifically, $\boldsymbol{S} \sim \boldsymbol{U}|(Y, \boldsymbol{X})$ follows a truncated distribution obtained by truncating the distribution of $\boldsymbol{U}|\boldsymbol{X} = \boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$ using the region $U_Y > U_j$ $(\forall j \neq Y)$, where $\boldsymbol{U} = [U_1, \ldots, U_J]^T$ and $\boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) = [V_1(\boldsymbol{X}_1, \boldsymbol{\beta}_1), \ldots, V_J(\boldsymbol{X}_J, \boldsymbol{\beta}_J)]^T$. We name S_j $(j \in \{1, \ldots, J\})$ as the surrogate response for alternative j or utility U_j. We define the surrogate residual $\boldsymbol{R}$ as

$$\boldsymbol{R} = \boldsymbol{S} - E_0(\boldsymbol{S}|\boldsymbol{X}) = \boldsymbol{S} - E(\boldsymbol{U}|\boldsymbol{X}) = \boldsymbol{S} - \boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) - \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon}),$$

where $E_0(\boldsymbol{S}|\boldsymbol{X})$ denotes the conditional mean of $\boldsymbol{S}$ given $\boldsymbol{X}$ under the null hypothesis that the model is specified correctly. In practice, given the observed data $\{y, \boldsymbol{x}\}$ and a fitted model, we estimate the conditional distribution $\boldsymbol{S} \sim \boldsymbol{U}|(y, \boldsymbol{x})$ by plugging in the estimated

For Peer Review Only

parameter $\hat{\boldsymbol{\beta}}$, and randomly draw a sample $\boldsymbol{s}$ from this plug-in distribution. Finally, the residual $\boldsymbol{r}$ is given by $\boldsymbol{r} = \boldsymbol{s} - \boldsymbol{V}(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) - \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})$.

We should note that $\boldsymbol{r}$ is a sample from $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$ rather than a realization of $\boldsymbol{R} \equiv \boldsymbol{R}_{\boldsymbol{\beta}}$, but we can show that when $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$ in probability, $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$ converges to $\boldsymbol{R}_{\boldsymbol{\beta}}$ in distribution. In addition, the properties of $\boldsymbol{R}_{\boldsymbol{\beta}}$ apply to $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$ asymptotically (see Theorems 3 and 4).

In practice, we need to draw a sample from $\boldsymbol{S}$, which follows a truncated J-dimensional distribution. As we know, *Inverse CDF Sampling* and *Rejection Sampling* are two effective methods to generate samples in one dimension. Due to the multidimensionality of $\boldsymbol{S}$, directly applying either method is quite difficult. In this paper, we introduced the Gibbs sampling method to generate this truncated multi-dimensional distribution by transforming the multidimensional sampling problem into a series of full conditional distribution (which is one-dimensional) sampling issues. Notice that $\boldsymbol{S}$ is a distribution subject to a specific choice, and hence the full conditional distribution of each dimension can be simplified and generated by *Inverse CDF Sampling* or any other regular univariate sampling algorithm. The specific procedure of generating samples from $\boldsymbol{S}$ is deferred to Appendix A in the supplementary material.

Remark 2. *As discussed in Section 2, some DCMs force $U_J \equiv 0$, and assume $\boldsymbol{\epsilon}$ follows a $(J-1)$-dimensional distribution $G(\epsilon_1, \ldots, \epsilon_{J-1})$. In this case, the c.d.f of the surrogate variable $\boldsymbol{S}$ follows a truncated distribution obtained by truncating the distribution of $\boldsymbol{U}|\boldsymbol{X} = \boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$ using the region $U_Y > U_j$, where $\boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) = [V_1(\boldsymbol{X}_1, \boldsymbol{\beta}_1), \ldots, V_{J-1}(\boldsymbol{X}_{J-1}, \boldsymbol{\beta}_{J-1})]^T$, $\boldsymbol{U} = [U_1, \ldots, U_{J-1}]^T$ and U_J is fixed as 0. This distribution is a special form of distribution of $\boldsymbol{S}$ above, in which $U_J \equiv 0$ can be seen as a degenerate distribution. Thus, the properties derived in Section 3.2 are still applicable in this case.*

3.2 Theoretical Properties

We investigate the theoretical properties of the surrogate variable $\boldsymbol{S}$ and the residual variable $\boldsymbol{R}$ in this subsection. These properties provide a theoretical foundation for model diagnostics. Firstly, suppose the true utility levels are

$$\tilde{U}_j = \tilde{V}_j(\boldsymbol{X}_j, \tilde{\boldsymbol{\beta}}_j) + \tilde{\epsilon}_j \quad j = 1, \ldots, J,$$

For Peer Review Only

where $\tilde{\boldsymbol{\epsilon}} = [\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_J]^T \sim \tilde{G}(\tilde{\boldsymbol{\epsilon}})$, and the true model for Y is

$$Pr(Y = y) = Pr(\tilde{U}_y > \tilde{U}_j, \forall j \neq y), \tag{3}$$

where $y = \{1, \ldots, J\}$. Then, the distribution of $\boldsymbol{S}$ is

$$\begin{aligned} Pr(\boldsymbol{S} \leq \boldsymbol{c}) &= \sum_{y=1}^{J} Pr(\boldsymbol{S} \leq \boldsymbol{c} | Y = y) Pr(Y = y) \\ &= \sum_{y=1}^{J} \frac{Pr(\boldsymbol{U} \leq \boldsymbol{c} \bigwedge U_y > U_j, \forall j \neq y)}{Pr(U_y > U_j, \forall j \neq y)} Pr(\tilde{U}_y > \tilde{U}_j, \forall j \neq y), \end{aligned} \tag{4}$$

where $\boldsymbol{U} = [U_1, \ldots, U_J]^T$ is the assumed utility levels in equation (1), $\tilde{\boldsymbol{U}} = [\tilde{U}_1, \ldots, \tilde{U}_J]^T$ is the true utility levels, and $\boldsymbol{c} = [c_1, \ldots, c_J]^T$ is an arbitrary but fixed point in $\mathbb{R}^J$. Then denote the area $(-\infty, \boldsymbol{c})$ by $\mathcal{C}$, $\{U_y > U_j, \forall j \neq y\}$ by $\mathcal{D}_y$, and $\{\tilde{U}_y > \tilde{U}_j, \forall j \neq y\}$ by $\tilde{\mathcal{D}}_y$. Now, the distribution (4) can be rewritten as

$$Pr(\boldsymbol{S} \leq \boldsymbol{c}) = \sum_{y=1}^{J} \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\boldsymbol{U} \in \mathcal{D}_y)} Pr(\tilde{\boldsymbol{U}} \in \tilde{\mathcal{D}}_y). \tag{5}$$

Equation (5) shows that the distribution of $\boldsymbol{S}$ is determined jointly by the assumed and the true model for Y. If the assumed model agrees with the true model for Y, i.e $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$, we have $\boldsymbol{U} = \tilde{\boldsymbol{U}}$ and $\mathcal{D}_y = \tilde{\mathcal{D}}_y$ for all $y = 1, \ldots, J$. Furthermore, noticing that $\{\mathcal{D}_1, \ldots, \mathcal{D}_J\}$ is a partition of $\mathbb{R}^J$, we have

$$\begin{aligned} Pr(\boldsymbol{S} \leq \boldsymbol{c}) &= \sum_{y=1}^{J} \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\tilde{\boldsymbol{U}} \in \tilde{\mathcal{D}}_y)} Pr(\tilde{\boldsymbol{U}} \in \tilde{\mathcal{D}}_y). \\ &= \sum_{y=1}^{J} Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y) = Pr\Big(\boldsymbol{U} \in \bigcup_{y=1}^{J} (\mathcal{C} \cap \mathcal{D}_y)\Big) \\ &= Pr\Big(\boldsymbol{U} \in \mathcal{C} \bigcap (\cup_{y=1}^{J} \mathcal{D}_y)\Big) = Pr(\boldsymbol{U} \leq \boldsymbol{c}), \end{aligned} \tag{6}$$

i.e, the surrogate variable $\boldsymbol{S}$ has the same distribution as $\boldsymbol{U}$. This equation immediately yields the following theorem.

Theorem 1. *If the assumed model (2) agrees with the true model (3) (i.e, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$), then the following results hold*

(1) The surrogate variable $\boldsymbol{S}$ has the same distribution as $\boldsymbol{U}$, i.e $\boldsymbol{S}|\boldsymbol{X} = \boldsymbol{U}|\boldsymbol{X} = \boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$.

(2) The residual variable $\boldsymbol{R}$, independent from X, follows the distribution $G\big(\boldsymbol{r} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})\big)$, i.e, $Pr(\boldsymbol{R} \leq \boldsymbol{r}) = Pr(\boldsymbol{R} \leq \boldsymbol{r}|\boldsymbol{X}) = G\big(\boldsymbol{r} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})\big)$.

Theorem 1 leads to the following result for checking the assumed model.

For Peer Review Only

Theorem 2. *If the assumed model (2) agrees with the true model (3) (i.e, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$), then the residual variable $\boldsymbol{R}$ has the following properties*

(1) (Zero mean) $E(\boldsymbol{R}|\boldsymbol{X}) = \boldsymbol{0}$.

(2) (Homogeneous variance) $Var(\boldsymbol{R}|\boldsymbol{X}) = Var(\boldsymbol{\epsilon})$ is a constant matrix, not depending on $\boldsymbol{X}$.

Theorems 1 and 2 establish the theoretical foundation of using $\boldsymbol{R}$ for model diagnosis. Our surrogate residual $\boldsymbol{R}$ has similar properties to the residual in the ordinary linear regression. By plotting the quantile-quantile plot of $\boldsymbol{r}$ with the reference distribution $G\big(\boldsymbol{r} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})\big)$, we can assess whether $\boldsymbol{R}$ agrees with $\tilde{\boldsymbol{R}}$. By drawing the residual-by-covariate plot, we can evaluate what aspects of our assumed models are poorly fitted.

Theorems 1 and 2 concern the surrogate residual variable $\boldsymbol{R} \equiv \boldsymbol{R}_{\boldsymbol{\beta}}$. Theorems 3 and 4 present the parallel results when $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + o_p(1)$, here $o_p(1)$ denotes a term vanishing to zero in probability.

Theorem 3. *If the assumed model (2) agrees with the true model (3) (i.e, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$), then $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$ converges to $G\big(\boldsymbol{r} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})\big)$ in distribution as $n \to \infty$, conditional or unconditional on $\boldsymbol{X}$, i.e, $Pr(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}} \le \boldsymbol{r}) = Pr(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}} \le \boldsymbol{r}|\boldsymbol{X}) = G\big(\boldsymbol{r} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon})\big) + o(1)$.*

Theorem 4. *If the assumed model (2) agrees with the true model (3) (i.e, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$), then $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$ has the following properties*

(1) $E(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}|\boldsymbol{X}) = o(1)$.

(2) $Var(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}|\boldsymbol{X})$ is a constant matrix, not depending on $\boldsymbol{X}$, except a vanishing term $o(1)$.

Appendix B in the supplementary material provides detailed proofs for Theorems 3 and 4.

3.3 Graphical Properties

In this subsection, using several numerical examples, we examine graphical properties of our proposed residual, when the model is specified correctly or misspecified. We first revisit Example 1.

For Peer Review Only

Example 1. (Continued) We choose alternative 2 as reference; thus, the true utilities are $U_1 = 1 + 2X + \epsilon_1$ and $U_2 \equiv 0$, corresponding to alternative 1 and 2 respectively. Here, ϵ_1 follows a logistic distribution. When the model is specified correctly, our surrogate residual is an one-dimensional random variable following a logistic distribution. We obtain the surrogate residual, and draw the residual-by-covariate plot and QQ plot in Figure 1 (e) and (f). As the residual-by-covariate plot shows, our surrogate residuals are randomly fluctuating around zero and independent from the covariate X, exhibiting no unusual pattern. Based on the QQ plot, we can see that the residuals are in accordance with the reference distribution.

We should note that Example 1 is a very special instance of DCMs as it has two alternatives only. Most DCMs, however, contain $J \geq 3$ alternatives, indicating that the surrogate residual is a J or $J-1$ (if $U_J \equiv 0$ is constrained) dimensional vector. Unlike the residuals in ordinary regression and other generalized linear models, this characteristic creates additional difficulties in model diagnostics. First, we need to evaluate whether every dimension of the residual vector satisfies the null hypothesis or not. Second, we need to synthesize the information across different residual dimensions and identify which part of the model assumptions is misspecified.

In the following, we use the multinomial logistic regression with a 3-alternative response as examples to illustrate how to detect a misspecified mean structure, check the proportionality assumption, and inspect interaction effects. Similar steps are applicable to general DCMs.

3.3.1 Detecting Misspecified Mean Structure

After fitting a discrete choice model it is important to determine whether the mean structures (i.e the deterministic parts $V_j(\boldsymbol{X}_j, \boldsymbol{\beta}_j)$) of all utilities are correctly specified before performing inference.

We can detect the misspecified mean structure based on the residual-by-covariate plots. If the assumed model holds, our surrogate residual should be randomly distributed around zero, independent from X; and if the mean structure is misspecificed, the patterns of our

For Peer Review Only

surrogate residual depend on X.

Example 2. We generate 2000 observations from a multinomial logistic regression with the following utility levels

$$U_j = \beta_{j0} + \beta_{j1}X + \beta_{j2}X^2 + \epsilon_j, \quad j = 1, 2, 3,$$

where ϵ_j follows a standard Gumbel distribution; X is an individual-specific covariate generated from $Unif(-5, 5)$; $\beta_{30} = \beta_{31} = \beta_{32} = 0$ as the reference; $\beta_{10} = 1$, $\beta_{11} = 7$, $\beta_{12} = -2$; and $\beta_{20} = 8$, $\beta_{21} = 2$, $\beta_{22} = -4$. In order to examine diagnostic power of our residuals when the mean structure is misspecificed, we do not include the quadratic term X^2 in the assumed model. Instead, we fit the model with only a linear term of X, i.e we assume the utilities $U_j = \beta_{j0} + \beta_{j1}X + \epsilon_j$, for $j = 1, 2, 3$. We obtain the surrogate residual $\boldsymbol{r} = [r_1, r_2, r_3]^T$ corresponding to the utilities 1, 2 and 3. The residual-by-covariate plots and QQ-plots of our surrogate residuals are shown in the left and right columns of Figure 2.

All the residual-by-covariate plots exhibit wave-like shapes, suggesting a misspecified mean structure. Specifically, we can see r_1 slants upward for $X \in (1, 3.5)$ and goes downward when $X > 3.5$. Similarly, r_2 seems to be positive for $X \in (-1, 1)$; r_3 negative when $X \in (-1, 3.5)$ and positive for $X > 3.5$. These patterns suggest that a linear form of X in the mean structure may not be enough, and a higher order term of X is needed. For comparison, we provide the residual-by-covariate plots of Pearson's and deviance residuals in Figure 3. Here, the Pearson's residual for multinomial logistic regression is given by $\boldsymbol{r}^{(P)} = [r_1^{(P)}, \ldots, r_J^{(P)}]$, with

$$r_j^{(P)} = (\tilde{y}_j - \hat{p}_j)/\sqrt{\hat{p}_j(1 - \hat{p}_j)}, \quad j = 1, \ldots, J,$$

and deviance residual $\boldsymbol{r}^{(D)} = [r_1^{(D)}, \ldots, r_J^{(D)}]$, with

$$r_j^{(D)} = \mathbb{I}(\tilde{y}_j - \hat{p}_j)\sqrt{2[(1 - \tilde{y}_j)\ln(1 - \hat{p}_j) - \tilde{y}_j \ln(\hat{p}_j)]}, \quad j = 1, \ldots, J,$$

where $\hat{p}_j$ is the estimated possibility of $Y = j$, and $\tilde{y}_j = 1$ if $Y = j$ and 0 otherwise. Both definitions are natural extension of the Pearson's and deviance residuals for the binary

For Peer Review Only

logistic regression. Similar to the surrogate residual, both are J-dimensional vectors corresponding to U_1 to U_J. As shown in Figure 3, both $\boldsymbol{r}^{(P)}$ and $\boldsymbol{r}^{(D)}$ detect a wave-like shape but indicate three separate parts. As a next step, we added a quadratic term of X. The updated surrogate residual plots are displayed in Figure S1 in the supplementary material. We can see that all plots exhibit no abnormal pattern.

