

Supplementary Materials for “Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach”

Part A: Figures and tables

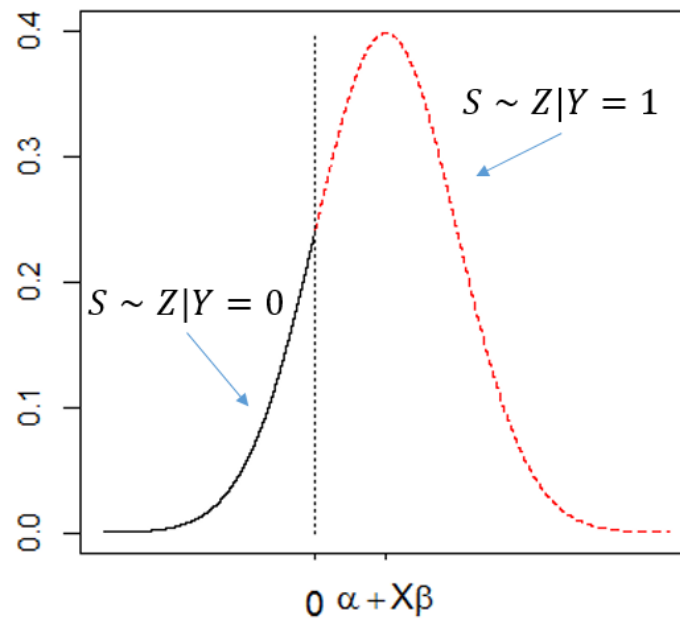


Figure 2: An illustration of the sampling procedure to generate a surrogate variable S .

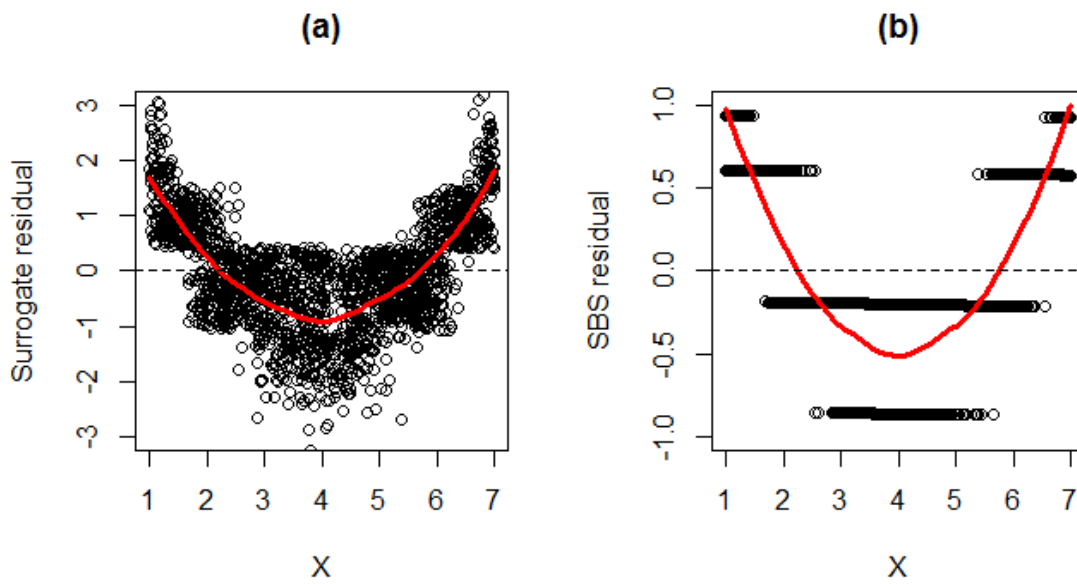


Figure 3: Residual-by-covariate plots using (a) the surrogate residual and (b) the SBS residual, when the mean structure is misspecified in the assumed model. Loess curves (red solid) are added as references.

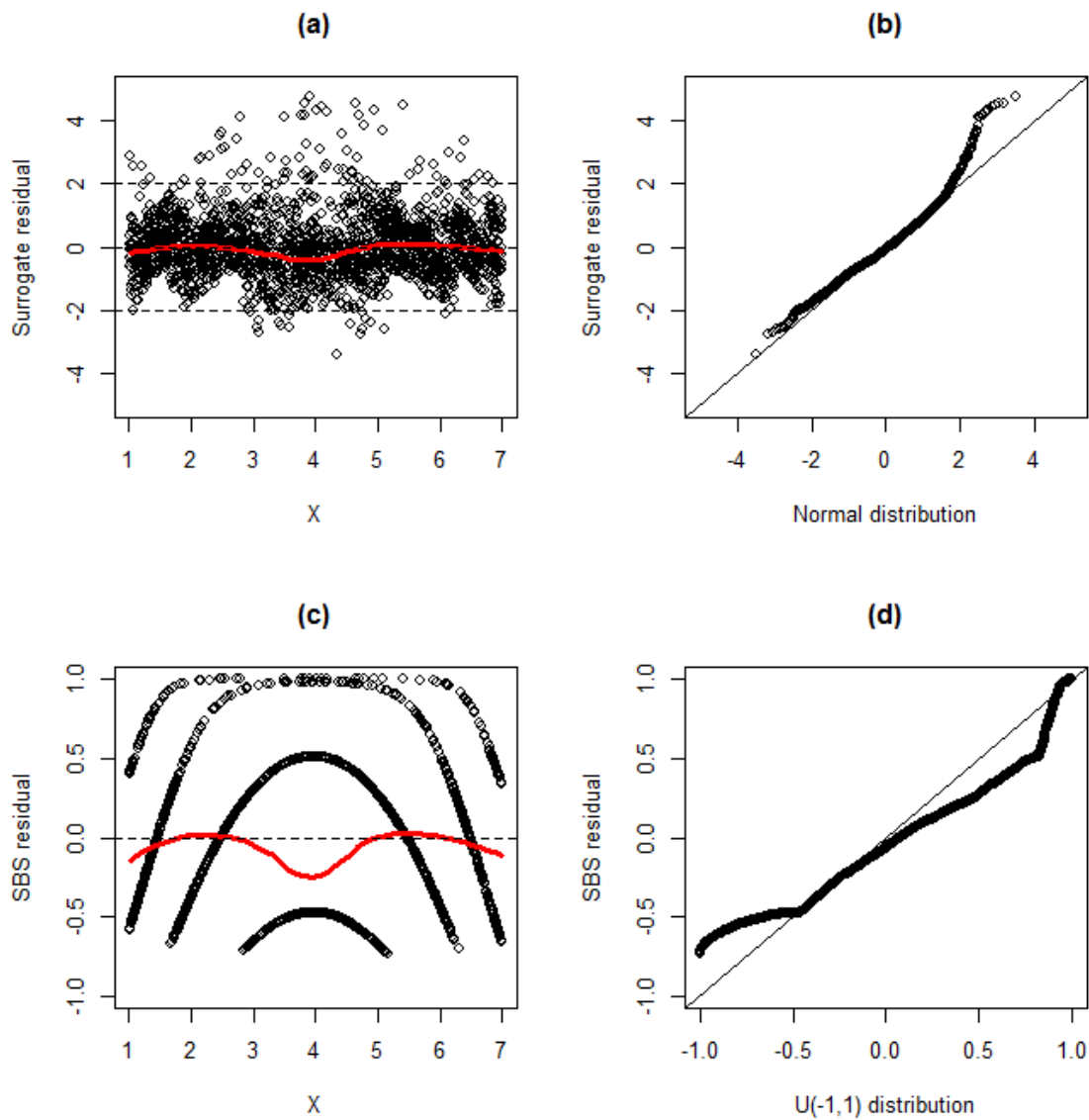


Figure 4: Model diagnostics when the link function in the assumed model is mis-specified. The figure includes (a) a plot of our residual versus the covariate X ; (b) a QQ-plot of our residual versus the standard normal distribution; (c) a plot of the SBS residual versus the covariate X ; and (d) a QQ-plot of the SBS residual versus the $U(-1,1)$ distribution. Loess curves (red solid) are added to plots (a) and (c) as references.

