Multidimensional Structural Regression Model for Causal Inference under Strongly Ignorable Treatment Assignment*

Yonghong Wang
School of Mathematics and Computer, Harbin University, Harbin 150086, China
Email:wangyonghong418@163.com

Abstract *In the study of epidemiology aetiology, we usually cannot measure exposed effect relative to an individual, but under some assumptions, we approximately replace the exposed effect by estimator of population average causal effect. A multidimensional structural regression model for causal inference is established to estimate the population average treatment effect under strongly ignorable treatment assignment. Under the normal distribution, the maximum likelihood estimator for population average treatment effect is proved to be consistent, unbiased and asymptotically normal.*

Keywords *Strongly ignorable treatment assignment Causal inference Population average treatment effect Multidimensional structural regression model Maximum likelihood estimator*

1. Introduction

The objective of epidemiology study is to search aetiology, and to measure its causal effect in the light of quantity, thereby prevent from the occurrence of disease[1]. The case -control study, which attempts to find a contrasted or exposed group which can be compared with the treated or exposed one, is the common method that we employ in the research of epidemiology aetiology. Except the difference of being exposed and unexposed, the ideal treated group and the contrasted group are expected to share the rest of the other features, which serves as the guiding principle in the epidemiology research[2-3]. The dummy truth model laid a theoretical foundation for us to demonstrate this principle[4]. Let U be the study population, and denote a generic individual in U by $u \in U$. A variable E is defined on each $u \in U$ so that $E(u)=1$ if u is exposed to the casual agent of interest and $E(u)=0$ if u is not so exposed. $Y_1(u)$ and $Y_0(u)$ indicate the diseased status of u as in the exposed and non-exposed cases respectively. Thus we can define the exposed causal effect of the individual u as $Y_1(u) — Y_0(u)$ .As we know, in any real epdidemilogic study, one does not observe $Y_1(u)$ and $Y_0(u)$ simultaneously. Therefore the casual effect of the exposure to the individual is not attainable. However, in some circumstances, we can get the population average causal effect, which can be defined as $\mathrm{E(Y_1(u)\text{-}Y_0(u))}$, while $E(\cdot)$ denotes expection or population average (over U). Under the circumstances of random tests, for the fact that the independence of the exposed E and other variants, the result should be $E \perp\!\!\!\perp (Y_1, Y_0)$, which indicates that the exposed E is independent of the random Vector (Y_1, Y_0) . Thus the population average causal effect is shown as $E(Y_1 - Y_0) = E(Y_1) - E(Y_0) = E(Y_1 | E=1) - E(Y_0 | E=0)$ .In other words, the causal effect can be estimated by the difference of the exposed population average diseased condition the non-exposed population average diseased condition[5]. The epidemiology, however, is an

*This work has been supported by the Science Development Foundation of Harbin University(HXKQ200701).

observatory science in nature, so the above formula is not to be established which necessitates the assumption of the state of being ignorable[5-6] .

Definition 1.1 Given the covariant X , the exposed E is called being strongly ignorable. If

$$\begin{cases} (1)\ (Y_1, Y_0) \perp\!\!\!\perp E|X \text{, i.e. the observed covariant X is given, the response variant is independent of the exposed E.} \\ (2)\ 0 < \Pr(E = e|X) < 1 \text{, i.e. for each and every X, they may receive various treatment.} \end{cases} \quad (1)$$

When the E is strongly ignorable, by getting the matching-differencing-averaging between the exposed and non-exposed groups, we can get the unbiased estimate of population average causal effect[7]. But the above method is complicated in calculation and the additional sparse-data problems, excess covariant, though few, may arise. In addition, the method of ranking the individuals with the estimated propensity score is an advisable way to solve the problem, but the calculation is still very complicated[8]. [9] Under strongly ignorable treatment assignment, a structural regression model for causal inference is established to estimates the population average treatment effect. They are the series of results achieved. with the response variant being one dimensional. The paper will discuss the inference problem with multi-dimensional response variants.

2. Multidimensional Structural Regression Model

Definition 2.1 In Condition of Strongly Ignorable Treatment Assignment, denote multidimensional Structural Regression model as follows:

$$\begin{cases} Y_{tj} = \alpha_t + \beta_t X_{tj} + e_{tj} \\ X_{tj} \sim N(\mu_x, \Sigma_x), e_{tj} \sim N(0, V_t), X_{tj} \text{ is independent of } e_{tj} \end{cases} \quad (2)$$

Here t=0,1 means two treatment; Y_{tj} is the random vector of order $p \times 1$,indicated the response variant of the jth observation for the tth treatment; X_{tj} is the covariant vector of order $k \times 1$; e_{tj} is the random error vector of order $p \times 1$; α_t is the parameter vector of order $p \times 1$; β_t is the parameter matrix of order $p \times k$; μ_x is the parameter vector of order $k \times 1$;The covariant matrix Σ_x和$\mathrm{V_t}$ of X_{tj} and e_{tj} are both positive definite.

