GENERALIZED SCORE TESTS FOR MISSING COVARIATE DATA

A Dissertation

by

LEI JIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Statistics

GENERALIZED SCORE TESTS FOR MISSING COVARIATE DATA

A Dissertation

by

LEI JIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Suojin Wang
Committee Members,	Michael T. Longnecker
	Daren B.H. Cline
	Jianhua Huang
	Jianxin Zhou
Head of Department,	Simon J. Sheather

August 2007

Major Subject: Statistics

ABSTRACT

Generalized Score Tests for Missing Covariate Data. (August 2007)
Lei Jin, B.S., Huazhong University of Science & Technology, P. R. China;
M.S., Zhejiang University, P. R. China
Chair of Advisory Committee: Dr. Suojin Wang

In this dissertation, the generalized score tests based on weighted estimating equations are proposed for missing covariate data. Their properties, including the effects of nuisance functions on the forms of the test statistics and efficiency of the tests, are investigated. Different versions of the test statistic are properly defined for various parametric and semiparametric settings. Their asymptotic distributions are also derived. It is shown that when models for the nuisance functions are correct, appropriate test statistics can be obtained via plugging the estimates of the nuisance functions into the appropriate test statistic for the case that the nuisance functions are known. Furthermore, the optimal test is obtained using the relative efficiency measure. As an application of the proposed tests, a formal model validation procedure is developed for generalized linear models in the presence of missing covariates. The asymptotic distribution of the data driven methods is provided. A simulation study in both linear and logistic regressions illustrates the applicability and the finite sample performance of the methodology. Our methods are also employed to analyze a coronary artery disease diagnostic dataset.

This work is dedicated to my parents and Rui.

ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to my advisor, Dr. Suojin Wang, for his direction, suggestion, encouragement, patience and continuous support toward my professional development. He guided me to think about statistical research problems, how to do research and how to write a paper step by step.

I would like to express my appreciation to all my committee members, Dr. Michael T. Longnecker, Dr. Daren B.H. Cline, Dr. Jianhua Huang and Dr. Jianxin Zhou. They gave me valuable comments and suggestions.

I also want to thank Dr. Samiran Sinha for his helpful comments and suggestions, the staff in the Department and all my friends who have given me help on various occasions.

Finally, I will express my appreciation to my family and Rui for their devoted love and support through my life.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation

Missing covariate data are very common in many applied areas, especially in the medical and social studies. Study designs are sometimes responsible for missing covariate data. For example, to obtain an optimal result in a fixed budget epidemiological study, researchers may employ a two stage study. Within the first stage, information on the response and some easily obtained variables is collected for all study subjects. During the second stage, information on other covariates is collected only for a subset of the study subjects depending on the observed attributes in stage one. The missingness can also be caused by happenstance. For example, respondents in a household survey may refuse to answer the questions regarding their income. In an industrial experimental process, some variables are not observed because of mechanical breakdowns.

A typical missing covariate data problem involves a response variable y, a vector of covariates ($\mathbf{x}$, $\mathbf{z}$) where the covariate $\mathbf{x}$ is not always observed, and a parametric model describing the relationship between y and $(\mathbf{x}, \mathbf{z})$. The parametric model may be specified by a conditional distribution $f(y|\mathbf{x}, \mathbf{z}; \boldsymbol{\beta})$ or a regression model

$$E(y|\mathbf{x}, \mathbf{z}) = g(\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}),$$

The format and style follow that of *Biometrics*.

where g is a known function and $\boldsymbol{\beta}$ is a p-dimensional parameter. Furthermore, an estimating equation condition

$$E\{\psi(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta})|\mathbf{x}, \mathbf{z}\} = 0$$

can be induced from the parametric model, where ψ is a p-dimensional estimating function. For example, ψ could be the score function

$$\psi = \frac{\partial \log f(y|\mathbf{x}, \mathbf{z}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}.$$

To indicate the missingness, we introduce an indicator random variable δ, which equals 1 if $\mathbf{x}$ is observed and 0 otherwise. According to Rubin (1976), the data are missing at random (MAR) if

$$\Pr(\delta = 1|y, \mathbf{x}, \mathbf{z}) = \Pr(\delta = 1|y, \mathbf{z}).$$

The data are missing completely at random (MCAR) if the missingness does not depend on the data values. Data analysis generally includes parameter estimation, hypothesis tests and the corresponding model validation. The focus of this dissertation centers on issues of testing composite hypotheses of $\boldsymbol{\beta}$ and formal model validation for the missing covariate data under the assumption of MAR.

Because standard techniques for statistical inferences usually require full covariate information, one simple way of handling missing covariate data is a complete-case analysis, which excludes observations with missing values and performs naive statistical analysis. Despite its convenience with existing statistical packages, the complete-case analysis discards information from the incomplete cases and may result in substantial efficiency loss. More importantly, it ignores the possible systematic difference between the complete cases and incomplete cases, and thus yields misleading results. Approaches for correctly analyzing missing covariate data may include likelihood

based approaches (Rubin, 1976; Little and Rubin, 2002; Ibrahim, Chen, and Lipsitz, 1999), multiple imputation (Rubin, 1996; Schafer, 1997; Little and Rubin, 2002) and weighted estimating equation methods (Flanders and Greenland, 1991; Zhao and Lipsitz, 1992; Robins, Rotnitzky, and Zhao, 1994; Wang, Wang, Zhao, and Ou, 1997; Lipsitz, Ibrahim, and Zhao, 1999). Compared with the other approaches, weighted estimating equation methods can provide consistent results under more flexible assumptions (Lipsitz et al., 1999; the discussion rejoinder in Scharfstein, Rotnitzky, and Robins, 1999; Van der Laan and Robins, 2003, Ibrahim, Chen, Lipsitz, and Herring, 2005).

Considering n independent observations, Robins et al. (1994) proposed the general weighted estimating equations (WEEs)

$$\begin{aligned} U(\boldsymbol{\beta}, \pi, \phi) &= \sum_{i=1}^{n} \mathbf{u}_i(\boldsymbol{\beta}, \pi_i, \phi) \\ &= \sum_{i=1}^{n} \{ \frac{\delta_i}{\pi_i} \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) + (1 - \frac{\delta_i}{\pi_i}) \phi(y_i, \mathbf{z}_i) \} = \mathbf{0}, \end{aligned} \tag{1.1}$$

where

$$\pi_i = \pi(y_i, \mathbf{z}_i) = \Pr(\delta_i = 1 | y_i, \mathbf{z}_i),$$

and ϕ is an arbitrary fixed $p \times 1$ function with finite second moments. For statistical inference in the WEE (1.1) setting, sandwich covariance estimates and Walds-type tests are widely used in the literature. As an alternative to Wald-type tests, generalized score tests (Rotnitzky and Jewell, 1990; Boos, 1992; Commenges and Jacqmin-Gadda, 1997; Thas and Rayner, 2005) are widely used to test a variety of hypotheses in a simple and unified way in estimating equation settings. Because of the invariance properties, the score type statistic and likelihood ratio statistic are often preferred to Wald statistics in standard parametric models (see Boos, 1992). One primary concern of this dissertation is the generalized score tests for testing composite hypotheses in

Table 1: Description of Duke Cardiac Catheterization Coronary Artery Disease Diagnostic Dataset; 3504 observations and 6 variables, maximum number of missing values (denoted by NAs):1246.

Name	Labels	Units	NAs
sex	Male = 1, Female = 0		0
age	Age	Year	0
cad.dur	Duration of Symptoms of Coronary Artery Disease		0
choleste	Cholesterol	mg%	1246
sigdz	Significant Coronary Disease by Cardiac Cath		0
tvdlm	Three Vessel or Left Main Disease by Cardiac Cath		3

the presence of missing covariates based under the WEE (1.1) setting. More specifically, we study the effects of nuisance functions π and ϕ and their estimates on the generalized score statistics, the efficiency issues and applications of the tests.

The following example motivates our study. It is of interest to use correct tools to analyze the Duke Cardiac Catheterization Coronary Artery Disease Diagnostic Dataset (Harrell, 2001, Chapter 10). The structure of the dataset is described in Table 1. The variable cholesterol is not observed among 1246 out of 3504 observations. Extensive complete-case analysis by Harrell (2001) included parameter estimation, tests of regression coefficients and the corresponding model validation. Since around one-third of the observations are incomplete, complete-case analysis might be invalid or inefficient. While it would be natural to recheck the validation of the logistic regression previously used for this dataset, to the best of our knowledge there are no such formal methods in the current literature that deal with this issue.

An assessment of model fit is an important part of any modeling procedure. In general, it evaluates how well the predicted outcomes coincide with the observed data. Model evaluation for missing data may include the detection of an incorrect assumption of missing-data mechanism, omitted important covariates, or inappropriate distributional assumptions. There are a few tests in the literature concerning a

model for the selection probability or missing-data mechanism. Lei and Wang (2001) developed two test statistics that focus on the validation of the MAR assumption. Lipsitz, Parzen, Molenberghs, and Ibrahim (2001) proposed a test for bias in WEEs caused by the missingness that is incorrectly modeled. For testing the adequacy of the primary regression function, González-Manteiga and Pérez-González (2006) proposed goodness-of-fit tests for linear models with missing responses under the MAR assumption. In the theoretical framework of Bayesian posterior predictive checks, Gelman, Van Mechelen, Verbeke, Heitjan, and Meulders (2005) proposed an informal missing data model checking method using graphical diagnostics. As is mentioned for a future research topic in the review paper by Ibrahim et al. (2005), it would be interesting to explore formal model validation methods in the presence of missing covariates. As an application the proposed generalized score tests, we develop a formal model validation procedure for generalized linear models in the presence of missing covariates.

1.2 Dissertation Structure

The dissertation is organized as follows. In Chapter II, a comprehensive review for missing data, weighted estimating equations, generalized score tests, etc. will be discussed. In Chapter III, we investigate the generalized score tests under the weighted estimating equation settings. In Chapter IV, we develop goodness-of-fit tests for generalized linear models when some covariates are partially missing. A simulation study and its results are presented in Chapter V. In Chapter VI, as an illustration we reanalyze the dataset discussed above. Some concluding remarks and comments on future research given in Chapter VII.

CHAPTER II

LITERATURE REVIEW

A main concern of this dissertation is the generalized score tests and their applications in the presence of missing covariates based on the weighted estimating equations. We review the missing-data mechanism, missing-data pattern, weighted estimating equation methods, generalized score tests, model validation methods and other relevant topics in this chapter.

2.1 Missing-data Mechanism and Pattern

Missing-data mechanism describes the relationship between the missingness and the values of variables. It is crucial because the properties of missing-data methods strongly depend on this mechanism. The data are missing completely at random if the missingness does not depend on the data values. Assumption of MCAR basically implies that the complete cases are a random subsample of the intended sample, and thus a complete-case analysis is valid. The data are missing at random if, conditional on the observed data, the missingness does not depend on the unobserved data. Clearly, MAR is a weaker assumption than MCAR. In this case of MAR, complete-case methods may not be valid because the complete cases are no longer a random sample of the intended sample. If the data are MAR and the missingness does not depend the response, then a complete-case analysis will lead to valid results. When neither MCAR nor MAR holds, we say the data are missing not at random (MNAR). In the likelihood setting, the missing-data mechanism MNAR is termed non-ignorable. Valid inferences generally require specifying the correct model for the missing-data

mechanism when the missingness is non-ignorable. The assumption of MAR has been widely used in the literature, as in this dissertation. Our methodology can be directly applied to MNAR cases if correct models for the selection probability are available, which generally requires additional information.

Missing-data pattern is another important concept regarding missing data, especially when there are multiple variables with missing values. It describes which values are observed and which values are missing in the data matrix. If the data matrix can be rearranged in such a way that there is a hierarchy of missingness, so that observing a particular variable for a subject implies that all other variables on the left-side of this variable are observed, then the missingness is said to be monotone. Little and Rubin (2002, Chapter I) described various missing-data patterns, including univariate nonresponse, multivariate two patterns, general missingness pattern, etc. Some methods for missing data are restricted to certain special patterns. In this dissertation, we assume the missing-data pattern is multivariate two patterns, where $\mathbf{x}$ is all missing if $\delta = 0$. The methodology can be applied to monotone missingness without difficulty because WEE (1.1) works for monotone missingness pattern.

2.2 Weighted Estimating Equations

2.2.1 Weighted Estimating Equations

Flanders and Greenland (1991), and Zhao and Lipsitz (1992) proposed an estimator based on the simple inverse probability weighted estimating equations

$$U_s(\boldsymbol{\beta}, \pi) = \sum_{i=1}^{n} \mathbf{u}_{si}(\boldsymbol{\beta}, \pi_i) = \sum_{i=1}^{n} \{\frac{\delta_i}{\pi_i} \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})\} = \mathbf{0} \tag{2.1}$$

for two-stage studies. Assuming that π is known, they showed that the estimator is consistent for $\boldsymbol{\beta}$ and asymptotically normal distributed. However, it is clear that the estimator is not efficient because Equation (2.1) has nothing to do with the incomplete

cases.

Robins et al. (1994) introduced weighted estimating equations (1.1) and defined a class of estimators indexed by ϕ under regularity conditions. For each ϕ, the estimator is the unique solution $\hat{\boldsymbol{\beta}}$ of Equation (1.1). Note that the solution, the equation and other relevant quantities depend on the nuisance functions π and ϕ. For notational convenience, we suppress this dependence throughout this dissertation when there is no confusion. The methods are quite general and can be applied to very large classes of models, including generalized linear models, proportional hazards model and nonlinear models. Robins et al. (1994) showed that $\hat{\boldsymbol{\beta}}$ are asymptotically normal and unbiased for $\boldsymbol{\beta}$ when (a) the data are MAR, (b) π is bounded away from 0, and (c) π is either known (in a designed study) or estimated via a correct model. The asymptotic variance of $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ can be consistently estimated by the corresponding sandwich estimator. It is clear that Equation (2.1) is a special case of WEE (1.1). They pointed out that estimators previously proposed by Horvitz and Thompson (1952), Breslow and Cain (1988), Flanders and Greenland (1991), and Zhao and Lipsitz (1992) are asymptotically equivalent to some inefficient estimators in their class. Misspecification of π could lead to a biased estimating equation (1.1), while the choice of ϕ affects the efficiency of the point estimators. The asymptotic variance of $\hat{\boldsymbol{\beta}}$ and WEE (1.1) is uniquely minimized in the positive definite sense when

$$\phi = \phi^*(y_i, \mathbf{z}_i) = E\{\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})|y_i, \mathbf{z}_i\}.$$

In fact, $\phi^*(y_i, \mathbf{z}_i)$ is the conditional mean score function. A sketch proof for this optimum property is given below.

Note that Equation (1.1) can be rewritten as

$$\begin{aligned} U(\boldsymbol{\beta}, \pi, \phi) &= \sum_{i=1}^{n} \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) + \sum_{i=1}^{n} (\frac{\delta_i}{\pi_i} - 1)\{\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) - \phi(y_i, \mathbf{z}_i)\} \\ &= U^F(\boldsymbol{\beta}) + U^M(\boldsymbol{\beta}, \pi, \phi). \end{aligned} \tag{2.2}$$

For all $\boldsymbol{\beta}$, $(\frac{\delta_i}{\pi_i} - 1)\{\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) - \phi(y_i, \mathbf{z}_i)\}$ has mean 0 given $(y_i, \mathbf{z}_i)$ and thus is uncorrelated with $\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})$. Hence

$$\begin{aligned} \mathrm{Var}\{U(\boldsymbol{\beta}, \pi, \phi)\} &= \mathrm{Var}\{U^F(\boldsymbol{\beta})\} + \mathrm{Var}\{U^M(\boldsymbol{\beta}, \pi, \phi)\} \\ &= \mathrm{Var}\{U^F(\boldsymbol{\beta})\} + E[\frac{(1-\pi)}{\pi} E\{[\psi_i - \phi(y_i, \mathbf{z}_i)][\psi_i - \phi(y_i, \mathbf{z}_i)]'\}], \end{aligned}$$

where $\phi_i = \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})$. It is clear that $E[\frac{(1-\pi)}{\pi} E\{[\psi_i - \phi(y_i, \mathbf{z}_i)][\psi_i - \phi(y_i, \mathbf{z}_i)]'\}]$ is minimized at ϕ^* in the positive definite sense.

According to Equation (2.2), $U(\boldsymbol{\beta}, \pi, \phi)$ can be decomposed to full data estimating estimating equations $U^F(\boldsymbol{\beta})$ and the noise $U^M(\boldsymbol{\beta}, \pi, \phi)$. Because $U^M(\boldsymbol{\beta}, \pi, \phi)$ has mean 0 and is uncorrelated with $U^F(\boldsymbol{\beta})$, $U^M(\boldsymbol{\beta}, \pi, \phi)$ is just random noise added to the full data estimating estimating equations due to the missingness and cannot help in estimation of $\boldsymbol{\beta}$. The variance of $U^M(\boldsymbol{\beta}, \pi, \phi)$ is a quantitative measure of the noise for not having observed all data. The penalty paid for missing data is minimized when the mean score function ϕ^* is used for extracting the information in the incomplete cases.

The estimator based on WEE (2.1) is biased if the selection probability is not appropriate while the estimator based on WEE (1.1) may not. Scharfstein et al. (1999) discussed doubly robust estimators based on general WEE (1.1). An estimator is doubly robust in the sense that it is consistent for $\boldsymbol{\beta}$ if either the model for the selection probability or the model for the conditional mean score function is correctly specified. For example, the estimators in Lipsitz et al. (1999) and Rotnitzky, Robins, and Scharfstein (1998) are doubly robust.

2.2.2 Parametric Setting

In WEE (1.1), the nuisance function π is often unknown and thus needs to be estimated. On the other hand, the efficient estimator $\boldsymbol{\beta}(\pi, \phi^*)$ is not feasible because

the conditional mean score function depends on the unknown conditional distribution $(\mathbf{x}|y, \mathbf{z})$.

Zhao, Lipsitz, and Lew (1996) introduced a joint estimating equation for regression analysis when some covariates are missing. They posed a logistic regression for the selection probability,

$$\pi_i = \pi_i(\boldsymbol{\alpha}) = \frac{\exp(-\boldsymbol{\alpha}'\mathbf{v}_i)}{1 + \exp(-\boldsymbol{\alpha}'\mathbf{v}_i)}, \tag{2.3}$$

where and $\mathbf{v}_i$ is a vector function of $(y_i, \mathbf{z}_i)$'s, such as $(y_i, \mathbf{z}_i')'$ and $(y_i^{\frac{1}{3}})$. The maximum likelihood equation for the logistic regression is

$$U_\pi(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{v}_i \{\delta_i - \pi_i(\boldsymbol{\alpha})\} = \mathbf{0}.$$

Assuming that $\boldsymbol{\kappa}$ is a necessary vector of unknown parameter in the model for ϕ^*, their joint estimating equation is

$$0 = \begin{pmatrix} U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ U_\pi(\boldsymbol{\alpha}) \end{pmatrix}, \tag{2.4}$$

where $U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ depends on assumptions regarding the conditional moments $E(\mathbf{x}_i|y_i, \mathbf{z}_i)$, $E(\mathbf{x}_i^2|y_i, \mathbf{z}_i)$, etc. The parameter $\boldsymbol{\kappa}$ in their setting is related to these conditional moments.

Lipsitz et al. (1999) proposed another joint estimating equations similar to maximum likelihood equations for missing covariate data. By assuming the conditional distribution $p(\mathbf{x}|\mathbf{z}; \boldsymbol{\kappa})$, they obtained a joint estimating equation similar to Equation (2.4) where

$$U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \sum_{i=1}^{n} [\frac{\delta_i}{\pi_i} \psi_{\phi^*}(\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\kappa}) + (1 - \frac{\delta_i}{\pi_i}) E\{\psi_{\phi^*}(\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\kappa})|y_i, \mathbf{z}_i\}]$$

and

$$\psi_{\phi^*}(\mathbf{x}, \mathbf{z}, \boldsymbol{\kappa}) = \{\partial \log p(\mathbf{x}|\mathbf{z}, \boldsymbol{\kappa})/\partial \boldsymbol{\kappa}\}'.$$

An EM-type algorithm was proposed to solve the joint estimating equation above. The estimate $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ is the solution of $U_\pi = 0$, while the estimates $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\kappa}}$ for $\boldsymbol{\kappa}$ can be obtained by EM-type iterative methods using U, U_{ϕ^*} and $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$. According to Lipsitz et al. (1999), $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ when at least one of the following is correctly specified: (a) model (2.3) for the selection probability or (b) the distributional assumptions on $f(y|\mathbf{x}, \mathbf{z}; \boldsymbol{\beta})$ and $p(\mathbf{x}|\mathbf{z}; \boldsymbol{\kappa})$. When π is correctly specified, they obtained

$$E\{\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\} = 0.$$

On the other hand, when $f(y|\mathbf{x}, \mathbf{z}; \boldsymbol{\beta})$ and $p(\mathbf{x}|\mathbf{z}; \boldsymbol{\kappa})$ are correctly specified, they showed

$$E\{\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})}{\partial \boldsymbol{\alpha}}\} = 0.$$

As a special case of Lipsitz et al. (1999), Parzen, Lipsitz, Ibrahim, and Lipshultz (2002) considered a weighted estimating equation for linear regression with missing covariate data. They proposed weighted estimating equations with the assumption that the missing covariates are multivariate normal, which might be incorrect. Via simulation, they compared their WEEs with the semiparametric efficient WEE with correct distribution assumption on the missing covariates as well as the maximum likelihood methods. They concluded that the methods work for many situations and the efficiency is high.

2.2.3 Semiparametric Setting

It is generally convenient to assume parametric models for the selection probability and the mean score function ϕ^*. However, it might be problematic when the parametric models (especially the model for the selection probability) are not correct.

To deal with this problem, Wang et al. (1997) proposed a semiparametric estimate of $\boldsymbol{\beta}$ in regression analysis with missing covariates. They considered the weighted estimating equation

$$U_s(\boldsymbol{\beta}, \hat{\pi}_N) = \sum_{i=1}^{n} \{ \frac{\delta_i}{\hat{\pi}_{Ni}} \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) \} = \mathbf{0},$$

where

$$\hat{\pi}_N(\mathbf{v}) = \frac{\sum_{i=1}^{n} \delta_i K_h(\mathbf{v} - \mathbf{v}_i)}{\sum_{i=1}^{n} K_h(\mathbf{v} - \mathbf{v}_i)}, \tag{2.5}$$

K is an sth-order kernel function, h is a proper bandwidth parameter, $K_h(\cdot) = K(\cdot/h)$, and $\mathbf{v}_i = (y_i, \mathbf{z}_i')'$. They concluded that (a) the semiparametric estimator $\hat{\boldsymbol{\beta}}$ is root-n consistent, though the nonparametric smoother $\hat{\pi}_N$ has slower rate than root-n consistency, and (b) the efficiency of estimating $\boldsymbol{\beta}$ may be gained via estimating the selection probability.

Wang and Wang (2001) investigated kernel assisted estimators in regression analysis in the presence of missing covariates. Smoothing techniques are employed in estimating π and ϕ^*. They proposed three kernel assisted semiparametric estimators and founded the asymptotic equivalence between these estimators. More specifically, the selection probability is estimated via (2.5) and the conditional mean score function is estimated by

$$\hat{\phi}_N^*(\mathbf{v}) = \frac{\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_N(\mathbf{v}_i)} \hat{\psi}_i K_h(\mathbf{v} - \mathbf{v}_i)}{\sum_{i=1}^{n} K_h(\mathbf{v} - \mathbf{v}_i)}, \tag{2.6}$$

where $\hat{\psi}_i = \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\beta}}$ is the solution of $U(\boldsymbol{\beta}, \hat{\pi}_N, \phi) = 0$.