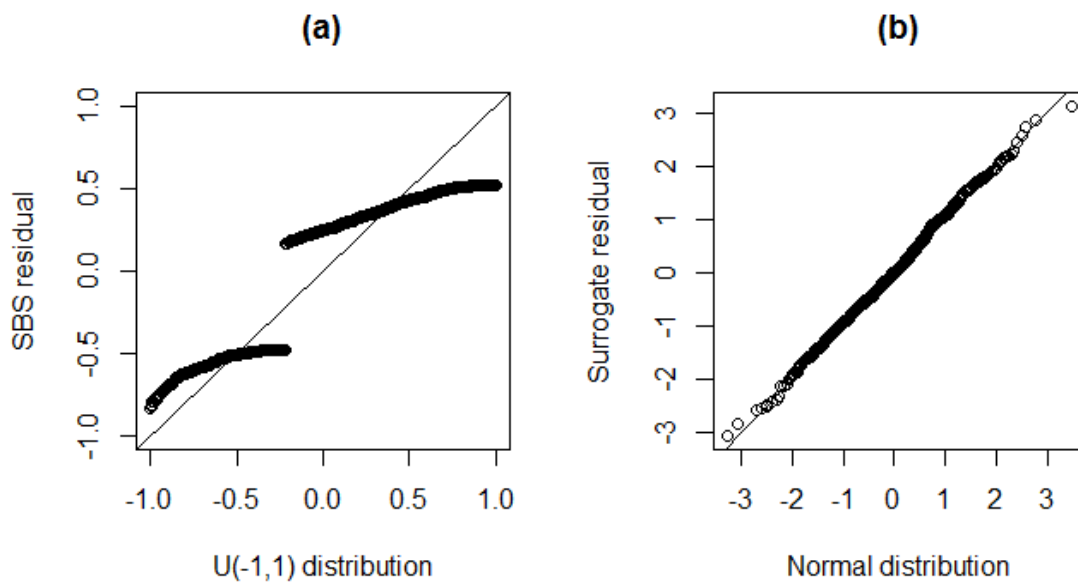


Figure 5: QQ plots in the same setting as Example 1 except the range of X restricted to $[3, 5]$. The figure includes (a) a QQ-plot of the SBS residual versus the $U(-1,1)$ distribution; and (b) a QQ-plot of our residual versus the standard normal distribution

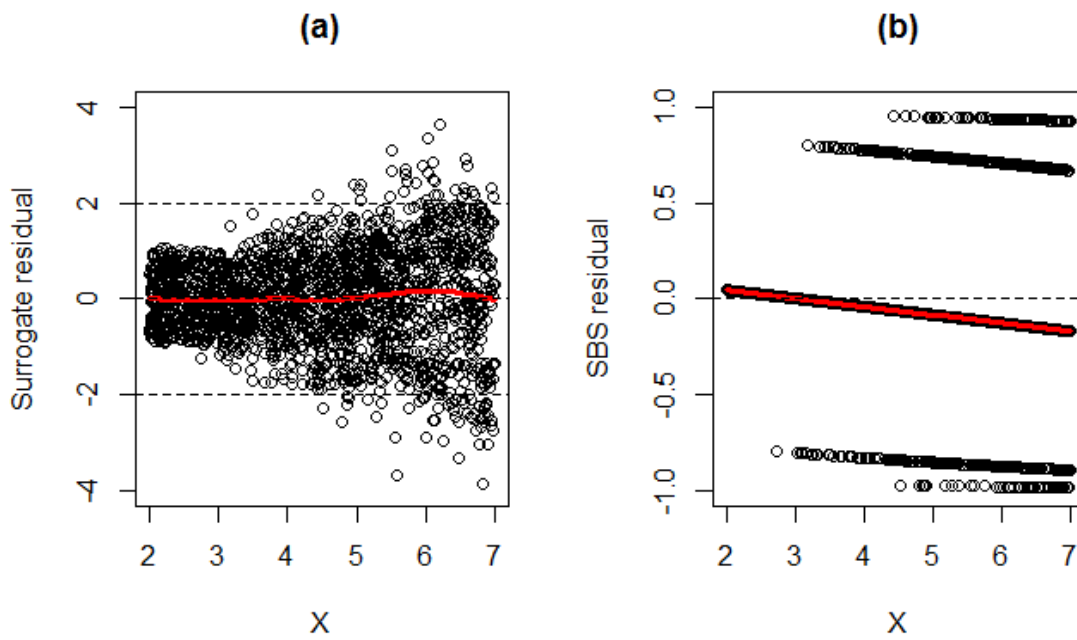


Figure 6: Residual-by-covariate plot of the surrogate residual (a) and the SBS residual (b) when there exists heteroscedasticity in the true model whereas a homoscedastic model is fitted to the data. Loess curves (red solid) are added as references.

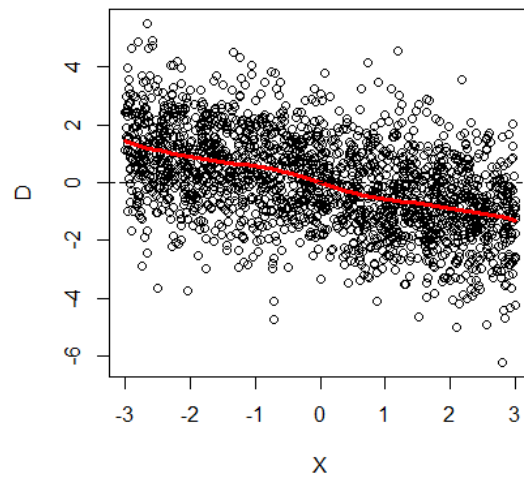


Figure 7: Diagnostic plot for checking the proportional assumption. Plotted are $D = S_2 - S_1$ versus X , where $D | X \sim N((\beta_1 - \beta_2)X, 2)$. A Loess curve (red solid) is added to illustrate a descending trend of the regression mean.

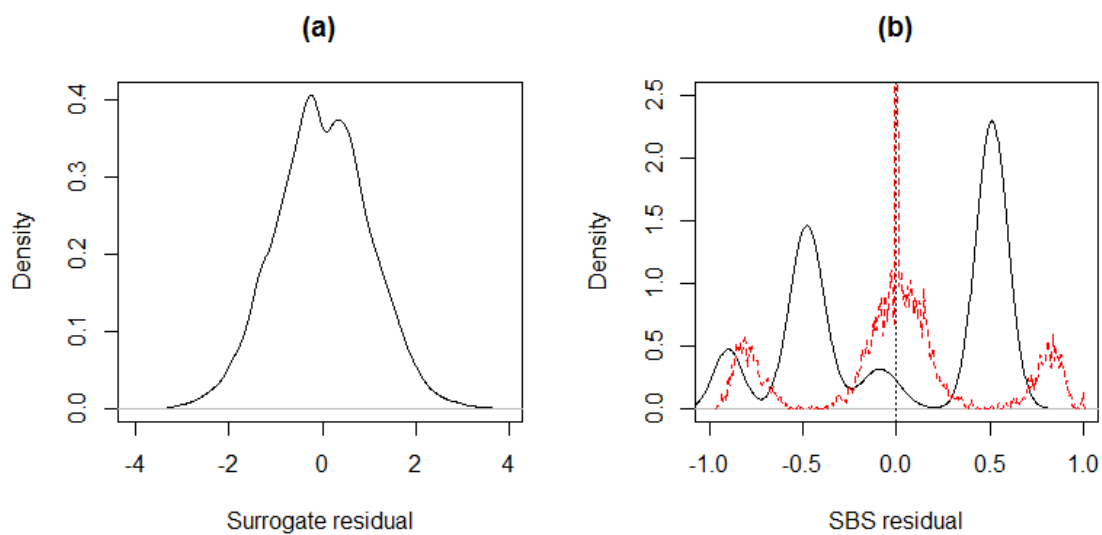


Figure 8: Density plot of the surrogate residual (a) and the SBS residual (b) when the assumed model excludes an indicator variable that represents a mixed-population effect. As references, we include in (b) the density curve when the assumed model is specified correctly (red dashed) and a vertical dotted line that represents the mean of the SBS residuals.

