# Regression models of Pearson correlation coefficient 

Abdisa G. Dufera, Tiantian Liu \& Jin Xu

To cite this article: Abdisa G. Dufera, Tiantian Liu \& Jin Xu (2023) Regression models of Pearson correlation coefficient, Statistical Theory and Related Fields, 7:2, 97-106, DOI: 10.1080/24754269.2023.2164970

To link to this article: https://doi.org/10.1080/24754269.2023.2164970

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor \& Francis Group

皿 Published online: 11 Jan 2023.


Article views: 652


View related articles

[^0]
# Regression models of Pearson correlation coefficient 

Abdisa G. Dufera ${ }^{\mathrm{a}}$, Tiantian Liu ${ }^{\mathrm{a}, \mathrm{b}}$ and Jin Xu ( ${ }^{\mathrm{a}, \mathrm{c}}$<br>${ }^{\text {a }}$ School of Statistics, East China Normal University, Shanghai, People's Republic of China; ${ }^{\text {b }}$ Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel; 'Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, East China Normal University, Shanghai, People's Republic of China


#### Abstract

We propose two simple regression models of Pearson correlation coefficient of two normal responses or binary responses to assess the effect of covariates of interest. Likelihood-based inference is established to estimate the regression coefficients, upon which bootstrap-based method is used to test the significance of covariates of interest. Simulation studies show the effectiveness of the method in terms of type-I error control, power performance in moderate sample size and robustness with respect to model mis-specification. We illustrate the application of the proposed method to some real data concerning health measurements.


## ARTICLE HISTORY

Received 6 September 2022
Revised 27 December 2022
Accepted 1 January 2023

## KEYWORDS

Binary responses; bivariate normal responses; Pearson correlation coefficient; regression

## 1. Introduction

Most regression models have been developed to describe the relationship between the expected value of response(s) and a number of covariates (predictors). In some situations, it is desired to understand the influence of covariates on the strength of Pearson correlation between two responses. For example, in some psychological study, it is of interest to understand the impact of age and/or sex on the association between physical functionality and mental functionality of the elder (Thomas et al., 2016).

One traditional measure for such association is the partial correlation coefficient, i.e., the correlation coefficient of two responses after removing the covariate effect (Anderson, 2003). However, its inference depends on the normality assumption and it misses the connection to the actual values of the covariates which is supposed to have different effects on the association. In terms of the rank-based counterpart, Liu et al. (2018) proposed covariateadjusted partial Spearman's rank correlation using probability scale residuals, which is particularly useful for ordinal responses.

Another method is to use a conditional approach to assess the correlation at different levels of the covariates. Bartlett (1993) proposed to model the Fish $z$-transformed correlation by a polynomial regression of a single covariate $z$. His model builds a linear regression up on paired observations of $\left(z_{k}, r\left(z_{k}\right)\right), k=1, \ldots, K$, where $r\left(z_{k}\right)$ is the sample correlation coefficient of two responses at $z_{k}$ over $K$ distinct levels of $z$. One clear drawback of this method is that it requires repeated measures of responses at $z_{k}$, which would lead to the demand of a large total sample size and even larger when multiple covariates are included. Other studies along this direction include Yu and Dunn (1982) and Paul (1989). Wilding et al. (2011) proposed a model to relate a transformed correlation coefficient of two normal responses ( $y_{1}$ and $y_{2}$ ) with a linear combination of covariates $\left(z_{1}, \ldots, z_{p}\right)$ through the probit link, i.e., $(1+\rho) / 2=\Phi\left(\gamma_{0}+\gamma_{1} z_{1}+\cdots+\gamma_{p} z_{p}\right)$, where $\rho=\operatorname{corr}\left(y_{1}, y_{2}\right)$ and $\Phi$ is the distribution function of the standard normal variable. (Restricted maximum likelihood corrected LRT is used to test the significance of covariates.) We point out that the linear transformed correlation $(1+\rho) / 2$ guarantees the range for a distribution function. However, it can lose efficiency when $\rho$ is known positive (since it can be modelled directly without transformation).

All the aforementioned methods consider the continuous responses. In many situations, there are only binary responses available, e.g., after dichotomization. In this paper, we propose two simple regression models of Pearson correlation coefficient of two normal responses and two binary responses (Section 2). Specifically, when the correlation coefficient is known positive, the logistic link function is used and when the correlation coefficient is arbitrary in $(-1,1)$, the hyperbolic tangent (or Fisher $z$-transformation) link function is used. Likelihood-based inference is developed to estimate the regression coefficients (Section 3). We demonstrate the performance of the proposed

[^1]method by simulation in Section 4 and illustrate the applications in great detail to a real data concerning physical functionality and mental functionality of the elder (Section 5). We defer the technical details in Appendix.

## 2. Model

Let $\left(y_{1}, y_{2}\right)$ be a pair of responses of a subject with correlation coefficient $\rho=\operatorname{corr}\left(y_{1}, y_{2}\right)$. Let $x_{1}, \ldots, x_{p}$ denote $p$ associated covariates of the paired responses, which follow a joint distribution $f\left(x_{1}, \ldots, x_{p}\right)$. Consider modelling $\rho$ as a monotonic function of the linear combination of these covariates through

$$
\begin{equation*}
\rho=h\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right), \tag{1}
\end{equation*}
$$

where $\mathbf{x}=\left(1, x_{1}, \ldots, x_{p}\right)^{\top}$ and $\boldsymbol{\beta}=\left(\beta_{0}, \beta_{1}, \ldots, \beta_{p}\right)^{\top}$. When $\rho>0$, we use the logistic function $h(x)=$ $\left(1+\mathrm{e}^{-\mathrm{x}}\right)^{-1}$, referred by link 1 . When $-1<\rho<1$, we use the hyperbolic tangent function $h(x)=\tanh (x)=$ $\left(e^{2 \mathrm{x}}-1\right) /\left(\mathrm{e}^{2 \mathrm{x}}+1\right)$, referred by link 2 . We adopt these two commonly used link functions to map the real line to intervals $(0,1)$ and $(-1,1)$, respectively, as they are analytically simple. Other links can also be used, e.g., the probit link for $\rho>0$. When the correlation is known positive, the logistic function is preferred as it is more efficient for estimation and easy for interpretation. Let $h^{\prime}$ and $h^{\prime \prime}$ denote the first and second derivatives of $h$, respectively. Our goal is to assess model (1).

### 2.1. Bivariate normal responses

Assume that $y_{1}$ and $y_{2}$ follow a bivariate normal distribution with marginal distributions $N\left(\mu_{1}, \sigma_{1}^{2}\right)$ and $N\left(\mu_{2}, \sigma_{2}^{2}\right)$, respectively. For $j=1,2$, we further model the expected value of $y_{j}$ by a linear regression, i.e., $\mathrm{E}\left(y_{j} \mid \mathbf{x}\right)=\mu_{j}=\mathbf{x}^{\top} \boldsymbol{\gamma}_{j}$ where $\boldsymbol{\gamma}_{j}=\left(\gamma_{0}, \gamma_{1}, \ldots, \gamma_{p}\right)^{\top}$. Denote $\boldsymbol{\eta}=\left(\boldsymbol{\gamma}_{1}^{\top}, \boldsymbol{\gamma}_{2}^{\top}, \sigma_{1}^{2}, \sigma_{2}^{2}\right)^{\top}$ as the collection of all nuisance parameters.

