



The Behaviour of the Dispersion Matrix of the Information Matrix Test under the Wrong Logistic Regression Model

Nuri H. Salem Badi

Faculty of Science, Statistical Department, University of Benghazi, Benghazi, Libya
Email: nuri.badi@uob.edu.ly

How to cite this paper: Badi, N.H.S. (2021) The Behaviour of the Dispersion Matrix of the Information Matrix Test under the Wrong Logistic Regression Model e. *Open Access Library Journal*, 8: e7183.
<https://doi.org/10.4236/oalib.1107183>

Received: January 26, 2021

Accepted: February 23, 2021

Published: February 26, 2021

Copyright © 2021 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Information Matrix Tests (IMT) considers as one of the important global goodness of fit test. The IMT provides a unified framework for specification goodness of fit tests for a wide variety of distribution, multivariate or univariate, discrete or continuous. Many researchers discussed the IMT in cases of the outcome covariate is a continuous variable which reported it has reasonable behaviour. This article considers using IMT as a goodness of fit test for the logistic regression mode, to investigate the behaviour of this statistic under the wrong model. Moreover, we are interested to examine the behaviour of the dispersion matrix under wrong logistic model and compute alternative formula of variance, empirical variance of IMT and examine it by simulation.

Subject Areas

Applied Statistical Mathematics

Keywords

Logistic Regression Model, Goodness of Fit Test, Information Matrix Test, Estimation of Parameters

1. Introduction

The IMT is a test for general misspecification, produced by [1] who pointed out that the properties of the Maximum likelihood estimator and the information matrix can be exploited to yield a family of useful tests for model mis-specification. The idea of the IMT is to compare two different estimators of the information matrix to assess model fit. The IMT is based on the information matrix equality that obtains when the model specification is correct. This equality implies the

asymptotic equivalence of the Hessian and the score forms of Fisher's information matrix [2]. As [1], points out, the IMT is designed to detect the failure of this equality and the failure implies the model misspecification. [3] discussed the information matrix test and showed that it is useful with binary data models. Many researchers, [4] [5] and [6] pointed out the behaviour of the asymptotic distribution of IMT statistic and dispersion matrix. The idea of the information matrix test is to compare $E\left(\frac{-\partial^2 \ell}{\partial \theta \partial \theta^T}\right)$ and $E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^T}\right)$, as these differ when the model is mis-specified but not when the model is correct. [7], pointed out, can be estimated the covariance matrix of IMT, dependent upon the IMT of [1], which can be estimated without the computation of analytic third derivatives of the density function. [4], discussed that, the IMT is sensitive to non-normality. Moreover, he proposed a simple computation procedure which employs the Outer Product of the Gradient (OPG) covariance matrix estimator of IMT statistic. However, [5] argue that, such a procedure maybe give unreliable inferences, related to the stochastic nature of the covariance matrix estimator which uses high sample moments to estimate high population moments. [6] purposed a simple calculation procedure for the test statistic, for general binary data models, which employs the ML covariance matrix estimator instead the OPG estimator. Moreover, [8], computed and examined IMT and found it had good power for logistic model.

Basic Idea of the IMT

Let us consider the density function $f(x_i, \theta)$ for individual observation and the data are independent, identically distribution so we have

$$\int f(x|\theta) dx = 1$$

and we consider $\ell(\theta) = \log f(x, \theta)$ to be the logarithm of a density function of x dependent upon p parameters θ , so the log-likelihood function in this case is

$$\ell_n(\theta) = \sum_{i=1}^n \log f(x_i, \theta)$$

Now, as we defined the idea of the IMT to compare two different matrix of expected the first and second partial derivatives of the $\ell_n(\theta)$, we have

$$\frac{\partial \ell}{\partial \theta} = \int \frac{\partial f(x|\theta)}{\partial \theta} dx = \int \frac{\partial \log f(x|\theta)}{\partial \theta} f(x|\theta) dx = E\left(\frac{\partial \log(f(x|\theta))}{\partial \theta}\right) = 0 \quad (1)$$