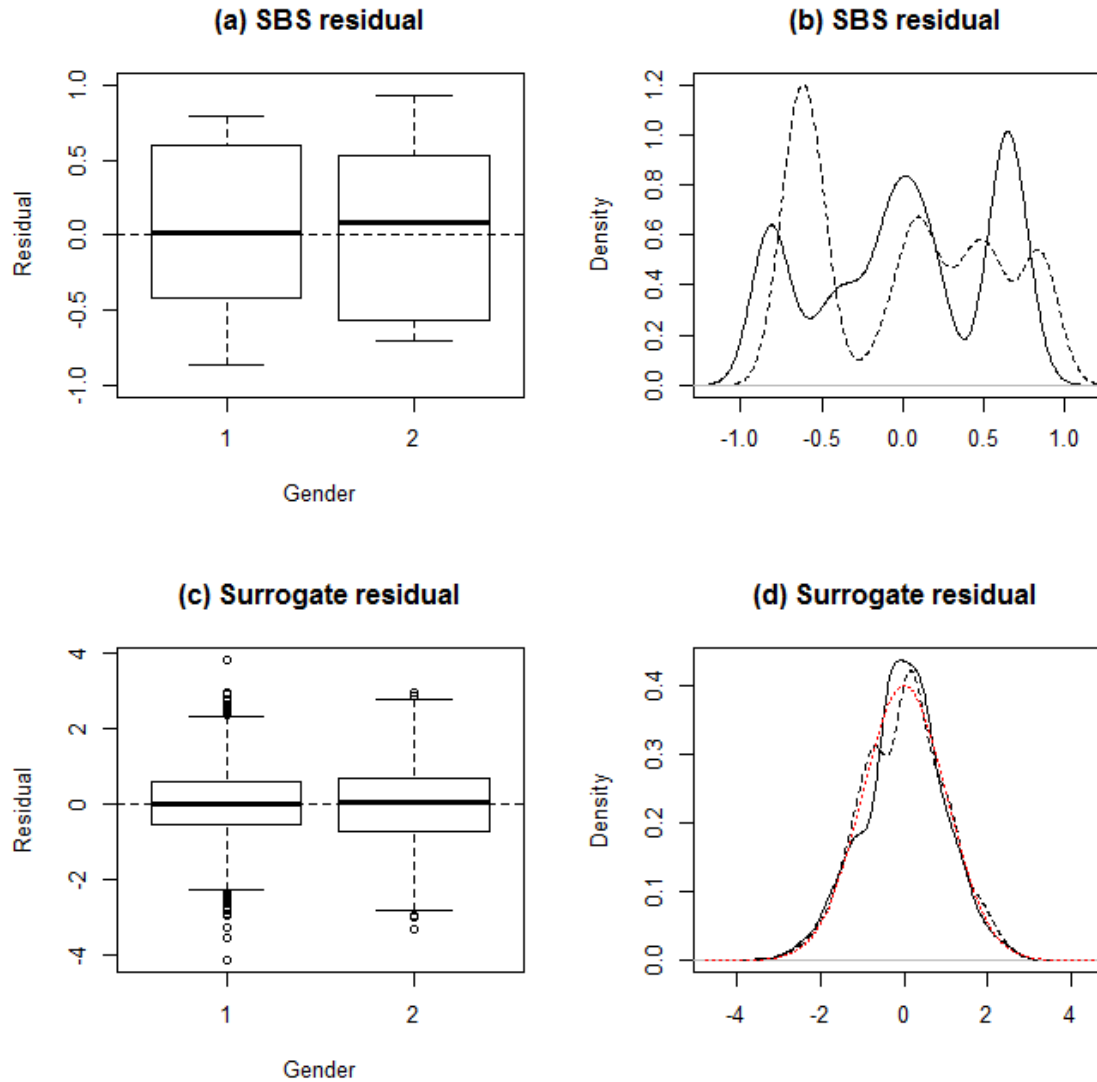


Figure 9: Residual-by-covariate plots for checking the ordered probit model fitted to the SAGE data. (a) Boxplot of the probability-based residual (1: male & 2: female); (b) Density plot of the probability-based residual (solid line: male & dashed line: female); (c) Boxplot of our residual; (d) Density plot of our residual (dotted line: the standard normal density). The sample size of male and female subjects is 1539 and 1858, respectively.