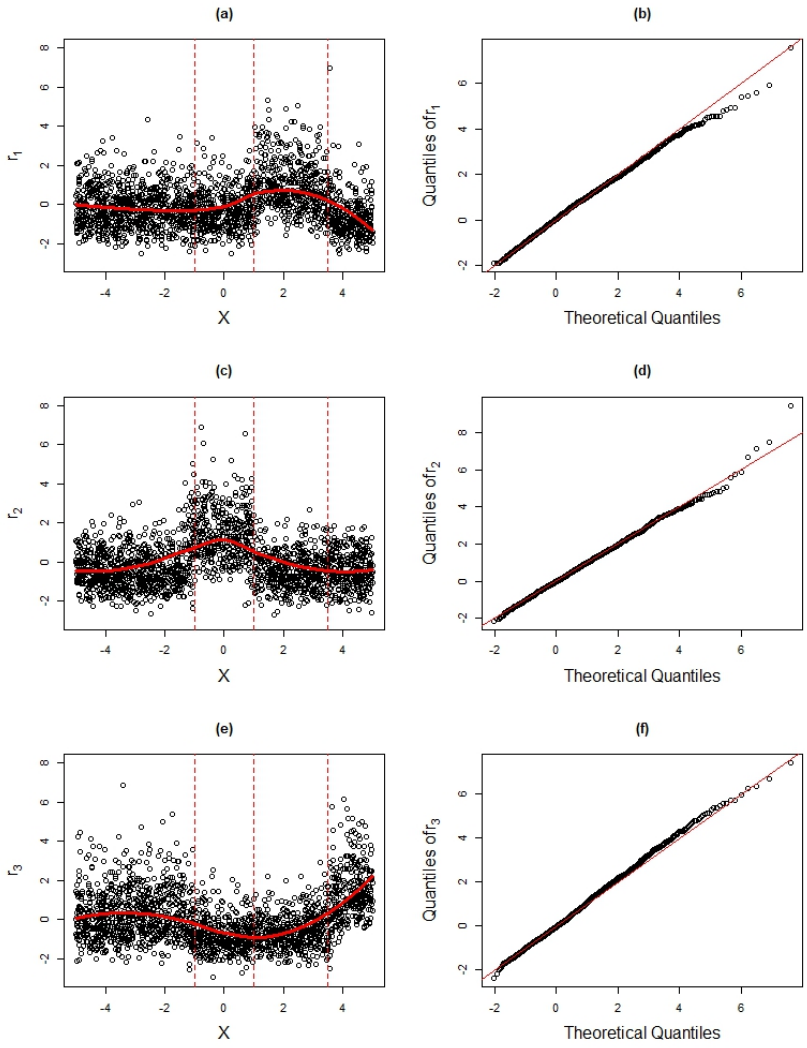

Figure 2: Model diagnostics when the mean structure in the assumed model is misspecified. Figures (a), (c) and (e) show the residuals-by-covariate plots of r_1, r_2 and r_3 respectively; vertical dashed lines $X = -1$, 1, and 3.5, and loess curves (red solid) are added as references. Figures (b), (d) and (f) are the corresponding QQ-plots (quantile-quantile plots) of above mentioned residuals.

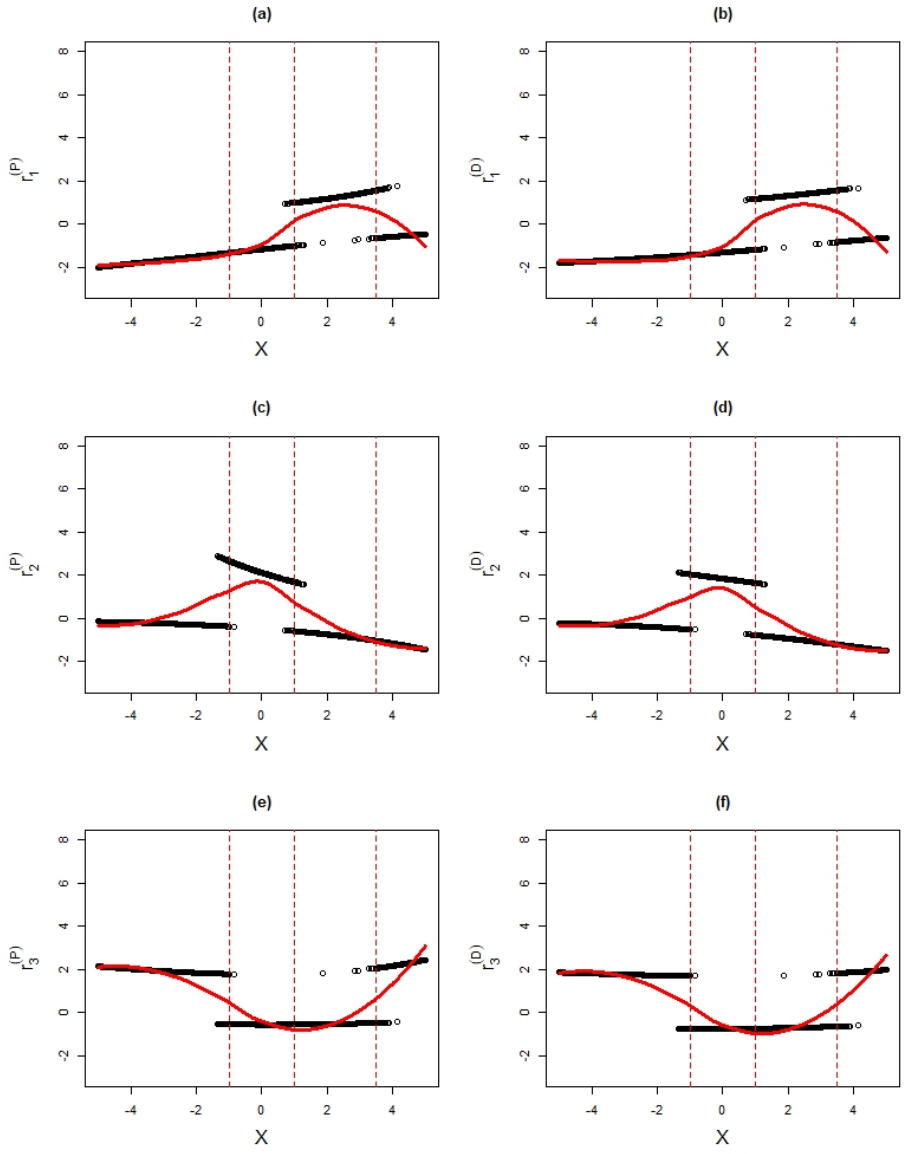

Figure 3: Model diagnostics when the mean structure in the assumed model is misspecified. The left and right columns display the residuals-by-covariate plots of Pearson's ($r_j^{(P)}$) and deviance ($r_j^{(D)}$) residuals, respectively. Vertical dashed lines $X = -1$, 1, and 3.5, and loess curves (red solid) are added as references.

3.3.2 Checking the individual-specific coefficients

The covariates and coefficients in the DCMs can be "individual-specific", i.e they are identical among all alternatives, or "alternative-specific", i.e they are different among al-

ternatives $1, \ldots, J$. Many empirical analyses of DCMs assume the coefficients of some alternative-specific variable are individual-specific, i.e., $\beta_{1k} = \beta_{2k} = \cdots = \beta_{Jk} =: \beta_k$ for alternative-specific variable $\tilde{\boldsymbol{X}}_k = [X_{1k}, \ldots, X_{Jk}]$. Such an assumption is adopted widely in practice to achieve parsimonious models. We demonstrate in the example below that our surrogate residual offers a simple way to check this assumption.

Example 3. We generated 2000 observations from a multinomial regression model with the following utility levels

$$U_j = \beta_{j0} + \beta_{j1} X_{j1} + \epsilon_j, \quad j = 1, 2, 3,$$

where ϵ_j follows a standard Gumbel distribution; $\boldsymbol{X}_1 = [X_{11}, X_{21}, X_{31}]^T$ are alternative-specific covariates independently generated from $Unif(-5, 5)$; $\beta_{10} = \beta_{20} = \beta_{30} = 0$ are the intercepts; $\beta_{11} = 10$, $\beta_{21} = 5$, $\beta_{31} = -10$ are alternative-specific coefficients. It is of interest to check if it is reasonable to assume $\beta_{11} = \beta_{21} = \beta_{31} =: \beta_1$. Now, we fit the observations based on an individual-specific coefficient, i.e., $U_j = \beta_{j0} + \beta_1 X_{j1} + \epsilon_j$ for $j = 1, 2, 3$.

If the true model contains an alternative-specific coefficient but the fitted model forces an individual-specific coefficient, we can detect this misspecification using the surrogate residuals. We simply draw r_j-by-x_{j1} plots, for $j = 1, 2, 3$, and detect if there is any trend inside these three plots. Or more intuitively, we simply draw s_j-by-x_{j1} plots, for $j = 1, 2, 3$, and compare the slopes. The slopes of s_j by x_{j1} for all $j = 1$, 2, 3 should be the same, or imply an alternative-specific β_1. Figure S2 in the supplementary material displays the surrogate variable versus covariate plots, which indicates a positive slope for s_1-by-x_{11} and s_2-by-x_{21} but a negative trend for s_3-by-x_{31}. Thus, an alternative-specific coefficient is detected. We also provide the surrogate-by-covariate plots of $\boldsymbol{r}^{(P)}$ and $\boldsymbol{r}^{(D)}$ in Figure S3 in the supplementary material. Here, the surrogate response for $\boldsymbol{r}^{(P)}$ and $\boldsymbol{r}^{(D)}$ are given by $s_j^{(P)} = V_j(\boldsymbol{X}_j, \hat{\boldsymbol{\beta}}_j) + r_j^{(P)}$ and $s_j^{(D)} = V_j(\boldsymbol{X}_j, \hat{\boldsymbol{\beta}}_j) + r_j^{(D)}$, $j = 1, \ldots, J$, respectively. All the scatter points of $s_j^{(P)}$ and $s_j^{(D)}$ have a clear positive trend but it is hard to detect the misspecified alternative-specific coefficient as they cluster in two strips.

For Peer Review Only

3.3.3 Detecting Interaction Effects

Another common issue in model development is to determine if there are important interactions between the covariates. We illustrate how to check whether or not such interaction effects are missing from the assumed model.

Example 4. We generated 2000 observations from a multinomial regression model with the following utility levels

$$U_j = \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \beta_{j3}X_1X_2 + \epsilon_j, \quad j = 1, 2, 3,$$

where ϵ_j follows a standard Gumbel distribution; $X_1 \sim Unif(-5, 5)$ and $X_2 \sim Bernoulli(0.5)$ are individual-specific variables; $\beta_{30} = \beta_{31} = \beta_{32} = \beta_{33} = 0$ are references; $\beta_{10} = -10$, $\beta_{11} = -5$, $\beta_{12} = 5$, $\beta_{13} = 3$; $\beta_{20} = -10$, $\beta_{21} = 5$, $\beta_{22} = -5$, $\beta_{23} = 6$. The indicator variable X_2 can be seen as a factor with levels "Treatment" and "Control."

To check if there are any interaction effects between X_1 and X_2, we just fit two models: The first model corresponds to the simulated control group, while the second model corresponds to the treatment group. Both models assume $U_j = \beta_{j0} + \beta_{j1}X_1 + \epsilon_j$ for $j = 1, 2, 3$. The surrogate response versus x_1 plots can help us detect if there are any interaction effects between X_1 and X_2. If the true model does not contain an interaction term X_1X_2, the slopes of s_j $(j = 1, 2, 3)$ by x_1 between the control and treatment groups should be exactly the same, otherwise indicating an interaction effect term between X_1 and X_2. In our surrogate-by-covariates plots (see Figure S4 in the supplementary material), the control group displays a stronger negative association between s_1 and x_1 comparing with the treatment group; but the treatment group shows a stronger positive association between s_2 and x_1 comparing with the control group. Therefore, the interaction effect should be considered based on our surrogate residuals. As shown in Figures S5 and S6 in the supplementary material, the surrogate responses of Pearson's and deviance residuals exhibit strange patterns and cannot be used to detect interaction terms.

3.4 Practical Remarks

In this subsection, we give two practical remarks related to our surrogate residuals.

For Peer Review Only

3.4.1 Normalized Surrogate Residual

From Section 3.2, we see that the null distribution of our surrogate residual depends on the distribution of the random part $\boldsymbol{\epsilon}$. Specifically, $\boldsymbol{R} \sim G(\boldsymbol{R} + \int \boldsymbol{\epsilon} dG(\boldsymbol{\epsilon}))$ if the assumed model holds. Lack of knowledge for the distribution $G(.)$ may lead to difficulties in practice. For instance, the null distribution of the surrogate residual for multinominal logistic regression is a Gumbel distribution, which is slightly positive skewed. Our inference may be biased if this positive-skewness is not considered. Here, for the convenience of the residual diagnosis, we may normalize the surrogate residuals to the standard normal distribution. Specifically, given the surrogate residual $\boldsymbol{r} = [r_1, \ldots, r_J]^T$, the normalized surrogate residual $\boldsymbol{r}^{(N)} = [r_1^{(N)}, \ldots, r_J^{(N)}]^T$ can be obtained by

$$r_j^{(N)} = \Phi^{-1}\big(G_j(r_j + m_j)\big) \quad j = 1, \ldots, J,$$

where $G_j(.)$ is the marginal distribution function of j-th element of $G(.)$, $m_j = \int \epsilon_j dG_j(\epsilon_j)$, and $\Phi^{-1}(.)$ denotes the inverse cumulative distribution function of the standard normal distribution. From Theorem 2, we can immediately derive that, if the assumed model agrees with the true model (i.e $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}, \boldsymbol{V} = \tilde{\boldsymbol{V}}, G = \tilde{G}$), every element $r_j^{(N)}$ of our normalized surrogate residual will follow a standard normal distribution.

The normalized surrogate residual is similar to the randomized quantile residual proposed by Dunn and Smyth (1996) who transformed the original residual to follow a standard normal distribution under the null hypothesis in generalized linear models with a continuous, binary or count response variable. In fact, for DCMs with a binary outcome where the choice set contains 2 alternatives, our proposed normalized surrogate residual is the same as the randomized quantile residual.

As a numerical illustration, Figure S7 in the supplementary material displays the residual-by-covariates and QQ-plots of our normalized residuals for the improved model of Example 2, and we can see the residual symmetrically distributed around 0, in comparison to the slightly positive-skewed patterns in Figure S1.

For Peer Review Only

3.4.2 Bootstrapping Residuals

The patterns of our surrogate residuals contain two sources of uncertainties: (1) modeling error and (2) sampling. The latter is introduced by conditional sampling from $\boldsymbol{S}$. The sampling uncertainty is even more important when the sample size is small. Therefore, we also recommend bootstrapping K copies of the empirical distributions of the residuals to reduce the sampling uncertainty for a small sample size. For more details about this bootstrapping approach, see Section 5 in Liu and Zhang (2018).

For comparison, We provide the residual-by-covariate plots of the surrogate residual ($\boldsymbol{r}$) and bootstrapping residual with $K = 10$ ($\boldsymbol{r}^{(B)}$) in the setting of Example 2 when the model is specified correctly, with a small sample size of 50. Specifically, three replications of $\boldsymbol{r}$ and $\boldsymbol{r}^{(B)}$ are realized, and each replication of r_1-by-x and $r_1^{(B)}$-by-x plots is presented in Supplementary Material Figure S8. We omit the second and third dimensions of the residuals because their patterns are similar. It is shown that every replication of the surrogate residual exhibits different shapes. In particular, the second and third replications provide a wave-like shape, which can be a false alarm of a misspecified mean structure. However, all the replications of the bootstrapping residuals are randomly distributed around 0, exhibiting no abnormal patterns.

3.5 Simulation Results under Small Sample Scenarios

In this subsection, a Monte Carlo simulation study with three scenarios were performed to investigate the finite-sample properties of Pearson, deviance and surrogate residuals. All three scenarios consider a multinomial logistic regression with a 3-alternative response $Y = \{1, 2, 3\}$, with the following utility levels

$$U_{ij} = \beta_{j0} + \beta_{j1} x_i + \epsilon_{ij}, \quad j = 1, 2, 3; i = 1, \ldots, n,$$

where x_i followed $Unif(0, 1)$, and β_{30} and β_{31} were fixed to zero as the reference. In Scenario I, $\beta_{10} = 0.8$, $\beta_{11} = -2$, $\beta_{21} = -1.2$ and $\beta_{22} = 2$, which resulted in the empirical probabilities of $Y = 1, 2, 3$ close to 33%, 33% and 34%, respectively. In Scenario II, $\beta_{10} = 1.6$, $\beta_{11} = -1.5$, $\beta_{21} = 2.5$ and $\beta_{22} = -5.5$, which led to the empirical probabilities of

For Peer Review Only

$Y = 1, 2, 3$ close to 50%, 25% and 25%, respectively. In Scenario III, $\beta_{10} = 1.1$, $\beta_{11} = 1.1$, $\beta_{21} = -2.2$ and $\beta_{22} = 4$, which yielded the empirical probabilities of $Y = 1, 2, 3$ close to 70%, 15% and 15%, respectively. All results were based on 10,000 Monte Carlo replications and $n = 20$. For each scenario, the covariate x_i's ($i = 1, \ldots, 20$) were held as constants over all replications. In each of the 10,000 replications, we fit the correct model and compute the Pearson's residual $r_{ij}^{(P)}$, deviance residual $r_{ij}^{(D)}$ and surrogate residual $r_{ij}^{(S)}$, for $i = 1, \ldots, 20$ and $j = 1, 2, 3$. Because the baseline for $r_{ij}^{(P)}$ and $r_{ij}^{(D)}$ were standard normal, but the null distribution of $r_{ij}^{(S)}$ was Gumbel, for easy comparison, we transformed $r_{ij}^{(S)}$ to the normalized surrogate residual, denoted by $r_{ij}^{(N)}$. Therefore, all $r_{ij}^{(P)}$, $r_{ij}^{(D)}$, $r_{ij}^{(N)}$ had a standard normal baseline.

Tables S1, S2 and S3 in the supplementary material present the empirical mean, variance, skewness and kurtosis coefficients for $r_{ij}^{(P)}$, $r_{ij}^{(D)}$, $r_{ij}^{(N)}$ in Scenario I, II, III, respectively. If the residual approximately follows the standard normal distribution, the average mean, variance, skewness and kurtosis coefficients should be close to 0, 1, 0 and 3, respectively. In all three scenarios, we can see all four characteristics of the normalized surrogate residual are close to their theoretical values, suggesting the normality of the residuals. The mean and variance of the Pearson's residual are also close to 0 and 1, respectively, but its skewness and kurtosis coefficients are far away from their limiting value, indicating that it may not be well approximated by the standard normal distribution. The deviance residual performed the worst compared with other residuals, as its characteristics were far away from its theoretical values. Tables S1 to S3 also provide the value of the Kolmogorov–Smirnov test (K-S test) statistic and p-value used to test whether the residual follows the standard normal distribution. Under the 5% significance level, the Pearson's and deviance residuals in all three scenarios were rejected from the standard normality hypothesis. However, most of K-S test p-values from the normalized surrogate residuals is greater than 5%, indicating the reasonableness of standard normality.