The log-likelihood function is

$$
\begin{equation*}
\ell(\boldsymbol{\beta}, \boldsymbol{\eta})=-\log (2 \pi)-\frac{T_{1}}{2\left(1-\rho^{2}\right)}+\frac{\rho T_{2}}{1-\rho^{2}}-\frac{\log \left(1-\rho^{2}\right)}{2} \tag{2}
\end{equation*}
$$

where

$$
\begin{equation*}
T_{1}=\left(\frac{y_{1}-\mu_{1}}{\sigma_{1}}\right)^{2}+\left(\frac{y_{2}-\mu_{2}}{\sigma_{2}}\right)^{2}, \quad T_{2}=\frac{\left(y_{1}-\mu_{1}\right)\left(y_{2}-\mu_{2}\right)}{\sigma_{1} \sigma_{2}} \tag{3}
\end{equation*}
$$

$\rho=h\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)$, and $\mu_{j}=\mathbf{x}^{\top} \boldsymbol{\gamma}_{j}$. The first and second derivatives of $\ell$ with respect to $\boldsymbol{\beta}$ are, respectively

$$
\begin{equation*}
\boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\eta})=\left\{a(\rho) T_{1}+b(\rho) T_{2}+c(\rho)\right\} \mathbf{x}, \quad H(\boldsymbol{\beta}, \boldsymbol{\eta})=\left\{u(\rho) T_{1}+v(\rho) T_{2}+w(\rho)\right\} \mathbf{x x}^{\top} \tag{4}
\end{equation*}
$$

where $a, b, c, u, v$ and $w$ are given in Table 1. By the fact that $\mathrm{E}\left(T_{1} \mid \mathbf{x}\right)=2$ and $\mathrm{E}\left(T_{2} \mid \mathbf{x}\right)=\rho$, we have $\mathrm{E}(\boldsymbol{s})=\mathbf{0}$, where $\mathbf{0}$ is the zero vector.

Table 1. Some functions under two link functions.

|  | Link 1 | Link 2 |
| :---: | :---: | :---: |
| $h(x)$ | $\left(1+\mathrm{e}^{-\mathrm{x}}\right)^{-1}$ | $\tanh (x)$ |
| $h^{\prime}(x)$ | $h(x)\{1-h(x)\}$ | $\{1+h(x)\}\{1-h(x)\}$ |
| $h^{\prime \prime}(x)$ | $h(x)\{1-h(x)\}\{1-2 h(x)\}$ | $-2 h(x)\{1+h(x)\}\{1-h(x)\}$ |
| $a$ | $-\frac{\rho^{2}}{(1-\rho)(1+\rho)^{2}}$ | $\frac{-\rho}{(1-\rho)(1+\rho)}$ |
| $b$ | $\frac{\left(1+\rho^{2}\right) \rho}{(1-\rho)(1+\rho)^{2}}$ | $\frac{\left(1+\rho^{2}\right)}{(1-\rho)(1+\rho)}$ |
| c | $\frac{\rho^{2}}{1+\rho}$ | $\rho$ |
| $u$ | $\frac{-\rho^{2}\left(2-\rho+\rho^{2}\right)}{(1-\rho)(1+\rho)^{3}}$ | $\frac{-\left(1+\rho^{2}\right)}{(1-\rho)(1+\rho)}$ |
| $v$ | $-\frac{\rho\left(-1+\rho-5 \rho^{2}+\rho^{3}\right)}{(1-\rho)(1+\rho)^{3}}$ | $\frac{4 \rho}{(1-\rho)(1+\rho)}$ |
| w | $\frac{\rho^{2}(1-\rho)(2+\rho)}{(1+\rho)^{2}}$ | $(1+\rho)(1-\rho)$ |

### 2.2. Bivariate binary responses

Assume that both $y_{1}$ and $y_{2}$ are binary responses. Let $\mathrm{E}\left(y_{1}\right)=p_{1}$ and $\mathrm{E}\left(y_{2}\right)=p_{2}$. For $a, b=0$, 1 , let $I_{a b}=I\left(y_{1}=\right.$ $a, y_{2}=b$ ), where $I(\cdot)$ is the indicator function, and $p_{a b}=\mathrm{E}\left(I_{a b}\right)=\mathrm{P}\left(y_{1}=a, y_{2}=b\right)$. Clearly, $p_{00}+p_{01}+p_{10}+$ $p_{11}=1$. By definition, we have

$$
p_{1}=p_{11}+p_{10}, \quad p_{2}=p_{11}+p_{01}, \quad \rho=\frac{p_{11}-p_{1} p_{2}}{\sqrt{p_{1}\left(1-p_{1}\right) p_{2}\left(1-p_{2}\right)}}
$$

Inversely, given $p_{1}, p_{2}$ and $\rho$, we can obtain

$$
\begin{equation*}
p_{11}=p_{1} p_{2}+\rho \sqrt{p_{1}\left(1-p_{1}\right) p_{2}\left(1-p_{2}\right)}, \quad p_{10}=p_{1}-p_{11}, \quad p_{01}=p_{2}-p_{11} \tag{5}
\end{equation*}
$$

For $j=1,2$, we further model the expected value of $y_{j}$ by a logistic regression through $p_{j}=\left\{1+\exp \left(-\mathbf{x}^{\top} \boldsymbol{\gamma}_{j}\right)\right\}^{-1}$, where $\boldsymbol{\gamma}_{j}$ is as defined in Section 2.1. Denote $\boldsymbol{\eta}=\left(\boldsymbol{\gamma}_{1}^{\top}, \boldsymbol{\gamma}_{2}^{\top}\right)^{\top}$. (Through (1) and (5), $p_{a b}$ is a function of $(\boldsymbol{\beta}, \boldsymbol{\eta})$.)

By (5), the log-likelihood is

$$
\ell(\boldsymbol{\beta}, \boldsymbol{\eta})=\sum_{a, b=0,1} I_{a b} \log p_{a b}=\sum_{a, b=0,1} I_{a b} \log \left(c_{a b}+d_{a b} \rho\right)
$$

where

$$
\begin{aligned}
& c_{11}=p_{1} p_{2}, \quad d_{11}=\sqrt{p_{1}\left(1-p_{1}\right) p_{2}\left(1-p_{2}\right)} \\
& c_{10}=p_{1}\left(1-p_{2}\right), \quad d_{10}=-d_{11} \\
& c_{01}=p_{2}\left(1-p_{1}\right), \quad d_{01}=-d_{11} \\
& c_{00}=\left(1-p_{1}\right)\left(1-p_{2}\right), \quad d_{00}=d_{11}
\end{aligned}
$$

Then, the first and second derivatives of $\ell$ with respect to $\boldsymbol{\beta}$ are, respectively

$$
\begin{equation*}
\boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\eta})=\sum_{a, b=0,1} I_{a b} e_{a b}(\rho) \mathbf{x}, \quad H(\boldsymbol{\beta}, \boldsymbol{\eta})=\sum_{a, b=0,1} I_{a b} f_{a b}(\rho) \mathbf{x x}^{\top} \tag{6}
\end{equation*}
$$

where

$$
e_{a b}(\rho)=\frac{d_{a b} h^{\prime}\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)}{c_{a b}+d_{a b} \rho}, \quad f_{a b}(\rho)=\frac{-d_{a b}^{2} h^{\prime 2}\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)+d_{a b}\left(c_{a b}+d_{a b} \rho\right) h^{\prime \prime}\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)}{\left(c_{a b}+d_{a b} \rho\right)^{2}}
$$

The detailed expressions of $e_{a b}$ and $f_{a b}$ under the two models of $h$ are given in Appendix. When $\mathrm{E}\left(I_{a b} \mid \mathbf{x}\right)=p_{a b}$, we have $\mathrm{E}(\boldsymbol{s})=\mathbf{0}$.

## 3. Inference

Let $\left\{\left(y_{1}, y_{2}, x_{1}, \ldots, x_{p}\right): i=1, \ldots, n\right\}$ denote independent samples of $\left(y_{1}, y_{2}, x_{1}, \ldots, x_{p}\right)$ of size $n$. For $i=1, \ldots, n$ and $j=1,2$, denote $\mathbf{x}_{i}=\left(1, x_{i 1}, \ldots, x_{i p}\right)^{\top}, \mathbf{y}_{j}=\left(y_{1 j}, \ldots, y_{n j}\right)^{\top}, \overline{\mathbf{x}}=n^{-1} \sum_{i=1}^{n} \mathbf{x}_{i}$ and $X_{n \times p}=\left(\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}\right)^{\top}$.