Under the regularity conditions $(C1)$–$(C5)$, they showed that

$$\begin{aligned} n^{-\frac{1}{2}} U(\boldsymbol{\beta}, \hat{\pi}_N, 0) &= n^{-\frac{1}{2}} U(\boldsymbol{\beta}, \pi, \phi^*) + O_p(\rho_n), \\ n^{-\frac{1}{2}} U_A(\boldsymbol{\beta}, \hat{\pi}_N, \hat{\phi}_N^*) &= O_p(\rho_n), \end{aligned} \tag{2.7}$$

where $U_A(\boldsymbol{\beta}, \pi, \phi) = \sum_{i=1}^{n} (1 - \frac{\delta_i}{\pi_i}) \phi(y_i, \mathbf{z}_i)$, and $\rho_n = \left\{ nh^{2s} + (nh^{2d})^{-1} \right\}^{\frac{1}{2}}$.

Liang, Wang, Robins, and Carroll (2004) used different nonparametric estimates for the selection probability in their local weighted estimating equations. They pointed out that the selection probability can be estimated via many nonparametric estimators (local polynomial, kernel methods with varying bandwidths, smoothing and regression splines, and so on) and the results are asymptotically equivalent under certain conditions similar to $(C1)-(C5)$.

2.2.4 Comparison with Other Methods

When the likelihood for the complete-data

$$p(y, \mathbf{x}|\mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\kappa}) = f(y|\mathbf{x}, \mathbf{z}; \boldsymbol{\beta})p(\mathbf{x}|\mathbf{z}; \boldsymbol{\kappa})$$

is available, the likelihood based methods (Rubin, 1976; Little and Rubin, 2002; Ibrahim et al., 1999) can be used for inference via the EM algorithm (Dempster et al., 1977). If the data are MAR, the likelihood based methods ignoring the missing-data mechanism is valid (Rubin, 1976). However, an appropriate likelihood is often difficult to obtain for missing data problems and the results are not robust to model misspecification.

Multiple imputation is another popular approach for handling missing covariate data. First, it creates multiple 'complete' datasets by making random draws from the predictive distribution $p(\mathbf{x}|y, \mathbf{z})$ of the missing values, which require essentially the same condition as likelihood based methods. Often multivariate normal models are used for covariates ($\mathbf{x}$,$\mathbf{z}$) because it is computationally tractable. Second, each of these 'complete' datasets are analyzed using standard methods. Finally, the results are combined which take uncertainty regarding the imputation into account. Multiple imputation is an attractive choice for missing data problems because of ease of use. For example, multiple imputation in SAS can be carried out in three simple

steps. First, the imputation is carried out by PROC MI. Next, standard methods are employed for complete-case analysis. Finally, the results are combined using PROC MIANALYZE. However, the predictive distribution $p(\mathbf{x}|y,\mathbf{z})$ must be proper to have consistent estimators and valid tests.

WEEs methods for missing covariate data without making strict parametric assumptions on the distribution of covariates. Without an appropriate assumption on $p(\mathbf{x}|\mathbf{z};\boldsymbol{\kappa})$, the result is still consistent if the selection probability is correctly specified or estimated. When $p(\mathbf{x}|\mathbf{z};\boldsymbol{\kappa})$ or other similar assumptions are correctly specified, the estimate is valid and efficient. Doubly robust estimators based on WEE (1.1) are preferred to MLE in many cases. If the missingness is non-ignorable, the maximum likelihood estimate will generally be inconsistent unless both the model for the selection probability and the model for the conditional mean score function are correctly specified. In two stage designs or samples surveys with a known selection probability, the doubly robust estimator is guaranteed to be consistent. In contrast, the maximum likelihood estimator may be inconsistent if the parametric model for all covariates is misspecified. However, for the general missing covariate data involving both continuous covariates and general missing pattern, the doubly robust estimators are difficult to obtain.

2.3 Generalized Score Tests

A comprehensive introduction for generalized score tests may be found in Boos (1992), which discussed the use of score tests in the general estimating equation setting for fully observed data. In the case of no missingness, WEE (1.1) reduces to $U^F(\boldsymbol{\beta})$, which is given in Equation (2.2) and free of π and ϕ. Typically, the generalized score tests are for testing

$$H_0\text{: } \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} \text{ vs } H_a\text{: } \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}, \tag{2.8}$$

where $\boldsymbol{\beta}_2$ is an $r \times 1$ sub-vector of $\boldsymbol{\beta}$ in Equation (1.1) such that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. All relevant vectors are partitioned accordingly, e.g., $U = (U_1', U_2')'$ where U_1 is $(p-r) \times 1$ and U_2 is $r \times 1$. Boos showed how generalized score statistics arise from Taylor expansion of the estimating equation. Let $\tilde{\boldsymbol{\beta}}$ be the solution of Equation (1.1) under H_0. By expanding $U_1(\tilde{\boldsymbol{\beta}})$ and $U_2(\tilde{\boldsymbol{\beta}})$ at the true value $\boldsymbol{\beta}$, and replace $\frac{\partial U_1}{\partial \boldsymbol{\beta}_1}$ and $\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}$ by their asymptotically equivalent versions $E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})$ and $E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})$, $U(\tilde{\boldsymbol{\beta}})$ can be written as

$$0 = U_1(\tilde{\boldsymbol{\beta}}) = U_1(\boldsymbol{\beta}) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + R_{n1},$$
$$U_2(\tilde{\boldsymbol{\beta}}) = U_2(\boldsymbol{\beta}) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + R_{n2},$$

where the remainders R_{n1}, R_{n2} are relatively negligibly small with order $O_p(1)$ under H_0 and the partial derivatives above are row vectors. By combining the two equations above, we have

$$U_2(\tilde{\boldsymbol{\beta}}) = (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}) + R_{n3}, \tag{2.9}$$

where $\mathbf{A} = E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}$, R_{n3} is a negligible remainder with order $O_p(1)$, and $\mathbf{I}_r$ is the $r \times r$ identity matrix. Therefore, Boos obtained one version of generalized score statistic

$$U_2(\tilde{\boldsymbol{\beta}})' \left\{(-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_U(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right\}^{-1} U_2(\tilde{\boldsymbol{\beta}}), \tag{2.10}$$

where $\tilde{\mathbf{A}} = \mathbf{A}|_{\tilde{\boldsymbol{\beta}}}$, $\boldsymbol{\mathcal{J}}_U = \sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\beta})\mathbf{u}_i'(\boldsymbol{\beta})$ and $\tilde{\boldsymbol{\mathcal{J}}}_U = \boldsymbol{\mathcal{J}}_U|_{\tilde{\boldsymbol{\beta}}}$. The test statistic follows χ^2_r asymptotically under H_0 and regularity conditions.

The efficiency of generalized score tests is another important issue. Tosteson and Tsiatis (1988) studied three score tests and their relative efficiency in a generalized linear model with surrogate covariates. We may follow Tosteson and Tsiatis (1988) to study efficiency issue of generalized score tests in the missing data setting.

2.4 Model Validation Procedures for Missing Data

In general, an assessment of model fit is an important part of any modeling procedure. Model evaluation for missing data may include the detection of the an incorrect assumption of missing-data mechanism, omitted important covariates, or inappropriate distributional assumptions.

There are a few tests in the literature concerning the assumption of missing at completely random. Chen and Little (1999) proposed a Wald-type test for missing at completely at random in generalized estimating equations with incomplete data. Strictly speaking, the proposed test statistic tests whether or not the data and the missing-data pattern are independent, which does not imply assumption of missing completely at random exactly. The test statistic follows a χ^2 distribution under H_0. They suggested that the an unadjusted generalized estimating equation is appropriate when H_0 is accepted. They employed an information decomposition and recombination procedure to construct the Walt-type test statistic.

Qu and Song (2002) proposed a generalized score-type test based on the quadratic inference for testing whether or not missing data in longitudinal data analysis are ignorable with regard to quasi likelihood or estimating equations approaches. In other words, they try to test unbiasedness of unadjusted estimating equations, which is almost the same null hypothesis as Chen and Little (1999). They used estimating equations U_Q based on $(p+r)\times 1$ dimensional unbiased function $s(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta})$. To construct the test statistic, they first separated the complete cases and incomplete cases into two groups. For each group of the data, one of the estimating equations U_{Q1} and U_{Q2}, such that $U_Q = (U'_{Q1}, U'_{Q2})'$, can be constructed. They defined the quadratic inference function as

$$Q = U'_{Q1}\mathbf{V}_{Q1}^{-1}U_{Q1} + U'_{Q2}\mathbf{V}_{Q2}^{-1}U_{Q2},$$

where $\mathbf{V}_{Q1}$ and $\mathbf{V}_{Q1}$ are consistent covariance estimate for U_{Q1} and U_{Q2}, respectively. The test statistic is $Q(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the minimizer of the quadratic inference function Q. The statistic $Q(\hat{\boldsymbol{\beta}})$ follows χ^2_r asymptotically. Furthermore, they also showed that the generalized score test is asymptotically equivalent to Chen and Little's Wald-type test.

Lei and Wang (2001) developed test statistics for bias of WEE (2.1) in the presence of missing covariates. Under the assumption that the primary regression model is correct, the test statistics focus on testing whether or not the data are MAR. The test statistics were developed based on partitioning the sample into disjoint q groups. For $k = 1, \cdots, q$, define

$$T_k = U_k(\hat{\boldsymbol{\beta}}, \hat{\pi})' \Sigma_k^{-1} U_k(\hat{\boldsymbol{\beta}}, \hat{\pi}),$$

where $\hat{\pi}$ is the estimated selection probability, U_k is WEE (2.1) which uses the kth group of data only, and Σ_k is a consistent estimator for the asymptotic covariance matrix of $U_k(\hat{\boldsymbol{\beta}}, \hat{\pi})$. The test statistic is

$$T = max_{1 \leq k \leq q}(T_1, \cdots, T_q).$$

In both parametric and semiparametric setting, they showed that the test statistics follow an asymptotic χ^2_p distribution when $q = 2$. Both the parametric and semiparametric tests performed well and similarly when (a) sample size is large enough and (b) the selection probability is correctly specified. When the parametric model for the selection probability is not correct, they suggested that the semiparametric test should be used.

Lipsitz et al. (2001) proposed a test for bias in WEE (2.1) caused by the missingness of the data that is not modeled correctly. More strictly, the null hypothesis is WEE and the full data estimates converge in probability to the same parameter. To

obtain the test statistic, the regression model of y given $\mathbf{z}$ was fitted using complete cases via WEE (2.1) as well as all data, and thus obtain two estimates of $\boldsymbol{\beta}_z$, say $\hat{\boldsymbol{\beta}}_{z,WEE}$ and $\hat{\boldsymbol{\beta}}_z$ respectively, where $\boldsymbol{\beta}_z$ the regression coefficient corresponding to $\mathbf{z}$. The test statistic is

$$(\hat{\boldsymbol{\beta}}_{z,WEE} - \hat{\boldsymbol{\beta}}_z)'[\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{z,WEE} - \hat{\boldsymbol{\beta}}_z)]^{-1}(\hat{\boldsymbol{\beta}}_{z,WEE} - \hat{\boldsymbol{\beta}}_z),$$

which is an approximate $\chi^2_{p_z}$ under H_0, where p_z is the dimension of covariate $\mathbf{z}$.

Regrading the primary regression function, González-Manteiga and Pérez-González (2006) proposed goodness-of-fit tests for a linear regression model with missing response only under the MAR assumption. The proposed test statistics are based on the L_2 distance between appropriate nonparametric and parametric estimates of the regression function under H_0. Because the convergence rate of the test statistics to the asymptotic distribution is slow, they proposed a bootstrap procedure for approximation of the critical values. Under MAR and no covariates are missing, there is no systematic difference between complete cases and incomplete cases. Therefore, the complete case analysis is valid.

2.5 Regularity Conditions

The regularity conditions given below are based on Wang et al. (1997) and Wang and Wang (2001) for kernel assisted estimators.

(C1) The function $\pi(\mathbf{v})$ is bounded away from 0 for all $\mathbf{v}$ in its domain.

(C2) The function $\pi(\mathbf{v})$ has s continuous and bounded partial derivatives with respect to the continuous components of $\mathbf{v}$.

(C3) The probability density function $p(\mathbf{v})$ and the conditional probability density function $p(\mathbf{v}|\delta)$ both have s continuous and bounded partial derivatives with

respect to the continuous components of $\mathbf{v}$.

(C4) The conditional mean score function $E\{\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})|y,\mathbf{z}\}$ and $E\{\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\psi'(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})|y,\mathbf{z}\}$ exist and have s continuous and bounded partial derivatives with respect to the continuous components of $\mathbf{v}$.

(C5) $E\{\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\psi'(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\}$ and $E\{\frac{\partial}{\partial\boldsymbol{\beta}}\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\}$ exits and are positive definite, and $\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})$ exists and is continuous in the parameter space.

The regularity conditions given below are for generalized score statistics.

(C6) The first and second moments of $\frac{\partial\psi}{\partial\boldsymbol{\beta}}$ exist and $\frac{\partial\psi}{\partial\boldsymbol{\beta}}$ is continuous in a neighborhood of the true value of $\boldsymbol{\beta}$.

(C7) $E(\frac{\partial U_1}{\partial\boldsymbol{\beta}_1})^{-1}$ exists, and $E(\frac{\partial U_2}{\partial\boldsymbol{\beta}_1})^{-1}$ has full row rank.

(C8) Estimating equation U is unbiased and has an unique solution, $E\{\psi(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\psi'(y,\mathbf{x},\mathbf{z},\boldsymbol{\beta})\}$ exists and is positive definite.

Note that $\frac{p(\mathbf{v}|\delta=0)}{p(\mathbf{v}|\delta=1)}$ is bounded in the domain, since $\pi(\mathbf{v})$ is bounded away from 0. Condition (C6) guarantees that $\frac{\partial U}{\partial\boldsymbol{\beta}} = E(\frac{\partial U}{\partial\boldsymbol{\beta}}) + O_p(n^{\frac{1}{2}})$. This means that $\frac{\partial U}{\partial\boldsymbol{\beta}}$ and $E(\frac{\partial U}{\partial\boldsymbol{\beta}})$ are asymptotically equivalent under condition (C6). Also note that, in parametric settings, the estimating equation U in (C6)-(C8) is matter for the whole corresponding joint estimating equation (e.g. U_J in 3.6) and $\boldsymbol{\beta}$ is for all parameters (e.g. $\boldsymbol{\tau}$ in 3.6) in the joint estimating equation.

CHAPTER III

GENERALIZED SCORE TESTS FOR MISSING COVARIATE DATA

3.1 Introduction

Generalized score methods provide a simple and unified way to test a variety of hypotheses in many statistical problems. For example, Rotnitzky and Jewell (1990) developed a generalized score test for regression coefficients in semiparametric generalized linear models for cluster correlated data. Boos (1992) discussed generalized score tests in a general estimating equation setting. Commenges and Jacqmin-Gadda (1997) derived a general form of the score statistic for the random effect in correlated random effects model. Thas and Rayner (2005) constructed a goodness-of-fit test using generalized score statistics to test for the zero-inflated Poisson distribution against general smooth alternatives. In this chapter, we study the generalized score tests for missing covariate data based on WEE (1.1).

WEE methods have been widely used for missing covariate data without making strict parametric assumptions. The estimator based on WEE (1.1)

$$\begin{aligned} \mathbf{0} = U(\boldsymbol{\beta}, \pi, \phi) &= \sum_{i=1}^{n} \mathbf{u}_i(\boldsymbol{\beta}, \pi_i, \phi) \\ &= \sum_{i=1}^{n} \{ \frac{\delta_i}{\pi_i} \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) + (1 - \frac{\delta_i}{\pi_i}) \phi(y_i, \mathbf{z}_i) \}, \end{aligned}$$

can be doubly robust. In the WEE (1.1) setting, Wald-type tests and sandwich covariance estimates are widely used in the literature. According to Boos (1992), the score type statistics are attractive because (a) they only require computation of the null estimates and (b) they could be invariant to nonlinear transformations of

the parameters whereas Wald statistics are not. More specifically, we are primarily concerned with generalized score tests for testing hypothesis (2.8)

$$H_0\text{: } \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} \text{ vs } H_a\text{: } \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$$

based on WEE (1.1) in different settings.

3.2 The Case of the Selection Probability π Being Known and ϕ Being Given

Zhao and Lipsitz (1992) concerned statistical inference of two-stage studies using WEE (2.1), which collects the data in two stages. In the first stage, the covariate $\mathbf{z}$ of n subjects are observed, and at the second stage covariate $\mathbf{x}$ is measured on a subset of the study subjects based on the design selection plan. Then it is reasonable to assume the selection probability is known for many applications. Recall that ϕ in WEE (1.1) could be an arbitrary fixed $p \times 1$ function with finite second moments. Therefore, based on WEE (1.1), a class of generalized score tests indexed by ϕ can be defined. In this section, we would like to investigate how the nuisance function ϕ affects the generalized score tests when the selection probability is known.

3.2.1 A Class of Generalized Score Tests

Since π is known, $U(\boldsymbol{\beta}, \pi, \phi)$ in WEE (1.1) reduces to $U(\boldsymbol{\beta}, \phi)$. Under the current setting and regularity conditions, an unique solution $\tilde{\boldsymbol{\beta}}$ of Equation (1.1) can be solved under H_0. Recall that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, $\boldsymbol{\beta}_1$ is $(p-r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$. In addition, we assume that $\boldsymbol{\beta}_1$ can be solved using U_1 given $\boldsymbol{\beta}_2$. Following the approach in Boos (1992) and by regularity condition (C6), we obtain

$$0 = U_1(\tilde{\boldsymbol{\beta}}, \phi) = U_1(\boldsymbol{\beta}, \phi) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\beta}}, \phi) = U_2(\boldsymbol{\beta}, \phi) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1),$$

under H_0, where E denotes the expectation with respect to $(\delta_i,\ y_i,\ \mathbf{x}_i,\ \mathbf{z}_i)$. In the case that $O_p(1)$ is a matrix or vector, $O_p(1)$ means that each element of the matrix or vector is of the order $O_p(1)$. Combining the two equations above, we have

$$U_2(\tilde{\boldsymbol{\beta}}, \phi) = (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \phi) + O_p(1), \tag{3.1}$$

where $\mathbf{A} = E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}$, which has the same form $\mathbf{A}$ in (2.9) while the meaning of E and the equation U are different. Let $\tilde{\mathbf{A}} = \mathbf{A}|_{\tilde{\boldsymbol{\beta}}}$, $\boldsymbol{\mathcal{J}}_U = \sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\beta}, \pi, \phi)\mathbf{u}_i'(\boldsymbol{\beta}, \pi, \phi)$ and $\tilde{\boldsymbol{\mathcal{J}}}_U = \boldsymbol{\mathcal{J}}_U|_{\tilde{\boldsymbol{\beta}}}$. Without confusion, we will continue to use $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\mathcal{J}}}_U$, etc. for the quantities evaluated at proper parameters or their estimates under other settings. By the root-n consistency of $\tilde{\boldsymbol{\beta}}$ under H_0 and (C6),

$$\mathbf{A} = \tilde{\mathbf{A}} + O(n^{-\frac{1}{2}}),$$
$$\boldsymbol{\mathcal{J}}_U = \tilde{\boldsymbol{\mathcal{J}}}_U + O_p(n^{\frac{1}{2}}).$$

Under the current setting,

$$\begin{aligned} E(1 - \frac{\delta_i}{\pi_i}|y_i, \mathbf{z}_i) &= 1 - \frac{E(\delta_i|y_i, \mathbf{z}_i)}{\pi_i} \\ &= 1 - \frac{\pi_i}{\pi_i} = 0. \end{aligned} \tag{3.2}$$

Therefore, the WEE is unbiased. Hence,

$$E\{U_2(\tilde{\boldsymbol{\beta}})\} = 0,$$

and

$$\mathrm{Cov}\{U_2(\tilde{\boldsymbol{\beta}})\} = (-\mathbf{A}, \mathbf{I}_r)\boldsymbol{\mathcal{J}}_U(-\mathbf{A}, \mathbf{I}_r)' + O(n^{\frac{1}{2}}).$$

By condition (C7), the matrix $(-\mathbf{A}, \mathbf{I}_r)\boldsymbol{\mathcal{J}}_U(-\mathbf{A}, \mathbf{I}_r)'$ is nonsingular. Therefore, an appropriate generalized score statistic is defined as

$$T_{GS} = U_2(\tilde{\boldsymbol{\beta}})' \left\{(-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_U(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right\}^{-1} U_2(\tilde{\boldsymbol{\beta}}). \tag{3.3}$$

By the Central Limit Theorem, it is clear that $T_{GS} \to \chi^2_r$ as $n \to \infty$ under H_0 and regularity conditions. Because ϕ could be different, a class of test statistics indexed by ϕ can be constructed. From the development above, we know that the asymptotic null distribution of T_{GS} does not depend on the choice of ϕ. Note that the WEE (1.1) generally depends on the nuisance functions π and ϕ, so does T_{GS}; we write it as $T_{GS}(\pi, \phi)$ symbolically if necessary. When π is misspecified, WEE (1.1) could be biased and $T_{GS}(\pi, \phi)$ may not be an appropriate test statistic for Hypotheses (2.8).

3.2.2 Relative Efficiency

Recall that the choice of ϕ affects the asymptotic variance of the estimating equation and the corresponding estimators. When

$$\phi = \phi^*(y_i, \mathbf{z}_i) = E\{\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})|y_i, \mathbf{z}_i\},$$

the asymptotic variance of the $\hat{\boldsymbol{\beta}}$ and the WEE (1.1) is uniquely minimized in the positive definite sense. We believe that some optimality holds for generalized score tests when $\phi = \phi^*(y_i, \mathbf{z}_i)$.

Consider a sequence of local alternatives $\boldsymbol{\beta}_2^{(n)}$, such that

$$n^{\frac{1}{2}}(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20}) \to \boldsymbol{\lambda}, \tag{3.4}$$

where $||\boldsymbol{\lambda}|| > 0$. It is nature to ask if the constrained estimate $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20})$ is still consistent for $\boldsymbol{\beta}$ under the local alternative. We have the following lemma:

Lemma 3.2.1. *Under the local alternative, the constrained estimate $\tilde{\boldsymbol{\beta}}$ is root-n consistent for $\boldsymbol{\beta}$.*

Proof. Let $\tilde{\boldsymbol{\beta}}_1^{(n)}$ be the solution of the estimating equation

$$U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2^{(n)}) = \mathbf{0}.$$

Since $\boldsymbol{\beta}_2^{(n)}$ is the true value of $\boldsymbol{\beta}_2$, it is clear that $\tilde{\boldsymbol{\beta}}_1^{(n)}$ is root-n consistent for the $\boldsymbol{\beta}_1$. In addition, by Equation (3.4),

$$\begin{aligned}
\mathbf{0} &= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2^{(n)}) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20}) + O_p(1) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20}) + O_p(1) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)} + \Delta_{\boldsymbol{\beta}}^{(n)}, \boldsymbol{\beta}_2) + O_p(1),
\end{aligned}$$

where $\Delta_{\boldsymbol{\beta}}^{(n)} = E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20})$. It is clear that $\Delta_{\boldsymbol{\beta}}^{(n)}$ has order $O_p(n^{-\frac{1}{2}})$. Therefore,

$$\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_1^{(n)} + O_p(n^{-\frac{1}{2}}),$$

and thus the constrained estimate under the local alternative is root-n consistent.