For Peer Review Only

4 Residual for General Models with a Categorical Outcome

In this section, we show that surrogate residuals can be expanded to define residuals for general statistical models with a categorical outcome. Suppose that the assumed model for a categorical outcome $Y \in \{1, \ldots, J\}$ is

$$Pr(Y = y) = F_a(y; \boldsymbol{X}, \boldsymbol{\beta}) \quad y = 1, 2, \ldots, J, \tag{7}$$

where $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are covariates and unknown parameters, and $F_a(.)$ is a nominal distribution function with J categories satisfying $F_a(j; \boldsymbol{X}, \boldsymbol{\beta}) \geq 0 \quad (\forall j \in \{1, \ldots, J\})$ and $\sum_{j=1}^{J} F_a(j; \boldsymbol{X}, \boldsymbol{\beta}) = 1$. This model is broad enough to cover nearly all regression models with categorical outcomes. For model (7), we can define the hypothetical variable $\boldsymbol{Z} = [Z_1, \ldots, Z_J]^T$ as

$$Z_j = \ln\big(F_a(j; \boldsymbol{X}, \boldsymbol{\beta})\big) + \epsilon_j \quad j = 1, 2, \ldots, J,$$

where $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_J]^T$ denotes a J-dimensional noise variable following a multivariate Gumbel distribution with c.d.f $MG(\boldsymbol{\epsilon}) = \exp(-\sum_{j=1}^{J} e^{-\epsilon_j})$. It is evident that ϵ_j's are mutually independent variables following a univariate standard Gumbel distribution. For simplicity, we abbreviate $\ln\big(F_a(j; \boldsymbol{X}, \boldsymbol{\beta})\big)$ as μ_j henceforth. Now, we need to build up a joint distribution of Y and $\boldsymbol{Z}$. Specifically, after derivation (see Appendix C in the supplementary material for more detail), we can show that

$$\Pr(Z_y > Z_j, \forall j \neq y) \equiv F_a(y; \boldsymbol{X}, \boldsymbol{\beta}) \quad \forall y = 1, \ldots, J.$$

Thus the joint distribution can be determined by setting $Y \triangleq y$ if and only if $Z_y > Z_j (\forall j \neq y)$. Now, let the surrogate variable $\boldsymbol{S}$ be the conditional distribution of $\boldsymbol{Z}$ given Y. More specifically, $\boldsymbol{S} \sim \boldsymbol{Z}|(Y, \boldsymbol{X})$ follows a truncated distribution obtained by truncating the distribution of $\boldsymbol{Z}|\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$ using the region $Z_Y > Z_j (\forall j \neq Y)$, where $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_J]^T$. Finally, we can define the surrogate residual $\boldsymbol{R}$ as

$$\boldsymbol{R} = \boldsymbol{S} - E_0(\boldsymbol{S}|\boldsymbol{X}) = \boldsymbol{S} - E(\boldsymbol{Z}|\boldsymbol{X}) = \boldsymbol{S} - \boldsymbol{\mu} - \int \boldsymbol{\epsilon} dMG(\boldsymbol{\epsilon}), \tag{8}$$

For Peer Review Only

where $\int \boldsymbol{\epsilon} dMG(\boldsymbol{\epsilon}) \approx \mathbf{0.5772}$ is the expectation of $MG(.)$. In practice, given the observed data $\{\boldsymbol{x}, y\}$ and a fitted model $F_a(y; \boldsymbol{X}, \hat{\boldsymbol{\beta}})$, we draw a sample $\boldsymbol{s}$ from $\boldsymbol{S}$ by plugging in the estimated parameter $\hat{\boldsymbol{\beta}}$, then let $\boldsymbol{r} = \boldsymbol{s} - \hat{\boldsymbol{\mu}} - \mathbf{0.5772}$ be the surrogate residual. If the assumed model agrees with the true model, say the true model for Y is $\tilde{F}(y; \boldsymbol{X}, \tilde{\boldsymbol{\beta}})$, the theorem below summarizes the properties of $\boldsymbol{R}$.

Theorem 5. *If the assumed model (7) agrees with the underlying true model (i.e, $F_a = \tilde{F}$, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$), the following results hold*

(1) The surrogate variable $\boldsymbol{S}$ has the same distribution as $\boldsymbol{Z}$, i.e $\boldsymbol{S} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ follows the distribution $MG(\boldsymbol{\epsilon})$.

(2) The residual variable $\boldsymbol{R}$, independent from $\boldsymbol{X}$, follows the distribution $MG(\boldsymbol{r}+\boldsymbol{m})$, i.e $Pr(\boldsymbol{R} \leq \boldsymbol{r}) = Pr(\boldsymbol{R} \leq \boldsymbol{r}|\boldsymbol{X}) = MG(\boldsymbol{r} + \boldsymbol{m})$, where $\boldsymbol{m} = \int \boldsymbol{\epsilon} dMG(\boldsymbol{\epsilon}) \approx \mathbf{0.5772}$. As a result, $E(\boldsymbol{R}|\boldsymbol{X}) = \mathbf{0}$, and $Var(\boldsymbol{R}|\boldsymbol{X}) = Var(\boldsymbol{\epsilon}) \approx 1.6449 \times I_{J\times J}$, where $I_{J\times J}$ denotes a $J \times J$ identity matrix.

Proof of Theorem 5 is similar to the one discussed in Section 3.2. Theorem 5(2) shows that the residual defined in (8) is similar to the residual defined for DCMs, has zero mean, homogeneous variance, and an explicit reference distribution. Therefore, it can be used for model diagnostics in a similar way to the surrogate residual defined for DCMs.

Remark 3. *For multinominal logistic regression, the residual defined in (8) is exactly equal to the residual given in Section 3.1. This reveals an interesting connection of the surrogate residual for general models with a categorical outcome to that for DCMs.*

5 MPN Data Analysis

We present a real data analysis to illustrate how to perform residual diagnosis in practice using the Netherlands Mobility Panel (MPN) data. The data were collected to investigate the travel behavior of Dutch residents. Since the MPN is a mobility panel, multiple waves are available since 2013. As an illustrative example, we only make use of the data from the first wave, which includes a total of 35,257 travel records from 5,297 individuals. See Hoogendoorn-Lanser et al. (2015) for more details.

For Peer Review Only

For travel behavior analysis, we focus on a categorical outcome containing six different travel modes (car as driver, car as passenger, bicycle, ebike, public transport, and walking). We use the discrete choice model to analyze the travel modes, and focus on the following covariates: travel distance (kilometers), departure time (day time, night time), driving license (with, without), gender, age, education (low, medium, high), work situation, and household income (low, medium, rich). All the covariates are individual-specific variables. Now, we illustrate how to use our surrogate residual to diagnose and improve model fitting.

Used in our initial analysis is a multinominal logistic regression, including all above-mentioned covariates in linear terms. We first investigate the association between the travel distance and travel mode choice, since the distance from the destination exerts direct and may be the most important influence to people's decision on which kind of transport to choose. Figure S9 in the supplementary material plots residual versus travel distance (we take travel distance under the logarithmic coordinate for better display effect). We see that the mean of scatter points in plots (a), (e) and (f), corresponding to the utilities of car as driver, public transport and walking, show a wave shape. Especially the points in Figure S9(a) have a positive mean shift to the right of the vertical dashed line; on the contrary, the points in Figure S9(e) have a negative mean shift in the same interval. Plus, plot (c) of Figure S9 reveals a wave shape, implying a misspecified mean structure. All these characteristics indicate that a linear form of travel distance is inadequate. We therefore tried using a natural logarithm form, quadratic form and piecewise linear form of travel distance in our multinominal logistic model, and found that the natural logarithm form could most effectively alleviate the mean shift. Therefore, we replaced the linear form of distance with a natural logarithm form from subsequent analysis. The updated residual-by-log(distance) plots are shown in Figure S10, which effectively smooth out the mean shifts. In fact, our updated model improved McFadden R-squared from 0.2581 to 0.3067. Moreover, a likelihood ratio test with a p-value less than 2.2×10^{-16} also suggests that a logarithm term of distance resulted in a better fit of the data.

Our further examination reveals that there exists an interaction effect between the driving license and travel distance. Specifically, we fit two models using the updated mean

For Peer Review Only

structure based on two groups. The first group is the subjects who have a driving license, while the second group is the subjects who do not have a driving license. We draw the plots of surrogate response versus log(distance) for above two groups in Figure S11. If the true model does not contain an interaction term log(distance)×license, the slopes of the surrogate responses by log(distance) should be all the same between these two data groups. However, our plots of surrogate responses by log(distance) present different slopes. Therefore, we introduce a new interaction term log(distance)×license to our model. This adjustment slightly improves McFadden R-squared from 0.3067 to 0.3073. The likelihood ratio test for including this interaction effect compared to the former model yields a p-value of 2.0×10^{-12}.

6 Discussion

In this article, we proposed the surrogate variables and residuals for the discrete choice models. We defined the surrogate residual, and comprehensively analyzed its theoretical and graphical properties. Numerical examples in a variety of settings confirm that our residual can be informative and useful in detecting misspecified mean structure, checking individual-specific coefficients, and detecting interaction effects. Our surrogate approach overcomes the problems in the Pearson's, deviance and other nonrandomized residuals, and enriches the diagnostic methods in DCMs. We extend our residuals for more general models with an unordered categorical response. Finally, our analysis of the MPN data confirms that our surrogate residual can help us improve the model fitting in practice.

Acknowledgments

Zhang's research was partially supported by grants NIH R01 MH116527 from National Institute of Mental Health and NSF DMS-1722544 from the National Science Foundation of United States of America.

For Peer Review Only

Supplementary Material

Appendices and supplementary figures: "`residual_DCM_supp.pdf`" presents appendices (Appendices A, B and C) and supplementary tables and figures unshown in the manuscript.

R-code for numerical examples: The R code for the numerical examples are shown at "`residual_DCM_codes.R`".

MPN data: The real data set used in Section 5 is available at `https://www.mpndata.nl`.

References

Ben-Akiva, M. E. and S. R. Lerman (1985). *Discrete choice analysis: theory and application to travel demand*, Volume 9. MIT press.

Chan, Y. H. (2005). Biostatistics 305. multinomial logistic regression. *Singapore Medical Journal 46*(6), 259–269.

Dunn, P. K. and G. K. Smyth (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics 5*(3), 236–244.

Goeman, J. J. and S. le Cessie (2006). A goodness-of-fit test for multinomial logistic regression. *Biometrics 62*(4), 980–985.

Gupta, A. K., T. Nguyen, and L. Pardo (2008). Residuals for polytomous logistic regression models based on φ-divergences test statistics. *Statistics 42*(6), 495–514.

Hausman, J. A. and D. McFadden (1981). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, 1219–1240.

Hoogendoorn-Lanser, S., N. T. Schaap, and M. J. OldeKalter (2015). The netherlands mobility panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation Research Procedia 11*, 311–329.

For Peer Review Only

Koppelman, F. S. and C. Bhat (2006). A self instructing course in mode choice modeling: multinomial and nested logit models. OnlineDoc: http://www.caee.utexas.edu/prof/bhat.

Liu, D. G. and H. P. Zhang (2018). Residuals and diagnostics for ordinal regression models: A surrogate approach. *Journal of the American Statistical Association 113*(522), 845–854.

Lovreglio, R., D. Borri, D. Luigi, and A. Ibeas (2014). A discrete choice model based on random utilities for exit choice in emergency evacuations. *Safety Science 62*, 418–426.

McFadden, D. (1978). Microarray data analysis: from disarray to consolidation and consensus. *Transportation Research Board 673*, 72–77.

Nagel, H. and R. Hatzinger (1992). Diagnostics in some discrete choice models. *Computational Statistics & Data Analysis 13*(4), 407–414.

Pagan, A. and F. Vella (1989). Diagnostic tests for models based on individual data: a survey. *Journal of Applied Econometrics 4*(S1), S29–S59.

Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics*, 230–239.

For Peer Review Only

Supplymentary Material: Surrogate Residuals for Discrete Choice Models

Chao Cheng

Department of Mathematical Sciences, Tsinghua University
Rui Wang
Department of Public Economics, School of Economics, Xiamen University
and
Heping Zhang*
Department of Biostatistics, School of Public Health, Yale University

August 26, 2019

This file presents the appendices (Appendices A, B and C) and supplementary tables and figures unshown in the manuscript.

Appendix A: Generating Samples from Truncated Multi-dimensional Distribution

In this section, we introduce a Gibbs method to generate samples from truncated multi-dimensional distribution like $\boldsymbol{S}$ in section 3.1. Before the multi-dimensional case, we at first present the method to generate truncated univariate distribution, because the ability to generate univariates is a central building block in the Gibbs method.

1. The univariate case

Let X follow a distribution G truncating on the interval (a, b), where G is a univariate continuous cumulative distribution function, a is the left truncated point and b is the right truncated

*Address for correspondence: Heping Zhang, Department of Biostatistics, School of Public Health, Yale University, New Haven, CT, USA. Email: heping.zhang@yale.edu

For Peer Review Only

point. The two truncated points are arbitrary but fixed on $\mathbb{R}$ and $-\infty \leq a < b \leq +\infty$. Then, the distribution function of X is

$$F(x) = \frac{G(x) - G(a)}{G(b) - G(a)} \quad a \leq x \leq b.$$

Based on the process of *inverse CDF method*, solving the equation $F(x) = u$, $u \in (0, 1)$, we get

$$x = G^{-1}\Big(u\big(G(b) - G(a)\big) + G(a)\Big).$$

Then, we simulate X as follows:

Step 1: Generate a sample u from Unif$(0, 1)$, where Unif$(0, 1)$ denotes the standard uniform distribution.

Step 2: Return $X = G^{-1}\Big(u\big(G(b) - G(a)\big) + G(a)\Big)$. Then, X follows the target distribution.

2. The multivariate case

Now, we propose a Gibbs sampling approach for drawing samples from a truncated multi-dimensional density like $\boldsymbol{S}$ in section 3.1. Gibbs sampling is a Markov chain Monte Carlo (MCMC) method for generation of samples from high dimensional densities by drawing the samples from full conditional densities. In other words, to implement the Gibbs sampling for an J-dimensional vector, we need to find full conditionals of its subvectors.

Let $\boldsymbol{X} = [X_1, \ldots, X_J]^T$ denote a variable that follows a J-dimensional distribution $G(x_1, \ldots, x_J)$ using the support area $\mathcal{D}_k$, where $G(x_1, \ldots, x_J)$ is a continuous cumulative distribution function and $\mathcal{D}_k$ is a support area in $\mathbb{R}^J$ such as $X_k > X_j$, $\forall j \neq k$. Denote by $\boldsymbol{x} = (x_1, x_2, \ldots, x_J)^T$ a sample from $\boldsymbol{X}$. We shall denote the c.d.f of the full conditionals $X_j|X_{-j}$, $j = 1, 2, \ldots, J$, where $X_{-j} = (X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_J)^T$, by $G_{X_j|X_{-j}}(x_j)$. The conditional density $X_j|X_{-j}$ is

$$F_{X_j|X_{-j}}(x_j|x_{-j}) = \frac{G_{X_j|X_{-j}}(x_j|x_{-j}) - G_{X_j|X_{-j}}(a|x_{-j})}{G_{X_j|X_{-j}}(b|x_{-j}) - G_{X_j|X_{-j}}(a|x_{-j})},$$

For Peer Review Only

where

$$\begin{cases} a = \max x_{-j}, b = +\infty & \text{if } j = k \\ a = -\infty, b = x_k & \text{if } j \neq k \end{cases}$$

and

$$G_{X_j|X_{-j}}(x_j|x_{-j}) = \int_{-\infty}^{x_j} \frac{g(x_1, \ldots, x_{j-1}, v_j, x_{j+1}, \ldots, x_J)}{\int_{-\infty}^{+\infty} g(x_1, \ldots, x_{j-1}, v_j, x_{j+1}, \ldots, x_J) dv_j} dv_j.$$

Here, $g(.)$ is the corresponding p.d.f of $G(.)$. The distribution above is nothing but a truncated univariate distribution whose samples can be generated by the previous subsection. Now we present the Gibbs sampling scheme that will be used to produce samples from $\boldsymbol{X}$. First, we get initial values $\{x_k^{(0)}\}_{k=2}^J$ from the support area $\mathcal{D}_k$, and then for $i = 1, 2, \ldots, M$ generate samples according to

$$\begin{aligned} X_1|x_2^{(i-1)}, x_3^{(i-1)}, \ldots, x_J^{(i-1)} &\longrightarrow x_1^{(i)} \\ X_2|x_1^{(i)}, x_3^{(i-1)}, \ldots, x_J^{(i-1)} &\longrightarrow x_2^{(i)} \\ \ldots & \\ X_{J-1}|x_1^{(i)}, \ldots, x_{J-2}^{(i)}, x_J^{(i-1)} &\longrightarrow x_{J-1}^{(i)} \\ X_J|x_1^{(i)}, x_2^{(i)}, \ldots, x_{J-1}^{(i)} &\longrightarrow x_J^{(i)} \end{aligned}$$

where the arrows represent sampling from the corresponding distribution, and each x_k is sampled according to the univariate scheme. We should discard the first B samples, where B is a given number in advance, which are called burn-in samples, and the remaining samples are considered as the samples from $\boldsymbol{X}$.