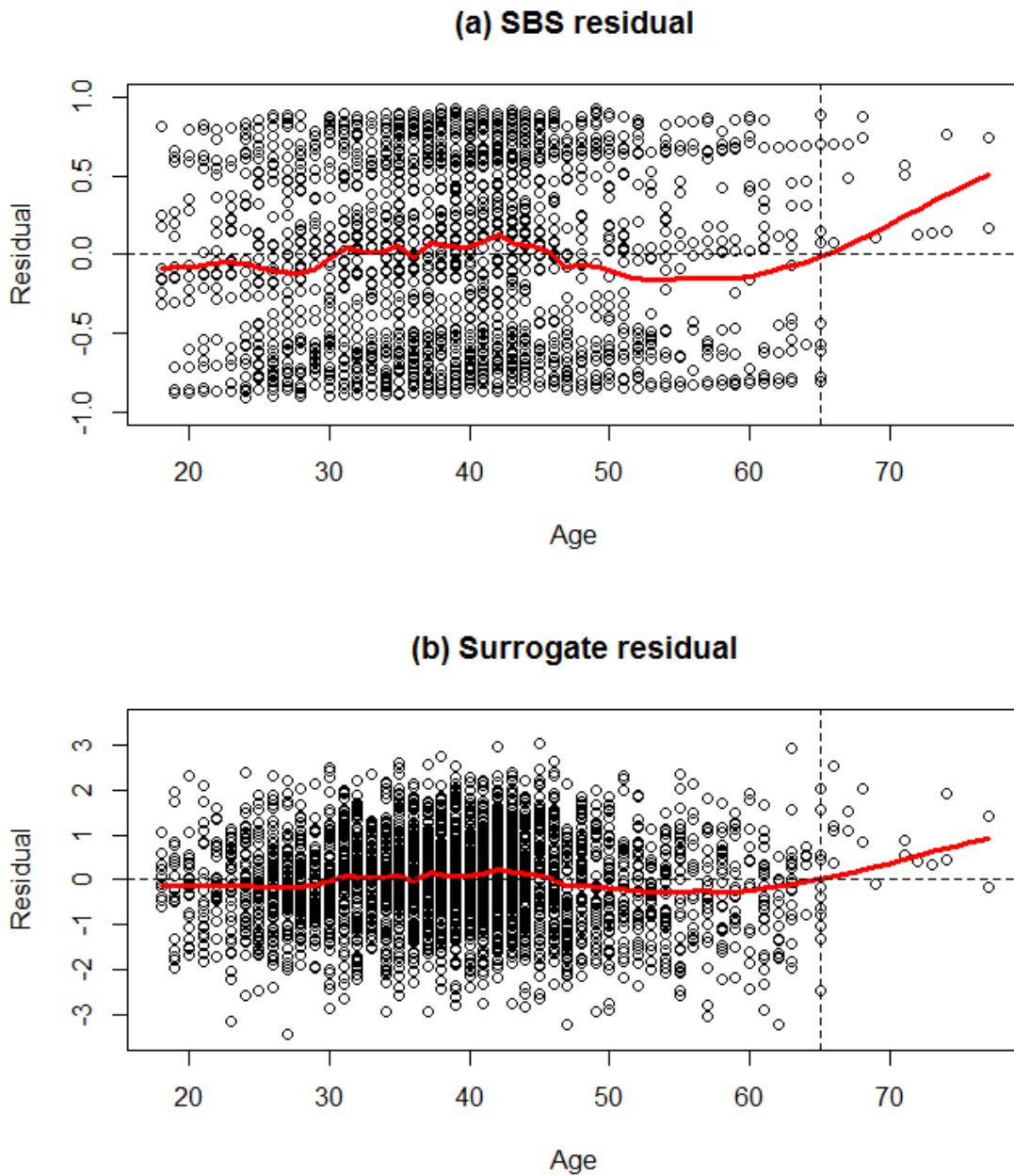


Figure 10: Residual-by-covariate plots when a linear term of age is included in the ordered probit model fitted to the SAGE data. Shown are the SBS (a) and our (b) residuals. A horizontal dashed line at $y = 0$, a vertical dashed line at $x = 65$ and a Loess curve (solid) are added for reference.

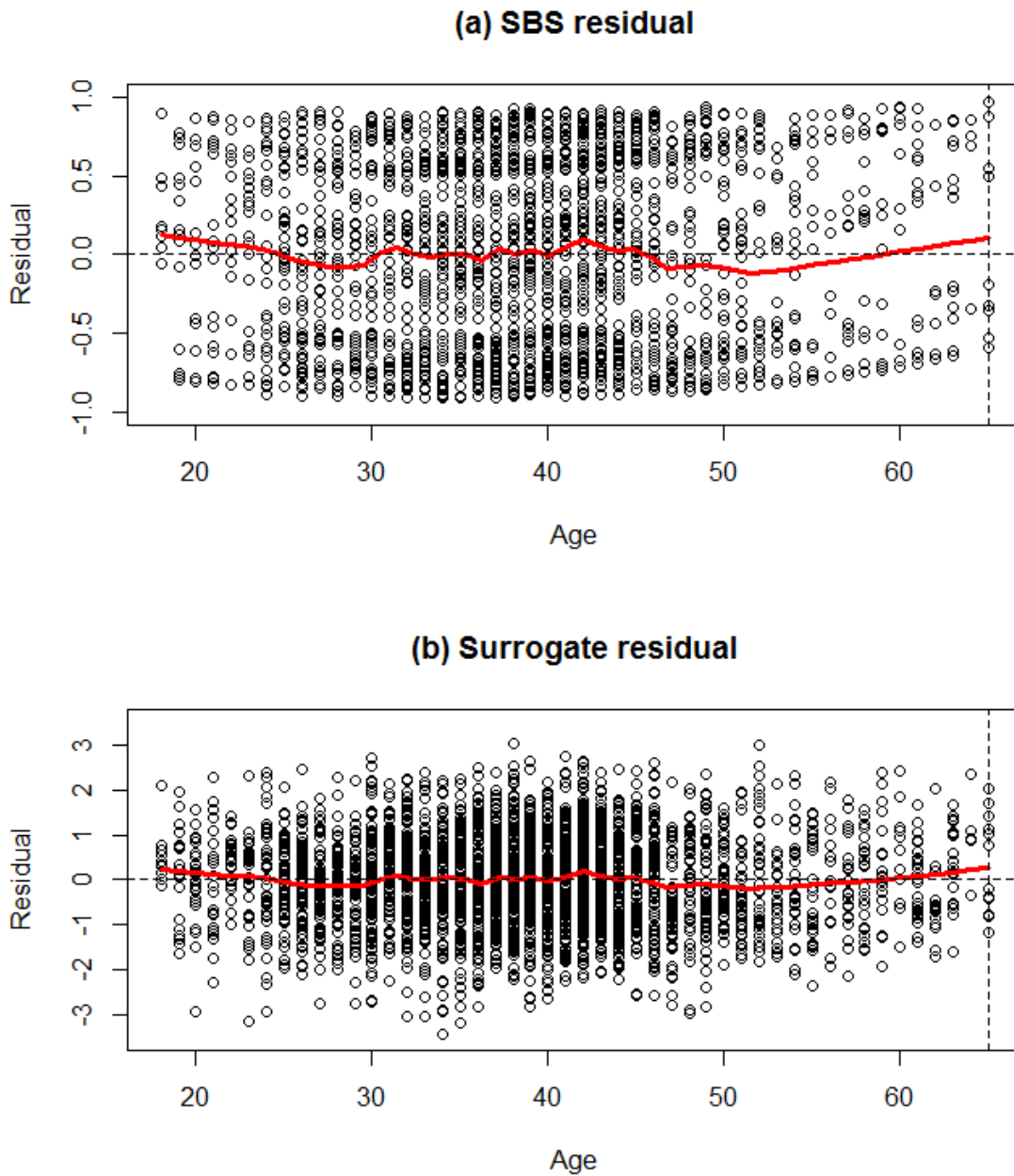


Figure 11: Residual-by-covariate plots when both linear and quadratic terms of age are included in the ordered probit model fitted to the SAGE data. Shown are the SBS (a) and our (b) residuals. A horizontal dashed line at $y = 0$, a vertical dashed line at $x = 65$ and a Loess curve (solid) are added for reference.

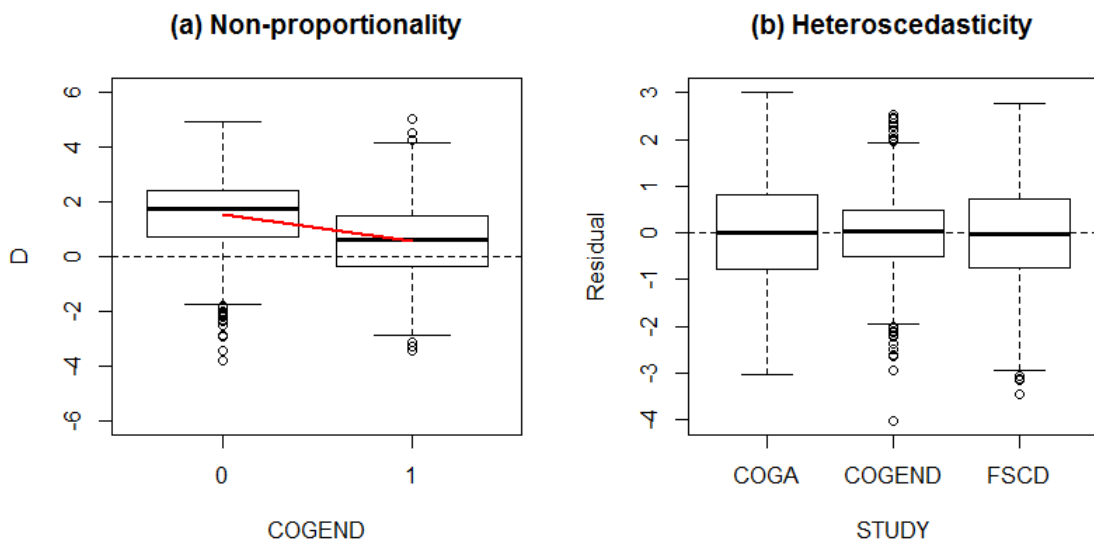


Figure 12: Evidence of non-proportionality and heteroscedasticity in the overarching model for all the three studies. Shown are boxplots for (a) D versus the indicator variable “COGEND”, where $D = S_3 - S_1$ is the difference between two surrogate variables; and (b) our residual versus the covariate study.