□

Since the constrained estimate $\tilde{\boldsymbol{\beta}}$ is root-n consistent for $\boldsymbol{\beta}$ under the local alternative, we expend $U_1(\tilde{\boldsymbol{\beta}})$ and $U_2(\tilde{\boldsymbol{\beta}})$ at $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2^{(n)})$:

$$\begin{aligned}
0 = U_1(\tilde{\boldsymbol{\beta}}) &= U_1(\boldsymbol{\beta}) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1) \\
U_2(\tilde{\boldsymbol{\beta}}) &= U_2(\boldsymbol{\beta}) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1).
\end{aligned}$$

The first equation above implies that

$$\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = -E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}\{U_1(\boldsymbol{\beta}) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1)\}.$$

Plugging it into $U_2(\tilde{\boldsymbol{\beta}})$, we obtain

$$\begin{aligned}
U_2(\tilde{\boldsymbol{\beta}}) &= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}) + \\
&\quad \{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1}E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2}) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1) \\
&= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}) + n^{-\frac{1}{2}}\{-\mathbf{A}E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2}) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})\}\boldsymbol{\lambda} + O_p(1) \\
&= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}) + n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda} + O_p(1), \qquad (3.5)
\end{aligned}$$

where $\mathbf{C} = (-\mathbf{A}, \mathbf{I}_r)E(\frac{\partial U}{\partial \boldsymbol{\beta}_2})$. Since $\mathbf{A}$ and $E(\frac{\partial U}{\partial \boldsymbol{\beta}_2})$ are both free of ϕ (see the proof of Lemma 3.2.2), $\mathbf{C}$ is also free of ϕ. By Equation (3.5),

$$E\{U_2(\tilde{\boldsymbol{\beta}})\} = n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda} + O(1),$$

and

$$\mathrm{Cov}\{U_2(\tilde{\boldsymbol{\beta}})\} = (-\mathbf{A}, \mathbf{I}_r)E(\boldsymbol{\mathcal{J}}_U)(-\mathbf{A}, \mathbf{I}_r)' + O(n^{\frac{1}{2}}),$$

under the sequence of alternatives $\boldsymbol{\beta}_2^{(n)}$. Comparing (3.1) and (3.5), we discover that the mean of $U_2(\tilde{\boldsymbol{\beta}})$ are asymptotically different while the variance of $U_2(\tilde{\boldsymbol{\beta}})$ are asymptotically equivalent under H_0 and the local alternative. In fact, the power of the test comes from the term $n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda}$.

As in Tosteson & Tsiatis (1988), the asymptotic relative efficiency of T_{GS} to $T^*_{GS} = T_{GS}(\pi, \phi^*)$ is

$$ARE(T_{GS}, T^*_{GS}) = G/G^*$$

where G and G^* are the non-centrality parameters for T_{GS} and T^*_{GS} under the sequence of alternatives $\boldsymbol{\beta}_2^{(n)}$.

Lemma 3.2.2. *When the selection probability π is known and ϕ is given, the asymptotic relative efficiency $ARE(T_{GS}, T^*_{GS}) \leq 1$. The equality holds iff $\phi = \phi^*$ a.e.*

Proof. By Equation (3.2),

$$\begin{aligned} E(\frac{\partial U}{\partial \boldsymbol{\beta}}) &= E\sum_{i=1}^{n}\{\frac{\delta_i}{\pi_i}\frac{\partial \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + (1 - \frac{\delta_i}{\pi_i})\frac{\partial \phi(y_i, \mathbf{z}_i)}{\partial \boldsymbol{\beta}}\} \\ &= E\sum_{i=1}^{n}\{\frac{\delta_i}{\pi_i}\frac{\partial \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\} + E_{y\mathbf{z}}\{E(1 - \frac{\delta_i}{\pi_i}|y_i, \mathbf{z}_i)\frac{\partial \phi(y_i, \mathbf{z}_i)}{\partial \boldsymbol{\beta}}\} \\ &= E\sum_{i=1}^{n}\{\frac{\delta_i}{\pi_i}\frac{\partial \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\} \end{aligned}$$

indicating $E(\frac{\partial U}{\partial \boldsymbol{\beta}})$ are free of ϕ for any $\boldsymbol{\beta}$. Consequently, the matrix $\mathbf{A}$ in (3.1) does not depend the choice of ϕ. By Equation (2.2),

$$E(\frac{\boldsymbol{\mathcal{J}}_U}{n}) = E(\mathbf{u}_1\mathbf{u}_1') = E(\psi\psi') + E[(1-\pi)\pi E\{(\psi - \phi)(\psi - \phi)'|y, \mathbf{z}\}$$

where $\psi = \psi(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta})$. Obviously, $E(\frac{1}{n}\boldsymbol{\mathcal{J}}_U)$ is minimized at $\phi = \phi^* = E\{\psi(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta})|y, \mathbf{z}\}$ in the positive definite sense. Hence, we can write

$$E(\boldsymbol{\mathcal{J}}_U) = E(\boldsymbol{\mathcal{J}}_U^*) + \mathbf{D},$$

where $\boldsymbol{\mathcal{J}}_U^*$ is $\boldsymbol{\mathcal{J}}_U$ evaluated at $\phi = \phi^*$ and $\mathbf{D}$ is a positive definite matrix when $\phi \neq \phi^*$.

The asymptotic non-centrality parameter for T_{GS} is

$$\begin{aligned} G &= \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'\{\mathbf{K}E(\boldsymbol{\mathcal{J}}_U)\mathbf{K}'\}^{-1}\mathbf{C}\boldsymbol{\lambda} \\ &= \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'\left[\mathbf{K}\{E(\boldsymbol{\mathcal{J}}_U^*) + \mathbf{D}\}\mathbf{K}'\right]^{-1}\mathbf{C}\boldsymbol{\lambda} \\ &= \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'(\mathbf{K}_J + \mathbf{K}_D)^{-1}\mathbf{C}\boldsymbol{\lambda}, \end{aligned}$$

where $\mathbf{K} = (-\mathbf{A}, \mathbf{I}_r)$, $\mathbf{K}_J = \mathbf{K}E(\boldsymbol{\mathcal{J}}_U^*)\mathbf{K}'$ and $\mathbf{K}_D = \mathbf{K}\mathbf{D}\mathbf{K}'$. The matrix $\mathbf{K}$ is also free of ϕ. Because $\mathbf{K}$ is full row rank, $\mathbf{K}_J$ and $\mathbf{K}_D$ are nonsingular when $\phi \neq \phi^*$. By Lemma 3.5.1,

$$\{\mathbf{K}_J + \mathbf{K}_D\}^{-1} = \mathbf{K}_J^{-1} - \mathbf{K}_J^{-1}(\mathbf{K}_J^{-1} + \mathbf{K}_D^{-1})^{-1}\mathbf{K}_J^{-1}.$$

Hence,

$$\begin{aligned} G &= \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'\{\mathbf{K}E(\boldsymbol{\mathcal{J}}_U^*)\mathbf{K}'\}^{-1}\mathbf{C}\boldsymbol{\lambda} - G_0 \\ &= G^* - G_0, \end{aligned}$$

where

$$G_0 = \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'\mathbf{K}_J^{-1}(\mathbf{K}_J^{-1} + \mathbf{K}_D^{-1})^{-1}\mathbf{K}_J^{-1}\mathbf{C}\boldsymbol{\lambda}.$$

When $\phi \neq \phi^*$, it is obvious that $(\mathbf{K}_J^{-1} + \mathbf{K}_D^{-1})^{-1}$ is positive definite, and thus $G_0 > 0$. It is clear that $G = G^*$ when $\phi = \phi^*$. Then

$$ARE(T_{GS}, T_{GS}^*) = \frac{G^* - G_0}{G^*} < 1,$$

completing the proof. □

The lemma implies that the asymptotic optimal test among all the choices of ϕ is achieved when $\phi = \phi^*$ in the current setting. As we will see later, the asymptotic relative efficiencies between different generalized score test statistics are given for both parametric and semiparametric settings, with the conclusion that T^*_{GS} and some other test statistics achieve the same asymptotic optimality in all these settings. Robins et al. (1994) showed the asymptotic variance of WEE (1.1) is uniquely minimized in the positive definite sense when $\phi = \phi^*$. From the development of the noncentrality parameter under the local alternative, we find out that the proposed tests keep this optimum property when $\phi = \phi^*$.

3.3 Parametric Setting

In many epidemiological studies, data are missing by happenstance rather than design, and thus the selection probability $\pi(y_i, \mathbf{z}_i)$ in WEE (1.1) is generally unknown and needs to be estimated. Furthermore, Robin et al. (1994) and Wang et al. (1997) showed that one can improve the efficiency of the inefficient estimators in their class by estimating the selection probability even when it is known.

On the other hand, by Lemma (3.2.2), it is intuitive to have ϕ^* or its estimate in WEE (1.1) to achieve good power in a generalized score test. Recall that ϕ^* is the mean score with respect to the conditional distribution $p(\mathbf{x}_i|y_i, \mathbf{z}_i)$, which is usually unknown too. Then additional models may be required to estimate ϕ^*.

One common approach is to assume parametric models for π and ϕ^*. In this section, we would like to obtain appropriate generalized score statistics in different parametric settings, study how parametric estimates of the selection probability and ϕ^* affect the test statistics and investigate the efficiency issues. In particular, we focus on the following special settings: (a) both π and ϕ^* are estimated via parametric models using joint estimating equation; (b) the selection probability is estimated via

a parametric model and ϕ is given; and (c) the selection probability is known and ϕ^* is estimated via a parametric model.

3.3.1 The Case of π and ϕ^* Being Estimated Using a Joint Estimating Equation

Zhao et al. (1996) introduced a joint estimating equation for regression analysis and Lipsitz et al.(1999) proposed proposed another joint estimating equation similar to the maximum likelihood equation for missing covariate data. They all assumed that the selection probability follows a logistic regression (2.3)

$$\pi_i = \pi_i(\boldsymbol{\alpha}) = \frac{\exp(-\boldsymbol{\alpha}'\mathbf{v}_i)}{1+\exp(-\boldsymbol{\alpha}'\mathbf{v}_i)},$$

where $\mathbf{v}_i$ is a vector function of $(y_i, \mathbf{z}_i)$ and $\boldsymbol{\alpha}$ is finite dimensional. Recall that the maximum likelihood estimate $\tilde{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ can be obtained using

$$U_\pi(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{v}_i \{\delta_i - \pi_i(\boldsymbol{\alpha})\} = \mathbf{0}.$$

However, they used different parametric models for ϕ^*. Zhao et al. (1996) used the assumptions regarding conditional moments and Lipsitz et al. (1999) used assumptions on the conditional distributions to build the parametric model for ϕ^*. To include both settings above and other possible situations in a unified way, we assume that $U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ is the estimating equation corresponding to a general model for ϕ^* with an additional finite dimensional parameter $\boldsymbol{\kappa}$. Therefore, to solve the parameter $\boldsymbol{\tau} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \boldsymbol{\kappa}')'$, we have a general joint estimating equation

$$0 = U_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \begin{pmatrix} U_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ U_\pi(\boldsymbol{\alpha}) \\ U_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \end{pmatrix}. \tag{3.6}$$

Let $\tilde{\boldsymbol{\tau}} = (\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\kappa}}')'$ be the solution of the joint estimating equation (3.6) under H_0. The estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\kappa}}$ can be obtained by iterative methods using U, U_{ϕ^*} and $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$.

Let $\hat{\pi} = \pi(\tilde{\alpha})$ and $\hat{\phi}^*$ be the corresponding estimate of ϕ^*. Note that the true value of $\boldsymbol{\kappa}$ might be meaningless when the model for ϕ^* is incorrect. In that situation, we assume that there is a vector $\boldsymbol{\kappa}^*$ such that $\tilde{\boldsymbol{\kappa}} \rightarrow \boldsymbol{\kappa}^*$ at a root-n rate, and let $\boldsymbol{\kappa}^*$ be the true value of $\boldsymbol{\kappa}$. Assume that ψ_J is the estimating function of U_J. When ψ_J is continuous with respect to $\boldsymbol{\tau}$ at a neighborhood of the solution for $E(\psi_J) = 0$ and the second moment of ψ_J exists, such a vector $\boldsymbol{\kappa}^*$ exists. In addition, a correct model for ϕ^* means that the estimate $\hat{\phi}^*$ is consistent whether the model for the selection probability is correct or not. For example, a correct model for ϕ^* in Lipsitz et al. (1999) requires that the distributional assumptions on $f(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\beta})$ and $p(\mathbf{x}_i|\mathbf{z}_i; \boldsymbol{\kappa})$ are correct. Following the ideas in Scharfstein et al. (1999), Zhao et al. (1996) and Lipsitz et al. (1999), it is easy to see that the estimator based on Equation (3.6) is doubly robust. The estimate $\tilde{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ under H_0 when at least one of the following is correctly specified: (a) the model for the selection probability or (b) the model for ϕ^*.

Because it is full parametric setting where nuisance functions π and ϕ^* are re-parametrized to finite dimensional nuisance parameters $\boldsymbol{\alpha}$and $\boldsymbol{\kappa}$, it seems that the generalized score statistics in Boos (1992) can be applied easily using U_J instead of U in test statistic (2.10). Assuming that $\tilde{\boldsymbol{\beta}}$ is consistent, the Boos's test statistic

$$T_{GSB} = U_2(\tilde{\boldsymbol{\tau}})' \left\{ (-\tilde{\tilde{\mathbf{A}}}, \mathbf{I}_r) \tilde{\boldsymbol{\mathcal{J}}}_{U_J} (-\tilde{\tilde{\mathbf{A}}}, \mathbf{I}_r)' \right\}^{-1} U_2(\tilde{\boldsymbol{\tau}}),$$

where

$$\tilde{\tilde{\mathbf{A}}} = -E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}}) E(\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1})^{-1},$$

$U_{J1} = (U_1', U_{\phi^*}', U_\pi')'$, and $\boldsymbol{\tau}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\alpha}', \boldsymbol{\kappa}')'$. However, two difficulties exist. First, because the model for the selection probability may not be appropriate, the estimating equation U_{ϕ^*} and U_J could be biased. Hence, $\tilde{\boldsymbol{\mathcal{J}}}_{U_J}$ may not be a consistent estimate of covariance matrix of U_J. Therefore, it is questionable that T_{GSB} still

follow χ^2_r asymptotically under H_0. Second, even given that this test statistics T_{GSB} is appropriate, the sub-matrices $E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}})$ and $E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}})$ of the matrix $E(\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1})$ could be extremely difficult to calculate because U_{ϕ^*} may not have a close form. We obtain relatively simple and appropriate test statistics in the following.

Before we introduce the test statistic and the main theorem, we would like to state two lemmas.

Lemma 3.3.1. *If a parametric model such as (2.3) is correct for the selection probability function in the setting using Equation (3.6), then*

$$E\{\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\} = 0.$$

Because the lemma has nothing to do with U_{ϕ^*}, the proof of the lemma is essential same as that of Lipsitz et al. (1999).

Lemma 3.3.2. *If the model for ϕ^* is correctly specified in (3.6), then*

$$E\{\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})}{\partial \boldsymbol{\alpha}}\} = 0.$$

When the model for the selection probability is correctly specified, as we will show in the proof of Theorem 3.3.1, an appropriate generalized score statistic is

$$T^*_{GSP} = U_2(\tilde{\boldsymbol{\tau}})' \left\{(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)'\right\}^{-1} U_2(\tilde{\boldsymbol{\tau}}), \tag{3.7}$$

where $\tilde{\mathbf{A}} = \mathbf{A}|_{\tilde{\boldsymbol{\tau}}}$, $\mathbf{B} = \{\mathbf{A}E(\frac{\partial U_1}{\partial \boldsymbol{\alpha}}) - E(\frac{\partial U_2}{\partial \boldsymbol{\alpha}})\}E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1}$, $\tilde{\mathbf{B}} = \mathbf{B}|_{\tilde{\boldsymbol{\tau}}}$, $U_R = (U_1', U_\pi', U_2')' = \sum_{i=1}^n \mathbf{u}_{Ri}$, $\boldsymbol{\mathcal{J}}_{U_R} = \sum_{i=1}^n \mathbf{u}_{Ri}(\boldsymbol{\tau})\mathbf{u}'_{Ri}(\boldsymbol{\tau})$ and $\tilde{\boldsymbol{\mathcal{J}}}_{U_R} = \boldsymbol{\mathcal{J}}_{U_R}|_{\tilde{\boldsymbol{\tau}}}$. An interesting finding here is that the test statistic T^*_{GSP} is free of U_{ϕ^*} and $\frac{\partial U}{\partial \boldsymbol{\kappa}}$. Equation (3.7) indicates that the estimates of π and ϕ^* have some effect on the generalized score statistics in the parametric setting. The test statistic T_{GS} is not an appropriate test statistic generally. Recall that $T_{GS}(\pi, \phi) = U_2(\tilde{\boldsymbol{\beta}})' \left\{(-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_U(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right\}^{-1} U_2(\tilde{\boldsymbol{\beta}})$ was developed in the

last subsection when the selection probability is known and ϕ is given. By Lemma 3.3.2, it is seen that if the model for ϕ^* is also correctly specified, T^*_{GSP} in (3.7) reduces to

$$T_{GS}(\hat{\pi}, \hat{\phi}^*) = U_2(\tilde{\boldsymbol{\tau}})' \left\{ (-\tilde{\mathbf{A}}, \mathbf{I}_r) \tilde{\boldsymbol{\mathcal{J}}}_U (-\tilde{\mathbf{A}}, \mathbf{I}_r)' \right\}^{-1} U_2(\tilde{\boldsymbol{\tau}})$$

since $\mathbf{B} = \mathbf{0}$. This implies that T_{GS} is still an appropriate generalized score statistic if both π and ϕ^* are estimated using correct parametric models. We now provide the main theorem:

Theorem 3.3.1. *Given the parametric setting based on the joint estimating equation (3.6) and assuming that the model for the selection probability is correctly specified, under suitable regularity conditions and H_0, we have $T^*_{GSP} \to \chi^2_r$ in distribution as $n \to \infty$.*

Proof. Let $\tilde{\boldsymbol{\tau}}_1 = (\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\kappa}}')'$. Under H_0 and condition (C6), by expanding $U_J(\tilde{\boldsymbol{\tau}})$ at the true value $\boldsymbol{\tau}$, we obtain

$$0 = U_{J1}(\tilde{\boldsymbol{\tau}}) = U_{J1}(\boldsymbol{\tau}) + E(\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1})(\tilde{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\tau}}) = U_2(\boldsymbol{\tau}) + E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_1})(\tilde{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1) + O_p(1).$$

Combining the equations above, we have the following results similar to (3.1):

$$U_2(\tilde{\boldsymbol{\tau}}) = \left\{ -E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}}) \boldsymbol{\mathcal{I}}^{-1}_{J11}, \mathbf{I}_r \right\} U_J(\boldsymbol{\tau}) + O_p(1), \tag{3.8}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{J11} &= E \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_1}{\partial \boldsymbol{\kappa}} & \frac{\partial U_1}{\partial \boldsymbol{\alpha}} \\ \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\alpha}} \\ \frac{\partial U_\pi}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_\pi}{\partial \boldsymbol{\kappa}} & \frac{\partial U_\pi}{\partial \boldsymbol{\alpha}} \end{pmatrix} \\ &= E \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \mathbf{0} & \frac{\partial U_1}{\partial \boldsymbol{\alpha}} \\ \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\alpha}} \\ \mathbf{0} & \mathbf{0} & \frac{\partial U_\pi}{\partial \boldsymbol{\alpha}} \end{pmatrix}. \end{aligned}$$

In addition, by Lemma(3.3.1),

$$U_2(\tilde{\boldsymbol{\tau}}) = \left\{ -E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \mathbf{0}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}}) \boldsymbol{\mathcal{I}}_{J11}^{-1}, \mathbf{I}_r \right\} U_J(\boldsymbol{\tau}) + O_p(1). \tag{3.9}$$

The inverse matrix is

$$\boldsymbol{\mathcal{I}}_{J11}^{-1} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} & \mathbf{0} & -E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} E(\frac{\partial U_1}{\partial \boldsymbol{\alpha}}) E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \\ * & * & * \\ \mathbf{0} & \mathbf{0} & E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \end{pmatrix}, \tag{3.10}$$

where $*$'s above are some constants. The detailed proof of Equation (3.10) is shown in Section 3.5 (Technical Detail). Therefore, (3.9) can be rewritten as

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}) &= \{-\mathbf{A}, \mathbf{0}, \mathbf{B}, \mathbf{I}_r\} U_J(\boldsymbol{\tau}) + O_p(1) \\ &= \{-\mathbf{A}, \mathbf{B}, \mathbf{I}_r\} U_R(\boldsymbol{\tau}) + O_p(1), \end{aligned}$$

where $\mathbf{A}$ was given in (3.1) and $\mathbf{B}$ in (3.7).

When the model for the selection probability is correctly specified, it is clear that $U_R(\boldsymbol{\tau})$ is an unbiased estimating equation and $\frac{1}{n}\tilde{\boldsymbol{\mathcal{J}}}_U$ is a root-n consistent estimate of the Cov$\{n^{-\frac{1}{2}} U_J(\boldsymbol{\tau})\}$. Then

$$E\{n^{-\frac{1}{2}} U_2(\tilde{\boldsymbol{\tau}})\} = O(n^{-\frac{1}{2}}),$$

and

$$\text{Cov}\{n^{-\frac{1}{2}} U_2(\tilde{\boldsymbol{\tau}})\} = \frac{1}{n}(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r) \tilde{\boldsymbol{\mathcal{J}}}_{U_R} (-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)' + O(n^{-\frac{1}{2}}).$$

By the Central Limit Theorem, $T_{GSP}^* \to \chi_r^2$ in distribution as $n \to \infty$. □

Note that the estimator based on $U_J(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ is doubly robust. Therefore, it is interesting to check if T_{GSP}^* has similar robust property. More specifically, we would like to check if T_{GSP}^* is still an appropriate generalized score when the model for the selection probability is misspecified but the model for ϕ^* is correctly specified.