The gradient (first derivative) and the Hessian matrix (second derivative) of the log-likelihood over $n$ independent samples with respect to $\boldsymbol{\beta}$ are $\boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\eta})=\sum_{i=1}^{n} \boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ and $H(\boldsymbol{\beta}, \boldsymbol{\eta})=\sum_{i=1}^{n} H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$, respectively, where $\boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ and $H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ are obtained by replacing $\left(y_{1}, y_{2}, \mathbf{x}\right)$ in (4) or (6) by $\left(y_{i 1}, y_{i 2}, \mathbf{x}_{i}\right)$ for the $i$ th subject.

In addition, let $G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})=\partial^{2} \ell_{i} / \partial \boldsymbol{\beta} \partial \boldsymbol{\eta}^{\top}$. Denote $\Theta=\mathrm{E}\left(H_{i}\right)$ and $\mathbf{J}=\mathrm{E}\left(G_{i}\right)$, where the expectation is taken with respect to the joint distribution of $\left(y_{i 1}, y_{i 2}, \mathbf{x}_{i}\right)$. Let $\mathbf{I}=\mathrm{E}\left\{\boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta}) \boldsymbol{s}_{i}^{\top}(\boldsymbol{\beta}, \boldsymbol{\eta})\right\}$ denote the Fisher information matrix with respect to $\boldsymbol{\beta}$. It is noted that unlike the usual regression of the mean value of response, the regularity condition does not hold for the model (1) (Tsiatis, 2006, Theorem 3.2). (It can be easily verified that $\mathbf{I} \neq-\Theta$.)

Let $\widehat{\boldsymbol{\eta}}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\eta}$ obtained from the marginal models. For instance, the marginal MLEs of the nuisance parameters under the bivariate normal response case are $\widehat{\boldsymbol{\gamma}}_{j}=\left(X^{\top} X\right)^{-1} X^{\top} \mathbf{y}_{j}, \widehat{\sigma}_{j}^{2}=$ $n^{-1} \sum_{i=1}^{n}\left(y_{j i}-\widehat{\mu}_{j i}\right)^{2}$, where $\widehat{\mu}_{j i}=\mathbf{x}_{i}^{\top} \widehat{\gamma}_{j}$. By large sample theory, we have $\widehat{\boldsymbol{\eta}} \xrightarrow{p} \eta$ and $\sqrt{n}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\eta}\right)$, where $\Sigma_{\eta}$ is the covariance matrix.

Next, we estimate $\boldsymbol{\beta}$ by using the Newton-Raphson method through iterating

$$
\begin{equation*}
\boldsymbol{\beta}^{(r+1)}=\boldsymbol{\beta}^{(r)}-H^{-1}\left(\boldsymbol{\beta}^{(r)}, \widehat{\boldsymbol{\eta}}\right) \boldsymbol{s}\left(\boldsymbol{\beta}^{(r)}, \widehat{\boldsymbol{\eta}}\right) \tag{7}
\end{equation*}
$$

where $\boldsymbol{\beta}^{(r)}$ is the estimate of $\boldsymbol{\beta}$ at the $r$ th iteration. The convergence of Newton-Raphson method depends on the initial point and the negative definitiveness of the Hessian matrix, which guarantees the (unique) global maximizer of the log-likelihood function. The initial estimate $\boldsymbol{\beta}^{(1)}$ can be chosen such that $h\left(\overline{\mathbf{x}}^{\top} \boldsymbol{\beta}^{(1)}\right)$ is in the range of $\rho$. However, the Hessian matrix involves the plug-in estimators of the nuisance parameters. It is not trivial to show its
negative definitiveness theoretically. Nevertheless, numerical study shows the condition holds with all eigenvalues being negative. Denote the root of the Newton-Raphson method by $\widehat{\boldsymbol{\beta}}$.

Theorem 3.1: Suppose that the conditions of Lemma A. 1 for the bivariate normal responses or those of Lemma A. 2 for the bivariate binary responses hold. (Lemmas A. 1 and A. 2 are given in Appendix.) Suppose the iteration (7) converges. (i) $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, i.e., $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, and (ii) $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, where $\Sigma$ is the covariance matrix given in the proof.

The significance of the overall model is assessed by testing the hypothesis $H_{0}: \beta_{1}=\cdots=\beta_{p}=0$. Denote the Wald statistic $W=\widehat{\boldsymbol{\beta}}_{-0}^{\top} \widehat{\operatorname{cov}}^{-1}\left(\widehat{\boldsymbol{\beta}}_{-0}\right) \widehat{\boldsymbol{\beta}}_{-0}$, where $\widehat{\boldsymbol{\beta}}_{-0}=\left(\widehat{\beta}_{1}, \ldots, \widehat{\beta}_{p}\right)^{\top}$ and $\widehat{\operatorname{cov}}\left(\widehat{\boldsymbol{\beta}}_{-0}\right)$ is the estimate of the covariance of $\widehat{\boldsymbol{\beta}}_{-0}$. By Theorem 3.1, under the null hypothesis, $W$ follows a chi-square distribution of degrees of freedom $p$ asymptotically. The significance of individual covariate $x_{j}(j=1, \ldots, p)$ is assessed by testing the hypothesis $H_{0 j}: \beta_{j}=0$ using the statistic $w_{j}=\widehat{\beta}_{j} / \widehat{\operatorname{se}}\left(\widehat{\beta}_{j}\right)$, where $\widehat{\operatorname{se}}\left(\widehat{\beta}_{j}\right)$ is the estimate of the standard deviation of $\widehat{\beta_{j}}$, which is the square root of the $(j, j)$ th element of $\widehat{\operatorname{cov}}\left(\widehat{\boldsymbol{\beta}}_{-0}\right)$. The null distribution of $w_{j}$ is asymptotically standard normal. Since the exact expression of $\Sigma$ is rather complicated, we use bootstrap to estimate it in practice.

## 4. Simulation study

We conduct simulations to examine the performance of the proposed model for making inference in estimation and testing.

### 4.1. Setup

Unlike possible large numbers of covariates for the mean model, a small number of relevant covariates are usually adequate for modelling the correlation coefficient. Consider the number of covariates $p$ to be two and three, respectively. Set the sample size $n$ to be 100, 200, 400 and 1000, respectively.

For bivariate normal responses, set the marginal variances $\sigma_{1}^{2}=\sigma_{2}^{2}=1$. For both bivariate normal responses and bivariate binary responses, we set both $\boldsymbol{\gamma}_{1}$ and $\boldsymbol{\gamma}_{2}$ to be a vector of $p+1$ zeros for the marginal mean models. It simplifies the data generation without simplifying the estimation procedure. Let $x_{1}, \ldots, x_{p}$ be independent uniform random variables in $(0,1)$.

For the correlation coefficient model, set $\beta_{0}=0.25$. Under the nonnull model, set $\beta_{1}, \ldots, \beta_{p}$ under two link functions as in Table 2, where only the first covariate is significant. Under the null model, simply set all $\beta_{1}, \ldots, \beta_{p}$ to be zeros. We set $(0.25,0, \ldots, 0)$ as the initial values for $\boldsymbol{\beta}$ under both link functions.

Throughout, we set the significant level to be 0.05 for both the global test and individual test, and set the number of replications, $B$, to be 1000 .

### 4.2. Results

First, denote the empirical root of mean square error of $\widehat{\boldsymbol{\beta}}$ by ERMSE $=B^{-1} \sum_{b=1}^{B}\left\{\sum_{j=0}^{p}\left(\widehat{\beta}_{j}^{(b)}-\beta_{j}\right)^{2}\right\}^{1 / 2}$, where $\widehat{\boldsymbol{\beta}}^{(b)}$ is the estimate of $\boldsymbol{\beta}$ in the $b$ th replication. It measures the performance of the estimation in terms of magnitude. Second, denote the directional consistency rate (CR) by CR $=B^{-1} \sum_{b=1}^{B}\left\{n^{-1} \sum_{i=1}^{n} I\left(\widehat{\rho}_{i}^{(b)} \rho_{i}^{(b)}>0\right)\right\}$, where $\rho_{i}^{(b)}$ is the true correlation coefficient in the $b$ th replication and $\widehat{\rho}_{i}^{(b)}=h\left(\mathbf{x}_{i}^{\top} \widehat{\boldsymbol{\beta}}^{(b)}\right)$. It measures the performance of estimation in terms of sign direction.