So, according to the ML method, we have

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = 0.$$

Differentiating (1) again we get

$$0 = \int \frac{\partial^2 \log f(x|\theta)}{\partial \theta \partial \theta^T} f(x|\theta) dx + \int \frac{\partial \log f(x|\theta)}{\partial \theta} \frac{\partial \log f(x|\theta)}{\partial \theta^T} f(x|\theta) dx \quad (2)$$

So

$$E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right) + E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^T}\right) = 0. \tag{3}$$

When the model is mis-specified, the above quantity will be not necessarily equal zero.

Asymptotic Distribution of $\hat{\theta}$

The asymptotic distribution of estimated parameters and the behaviour of the MLE under the wrong model discussed by [9] and more investigated considered by [10]. [11], pointed out the estimation the parameters of a given regression model. In the limit for each value of the parameter vector θ ,

$$n^{-1} \ell_n(\theta) \rightarrow \int g(Y) \log f(Y|\theta) dY = E(\log f(Y|\theta))$$

where $g(Y)$ denoted to the true model and $f(Y|\theta)$ is the fitted model. Also, consider the Kullback-Leibler divergence (KL) from the true to the approximating model conditional on X , under the wrong model. In this case $\hat{\theta} \rightarrow \theta^*$, where θ^* is the least false value (LF). Note that the least false value θ^* minimizes the KL divergence, because the derivative of the KL is

$$E\left(\frac{\partial \log f(Y,\theta)}{\partial \theta}\right) = \int g(Y) \frac{\partial \log f(Y,\theta)}{\partial \theta} dY = 0.$$

Also, if we need define

$$J = -E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right)$$

and

$$K = \text{var}\left(\frac{\partial \log f(Y,\theta)}{\partial \theta}\right) = E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^T}\right)$$

these matrixes are identical when $g(Y) = \frac{\partial \log f(Y,\theta)}{\partial \theta}$ for all Y . As explained in [11], the distribution of the $\hat{\theta}$, in this case from the central limit theorem there is convergence in distribution

$$\sqrt{n} \bar{U}_n \rightarrow U' \sim N_p(0, K)$$

where, $\bar{U} = n^{-1} \sum_{i=1}^n u(Y_i, \theta^*)$, which is leads to

$$\sqrt{n}(\hat{\theta} - \theta^*) \rightarrow J^{-1} U' \sim N_p(0, J^{-1} K J^{-1}).$$

So, we can say, the asymptotic MLE distribution under the null hypotheses H_0 , in this case

$$\sqrt{n} \hat{\theta} \sim N(\theta_0, J^{-1})$$

where, θ_0 is the true value. And the asymptotic distribution of $\hat{\theta}$ under alternative hypotheses H_1 is

$$\sqrt{n} \hat{\theta} \sim N(\theta^*, J^{-1} K J^{-1})$$

So, that is meaning ($J = K$) if and only if when fitted the correct model (*i.e.* under H_0).

2. The IMT under Missing Covariates for Logistic Regression Model

In this part, we apply the procedure of the IMT statistic under missing covariates for a logistic regression model. If X_i is a p -dimensional vector of covariates draw from normal distribution and Y_i is binary with

$$P(Y_i = 1 | X_i) = \text{expit}(\alpha + \beta^T X_i). \tag{4}$$

In the following we treat the simple case where the fitted model is

$$P(Y_i = 1 | X_i) = \text{expit}(\alpha + \beta_1 X_{1i}) \tag{5}$$

for a scalar X_1 and that the true model has

$$P(Y_i = 1 | X_i) = \text{expit}(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i}), \tag{6}$$

where X_2 is also a scalar. We have the log-likelihood function contribution for the i^{th} element (Y_i, X_i) is

$$\ell(Y_i, X_i) = Y_i(\alpha + \beta^T X_i) - \log(1 + \exp(\alpha + \beta^T X_i)) \tag{7}$$