Appendix B: Proofs for Theorems 3 and 4

In this section, we prove Theorems 3 and 4. First, we derive the distribution of $\hat{\boldsymbol{S}}$ that follows the truncated distribution obtained by truncating the distribution of $\hat{\boldsymbol{U}} = \hat{\boldsymbol{V}}(\boldsymbol{X}, \hat{\boldsymbol{\beta}}) + \boldsymbol{\epsilon}$ using the area $\{\hat{U}_y > \hat{U}_j, \forall j \neq y\}$, denoted by $\hat{\mathcal{D}}_y$ henceforth, given $Y = y$. Noting that $\{\hat{U}_y > \hat{U}_j, \forall j \neq y\}$ only represents the order information, we have $\hat{\mathcal{D}}_y \equiv \mathcal{D}_y$, whether $\hat{\boldsymbol{\beta}}$ is a consistent estimate or not. When $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + o_p(1)$ and under the assumption that $\boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta})$ is a continuous function, we have $\boldsymbol{V}(\boldsymbol{X}, \hat{\boldsymbol{\beta}}) = \boldsymbol{V}(\boldsymbol{X}, \boldsymbol{\beta}) + o_p(1)$, thus $\hat{\boldsymbol{U}} = \boldsymbol{U} + o_p(1)$. Therefore, for all continuity sets $\mathcal{B} \in \mathbb{R}^J$, we have $Pr(\hat{\boldsymbol{U}} \in \mathcal{B}) = Pr(\boldsymbol{U} \in \mathcal{B}) + o(1)$, which implies $Pr(\hat{\boldsymbol{U}} \in \mathcal{C} \cap \mathcal{D}_y) = Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y) + o(1)$

For Peer Review Only

and $Pr(\hat{\boldsymbol{U}} \in \mathcal{D}_y) = Pr(\boldsymbol{U} \in \mathcal{D}_y) + o(1)$. Now, conditional on $\boldsymbol{X}$, for any arbitrary but fixed $\boldsymbol{c} = [c_1, \ldots, c_J]$,

$$\begin{aligned}
& |Pr(\hat{\boldsymbol{S}} \le \boldsymbol{c}) - Pr(\boldsymbol{S} \le \boldsymbol{c})| \\
= & \left| \sum_{y=1}^{J} Pr(\hat{\boldsymbol{S}} \le \boldsymbol{c}|Y=y)Pr(Y=y) - \sum_{y=1}^{J} Pr(\boldsymbol{S} \le \boldsymbol{c}|Y=y)Pr(Y=y) \right| \\
\le & \sum_{y=1}^{J} \left| Pr(\hat{\boldsymbol{S}} \le \boldsymbol{c}|Y=y) - Pr(\boldsymbol{S} \le \boldsymbol{c}|Y=y) \right| Pr(Y=y) \\
= & \sum_{y=1}^{J} \left| \frac{Pr(\hat{\boldsymbol{U}} \in \mathcal{C} \cap \hat{\mathcal{D}}_y)}{Pr(\hat{\boldsymbol{U}} \in \hat{\mathcal{D}}_y)} - \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\boldsymbol{U} \in \mathcal{D}_y)} \right| Pr(Y=y) \\
= & \sum_{y=1}^{J} \left| \frac{Pr(\hat{\boldsymbol{U}} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\hat{\boldsymbol{U}} \in \mathcal{D}_y)} - \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\boldsymbol{U} \in \mathcal{D}_y)} \right| Pr(Y=y) \\
= & \sum_{y=1}^{J} \left| \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y) + o(1)}{Pr(\boldsymbol{U} \in \mathcal{D}_y) + o(1)} - \frac{Pr(\boldsymbol{U} \in \mathcal{C} \cap \mathcal{D}_y)}{Pr(\boldsymbol{U} \in \mathcal{D}_y)} \right| Pr(Y=y) \\
= & \; o(1).
\end{aligned}$$

Thus, we can establish $Pr(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}} \le \boldsymbol{r}|\boldsymbol{X}) = Pr(\boldsymbol{R} \le \boldsymbol{r}|\boldsymbol{X}) + o(1)$. The unconditional distribution of $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}$

$$Pr(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}} \le \boldsymbol{r}) = \int Pr(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}} \le \boldsymbol{r}|\boldsymbol{x}) d\mu(\boldsymbol{x}) \to \int Pr(\boldsymbol{R} \le \boldsymbol{r}|\boldsymbol{x}) d\mu(\boldsymbol{x}) = Pr(\boldsymbol{R} \le \boldsymbol{r}),$$

based on Lebesgue's Dominated Convergence Theorem. Now, Theorem 3 has been proved. Moreover, based on Theorem 3, we have $E(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}|\boldsymbol{X}) = E(\boldsymbol{R}|\boldsymbol{X}) + o(1) = o(1)$ and $Var(\hat{\boldsymbol{R}}_{\hat{\boldsymbol{\beta}}}|\boldsymbol{X}) = Var(\boldsymbol{R}|\boldsymbol{X}) + o(1)$. This completes the proof of Theorem 4.

Appendix C: Proof for $Pr(Z_y > Z_j; \forall j \neq y) = F_a(y; \boldsymbol{X}, \boldsymbol{\beta})$

We now prove that under the assumed model (4.11), $Pr(Z_y > Z_j; \forall j \neq y) = F_a(y; \boldsymbol{X}, \boldsymbol{\beta})$. At first, abbreviating $Pr(Z_y > Z_j; \forall j \neq y)$ as p_y, we have

$$\begin{aligned}
p_y &= \Pr\big(Z_y > Z_j, \forall j \neq y\big) \\
&= \Pr\big(\mu_y + \epsilon_y > \mu_j + \epsilon_j, \forall j \neq y\big) \\
&= \Pr\big(\epsilon_j < \mu_y - \mu_j + \epsilon_y, \forall j \neq y\big).
\end{aligned}$$

If ϵ_y is considered given, this equation is the cumulative density function for each ϵ_j evaluated at $\mu_y - \mu_j + \epsilon_y$, according to the c.d.f of standard Gumbel distribution, which is $\exp\big(-\exp(-(\mu_y - \mu_j + \epsilon_y))\big)$. Since all ϵ_y are independent, the cumulative distribution over all $y \neq j$ is the product

For Peer Review Only

of the individual cumulative distributions

$$p_y|\epsilon_y = \prod_{j\neq y} \exp\big(-\exp(-(\mu_y - \mu_j + \epsilon_y))\big).$$

But in this question, ϵ_y is unknown, so the probability p_y is the integral of $p_y|\epsilon_y$ weighted by the density of ϵ_y

$$p_y = \int \Big(\prod_{j\neq y} e^{-\exp(-(\mu_y-\mu_j+\epsilon_y))}\Big) e^{-\epsilon_y - \exp(-\epsilon_y)} \mathrm{d}\epsilon_y. \tag{1}$$

Mark ϵ_y as t and note that $\mu_y - \mu_y = 0$. Thus, collecting the terms in the exponent of e, we have

$$\begin{aligned} p_y &= \int_{t=-\infty}^{+\infty} \Big(\textstyle\prod_j e^{-\exp(-(\mu_y-\mu_j-t))}\Big) e^{-t}\mathrm{d}t \\ &= \int_{t=-\infty}^{+\infty} \exp\Big(-\textstyle\sum_j e^{-\mu_y+\mu_j+t}\Big) e^{-t}\mathrm{d}t \\ &= \int_{t=-\infty}^{+\infty} \exp\Big(-e^{-t}\textstyle\sum_j e^{-\mu_y+\mu_j}\Big) e^{-t}\mathrm{d}t. \end{aligned}$$

Define $s = e^{-t}$ such that $e^{-t}\mathrm{d}t = -\mathrm{d}s$, using the new term, we have

$$\begin{aligned} p_j &= \int_0^{+\infty} \exp\Big(-s\textstyle\sum_j e^{-\mu_y+\mu_j}\Big)\mathrm{d}t = \left(\frac{\exp\left(-s\sum_j e^{-\mu_y+\mu_j}\right)}{-\sum_j e^{-\mu_y+\mu_j}}\right)_0^{+\infty} \\ &= \frac{\exp(\mu_y)}{\sum_{j=1}^J \exp(\mu_j)} = \frac{F_a(y;\boldsymbol{X},\boldsymbol{\beta})}{\sum_{j=1}^J F_a(j;\boldsymbol{X},\boldsymbol{\beta})} \\ &= F_a(y;\boldsymbol{X},\boldsymbol{\beta}). \end{aligned}$$

This completes the proof.

Supplementary Tables and Figures

For Peer Review Only

Table 1: Comparison of Pearson's residual ($r_{ij}^{(P)}$), deviance residual ($r_{ij}^{(D)}$), normalized surrogate residual ($r_{ij}^{(N)}$), for Scenario I.

j	i	mean			variance			skewness			kurtosis			K-S test		
		$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$
1	1	0.02	-0.06	-0.00	1.01	1.23	1.02	0.59	0.47	0.04	1.93	1.45	3.01	0.230(0.000)	0.284(0.000)	0.009(0.456)
	2	-0.03	-0.17	-0.02	0.72	0.52	0.97	3.09	2.29	-0.01	13.06	7.37	3.09	0.400(0.000)	0.400(0.000)	0.018(0.003)
	3	-0.00	-0.15	-0.02	0.77	0.56	1.00	2.85	2.15	0.02	11.11	6.56	2.90	0.388(0.000)	0.388(0.000)	0.014(0.039)
	4	-0.02	-0.03	-0.02	0.99	1.26	1.00	0.07	0.06	-0.01	1.71	1.28	2.98	0.170(0.000)	0.226(0.000)	0.009(0.430)
	5	-0.00	-0.02	0.00	1.01	1.28	1.01	0.13	0.11	-0.03	1.67	1.27	2.95	0.180(0.000)	0.238(0.000)	0.009(0.453)
	6	-0.00	-0.06	-0.00	1.02	1.25	1.00	0.52	0.42	-0.03	1.86	1.41	2.93	0.230(0.000)	0.285(0.000)	0.006(0.894)
	7	-0.02	-0.00	-0.00	0.99	1.23	1.01	-0.15	-0.10	0.05	1.89	1.33	2.90	0.151(0.000)	0.206(0.000)	0.009(0.401)
	8	0.01	-0.14	0.01	0.89	0.78	0.99	1.99	1.59	-0.03	6.05	4.05	2.96	0.332(0.000)	0.332(0.000)	0.010(0.233)
	9	0.02	-0.12	0.02	0.95	0.91	1.02	1.60	1.29	0.03	4.47	3.07	2.97	0.292(0.000)	0.308(0.000)	0.016(0.016)
	10	0.01	-0.09	0.01	1.01	1.17	0.99	0.85	0.70	0.01	2.35	1.74	3.00	0.247(0.000)	0.294(0.000)	0.009(0.366)
	11	-0.05	-0.18	-0.01	0.65	0.45	0.96	3.47	2.51	0.01	16.14	8.79	2.93	0.415(0.000)	0.415(0.000)	0.008(0.519)
	12	-0.02	-0.16	-0.00	0.84	0.73	0.99	2.13	1.70	-0.01	6.74	4.48	3.00	0.346(0.000)	0.346(0.000)	0.006(0.912)
	13	0.01	-0.07	0.00	1.01	1.22	0.99	0.66	0.54	-0.01	2.02	1.52	2.92	0.240(0.000)	0.293(0.000)	0.008(0.598)
	14	0.01	-0.13	0.01	0.97	0.99	1.02	1.38	1.12	0.05	3.69	2.59	2.98	0.275(0.000)	0.305(0.000)	0.009(0.427)
	15	0.00	0.06	-0.01	0.93	1.09	0.99	-0.63	-0.40	0.01	2.95	1.66	3.06	0.168(0.000)	0.211(0.000)	0.010(0.323)
	16	0.00	0.01	-0.01	1.01	1.26	1.00	-0.02	-0.03	-0.01	1.84	1.30	2.94	0.156(0.000)	0.213(0.000)	0.012(0.124)
	17	0.00	0.04	0.01	0.98	1.20	0.98	-0.31	-0.22	0.00	2.12	1.41	2.99	0.165(0.000)	0.218(0.000)	0.008(0.607)
	18	-0.04	-0.17	-0.00	0.68	0.47	0.98	3.38	2.44	-0.01	15.40	8.32	2.97	0.410(0.000)	0.410(0.000)	0.010(0.271)
	19	0.02	-0.13	0.01	0.93	0.84	1.00	1.82	1.45	0.00	5.26	3.58	2.99	0.314(0.000)	0.314(0.000)	0.005(0.924)
	20	-0.04	-0.17	-0.02	0.72	0.54	1.00	2.92	2.21	0.02	11.69	6.96	2.96	0.395(0.000)	0.395(0.000)	0.011(0.189)
2	1	-0.00	-0.13	-0.00	0.96	1.01	0.99	1.32	1.06	0.01	3.60	2.47	3.01	0.273(0.000)	0.307(0.000)	0.007(0.767)
	2	0.00	0.06	-0.00	0.99	1.21	0.96	-0.51	-0.38	-0.01	2.11	1.45	3.02	0.200(0.000)	0.252(0.000)	0.007(0.678)
	3	-0.03	0.02	-0.01	0.99	1.22	1.00	-0.41	-0.29	0.03	2.03	1.39	3.03	0.183(0.000)	0.235(0.000)	0.013(0.070)
	4	-0.00	-0.15	0.01	0.88	0.75	0.99	2.17	1.62	-0.00	7.34	4.27	2.92	0.336(0.000)	0.336(0.000)	0.008(0.620)
	5	0.00	-0.14	0.00	0.89	0.81	1.03	1.88	1.47	-0.02	5.70	3.69	3.01	0.317(0.000)	0.319(0.000)	0.009(0.339)
	6	0.01	-0.12	0.02	0.96	0.97	0.98	1.41	1.12	0.03	3.88	2.63	2.96	0.275(0.000)	0.303(0.000)	0.009(0.361)
	7	-0.04	-0.18	-0.02	0.78	0.63	0.99	2.54	1.92	-0.04	9.39	5.54	2.94	0.370(0.000)	0.370(0.000)	0.010(0.285)
	8	-0.00	-0.01	0.00	1.02	1.31	0.99	0.03	0.04	-0.02	1.52	1.21	3.01	0.191(0.000)	0.248(0.000)	0.006(0.865)
	9	0.00	-0.04	0.00	1.00	1.28	1.01	0.32	0.25	0.02	1.59	1.26	2.98	0.214(0.000)	0.271(0.000)	0.006(0.818)
	10	0.01	-0.11	0.03	0.99	1.11	1.00	1.04	0.83	-0.00	2.79	1.97	2.99	0.255(0.000)	0.298(0.000)	0.013(0.061)
	11	0.02	0.08	0.00	0.97	1.15	0.99	-0.62	-0.47	0.03	2.57	1.60	3.03	0.205(0.000)	0.250(0.000)	0.008(0.544)
	12	0.00	0.00	0.02	1.02	1.31	1.01	-0.01	-0.00	0.05	1.54	1.21	2.98	0.185(0.000)	0.242(0.000)	0.011(0.152)
	13	0.01	-0.12	0.01	0.98	1.03	0.99	1.25	1.00	0.04	3.37	2.33	2.97	0.267(0.000)	0.302(0.000)	0.006(0.866)
	14	0.01	-0.06	-0.00	1.01	1.25	0.98	0.50	0.42	0.01	1.78	1.39	3.04	0.226(0.000)	0.281(0.000)	0.010(0.254)
	15	-0.03	-0.16	-0.00	0.71	0.49	0.99	3.33	2.31	-0.03	15.53	7.74	2.95	0.403(0.000)	0.403(0.000)	0.005(0.951)
	16	-0.01	-0.15	-0.00	0.86	0.73	1.01	2.19	1.67	-0.01	7.41	4.46	2.96	0.344(0.000)	0.344(0.000)	0.006(0.908)
	17	-0.01	-0.15	0.00	0.82	0.62	1.00	2.71	1.93	-0.01	10.76	5.61	2.99	0.369(0.000)	0.369(0.000)	0.008(0.546)
	18	0.01	0.07	-0.00	0.97	1.17	0.99	-0.61	-0.44	-0.00	2.40	1.55	3.08	0.202(0.000)	0.249(0.000)	0.010(0.278)
	19	-0.02	-0.05	-0.02	1.00	1.29	1.00	0.18	0.16	0.02	1.52	1.23	2.93	0.213(0.000)	0.270(0.000)	0.012(0.117)
	20	0.01	0.06	-0.01	0.99	1.22	1.03	-0.44	-0.33	0.01	2.04	1.40	3.05	0.198(0.000)	0.252(0.000)	0.009(0.333)
3	1	-0.01	-0.08	-0.01	1.00	1.23	0.98	0.59	0.49	0.01	1.94	1.46	2.99	0.244(0.000)	0.297(0.000)	0.014(0.050)
	2	0.01	-0.10	-0.00	0.98	1.07	1.00	1.09	0.80	-0.04	3.28	2.02	2.98	0.246(0.000)	0.283(0.000)	0.007(0.731)
	3	0.01	-0.09	0.01	0.98	1.09	1.00	1.03	0.76	0.04	3.13	1.94	2.98	0.240(0.000)	0.278(0.000)	0.008(0.606)
	4	0.02	-0.06	0.02	1.01	1.18	1.01	0.78	0.57	-0.03	2.48	1.63	2.99	0.223(0.000)	0.266(0.000)	0.018(0.002)
	5	-0.01	-0.09	0.01	1.00	1.18	1.01	0.76	0.59	-0.02	2.39	1.64	2.94	0.237(0.000)	0.284(0.000)	0.010(0.239)
	6	-0.01	-0.08	-0.01	1.00	1.23	1.01	0.61	0.49	0.00	2.01	1.48	2.99	0.239(0.000)	0.292(0.000)	0.007(0.666)
	7	0.04	-0.05	0.02	0.99	1.13	1.00	0.85	0.61	-0.04	2.75	1.73	3.04	0.213(0.000)	0.251(0.000)	0.012(0.103)
	8	-0.01	-0.10	-0.01	1.02	1.21	1.00	0.79	0.62	0.00	2.37	1.63	2.96	0.254(0.000)	0.302(0.000)	0.008(0.532)
	9	-0.02	-0.10	-0.02	0.99	1.22	1.00	0.66	0.56	0.01	1.98	1.53	3.03	0.257(0.000)	0.309(0.000)	0.010(0.317)
	10	-0.01	-0.08	0.00	1.01	1.25	1.01	0.55	0.46	0.01	1.85	1.42	3.04	0.248(0.000)	0.301(0.000)	0.005(0.940)
	11	0.00	-0.10	0.00	0.96	1.04	0.99	1.16	0.84	-0.02	3.70	2.15	3.00	0.246(0.000)	0.277(0.000)	0.006(0.877)
	12	0.00	-0.09	0.00	1.01	1.20	1.02	0.78	0.61	-0.06	2.33	1.62	2.95	0.249(0.000)	0.295(0.000)	0.009(0.403)
	13	-0.01	-0.08	-0.00	1.02	1.25	0.99	0.59	0.47	0.00	1.95	1.44	3.01	0.243(0.000)	0.298(0.000)	0.004(0.989)
	14	-0.01	-0.09	-0.00	1.00	1.24	1.01	0.60	0.50	0.02	1.89	1.45	3.01	0.255(0.000)	0.308(0.000)	0.005(0.948)
	15	0.00	-0.10	-0.01	0.90	0.97	1.00	1.20	0.80	-0.01	4.24	2.22	2.99	0.222(0.000)	0.246(0.000)	0.007(0.696)
	16	-0.01	-0.10	-0.02	0.98	1.14	0.99	0.84	0.64	-0.01	2.65	1.75	2.90	0.236(0.000)	0.278(0.000)	0.012(0.123)
	17	-0.01	-0.11	-0.00	0.96	1.07	0.99	1.02	0.74	0.01	3.32	1.97	3.05	0.233(0.000)	0.269(0.000)	0.009(0.352)
	18	0.00	-0.10	-0.01	0.94	1.04	1.02	1.12	0.82	0.03	3.49	2.10	3.04	0.245(0.000)	0.278(0.000)	0.011(0.182)
	19	0.00	-0.08	-0.00	0.99	1.21	1.00	0.68	0.56	0.04	2.04	1.53	2.94	0.247(0.000)	0.296(0.000)	0.008(0.527)
	20	0.01	-0.10	-0.01	0.99	1.10	1.00	1.07	0.77	0.03	3.44	1.94	2.96	0.246(0.000)	0.284(0.000)	0.009(0.445)

Note: The numbers in parentheses represent the p-value of the K-S test for r_{ij}.