Under model （2）

$$E\begin{pmatrix} X_{tj} \\ Y_{tj} \end{pmatrix} = \begin{pmatrix} \mu_x \\ \alpha_t + \beta_t \mu_x \end{pmatrix} \;,\; A_t \triangleq Cov\begin{pmatrix} X_{tj} \\ Y_{tj} \end{pmatrix} = \begin{pmatrix} \Sigma_x & \Sigma_x \beta_t' \\ \beta_t \Sigma_x & \beta_t \Sigma_x \beta_t' + V_t \end{pmatrix}$$

We can easily find $|A_t| = |\Sigma_x| \cdot |V_t|$;because Σ_x and V_t are both positive definite, we can know $|\Sigma_x| \neq 0$, $|V_t| \neq 0$; so $|A_t| \neq 0$;therefore A_t^{-1} is valid, and we denote it as B_t .again based on the principle of block matrix inversion, we can get

$$B_t = \begin{pmatrix} \Sigma_x^{-1} + \beta_t' V_t^{-1} \beta_t & -\beta_t' V_t^{-1} \\ V_t^{-1} \beta_t & V_t^{-1} \end{pmatrix}$$

$$\text{So} \quad \begin{pmatrix} X_{tj} \\ Y_{tj} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \alpha_t + \beta_t \mu_x \end{pmatrix}, A_t \right)$$

Where: t=0，1 ； j=1，……，n_t .

We also get observation logarithm likelihood function of the sample (x_{tj}, y_{tj}) with the capacity n_t (t=0，1； j=1，……，n_t)

$$-2\ln L=\sum_{t=0}^{1}n_t\ln|A_t|+\sum_{t=0}^{1}\sum_{j=1}^{n_t}\begin{pmatrix}x_{tj}-\mu_x\\ y_{tj}-\alpha_t-\beta_t\mu_x\end{pmatrix}^{'}B_t\begin{pmatrix}x_{tj}-\mu_x\\ y_{tj}-\alpha_t-\beta_t\mu_x\end{pmatrix}$$

Furthermore，

$$\begin{aligned}-2\ln L=&\sum_{t=0}^{1}n_t(\ln|\Sigma_x|+\ln|V_t|)+\sum_{t=0}^{1}\sum_{j=1}^{n_t}(x_{tj}^{'}\Sigma_x^{-1}x_{tj}-2x_{tj}^{'}\Sigma_x^{-1}\mu_x\\ &+\mu_x^{'}\Sigma_x^{-1}\mu_x+x_{tj}^{'}\beta_t^{'}V_t^{-1}\beta_t x_{tj}-2x_{tj}^{'}\beta_t^{'}V_t^{-1}y_{tj}\\ &+2x_{tj}^{'}\beta_t^{'}V_t^{-1}\alpha_t+y_{tj}^{'}V_t^{-1}y_{tj}-2y_{tj}^{'}V_t^{-1}\alpha_t+\alpha_t^{'}V_t^{-1}\alpha_t)\end{aligned} \qquad (3)$$

3. Conclusion

3.1 Two Lemmas

In order to get the likelihood estimates of all parameters in model (2) and the population average causal effect, we give the following two lemmas, of which lemma3.1 is shown at [10]. First, suppose X is the matrix of order $m\times n$, $f(x)$ is the real-valued function of matrix X, $\frac{\partial f(x)}{\partial x}$ indicate real-valued function $f(x)$ settling partial derivation for each element of matrix X. Thus the following lemmas are valid:

Lemma 3.1（1） $\frac{\partial tr(AXB)}{\partial X}=A^{'}B^{'}$

（2） $\frac{\partial tr(X^{'}AXB)}{\partial X}=AXB+A^{'}XB^{'}$

（3） $\frac{\partial \ln|X|}{\partial X}=(X^{-1})^{'}$

Of which A and B are two matrixes that can make matrix calculation with X, $tr(X)$ means the trace of matrix X.

Lemma 3.2 If model (2) is valid, when the covariant matrix and the random error matrix are independent of each other, then

$$\overline{X}_t=\frac{1}{n_t}\sum_{j=1}^{n_t}X_{tj}$$

and

$$M_{X_tX_t^{'}}=\frac{1}{n_t}\sum_{j=1}^{n_t}(X_{tj}-\overline{X}_t)(X_{tj}-\overline{X}_t)^{'}，\quad M_{e_tX_t^{'}}=\frac{1}{n_t}\sum_{j=1}^{n_t}(e_{tj}-\overline{e}_t)(X_{tj}-\overline{X}_t)^{'}$$

are independent of each other, where $\overline{e}_t=\frac{1}{n_t}\sum_{j=1}^{n_t}e_{tj}$.