and so,

$$\frac{\partial \ell_i}{\partial \alpha} = Y_i - \pi_i; \quad \frac{\partial \ell_i}{\partial \beta_1} = (Y_i - \pi_i) X_{1i}$$

and note that we only consider fitting the model with X_1 , even if the true model also includes X_2 (*i.e.* $\beta_2 \neq 0$). From this we get:

$$\frac{\partial^2 \ell_i}{\partial \theta \partial \theta^T} = \begin{bmatrix} -\pi_i(1-\pi_i) & -\pi_i(1-\pi_i)X_i \\ -\pi_i(1-\pi_i)X_i & -\pi_i(1-\pi_i)X_i^2 \end{bmatrix}$$

Also,

$$\frac{\partial \ell_i}{\partial \theta} \frac{\partial \ell_i}{\partial \theta^T} = \begin{bmatrix} (Y_i - \pi_i)^2 & (Y_i - \pi_i)^2 X_i \\ (Y_i - \pi_i)^2 X_i & (Y_i - \pi_i)^2 X_i^2 \end{bmatrix}$$

using,

$$(Y_i - \pi_i)^2 - \pi_i(1-\pi_i) = (Y_i - \pi_i)(1-2\pi_i),$$

as Y_i^2 is Y_i , and so we get that

$$d_g(y_i, \theta) = (Y_i - \pi_i)(1-2\pi_i) \begin{bmatrix} 1 \\ X_i \\ X_i^2 \end{bmatrix}. \tag{8}$$

3. An Alternative Formulae of Variance

In this part we are interested to find a formulae of the variance of d statistic, even when the model is mis-specified. To perform the IMT we need to find the mean and variance of

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^n d_{gi}$$

Under H_0 $E(d_{gi}) = 0$, and so the IMT could be written as

$$T^T \text{var}(T)^{-1} T$$

which will have a χ^2 -distribution on $\text{rank}(\text{var}(T))$ d.f. as T is asymptotically Normal. However, the test statistic has to be evaluated at the MLE $\hat{\theta}$ and this introduces a complication. The MLE $\hat{\theta}$ is the solution to

$$S = \frac{1}{\sqrt{n}} \nabla \ell = \frac{1}{\sqrt{n}} \sum_{i=1}^n \nabla \ell_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n (y_i - \pi_i) \begin{bmatrix} 1 \\ X_i \end{bmatrix} = 0.$$

The expression for T is

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^n (y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}$$

and this is clearly going to be highly correlation with S . Therefore, the appropriate variance for the IMT is $\text{var}(T | S = 0)$. As T and S are sums of independent elements, the Central limit Theorem implies that $(T, S)^T$ is asymptotically Normal and so we can use

$$\text{var}(T | S = 0) = \text{var}(T) - \text{cov}(T, S) \text{var}(S)^{-1} \text{cov}(T, S)^T. \tag{9}$$

To work out $\text{var}(T | S = 0)$, so, in this case we can write

$$\text{var}(T) = \text{var}([\![d_{g1} + d_{g2} + \dots + d_{gn}]\!] / \sqrt{n}) = \text{var}(d_{g1}),$$

and similarly

$$\text{var}(S) = \text{var}(\nabla \ell_1), \text{cov}(T, S) = \text{cov}(d_{g1}, \nabla \ell_1).$$

3.1. The Variance of IMT under Missing Covariates for Logistic Regression Model

We now need to find expressions for $\text{var}(d_{g1})$, $\text{var}(\nabla \ell_1)$ and $\text{cov}(d_{g1}, \nabla \ell_1)$