If the model for ϕ^* is correct, by Lemma (3.3.1), (3.3.2) and expanding $U_J(\tilde{\boldsymbol{\tau}})$ at the true value $\boldsymbol{\tau}$ and replacing $\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1}$ and $\frac{\partial U_2}{\partial \boldsymbol{\tau}_1}$ by their asymptotically equivalent versions $E(\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1})$ and $E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_1})$ under H_0, we obtain

$$U_2(\tilde{\boldsymbol{\tau}}) = \left\{ -E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \mathbf{0})\boldsymbol{\mathcal{I}}_{J11}^{-1}, \mathbf{I}_r \right\} U_J(\boldsymbol{\tau}) + O_p(1),$$

where

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{J11} &= E \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_1}{\partial \boldsymbol{\kappa}} & \frac{\partial U_1}{\partial \boldsymbol{\alpha}} \\ \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\alpha}} \\ \frac{\partial U_{\pi}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\pi}}{\partial \boldsymbol{\kappa}} & \frac{\partial U_{\pi}}{\partial \boldsymbol{\alpha}} \end{pmatrix} \\ &= E \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_1}{\partial \boldsymbol{\kappa}} & \mathbf{0} \\ \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\alpha}} \\ \mathbf{0} & \mathbf{0} & \frac{\partial U_{\pi}}{\partial \boldsymbol{\alpha}} \end{pmatrix}. \end{aligned}$$

In general, $\frac{\partial U_1}{\partial \boldsymbol{\kappa}}$ is not equal $\mathbf{0}$ when the model for the selection probability is misspecified. With some algebra, we will see that $U_2(\tilde{\boldsymbol{\tau}})$ is not free of $\frac{\partial U_1}{\partial \boldsymbol{\kappa}}$ and U_{ϕ^*}. Therefore, T^*_{GSP} is definitely not appropriate generalized score statistics when the model for the selection probability is misspecified but the model for ϕ^* is correctly specified. From the development above, an appropriate generalized score statistic generally depends on $\frac{\partial U_1}{\partial \boldsymbol{\kappa}}$, $\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1}$, and $\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}}$ which are difficult to obtain in this setting. Therefore, generalized score tests are not very useful when the model for the selection probability is inappropriate.

It is also interesting to study the asymptotic efficiency of test statistic T^*_{GSP} given appropriate assumptions. First, following the the proof of Lemma (3.2.1), we can show that the constrained estimate $\tilde{\boldsymbol{\tau}} = (\tilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20}, \tilde{\boldsymbol{\kappa}}, \tilde{\boldsymbol{\alpha}})$ is still consistent for $\boldsymbol{\tau}$ under the sequence of local alternatives $\boldsymbol{\beta}_2^{(n)}$ in (3.4). Therefore, by expanding $U_J(\tilde{\boldsymbol{\tau}})$

at the true value $\boldsymbol{\tau}$, we obtain

$$0 = U_{J1}(\tilde{\boldsymbol{\tau}}) = U_{J1}(\boldsymbol{\tau}) + E(\frac{\partial U_{J1}}{\partial \boldsymbol{\tau}_1})(\tilde{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1) + E(\frac{\partial U_{J1}}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\tau}}) = U_2(\boldsymbol{\tau}) + E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_1})(\tilde{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$

and hence

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}) &= \left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{J11}^{-1}, \mathbf{I}_r\right\} U_J(\boldsymbol{\tau}) + \\ &\quad \left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{J11}^{-1}, \mathbf{I}_r\right\} E\{\frac{\partial U_J(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1), \end{aligned}$$

under the local alternative. From the proof of Theorem 3.3.1, it is easy to see that

$$\left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\kappa}}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{J11}^{-1}, \mathbf{I}_r\right\} = (-\mathbf{A}, \mathbf{0}, \mathbf{B}, \mathbf{I}_r).$$

Since $\partial U_\pi / \partial \boldsymbol{\beta}_2 = \mathbf{0}$,

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}) &= (-\mathbf{A}, \mathbf{0}, \mathbf{B}, \mathbf{I}_r)U_J(\boldsymbol{\tau}) + \\ &\quad (-\mathbf{A}, \mathbf{0}, \mathbf{B}, \mathbf{I}_r)E\{\frac{\partial U_J(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1) \\ &= (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r)U_R(\boldsymbol{\tau}) + \\ &\quad (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1). \end{aligned} \tag{3.11}$$

Therefore, it is clear that

$$\begin{aligned} E\{U_2(\tilde{\boldsymbol{\tau}})\} &= E\{(-\mathbf{A}, \mathbf{B}, \mathbf{I}_r)U_R(\boldsymbol{\tau})\} + (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O(1) \\ &= n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda} + O(1), \end{aligned}$$

and

$$\mathrm{Cov}\{U_2(\tilde{\boldsymbol{\tau}})\} = (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r)E(\boldsymbol{\mathcal{J}}_{U_R})(-\mathbf{A}, \mathbf{B}, \mathbf{I}_r)' + O(n^{\frac{1}{2}}).$$

By the results above, we can obtain the noncentrality parameters for T_{GSP}^* and results concerning efficiency. We have the following lemma:

Lemma 3.3.3. *If the models for π and ϕ^* are correctly specified in the parametric setting using joint estimating equation (3.6), then*

$$ARE(T^*_{GSP}, T^*_{GS}) = 1.$$

Proof. Since both the models for π and ϕ^* are correct, $\mathbf{B} = 0$ and

$$U(\boldsymbol{\tau}) = U(\boldsymbol{\beta}, \pi, \phi^*),$$

Therefore, Equation (3.11) reduces to

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}) &= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \pi, \phi^*) + \\ &\quad (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1), \end{aligned}$$

which is asymptotically equivalent to $U_2(\tilde{\boldsymbol{\beta}})$ under the local alternative when π is known and ϕ^* is given. Therefore, $ARE(T^*_{GSP}, T_{GS}*) = 1$.

3.3.2 The Case of π Being Estimated Parametrically and ϕ Being Given

Recall that ϕ^* depends on the conditional distribution $p(\mathbf{x}_i|y_i, \mathbf{z}_i)$, which is unknown in general. An appropriate model for ϕ^* usually is complicated. Because of the simplicity, the weighted estimating equations with $\phi = 0$, which was proposed by Zhao and Lipsitz (1992), are also widely used. We consider one reduced parametric setting in which the ϕ is given and the selection probability is estimated in this subsection.

We keep using the assumption that the selection probability follows logistic regression (2.3). We have the joint estimating equation

$$0 = U_R(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi) = \begin{pmatrix} U_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi) \\ U_\pi(\boldsymbol{\alpha}) \\ U_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi) \end{pmatrix}, \tag{3.12}$$

where ϕ is given. More precisely, ϕ can still be estimated from the data and plugged in $U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)$ in this setting, as long as the model for ϕ is free of $\boldsymbol{\beta}$. Therefore, the ϕ can be solved without $U(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)$ and thus no iteration algorithm is necessary. Note that such a model for ϕ can not be the correct model for ϕ^*, because the correct model for ϕ^* involves $\boldsymbol{\beta}$ in general. Equation $U_R(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)$ actually is $U_R(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Let $\boldsymbol{\tau}_R = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ and $\tilde{\boldsymbol{\tau}}_R = (\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\alpha}}')'$ be the solution of the equation under H_0. It is clear that $\tilde{\boldsymbol{\beta}}$ is root-n consistent if the model for the selection probability is correct.

Under the current setting, an appropriate generalized score statistic is

$$T_{GSP} = U_2(\tilde{\boldsymbol{\tau}}_R)' \left\{ (-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r) \tilde{\boldsymbol{\mathcal{J}}}_{U_R} (-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)' \right\}^{-1} U_2(\tilde{\boldsymbol{\tau}}_R), \tag{3.13}$$

where $\tilde{\mathbf{A}} = \mathbf{A}|_{\tilde{\boldsymbol{\tau}}_R}$, $\tilde{\mathbf{B}} = \mathbf{B}|_{\tilde{\boldsymbol{\tau}}_R}$, and $\tilde{\boldsymbol{\mathcal{J}}}_{U_R} = \boldsymbol{\mathcal{J}}_{U_R}|_{\tilde{\boldsymbol{\tau}}_R}$.

Theorem 3.3.2. *Assuming that the model for the selection probability is correctly specified in the parametric setting based on Equation (3.12), under suitable regularity conditions and H_0, $T_{GSP} \to \chi^2_r$ in distribution as $n \to \infty$.*

Proof. Let $U_{R1} = (U_1', U_\pi')'$, $\boldsymbol{\tau}_{R1} = (\boldsymbol{\beta}_1', \boldsymbol{\alpha}')'$ and $\tilde{\boldsymbol{\tau}}_{R1} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\alpha}}')'$. Under H_0, by expanding $U_R(\tilde{\boldsymbol{\tau}}_R)$ at the true value $\boldsymbol{\tau}_R$, we obtain

$$\begin{aligned} 0 = U_{R1}(\tilde{\boldsymbol{\tau}}_R) &= U_{R1}(\boldsymbol{\tau}_R) + E(\frac{\partial U_{R1}}{\partial \boldsymbol{\tau}_{R1}})(\tilde{\boldsymbol{\tau}}_{R1} - \boldsymbol{\tau}_{R1}) + O_p(1), \\ U_2(\tilde{\boldsymbol{\tau}}_R) &= U_2(\boldsymbol{\tau}_R) + E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_{R1}})(\tilde{\boldsymbol{\tau}}_{R1} - \boldsymbol{\tau}_{R1}) + O_p(1). \end{aligned}$$

Combining the equations above, we have

$$U_2(\tilde{\boldsymbol{\tau}}_R) = \left\{ -E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}}) \boldsymbol{\mathcal{I}}_{R11}^{-1}, \mathbf{I}_r \right\} U_R(\boldsymbol{\tau}_R) + O_p(1), \tag{3.14}$$

where

$$\boldsymbol{\mathcal{I}}_{R11} = \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_1}{\partial \boldsymbol{\alpha}} \\ \mathbf{0} & \frac{\partial U_\pi}{\partial \boldsymbol{\alpha}} \end{pmatrix}.$$

By Lemma (3.5.2),

$$\boldsymbol{\mathcal{I}}_{R11}^{-1} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} & -E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} E(\frac{\partial U_1}{\partial \boldsymbol{\alpha}}) E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \\ \mathbf{0} & E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \end{pmatrix},$$

Therefore, (3.14) can be rewritten as

$$U_2(\tilde{\boldsymbol{\tau}}_R) = \{-\mathbf{A}, \mathbf{B}, \mathbf{I}_r\} U_R(\boldsymbol{\tau}_R) + O_p(1).$$

When the model for the selection probability is correctly specified, it is clear that $U_R(\boldsymbol{\tau}_R)$ is an unbiased estimating equation. Therefore,

$$E\{n^{-\frac{1}{2}} U_2(\tilde{\boldsymbol{\tau}}_R)\} = O(n^{-\frac{1}{2}}),$$

and

$$\mathrm{Cov}\{n^{-\frac{1}{2}} U_2(\tilde{\boldsymbol{\tau}}_R)\} = \frac{1}{n}(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)' + O(n^{-\frac{1}{2}}).$$

By the Central Limit Theorem, $T_{GSP} \to \chi_r^2$ in distribution as $n \to \infty$. □

In practice, ϕ may takes 0 for simplicity. With the additional assumption that $\phi = 0$, we have

$$T_{GSP} = U_2(\tilde{\boldsymbol{\tau}}_R)' \left[(-\tilde{\mathbf{A}}, \mathbf{I}_r)\left\{\tilde{\boldsymbol{\mathcal{J}}}_U - \tilde{\mathbf{F}}\right\}(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right]^{-1} U_2(\tilde{\boldsymbol{\tau}}_R) + O_p(n^{-\frac{1}{2}}), \qquad (3.15)$$

where $\mathbf{F} = E(\frac{\partial U}{\partial \boldsymbol{\alpha}}) E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} E(\frac{\partial U}{\partial \boldsymbol{\alpha}})'$ and $\tilde{\mathbf{F}} = \mathbf{F}|_{\tilde{\boldsymbol{\tau}}_R}$. The proof of the Equation (3.15) is given as following.

Proof. First express the matrix

$$\tilde{\boldsymbol{\mathcal{J}}}_{U_R} = \begin{bmatrix} \tilde{\boldsymbol{\mathcal{J}}}_{11} & \tilde{\boldsymbol{\mathcal{J}}}_{1\pi} & \tilde{\boldsymbol{\mathcal{J}}}_{12} \\ \tilde{\boldsymbol{\mathcal{J}}}_{\pi 1} & \tilde{\boldsymbol{\mathcal{J}}}_{\pi\pi} & \tilde{\boldsymbol{\mathcal{J}}}_{\pi 2} \\ \tilde{\boldsymbol{\mathcal{J}}}_{21} & \tilde{\boldsymbol{\mathcal{J}}}_{2\pi} & \tilde{\boldsymbol{\mathcal{J}}}_{22} \end{bmatrix},$$

where $\tilde{\boldsymbol{\mathcal{J}}}_{11}$ is a $(p-r)\times(p-r)$ matrix, and $\tilde{\boldsymbol{\mathcal{J}}}_{22}$ is an $r \times r$ matrix. Since the model for the selection probability is correct and $U_\pi(\boldsymbol{\alpha})$ is the maximum likelihood equations,

$$\begin{aligned} E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}}) &= Cov\{U_\pi(\boldsymbol{\alpha})\} \\ &= \tilde{\boldsymbol{\mathcal{J}}}_{\pi\pi} + O(n^{\frac{1}{2}}). \end{aligned}$$

In addition,

$$E(\frac{\partial U}{\partial \boldsymbol{\alpha}})' = (\tilde{\boldsymbol{\mathcal{J}}}'_{1\pi}, \tilde{\boldsymbol{\mathcal{J}}}'_{2\pi}) + O_p(n^{\frac{1}{2}}).$$

In fact,

$$\begin{aligned} \mathbf{B} &= \{\mathbf{A}E(\frac{\partial U_1}{\partial \boldsymbol{\alpha}}) - E(\frac{\partial U_2}{\partial \boldsymbol{\alpha}})\}E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \\ &= (\mathbf{A}, -\mathbf{I}_r)E(\frac{\partial U}{\partial \boldsymbol{\alpha}})E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1}. \end{aligned}$$

Therefore, $\tilde{\mathbf{V}}^*_{GSP} = \left\{(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(-\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{I}_r)'\right\}$ can be rewritten as

$$\begin{aligned} \tilde{\mathbf{V}}^*_{GSP} &= (-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_U(-\tilde{\mathbf{A}}, \mathbf{I}_r)' + (\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})' + (-\tilde{\mathbf{A}}, \mathbf{0}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})' \\ &\quad +(\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(-\tilde{\mathbf{A}}, \mathbf{0}, \mathbf{I}_r)' \\ &= (-\tilde{\mathbf{A}}, \mathbf{I}_r)(\tilde{\boldsymbol{\mathcal{J}}}_U + \tilde{\mathbf{F}})(-\tilde{\mathbf{A}}, \mathbf{I}_r)' + (-\tilde{\mathbf{A}}, \mathbf{0}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})' \\ &\quad +(\mathbf{0}, \tilde{\mathbf{B}}, \mathbf{0})\tilde{\boldsymbol{\mathcal{J}}}_{U_R}(-\tilde{\mathbf{A}}, \mathbf{0}, \mathbf{I}_r)' \\ &= (-\tilde{\mathbf{A}}, \mathbf{I}_r)(\tilde{\boldsymbol{\mathcal{J}}}_U + \tilde{\mathbf{F}})(-\tilde{\mathbf{A}}, \mathbf{I}_r)' - 2(-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\mathbf{F}}(-\tilde{\mathbf{A}}, \mathbf{I}_r)' + O_p(n^{\frac{1}{2}}) \\ &= (-\tilde{\mathbf{A}}, \mathbf{I}_r)(\tilde{\boldsymbol{\mathcal{J}}}_U - \tilde{\mathbf{F}})(-\tilde{\mathbf{A}}, \mathbf{I}_r)' + O_p(n^{\frac{1}{2}}). \end{aligned} \tag{3.16}$$

Since the matrix $\tilde{\mathbf{V}}^*_{GSP}$ is $O_p(n)$, Equation (3.7) holds. □

We may have other appropriate generalized score statistic when $\phi = 0$. Let

$$T_{GSP0} = U_2(\tilde{\boldsymbol{\tau}}_R)' \left[(-\tilde{\mathbf{A}}, \mathbf{I}_r)\left\{\tilde{\boldsymbol{\mathcal{J}}}_U - \tilde{\mathbf{F}}\right\}(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right]^{-1} U_2(\tilde{\boldsymbol{\tau}}_R).$$

We have the following corollary:

Corollary 3.3.1. *Assuming that the model for the selection probability is correctly specified and $\phi = 0$ in the parametric setting based on Equation (3.12), under suitable regularity conditions and H_0, $T_{GSP0} \to \chi^2_r$ in distribution as $n \to \infty$.*

The corollary implies that T_{GSP0} is an appropriate generalized score statistic when $\phi = 0$ under the current parametric setting. If $E(\frac{\partial U}{\partial \boldsymbol{\alpha}})$ is of full row rank, we can show that

$$T_{GSP0} = T_{GS}(\pi, 0) - C_\Delta,$$

where C_Δ is a positive number.

In addition, we would like to study the asymptotic efficiency of the proposed test based on U_R in Equation (3.12). Following the the proof of Lemma (3.2.1), it is easy to see that the constrained estimate $\tilde{\boldsymbol{\tau}}_R = (\tilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20}, \tilde{\boldsymbol{\alpha}})$ is consistent for $\boldsymbol{\tau}_R$ under the sequence of local alternatives $\boldsymbol{\beta}_2^{(n)}$ in (3.4). Therefore, we obtain

$$0 = U_{R1}(\tilde{\boldsymbol{\tau}}_R) = U_{R1}(\boldsymbol{\tau}_R) + E(\frac{\partial U_{R1}}{\partial \boldsymbol{\tau}_{R1}})(\tilde{\boldsymbol{\tau}}_{R1} - \boldsymbol{\tau}_{R1}) + E(\frac{\partial U_{R1}}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\tau}}_R) = U_2(\boldsymbol{\tau}_R) + E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_{R1}})(\tilde{\boldsymbol{\tau}}_{R1} - \boldsymbol{\tau}_{R1}) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1)$$

by expanding $U_R(\tilde{\boldsymbol{\tau}}_R)$ at the true value $\boldsymbol{\tau}_R$ under the local alternative. Hence,

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}_R) &= \left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{R11}^{-1}, \mathbf{I}_r\right\} U_R(\boldsymbol{\tau}_R) + \\ &\quad \left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{R11}^{-1}, \mathbf{I}_r\right\} E\{\frac{\partial U_R(\boldsymbol{\tau}_R)}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1). \end{aligned}$$

From the proof of Theorem 3.3.2, it is readily to see that

$$\left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \frac{\partial U_2}{\partial \boldsymbol{\alpha}})\boldsymbol{\mathcal{I}}_{R11}^{-1}, \mathbf{I}_r\right\} = (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r).$$

Since $\partial U_\pi / \partial \boldsymbol{\beta}_2 = \mathbf{0}$,

$$\begin{aligned} U_2(\tilde{\boldsymbol{\tau}}_R) &= (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r) U_R(\boldsymbol{\tau}_R) + \\ &\quad (-\mathbf{A}, \mathbf{I}_r) E\{\frac{\partial U(\boldsymbol{\tau}_R)}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1). \end{aligned} \tag{3.17}$$

By the equation above, it is easy to obtain the noncentrality parameters for T_{GSP} and T_{GSP0}. It follows a lemma of asymptotic relative efficiency:

Lemma 3.3.4. *Assuming that the model for the selection probability is correctly specified in the parametric setting based on Equation (3.12) and $E(\frac{\partial U}{\partial \boldsymbol{\alpha}})$ is of full row rank, we have the asymptotic relative efficiency*

$$ARE(T_{GS}(\pi, 0), T_{GSP0}) < 1.$$

Proof. By Equation (3.17) and asymptotic unbiasedness of $U_R(\boldsymbol{\tau}_R)$,

$$\begin{aligned} E\{U_2(\tilde{\boldsymbol{\tau}}_R)\} &= (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r) E\{U_R(\boldsymbol{\tau}_R)\} + \\ &\quad (-\mathbf{A}, \mathbf{I}_r) E\{\frac{\partial U(\boldsymbol{\tau}_R)}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1) \\ &= n^{-\frac{1}{2}} \mathbf{C}\boldsymbol{\lambda} + O(1), \end{aligned}$$

and

$$\begin{aligned} \mathrm{Cov}\{U_2(\tilde{\boldsymbol{\tau}}_R))\} &= (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r) \mathrm{Cov}\{U_R(\boldsymbol{\tau}_R)\}(-\mathbf{A}, \mathbf{B}, \mathbf{I}_r)' + O(n^{\frac{1}{2}}) \\ &= (-\mathbf{A}, \mathbf{B}, \mathbf{I}_r) E(\boldsymbol{\mathcal{J}}_{U_R})(-\mathbf{A}, \mathbf{B}, \mathbf{I}_r) + O(n^{\frac{1}{2}}), \end{aligned}$$

where $\boldsymbol{\lambda}$ is given in Equation and $\mathbf{C}$ in Equation (3.5). Since $\phi = 0$, by Equation (3.16),

$$\mathrm{Cov}\{U_2(\tilde{\boldsymbol{\tau}}_R))\} = (-\tilde{\mathbf{A}}, \mathbf{I}_r) \left\{E(\boldsymbol{\mathcal{J}}_U) - \mathbf{F}\right\} (-\tilde{\mathbf{A}}, \mathbf{I}_r)' + O(n^{\frac{1}{2}}).$$

The noncentrality parameter for T_{GSP0} is

$$\begin{aligned} G_{GSP0} &= \frac{1}{2n} \boldsymbol{\lambda}' \mathbf{C}' [(-\mathbf{A}, \mathbf{I}_r) \left\{E(\boldsymbol{\mathcal{J}}_U) - \mathbf{F}\right\} (-\mathbf{A}, \mathbf{I}_r)']^{-1} \mathbf{C}\boldsymbol{\lambda} \\ &= \frac{1}{2n} \boldsymbol{\lambda}' \mathbf{C}' (\mathbf{K}_J - \mathbf{K}_F)^{-1} \boldsymbol{\lambda}' \mathbf{C}', \end{aligned}$$

where $\mathbf{K}_J = (-\mathbf{A}, \mathbf{I}_r) E(\boldsymbol{\mathcal{J}}_U)(-\mathbf{A}, \mathbf{I}_r)'$ and $\mathbf{K}_F = (-\mathbf{A}, \mathbf{I}_r)\mathbf{F}(-\mathbf{A}, \mathbf{I}_r)'$. Because $E(\frac{\partial U}{\partial \boldsymbol{\alpha}})$ is of full row rank, $\mathbf{K}_F$ is positive definite and hence

$$(\mathbf{K}_J - \mathbf{K}_F)^{-1} = \mathbf{K}_J^{-1} - (\mathbf{K}_J - \mathbf{K}_F)^{-1} \{(\mathbf{K}_J - \mathbf{K}_F)^{-1}) + \mathbf{K}_F)^{-1}\}^{-1} (\mathbf{K}_J - \mathbf{K}_F)^{-1}.$$

Therefore,

$$\begin{aligned} G_{GSP0} &= \frac{1}{2n} \boldsymbol{\lambda}' \mathbf{C}' \mathbf{K}_J^{-1} \mathbf{C}\boldsymbol{\lambda} - \\ &\quad \frac{1}{2n} \boldsymbol{\lambda}' \mathbf{C}' (\mathbf{K}_J - \mathbf{K}_F)^{-1} \{(\mathbf{K}_J - \mathbf{K}_F)^{-1} + \mathbf{K}_F^{-1}\}^{-1} (\mathbf{K}_J - \mathbf{K}_F)^{-1} \mathbf{C}\boldsymbol{\lambda} \\ &= G_{GS0} - C_F, \end{aligned}$$

where G_{GS0} is noncentrality parameter for $T_{GS}(\pi, 0)$ and $C_F = \frac{1}{2n}\boldsymbol{\lambda}'\mathbf{C}'(\mathbf{K}_J - \mathbf{K}_F)^{-1}\{(\mathbf{K}_J - \mathbf{K}_F)^{-1} + \mathbf{K}_F^{-1}\}^{-1}(\mathbf{K}_J - \mathbf{K}_F)^{-1}\mathbf{C}\boldsymbol{\lambda}$. It is clear that $(\mathbf{K}_J - \mathbf{K}_F)^{-1}\{(\mathbf{K}_J - \mathbf{K}_F)^{-1} + \mathbf{K}_F^{-1}\}^{-1}(\mathbf{K}_J - \mathbf{K}_F)^{-1}$ is positive define, and thus $C_F > 0$. Therefore, $G_{GSP0} > G_{GS0}$, completing the proof.