For Peer Review Only

Table 2: Comparison of Pearson's residual ($r_{ij}^{(P)}$), deviance residual ($r_{ij}^{(D)}$), normalized surrogate residual ($r_{ij}^{(N)}$), for Scenario I.

j	i	mean			variance			skewness			kurtosis			K-S test		
		$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$
1	1	-0.02	0.03	-0.01	1.01	1.26	1.01	-0.41	-0.32	0.00	1.74	1.33	3.10	0.206(0.000)	0.262(0.000)	0.010(0.293)
	2	-0.07	-0.08	-0.03	0.98	1.25	1.01	0.09	0.11	-0.02	1.65	1.26	3.00	0.198(0.000)	0.255(0.000)	0.014(0.046)
	3	-0.05	0.00	-0.02	1.02	1.27	0.99	-0.41	-0.31	0.02	1.72	1.31	3.07	0.195(0.000)	0.250(0.000)	0.012(0.128)
	4	0.01	-0.07	0.01	1.00	1.19	0.99	0.72	0.54	0.03	2.37	1.59	3.00	0.225(0.000)	0.274(0.000)	0.006(0.800)
	5	0.05	-0.04	0.03	1.04	1.20	0.99	0.74	0.52	0.00	2.47	1.57	3.00	0.214(0.000)	0.261(0.000)	0.014(0.043)
	6	-0.04	0.01	-0.00	1.03	1.28	1.02	-0.38	-0.27	0.02	1.77	1.30	2.94	0.195(0.000)	0.251(0.000)	0.009(0.399)
	7	0.04	-0.07	0.02	0.99	1.08	1.03	1.08	0.77	-0.03	3.29	1.97	2.99	0.238(0.000)	0.270(0.000)	0.017(0.006)
	8	-0.02	0.03	-0.02	1.01	1.26	1.01	-0.39	-0.31	0.00	1.73	1.32	2.89	0.204(0.000)	0.260(0.000)	0.012(0.096)
	9	0.06	-0.03	0.05	1.06	1.17	1.01	0.91	0.60	0.02	3.15	1.69	2.98	0.218(0.000)	0.259(0.000)	0.021(0.000)
	10	0.08	0.09	0.04	0.95	1.15	1.02	-0.01	-0.08	-0.04	2.25	1.47	2.94	0.144(0.000)	0.200(0.000)	0.025(0.000)
	11	-0.02	-0.06	-0.01	1.00	1.25	0.98	0.37	0.31	-0.01	1.78	1.35	3.04	0.205(0.000)	0.261(0.000)	0.007(0.647)
	12	-0.04	-0.08	-0.02	0.99	1.24	0.99	0.39	0.32	-0.00	1.92	1.37	2.99	0.215(0.000)	0.270(0.000)	0.014(0.053)
	13	0.05	0.06	0.03	0.98	1.21	1.02	-0.13	-0.14	-0.05	2.02	1.38	2.97	0.164(0.000)	0.220(0.000)	0.018(0.004)
	14	-0.04	0.01	-0.00	1.00	1.26	0.99	-0.35	-0.28	0.00	1.68	1.30	2.96	0.196(0.000)	0.251(0.000)	0.009(0.397)
	15	-0.06	-0.09	-0.03	0.99	1.25	0.99	0.29	0.26	0.03	1.75	1.33	3.03	0.212(0.000)	0.268(0.000)	0.019(0.001)
	16	0.04	0.05	0.01	0.95	1.18	1.00	-0.04	-0.08	-0.02	2.08	1.40	3.04	0.146(0.000)	0.201(0.000)	0.008(0.602)
	17	0.05	-0.04	0.02	1.03	1.15	1.01	0.94	0.65	-0.01	3.00	1.75	3.00	0.223(0.000)	0.263(0.000)	0.015(0.017)
	18	0.05	-0.06	0.02	1.02	1.08	1.03	1.13	0.78	-0.02	3.59	2.00	2.96	0.235(0.000)	0.266(0.000)	0.015(0.023)
	19	0.02	0.04	0.01	1.02	1.26	1.02	-0.22	-0.19	0.02	1.90	1.34	3.01	0.178(0.000)	0.234(0.000)	0.010(0.300)
	20	-0.06	-0.00	-0.02	1.03	1.28	1.02	-0.39	-0.29	0.02	1.69	1.30	3.02	0.193(0.000)	0.248(0.000)	0.012(0.093)
2	1	-0.03	-0.15	0.00	0.60	0.47	0.98	2.81	2.28	-0.02	10.12	7.49	2.98	0.408(0.000)	0.408(0.000)	0.007(0.649)
	2	0.08	0.01	0.05	0.96	1.17	1.03	0.51	0.39	-0.03	1.87	1.46	3.04	0.156(0.000)	0.202(0.000)	0.029(0.000)
	3	0.01	-0.12	0.03	0.74	0.64	1.01	2.11	1.74	-0.00	6.37	4.81	3.02	0.362(0.000)	0.362(0.000)	0.016(0.014)
	4	0.00	0.03	-0.00	1.00	1.26	1.01	-0.23	-0.18	-0.00	1.76	1.30	3.00	0.184(0.000)	0.238(0.000)	0.005(0.939)
	5	-0.03	0.00	-0.01	1.02	1.26	0.99	-0.28	-0.19	0.01	1.84	1.31	3.07	0.179(0.000)	0.232(0.000)	0.010(0.238)
	6	-0.03	-0.14	-0.01	0.56	0.43	0.99	3.06	2.46	-0.03	11.78	8.52	2.91	0.418(0.000)	0.418(0.000)	0.007(0.729)
	7	-0.03	0.04	-0.02	0.99	1.18	1.01	-0.62	-0.45	-0.01	2.27	1.54	2.94	0.195(0.000)	0.239(0.000)	0.011(0.214)
	8	-0.03	-0.15	-0.02	0.58	0.44	1.00	3.00	2.43	0.01	11.31	8.36	3.04	0.417(0.000)	0.417(0.000)	0.011(0.150)
	9	-0.06	-0.01	-0.02	1.03	1.24	1.00	-0.44	-0.28	-0.01	2.06	1.38	2.97	0.177(0.000)	0.226(0.000)	0.017(0.007)
	10	-0.04	-0.12	-0.01	0.34	0.18	0.98	5.93	4.21	-0.05	41.79	23.79	2.97	0.471(0.000)	0.471(0.000)	0.006(0.890)
	11	0.03	0.01	0.02	1.02	1.28	1.02	0.15	0.10	0.01	1.68	1.28	2.97	0.163(0.000)	0.220(0.000)	0.011(0.204)
	12	0.05	0.02	0.04	1.01	1.26	0.99	0.17	0.12	-0.01	1.69	1.28	2.96	0.162(0.000)	0.219(0.000)	0.019(0.002)
	13	-0.05	-0.14	-0.02	0.37	0.22	0.97	5.22	3.78	-0.07	32.64	19.42	2.99	0.464(0.000)	0.464(0.000)	0.009(0.437)
	14	-0.03	-0.15	-0.00	0.58	0.45	0.98	2.94	2.38	-0.03	10.93	8.04	2.89	0.414(0.000)	0.414(0.000)	0.007(0.694)
	15	0.06	0.03	0.06	1.00	1.25	1.01	0.26	0.18	-0.02	1.73	1.31	2.99	0.161(0.000)	0.217(0.000)	0.028(0.000)
	16	-0.05	-0.13	0.00	0.35	0.20	0.97	5.51	3.95	0.00	36.15	21.07	3.02	0.467(0.000)	0.467(0.000)	0.010(0.272)
	17	-0.04	0.01	-0.04	1.00	1.21	1.02	-0.48	-0.34	0.01	2.05	1.42	2.99	0.185(0.000)	0.234(0.000)	0.024(0.000)
	18	-0.05	0.02	-0.03	1.02	1.18	1.02	-0.68	-0.46	0.07	2.53	1.57	3.04	0.189(0.000)	0.233(0.000)	0.021(0.000)
	19	-0.04	-0.14	-0.02	0.45	0.29	0.98	4.25	3.27	-0.04	21.56	14.40	2.96	0.451(0.000)	0.451(0.000)	0.008(0.483)
	20	0.00	-0.12	0.04	0.72	0.64	0.98	2.12	1.75	0.03	6.43	4.87	2.98	0.364(0.000)	0.364(0.000)	0.016(0.014)
3	1	0.03	-0.07	0.01	1.02	1.16	1.01	0.90	0.70	-0.00	2.54	1.76	2.95	0.248(0.000)	0.296(0.000)	0.010(0.310)
	2	-0.03	-0.17	-0.01	0.80	0.69	0.99	2.19	1.75	-0.03	7.01	4.79	2.99	0.355(0.000)	0.355(0.000)	0.010(0.297)
	3	0.03	-0.09	0.04	1.00	1.09	1.01	1.10	0.88	-0.01	2.89	2.08	2.94	0.261(0.000)	0.297(0.000)	0.018(0.002)
	4	-0.03	-0.16	-0.02	0.68	0.48	0.99	3.08	2.42	-0.04	12.24	8.05	2.94	0.409(0.000)	0.409(0.000)	0.011(0.180)
	5	-0.04	-0.17	-0.02	0.64	0.45	0.97	3.25	2.53	-0.01	13.65	8.70	3.00	0.416(0.000)	0.416(0.000)	0.012(0.139)
	6	0.04	-0.05	0.03	1.04	1.20	1.00	0.80	0.60	-0.00	2.37	1.62	3.06	0.229(0.000)	0.280(0.000)	0.017(0.005)
	7	-0.02	-0.15	-0.01	0.65	0.40	0.99	3.79	2.80	-0.02	18.79	10.46	2.93	0.428(0.000)	0.428(0.000)	0.008(0.530)
	8	0.03	-0.07	0.02	1.03	1.18	1.01	0.83	0.64	0.01	2.40	1.68	2.92	0.240(0.000)	0.289(0.000)	0.015(0.019)
	9	-0.02	-0.15	0.00	0.66	0.44	0.96	3.44	2.62	-0.05	15.26	9.24	2.95	0.420(0.000)	0.420(0.000)	0.011(0.153)
	10	-0.07	-0.09	-0.04	0.94	1.14	1.02	0.18	0.19	0.01	2.33	1.51	3.05	0.154(0.000)	0.208(0.000)	0.017(0.006)
	11	-0.04	-0.17	-0.02	0.71	0.56	0.97	2.67	2.12	-0.01	9.57	6.45	2.90	0.389(0.000)	0.389(0.000)	0.010(0.303)
	12	-0.02	-0.16	-0.01	0.76	0.60	0.97	2.56	2.02	-0.01	9.02	5.96	2.98	0.380(0.000)	0.380(0.000)	0.010(0.299)
	13	-0.03	-0.06	-0.03	0.98	1.19	1.00	0.32	0.27	-0.03	2.17	1.46	3.06	0.175(0.000)	0.229(0.000)	0.013(0.084)
	14	0.05	-0.05	0.03	1.02	1.18	1.02	0.81	0.62	-0.04	2.35	1.65	3.04	0.233(0.000)	0.281(0.000)	0.015(0.023)
	15	-0.03	-0.17	-0.01	0.74	0.61	0.99	2.43	1.93	-0.02	8.27	5.59	3.02	0.374(0.000)	0.374(0.000)	0.007(0.647)
	16	-0.03	-0.05	-0.00	0.95	1.17	1.00	0.21	0.19	-0.02	2.12	1.45	2.96	0.154(0.000)	0.209(0.000)	0.007(0.664)
	17	-0.03	-0.16	-0.01	0.64	0.41	0.98	3.61	2.73	0.03	16.50	10.02	2.97	0.425(0.000)	0.425(0.000)	0.009(0.352)
	18	-0.02	-0.14	-0.00	0.64	0.40	0.97	3.76	2.78	-0.04	18.09	10.35	2.98	0.427(0.000)	0.427(0.000)	0.006(0.844)
	19	-0.00	-0.06	0.00	1.00	1.22	1.00	0.46	0.36	0.01	2.03	1.46	2.99	0.191(0.000)	0.245(0.000)	0.006(0.904)
	20	0.04	-0.08	0.01	1.02	1.10	1.05	1.08	0.87	-0.03	2.85	2.04	2.87	0.256(0.000)	0.292(0.000)	0.015(0.029)

Note: The numbers in parentheses represent the p-value of the K-S test for r_{ij}.

For Peer Review Only

Table 3: Comparison of Pearson's residual ($r_{ij}^{(P)}$), deviance residual ($r_{ij}^{(D)}$), normalized surrogate residual ($r_{ij}^{(N)}$), for Scenario I.