Table 1: Estimates of regression coefficients in separate binary models

Binary models	SNP	Gender	Race	COGA	COGEND	Age	Age ²
$\Pr\{Y \leq 1\}$	-0.094	-0.679	0.110	-0.009	0.082	0.058	-0.001
$\Pr\{Y \leq 2\}$	-0.099	-0.716	-0.127	0.422	-0.286	0.060	-0.001
$\Pr\{Y \leq 3\}$	-0.095	-0.435	0.121	0.271	-0.756	0.096	-0.001

Remark: Shown are maximum likelihood estimates of the coefficients β_j in the binary probit models $\Pr\{Y \leq j\} = \Phi(\alpha_j + \mathbf{X}\beta_j)$, $j = 1, 2, 3$.

Part B: Technical details for multiple sampling

The k -th copy of the empirical distribution is obtained by a two-step sampling:

Step 1. Bootstrap the ordinal outcome Y together with its corresponding covariates \mathbf{X} , and obtain a bootstrap sample $\{(\mathbf{X}_{1,k}^*, Y_{1,k}^*), \dots, (\mathbf{X}_{n,k}^*, Y_{n,k}^*)\}$.

Step 2. Perform the conditional sampling in Section 3.1, using the bootstrapped sample $\{(\mathbf{X}_{1,k}^*, Y_{1,k}^*), \dots, (\mathbf{X}_{n,k}^*, Y_{n,k}^*)\}$ and obtain a sample of residuals $\{R_{1,k}^*, \dots, R_{n,k}^*\}$.

Remark 4. *Our procedure of bootstrapping residuals is similar to that for linear regression models. To bootstrap residuals for a linear model $Y = X\theta + \varepsilon$, we fit the model to the original data and obtain a parameter estimate $\hat{\theta}$. By defining $\hat{\varepsilon}_i = y_i - x_i\hat{\theta}$, we have a set of residuals $\{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n\}$. Resampling randomly from this residual set (or observation index) yields a bootstrap copy of residuals (Efron and Tibshirani, 1994, pp.111-112). Bootstrapping residuals does not require that we resample the outcome-covariate pairs (\mathbf{X}, Y) or repeatedly re-estimate θ . This makes such a procedure sensitive to model assumptions (Efron and Tibshirani, 1994, pp.113-115), which is ideal for model diagnostics. For ordinal regression models, the only difference is that bootstrapping residuals requires an additional sampling according to a distribution that depends on the value of (\mathbf{X}, Y) . This is why we resample the outcome-covariate pairs as well.*

5.1 Diagnostic plots

For residual-by-covariate plots, we scatter all the $K \times n$ residuals on the same plot. It is valid since each residual is independently drawn from the empirical distribution of (\mathbf{X}, Y) (nonparametric bootstrap) and the conditional model (parametric bootstrap). For QQ plots, we use the “median” of the K empirical distributions,

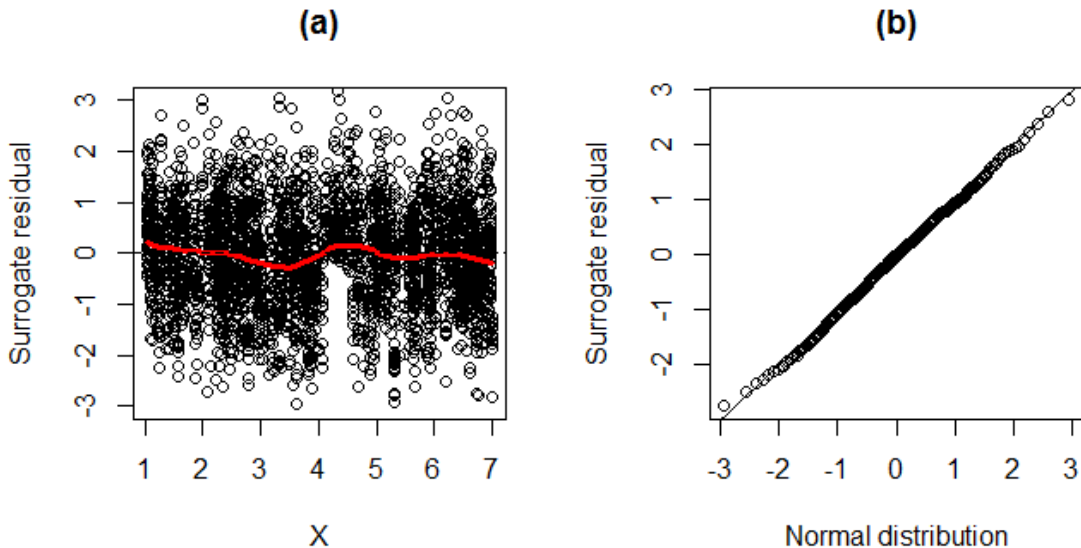


Figure 13: Diagnostic plots using our residuals based on multiple sampling ($K=10$) when the model is specified correctly ($n=300$). The figure includes (a) a plot of our residual versus the covariate X , with a Loess curve (red solid); and (b) a QQ-plot of our residual versus the standard normal distribution.

namely, $Q_{n, Mid}^*(t) \triangleq \text{Median}\{Q_{n,1}^*(t), \dots, Q_{n,K}^*(t)\}$. For any K , $Q_{n, Mid}^*(t)$ converges to $G(c + \int u dG(u))$ under the null, in view of Theorem 3(c).

For illustration, Figure 13 shows the residual-by-covariate plot and QQ plot in the setting of Example 1 where the model is specified correctly. Figures 14(a) and (b) show the residual-by-covariate plot and QQ plot in the setting of Examples 2 and 3, respectively, where the model is misspecified in the mean structure and link function. In all the settings, the sample size $n = 300$ and $K = 10$. The plots all look similar to those in Examples 1-3 where $n = 2000$. We note that the purpose of multiple samplings is to minimize the sampling error and stabilize the patterns. It should not be confused with generating more data (evidence). For example, the evidence of deviation in Figure 14(b) ($n \times K = 3000$) is weaker than that in Figure 4(b) ($n=2000$).

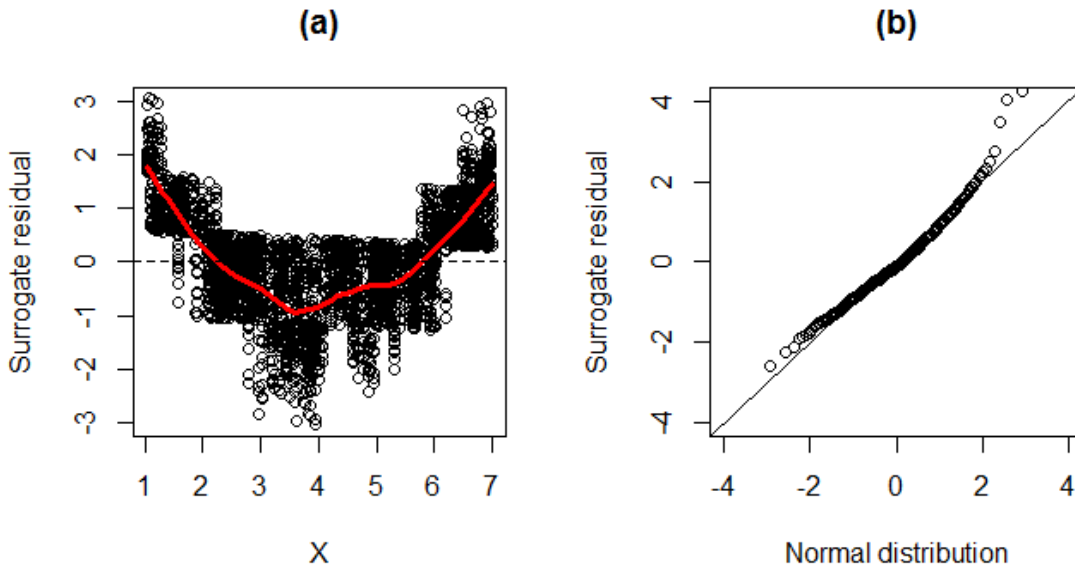


Figure 14: Diagnostic plots using our residuals based on multiple sampling ($K=10$) when the model is not specified correctly ($n=300$). The figure includes (a) a plot of our residual versus the covariate X when the mean structure is misspecified, with a Loess curve (red solid); and (b) a Q-Q-plot of our residual versus the standard normal distribution when the link function is misspecified.