We already have that

$$d_g = (y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}$$

and

$$\nabla \ell_i = (y_i - \pi_i) \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

so, the variance is

$$\text{var}(d_g) = E(d_g d_g^T) - E(d_g) E(d_g^T) \tag{10}$$

and we have

$$d_g d_g^T = (y - \pi)^2 (1 - 2\pi)^2 \begin{bmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{bmatrix} \tag{11}$$

taking expectation $E_{Y|X}$ we obtain

$$E(d_{g1}) = E_X \left[(\pi_i - \pi)(1 - 2\pi) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix} \right] \tag{12}$$

and,

$$E(d_{g1}d_{g1}^T) = E_X \left[(\pi_i(1 - 2\pi) + \pi^2)(1 - 2\pi)^2 \begin{bmatrix} 1 & X & X^2 \\ X & X^2 & X^3 \\ X^2 & X^3 & X^4 \end{bmatrix} \right]. \tag{13}$$

Now we need to compute $\text{cov}(d_g, \nabla \ell)$. In fact $E(\nabla \ell) = 0$, not only if the model is correct but also when evaluated at the least false value θ^* (under wrong model), so in this case

$$\text{cov}(d_{g1}, \nabla \ell_1) = E(d_g \nabla \ell)^T.$$

and we have

$$\begin{aligned} d_{g1} \nabla \ell_1^T &= (y - \pi)(1 - 2\pi) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix} (y - \pi) [1 \quad x_i] \\ &= (y - \pi)^2 (1 - 2\pi) \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \\ x_i^2 & x_i^3 \end{bmatrix} \end{aligned}$$

then,

$$E(d_{g1} \nabla \ell_1^T) = E_X \left[(\pi_i(1 - 2\pi) + \pi^2)(1 - 2\pi) \begin{bmatrix} 1 & X \\ X & X^2 \\ X^2 & X^3 \end{bmatrix} \right]. \tag{14}$$

Now we will work out $\text{var}(\nabla \ell)$, as before, since $E(\nabla \ell) = 0$, so

$$\text{var}(\nabla \ell_1) = E(\nabla \ell \nabla \ell^T) = E_X E_{Y|X} \begin{bmatrix} (Y - \pi)^2 & (Y - \pi)^2 X \\ (Y - \pi)^2 X & (Y - \pi)^2 X^2 \end{bmatrix}$$

and note that

$$E_{Y|X} (Y - \pi)^2 = E_{Y|X} (Y(1 - 2\pi) + \pi^2) = \pi_i(1 - 2\pi) + \pi^2,$$

where, π_i is $E(Y)$ under the true model. So,

$$E(\nabla \ell \nabla \ell^T) = E_X \begin{bmatrix} \pi_i(1 - 2\pi) + \pi^2 & (\pi_i(1 - 2\pi) + \pi^2)X \\ (\pi_i(1 - 2\pi) + \pi^2)X & (\pi_i(1 - 2\pi) + \pi^2)X^2 \end{bmatrix}. \tag{15}$$

Hence, the required variance (9)

$$E(d_g d_g^T) - E(d_g)E(d_g^T) - E(d_g \nabla \ell^T)E(\nabla \ell \nabla \ell^T)^{-1}E((\nabla \ell)d_g^T) \tag{16}$$

and we have expressions for each component from (12), (13), (14) and (15) We need to evaluate these components by simulation.

3.2. The Dispersion Matrix under Wrong Model

In fact, may be some elements of the covariance matrix of the IMT are linear combinations of others leading to singularity of the estimated covariance matrix, this point discussed by [1] and [12]. We are interested to compute the $\text{var}(T | S = 0)$, even when the wrong model has been fitted. We will compute each of the components of this variance separately. We see from Section 3.1 that we need to evaluate, e.g.

$$E(d) = E_X \left((\pi_i - \pi)(1 - 2\pi) \begin{bmatrix} 1 \\ X \\ X^2 \end{bmatrix} \right)$$

and also,

$$E(dd^T) = E_X \left(\left[\pi_i(1 - 2\pi) + \pi^2 \right] (1 - 2\pi)^2 \begin{bmatrix} 1 & X & X^2 \\ X & X^2 & X^3 \\ X^2 & X^3 & X^4 \end{bmatrix} \right).$$