Table 3 presents the type-I error of the global test and individual tests under the null model. It is seen that the proposed resampling test controls the type-I error well at the nominal level.

Columns 4-5 and 9-10 of Table 4 present the ERMSE and CR under the nonnull model considered in Section 4.1. It is seen that as the sample size increases the ERMSE decreases while the CR increases as expected. The CR ranges from $85 \%$ to $100 \%$ indicating that the proposed model yields a good fit in the sign direction of the correlation coefficient. Columns 7-9 and 12-15 of Table 4 present the powers of the global test and individual tests of the covariates. It is seen that both the global test and the individual test for the significant covariate (i.e., $x_{1}$ ) gain power

Table 2. Specifications of $\beta_{1}, \ldots, \beta_{p}$ under $p=2,3$ and two link functions.

|  | $\left(\beta_{1}, \beta_{2}\right)$ | $\left(\beta_{1}, \beta_{2}, \beta_{3}\right)$ |
| :--- | :---: | :---: |
| Link 1 | $(1,0)$ | $(1,0,0)$ |
| Link 2 | $(-1,0)$ | $(-1,0,0)$ |

Table 3. Type-I errors (in \%) of the global test (for the model significance) and individual test (for the significance of the covariates) under two types of responses, two link functions and various sample sizes.

|  |  |  | $p=2$ |  |  |  | $p=3$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Response | Link | $n$ | Global | $\beta_{1}$ | $\beta_{2}$ |  | Global | $\beta_{1}$ | $\beta_{2}$ | $\beta_{3}$ |
| Normal | 1 | 100 | 2.5 | 2.6 | 2.3 |  | 1.6 | 2.6 | 2.1 | 2.4 |
|  |  | 200 | 4.5 | 5.1 | 3.9 |  | 3.7 | 4.2 | 3.0 | 3.9 |
|  |  | 400 | 5.5 | 4.3 | 5.0 |  | 4.1 | 3.8 | 3.8 | 5.8 |
|  |  | 1000 | 6.0 | 4.4 | 5.4 |  | 5.8 | 5.1 | 5.3 | 5.3 |
|  | 2 | 100 | 5.6 | 4.3 | 5.2 |  | 6.6 | 5.2 | 5.3 | 4.7 |
|  |  | 200 | 5.7 | 3.9 | 6.3 |  | 7.5 | 6.0 | 6.3 | 5.3 |
|  |  | 400 | 6.1 | 6.1 | 5.5 |  | 5.5 | 5.1 | 3.7 | 5.1 |
|  |  | 1000 | 5.7 | 4.6 | 4.6 |  | 4.5 | 4.0 | 5.0 | 4.7 |
| Binary | 1 | 100 | 2.6 | 2.5 | 2.5 |  | 4.6 | 3.5 | 3.5 | 4.4 |
|  |  | 200 | 0.1 | 0.8 | 1.0 |  | 0.1 | 0.5 | 0.6 | 0.4 |
|  |  | 400 | 1.1 | 2.2 | 2.1 |  | 0.3 | 1.8 | 2.9 | 1.4 |
|  |  | 1000 | 5.5 | 5.2 | 6.1 |  | 4.4 | 3.7 | 5.8 | 4.7 |
|  | 2 | 100 | 0.6 | 1.1 | 1.2 |  | 0.5 | 0.9 | 0.8 | 0.7 |
|  |  | 200 | 3.7 | 3.3 | 3.9 |  | 3.3 | 4.1 | 5.0 | 3.7 |
|  |  | 400 | 4.9 | 5.1 | 4.0 |  | 4.9 | 4.8 | 5.1 | 6.3 |
|  |  | 1000 | 5.1 | 4.6 | 4.6 |  | 5.6 | 5.5 | 4.7 | 6.1 |

Table 4. ERMSE, CR and testing power of $\beta^{\prime}$ 's of models considered in Table 1 with various sample sizes.

| Model | Link | $n$ | $p=2$ |  |  |  |  | $p=3$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | RMSE | CR(\%) | Power |  |  | RMSE | CR(\%) | Power |  |  |  |
|  |  |  |  |  | Global | $\beta_{1}$ | $\beta_{2}$ |  |  | Global | $\beta_{1}$ | $\beta_{2}$ | $\beta_{3}$ |
| Normal | 1 | 100 | 1.149 | 100 | 15.0 | 18.7 | 7.8 | 1.296 | 100 | 14.5 | 18.6 | 13.5 | 13.2 |
|  |  | 200 | 0.848 | 100 | 36.2 | 43.5 | 6.0 | 1.039 | 100 | 29.6 | 40.8 | 7.9 | 8.0 |
|  |  | 400 | 0.619 | 100 | 69.4 | 75.8 | 6.3 | 0.737 | 100 | 63.2 | 77.1 | 5.4 | 5.8 |
|  |  | 1000 | 0.373 | 100 | 98.6 | 99.3 | 4.8 | 0.447 | 100 | 97.4 | 99.0 | 4.3 | 5.1 |
|  | 2 | 100 | 0.583 | 87.1 | 69.0 | 77.4 | 4.7 | 0.74 | 85.1 | 60.0 | 68.4 | 4.6 | 3.3 |
|  |  | 200 | 0.401 | 90.5 | 95.5 | 97.5 | 5.4 | 0.493 | 89.1 | 94.0 | 97.3 | 4.6 | 4.8 |
|  |  | 400 | 0.274 | 93.5 | 100 | 100 | 5.3 | 0.336 | 92.3 | 100 | 100 | 5.7 | 5.1 |
|  |  | 1000 | 0.165 | 96.0 | 100 | 100 | 5.9 | 0.207 | 95.3 | 100 | 100 | 5.1 | 4.6 |
| Binary | 1 | 100 | 2.924 | 100 | 6.0 | 7.0 | 4.7 | 3.529 | 100 | 10.5 | 10.1 | 8.7 | 8.8 |
|  |  | 200 | 1.766 | 100 | 2.9 | 5.7 | 1.8 | 2.319 | 100 | 2.8 | 6.3 | 2.0 | 1.3 |
|  |  | 400 | 1.066 | 100 | 7.0 | 16.5 | 1.5 | 1.337 | 100 | 3.1 | 13.8 | 1.2 | 1.2 |
|  |  | 1000 | 0.620 | 100 | 52.5 | 68.8 | 3.8 | 0.772 | 100 | 38.9 | 69.3 | 3.3 | 2.9 |
|  | 2 | 100 | 0.696 | 86.2 | 20.8 | 40.7 | 1.6 | 0.871 | 83.9 | 4.7 | 26.1 | 0.7 | 0.6 |
|  |  | 200 | 0.443 | 89.8 | 90.4 | 94.8 | 3.8 | 0.556 | 88.4 | 80.5 | 94.4 | 4.8 | 4.7 |
|  |  | 400 | 0.308 | 93 | 99.7 | 99.7 | 4.7 | 0.367 | 91.7 | 99.6 | 100 | 4.4 | 4.6 |
|  |  | 1000 | 0.196 | 95.5 | 99.9 | 99.9 | 5.1 | 0.235 | 94.8 | 100 | 100 | 5.0 | 5.0 |

as the sample size increases as desired. The rejection rates for the insignificant covariates (i.e., $x_{2}$ and $x_{3}$ ) are close to the nominal level as desired. We note that the power performance depends on the actual model. In general, we see the proposed method works well under moderate sample size.