j	i	mean			variance			skewness			kurtosis			K-S test		
		$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$	$r_{ij}^{(P)}$	$r_{ij}^{(D)}$	$r_{ij}^{(N)}$
1	1	-0.03	0.06	-0.01	1.02	1.20	0.99	-0.74	-0.59	-0.01	2.20	1.60	2.98	0.244(0.000)	0.288(0.000)	0.007(0.694)
	2	-0.01	0.08	0.00	1.01	1.19	1.00	-0.75	-0.60	-0.04	2.27	1.63	2.97	0.247(0.000)	0.292(0.000)	0.007(0.746)
	3	-0.04	0.10	-0.02	1.03	1.04	1.00	-1.33	-1.08	0.00	3.49	2.45	2.96	0.289(0.000)	0.325(0.000)	0.013(0.071)
	4	-0.04	0.07	-0.00	1.02	1.17	0.99	-0.87	-0.70	0.01	2.36	1.73	2.98	0.259(0.000)	0.300(0.000)	0.006(0.924)
	5	-0.07	0.07	-0.00	1.07	1.07	1.01	-1.29	-1.05	0.04	3.30	2.34	2.99	0.288(0.000)	0.320(0.000)	0.010(0.291)
	6	-0.04	0.10	-0.03	1.02	1.04	1.00	-1.29	-1.03	0.04	3.39	2.37	2.99	0.282(0.000)	0.317(0.000)	0.016(0.020)
	7	0.19	0.23	0.10	0.82	1.02	0.97	-0.42	-0.42	-0.03	2.55	1.73	2.97	0.204(0.000)	0.257(0.000)	0.049(0.000)
	8	0.11	0.19	0.03	0.78	0.92	0.98	-1.01	-0.84	-0.04	3.47	2.31	2.98	0.252(0.000)	0.285(0.000)	0.020(0.002)
	9	0.03	0.14	0.01	0.90	0.99	0.98	-1.17	-0.96	-0.01	3.31	2.34	3.04	0.272(0.000)	0.306(0.000)	0.010(0.328)
	10	-0.04	0.09	-0.00	1.02	1.06	1.00	-1.23	-0.97	-0.00	3.29	2.26	2.98	0.272(0.000)	0.307(0.000)	0.007(0.753)
	11	-0.02	0.07	0.00	1.00	1.17	1.00	-0.79	-0.64	-0.01	2.24	1.67	2.98	0.254(0.000)	0.297(0.000)	0.004(0.996)
	12	-0.06	0.07	-0.01	1.06	1.14	1.01	-1.04	-0.83	0.06	2.72	1.92	3.03	0.270(0.000)	0.311(0.000)	0.013(0.103)
	13	-0.06	0.08	-0.01	1.07	1.05	1.00	-1.33	-1.07	-0.01	3.47	2.41	3.00	0.287(0.000)	0.321(0.000)	0.010(0.361)
	14	-0.07	0.07	-0.01	1.07	1.07	1.00	-1.27	-1.05	0.07	3.23	2.33	3.05	0.288(0.000)	0.319(0.000)	0.014(0.051)
	15	0.02	0.10	0.01	0.97	1.17	1.03	-0.67	-0.55	-0.01	2.22	1.61	3.06	0.240(0.000)	0.287(0.000)	0.008(0.557)
	16	-0.07	0.08	-0.03	1.09	1.07	1.03	-1.30	-1.06	0.01	3.33	2.36	2.98	0.288(0.000)	0.320(0.000)	0.017(0.008)
	17	-0.05	0.07	-0.04	1.04	1.16	1.01	-0.95	-0.76	0.00	2.52	1.82	2.90	0.265(0.000)	0.306(0.000)	0.021(0.001)
	18	0.05	0.15	0.02	0.86	0.98	0.98	-1.08	-0.89	0.00	3.17	2.28	3.02	0.256(0.000)	0.290(0.000)	0.015(0.035)
	19	0.09	0.14	0.05	0.92	1.14	0.97	-0.51	-0.46	0.00	2.15	1.59	2.96	0.218(0.000)	0.268(0.000)	0.023(0.000)
	20	0.13	0.22	0.04	0.78	0.91	0.97	-1.06	-0.88	-0.00	3.48	2.41	3.07	0.258(0.000)	0.291(0.000)	0.022(0.000)
2	1	0.04	-0.09	0.03	1.02	1.02	1.01	1.34	1.05	-0.01	3.63	2.48	3.00	0.280(0.000)	0.315(0.000)	0.014(0.053)
	2	0.02	-0.12	0.00	0.98	1.00	1.02	1.36	1.07	0.03	3.75	2.55	2.94	0.287(0.000)	0.326(0.000)	0.008(0.652)
	3	-0.04	-0.15	-0.01	0.51	0.34	0.98	3.71	3.00	-0.04	16.37	11.92	2.98	0.440(0.000)	0.440(0.000)	0.009(0.450)
	4	0.04	-0.11	0.02	0.99	0.93	0.99	1.59	1.26	0.02	4.40	2.99	3.02	0.304(0.000)	0.332(0.000)	0.009(0.440)
	5	0.03	-0.12	0.02	0.83	0.57	1.01	2.70	2.22	-0.04	9.34	6.72	2.98	0.392(0.000)	0.392(0.000)	0.014(0.058)
	6	-0.03	-0.14	-0.01	0.48	0.32	1.00	3.79	3.05	-0.07	17.09	12.38	2.96	0.442(0.000)	0.442(0.000)	0.007(0.783)
	7	-0.17	-0.24	-0.09	0.75	0.91	0.95	0.89	0.71	0.04	3.50	2.21	2.99	0.224(0.000)	0.272(0.000)	0.043(0.000)
	8	-0.07	-0.14	-0.02	0.25	0.18	0.95	5.33	3.66	-0.08	35.26	20.37	2.95	0.469(0.000)	0.469(0.000)	0.012(0.134)
	9	-0.06	-0.14	-0.03	0.34	0.24	0.98	4.50	3.43	-0.07	24.22	16.29	2.87	0.459(0.000)	0.459(0.000)	0.011(0.177)
	10	-0.04	-0.14	-0.01	0.42	0.29	0.99	4.08	3.22	-0.03	19.67	13.93	2.95	0.449(0.000)	0.449(0.000)	0.009(0.393)
	11	0.03	-0.11	0.03	0.99	0.97	1.02	1.45	1.14	0.01	3.99	2.68	3.03	0.295(0.000)	0.328(0.000)	0.012(0.137)
	12	0.07	-0.09	0.03	1.04	0.88	1.06	1.80	1.42	0.04	5.14	3.45	3.00	0.308(0.000)	0.316(0.000)	0.015(0.033)
	13	-0.00	-0.13	0.00	0.65	0.43	1.01	3.23	2.66	-0.03	12.64	9.32	2.94	0.421(0.000)	0.421(0.000)	0.013(0.094)
	14	0.03	-0.12	0.02	0.82	0.56	1.02	2.72	2.25	-0.04	9.47	6.89	2.98	0.395(0.000)	0.395(0.000)	0.016(0.023)
	15	-0.01	-0.13	-0.01	0.95	1.02	1.01	1.21	0.95	0.01	3.40	2.33	2.91	0.270(0.000)	0.314(0.000)	0.011(0.241)
	16	0.02	-0.12	0.03	0.79	0.55	1.01	2.77	2.28	-0.03	9.78	7.05	2.87	0.397(0.000)	0.397(0.000)	0.017(0.007)
	17	0.08	-0.08	0.03	1.09	0.95	1.05	1.64	1.29	-0.03	4.55	3.05	2.89	0.289(0.000)	0.314(0.000)	0.020(0.001)
	18	-0.06	-0.14	-0.02	0.29	0.21	0.98	4.86	3.52	-0.04	29.09	17.78	2.96	0.463(0.000)	0.463(0.000)	0.010(0.361)
	19	-0.07	-0.16	-0.05	0.88	1.01	0.97	0.95	0.78	-0.00	3.01	2.13	2.97	0.245(0.000)	0.290(0.000)	0.021(0.001)
	20	-0.07	-0.15	-0.01	0.23	0.17	0.94	5.53	3.71	-0.11	38.38	21.69	3.02	0.473(0.000)	0.473(0.000)	0.014(0.067)
3	1	-0.03	-0.17	-0.01	0.71	0.53	1.00	2.85	2.29	-0.03	10.49	7.21	3.09	0.398(0.000)	0.398(0.000)	0.009(0.453)
	2	-0.02	-0.16	0.01	0.73	0.54	0.99	2.79	2.26	-0.03	10.11	7.02	2.87	0.396(0.000)	0.396(0.000)	0.010(0.317)
	3	0.06	-0.11	0.03	1.04	0.89	1.03	1.78	1.43	0.01	5.05	3.44	2.96	0.307(0.000)	0.323(0.000)	0.018(0.006)
	4	-0.01	-0.16	-0.01	0.78	0.58	0.98	2.67	2.18	-0.00	9.29	6.55	2.98	0.389(0.000)	0.389(0.000)	0.010(0.293)
	5	0.04	-0.13	0.02	1.01	0.77	1.01	2.19	1.78	-0.02	6.77	4.62	3.01	0.348(0.000)	0.348(0.000)	0.018(0.005)
	6	0.05	-0.11	0.02	1.03	0.91	1.03	1.70	1.35	0.01	4.76	3.24	2.93	0.297(0.000)	0.320(0.000)	0.017(0.007)
	7	-0.06	-0.18	-0.02	0.51	0.37	0.96	3.65	2.61	-0.05	17.63	10.17	2.99	0.430(0.000)	0.429(0.000)	0.013(0.073)
	8	-0.08	-0.18	-0.05	0.75	0.85	0.97	1.22	0.96	-0.03	4.06	2.65	3.03	0.257(0.000)	0.274(0.000)	0.024(0.000)
	9	-0.01	-0.14	-0.01	0.88	0.90	1.00	1.45	1.16	-0.04	4.12	2.85	2.96	0.277(0.000)	0.305(0.000)	0.006(0.927)
	10	0.06	-0.09	0.04	1.04	0.95	1.02	1.56	1.22	-0.06	4.32	2.91	2.90	0.279(0.000)	0.311(0.000)	0.024(0.000)
	11	-0.02	-0.16	-0.01	0.75	0.55	1.01	2.78	2.26	-0.02	10.03	7.00	2.93	0.395(0.000)	0.395(0.000)	0.007(0.731)
	12	-0.02	-0.18	0.00	0.79	0.58	1.00	2.70	2.21	-0.00	9.43	6.68	3.02	0.390(0.000)	0.390(0.000)	0.005(0.985)
	13	0.06	-0.11	0.03	1.06	0.85	1.05	1.95	1.56	-0.00	5.71	3.85	2.93	0.324(0.000)	0.325(0.000)	0.020(0.001)
	14	0.05	-0.12	0.04	1.03	0.78	1.02	2.15	1.76	0.03	6.54	4.54	2.95	0.346(0.000)	0.346(0.000)	0.016(0.013)
	15	-0.04	-0.18	-0.03	0.66	0.49	0.99	3.00	2.38	0.01	11.63	7.83	3.00	0.407(0.000)	0.407(0.000)	0.014(0.065)
	16	0.05	-0.12	0.02	1.04	0.79	1.04	2.14	1.75	0.04	6.49	4.50	2.96	0.345(0.000)	0.345(0.000)	0.015(0.034)
	17	-0.04	-0.19	-0.02	0.73	0.54	0.98	2.82	2.30	-0.04	10.25	7.19	2.99	0.398(0.000)	0.398(0.000)	0.012(0.158)
	18	-0.04	-0.15	-0.02	0.83	0.90	0.99	1.32	1.04	-0.05	3.86	2.68	2.94	0.263(0.000)	0.282(0.000)	0.007(0.722)
	19	-0.05	-0.18	-0.02	0.60	0.44	0.97	3.32	2.48	-0.04	14.74	8.65	3.03	0.415(0.000)	0.415(0.000)	0.014(0.056)
	20	-0.10	-0.20	-0.06	0.76	0.85	0.97	1.28	1.00	-0.03	4.17	2.75	3.03	0.265(0.000)	0.282(0.000)	0.028(0.000)

Note: The numbers in parentheses represent the p-value of the K-S test for r_{ij}.

For Peer Review Only

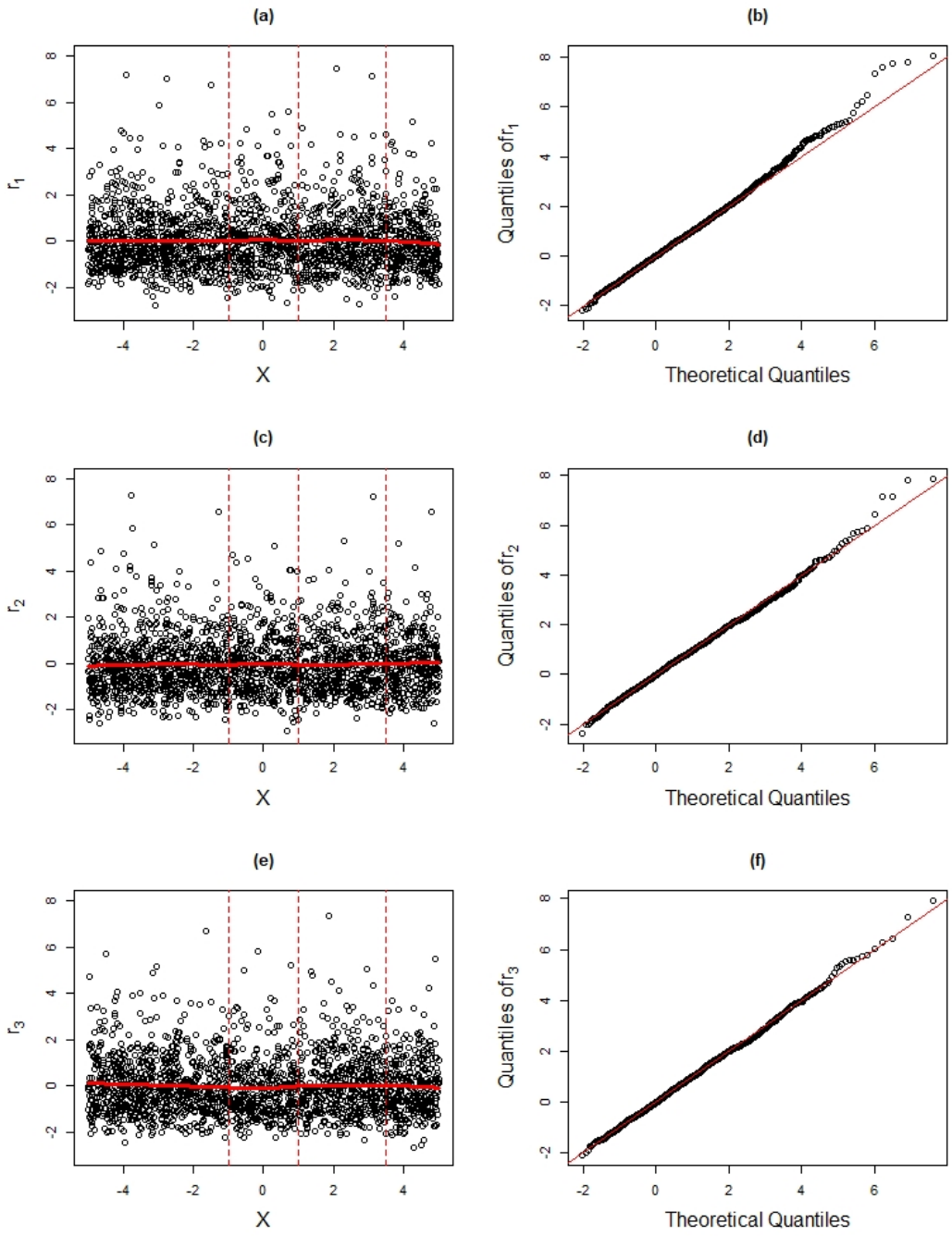

Figure S1: Model diagnostics for the improved model in Example 2 with the surrogate residual. Figures (a), (c) and (e) show the residuals-by-covariate plots of r_1, r_2 and r_3 respectively; vertical dashed lines $X = -1$, 1, and 3.5, and loess curves (red solid) are added as references. Figures (b), (d) and (f) are the corresponding QQ-plots (quantile-quantile plots) of above mentioned residuals.

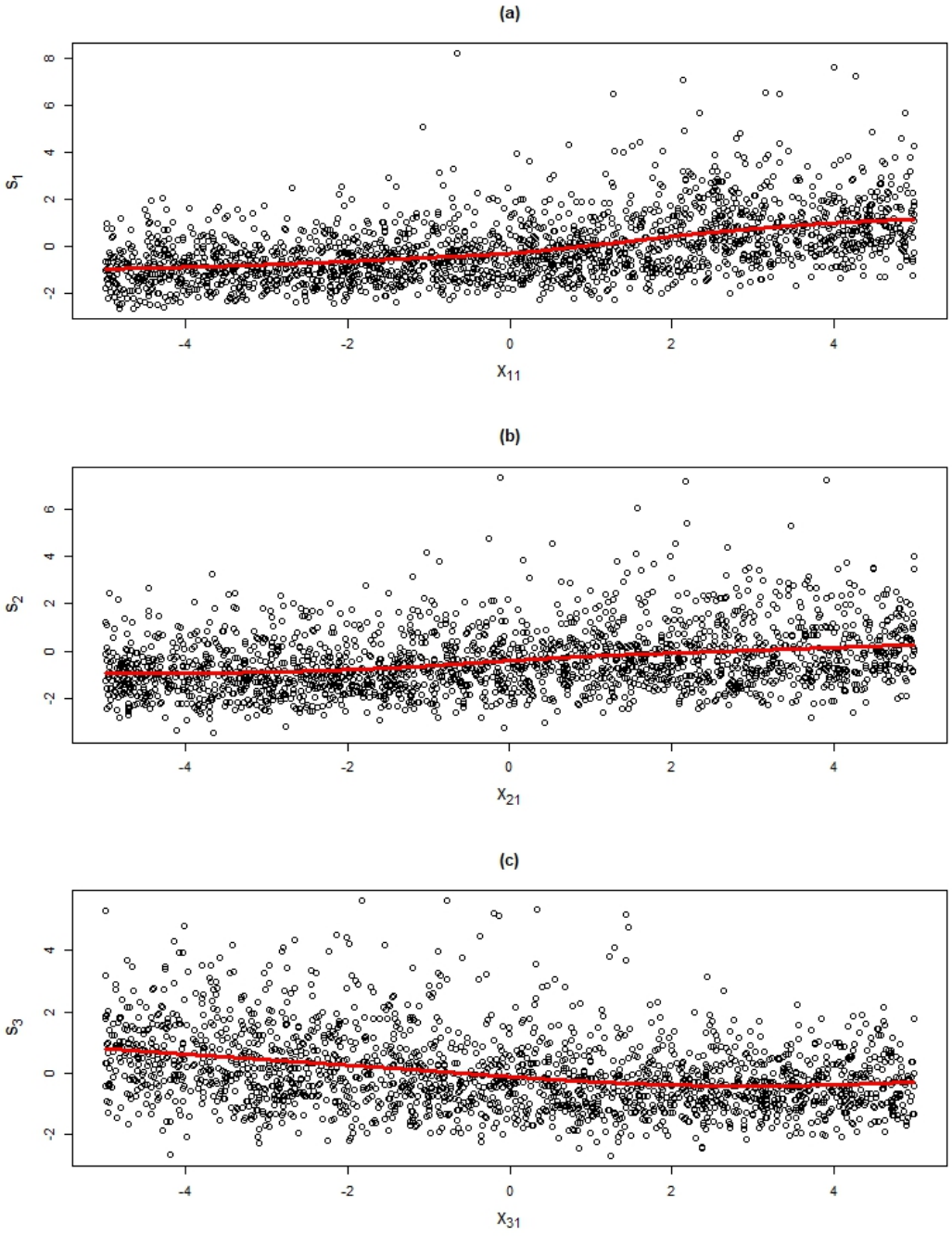

Figure S2: Scatterplots of the surrogate response $\boldsymbol{s}$ versus x_1 with a nonparametric loess smooth (red line). (a): s_1-by-x_{11}; (b): s_2-by-x_{21}; (c): s_3-by-x_{31}.