This cannot be done analytically so we simulate 5000 values of X and replace the $E(d)$ by the mean of these 5000 values. In evaluating π_i we use the values of the parameters α_i , β_{1i} and β_{2i} . What do we use for π ? We need to evaluate $\pi(\alpha, \beta_1)$ at the least false values α^* and β_1^* for α and β_1 . So, e.g, the first element of $E(d)$ is found by simulation from

$$E_X \left[\left(\text{expit}(\alpha_i + \beta_{1i}X_1 + \beta_{2i}X_2) - \text{expit}(\alpha^* + \beta_1^*X_1) \right) \left(1 - 2\text{expit}(\alpha^* + \beta_1^*X_1) \right) \right]$$

where,

$$\alpha^* = \frac{\alpha_i + \beta_{12}(\mu_2 - \rho\mu_1)}{\sqrt{1 + k^2\beta_{12}^2\sigma^2(1 - \rho^2)}}, \tag{17}$$

$$\beta_1^* = \frac{\beta_{11} + \rho\beta_{12}}{\sqrt{1 + k^2\beta_{12}^2\sigma^2(1 - \rho^2)}} \tag{18}$$

and X draw from bivariate normal distribution with $\mu = (\mu_1, \mu_2)$, and $\sigma_1^2 = \sigma_2^2$. The formulae of the least false values α^* and β_1^* has been discussed and calculated by [10].

4. Empirical Variance of IMT

The expression in (16) is the variance V of d at $\hat{\theta}$ but we need an estimate, \hat{V} . If we have a sample $\{(y_i, x_{1i}) | i = 1, \dots, n\}$ how can we estimate V consistently? One candidate would be to compute

$$d_i = (y_i - \hat{\pi}_i)(1 - 2\hat{\pi}_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}, i = 1, \dots, n$$

and

$$\nabla \ell_i = (y_i - \hat{\pi}_i) \begin{bmatrix} 1 \\ x_i \end{bmatrix}, i = 1, \dots, n$$

where, $\hat{\pi}_i$ is the fitted value from the model with just x_1 . Now compute

$$\hat{W}_n = \frac{1}{n} \sum_{i=1}^n d_i d_i^T - \left(\frac{1}{n} \sum_{i=1}^n d_i \right) \left(\frac{1}{n} \sum_{i=1}^n d_i^T \right)$$

and

$$\hat{B}_n = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\pi}_i)^2 \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix},$$

$$\hat{C}_n = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\pi}_i)^2 (1 - 2\hat{\pi}_i) \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \\ x_i^2 & x_i^3 \end{bmatrix}$$

Then use

$$\hat{V} = \hat{W}_n - \hat{C}_n \hat{B}_n^{-1} \hat{C}_n^T \tag{19}$$

as an estimate of V , we will assess this by simulation.

5. Simulation Study

This simulation examines the correctness of the form of the dispersion matrix V in (16) and (19). To achieve the aim of this simulation, we will consider a logistic regression model which has two covariates draw from bivariate normal distribution with mean zero and covariance matrix Σ as:

$$\pi_i = \text{expit}(\alpha_i + \beta_{i1}x_1 + \beta_{i2}x_2)$$

and the fitted model is

$$\pi = \text{expit}(\alpha + \beta_1 x_1)$$

- Apply in two cases of logistic model,
- The fitted is the true logistic model (*i.e.* $\beta_{i2} = 0$)
- The fitted model is mis-specified (*i.e.* $\beta_{i2} \neq 0$).
- Use variance ($\sigma_1^2 = \sigma_2^2 = 2$) and correlation $\rho = 0.1$.
- We choose some different components of parameters α_i , β_{i1} and β_{i2} to calculate π_i .
- We compute the least false values α^* and β_1^* by formulae to calculate π .
- We compute the true variance by simulating d_i and take the variance to be $\text{var}(\sqrt{nd}) = V_{tr}$.
- We compute the theoretical variance $\text{var}(d) = V_T$ at the least false value and calculate $E(d_1)$ and $E(d_1 d_1^T)$ as described in section 3.2.
- Finally, for each simulation we compute the empirical variance V_E and take the mean over the simulations.
- We make comparison between the diagonal elements of dispersion matrix V_E, V_T vs. V_{tr} respectively.
- Apply on different sample size $n = 500, 1000$ and $N = 5000$ number of simulations.