### 4.3. Sensitivity analysis

We adopt the simulation model from Wilding et al. (2011) for bivariate normal responses, where $y_{1} \sim N(1+$ $\left.2 x_{1}, 1\right), y_{2} \sim N\left(1.5+1.2 x_{1}, 1\right)$ and their correlation coefficient is given by $\rho\left(y_{1}, y_{2}\right)=2 \Phi\left(0.5+\beta_{1} x_{1}\right)-1$ with $x_{1}$ being a uniform random variable in $(0,1)$. Consider $\beta_{1}$ to be $0,0.5$ and 1 , respectively, to represent the null case and two nonnull cases. Set $n$ to be 30, 60 and 150, respectively.

We compute the rejection rate for testing the significance of $x_{1}$ using the proposed method with link 2 in comparison with Wilding et al. (2011)'s method by which the model is correctly specified. This serves as a sensitivity analysis under model mis-specification. Table 5 reports the rejection rates under the considered cases. It is seen that under the null case when $\beta_{1}=0$ the proposed method controls the type-I error well. Under the nonnull cases when $\beta_{1}=0.5$ and 1 the proposed method is less powerful than Wilding et al.'s method when the sample size is 30 and becomes comparable in power when the sample size gets larger $(\geqslant 60)$.

## 5. Real data application

Our data are taken from the survey of National Health Measurement Study in 2005-2006 (Fryback, 2009). It collects responses to health-related quality of life questionnaires and health status questions through Short Form SF-36 questionnaires from 3844 older adults in the continental United States ( 1641 males and 2203 females).

Table 5. Rejection rates (in \%) of Wilding et al. (2011)'s method and the proposed method for testing the significance of the covariate.

| $n$ | $\beta_{1}$ | Wilding et al. | Proposed |
| :--- | :---: | :---: | :---: |
| 30 | 0 | 6.1 | 4.4 |
|  | 0.5 | 10.6 | 7.2 |
|  | 1 | 31.5 | 13.6 |
| 60 | 0 | 5.2 | 4.8 |
|  | 0.5 | 17.2 | 17.1 |
|  | 1 | 56.2 | 53.6 |
| 150 | 0 | 5.3 | 6.1 |
|  | 0.5 | 43.2 | 41.8 |
|  | 1 | 95.3 | 95.5 |

### 5.1. Bivariate normal responses

First, we illustrate the application of the proposed model to the bivariate normal responses case. The SF-36 items were aggregated into eight health concepts, which were further summarized into the physical component summary (PCS) and the mental component summary (MCS), both in the range of $(0,100)$ with higher scores indicating better physical and mental health functions, respectively [28]. Several medical studies showed that the PCS declined significantly with age, but that the MCS did not change with age (Kim, 2016; Ware Jr et al., 1994). To continue along these lines, we applied the proposed method to further investigate the effect of gender and marital status. We used the Fisher $z$-transformation (i.e., tanh link) to accommodate negative value of correlation. The estimate model for the correlation between PCS and age is found to be

$$
\tanh (\rho(\mathrm{PCS}, \text { age }))=-0.203-0.145 \times \mathrm{MARRIED}+0.018 \times \mathrm{SEX}
$$

where MARRIED is 1 if married or living with a partner and 0 otherwise, and SEX is 1 if female and 0 if male. The corresponding $p$-values for marital status and gender are $<0.001$ and 0.592 , respectively, indicating strong significance of marriage and insignificance of gender. The negative coefficient of MARRIED implies that the status of being married or living with a partner aggravates the declination of PCS with age ( -0.283 for people married or living with a partner and -0.209 for people living by themselves). In other words, the people who are married or living with a partner have relatively lower physical functionality than the people who live by themselves when they age in general. On the other hand, our model concurred with the conclusion of insignificant correlation between MCS and age. In addition, neither gender nor marital status has significant effect on this correlation. (The details are omitted.)

Moreover, we applied the proposed model to study the age effect on the correlation of PCS and MCS. (Throughout this section the observations of age are standardized, denoted by AGE*, before fitting the model as covariate.) Insignificant result is found and supported by the visual inspection of the correlations over 20 age intervals determined by the percentiles (Figure 1(a)). This result is different from the finding of Wilding et al. (2011), who showed a nonlinear downward trend of the correlation coefficient over age based on the 2003-2004 Health Outcome Survey data. When we further include gender and marital status as covariates, there is no significant gender effect


Figure 1. (a) Correlation of PCS and MCS over 20 age groups, (b) correlation of PCS and MCS over 20 age groups with different genders, (c) correlation of PCS and MCS over 20 age groups with different marital status.


Figure 2. (a) Correlation of STROKE and DIABETES over 10 age groups, (b) correlation of STROKE and DIABETES over 10 age groups with different genders, (c) correlation of STROKE and DIABETES over 10 age groups with different marital status.
either ( $p$-value 0.312 ) as also seen in Figure 1(b). However, there appears a very mild effect of marital status ( $p$ value 0.236 ) in the way that the status of being married or living with a partner slightly reduces the correlation from otherwise slight positive to the level of near zero (Figure 1(c)). (The fitted model is $\tanh (\rho$ (PCS, MCS)) $=$ $0.148-0.016 \times$ AGE $^{*}-0.063 \times$ MARRIED $-0.052 \times$ SEX. $)$

### 5.2. Bivariate binary responses

Second, we applied the proposed method to model the correlation of two binary indicators of stroke (STR) and diabetes (DIA) using the covariates of age, gender and marital status. Since the correlation is known positive, the logistic link is used. The estimated model which includes the second-order age effect is

$$
\begin{aligned}
\operatorname{logistic}(\rho(\mathrm{STR}, \mathrm{DIA}))= & -2.050-1.352 \times \mathrm{AGE}^{*}-0.090 \times \mathrm{AGE}^{* 2} \\
& +0.938 \times \operatorname{MARRIED}+0.270 \times \mathrm{SEX}
\end{aligned}
$$

with strong significance of the linear age effect ( $p$-value $<0.001$ ) and marriage effect ( $p$-value 0.028 ) and insignificance of the second-order term of age ( $p$-value 0.786 ) and gender effect ( $p$-value 0.675 ). It is seen from Figure 2(a) that the correlation exhibits a parabola shape in age where the largest value (about 0.25 ) occurs around age 55 and then starts to decline afterwards. This indicates that the association of stroke and diabetes is strongly age dependent. The diabetes becomes a less significant risk factor to stroke after a certain age, e.g., 70 (Kim, 2016). The status of married or living with a partner contributes significantly to the positive correlation ( 0.145 for married vs 0.106 for otherwise) especially for people who are more than 50 years old (Figure 2(c)), which could be explained by the evidence of low physical functionality found before.

## 6. Concluding remarks

We propose two simple regression models of Pearson correlation coefficient of two continuous responses or two binary responses. Likelihood-based inference is developed to estimate the model and test the significance of covariates. Our method for the binary response case is new to the literature and is useful for analysing data when outcomes are observed in binary form (such as yes or no) as seen in many psychological and sociological studies. The proposed method is easy to implement and computationally affordable. (The Newton-Raphson iteration converges in a few steps.) We have made R package available upon request.

It is noted that for the binary response case, the correlation $\rho$ actually has a restricted range given by $\max \left\{-\psi_{1} \psi_{2},-\left(\psi_{1} \psi_{2}\right)^{-1}\right\} \leqslant \rho \leqslant \min \left\{\psi_{1} \psi_{2}^{-1}, \psi_{2} \psi_{1}^{-1}\right\}$, where $\psi_{1}=\sqrt{p_{1} /\left(1-p_{1}\right)}$ and $\psi_{2}=\sqrt{p_{2} /\left(1-p_{2}\right)}$ (Qaqish, 2003). A more precise model than the Fisher $z$-transformation is warranted to further improve the efficiency.

The limitation of the present paper is that the proposed method does not handle the case of mixed responses, i.e., one continuous response and one binary response, or more generally an ordinal response categorized from a latent variable, such as 'mild', 'moderate', 'severe'. Regression model of correlation involving latent variable(s) is worth investigation. We will report the result in a separate work.

## Disclosure statement

All authors declare no conflict of interest.