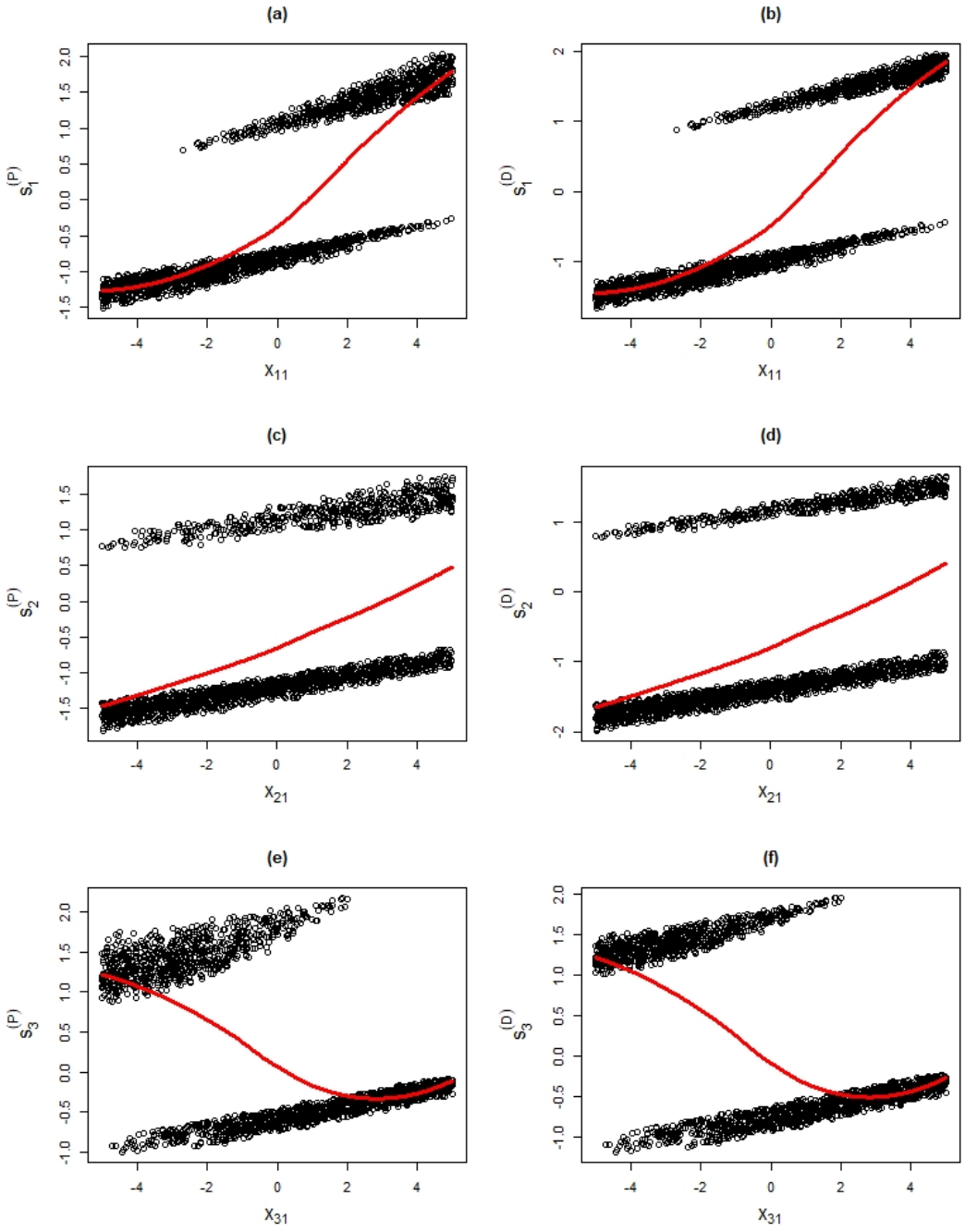

Figure S3: Scatterplots of the surrogate response by Pearson's residual ($\boldsymbol{s}^{(P)}$) and deviance residual ($\boldsymbol{s}^{(D)}$) versus x_1 with a nonparametric loess smooth (red line). Left column: $\boldsymbol{s}^{(P)}$ by x_1; Right column: $\boldsymbol{s}^{(D)}$ by x_1.

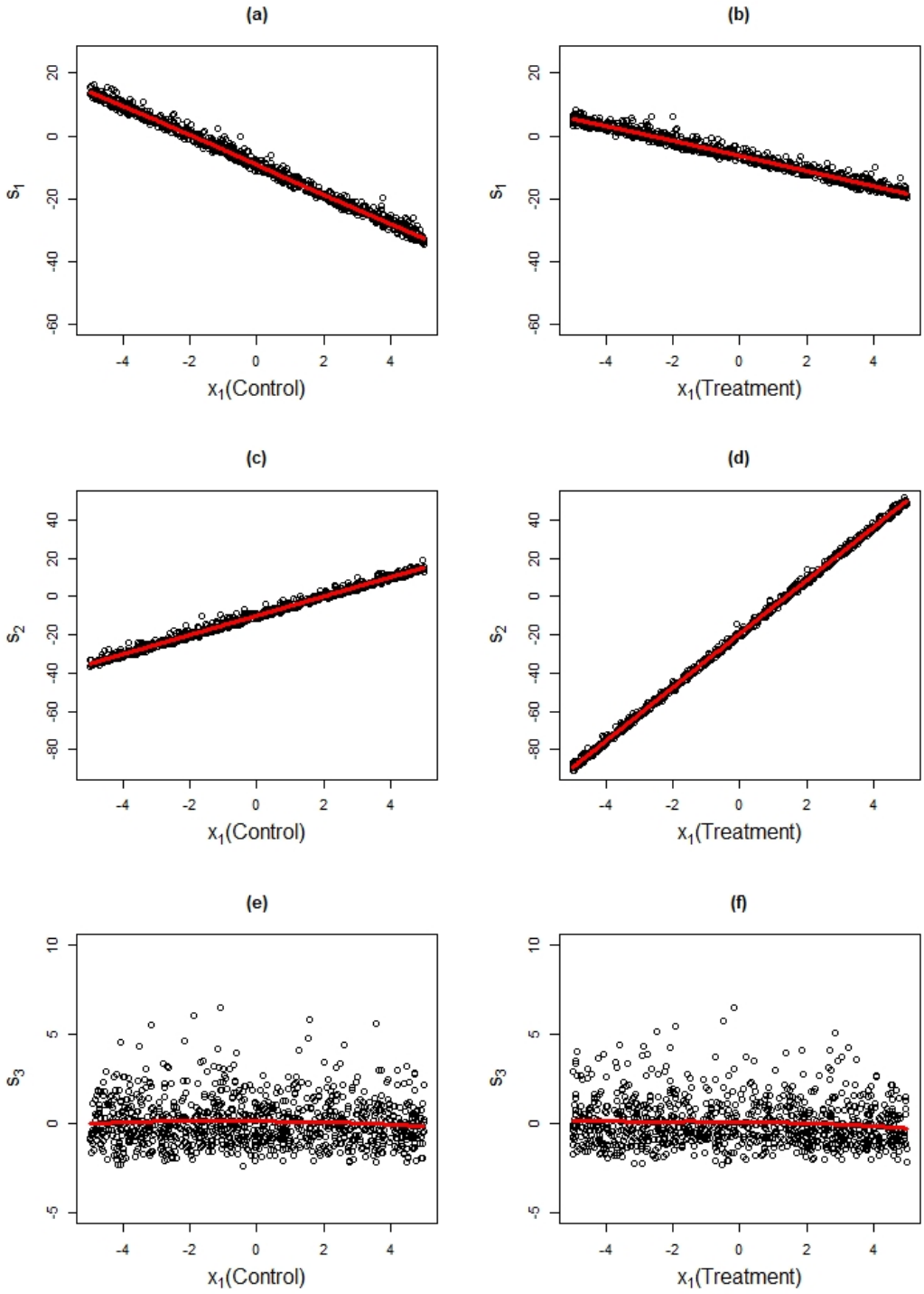

Figure S4: Scatterplots of the surrogate response $\boldsymbol{s}$ versus x_1 with a nonparametric loess smooth (red line). Left column: control group; Right column: treatment group.

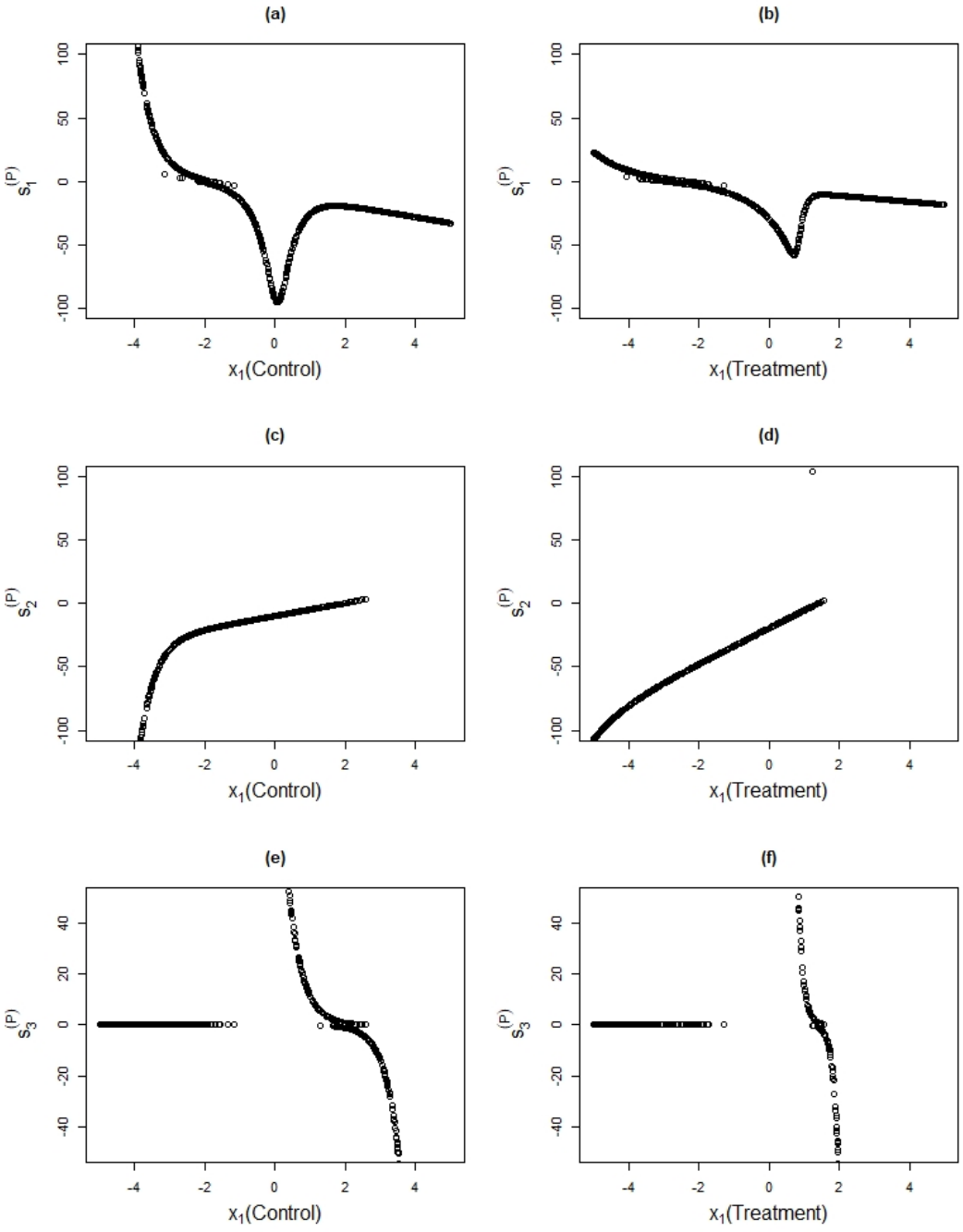

Figure S5: Scatterplots of the surrogate response of Pearson's residual ($\boldsymbol{s}^{(P)}$) versus x_1. Left column: control group; Right column: treatment group. The loess curve was removed, as some points tend to infinity and a loess curve is hard to compute.

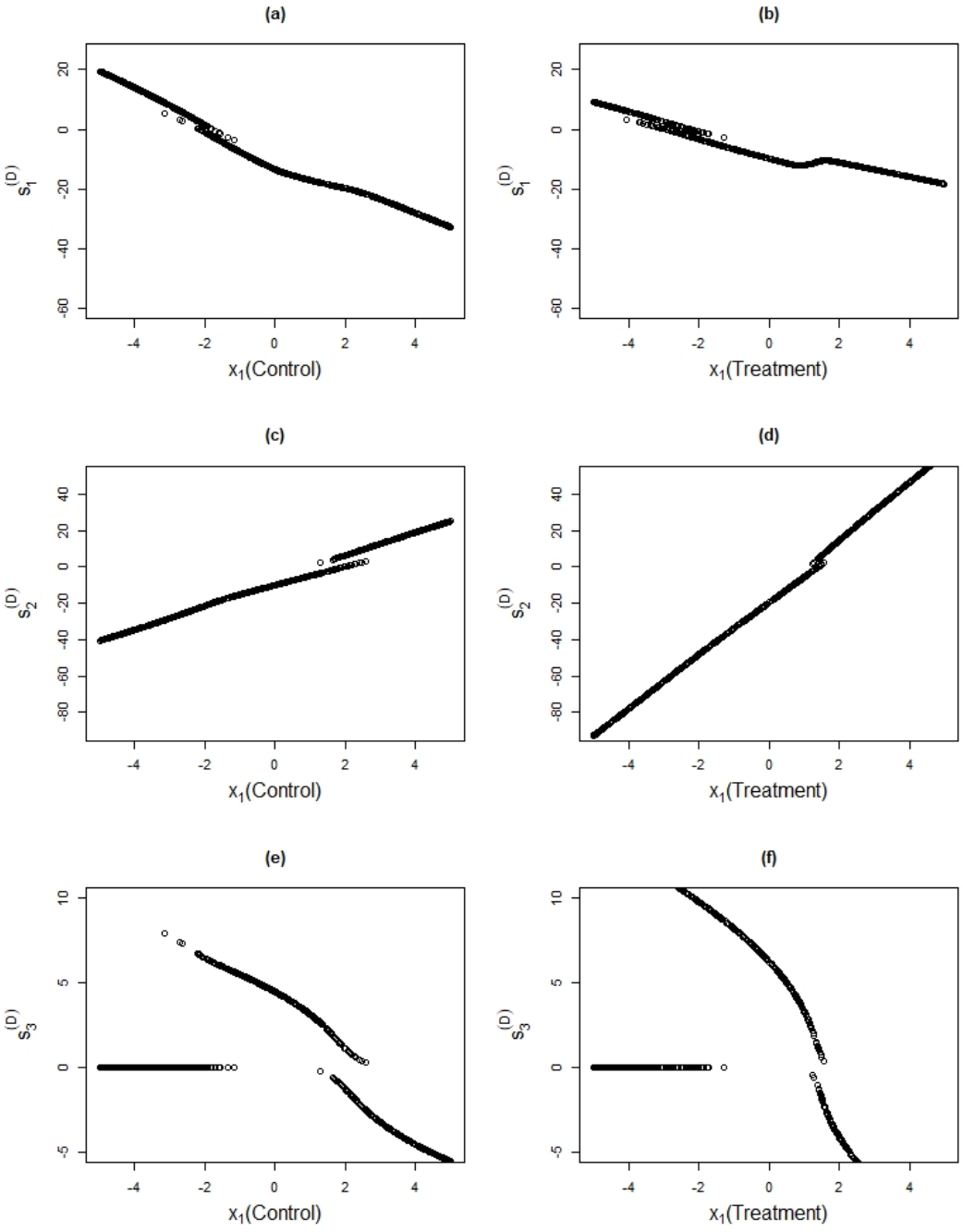

Figure S6: Scatterplots of the surrogate response of deviance residual ($\boldsymbol{s}^{(P)}$) versus x_1. Left column: control group; Right column: treatment group. The loess curve was removed, as some points tend to infinity and a loess curve is hard to compute.

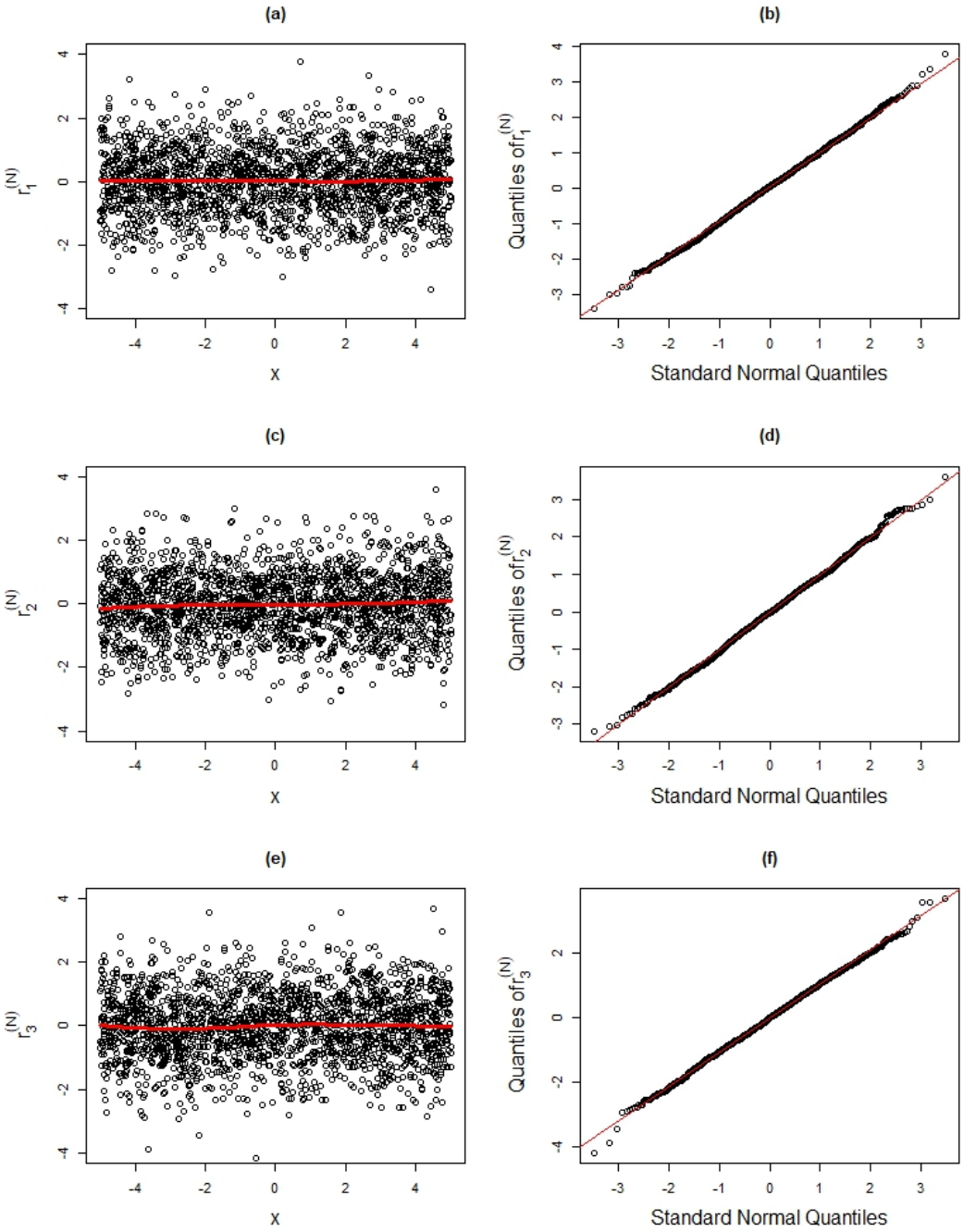

Figure S7: Model diagnostics for the improved model in Example 2 with the normalized surrogate residual defined in Section 3.4.1. Figures (a), (b) and (c) show the residuals-by-covariate plots of $r_1^{(N)}$, $r_2^{(N)}$ and $r_3^{(N)}$ respectively; loess curves (red solid) are added as references. Figures (b), (d) and (f) are the corresponding QQ-plots (quantile-quantile plots) of above mentioned residuals.