6. Results and Discussion

The results were reported in tables, which show the diagonal elements of the va-

variance matrix: V_E denotes the empirical variance, V_T denotes the theoretical variance and V_{ir} denotes the true variance. The true parameters appear as α_i , β_{i1} , and β_{i2} ; Rn_E and Rn_T denote to the rank of the covariance matrix empirical and theoretical respectively. The Ratio R_E and R_T are $\sqrt{\frac{V_E}{V_{ir}}}$, $\sqrt{\frac{V_T}{V_{ir}}}$ respectively. $S.D(\pi_i)$ denotes the standard deviation over a sample where π_i is the true model. In our simulation we consider two covariates, so in this case the dispersion matrix of d is a 3×3 dimensional matrix.

Firstly, we consider the results under true logistic model, **Table 1**, shows the results of simulation, which appeared the diagonal elements of matrix V , the empirical version and theoretical form comparing with true variance, which use $\rho = 0.1$ in case of $\sigma_1^2 = \sigma_2^2 = 2$ by sample size $n = 500$. **Table 2**, reported the results by sample size $n = 1000$, with equal variance $\sigma_1^2 = \sigma_2^2 = 2$. We can see clearly, that all diagonal elements appeared small in value in two different cases of sample size. The first element was much closer to zero than of the rest. In almost cases the results appeared reasonable ratio which is meaning the theoretical variance and empirical variance are close to the true value. There are some slightly strange ratio almost in case of sample size $n = 500$, the reason may be affected by small value of standard deviation of π_i , $S.D(\pi_i)$, otherwise the ratio is close to one. In case of sample size $n = 1000$, the behaviour of results shows almost the same pattern, with the ratio close to one and that is meaning the formulae of the variance works well. In a few cases with small values of $S.D(\pi_i)$ which affected on the ratio where the first two elements were more sensitive. Overall, we have reasonable results to say that, the alternative formulae of variance works well and the two first elements still more sensitive which appeared tend to zero.

Secondly, we consider the results when the missing covariate logistic model has been fitted. That is meaning when the variance of IMT computed under H_1 and uses the least false values. **Table 3**, shows the results of sample size $n = 500$. **Table 4**, shows the results of sample size 1000. In general, the behaviour of ratio

Table 1. Simulation results of the variance (V_{ir}), (V_E) and (V_T) in case of fitted true model, with sample size $n = 500$ and $\sigma_1^2 = \sigma_2^2 = 2$.

Diagonal component of variance IMT and Ratio										
α_i	β_{i1}	π_i	$S.D(\pi_i)$	Rn_E	Rn_T	V_E	V_T	V_{ir}	R_1	R_2
0.80	0.50	0.68	0.14	3	3	2.4323e ⁻⁰⁴	2.4817e ⁻⁰⁴	2.4979e ⁻⁰⁴	0.99	0.99
-	-	-	-	-	-	4.2714e ⁻⁰²	4.6277e ⁻⁰²	4.3906e ⁻⁰²	0.99	1.02
-	-	-	-	-	-	1.9446e ⁻⁰¹	2.1171e ⁻⁰¹	2.0659e ⁻⁰¹	0.97	1.01
1.20	2.20	0.63	0.36	3	3	1.4666e ⁻⁰³	1.6904e ⁻⁰³	1.6126e ⁻⁰³	0.95	1.02
-	-	-	-	-	-	1.8138e ⁻⁰²	2.0199e ⁻⁰²	1.9872e ⁻⁰²	0.96	1.01
-	-	-	-	-	-	2.5199e ⁻⁰²	2.9136e ⁻⁰²	2.8595e ⁻⁰²	0.94	1.01

Table 2. Simulation results of the variance (V_{ir}), (V_E) and (V_T) in case of fitted true model, with sample size $n=1000$ and $\sigma_1^2 = \sigma_2^2 = 2$.