## Data availability statement

The data that support the findings of this study are available from the corresponding author upon request.

## ORCID

Jin Xu (©) http://orcid.org/0000-0003-2434-7183

## References

Anderson, T. (2003). An introduction to statistical multivariate analysis. 3rd ed. Wiley.
Bartlett, R. F. (1993). Linear modelling of Pearson's product moment correlation coefficient: An application of Fisher's $z$-transformation. Journal of the Royal Statistical Society: Series D, 42(1), 45-53.
Fryback, D. G. (2009). United States National Health Measurement Study, 2005-2006. Inter-University Consortium for Political and Social Research.
Kim, D. (2016). Correlation between physical function, cognitive function, and health-related quality of life in elderly persons. Journal of Physical Therapy Science, 28(6), 1844-1848. https://doi.org/10.1589/jpts.28.1844
Liu, Q., Li, C., Wanga, V., \& Shepherd, B. E. (2018). Covariate-adjusted Spearman rank correlation with probability-scale residuals. Biometrics, 74(2), 595-605. https://doi.org/10.1111/biom.v74.2
Newey, W. K., \& McFadden, D. (1994). Large sample estimation and hypothesis testing. Handbook of Econometrics, 4, 2111-2245. https://doi.org/10.1016/S1573-4412(05)80005-4
Paul, S. (1989). Test for the equality of several correlation coefficients. Canadian Journal of Statistics, 17(2), 217-227. https://doi.org/10.2307/3314850
Qaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. Biometrika, 90(2), 455-463. https://doi.org/10.1093/biomet/90.2.455
Thomas, P., Jolanda, L., Antonius, J. M. d. C., Joris, P. J. S., \& Rudi, G. J. W. (2016). Impact of physical and mental health on life satisfaction in old age: A population based observational study. BMC Geriatrics, 16(1), 194. https://doi.org/10.1186/s12877-016-0365-4
Tsiatis, A. A. (2006). Semiparametric theory and missing data. Springer.
Ware Jr, J. E., Kosinski, M., \& Keller, S. D. (1994). SF-36 physical and mental summary scales: A user's manual. The Health Institute, New England Medical Center.
Wilding, G. E., Cai, X., Hutson, A., \& Yu, Z. (2011). A linear model-based test for the heterogeneity of conditional correlations. Journal of Applied Statistics, 38(10), 2355-2366. https://doi.org/10.1080/02664763.2011.559201
Yu, M., \& Dunn, O. (1982). Robust tests for the equality of two correlation coefficients: A monte carlo study. Educational and Psychological Measurement, 42(4), 987-1004. https://doi.org/10.1177/001316448204200407

## Appendix. Technical details

## A. 1 Details of $e_{a b}$ and $f_{a b}$

The expressions of $e_{a b}$ and $f_{a b}$ under two link functions of $h$ are, respectively

$$
e_{a b}(\rho)= \begin{cases}\frac{d_{a b} \rho(1-\rho)}{c_{a b}+d_{a b} \rho}, & \text { link 1, } \\ \frac{d_{a b}(1+\rho)(1-\rho)}{c_{a b}+d_{a b} \rho}, & \text { link 2, }\end{cases}
$$

and

$$
f_{a b}(\rho)= \begin{cases}\frac{\left\{-d_{a b}^{2}+d_{a b}\left(c_{a b}+d_{a b} \rho\right)(1-2 \rho)\right\} \rho^{2}(1-\rho)^{2}}{\left(c_{a b}+d_{a b} \rho\right)^{2}}, & \text { link } 1, \\ \frac{\left\{-d_{a b}^{2}-d_{a b}\left(c_{a b}+d_{a b} \rho\right) 2 \rho\right\}(1+\rho)^{2}(1-\rho)^{2}}{\left(c_{a b}+d_{a b} \rho\right)^{2}}, & \text { link } 2 .\end{cases}
$$

## A. 2 Lemmas

Lemma A.1: Assume that under the bivariate normal response model in Section 2.1, there exist constants $c_{1}, c_{2}, c_{3}$, and $c_{4}$ such that $0<c_{1} \leqslant|\rho| \leqslant c_{2}<1,0 \leqslant\left|\widehat{\mu}_{j}\right|<c_{3},\left|\widehat{\sigma}_{j}\right| \geqslant c_{4}>0$ for $j=1$, 2. Then, (i) $n^{-1} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right) \xrightarrow{p} \Theta$, and (ii) $n^{-1} \sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right) \xrightarrow{p} \mathbf{J}$.

Proof: Note that we cannot apply the weak law of large number as $\ell_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right)$ s and $\ell_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right)$ s are neither independent nor identically distributed.
(i) Let $\widehat{T}_{1 i}=\left(\frac{y_{1 i} i}{\widehat{\mu}_{1 i}}\right)^{2}+\left(\frac{y_{2 i}-\widehat{\mu}_{2 i}}{\widehat{\sigma}_{2}}\right)^{2}$ and $\widehat{T}_{2 i}=\frac{\left(y_{1 i}-\widehat{\mu}_{1 i 1}\right)\left(y_{2 i}-\widehat{\mu}_{2 i}\right)}{\widehat{\sigma}_{1} \widehat{\sigma}_{2}}$. First, by the assumptions of $0 \leqslant\left|\widehat{\mu}_{j}\right|<c_{3}$ and $\left|\widehat{\sigma}_{j}\right| \geqslant c_{4}>0$ for $j=1,2$, we have

$$
\begin{aligned}
& \left|\widehat{T}_{1 i}\right|=\left|\left(\frac{y_{1 i}-\widehat{\mu}_{1}}{\widehat{\sigma}_{1}}\right)^{2}+\left(\frac{y_{2 i}-\widehat{\mu}_{2}}{\widehat{\sigma}_{2}}\right)^{2}\right| \leqslant \frac{y_{1 i}^{2}+c_{3}^{2}+2 c_{3}\left|y_{1 i}\right|}{c_{4}^{2}}+\frac{y_{2 i}^{2}+c_{3}^{2}+2 c_{3}\left|y_{2 i}\right|}{c_{4}^{2}} ; \\
& \left|\widehat{T}_{2 i}\right|=\left|\frac{\left(y_{1 i}-\widehat{\mu}_{1}\right)\left(y_{2 i}-\widehat{\mu}_{2}\right)}{\widehat{\sigma}_{1} \widehat{\sigma}_{2}}\right| \leqslant \frac{\left(\left|y_{1 i}\right|+c_{3}\right)\left(\left|y_{2 i}\right|+c_{3}\right)}{c_{4}^{2}}
\end{aligned}
$$

Second, by the assumption of $0<c_{1} \leqslant|\rho| \leqslant c_{2}<1$, there exists a positive constant $M_{1}$ such that $|u(\rho)|,|v(\rho)|$, and $|w(\rho)|$ are all bounded by $M_{1}$. Since $H_{i}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})=\left\{u\left(\rho_{i}\right) \widehat{T}_{1 i}+v\left(\rho_{i}\right) \widehat{T}_{2 i}+w\left(\rho_{i}\right)\right\} \mathbf{x}_{i} \mathbf{x}_{i}^{\top}$, then, there exists a function $g(\mathbf{x}, y)$ such that $\left\|H_{i}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})\right\| \leqslant g\left(\mathbf{x}_{i}, y_{i}\right)$ and $\mathrm{E}\left\{g\left(\mathbf{x}_{i}, y_{i}\right)\right\} \leqslant \infty$, where $\|\cdot\|$ is the Euclidean norm (i.e., $\left.\|A\|=\left\{\operatorname{tr}\left(A A^{\top}\right)\right\}^{1 / 2}\right)$. Clearly, $H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$.