This Lemma indicates that the generalized score test may gain some efficiency if the selection probability is estimated via a correct parametric model even if the true selection probability is given. If we know the correct model for the selection probability, π should be estimated to improve the power of the generalized score test. However, if the model for the selection probability is not correct, then the test would be invalid.

3.3.3 The Case of π Being Known and ϕ^* Being Estimated Parametrically

In a two-stage study, it would be safe to use the true selection probability instead of the estimated one when a correct model for the selection probability is not guaranteed. As we stated before, using the estimated ϕ^* may gain efficiency. In this subsection, we study generalized score tests

$$0 = U_T(\boldsymbol{\beta}, \boldsymbol{\kappa}) = \begin{pmatrix} U_1(\boldsymbol{\beta}, \boldsymbol{\kappa}) \\ U_{\phi^*}(\boldsymbol{\beta}, \boldsymbol{\kappa}) \\ U_2(\boldsymbol{\beta}, \boldsymbol{\kappa}) \end{pmatrix}, \tag{3.18}$$

where π is known. Let $\boldsymbol{\tau}_T = (\boldsymbol{\beta}', \boldsymbol{\kappa}')'$, $\tilde{\boldsymbol{\tau}}_T = (\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\kappa}}')'$ be the solution of the equation under H_0, and $\hat{\phi}^*$ is the estimate of ϕ^*. Since π is given, $U(\boldsymbol{\tau}_T)$ is unbiased and $\tilde{\boldsymbol{\beta}}$ is root-n consistent for $\boldsymbol{\beta}$. In addition, it is easy to see that

$$E\{\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\} = 0.$$

Under the current setting, an appropriate generalized score statistic is $T_{GS}(\pi, \hat{\phi}^*)$. Recall that $T_{GS}(\pi, \phi)$ is an appropriate generalized score statistic when the selection

probability is known and ϕ is given. We have the following theorem:

Theorem 3.3.3. *Assuming that the model for π is given in the parametric setting based on Equation (3.18), under suitable regularity conditions and H_0, $T_{GS}(\pi, \hat{\phi}^*) \to \chi^2_r$ in distribution as $n \to \infty$.*

Proof. Let $U_{T1} = (U_1', U_\phi')'$, $\boldsymbol{\tau}_{T1} = (\boldsymbol{\beta}_1', \boldsymbol{\kappa}')'$ and $\tilde{\boldsymbol{\tau}}_{T1} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\kappa}}')'$. Under H_0, by expanding $U_T(\tilde{\boldsymbol{\tau}}_T)$ at the true value $\boldsymbol{\tau}_T$, we obtain

$$0 = U_{T1}(\tilde{\boldsymbol{\tau}}_T) = U_{T1}(\boldsymbol{\tau}_T) + E(\frac{\partial U_{T1}}{\partial \boldsymbol{\tau}_{T1}})(\tilde{\boldsymbol{\tau}}_{T1} - \boldsymbol{\tau}_{T1}) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\tau}}_T) = U_2(\boldsymbol{\tau}_T) + E(\frac{\partial U_2}{\partial \boldsymbol{\tau}_{T1}})(\tilde{\boldsymbol{\tau}}_{T1} - \boldsymbol{\tau}_{T1}) + O_p(1).$$

Combining the equations above and using the fact that $E\{\frac{\partial U(\boldsymbol{\beta},\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\} = 0$, we have

$$U_2(\tilde{\boldsymbol{\tau}}_T) = \left\{-E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}, \mathbf{0})\boldsymbol{\mathcal{I}}_{T11}^{-1}, \mathbf{I}_r\right\} U_T(\boldsymbol{\tau}_T) + O_p(1), \tag{3.19}$$

where

$$\boldsymbol{\mathcal{I}}_{T11} = \begin{pmatrix} \frac{\partial U_1}{\partial \boldsymbol{\beta}_1} & \mathbf{0} \\ \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1} & \frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}} \end{pmatrix}.$$

By Lemma (3.5.2),

$$\boldsymbol{\mathcal{I}}_{T11}^{-1} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} & 0 \\ * & * \end{pmatrix},$$

where $*$'s represent some constants. Therefore, (3.19) can be rewritten as

$$U_2(\tilde{\boldsymbol{\tau}}_T) = \{-\mathbf{A}, \mathbf{I}_r\} U(\boldsymbol{\tau}_T) + O_p(1).$$

By the Central Limit Theorem and unbiasedness of $U(\boldsymbol{\tau}_T)$, $T_{GS}(\pi, \hat{\phi}^*) \to \chi^2_r$ in distribution as $n \to \infty$. □

In this case, $T_{GS}(\pi, \hat{\phi}^*)$ is an appropriate test statistic. The theorem indicates that plugging the estimate of ϕ^* into WEE(1.1) has no effect on the test statistics when the selection probability is given. Similarly, we can show that $ARE(T_{GS}(\pi, \hat{\phi}^*), T^*_{GS}) \leq 1$. The equation holds when the model for ϕ^* is correctly specified.

3.3.4 The Case of $(y, \mathbf{z}')$ Being Categorical

In this subsection, we consider the special case that $(y, \mathbf{z}')$ is categorical and from a finite set. In this case, the parametric settings above may not work because the logistic model (2.3) for the selection probability is not suitable when $\mathbf{v} = (y, \mathbf{z}')$ is discrete. Without loss of generality, we assume that the first k elements, $\mathbf{v}_i = (y_i, \mathbf{z}_i')$, $i = 1, \cdots, k$, are different from each other. Therefore, these first k elements could be the representatives for all categories. The selection probability can be estimated via

$$\hat{\pi}(\mathbf{v}_i) = \frac{\sum_{j=1}^n \delta_i I(\mathbf{v}_j = \mathbf{v}_i)}{\sum_{j=1}^n I(\mathbf{v}_j = \mathbf{v}_i)}, \tag{3.20}$$

and the conditional mean score ϕ^* by

$$\hat{\phi}^*(\mathbf{v}_i) = \frac{\sum_{j=1}^n \delta_i \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\boldsymbol{\beta}}) I(\mathbf{v}_j = \mathbf{v}_i)}{\sum_{j=1}^n \delta_i I(\mathbf{v}_j = \mathbf{v}_i)}, \tag{3.21}$$

where $\tilde{\boldsymbol{\beta}}$ is solution of WEE (1.1) under H_0. Because the number of categories is finite, $\hat{\pi}$ and $\hat{\phi}^*$ are root-n consistent for π and ϕ^*, respectively. Then we have

$$\begin{aligned} U(\boldsymbol{\beta}, \hat{\pi}, \hat{\phi}^*) &= U(\boldsymbol{\beta}, \pi, \hat{\phi}^*) - \sum_{i=1}^n \{ \frac{\delta_i(\hat{\pi}_i - \pi_i)}{\hat{\pi}_i \pi_i} (\psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\boldsymbol{\beta}}) - \hat{\phi}^*) \} \\ &= U(\boldsymbol{\beta}, \pi, \hat{\phi}^*) \\ &\quad + \sum_{i=1}^k \left[\frac{(\hat{\pi}_i - \pi_i)}{\hat{\pi}_i \pi_i} \{ \frac{\sum_{j=1}^n \delta_i \Delta_i(\psi) I(\mathbf{v}_j = \mathbf{v}_i)}{\sum_{j=1}^n \delta_i I(\mathbf{v}_j = \mathbf{v}_i)} \sum_{j=1}^n I(\mathbf{v}_j = \mathbf{v}_i) \} \right] \\ &= U(\boldsymbol{\beta}, \pi, \hat{\phi}^*) + O_p(1), \end{aligned} \tag{3.22}$$

where $\Delta_i(\psi) = \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) - \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\boldsymbol{\beta}})$. By (3.22), it is clear that $\tilde{\boldsymbol{\beta}}$ is root-n consistent. Expanding $U(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*)$ at $\boldsymbol{\beta}$ under H_0, we have

$$0 = U_1(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = U_1(\boldsymbol{\beta}, \hat{\pi}, \hat{\phi}^*) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1),$$

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = U_2(\boldsymbol{\beta}, \hat{\pi}, \hat{\phi}^*) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1).$$

Combining the equations above and by Equation (3.22), we have

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = (-\mathbf{A}, \mathbf{I}_r) U(\boldsymbol{\beta}, \pi, \hat{\phi}^*) + O_p(1).$$

Therefore, we have the following lemma:

Lemma 3.3.5. *If $(y, \mathbf{z}')$ is categorical and from a finite set, under regularity conditions and H_0, $T_{GS}(\hat{\pi}, \hat{\phi}^*) \to \chi^2_r$ in distribution as $n \to \infty$.*

We would like to study the asymptotic efficiency of the proposed test when $(y, \mathbf{z}')$ is categorical. Under the sequence of local alternatives $\boldsymbol{\beta}_2^{(n)}$ in (3.4), we expend $U_1(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*)$ and $U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*)$ at $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2^{(n)})$:

$$0 = U_1(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = U_1(\boldsymbol{\beta}, \hat{\pi}, \hat{\phi}^*) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1)$$

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = U_2(\boldsymbol{\beta}, \hat{\pi}, \hat{\phi}^*) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1).$$

By Equation (3.22), we have

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*) = (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \pi, \hat{\phi}^*) + n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda} + O_p(1).$$

Therefore,

$$E\{U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*)\} = n^{-\frac{1}{2}}\mathbf{C}\boldsymbol{\lambda} + O(1),$$

and

$$\mathrm{Cov}\{n^{-\frac{1}{2}}U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}, \hat{\phi}^*)\} = \frac{\boldsymbol{\mathcal{J}}^*}{n} + O(n^{-\frac{1}{2}}),$$

since $\hat{\phi}^*$ is root-n consistent. This implies that the noncentrality parameter is equivalent to G^* based on Equation (3.5) at $\phi = \phi^*$. Therefore, we have the following lemma:

Lemma 3.3.6. *If $(y, \mathbf{z}')$ is categorical,*

$$ARE(T_{GS}(\hat{\pi}, \hat{\phi}^*), T^*_{GS}) = 1.$$

When $(y, \mathbf{z}')$ is categorical and from a finite set, the proposed generalized score test obtain the optimal power asymptotically if the selection probability is estimated via (3.20) and ϕ^* is estimated via (3.21). If the $(y, \mathbf{z}')$ is discrete and from an infinite set, both Lemma 3.3.5 and 3.3.6 may be invalid because $\frac{(\hat{\pi}_i - \pi_i)}{\hat{\pi}_i \pi_i}$ may be unbounded.

3.3.5 Discussion

When the model for the selection probability is correctly specified, T^*_{GSP} is an appropriate generalized score statistic, which is free of U_{ϕ^*} and $\frac{\partial U}{\partial \boldsymbol{\kappa}}$ given the estimate of ϕ^*. Therefore, the generalized score statistic is easy to calculate in this case. If the parametric model for ϕ^* is also correctly specified, $T_{GS}(\hat{\pi}, \hat{\phi})$ is an appropriate generalized score statistic and $ARE(T_{GS}(\hat{\pi}, \hat{\phi}), T^*_{GS}) = 1$. Therefore, we may use the difference between T^*_{GSP} and $T_{GS}(\hat{\pi}, \hat{\phi})$ to informally check whether the model for ϕ^* is correct or not. If the model for the selection probability is not correct while the model for ϕ^* is correctly specified, an appropriate generalized score statistic generally depends on U_{ϕ^*} and $\frac{\partial U}{\partial \boldsymbol{\kappa}}$, which are extremely difficult to obtain; it is not feasible to use the generalized score test in this case. Moreover, $ARE(T_{GS}(\pi, 0), T_{GSP0}) < 1$ when $\phi = 0$. This indicates that the tests may gain some efficiency if the selection probability is estimated via a correct parametric model even if the true selection probability is given.

3.4 Semiparametric Setting

It is generally convenient to assume parametric models for the selection probability and the mean score function ϕ^*. However, it might be problematic when the parametric models (especially the model for the selection probability) are not correct. To deal with this problem, we may alternatively estimate the nuisance functions nonparametrically. Assume that the selection probability may be estimated by (2.5)

$$\hat{\pi}_N(\mathbf{v}) = \frac{\sum_{i=1}^n \delta_i K_h(\mathbf{v} - \mathbf{v}_i)}{\sum_{i=1}^n K_h(\mathbf{v} - \mathbf{v}_i)},$$

and ϕ^* by (2.6)

$$\hat{\phi}^*_N(\mathbf{v}) = \frac{\sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_N(\mathbf{v}_i)} \tilde{\psi}_i K_h(\mathbf{v} - \mathbf{v}_i)}{\sum_{i=1}^n K_h(\mathbf{v} - \mathbf{v}_i)}.$$

where $\tilde{\boldsymbol{\beta}}$ is the solution of $U(\boldsymbol{\beta}, \hat{\pi}_N, \phi) = 0$ under H_0 and $\tilde{\psi}_i = \psi(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\boldsymbol{\beta}})$. Recall that K is an sth-order kernel function, h is a proper bandwidth parameter, $K_h(\cdot) = K(\cdot/h)$, $\rho_n = \left\{nh^{2s} + (nh^{2d})^{-1}\right\}^{\frac{1}{2}}$, and $\mathbf{v}_i = (y_i, \mathbf{z}_i')'$. Let d be the number of continuous components of $\mathbf{v}_i$. As a concrete example, if y_i is binary, $\mathbf{z}_i$ is univariate and continuous, and K is a 2nd-order kernel function, then $d = 1$, $s = 2$ and the optimal bandwidth is $h = O(n^{-\frac{1}{3}})$. In this situation, $\rho_n = O_p(n^{-\frac{1}{6}})$.

Let the selection probability be estimated via (2.5). Define

$$T_{GSN} = U_2'(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)\tilde{\mathbf{V}}_{GS}^{-1}U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) \tag{3.23}$$

for both $\phi = 0$ and $\phi = \hat{\phi}_N^*$, where $\tilde{\mathbf{V}}_{GS} = (-\tilde{\mathbf{A}}, \mathbf{I}_r)\hat{\Sigma}_N(-\tilde{\mathbf{A}}, \mathbf{I}_r)'$, $\tilde{\mathbf{A}} = \mathbf{A}|_{(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)}$, and $\hat{\Sigma}_N$ is any consistent estimate of $\mathrm{Cov}\{U(\boldsymbol{\beta}, \pi, \phi^*)\}$ with a converge rate not slower than $O_p(n\rho_n)$. Here, without confusion, we continue to use $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\mathcal{J}}}_U$, etc. for the quantities evaluated at properly estimated parameters.

By Equation (2.7) and Equation (2.2),

$$\begin{aligned}
\mathrm{Cov}\{U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)\} &= \mathrm{Cov}\{U_2(\tilde{\boldsymbol{\beta}}, \pi, \phi) + O_p(n^{-\frac{1}{2}}\rho_n)\} \\
&= \mathrm{Cov}\{U^F(\boldsymbol{\beta})\} + \mathrm{Cov}\{U^M(\boldsymbol{\beta}, \pi, \phi)\} + O(n\rho_n)
\end{aligned}$$

for both $\phi = 0$ and $\phi = \hat{\phi}_N^*$. Therefore, one possible choice of $\hat{\Sigma}_N$ is

$$\hat{\Sigma}_{N1} = \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_N^2(\mathbf{v}_i)} \left\{\tilde{\psi}_i - \hat{\phi}_N^*(\mathbf{v}_i)\right\} \left\{\tilde{\psi}_i - \hat{\phi}_N^*(\mathbf{v}_i)\right\}' + \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_N(\mathbf{v}_i)} \hat{\phi}_N^*(\mathbf{v}_i)\hat{\phi}_N^*(\mathbf{v}_i)'.$$

When $\phi = \hat{\phi}_N^*$, another possible choice of $\hat{\Sigma}_N$ is $\tilde{\boldsymbol{\mathcal{J}}}_U$, since asymptotic unbiasedness of $U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)$. It is easy to see that one version of T_{GSN} is

$$T_{GS}(\hat{\pi}_N, \hat{\phi}_N^*) = U_2'(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \hat{\phi}_N^*) \left\{(-\tilde{\mathbf{A}}, \mathbf{I}_r)\tilde{\boldsymbol{\mathcal{J}}}_U(-\tilde{\mathbf{A}}, \mathbf{I}_r)'\right\}^{-1} U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \hat{\phi}_N^*).$$

This implies $T_{GS}(\hat{\pi}_N, \hat{\phi}_N^*)$ is an appropriate generalized score test statistic when both $\hat{\pi}_N$ and $\hat{\phi}_N^*$ are proper nonparametric estimates.

Theorem 3.4.1. *Assume that the bandwidths h in (2.5) and (2.6) satisfy $nh^{2s} \to 0$ and $nh^{2d} \to \infty$. Under H_0 and suitable regularity conditions, $T_{GSN} \to \chi^2_r$ in distribution as $n \to \infty$.*

Proof. Since $\tilde{\boldsymbol{\beta}}$ is root-n consistent, by expanding $U_1(\tilde{\boldsymbol{\beta}})$ and $U_2(\tilde{\boldsymbol{\beta}})$ at the true value $\boldsymbol{\beta}$ and replacing $\frac{\partial U_1}{\partial \boldsymbol{\beta}_1}$ and $\frac{\partial U_2}{\partial \boldsymbol{\beta}_1}$ by their asymptotically equivalent versions $E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})$ and $E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})$ under H_0, we obtain

$$0 = U_1(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1),$$
$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = U_2(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + O_p(1),$$

and thus

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + O_p(1)$$

for any fixed ϕ. By Equation (2.7), when ϕ is either 0 or $\hat{\phi}^*_N$, $U(\boldsymbol{\beta}, \hat{\pi}_N, \phi)$ and $U(\boldsymbol{\beta}, \pi, \phi^*)$ are asymptotically equivalent. Therefore,

$$n^{-\frac{1}{2}}U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = (-\mathbf{A}, \mathbf{I}_r)\{n^{-\frac{1}{2}}U(\boldsymbol{\beta}, \pi, \phi^*)\} + O_p(\rho_n).$$

Since $U(\boldsymbol{\beta}, \pi, \phi^*)$ is an unbiased estimating equation and $\tilde{\mathbf{A}}$ is root-n consistent,

$$E\{n^{-\frac{1}{2}}U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)\} = O(\rho_n),$$

and

$$\begin{aligned}\mathrm{Cov}\{n^{-\frac{1}{2}}U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)\} &= (-\mathbf{A}, \mathbf{I}_r)\mathrm{Cov}\{n^{-\frac{1}{2}}U(\boldsymbol{\beta}, \pi, \phi^*)\}(-\mathbf{A}, \mathbf{I}_r)' + O(\rho_n) \\ &= (-\tilde{\mathbf{A}}, \mathbf{I}_r)\mathrm{Cov}\{n^{-\frac{1}{2}}U(\boldsymbol{\beta}, \pi, \phi^*)\}(-\tilde{\mathbf{A}}, \mathbf{I}_r)' + O(\rho_n).\end{aligned}$$

It is clear that $\rho_n \to 0$ because $nh^{2s} \to 0$ and $nh^{2d} \to \infty$. It is readily shown that $T_{GSN} \to \chi^2_r$ in distribution as $n \to \infty$ under H_0 by the Central Limit Theorem. □

In addition, we would like to study the asymptotic efficiency of the proposed semiparametric test. First we need the following lemma:

Lemma 3.4.1. *Under the local alternative, the semiparametric constrained estimate* $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20})'$ *is root-n consistent for* $\boldsymbol{\beta}$.

Proof. Let $\tilde{\boldsymbol{\beta}}_1^{(n)}$ be the solution of the estimating equation such that

$$U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2^{(n)}, \hat{\pi}_N, \phi) = \mathbf{0}.$$

By Equation (2.7) and the fact that $\boldsymbol{\beta}_2^{(n)}$ is the true value of $\boldsymbol{\beta}_2$, $\tilde{\boldsymbol{\beta}}_1^{(n)}$ is root-n consistent for the $\boldsymbol{\beta}_1$.

Under the local alternative,

$$\begin{aligned}
\mathbf{0} &= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2^{(n)}, \hat{\pi}_N, \phi) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2, \hat{\pi}_N, \phi) + \frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_2}(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20}) + O_p(1) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)}, \boldsymbol{\beta}_2, \hat{\pi}_N, \phi) + \\
&\quad \frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_1}\left[\{\frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_1}\}^{-1}\frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_2}(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20})\right] + O_p(1) \\
&= U_1(\tilde{\boldsymbol{\beta}}_1^{(n)} + \Delta_{\boldsymbol{\beta}}^{(n)}, \boldsymbol{\beta}_2, \hat{\pi}_N, \phi) + O_p(1),
\end{aligned}$$

where $\Delta_{\boldsymbol{\beta}}^{(n)} = \{\frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_1}\}^{-1}\frac{\partial U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi)}{\partial \boldsymbol{\beta}_2}(\boldsymbol{\beta}_2^{(n)} - \boldsymbol{\beta}_{20})$. Therefore,

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_1 &= \tilde{\boldsymbol{\beta}}_1^{(n)} + \Delta_{\boldsymbol{\beta}}^{(n)} + O_p(n^{-\frac{1}{2}}) \\
&= \tilde{\boldsymbol{\beta}}_1^{(n)} + O_p(n^{-\frac{1}{2}}),
\end{aligned}$$

and thus the constrained estimate under the local alternative is root-n consistent for $\boldsymbol{\beta}$. □

Since the constrained estimate is consistent for $\boldsymbol{\beta}$ under the sequence of local alternatives $\boldsymbol{\beta}_2^{(n)}$ in (3.4), by expanding $U(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi)$ at the true value $\boldsymbol{\beta}$, we obtain

$$0 = U_1(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = U_1(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$

$$U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) = U_2(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_1})(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + E(\frac{\partial U_2}{\partial \boldsymbol{\beta}_2})(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$

and hence

$$\begin{aligned} U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) &= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \hat{\pi}_N, \phi) + \\ &\quad (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1). \end{aligned}$$

By Equation (2.7),

$$\begin{aligned} U_2(\tilde{\boldsymbol{\beta}}, \hat{\pi}_N, \phi) &= (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \pi, \phi) + \\ &\quad (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(n^{-\frac{1}{2}}\rho_n). \end{aligned}$$

By the equation above, it is easy to obtain the noncentrality parameters for T_{GSN} and results of relative efficiency.