Diagonal component of variance IMT and Ratio										
α_i	β_{i1}	π_i	$S.D(\pi_i)$	Rn_E	Rn_T	V_E	V_T	V_{ir}	R_1	R_2
0.80	0.50	0.67	0.14	3	3	$2.4038e^{-04}$	$2.4441e^{-04}$	$2.4915e^{-04}$	0.98	0.99
-	-	-	-	-	-	$4.3783e^{-02}$	$4.4257e^{-02}$	$4.4706e^{-02}$	0.99	0.99
-	-	-	-	-	-	$1.9858e^{-01}$	$1.9486e^{-01}$	$1.0478e^{-01}$	0.98	0.98
1.20	2.20	0.64	0.35	3	3	$1.5709e^{-03}$	$1.6876e^{-03}$	$1.6469e^{-03}$	0/98	1.01
-	-	-	-	-	-	$1.9049e^{-02}$	$2.0225e^{-02}$	$2.0199e^{-02}$	0.97	1.00
-	-	-	-	-	-	$2.6726e^{-02}$	$2.9664e^{-02}$	$2.7877e^{-02}$	0.98	1.03

Table 3. Simulation results of the variance (V_{ir}), (V_E) and (V_T) in case of fitted missing covariates model, with sample size $n=500$ and $\sigma_1^2 = \sigma_2^2 = 2$.

Diagonal component of variance IMT and Ratio											
α_i	β_{i1}	β_{i2}	π_i	$S.D(\pi_i)$	Rn_E	Rn_T	V_E	V_T	V_{ir}	R_1	R_2
0.80	0.50	0.4	0.66	0.18	3	3	$2.3400e^{-04}$	$2.5316e^{-04}$	$2.3730e^{-04}$	0.99	1.03
-	-	-	-	-	-	-	$4.3984e^{-02}$	$5.1198e^{-02}$	$4.5628e^{-02}$	0.98	1.05
-	-	-	-	-	-	-	$1.9280e^{-01}$	$2.3559e^{-01}$	$2.0038e^{-01}$	0.98	1.08
1.20	2.20	0.8	0.62	0.36	3	3	$1.2620e^{-03}$	$1.4252e^{-03}$	$1.3850e^{-03}$	0.95	1.01
-	-	-	-	-	-	-	$2.1819e^{-02}$	$2.3741e^{-02}$	$2.3411e^{-02}$	0.97	1.01
-	-	-	-	-	-	-	$3.2250e^{-02}$	$3.5432e^{-02}$	$3.6766e^{-02}$	0.94	0.98

Table 4. Simulation results of the variance (V_{ir}), (V_E) and (V_T) in case of fitted missing covariates model, with sample size $n=1000$ and $\sigma_1^2 = \sigma_2^2 = 2$.

Diagonal component of variance IMT and Ratio											
α_i	β_{i1}	β_{i2}	π_i	$S.D(\pi_i)$	Rn_E	Rn_T	V_E	V_T	V_{ir}	R_1	R_2
0.80	0.50	0.4	0.66	0.18	3	3	$2.3253e^{-04}$	$2.6095e^{-04}$	$2.2984e^{-04}$	1.01	1.06
-	-	-	-	-	-	-	$4.5129e^{-02}$	$4.8176e^{-02}$	$4.4637e^{-02}$	1.01	1.03
-	-	-	-	-	-	-	$1.9837e^{-01}$	$2.3374e^{-01}$	$2.0511e^{-01}$	0.98	1.06
1.20	2.20	0.8	0.59	0.36	3	3	$1.3404e^{-03}$	$1.4576e^{-03}$	$1.4116e^{-03}$	0.97	1.01
-	-	-	-	-	-	-	$2.2914e^{-02}$	$2.4611e^{-02}$	$2.3911e^{-02}$	0.98	1.01
-	-	-	-	-	-	-	$3.3481e^{-02}$	$3.7438e^{-02}$	$3.4621e^{-02}$	0.98	1.03

appeared the same behaviour which found in case of $\beta_{2i} = 0$, the two cases of different sample size appeared reasonable ratio which is close to one. A few cases shows low ratio, the reason is as discussed before concerning to the small value of $S.D(\pi_i)$.