By Lemma 2.4 of Newey and McFadden (1994), we have $\sup _{\boldsymbol{\beta}, \eta}\left\|n^{-1} \sum_{i=1}^{n} H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})-\Theta\right\| \xrightarrow{p} 0$. Since $\boldsymbol{\beta}^{*} \xrightarrow{p} \boldsymbol{\beta}$ (because $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ and $\boldsymbol{\beta}^{*}$ is between $\boldsymbol{\beta}$ and $\left.\widehat{\boldsymbol{\beta}}\right), \widehat{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}$ and $\Theta$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$, by Tsiatis (2006, Section 3.2), the result of (i) follows.
(ii) By (2) and (3), we have

$$
\begin{aligned}
& \frac{\partial T_{1 i}}{\partial \boldsymbol{\gamma}_{1}^{\top}}=-\frac{2\left(y_{1 i}-\mu_{1}\right)}{\sigma_{1}^{2}} \mathbf{x}_{i}^{\top}, \quad \frac{\partial T_{2 i}}{\partial \boldsymbol{\gamma}_{1}^{\top}}=-\frac{y_{2 i}-\mu_{2}}{\sigma_{1} \sigma_{2}} \mathbf{x}_{i}^{\top} \\
& \frac{\partial T_{1 i}}{\partial \boldsymbol{\gamma}_{2}^{\top}}=-\frac{2\left(y_{2 i}-\mu_{2}\right)}{\sigma_{2}^{2}} \mathbf{x}_{i}^{\top}, \quad \frac{\partial T_{2 i}}{\partial \boldsymbol{\gamma}_{2}^{\top}}=-\frac{y_{1 i}-\mu_{1}}{\sigma_{1} \sigma_{2}} \mathbf{x}_{i}^{\top}, \\
& \frac{\partial T_{1 i}}{\partial \sigma_{1}^{2}}=-\frac{\left(y_{1 i}-\mu_{1}\right)^{2}}{\sigma_{1}^{4}}, \quad \frac{\partial T_{2 i}}{\partial \sigma_{1}^{2}}=-\frac{\left(y_{1 i}-\mu_{1}\right)\left(y_{2 i}-\mu_{2}\right)}{2 \sigma_{1}^{3} \sigma_{2}}, \\
& \frac{\partial T_{1 i}}{\partial \sigma_{2}^{2}}=-\frac{\left(y_{2 i}-\mu_{2}\right)^{2}}{\sigma_{2}^{4}}, \quad \frac{\partial T_{2 i}}{\partial \sigma_{2}^{2}}=-\frac{\left(y_{1 i}-\mu_{1}\right)\left(y_{2 i}-\mu_{2}\right)}{2 \sigma_{1} \sigma_{2}^{3}},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^{2} \ell_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_{1}^{\top}} & =\left\{-a(\rho) \frac{2\left(y_{1 i}-\mu_{1}\right)}{\sigma_{1}^{2}}-b(\rho) \frac{y_{2 i}-\mu_{2}}{\sigma_{1} \sigma_{2}}\right\} \mathbf{x}_{i} \mathbf{x}_{i}^{\top}, \\
\frac{\partial^{2} \ell_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_{2}^{\top}} & =\left\{a(\rho) \frac{2\left(y_{2 i}-\mu_{2}\right)}{\sigma_{2}^{2}}-b(\rho) \frac{y_{1 i}-\mu_{1}}{\sigma_{1} \sigma_{2}}\right\} \mathbf{x}_{i} \mathbf{x}_{i}^{\top}, \\
\frac{\partial^{2} \ell_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta} \partial \sigma_{1}^{2}} & =\left\{a(\rho) \frac{\left(y_{1 i}-\mu_{1}\right)^{2}}{\sigma_{1}^{4}}-b(\rho) \frac{\left(y_{1 i}-\mu_{1}\right)\left(y_{2 i}-\mu_{2}\right)}{2 \sigma_{1}^{3} \sigma_{2}}\right\} \mathbf{x}_{i}, \\
\frac{\partial^{2} \ell_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\beta} \partial \sigma_{2}^{2}} & =\left\{-a(\rho) \frac{\left(y_{2 i}-\mu_{2}\right)^{2}}{\sigma_{2}^{4}}-b(\rho) \frac{\left(y_{1 i}-\mu_{1}\right)\left(y_{2 i}-\mu_{2}\right)}{2 \sigma_{1} \sigma_{2}^{3}}\right\} \mathbf{x}_{i} .
\end{aligned}
$$

By the assumptions, we can find a function $k(\mathbf{x}, \boldsymbol{y})$ such that $\left\|D_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})\right\| \leqslant k\left(\mathbf{x}_{i}, y_{i}\right)$ and $\mathrm{E}\left\{\left|k\left(\mathbf{x}_{i}, y_{i}\right)\right|\right\}<\infty$. And $D_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$. Using the similar argument as in the proof for (i), by Lemma 2.4 of Newey and McFadden (1994) $\sup _{\boldsymbol{\beta}, \boldsymbol{\eta}}\left\|n^{-1} \sum_{i=1}^{n} G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})-\mathbf{J}\right\| \xrightarrow{p} 0$. Since $\boldsymbol{\eta}^{*} \xrightarrow{p} \boldsymbol{\eta}$ and $\mathbf{J}$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$, the result of (ii) follows.

Lemma A.2: Assume that under the bivariate Bernoulli response model in Section 2.2 there exist constants $d_{1}, \ldots, d_{6}$ such that $\left|h^{\prime}\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)\right| \leqslant d_{1},\left|h^{\prime \prime}\left(\mathbf{x}^{\top} \boldsymbol{\beta}\right)\right| \leqslant d_{2}$ for all $\mathbf{x}, 0<d_{3}<p_{j}<d_{4}<1$, for $j=1,2,0<d_{5}<|\rho|<d_{6}<1$. Then, (i) $n^{-1} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right) \xrightarrow{p} \Theta$, and (ii) $n^{-1} \sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right) \xrightarrow{p} \mathbf{J}$.

Proof: (i) By the assumptions, there exists $M_{2}$ such that $\left|f_{\text {iab }}(\boldsymbol{\beta}, \boldsymbol{\eta})\right| \leqslant M_{2}$ for all $a, b=0,1$ and $i=1, \ldots, n$. Then, there exists a function $g(\mathbf{x}, y)$ such that $\left\|H_{i}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})\right\| \leqslant g_{1}\left(\mathbf{x}_{i}, y_{i}\right)$ and $\mathrm{E}\left\{g_{1}\left(\mathbf{x}_{i}, y_{i}\right)\right\} \leqslant \infty$. Clearly, $H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$. By Lemma 2.4 of Newey and McFadden (1994), $\sup _{\boldsymbol{\beta}, \eta}\left\|n^{-1} \sum_{i=1}^{n} H_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})-\Theta\right\| \xrightarrow{p} 0$.

Since $\boldsymbol{\beta}^{*} \xrightarrow{p} \boldsymbol{\beta}, \widehat{\boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{\eta}$ and $\Theta$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$, by Tsiatis (2006) Section 3.2, the result of (i) follows.
(ii) Observe that

$$
\begin{aligned}
G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta}) & =\sum_{a, b=0,1} I_{i a b} \frac{\partial e_{i a b}(\boldsymbol{\beta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\top}} \mathbf{x}_{i} \\
& =\sum_{a, b=0,1} I_{i a b} \frac{\frac{\partial d_{i a b}}{\partial \boldsymbol{\eta}^{\top}}\left(c_{i a b}+d_{i a b} \rho_{i}\right)-\left(\frac{\partial c_{i a b}}{\partial \boldsymbol{\eta}^{\top}}-\frac{\partial d_{i a b}}{\partial \boldsymbol{\eta}^{\top}} \rho_{i}\right) d_{i a b}}{\left(c_{i a b}+d_{i a b} \rho_{i}\right)^{2}} h^{\prime}\left(\mathbf{x}_{i}^{\top} \boldsymbol{\beta}\right) \mathbf{x}_{i} .
\end{aligned}
$$

By the logistic model defined in Section 2.2, $\partial p_{j i} / \partial \boldsymbol{\eta}_{j}^{\top}=p_{j}\left(1-p_{j}\right) \mathbf{x}_{i}^{\top}$. The expressions of $\partial d_{i a b} / \partial \boldsymbol{\eta}^{\top}$ and $\partial c_{i a b} / \partial \boldsymbol{\eta}^{\top}$ are obtained as follows.