5.2 Goodness-of-fit measure based on distance notions

As a continuous variable, our residual enables us to make use of existing goodness-of-fit measures. Let $d(F, G)$ be a metric on the space of distribution functions, measuring the discrepancy between two distribution functions $F(t)$ and $G(t)$. A classical example is the Kolmogorov-Smirnov distance $d_{KS}(F, G) = \sup_{t \in \mathbb{R}} |F(t) - G(t)|$, which gives rise to a goodness-of-fit statistic

$$T_n(R_1, \dots, R_n) \equiv n^{1/2} d_{KS}(Q_n(t; R_1, \dots, R_n), G(t)).$$

If the assumed model (3) agrees with the true model (5), then $\{R_1, \dots, R_n\}$ is an i.i.d. sample from the distribution G , and thus we have that $\Pr\{T_n \leq t\} \rightarrow H(t) \equiv 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp(-2j^2 t^2)$ as $n \rightarrow \infty$ (Lehmann and Romano, 2005, pp.585). This result can be used to examine goodness-of-fit for the ordinal regression model

(3).

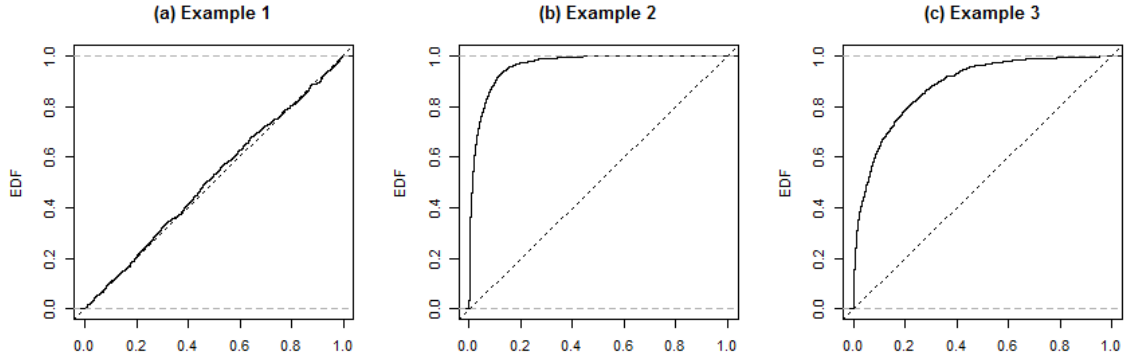


Figure 15: Empirical distribution functions (solid lines) of the bootstrapped goodness-of-fit test statistics $\{1-H(T_{n,1}^*), \dots, 1-H(T_{n,K}^*)\}$. The illustrations (a)-(c) correspond to Examples 1-3, respectively, when the sample size $n = 500$, $X \sim N(3, 4)$ and the bootstrap repetition $K = 1000$. The diagonal lines (dotted) represent the CDF of the $U(0,1)$ distribution and serve as references.

Given K bootstrap copies of $\{R_{1,k}^*, \dots, R_{n,k}^*\}$, we bootstrap goodness-of-fit statistics $\{T_{n,1}^*, \dots, T_{n,K}^*\}$ where $T_{n,k}^* = T_n(R_{1,k}^*, \dots, R_{n,k}^*) = n^{1/2}d_{KS}(Q_{n,k}^*, G)$. Considering that the empirical distribution function (EDF) of $\{T_{n,1}^*, \dots, T_{n,K}^*\}$

$$H_{n,K}(t) = \frac{1}{K} \sum_{k=1}^K I(T_{n,k}^* \leq t) \rightarrow \Pr\{T_n \leq t\},$$

as both n and $K \rightarrow \infty$, we compare $H_{n,K}(t)$ with $H(t)$. Equivalently, we compare the EDF of $\{1 - H(T_{n,1}^*), \dots, 1 - H(T_{n,K}^*)\}$ with the CDF of the uniform distribution $U(0,1)$. Figures 15(a)-(c) display such EDFs (solid curves) for Examples 1-3, respectively, together with the CDF of the $U(0,1)$ distribution (dotted curves) as a reference. The discrepancy between the solid and dotted curves in Figures 15(b)-(c) is a clear indication of model misspecification. We note that for hypothesis testing, the approximations such as $H_{n,K}(t) \rightarrow \Pr\{T_n \leq t\}$ and $\Pr\{T_n \leq t\} \rightarrow H(t)$ may not be adequate for controlling the type I error, when parameter estimates are used to derive residuals. To solve this issue, we use some resampling methods to better

approximate the null distribution and adjust the p -value. The details, together with power analysis, are presented below.

5.3 Power analysis of goodness-of-fit tests

To better approximate the null distribution and improve the power, we propose (i) instead of using the $U(0,1)$ distribution as seen in Figure 15, to obtain the null distribution of $\{T_{n,1}^*, \dots, T_{n,K}^*\}$ by bootstrapping $\{(\mathbf{X}_{1,k}^*, Y_{1,k}^*), \dots, (\mathbf{X}_{n,k}^*, Y_{n,k}^*)\}$ in Step 1 under the null model; and (ii) to use the adjusted p -value $p_{adj} = \Pr\{P^* \leq p \mid \hat{F}_0\}$ where we estimate the distribution of the p -value random variable P by resampling from the null model (Davison and Hinkley, 1997, pp.175-177).

We study the power of our goodness-of-fit test. We compare it with a test derived from the expectation-based residual in Proposition 2(b). Specifically, the test statistic for comparison is $V = E\{R^2 \mid X, Y\}$ and its null distribution is the distribution of $E\{R^2 \mid X, F_0\}$, which can be simulated by bootstrapping. We call such a test a SBS-type test, since the expectation-based residual captures information similar to the SBS residual (Li and Shepherd, 2012). We also include in the comparison a likelihood ratio test (Lipsitz et al., 1996) and a Hosmer-Lemeshow-type test (Fagerland and Hosmer, 2013). The result for $n = 200$ is reported in Table 2.

The third column of Table 2 shows that the SBS-type residual virtually has no power to detect model misspecification in all the examples considered. In the situation of the null model (Example 1), the SBS-type test has type I error rate (power) close to the nominal level only when the true (unknown) value of parameters are used to derive the null distribution. If parameter estimates are used instead as in our simulation, its type I error rate becomes zero, so does its power for alternative models. This observation remains as the sample size increases to $n = 5000$. The result suggests that the SBS-type test is not sensitive to model misspecification, as the SBS-type residual

Table 2: Power analysis of goodness-of-fit tests

	Surrogate residual (Multiple-sampling)	Expected residual (SBS-type test)	Lipsitz et al.'s (Likelihood ratio)	Fagerland-Hosmer (Chi-square)
Significance level	Example 1 (The null model)			
0.05	0.050	0.000	0.031	0.016
0.10	0.091	0.000	0.066	0.018
	Example 2 (Missing quadratic term)			
0.05	0.996	0.000	1.000	1.000
0.10	0.996	0.000	1.000	1.000
	Example 3 (Wrong link function)			
0.05	0.919	0.000	0.290	0.950
0.10	0.940	0.000	0.383	0.962
	Example 4 (Presence of heteroscedasticity)			
0.05	0.460	0.000	0.078	1.000
0.10	0.559	0.000	0.130	1.000
	Example 5 (Violation of proportionality)			
0.05	0.193	0.000	0.288	0.209
0.10	0.269	0.000	0.407	0.250
	Example 6 (Omission of mixed populations)			
0.05	0.066	0.000	0.076	0.161
0.10	0.141	0.000	0.138	0.257

only has the first-moment property and its variance/range/distribution varies across the values of X as summarized in the properties (\mathcal{P} -1) and (\mathcal{P} -2). The considerable difference between the second and third columns of Table 2 demonstrates that our surrogate residual is more powerful than the SBS-type residual, as our method makes full use of distributional information.