Lemma 3.4.2. *If the models for π and ϕ^* are correctly specified in the parametric setting using joint estimating equation (3.6), then*

$$ARE(T_{GSN}, T_{GS}^*) = 1.$$

Proof. Since both the models for π and ϕ^* are correct, $\mathbf{B} = 0$ and

$$U(\boldsymbol{\tau}) = U(\boldsymbol{\beta}, \pi, \phi^*),$$

Therefore, Equation (3.11) reduces to

$$U_2(\tilde{\boldsymbol{\tau}}) = (-\mathbf{A}, \mathbf{I}_r)U(\boldsymbol{\beta}, \pi, \phi^*) + (-\mathbf{A}, \mathbf{I}_r)E\{\frac{\partial U(\boldsymbol{\tau})}{\partial \boldsymbol{\beta}_2}\}(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2^{(n)}) + O_p(1),$$

which is asymptotically equivalent to $U_2(\tilde{\boldsymbol{\beta}})$ under the local alternative when π is known and ϕ^* is given. Therefore, $ARE(T_{GSP}^*, T_{GS}*) = 1$. □

Lemma 3.4.2 implies that the optimal power can be obtained asymptotically by using an appropriate nonparametric estimate of π and ϕ in WEE (1.1). However, it is easy to see that T_{GSN} converges to a χ_r^2 distribution with a rate of $O_p(\rho_n)$, which is

generally slower than that in the parametric setting. When the sample size is small and an appropriate model for π is available, we should use the parametric generalized score tests rather than the semiparametric generalized score test. On the other hand, when the sample size is reasonably large, the semiparametric tests are often preferred to the parametric tests because the semiparametric tests would obtain the optimal power asymptotically and there is no worry about the model misspecification for the selection probability.

3.5 Technical Detail

Lemma 3.5.1. *If all necessary inverses exist, then for matrices* $\mathbf{Q}_a(p\times p)$, $\mathbf{Q}_b(p\times n)$, $\mathbf{Q}_c(n\times n)$, *and* $\mathbf{Q}_d(n\times p)$,

$$(\mathbf{Q}_a+\mathbf{Q}_b\mathbf{Q}_c\mathbf{Q}_d)^{-1}=\mathbf{Q}_a^{-1}+\mathbf{Q}_a^{-1}\mathbf{Q}_b(\mathbf{Q}_c+\mathbf{Q}_d\mathbf{Q}_a^{-1}\mathbf{Q}_b)^{-1}\mathbf{Q}_d\mathbf{Q}_a^{-1}.$$

Proof. See Mardia, Kent, and Bibby (1979), (page 458). □

Lemma 3.5.2. *Given that the* 2×2 *block matrix is nonsingular, the inverse matrix*

$$\begin{bmatrix}\mathbf{Q}_{11} & \mathbf{Q}_{12}\\ \mathbf{Q}_{21} & \mathbf{Q}_{22}\end{bmatrix}^{-1}=\begin{bmatrix}\mathbf{Q}_{11}^{-1}+\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{S}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} & -\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}\mathbf{S}^{-1}\\ -\mathbf{S}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} & \mathbf{S}^{-1}\end{bmatrix},$$

where the quantity $\mathbf{S}=\mathbf{Q}_{22}-\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}$ *is called the Schur complement of* $\mathbf{Q}_{11}$.

Proof. It can be verified directly by checking that the product of the matrix and its inverse reduces to the identity matrix. □

Proof of Equation (3.10). We have the following partition of the matrix

$$\boldsymbol{\mathcal{I}}_{J11}=\begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{0} & E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})\end{pmatrix},$$

where

$$\mathbf{H}_{11} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1}) & 0 \\ E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1}) & E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}}) \end{pmatrix} \text{ and } \mathbf{H}_{12} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\alpha}}) \\ E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\alpha}}) \end{pmatrix}.$$

Using Lemma 3.5.2 twice, we have

$$\boldsymbol{\mathcal{I}}_{J11}^{-1} = \begin{pmatrix} \mathbf{H}_{11}^{-1} & -\mathbf{H}_{11}^{-1}\mathbf{H}_{12}E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \\ 0 & E(\frac{\partial U_\pi}{\partial \boldsymbol{\alpha}})^{-1} \end{pmatrix},$$

where

$$\mathbf{H}_{11}^{-1} = \begin{pmatrix} E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} & 0 \\ -E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}})^{-1}E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\beta}_1})E(\frac{\partial U_1}{\partial \boldsymbol{\beta}_1})^{-1} & E(\frac{\partial U_{\phi^*}}{\partial \boldsymbol{\kappa}})^{-1} \end{pmatrix}.$$

Plugging $\mathbf{H}_{11}^{-1}$ into $\boldsymbol{\mathcal{I}}_{J11}^{-1}$ leads to (3.10). □

CHAPTER IV

GOODNESS OF FIT TESTS FOR GENERALIZED LINEAR MODELS WHEN SOME COVARIATES ARE PARTIALLY MISSING

4.1 Introduction

First introduced by Nelder and Wedderburn (1972), generalized linear models provide a unified approach for a broad class of regression models in applied statistics. They are designed for applications with independent observations having a density:

$$f(y_i|\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}_1, \varsigma) = \exp\left\{\frac{\theta_i y_i - b(y_i, \varsigma)}{a_i(\varsigma)} + c(y, \varsigma)\right\} \tag{4.1}$$

where θ_i is known as the canonical parameter and ς is a scale parameter. The functions $a_i(\varsigma)$ are commonly of the form $a_i(\varsigma) = \frac{\varsigma}{k_i}$, where k_i's are known weights. In addition, the p-dimensional covariate $\mathbf{w}_i = (1, \mathbf{x}_i', \mathbf{z}_i')'$ is related to θ_i through the link function $\eta_i = l(\theta_i)$ and the linear component

$$\eta_i = \beta_0 + \mathbf{x}_i\boldsymbol{\beta}_x + \mathbf{z}_i\boldsymbol{\beta}_z = \mathbf{w}_i\boldsymbol{\beta}_1,$$

where $\boldsymbol{\beta}_1 = (\beta_0, \boldsymbol{\beta}_x', \boldsymbol{\beta}_z')'$ is a $q \times 1$ vector of regression coefficients and g is a monotone differentiable function. See McCullagh and Nelder (1989) for more details about the generalized linear models.

The likelihood equations for $\boldsymbol{\beta}$ are

$$\sum_{i=1}^{n} k_i(\frac{d\theta_i}{\eta_i})\mathbf{w}_i'\{y_i - l^{-1}(\mathbf{w}_i\boldsymbol{\beta}_1)\} = 0,$$

which are generally nonlinear in $\boldsymbol{\beta}$. The parameters can be solved via iteratively reweighted least squares (IRLS) algorithm. An important component of any modelling procedure is an assessment of model fit, which evaluates how well model-based

predicted outcomes coincide with the observed data. For fully observed data, the scaled deviance and Pearson's chi-square statistic are helpful in assessing the goodness of fit of a given generalized linear model. However, these methods may be problematic when some covariates are partially missing. In this chapter, we propose an formal model validation procedure for the generalized linear models in the presence of missing covariates.

4.2 Goodness of Fit Test

We focus on testing the linearity of primary regression model (4.1). In general, the hypotheses are

$$H_0\text{: } \eta_i = \mathbf{w}_i\boldsymbol{\beta}_1 \text{ vs } H_a\text{: } \eta_i \neq \mathbf{w}_i\boldsymbol{\beta}_1. \tag{4.2}$$

Strictly speaking, the alternative depends on the the situations of applications and settings. The rejection of the null hypothesis implies several possibilities: (a) misspecification of the primary regression model, including the linear component and the link function l; (b) violation in the MAR assumption, or (c) a model misspecification for the selection probability. It is possible to test (b) and (c) using the methods proposed by Lipsitz et al. (2001) or a global test statistic for model (2.3). Though we do not detect the misspecification of the link function directly, a misspecification in the the link function will reflect as a misspecification in the linear component.

Let $M_{(0)}$ be the model under H_0. We consider the alternative model $M_{(r)}$ with r more parameters than $M_{(0)}$. More specifically, model $M_{(r)}$ has a linear component

$$\eta_i = \mathbf{w}_i\boldsymbol{\beta}_1 + \sum_{j=1}^{r} f_j(\mathbf{w}_i)\boldsymbol{\beta}_{q+j} = \mathbf{w}_i\boldsymbol{\beta}_1 + \mathbf{u}_{ri}\boldsymbol{\beta}_{2(r)}, \tag{4.3}$$

where the vector of parameters $\boldsymbol{\beta}_{2(r)} = (\beta_{q+1}, \cdots, \beta_{q+r})'$, $\mathcal{F} = \{f_1, f_2, \cdots\}$ is a sequence of $\Re^q \to \Re$ mutually linear independent functions, and $\mathbf{u}_{ri} = (f_1(\mathbf{w}_i), \cdots, f_r(\mathbf{w}_i))'$, which is the r-dimensional supplement covariate and may contain missing values if

$\delta_i = 0$. Obviously, model $M_{(i)}$ is nested in model $M_{(j)}$ if $i < j$. Theoretically, with a proper choice of the sequence $\mathcal{F}$, such as a certain complete basis in the covariate space, model $M_{(r)}$ eventually includes any alternative of interest as $r \to \infty$. If the alternative is (a) a partially linear (or single index) model

$$H_a\text{: } \eta = \beta_0 + x\boldsymbol{\beta}_x + v_z(z),$$

where v_z is a continuous function; or (b) the model $M_{(0)}$ only has one univariate covariate, then $\mathcal{F}$ could be orthonormal polynomials or the cosine system. Orthonormal polynomials of order greater than 2 may be easily computed recursively by using the Emerson recurrence formula (Emerson, 1968). Otherwise, both orthonormal polynomials and the cosine system in the high dimensional space are too complex to be suitable. Therefore, in general we suggest generating supplement covariates based on space partitioning described in Barnhart and Williamson (1998). First the covariate space is partitioned into $(r+1)$ distinct regions, and then define the $r \times 1$ supplement covariates $\mathbf{u}_{ri} = \{I_{i1}, \cdots, I_{ir}\}$, where $I_{im} = 1$, $m = 1, 2, \cdots, r$, if $\mathbf{w}_i$ is in the mth region, 0 if not. If $\delta_i = 0$, it might be impossible to determine whether $\mathbf{w}_i$ is in the mth region, then the value I_{im} is missing.

If we are interested in testing certain types of departures from H_0, such as a low frequency departure, 2-way interaction, etc, it is not difficult to specify a number R with a proper choice of the sequence $\mathcal{F}$, such that the model $M_{(R)}$ approximately captures the departure. Then we can use the generalized score statistic to detect the departure by testing $\boldsymbol{\beta}_{2(R)} = 0$ with model $M_{(R)}$ being a plausible alternative. The power of the test depends on the plausible alternative in two aspects. First, the power of the test depends on how much the true regression function can be approximated by model $M_{(R)}$. If $M_{(R)}$ could not capture the departure well, such as a departure orthogonal to the space spanned by the linear component of $M_{(R)}$, then the test has

low power even the departure is very strong. It is more likely the $M_{(R)}$ well capture the departure when the R is getting larger. A larger R is desired in this sense. On the other hand, under H_0 and the selection probability is correctly specified, the test statistic follow a χ^2_R distribution. It is clear that critical value increases when R increases. As a consequence, the test would lose sensitivity to low frequency departures when R is getting larger. Therefore, we can improve the power of the test if we use a smaller model $M_{(r)}$, $r \leq R$, which captures the departure well, as the plausible alternative. In the next section, we will introduce data driven methods to find an optimal plausible alternative model in the sequence alternatives.

4.3 Data Driven Methods

Model $M_{(R)}$ is often not optimal for the goodness of fit test. It is possible that a nested model $M_{(r)}$, $r \leq R$, is better than $M_{(R)}$ alone for testing. We use the following WEEs to obtain $T_{GS}(r)$ for testing $\boldsymbol{\beta}_{2(r)} = 0$ $(r = 1, 2, \cdots, R)$:

$$U_{(r)}(\boldsymbol{\beta}_{(r)}, \pi, \phi) = \sum_{i=1}^{n} \{ \frac{\delta_i k_i}{\pi_i} (\frac{d\theta_i}{\eta_i}) (\mathbf{w}_i', \mathbf{u}_{ri}')' (y_i - l^{-1}(\mathbf{w}_i \boldsymbol{\beta}_1 + \mathbf{u}_{ri} \boldsymbol{\beta}_{2(r)})) + (1 - \frac{\delta_i}{\pi_i}) \phi(y_i, \mathbf{z}_i) \}, \tag{4.4}$$

where $\boldsymbol{\beta}_{(r)} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_{2(r)}')'$. Note that we let ϕ in Equation (4.4) to be independent of r because (a) the observed supplement covariates are transformations of $\mathbf{z}_i$, and (b) we would like to see the role of supplement covariates in the regression models instead of their effects on the mean score functions. One way to automatically choose r is to use a data-driven method following the idea in Aerts et al., (1999, 2000). It is to choose $r = \hat{r}$ that maximizes the penalized score criterion

$$SIC(r) = T_{GS}(r) - 2r, \ (r = 1, \cdots, R)$$

and the data driven test statistic is

$$T_{MGS}(R) = SIC(\hat{r}).$$

We know that the generalized score statistics defined in the previous chapter all have the form $\tilde{U}_2\tilde{\mathbf{V}}^{-1}\tilde{U}_2$, where $\tilde{U}_2$ and $\tilde{\mathbf{V}}$ are the vector and matrix corresponding to each of the three settings considered in (3.3), (3.7) and (3.23). Similarly, for each generalized score statistic $T_{GS}(r)$ above, we have $T_{GS}(r) = \tilde{U}_{2(r)}\tilde{\mathbf{V}}_{(r)}^{-1}\tilde{U}'_{2(r)}$, whose components are indexed by r. To introduce the asymptotic distribution of $T_{MGS}(R)$, we define $\Gamma_R = max_{1\le r\le R}(\sum_{k=1}^{r} Z_k^2 - 2r)$, where Z_i's are a sequence of independent and identically distributed standard norm random variables. We used $200,000$ runs simulation for the critical values of the random variable Γ_R. The critical value is 3.57 when $R = 5$ for a 0.05 significant level test.

Theorem 4.3.1. *Assume that the selection probability in WEE (4.4) is either known, appropriately estimated via a correct parametric model or estimated nonparametrically via (2.5). Under H_0 and regularity conditions, $T_{MGS}(R)$ converges to Γ_R in distribution.*

Proof. Assuming that $\tilde{\boldsymbol{\beta}}_{(r)}$ is the solution of $U_{(r)}(\boldsymbol{\beta}_{(r)}, \pi, \phi) = 0$ under H_0 for $r = 1, \cdots, R$, it is clear that $\tilde{\boldsymbol{\beta}}_{(r+1)} = (\tilde{\boldsymbol{\beta}}'_{(r)}, 0)'$. Plugging $\tilde{\boldsymbol{\beta}}_{(r)}$ into $U_{(r)}(\boldsymbol{\beta}_{(r)}, \pi, \phi)$, for $r = 1, \cdots, R$, respectively, it is seen that (a) $\tilde{U}_{2(r)}$ is the first $r \times 1$ subvector of $\tilde{U}_{2(R)}$, and (b) $\tilde{\mathbf{V}}_{(r)}$ is the upper $r \times r$ submatix of $\tilde{\mathbf{V}}_{(R)}$, for $r = 1, \cdots, R$.

Write

$$\tilde{U}_{2(R)} = \begin{pmatrix} \tilde{U}_{2(R-1)} \\ Q_R \end{pmatrix}$$

and

$$\tilde{\mathbf{V}}_{(R)} = \begin{pmatrix} \tilde{\mathbf{V}}_{(R-1)} & \mathbf{a}_{(R)}^{(n)} \\ \mathbf{a}_{(R)}^{\prime(n)} & d_{(R)}^{(n)} \end{pmatrix}.$$

By Lemma 3.5.2, we have

$$\begin{pmatrix} \tilde{\mathbf{V}}_{(R-1)} & \mathbf{a}_{(R)}^{(n)} \\ \mathbf{a}_{(R)}^{\prime(n)} & d_{(R)}^{(n)} \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{\mathbf{V}}_{(R-1)}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{c_{(R)}^{(n)}} \begin{pmatrix} \tilde{\mathbf{V}}_{(R-1)}^{-1}\mathbf{a}_{(R)}^{(n)}\mathbf{a}_{(R)}^{\prime(n)}\tilde{\mathbf{V}}_{(R-1)}^{-1} & -\tilde{\mathbf{V}}_{(R-1)}^{-1}\mathbf{a}_{(R)}^{(n)} \\ -\mathbf{a}_{(R)}^{\prime(n)}\tilde{\mathbf{V}}_{(R-1)}^{-1} & 1 \end{pmatrix},$$

where

$$c_{(R)}^{(n)} = d_{(R)}^{(n)} - \mathbf{a}_{(R)}^{\prime(n)}\tilde{\mathbf{V}}_{(R-1)}^{-1}\mathbf{a}_{(R)}^{(n)}.$$

Obviously,

$$T_{GS}(R) = T_{GS}(R-1) + \frac{1}{c_{(R)}^{(n)}} W_{(R)}^2,$$

where

$$W_{(R)} = Q_R - \mathbf{a}_{(R)}^{\prime(n)}\tilde{\mathbf{V}}_{(R-1)}^{-1}\tilde{U}_{2(R-1)}.$$

It is easy to see that

$$\mathrm{Cov}\{W_{(R)}, \tilde{U}_{2(R-1)}\}/n \to 0$$

and

$$\frac{1}{c_{(R)}^{(n)}} W_{(R)}^2 \to \chi_1^2 \text{ in distribution.}$$

Since $(\tilde{U}_{2(R-1)}^{\prime}, W_{(R)})^{\prime}$ is asymptotically normal, $\tilde{U}_{2(R-1)}$ and $W_{(R)}$ are asymptotically independent. This implies that $T_{GS}(R-1)$ and $T_{GS}(R)-T_{GS}(R-1)$ are asymptotically independent. Using the same argument iteratively, one can show that

$$T_{GS}(1), T_{GS}(2) - T_{GS}(1), \cdots, T_{GS}(R) - T_{GS}(R-1)$$

are all asymptotically χ_1^2 distributed and asymptotically independent of each other. Since $T_{MGS}(R)$ is a continuous function of $T_{GS}(r)$ for $r = 1, \cdots, R$, it converges to Γ_R in distribution. □

CHAPTER V

SIMULATION STUDIES

5.1 Introduction

Simulation studies were performed to investigate the finite sample properties of the proposed tests, by assessing the adequacy of the asymptotic null distribution of the test statistics and the power to detect unknown primary model misspecifications.

The generalized score statistics used in the simulations are (a) F_{GS} for full data; (b) CC using complete cases only; (c) T_{GS} using the known selection probability and $\phi = 0$; (d) T_{GSP}, when the selection probability is estimated by the correct parametric model (2.3) and $\phi = 0$; (e) T_{GSN}, when the selection probability is estimated non-parametrically via (2.5) and $\phi = 0$; and (f) T^*_{GSP}, when both the selection probability and $\phi = \phi^*$ are estimated via correct parametric models. Note that Monte Carlo methods are typically used for estimating the mean score function (see Lipsitz et al., 1999) in Equation (3.6) when $\mathbf{x}_i$ contains continuous components. We implemented T^*_{GSP} in the table on page 65 only, where the mean score function can be estimated by a direct calculation of the conditional expectation without relying on Monte Carlo methods since $\mathbf{x}_i$ is univariate and binary. In each part of the simulation study, we used the triangular system $\mathcal{F} = \{cos(2\pi t), sin(2\pi t), cos(4\pi t), sin(4\pi t), \cdots\}$ for $t \in (0, 1)$ to generate a plausible alternative $M_{(R)}$ with $R = 5$ supplement covariates. We investigated the numerical performance for different choices of R. Our limited numerical experience suggests that $R = 5$ is sufficient for low frequency alternatives. For high frequency alternatives, larger R values would be desired. We used 1,000

simulation replications for each case considered below. Under H_0 and the significance level $\alpha = 5\%$, a correct rejection ratio should be around 5% with a Monte Carlo error of $\sqrt{0.05 \times (1 - 0.05)/1000} = 0.007$. Under an alterative, a rejection ratio is reported with a Monto Carlo error no large than $\sqrt{0.5 \times (1 - 0.5)/1000} = 0.016$. When computing T_{GSN} in our simulation study, π and ϕ^* were estimated via a local linear estimate (by the *locfit* function in R), which is equivalent and possibly more stable than the kernel estimation in (2.5) and (2.6), with the bandwidth $h = d(100/n)^{-\frac{1}{3}}$, where d is a constant. We experimented with different d from $[0.5, 1.2]$. The results are stable, so we chose $d = 0.7$.

5.2 General Linear Models

In this section, we study the power and type I error of the proposed score tests for testing adequacy of a general linear model with one or two covariates. We first stimulated data from the following model with possibly missing univariate covariate x_i and the response variable y_i:

$$y_i = \beta_0 + \beta_x x_i + c x_i^2 + e_i, \tag{5.1}$$

where e_i's are independent error terms. The missingness indicator δ_i follows the logistic regression model

$$\text{logit}\{\Pr(\delta_i = 1|y_i)\} = \alpha_0 + \alpha_1 y_i^{\frac{1}{3}}. \tag{5.2}$$

The hypotheses are

$$H_0\text{: } E(y_i|x_i) = \beta_0 + \beta_x x_i$$

against

$$H_a\text{: } E(y_i|x_i) \neq \beta_0 + \beta_x x_i.$$

Note that H_a is a general alternative. To investigate how robust the tests are under various situations, we considered two cases for the error terms: (a) $e_i \sim N(0, \frac{2}{3}(1.5 - x_i^2))$ and (b) $e_i \sim \text{Gamma}(1,1)$. The covariate x was generated from $\frac{1}{3}\text{Unif}[-1,0] + \frac{2}{3}\text{Unif}[-1,1]$. True values $\beta_0 = 0$, $\beta_x = 2$, $\alpha_0 = 1$ and $\alpha_1 = 1$ were used in generating y_i and δ_i. To generate the supplement covariate $\mathbf{u}$ for generalized score tests, we sorted all complete cases to

$$\{(y_{(1)}, x_{(1)}), \cdots, (y_{(N_0)}, x_{(N_0)})\}$$

such that $x_{(j)} \leq x_{(k)}$ if $j < k$, where N_0 is total number of complete cases. Then the supplement covariate for the ith observation is

$$\mathbf{u}_i = (cos(2\pi\frac{L_i}{N_0}), sin(2\pi\frac{L_i}{N_0}), cos(4\pi\frac{L_i}{N_0}), sin(4\pi\frac{L_i}{N_0}), cos(6\pi\frac{L_i}{N_0}))',$$

where L_i is the position of the i-th observation after sorting. In addition, we used the weighted estimating equations (1.1) with equal variance assumption to construct the generalized score statistics for both cases. The sample sizes are $n = 100$, 300 and 500. The results for Case (a) are given in Table 2 and Figure 1; the results for Case (b) are given in Table 3.

Under H_0, in Case (a), F_{GS}, T_{GS}, and T_{GSP} have rejection rates close to the nominal level of 5%, the CC method has a rejection rate much higher than the nominal level when the sample size is large, and T_{GSN} has a somewhat higher rejection rate than the nominal level when the sample size is not large and the rejection rate is close to the nominal level when n is large ($n = 500$). In Case (b), F_{GS}, T_{GS} and T_{GSP} also have a rejection rate close to the nominal level, while T_{GSN} tends to have a slightly higher rejection rate when the sample size is small and has a rejection rate close to the nominal level when $n = 300$ and 500. Under the alternatives, T_{GSN} has higher power than T_{GS} and T_{GSP}, and the CC method has significantly lower power than others. The proposed test statistics appear to work well in both cases.

Table 2: Comparisons of generalized score tests for testing adequacy of a simple linear model. Data were generated from a model with an additional quadratic term and heteroscedastic normal error. About 64% observations are fully observed.