$$
\begin{aligned}
& \frac{\partial d_{i 11}}{\partial \boldsymbol{\eta}_{j}^{\top}}=\left\{p_{1 i}\left(1-p_{1 i}\right) p_{2 i}\left(1-p_{2 i}\right)\right\}^{1 / 2}\left(\frac{1}{2}-p_{j i}\right) \mathbf{x}_{i}^{\top}, \quad j=1,2 ; \\
& \frac{\partial d_{i 10}}{\partial \boldsymbol{\eta}_{j}^{\top}}=\frac{\partial d_{i 01}}{\partial \boldsymbol{\eta}_{j}^{\top}}=-\frac{\partial d_{i 11}}{\partial \boldsymbol{\eta}_{j}^{\top}}, \quad \frac{\partial d_{i 00}}{\partial \boldsymbol{\eta}_{j}^{\top}}=\frac{\partial d_{i 11}}{\partial \boldsymbol{\eta}_{j}^{\top}}, \quad j=1,2 ; \\
& \frac{\partial c_{i 11}}{\partial \boldsymbol{\eta}_{j}^{\top}}=p_{1 i} p_{2 i}\left(1-p_{j i}\right) \mathbf{x}_{i}^{\top}, \quad j=1,2 ; \\
& \frac{\partial c_{i 10}}{\partial \boldsymbol{\eta}_{1}^{\top}}=\frac{\partial p_{1 i}}{\partial \boldsymbol{\eta}_{1}^{\top}}-p_{2 i} \frac{\partial p_{1 i}}{\partial \boldsymbol{\eta}_{1}^{\top}}=p_{1 i}\left(1-p_{1 i}\right)\left(1-p_{2 i}\right) \mathbf{x}_{i}^{\top} ; \\
& \frac{\partial c_{i 10}}{\partial \boldsymbol{\eta}_{2}^{\top}}=-p_{1 i} \frac{\partial p_{2 i}}{\partial \boldsymbol{\eta}_{2}^{\top}}=-p_{1 i} p_{2 i}\left(1-p_{2 i}\right) \mathbf{x}_{i}^{\top} ; \\
& \frac{\partial c_{i 01}}{\partial \boldsymbol{\eta}_{1}^{\top}}=-p_{2 i} \frac{\partial p_{1 i}}{\partial \boldsymbol{\eta}_{1}^{\top}}=-p_{2 i} p_{1 i}\left(1-p_{1 i}\right) \mathbf{x}_{i}^{\top} ; \\
& \frac{\partial c_{i 01}}{\partial \boldsymbol{\eta}_{2}^{\top}}=\frac{\partial p_{2 i}}{\partial \boldsymbol{\eta}_{2}^{\top}}-p_{1 i} \frac{\partial p_{2 i}}{\partial \boldsymbol{\eta}_{2}^{\top}}=p_{2 i}\left(1-p_{2 i}\right)\left(1-p_{1 i}\right) \mathbf{x}_{i}^{\top} ; \\
& \frac{\partial c_{i 00}}{\partial \boldsymbol{\eta}_{j}^{\top}}=-p_{j i}\left(1-p_{1 i}\right)\left(1-p_{2 i}\right) \mathbf{x}_{i}^{\top}, \quad j=1,2 .
\end{aligned}
$$

By the assumptions, we can find a function $\omega(x, y)$ such that $\left\|G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})\right\| \leqslant \omega\left(\mathbf{x}_{i}, y_{i}\right)$ and $\mathrm{E}\left\{\omega\left(\mathbf{x}_{i}, y_{i}\right)\right\}<\infty$. And, $G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$. Using the similar argument as in the proof for (i), by Lemma 2.4 of Newey and McFadden (1994) $\sup _{\boldsymbol{\beta}, \boldsymbol{\eta}}\left\|n^{-1} \sum_{i=1}^{n} G_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})-\mathbf{J}\right\| \xrightarrow{p} 0$. Since $\boldsymbol{\eta}^{*} \xrightarrow{p} \eta$ in probability and $\mathbf{J}$ is continuous in $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\eta}^{\top}\right)^{\top}$, the result of (ii) follows.

## A. 3 Proof of Theorem 3.1

By the standard theory of regression, we have $\sqrt{n}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}) \rightarrow N\left(0, \Sigma_{\eta}\right)$.
By the fact that $\widehat{\boldsymbol{\beta}}$ solves $\boldsymbol{s}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})=\mathbf{0}$, expand $\boldsymbol{s}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}})$ (with respect to $\boldsymbol{\beta}$ ) around $\boldsymbol{\beta}$ as

$$
\begin{equation*}
0=\boldsymbol{s}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}})=\boldsymbol{s}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})+\sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}), \tag{A1}
\end{equation*}
$$

where $\boldsymbol{\beta}^{*}$ is between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ component-wise.
Second, expand $\boldsymbol{s}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})$ (with respect to $\boldsymbol{\eta}$ ) around $\boldsymbol{\eta}$ as

$$
\begin{equation*}
\boldsymbol{s}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})=\sum_{i=1}^{n} \boldsymbol{s}_{i}(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}})=\sum_{i=1}^{n} \boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})+\sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right)(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}), \tag{A2}
\end{equation*}
$$

where $\eta^{*}$ is between $\widehat{\eta}$ and $\eta$ component-wise.
Combining (A1) and (A2), $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is expressed as

$$
\begin{equation*}
-\sqrt{n}\left\{\frac{1}{n} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right)\right\}^{-1}\left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta})+\left\{\frac{1}{n} \sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right)\right\}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\right] . \tag{A3}
\end{equation*}
$$

By Lemma A. 1 for the bivariate normal response case or Lemma A. 2 for the bivariate Bernoulli response case,

$$
\begin{equation*}
\frac{1}{n} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right) \xrightarrow{p} \Theta \text { and } \frac{1}{n} \sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right) \xrightarrow{p} \mathbf{J} . \tag{A4}
\end{equation*}
$$

And by the central limit theorem, $n^{-1 / 2} \sum_{i=1}^{n} \boldsymbol{s}_{i}(\boldsymbol{\beta}, \boldsymbol{\eta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$. Then, by Slusky theorem,

$$
\begin{equation*}
-\left\{\frac{1}{n} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right)\right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_{i}(\boldsymbol{\beta}, \boldsymbol{\eta}) \xrightarrow{d} N\left(\mathbf{0}, \Theta^{-1} \mathbf{I} \Theta^{-1}\right), \tag{A5}
\end{equation*}
$$

and

$$
\begin{equation*}
-\left\{\frac{1}{n} \sum_{i=1}^{n} H_{i}\left(\boldsymbol{\beta}^{*}, \widehat{\boldsymbol{\eta}}\right)\right\}^{-1}\left\{\frac{1}{n} \sum_{i=1}^{n} G_{i}\left(\boldsymbol{\beta}, \boldsymbol{\eta}^{*}\right)\right\} \sqrt{n}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}) \xrightarrow{d} N\left(\mathbf{0}, \Theta^{-1} \mathbf{J} \Sigma_{\eta} \mathbf{J}^{\top} \Theta^{-1}\right) \tag{A6}
\end{equation*}
$$

Combining (A3), (A5), and (A6), $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}$ ) is asymptotically normal with zero mean vector and covariance matrix given by

$$
\Sigma=\Theta^{-1} \mathbf{I} \Theta^{-1}+\Theta^{-1} \mathbf{J} \Sigma_{\eta} \mathbf{J}^{\top} \Theta^{-1}+\operatorname{cov}\{\text { LHS of (A5), LHS of (A6) }\} .
$$


[^0]:    View Crossmark data

[^1]:    CONTACT Jin Xu jxu@stat.ecnu.edu.cn School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China; Key Laboratory of Advanced Theory and Application in Statistics and Data Science - MOE, East China Normal University, Shanghai 200062, People's Republic of China
    *These authors contributed equally.