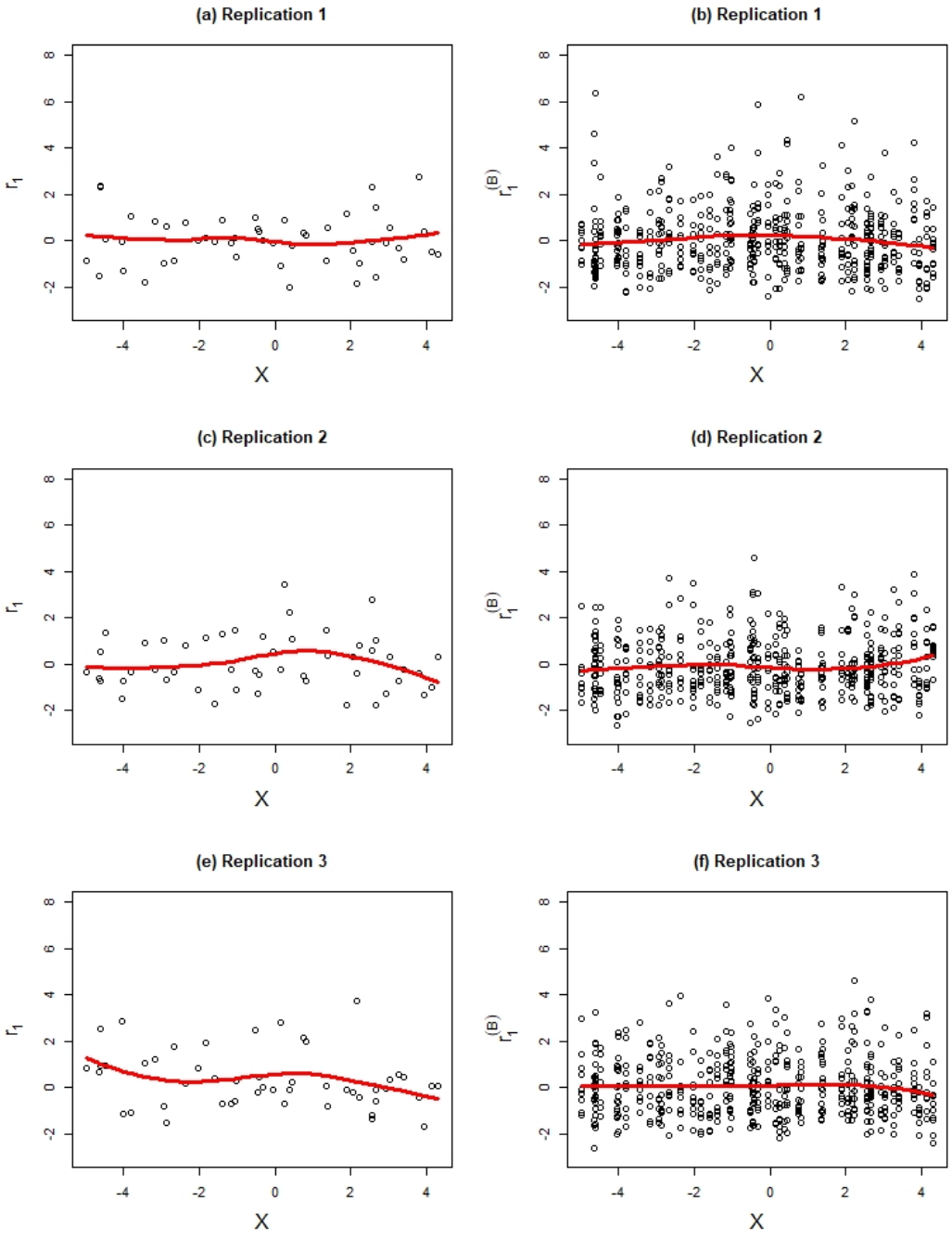

Figure S8: Three replications of r_1-by-x plots by the surrogate residual (r_1) and bootstrapping residual ($r_1^{(B)}$) with 10 copies. Upper row: the first replication; Middle row: the second replication; Bottom row: the third replication.

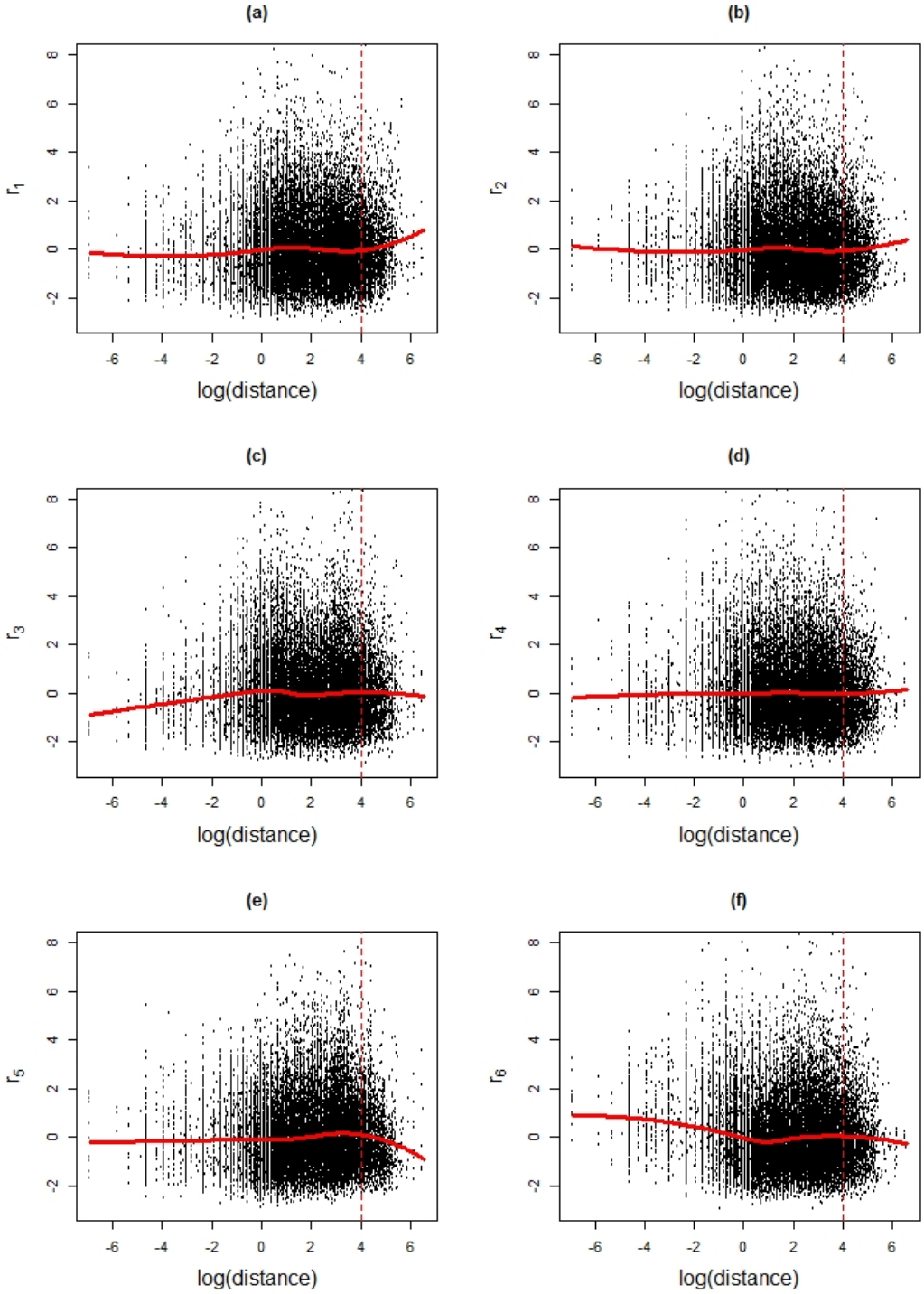

Figure S9: Residual-by-log(distance) plots for our initial multinomial logistic regression fitted to the MPN data. r_1 to r_6 represent the surrogate residuals for the utilities of car as driver, car as passenger, bicycle, ebike, public transport, and walking, respectively. A vertical dashed line at $X = 4$ and a loess curve (solid) are added for reference.

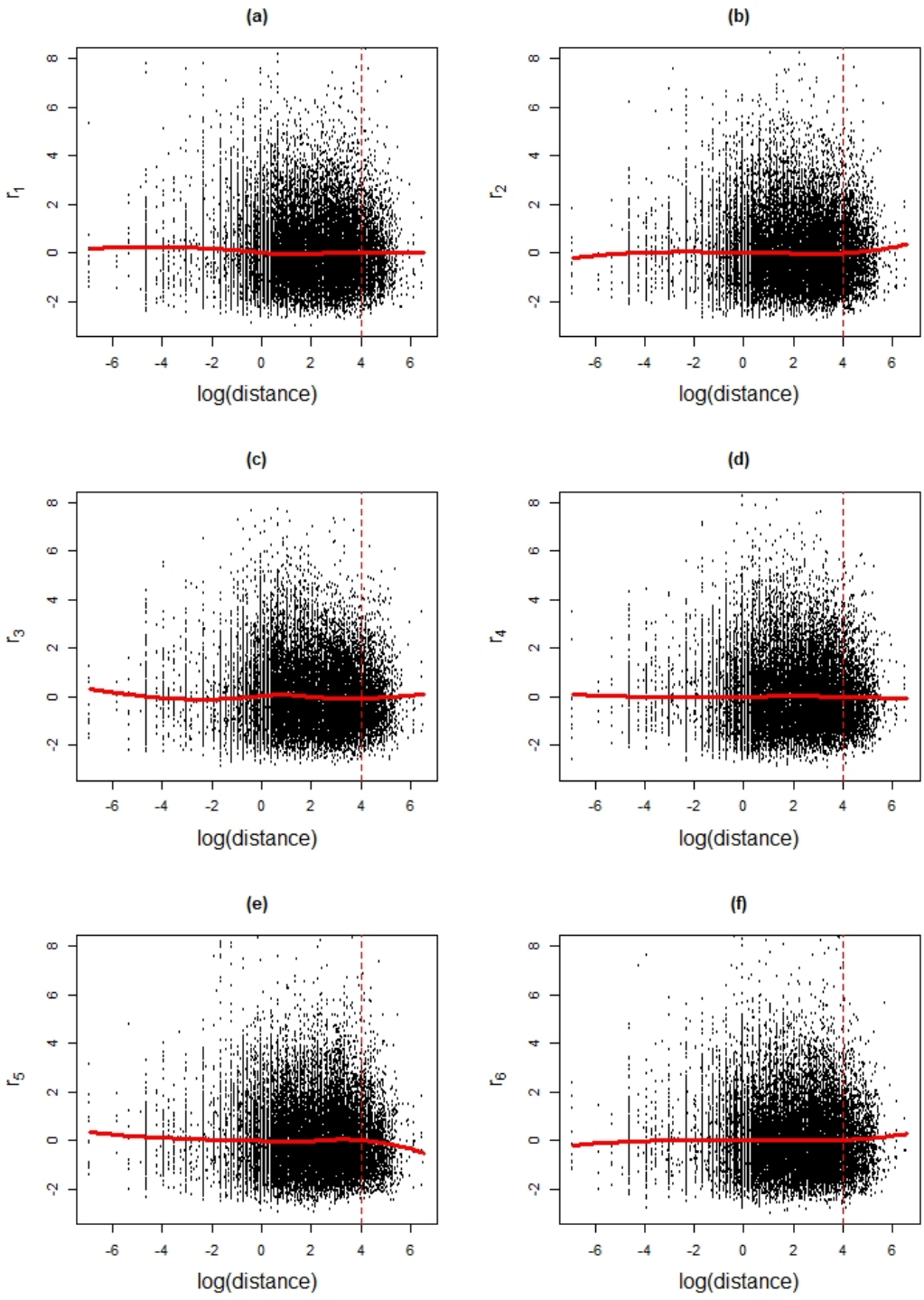

Figure S10: Residual-by-log(distance) plots when natural logarithm term of distance are applied in the multinomial logistic regression fitted to the MPN data. r_1 to r_6 represent the surrogate residuals for the utilities of car as driver, car as passenger, bicycle, ebike, public transport, and walking, respectively. A vertical dashed line at $X = 4$ and a loess curve (solid) are added for reference.

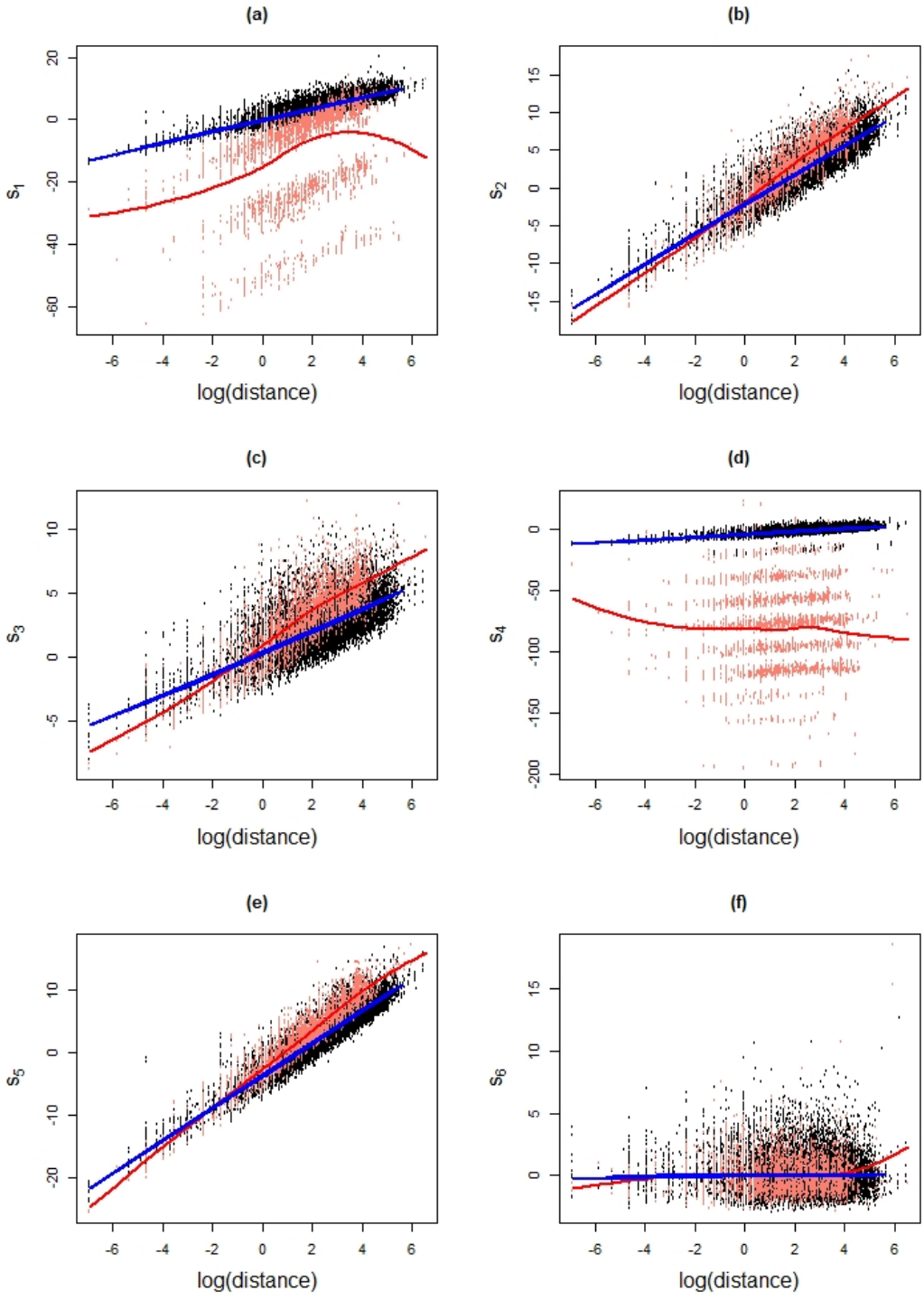

Figure S11: Plots of surrogate response versus log(distance) for the subjects who have a driving license and who do not have a driving license. s_1 to s_6 represent the surrogate responses corresponding to the utilities of car as driver, car as passenger, bicycle, ebike, public transport, and walking, respectively. The black points with a blue loess curve correspond to data from the subjects who have a driving license, while the orange points with a red loess curve correspond to data from the subjects who do not have a driving license.