Table 2 shows that our test is comparable to Lipsitz et al.'s and Fagerland-Hosmer methods. Consistent with Fagerland and Hosmer (2013), our result shows that there

may not exist a method that is always the most powerful in all circumstances. The testing power often depends on the alternative models.

References

- Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, Cambridge university press.
- Efron, B. and Tibshirani, R. J. (1994), *An introduction to the bootstrap*, CRC press.
- Fagerland, M. W. and Hosmer, D. W. (2013), “A goodness-of-fit test for the proportional odds regression model,” *Statistics in Medicine*, 32, 2235–2249.
- Lehmann, E. L. and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York:Springer, 3rd ed.
- Lipsitz, S. R., Fitzmaurice, G. M., and Molenberghs, G. (1996), “Goodness-of-fit tests for ordinal response regression models,” *Applied Statistics*, 45, 175–190.

Part C: Proofs

Proof of Theorem 1. If $f_a(y) \equiv f_0(y)$ almost surely, then $S \sim \int f_a(z | y)f_0(y)dy/m_c = \int f_a(z | y)f_a(y)dy/m_c = \int f_a(z, y)dy/m_c = f_a(z)$. This completes the proof. \square

Proof of Theorem 2. First we derive the distribution of S when the true model for Y is (5). Conditional on X , for any arbitrary but fixed c such that $\alpha_{k-1} \leq c < \alpha_k$, $1 \leq k \leq J$,

$$\begin{aligned} \Pr\{S \leq c\} &= \sum_{j=1}^J \Pr\{S \leq c | Y = j\} \Pr\{Y = j\} \\ &= \Pr\{Y \leq k-1\} + \Pr\{S \leq c | Y = k\} \Pr\{Y = k\} \\ &= \Pr\{Y \leq k-1\} + \frac{\Pr\{\alpha_{k-1} < Z \leq c\}}{\Pr\{\alpha_{k-1} < Z \leq \alpha_k\}} \Pr\{Y = k\}. \end{aligned}$$

Since $\Pr\{Y \leq k\} = G_0(\tilde{\alpha}_k + f_0(\mathbf{X}, \tilde{\boldsymbol{\beta}}))$ and $\Pr\{Z \leq c\} = G(c + f(\mathbf{X}, \boldsymbol{\beta}))$, we obtain the equations (6) and (7). If the true underlying model (5) agrees with the assumed model (3), we obtain from (6) that $\Pr\{S \leq c\} = G(c + f(\mathbf{X}, \boldsymbol{\beta}))$, showing that S has the same distribution as Z . It follows immediately that the residual R follows the distribution $G(c + \int u dG(u))$ conditional on any \mathbf{X} . Since this distribution function $G(c + \int u dG(u))$ does not rely on \mathbf{X} , the unconditional distribution of R remains the same. This completes the proof of Theorem 2. \square

Proof of Theorem 3. Parts(a) and (b) are direct results from Theorem 2. Part(c) follows Glivenko-Cantelli Theorem (Lehmann and Romano, 2005, pp 441). \square

Proof of Proposition 1. This is a direct result from the observation $s_k < s_j$. \square

Proof of Proposition 2. Part (b) is obvious. We only derive Part (a). Under the assumed model (3), we have $\Pr\{y > Y\} = G(\alpha_{y-1} + f(\mathbf{X}, \boldsymbol{\beta}))$. Notice that $\alpha_{y-1} = \min(S | y) = \min(R | y) - f(\mathbf{X}, \boldsymbol{\beta}) + \int u dG(u)$. Thus $\Pr\{y > Y\} = G(\min(R | y) + \int u dG(u))$.

Similarly, we can derive $\Pr\{y < Y\} = 1 - G(\max(R | y) + \int udG(u))$. Thus the formula for r^P follows. The statements of its conditional and unconditional expectation follow the results in Li and Shepherd (2012). \square

Proof of Theorem 4. The result in Part(a) follows from the definition of R . For Part(b), it is straightforward to show that $S | \mathbf{X} \sim U(0, 1)$. Thus, we have $R | \mathbf{X} \sim U(-1/2, 1/2)$. \square

Proof of Proposition 3. We write $2E\{S | Y = y, \mathbf{X}\} = F_a(y-1) + F_a(y) = \Pr_{F_a}\{Y < y\} + (1 - \Pr_{F_a}\{Y > y\}) = R^P + 1$. Thus, we have $E_0\{S | \mathbf{X}\} = 1$ and $R^{SBS} = 2E\{R | Y, \mathbf{X}\}$. \square

Part D: Properties of $\hat{R}_{\hat{\alpha}, \hat{\beta}}$

Assume that Y follows the cumulative link regression model

$$G^{-1}(\Pr\{Y \leq j\}) = \hat{\alpha}_j + f(\mathbf{X}, \hat{\beta}), \quad (10)$$

where the parameter estimates $\hat{\alpha} = \alpha + o_p(1)$, $\hat{\beta} = \beta + o_p(1)$. Under the assumption that $f(\mathbf{X}, \hat{\beta}) = f(\mathbf{X}, \beta) + o_p(1)$, we have the following results for $\hat{R}_{\hat{\alpha}, \hat{\beta}}$, which are parallel to Theorems 2-3 for the residual variable R as presented in Section 3.2.

Theorem 2'. If the true underlying model (5) agrees with the assumed model (10) (i.e., $\alpha = \tilde{\alpha}, \beta = \tilde{\beta}, G = G_0, f = f_0$), then the distribution of $\hat{R}_{\hat{\alpha}, \hat{\beta}}$ converges weakly to $G(c + \int u dG(u))$ as $n \rightarrow \infty$, conditional or unconditional on X , i.e., $\Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c \mid \mathbf{X}\} = \Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c\} = G(c + \int u dG(u)) + o(1)$.

Theorem 3'. If the true underlying model (5) agrees with the assumed model (10) (i.e., $\alpha = \tilde{\alpha}, \beta = \tilde{\beta}, G = G_0, f = f_0$), then $\hat{R}_{\hat{\alpha}, \hat{\beta}}$ has the following properties:

- (a) $E\{R \mid \mathbf{X}\} = o(1)$.
- (b) $Var\{R \mid \mathbf{X}\}$ is a constant, not depending on \mathbf{X} , except a vanishing term $o(1)$.
- (c) $\sup_{c \in \mathbb{R}} |\hat{Q}_n(c) - G(c + \int u dG(u))| \rightarrow 0$ almost surely, where $\hat{Q}_n(c) = \frac{1}{n} \sum_{i=1}^n I(\hat{R}_i \leq c)$ is the empirical cumulative distribution of $\{\hat{R}_1, \dots, \hat{R}_n\}$.