	Method	F_{GS}	CC	T_{GS}	T_{GSP}	T_{GSN}	T^{Δ}_{GSP1}	T^{Δ}_{GSP2}
n	c	Rejection Rate (%), Test Level 0.05						
100	0.0	4.7	5.6	4.5	5.8	8.5	8.3	6.2
	0.5	18.7	4.8	6.8	6.5	15.7	6.3	6.1
	1.0	67.7	19.8	29.0	27.8	50.4	23.1	27.0
	1.5	97.2	61.2	72.1	72.8	86.4	66.1	70.7
300	0.0	4.8	16.1	5.9	5.7	8.2	9.3	6.5
	0.5	75.1	14.0	42.8	42.7	55.9	29.8	39.4
	1.0	100.0	93.0	99.1	99.1	98.9	97.6	99.0
500	0.0	4.0	31.6	4.5	4.2	6.4	8.0	4.7
	0.5	98.9	37.5	84.2	85.5	88.5	73.5	83.6
	0.8	100.0	83.0	99.3	99.4	98.9	97.9	99.2

To investigate the issue of the model misspecification of the selection probability, we used the following misspecified models for the selection probability:

$$\text{logit}\{\Pr(\delta_i = 1|y_i)\} = m_0 + m_y y_i. \tag{5.3}$$

and

$$\text{logit}\{\Pr(\delta_i = 1|y_i)\} = m_0 + m_y y_{ti}, \tag{5.4}$$

where $y_t = \sqrt{y}$ if $y > 0$, $-\sqrt{-y}$ otherwise. Intuitively, Model (5.4) is more appropriate for the selection probability than Model (5.3), though both models are not exactly correct. We use T^{Δ}_{GSP1} and T^{Δ}_{GSP2} to denote the generalized score test statistics when the selection probability is estimated by the misspecified model (5.3) and (5.4, respectively. For brevity, we present only the normal error case. In this setting, the rejection rate of T^{Δ}_{GSP1} is slightly higher than the nominal level under H_0, and its power is much lower than those of T_{GS}, T_{GSP} and T_{GSN} under alternatives. On the other hand, the test statistic T^{Δ}_{GSP2} has almost same performance as T_{GS}, and T_{GSP}. The results are in Table 2 and Figure 2.

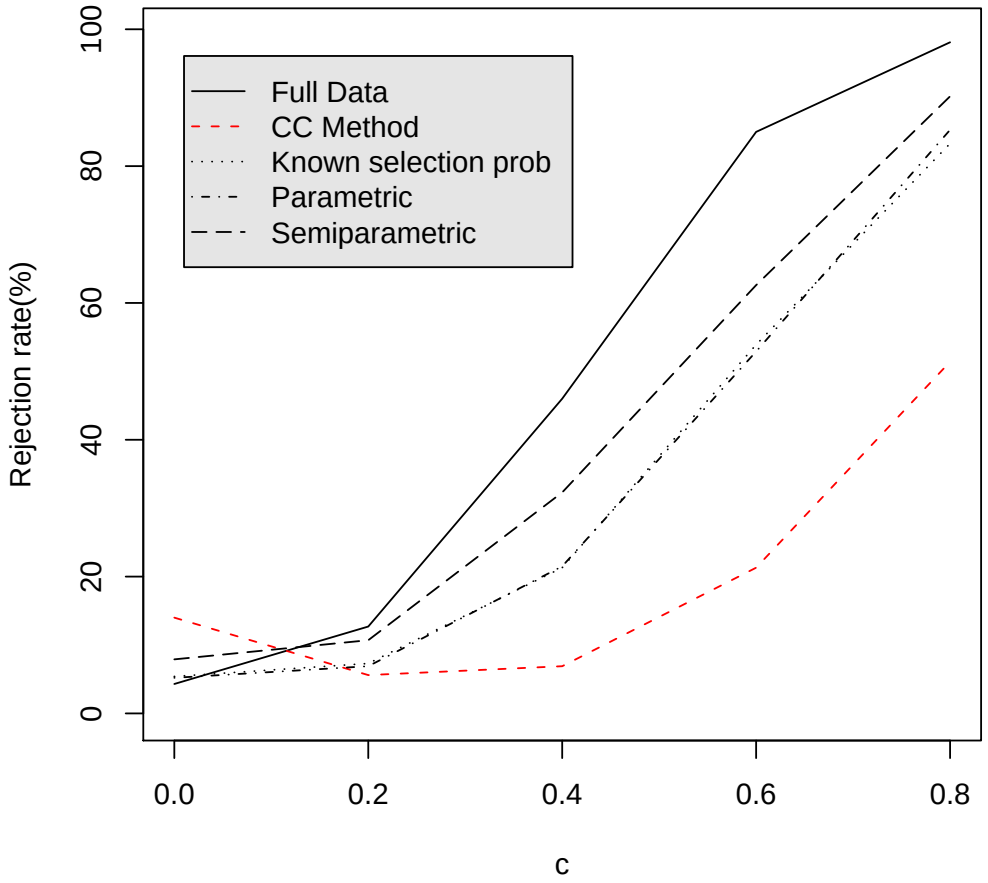

Figure 1: Comparisons of generalized score tests for testing adequacy of a simple linear model. Data were generated from a model with an additional quadratic term and heteroscedastic normal error terms. About 64% observations are fully observed. The sample size is 300.

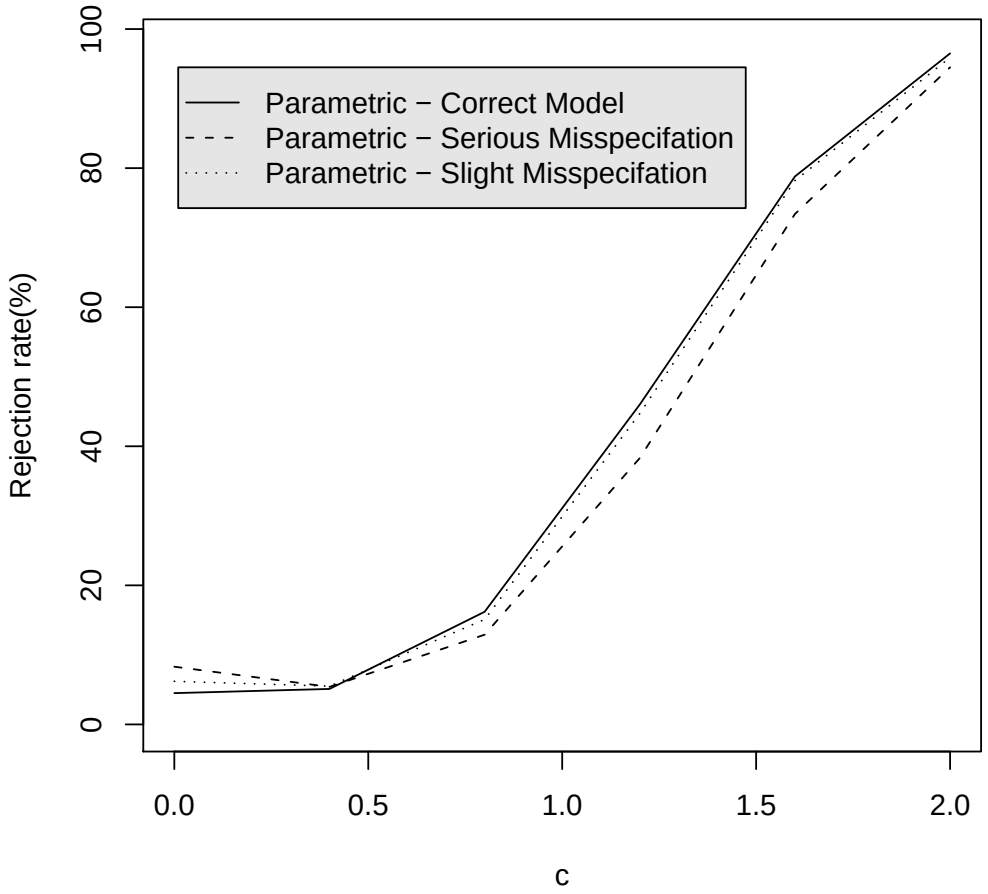

Figure 2: The effect of model misspecification in the selection probability on the generalized score tests. Data were generated from a model with an additional quadratic term and heteroscedastic normal error terms. About 64% observations are fully observed. The sample size is 100.

Table 3: Comparisons of generalized score tests for testing adequacy of a simple linear model. Data were generated from a model with an additional quadratic term and homoscedastic $gamma(1,1)$ error terms. About 64% observations are fully observed.

	Method	F_{GS}	CC	T_{GS}	T_{GSP}	T_{GSN}
n	c	Rejection Rate (%), Test Level 0.05				
100	0.0	3.6	3.7	3.3	4.1	7.1
	0.5	13.4	6.4	8.1	8.4	13.1
	1.0	55.1	25.3	33.4	32.8	45.1
	1.5	93.0	62.1	75.3	75.7	83.8
300	0.0	4.9	4.7	4.5	4.6	5.3
	0.5	45.9	18.9	32.8	32.1	37.5
	1.0	98.9	81.1	95.0	95.3	97.4
500	0.0	5.1	5.3	4.7	4.9	5.7
	0.5	71.5	32.1	56.5	56.4	63.3
	0.8	99.1	83.2	96.1	96.2	97.7

To investigate how the estimated mean score function improves the power of the tests, we considered the situation that ϕ^* was estimated via a correct parametric model. The data were generated from the linear model with a univariate binary covariate x_i, a univariate continuous covariate z_i and the response variable y_i:

$$y_i = \beta_0 + \beta_x x_i + \beta_z z_i + c z_i^2 + e_i. \tag{5.5}$$

where the error terms follow identical and independent $N(0,1)$, and δ_i follows the logistic model (5.2) with $\alpha_0 = \alpha_1 = 1$ as before. The covariate $z_i = -1 + (2i)/n$, $x_i \sim \text{Bernoulli}(p_{xi})$, and $\text{logit}(p_{xi}) = \kappa_0 + \kappa_z z_i$. True values $\beta_0 = 0$, $\beta_z = 1.0$, $\beta_x = 0.2$ or 0.8 and $\kappa_0 = \kappa_z = 0$ were used. In this case, the supplement covariate for the ith observation is

$$\mathbf{u}_i = (cos(2\pi t_i), sin(2\pi t_i), cos(4\pi t_i), sin(4\pi t_i), cos(6\pi t_i))',$$

where $t_i = i/n$. The results are given in Table 4 and Figure 3.

Under H_0, both non data driven and data driven T^*_{GSP} have rejection rates close to the nominal level 5% for both sample sizes of $n = 100$ and 300. Under the

Table 4: Comparisons of generalized score tests and their data driven version tests for testing adequacy of a linear model with covariates x and z. Data were generated from a model with an additional quadratic term cz^2. The error terms follow identical and independent $N(0,1)$. Around 63% observations are fully observed.

Method		Non data Driven				Data Driven		
		F_{GS}	CC	T_{GSP}	T^*_{GSP}	F_{GS}	T_{GSP}	T^*_{GSP}
n	c	Rejection Rate (%), Test Level 0.05						
		$\beta_x = 0.2$						
100	0.0	4.6	5.0	6.3	4.1	4.8	6.8	5.0
	0.5	14.8	7.7	8.7	14.6	24.0	12.6	23.5
	1.0	56.2	24.3	25.4	52.7	74.4	44.4	71.3
300	0.0	4.6	5.0	4.9	4.1	5.2	5.3	5.2
	0.5	43.2	15.3	24.3	41.1	58.7	32.9	57.1
	1.0	98.1	76.3	81.9	97.9	99.3	90.8	100.0
		$\beta_x = 0.8$						
100	0.0	4.6	5.1	6.9	5.0	4.8	6.7	5.3
	0.5	14.8	9.2	9.8	14.1	24.0	14.0	23.0
	1.0	56.2	30.9	29.3	51.4	74.4	47.8	68.7
300	0.0	4.6	4.6	5.5	4.0	5.2	5.3	4.6
	0.5	43.2	21.7	25.2	39.8	58.7	34.2	54.3
	1.0	98.1	85.5	84.5	97.3	99.3	92.5	100.0

alternatives, (a) T^*_{GSP} is much more powerful than T_{GSP}; (b) the data driven version of tests is more powerful than their non data driven version tests. This implies that the tests with an appropriate model for ϕ^* are much more efficient than the tests with $\phi = 0$, the result. When $\beta_x = 0.2$, the covariate x is not very useful in the regression model and the missingness in x caused little information loss. In this case T^*_{GSP} has the power close to that of F_{GS}. On the other hand, when $\beta_x = 0.8$, the missingness in x led to more information loss. Consequently the power of T^*_{GSP} is reduced slightly more from that of F_{GS}.

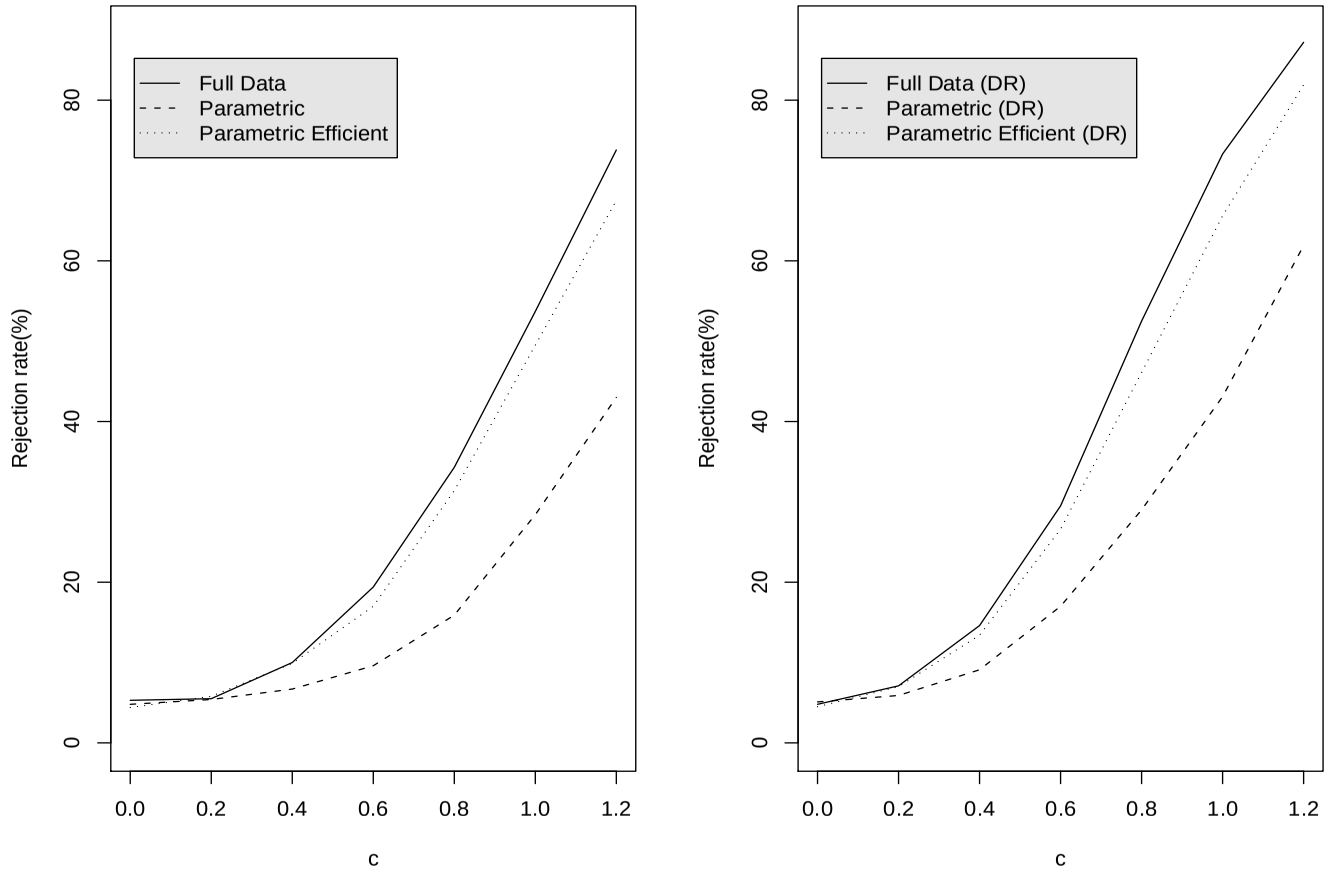

Figure 3: Comparisons of generalized score tests with their data driven version tests for testing adequacy of a linear model with two covariates. Data were generated from a model with an additional quadratic term cz^2. The error terms follow identical and independent $N(0,1)$. The sample size is 100.

5.3 Logistic Regression

In this part of the simulation study, we considered the proposed tests for testing adequacy of logistic regression models. The data were generated from the following models:

Model I:

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_x x_i + \beta_z z_i + c z_i^2,$$

Model II:

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_x x_i + \beta_z z_i + c x_i \times z_i,$$

where y_i is a binary response variable, z_i and x_i are two univariate covariates. The missing indicator δ_i follows the logistic regression model with covariates y_i and z_i:

$$\text{logit}\{\Pr(\delta_i = 1|y_i, z_i)\} = y_i(3z_i^2 - 0.4).$$

For both models, the values of $\boldsymbol{\beta}$ used to generate the data were $\beta_0 = \beta_x = 0$ and $\beta_z = 1$; the covariate x_i was generated from Unif$[-1, 1]$ and z_i were equal space points between -1 and 1. According to applications, the alternative could be a partial linear model or a general alternative.

5.3.1 Partially Linear Alternatives

Partially linear models are widely used for missing covariate data because of their flexibilities. In this subsection, we would like to test

$$H_0\text{: logit}\{\Pr(y_i|x_i, z_i)\} = \beta_0 + \beta_x x_i + \beta_z z_i$$

against its partially linear alternative

$$H_a\text{: logit}\{\Pr(y_i|x_i, z_i)\} = \beta_0 + \beta_x x_i + f_z(z_i),$$

where f_z is a smooth function. The data were simulated using the model I and model II. The supplement covariate for the ith observation is

$$\mathbf{u}_i = (cos(2\pi t_i), sin(2\pi t_i), cos(4\pi t_i), sin(4\pi t_i), cos(6\pi t_i))',$$

where $t_i = i/n$. The simulation results based on Model I are given in Table 5 and the simulation results based on Model II are given in Table 6

The data were simulated form Model I: under H_0, F_{GS}, T_{GS} and T_{GSP} have rejection rates close to the nominal level 5% for both sample sizes $n = 200$ and 500; the CC method has a rejection rate of 35% when $n = 500$, indicating the CC method is severely biased; T_{GSN} has a somewhat high rate of 8.4% when the sample size $n = 200$ and a rejection rate close to the nominal level when sample size $n = 500$, reflecting the fact that T_{GSN} converges to the asymptotic distribution slowly. Under the alternatives, T_{GS} and T_{GSP} have similar power. The data driven version tests are also given in Table 5. The data driven procedures have similar rejection rates to those of the non data driven tests under H_0, with the exception of their noticably higher power under the alternatives.

The data were simulated form Model II: under H_0 and H_a, F_{GS}, T_{GS} and T_{GSP} have rejection rates close to 5% for sample size $n = 200$ and 500. This indicates the tests may have no power to detect an interaction departure if the alternative is a partial linear model. It is not a good idea to use a partial linear alternative if you would like to detect a general signal.

5.3.2 General Alternatives

Partial alternatives may be inadequate in some applications. In this subsection, we would like to test

$$H_0\text{: logit}\{\Pr(y_i|x_i, z_i)\} = \beta_0 + \beta_x x_i + \beta_z z_i$$

against a general alternative

$$H_a\text{: logit}\{\Pr(y_i|x_i, z_i)\} = f_{xz}(x_i, z_i),$$

where f_{xz} is a general bivariate smooth function. To construct the proposed goodness-of-fit tests, the following partitioning was utilized. The regions of covariate space were automatically partitioned into 6 parts:

$$\begin{aligned}
\text{Part I} &= \{x < q_{x,0.5}, z < q_{z,0.33}\},\\
\text{Part II} &= \{x \geq q_{x,0.5}, z < q_{z,0.33}\},\\
\text{Part III} &= \{x < q_{x,0.5}, q_{z,0.33} \geq z < q_{z,0.66}\},\\
\text{Part IV} &= \{x \geq q_{x,0.5}, q_{z,0.33} \geq z < q_{z,0.66}\},\\
\text{Part V} &= \{x < q_{x,0.5}, q_{z,0.66} \geq z\},\\
\text{Part VI} &= \{x \geq q_{x,0.5}, q_{z,0.66} \geq z),
\end{aligned}$$

where $q_{x,t1}$, $q_{x,t2}$ are the t_1 and t_2 quantile of the variables x and z, respectively. The corresponding supplement covariates are 5×1 vectors. The results of the simulation study are given in the Table 7 and Figure 4. Under H_0, F_{GS}, T_{GS} and T_{GSP} have rejection rates close to the nominal level 5% for both sample sizes $n = 200$ and 500; the CC method has a rejection rate of 16.1% when $n = 200$ and has a rejection rate of 43.1% when $n = 500$, indicating the CC method is severely biased; T_{GSN} has a somewhat high rate of 9.1% when the sample size $n = 200$ and a rejection rate close to the nominal level when sample size $n = 500$. Under the alternatives, T_{GS} and T_{GSP} have similar power; the power of T_{GSN} is higher than that of T_{GS} and T_{GSP} while type I error are similar when $n = 500$. Recall the results in Table 6, using the same simulated data, the tests almost have no power when the alternative is a partial linear model while the tests have reasonable power when the alternative is a general one.

Table 5: Comparisons of generalized score tests and their data driven version tests for testing adequacy of a logistic regression against a partial linear alternative. The responses were generated from a logistic regression model (Model I) with an additional quadratic term cz^2 . Around 66% observations are fully observed.

Method	F_{GS}	CC	T_{GS}	T_{GSP}	T_{GSN}
c	Rejection Rate (%), Test Level 0.05				
$n = 200$					
Non data Driven Methods					
0.0	4.0	6.4	5.8	4.8	8.4
0.5	11.0	20.0	11.0	8.8	16.8
1.0	29.4	44.8	23.6	21.2	36.6
2.0	82.8	81.8	64.2	63.8	86.2
Data Driven Methods					
0.0	3.0	35.0	4.8	3.2	6.0
0.5	14.4	63.0	11.6	9.2	15.2
1.0	39.2	84.4	26.0	24.4	40.6
2.0	91.6	98.8	69.6	69.6	91.4
$n = 500$					
Non data Driven Methods					
0.0	6.0	35.6	4.8	4.6	5.4
0.5	19.8	75.4	17.0	14.6	19.8
1.0	67.0	97.0	44.8	41.6	65.4
2.0	100.0	100.0	95.8	95.6	100.0
Data Driven Methods					
0.0	5.4	75.4	5.0	5.0	5.8
0.5	25.0	97.6	19.4	15.6	22.8
1.0	77.2	100.0	52.6	49.8	74.2
2.0	100.0	100.0	97.4	97.0	100.0

Table 6: Comparisons of generalized score tests (non data driven method) for testing adequacy of a logistic regression against a partial linear alternative. The responses were generated from a logistic regression model (Model II) with an additional interaction term $c\ x \times z$. Around 65% observations are fully observed.