7. Conclusion

This paper carried out to investigate the behaviour of IMT and compute the covariance matrix under the wrong logistic regression model. As result, we can see that the alternative formula of the variance appeared reasonable results under the true and missing covariate model. As we computed the final form of the variance of IMT, we can see clearly it is dependent on $E(d)$. As we know, we made some notes on the first two elements of $E(d)$, which may be quite close to zero under true model and use the least false value, the $E(\pi_i - \pi) = E((\pi_i - \pi)X) = 0$ related to the log likelihood functions. So, these elements leading to singularity of the estimated covariance matrix, and have effect on the behaviour of the dispersion matrix of the IMT.

Acknowledgements

I am very grateful to Professor J. N. S. Matthews, School of Mathematics and Statistics, Newcastle University for academic supporting and Dr. Hamza M. A. Boauod, (FCOPHTH) (SA), Consultant Ophthalmologist, Eye Department, Klerksdorp Hospital. South Africa for his financial support. Also thank the referees, associate editor and joint editor for their helpful comments and additional references.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] White, H. (1982) Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, **50**, 1-25. <https://doi.org/10.2307/1912526>
- [2] Hausman, J. A. (1978) Specification Tests in Econometrics. *Econometrica*, **46**, 1251-1271. <https://doi.org/10.2307/1913827>
- [3] Chesher, A. (1984) Testing for Neglected Heterogeneity. *Econometrica*, **52**, 865-872. <https://doi.org/10.2307/1911188>
- [4] Newey, W.K. (1984) Maximum Likelihood Specification Testing and Conditional Moment Tests. *Econometrica*, **53**, 1047-1070. <https://doi.org/10.2307/1911011>
- [5] Davidson, R. and Mackinnon, J.G. (1984) Convenient Specification Tests for Logit and Probit Models. *Journal of Econometrics*, **25**, 241-262. [https://doi.org/10.1016/0304-4076\(84\)90001-0](https://doi.org/10.1016/0304-4076(84)90001-0)
- [6] Orme, C. (1988) The Calculation of the Information Matrix Test for Binary Data Models. *The Manchester School*, **56**, 370-376. <https://doi.org/10.1111/j.1467-9957.1988.tb01339.x>
- [7] Lancaster, T. (1984) Covariance Matrix of the Information Matrix Test. *Econometrica*, **52**, 1051-1053. <https://doi.org/10.2307/1911198>
- [8] Kuss, O. (2002) Global Goodness-of-Fit Tests in Logistic Regression with Sparse Data. *Statistics in Medicine*, **21**, 3789-3801. <https://doi.org/10.1002/sim.1421>
- [9] Matthews, J.N.S. and Badi, N.H. (2015) Inconsistent Treatment Estimates from

Mis-Specified Logistic Regression Analyses of Randomized Trials. *Statistics in Medicine*, **34**, 2681-2694. <https://doi.org/10.1002/sim.6508>

- [10] Badi, N.H.S. (2017) Properties of the Maximum Likelihood Estimates and Bias Reduction for Logistic Regression Model. *Open Access Library Journal*, **4**, e3625.
- [11] Claeskens, G. and Hjort, N.L. (2008) Model Selection and Model Averaging. Cambridge University Press, Cambridge.
- [12] Lin, D.Y. and Wei, L.J. (1991) Goodness-of-Fit Tests for the General Cox Regression Model. *Statistica Sinica*, **1**, 1-17.