We prove Theorems 2'-3' in the following. First, we derive the distribution of \hat{S} that follows the truncated distribution obtained by truncating the distribution of $\hat{Z} = -f(\mathbf{X}, \hat{\beta}) + \epsilon$ using the interval $(\hat{\alpha}_{y-1}, \hat{\alpha}_y)$ given $Y = y$. Conditional on X , for

any arbitrary but fixed c such that $\alpha_{k-1} \leq c < \alpha_k$, $1 \leq k \leq J$,

$$\begin{aligned}
& | \Pr\{\hat{S} \leq c\} - \Pr\{S \leq c\} | \\
& \leq \sum_{j=1}^J | \Pr\{\hat{S} \leq c \mid Y = j\} - \Pr\{S \leq c \mid Y = j\} | \Pr\{Y = j\} \\
& \leq \Pr\{c < \hat{Z} \leq \hat{\alpha}_{k-1}\} \Pr\{Y = k-1\} + \Pr\{\hat{\alpha}_k < \hat{Z} \leq c\} \Pr\{Y = k+1\} \\
& \quad + \left(\frac{\Pr\{\hat{\alpha}_{k-1} < \hat{Z} \leq c\}}{\Pr\{\hat{\alpha}_{k-1} < \hat{Z} \leq \hat{\alpha}_k\}} - \frac{\Pr\{\alpha_{k-1} < Z \leq c\}}{\Pr\{\alpha_{k-1} < Z \leq \alpha_k\}} \right) \Pr\{Y = k\} \\
& \leq \Pr\{\alpha_{k-1} < Z + o_p(1) \leq \alpha_{k-1} + o_p(1)\} \Pr\{Y = k-1\} \\
& \quad + \Pr\{\alpha_k + o_p(1) < Z + o_p(1) \leq \alpha_k\} \Pr\{Y = k+1\} \\
& \quad + \left(\frac{\Pr\{\alpha_{k-1} + o_p(1) < Z + o_p(1) \leq c\}}{\Pr\{\alpha_{k-1} + o_p(1) < Z + o_p(1) \leq \alpha_k + o_p(1)\}} - \frac{\Pr\{\alpha_{k-1} < Z \leq c\}}{\Pr\{\alpha_{k-1} < Z \leq \alpha_k\}} \right) \Pr\{Y = k\} \\
& \leq o(1) + o(1) + \left(\frac{\Pr\{\alpha_{k-1} < Z \leq c\} + o(1)}{\Pr\{\alpha_{k-1} < Z \leq \alpha_k\} + o(1)} - \frac{\Pr\{\alpha_{k-1} < Z \leq c\}}{\Pr\{\alpha_{k-1} < Z \leq \alpha_k\}} \right) \Pr\{Y = k\} \\
& = o(1).
\end{aligned}$$

Thus, we can establish $\Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c \mid \mathbf{X}\} = \Pr\{R \leq c \mid \mathbf{X}\} + o(1)$. The unconditional distribution of $\hat{R}_{\hat{\alpha}, \hat{\beta}}$

$$\Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c\} = \int \Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c \mid \mathbf{x}\} d\mu(\mathbf{x}) \rightarrow \int \Pr\{R_{\alpha, \beta} \leq c \mid \mathbf{x}\} d\mu(\mathbf{x}) = \Pr\{R \leq c\},$$

by Lebesgue's Dominated Convergence Theorem. Theorem 2' has been proved.

Based on Theorem 2', we have $E\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \mid \mathbf{X}\} = E\{R \mid \mathbf{X}\} + o(1)$ and $Var\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \mid \mathbf{X}\} = Var\{R \mid \mathbf{X}\} + o(1)$. Thus, Theorem 3'(a)-(b) follow. To see Theorem 3'(c),

$$\begin{aligned}
& \sup_{c \in \mathbb{R}} | \hat{Q}_n(c) - G(c + \int u dG(u)) | \\
& \leq \sup_{c \in \mathbb{R}} | \hat{Q}_n(c) - \Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c\} | + \sup_{c \in \mathbb{R}} | \Pr\{\hat{R}_{\hat{\alpha}, \hat{\beta}} \leq c\} - G(c + \int u dG(u)) | \\
& \rightarrow 0 \quad \text{almost surely,}
\end{aligned}$$

by Glivenko-Cantelli Theorem and Polyá's Theorem (Lehmann and Romano, 2005, pp 429). This completes the proof of Theorem 3'. \square

Part E: An example in support of Remark 3

We show that a QQ-plot of the SBS residuals mentioned in Remark 3 may not be informative for diagnostics. Specifically, we simulate two data sets ($n = 2000$) from the following two models:

$$\Pr\{Y \leq j\} = \Phi(\alpha_j + 8X - X^2), \quad j = 1, 2, 3, 4, \quad (\alpha_1, \alpha_2, \alpha_3) = (-16, -12, -8) \quad (11)$$

$$\Pr\{Y \leq j\} = \Phi(\alpha_j - 0.01X), \quad j = 1, 2, 3, 4, \quad (\alpha_1, \alpha_2, \alpha_3) = (-1.05, 0.46, 1.54) \quad (12)$$

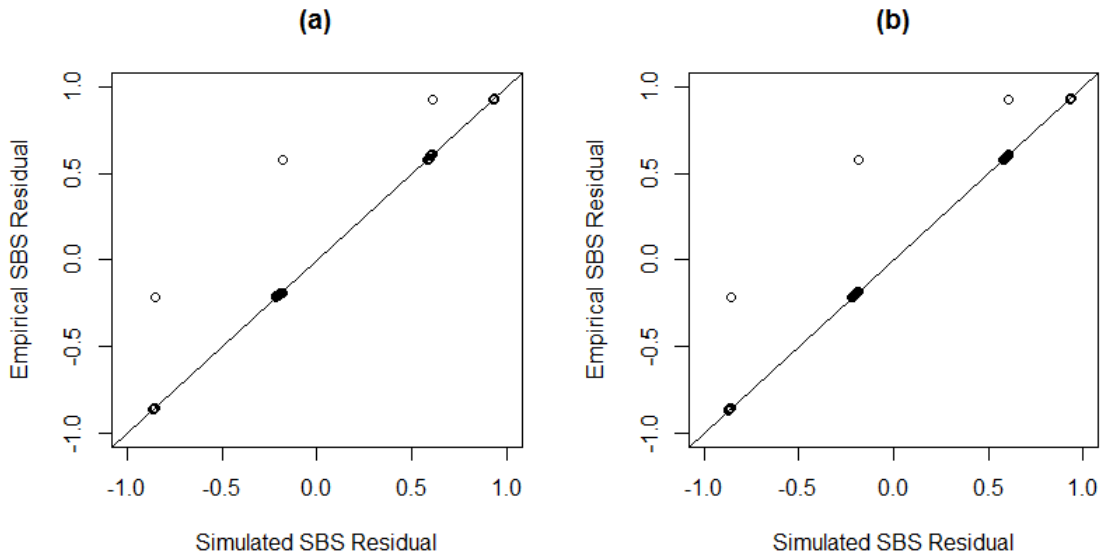


Figure 16: QQ-plots of the simulated versus empirical SBS residuals. The mean structure of the fitted model is linear while the mean structure of the true model is (a) quadratic or (b) linear.

We fit both data sets using the model $\Pr\{Y \leq j\} = \Phi(\alpha_j + \beta X)$ with a linear mean structure. From this assumed model (with estimated parameters), we simulate SBS residuals and compare this null distribution to its empirical distribution using QQ-plots. Figure 16 shows the results. Specifically, Figure 16(a) is the QQ plot when the true model has a quadratic mean structure (11), i.e., the fitted model is misspecified. Figure 16(b) is when both the true and fitted models have a linear mean structure

(12), i.e., the fitted mode is correct. Examining Figure 16(a) alone, we see that it is not obvious to conclude that the mean structure is misspecified. Furthermore, comparing Figure 16(a) with (b), we see that the two QQ-plots are quite similar. In other words, no matter the true model has a linear or quadratic mean structure, we end up with quite similar QQ-plots. This implies that the approach of using simulated SBS residuals as a reference is not able to tell the true form of the mean structure.