Method	F_{GS}	CC	T_{GS}	T_{GSP}	T_{GSN}
c	Rejection Rate (%), Test Level 0.05				
			$n = 200$		
0.0	3.6	21.7	5.6	5.5	7.9
1.0	4.5	21.0	5.6	5.5	7.0
2.0	3.6	22.4	6.0	5.2	7.3
			$n = 500$		
0.0	4.1	60.9	4.8	4.0	4.4
1.0	5.2	58.8	5.4	3.8	4.1
2.0	6.4	59.6	5.5	3.8	6.7

Table 7: Comparisons of generalized score tests (data driven) for testing testing adequacy of a logistic regression against a general alternative. The responses were generated from a logistic regression model (Model II) with an additional interaction term $cx \times z$. Around 65% observations are fully observed.

Method	F_{GS}	CC	T_{GS}	T_{GSP}	T_{GSN}
c	Rejection Rate (%), Test Level 0.05				
			$n = 200$		
0.0	4.7	16.7	7.5	6.3	9.1
0.5	7.1	27.6	10.3	8.9	12.5
1.0	14.3	38.1	14.2	11.9	15.7
1.5	29.5	52.2	19.3	17.3	27.9
2.0	47.7	64.4	27.0	25.3	40.8
			$n = 500$		
0.0	5.1	43.1	5.0	3.9	4.0
0.5	10.3	69.4	8.6	5.5	7.7
1.0	34.0	88.3	20.1	15.9	24.5
1.5	67.8	96.3	40.2	35.0	54.8
2.0	89.7	98.9	62.0	57.6	79.4

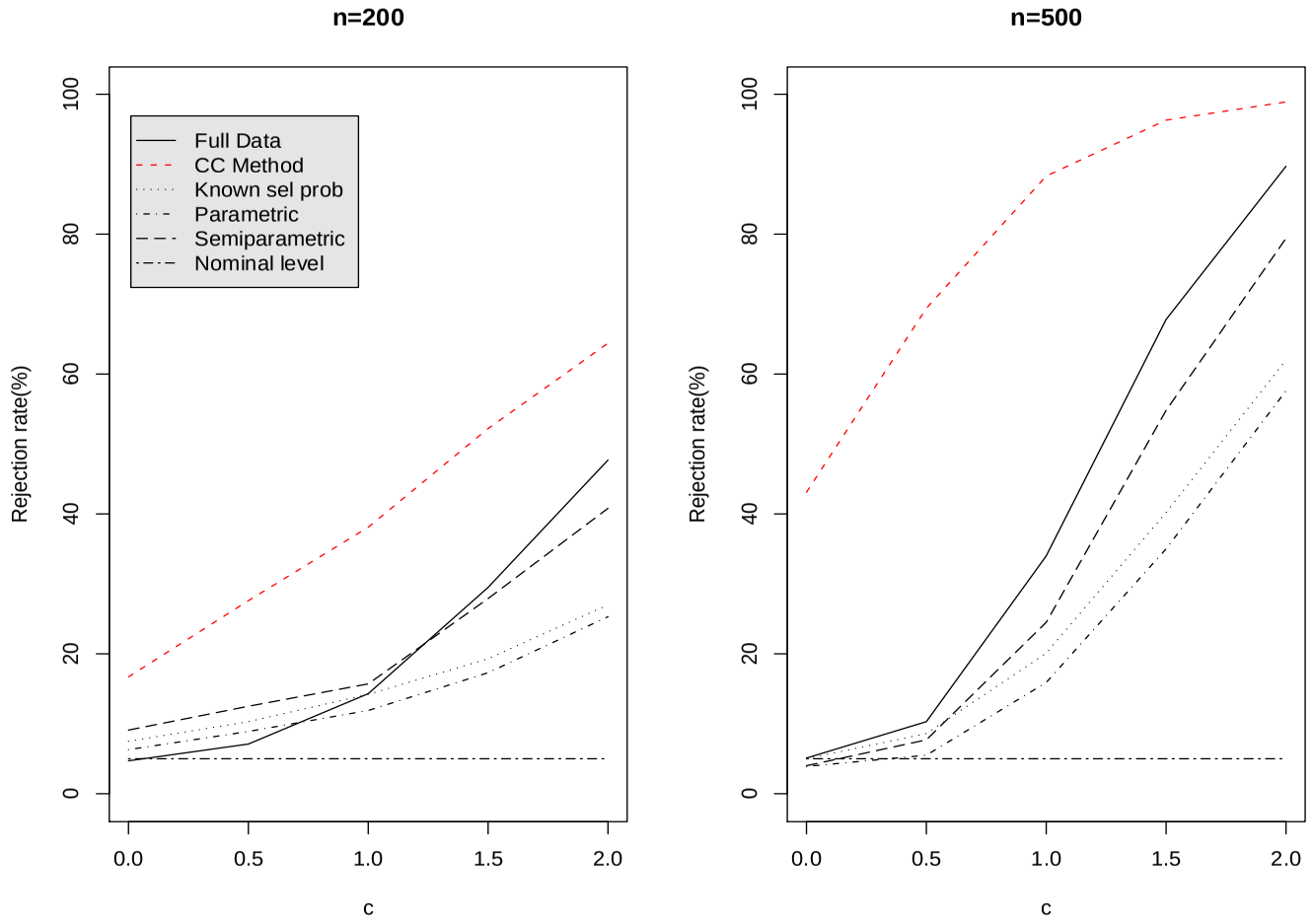

Figure 4: Comparisons of generalized score tests (data driven) for testing testing adequacy of a logistic regression against a general alternative. The responses were generated from a logistic regression model (Model II) with an additional interaction term $cx \times z$. Around 65% observations are fully observed. The sample size $n = 200$ and 500.

Table 8: Comparisons of the generalized score test, its data driven test and the adaptive Neyman test for testing testing adequacy of a simple linear model against a general alternative when no missingness occurs. Data were generated from a model with an additional quadratic term cz^2. The error terms follow identical and independent $N(0,1)$.

Method	F_{GS}	DR	$Neyman$	F_{GS}	DR	$Neyman$	F_{GS}	DR	$Neyman$
	$N = 100$			$N = 300$			$N = 500$		
c	Reject Ratio (%), Test Level 0.05, $r = 5$								
0.0	4.7	5.3	5.1	3.9	4.2	3.7	4.9	4.2	5.0
0.2	5.2	6.5	9.9	8.5	10.5	12.4	13.0	17.7	17.6
0.4	8.1	10.6	15.2	25.5	33.5	34.9	48.9	61.3	61.0
0.6	12.1	19.6	25.0	55.6	67.0	68.4	86.5	92.8	92.3
0.8	21.0	31.8	38.9	85.0	91.2	91.1	99.2	99.9	99.9
1.0	33.3	46.7	53.0	97.0	98.5	98.9	100.0	100.0	100.0

5.4 Comparisons between Tests When No Missingness Occurs

Our proposed test statistics simplify to generalized score statistics in Boos (1992) and their data driven versions similar to Aerts et al. (2000) when no missingness occurs. Under some circumstances, there might be some optimal or nearly optimal goodness-fit-tests, such as the adaptive Neyman test in Fan and Huang (2001) for testing adequacy of a simple linear model. In this section, we compare the performance between the generalized score test, its data driven test and the adaptive Neyman test when no missingness occurs, to indirectly gain further understanding of the performance of our proposed test statistics for missing covariate data.

We simulated the response variable using (5.1) with $\beta_0 = 0$, $\beta_x = 2$, and two types of error terms (a) $N(0,1)$, (b) $N(0, \frac{2}{3}(1.5 - x_i^2))$. The results are depicted in Tables 8, 9 and Figure 5. In case (a), the adaptive Neyman test has better performance than the data driven test when $n = 100$, and almost the same performance as the data driven test when $n = 200$, $n = 300$ and $n = 500$. However, the adaptive Neyman test might not be as good as the data driven test with heterogeneity of variance.

Table 9: Comparisons of the generalized score test, its data driven test and the adaptive Neyman test for testing testing adequacy of a simple linear model against a general alternative when no missingness occurs. Data were generated from a model with an additional quadratic term cz^2 and heteroscedastic normal error terms.

Method	F_{GS}	DR	$Neyman$	F_{GS}	DR	$Neyman$	F_{GS}	DR	$Neyman$
	$N = 100$			$N = 300$			$N = 500$		
c	Reject Ratio (%), Test Level 0.05, $r = 5$								
0.0	4.2	4.6	4.3	4.6	5.1	5.1	4.1	5.2	5.6
0.2	5.5	7.7	5.2	17.6	19.7	8.8	30.1	35.6	11.6
0.4	11.7	16.0	8.0	59.7	64.5	29.0	89.5	91.1	62.4
0.6	24.4	31.3	14.8	95.2	96.9	72.1	100.0	100.0	98.5
0.8	44.6	52.5	26.0	100.0	100.0	97.1	100.0	100.0	100.0
1.0	65.0	72.8	44.9	100.0	100.0	100.0	100.0	100.0	100.0

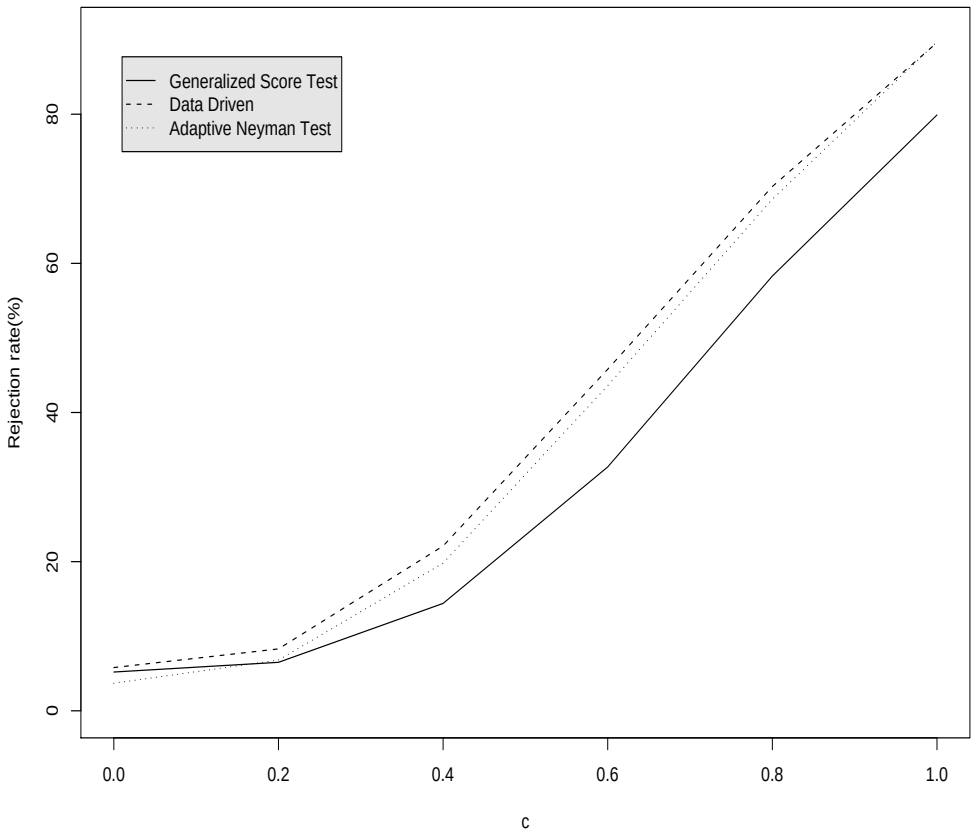

Figure 5: Comparisons of the generalized score test, its data driven test and the adaptive Neyman test for testing testing adequacy of a simple linear model against a general alternative when no missingness occurs. Data were generated from a model with an additional quadratic term cz^2. The error terms follow identical and independent $N(0, 1)$. The sample size is 200.

CHAPTER VI

AN EXAMPLE OF DATA ANALYSIS

6.1 Introduction

In this chapter, we consider the dataset mentioned in Chapter I, which is from the Duke University Cardiovascular Disease Databank. The patients were referred to Duke University Medical Center for chest pain. It was found that 2332 patients have significant ($\geq 75\%$ diameter narrowing of at least one major coronary artery) coronary disease (sigdz = 1) by Cardiac Catheterization. Among these 2332 patients, 1129 patients have severe coronary disease (three-vessel or left main disease, tvdlm = 1). It is very interesting to predict the probability of significant coronary disease and the probability of severe coronary disease given the information of cholesterol ,age, and so on. The content of the dataset was described in Chapter I. As was stated earlier, it consists of 3504 patients and 6 variables. The covariate cholesterol is not observed among 1246 out of 3504 observations. Harrell (2001) analyzed the dataset extensively. However, complete-case analysis was used when the covariate cholesterol is involved in his analysis. It is well know that complete-case analysis may be misleading if the missing-data mechanism is not MCAR. In this chapter, we reanalyze the Duke Cardiac Catheterization Coronary Artery Disease Diagnostic Dataset to illustrate our methodology.

Table 10: Fit the missingness on sigdz, age and sigdz*age.

	Estimate	Std. Error	p-value
(Intercept)	1.862555	0.323827	8.84e-09
sigdz	1.774492	0.425385	3.03e-05
age	-0.024721	0.006418	0.000117
sigdz × age	-0.031378	0.008087	0.000104

6.2 Data Analysis

We are interested in investigating the relationship between sigdz (y) and covariates cholesterol (x) and age (z) while one third of cholesterol values are missing. Due to the missingness, it is crucial to identify the relationship between the missingness and the values of variable. Assume that the data are MAR, we characterized the missing-data mechanism by fitting the logistic regression

$$\text{logit}\{\Pr(\delta_i = 1|y_i, z_i)\} = \alpha_0 y_i + \alpha_z z_i + \alpha_{yz} y_i \times z_i.$$

The results are shown in Table 10. Significant dependence of the missingness on the data is apparent because all terms above are significant (p-value < 0.001), indicating that the data are not MCAR and the missingness depends on y_i and z_i. This suggests that the previous complete-case analysis (Harrell 2001) might be problematic for this dataset.

As in Harrell's (2001) analysis, we assumed a logistic linear regression model

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_x x_i + \beta_z z_i. \tag{6.1}$$

Because the sample size is relatively large and we don't have much knowledge about the selection probability, we used the semiparametric approach in Wang et al. (1997) to estimate the parameters. The estimate of the parameter $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z)'$ is the solution of

$$U_s(\boldsymbol{\beta}, \hat{\pi}) = \sum_{i=1}^{3504} \{\frac{\delta_i}{\hat{\pi}_i}(1, x_i, z_i)'(y_i - \hat{p}_i)\},$$

where $\hat{\pi}$ is the local linear estimate for the selection probability and

$$\hat{p}_i = \frac{1}{1 + e^{-(\beta_0 + \beta_x x_i + \beta_z z_i)}}.$$

The estimate of the parameters is

$$(\hat{\beta}_0, \hat{\beta}_x, \hat{\beta}_z) = (-3.28, 0.049, 0.0064).$$

Furthermore, using the semiparametric generalized score statistic T_{GSN} to test each of the following three null hypotheses: $H_0 : \beta_0 = 0$, $H_0 : \beta_x = 0$ and $H_0 : \beta_z = 0$, we found that each term is significant (p-value < 0.001). Before we use these results to explain the relationship between the disease and the covariates age and cholesterol, it is natural to ask if model (6.1) is adequate. To investigate this issue, let the null hypothesis H_0 be the model in (6.1). Possible alternatives are:

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_z z_i + f_1(x_i), \tag{6.2}$$

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_x x_i + f_2(z_i), \tag{6.3}$$

or

$$\text{logit}\{\Pr(y_i = 1|x_i, z_i)\} = g(x_i, z_i), \tag{6.4}$$

where f_1, f_2 and g are smoothing functions. Alternatives (6.2) and (6.3) are partially linear models, while alternative (6.4) is a general one. The different supplement covariates should be used for different alternatives. We used the cosine system to generate plausible alternative models with $R = 5$ supplement covariates for the partially linear alternatives. We sorted the observations according to the value of x and z from smallest to largest, respectively, in order to generate the two supplement covariates for ith observation for alternatives (6.2) and (6.3). Let L_i^x and L_i^z be the positions of the ith observation after sorting, accordingly. Then

$$(cos(2\pi\frac{L_i^x}{N_0}), sin(2\pi\frac{L_i^x}{N_0}), cos(4\pi\frac{L_i^x}{N_0}), sin(4\pi\frac{L_i^x}{N_0}), cos(6\pi\frac{L_i^x}{N_0}))'$$

and

$$(cos(2\pi\frac{L_i^z}{n}), sin(2\pi\frac{L_i^z}{n}), cos(4\pi\frac{L_i^z}{n}), sin(4\pi\frac{L_i^z}{n}), cos(6\pi\frac{L_i^z}{n}))'$$

are the supplement covariates for the ith observation for alternatives (6.2) and (6.3), respectively, where $N_0 = 2258$ and $n = 3504$. For alternative (6.4), the corresponding plausible alternative $M_{(R)}$, $R = 5$ was generated based on partitioning the covariate space into 6 distinct regions. The observed test statistics are 5.77, 16.03 and 33.06 for the three alternatives above, respectively. The asymptotic critical value is 11.07. Therefore, the conclusion is that the linear relationship described by (6.1) between the disease and covariates age and cholesterol is not adequate. On the other hand, partially linear model (6.3) or a fully nonparametric regression (6.4) may be more adequate to describe such a relationship. To further analyze this dataset, it is possible to apply the methodologies developed in Liang et al. (2004) and Wang, Wang, Gutierrez, and Carroll (1998) for partially linear models and fully nonparametric techniques in generalized linear models, respectively, when some covariates are partially missing.

CHAPTER VII

SUMMARY AND FUTURE RESEARCH

7.1 Summary

In this dissertation, we have studied the generalized score tests based on WEE (1.1) with two nuisance functions π and ϕ for missing covariate data. Different versions of the test statistic have been properly defined according to different settings, and their asymptotic distributions have been derived. The proposed parametric tests appear to give proper type I error rates and reasonable power for different sample sizes and obtain the asymptotically optimal power within the class when the parametric models for π and ϕ are correctly specified, while the proposed semiparametric tests appear to work well when sample size is sufficiently large. Moreover, the optimal power can also be obtained asymptotically by using an appropriate nonparametric estimate of π using the simplified WEE with $\phi = 0$. As an important application, we have investigated the model assessment procedures for generalized linear models when some covariates are partially missing. Our empirical study suggests that, with a proper choice of the function sequence $\mathcal{F}$ and the number of supplement covariates, the tests have good power in testing certain types of departures from the null models.

7.2 Future Research

As a future research problem, it would be interesting to extend the proposed methodology by employing generalized weighted estimating equations for correlated data such as longitudinal data with missing covariates. It would also be interesting to develop

generalized score tests for testing overdispersion, correlation and heterogeneity over mixed effects in the presence of missing covariates.

Another problem is that the asymptotic null distribution seems to be inadequate for the semiparametric tests when the sample size is not large enough. One possible remedy for this is to develop bootstrap methods to approximate the critical values of the null distributions. Another possibility is to investigate the effects of Bartlett corrections in an attempt to improve the accuracy of approximate null distributions.

REFERENCES

Aerts, M., Claeskens, G., and Hart, J. D. (1999). Testing the fit of a parametric function. *Journal of the American Statistical Association* **94**, 869–879.

Aerts, M., Claeskens, G., and Hart, J. D. (2000). Testing lack of fit in multiple regression.. *Biometrika* **87**, 405–424.

Barnhart, H. X. and Williamson, J. M. (1998). Goodness-of-fit tests for GEE modeling with binary data. *Biometrics* **54**, 720–729.

Boos, D. D. (1992). On generalized score tests. *American Statistician* **46**, 327–333.

Breslow, N. E. and Cain, K. C. (1988). Logistic regression for two-stage case-control data. *Biometrika* **75**, 11–20.

Chen, H. Y. and Little, R. (1999). A test of missing completely at random for generalized estimating equations with missing data. *Biometrika* **86**, 1–13.

Commenges, D. and Jacqmin-Gadda, H. (1997). Generalized score test of homogeneity based on correlated random effects models. *Journal of the Royal Statistical Society, Series B* **59**, 157–171.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

Emerson, L. P. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics* **24**, 695–701.

Fan, J. and Huang, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association* **96**, 640–652.

Flanders, W. D. and Greenland, S. (1991). Analytic methods for two-stage case-control studies and other stratified designs. *Statistics in Medicine* **10**, 739–747.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* **61**, 74–85.

González-Manteiga, W. and Pérez-González, A. (2006). Goodness-of-fit tests for linear regression models with missing response data. *The Canadian Journal of Statistics* **34**, 149–170.

Harrell, F. E. (2001). *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer-Verlag.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (1999). Incomplete data in generalized linear models. *Biometrics* **55**, 591–596.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., and Herring, A. (2005). Missing-data methods for generalised linear models: A comparative review. *Journal of the American Statistical Association* **100**, 332–346.

Lei, S. Y. and Wang, S. (2001). Diagnostic tests for bias of estimating equations in weighted regression with missing covariates. *The Canadian Journal of Statistics* **29**, 239–250.

Liang, H., Wang, S., Robins, J. M., and Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association* **99**, 357–367.

Lipsitz, S., Parzen, M., Molenberghs, G., and Ibrahim, J. (2001). Testing for bias in weighted estimating equations. *Biostatistics* **2**, 295–307.

Lipsitz, S. R., Ibrahim, J. G., and Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association* **94**, 1147–1160.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (Second Edition)*. Chichester: Wiley.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. New York: Elsevier.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (Second Edition)*. London: Chapman and Hall.

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

Parzen, M., Lipsitz, S. R., Ibrahim, J. G., and Lipshultz, S. (2002). A weighted estimating equation for linear regression with missing covariate data. *Statistics in Medicine* **21**, 2421–2436.

Qu, A. and Song, P.-K. (2002). Testing ignorable missingness in estimating equation approaches for longitudinal data. *Biometrika* **89**, 841–850.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

Rotnitzky, A. and Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Rubin, D. B. (1996). Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association* **91**, 473–489.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with comments). *Journal of the American Statistical Association* **94**, 1096–1146.

Thas, O. and Rayner, J. C. (2005). Smooth tests for the zero-inflated poisson distribution. *Biometrics* **61**, 808–815.

Tosteson, T. D. and Tsiatis, A. A. (1988). The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika* **75**, 507–514.

Van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer-Verlag.

Wang, C. Y., Wang, S., Gutierrez, R. G., and Carroll, R. J. (1998). Local linear regression for generalized linear models with missing data. *Annals of Statistics* **26**, 1028–1050.

Wang, C. Y., Wang, S., Zhao, L.-P., and Ou, S.-T. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association* **92**, 512–525.

Wang, S. and Wang, C. Y. (2001). A note on kernel assisted estimators in missing covariate regression. *Statistics and Probability Letters* **55**, 439–449.

Zhao, L. P. and Lipsitz, S. R. (1992). Designs and analysis of two-stage studies. *Statistics in Medicine* **11**, 769–782.

Zhao, L. P., Lipsitz, S. R., and Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics* **52**, 1165–1182.

VITA

Lei Jin was born in Yiwu, Zhejiang, China. He received his B.S. in applied mathematics from Huazhong University of Science and Technology, Wuhan, China, June 1998. He received a M.S. in applied mathematics from zhejiang University, Hangzhou, China, June 2001. He received his Ph.D. in statistics from Texas A&M University in August 2007.

Permanent Address:

19 Gonghe Lane, Fotang,

Yiwu, Zhejiang, P.R China, 322002