Proof: First to prove $\overline{X}_t$ and $M_{X_tX_t^{'}}$ are independent.

Let
$$X_t = \begin{pmatrix} X_{t11} & X_{t12} & \cdots & X_{t1k} \\ X_{t21} & X_{t22} & \cdots & X_{t2k} \\ \cdots & \cdots & \cdots & \cdots \\ X_{tn_t1} & X_{tn_t2} & \cdots & X_{tn_tk} \end{pmatrix}$$

Set $\overline{X}_t = \frac{1}{n} X_t^{'} \mathbf{1}$ also

$$M_{X_t X_t^{'}} = \frac{1}{n_t}(X_t^{'} - \frac{1}{n_t} X_t^{'} \mathbf{11}^{'})(X_t - \frac{1}{n_t}\mathbf{11}^{'} X_t)$$
$$= \frac{1}{n_t} X_t^{'} (\mathbf{I}_{n_t} - \frac{1}{n_t}\mathbf{11}^{'}) \ X_t$$

Which **1** is the column vector of $n_t \times 1$.

We can do the vector straight operations to matrix X_t, then $Cov(Vec(X)) = \Sigma_x \otimes \mathbf{I}_{n_t}$,

also, as for $\overline{X}_t$ and each at the element m_{ij} in $M_{X_t X_t^{'}}$

follow $\overline{X}_t = \frac{1}{n_t}(\boldsymbol{I_k} \otimes \mathbf{1}^{'}) \ Vec(X_t)$

$$m_{ij} = Vec(X_t)^{'}[\frac{1}{2}(E_{ij} + E_{ij}^{'}) \otimes (\mathbf{I}_{n_t} - \frac{1}{n_t}\mathbf{11}^{'})]Vec(X_t)$$

Of which $\otimes$ indicates the Kronecker product of matrix, E_{ij} indicates matrix of order $k \times k$ with $(i,j)^{th}$ is 1, and the rest is 0. Now we can draw the following conclusion by making use of the characteristic of multidimensional normal random variant and the Kronecker product of the matrix:

$$\frac{1}{n_t}(\boldsymbol{I_k} \otimes \mathbf{1}^{'})(\Sigma_x \otimes \mathbf{I}_{n_t})[\frac{1}{2}(E_{ij} + E_{ij}^{'}) \otimes (\mathbf{I}_{n_t} - \frac{1}{n_t}\mathbf{11}^{'})]$$
$$= [\frac{1}{2}\Sigma_x (E_{ij} + E_{ij}^{'})] \otimes [\frac{1}{n_t} 1^{'} (\mathbf{I}_{n_t} - \frac{1}{n_t}\mathbf{11}^{'})] = 0$$

Thus we can prove m_{ij} and $\overline{X}_t$ are independent, and it's valid to any *i*, *j*=*1*, ……, *k*. so we have proved $\overline{X}_t = \frac{1}{n_t}\sum_{j=1}^{n_t} X_{tj}$ and $M_{X_t X_t^{'}} = \frac{1}{n_t}\sum_{j=1}^{n_t}(X_{tj} - \overline{X}_t)(X_{tj} - \overline{X}_t)^{'}$ are independent of each other.

Suppose
$$e_t = \begin{pmatrix} e_{t11} & e_{t12} & \cdots & e_{t1p} \\ e_{t21} & e_{t22} & \cdots & e_{t2p} \\ \cdots & \cdots & \cdots & \cdots \\ e_{tn_t1} & e_{tn_t2} & \cdots & e_{tn_tp} \end{pmatrix}$$

$Z_t = (X_t, e_t)$, in the same way , and by mutual independence of X_t and e_t, we can prove $\overline{X}_t$ and $M_{e_t X_t^{'}}$ are also independent of each other.

Lemma3.2 is proved!

3.2 Estimate of Parameter

Theorem 3.1 Suppose model（2）is valid, we can get the maximum likelihood estimates of all the parameter:

$$\hat{\alpha}_t = \bar{y}_t - M_{y_t x_t^{'}} (M_{x_t x_t^{'}})^{-1} \bar{x}_t \ ; \hat{\beta}_t = M_{y_t x_t^{'}} (M_{x_t x_t^{'}})^{-1} \ ; \hat{\mu}_x = \frac{1}{n_0 + n_1} \sum_{t=0}^{1} \sum_{j=1}^{n_t} x_{tj} \ ;$$

$$\hat{\Sigma}_x = \frac{1}{n_0 + n_1} \sum_{t=0}^{1} \sum_{j=1}^{n_t} (x_{tj} - \hat{\mu}_x)(x_{tj} - \hat{\mu}_x)^{'} ; \hat{V}_t = \frac{1}{n_t} \sum_{j=1}^{n_t} (y_{tj} - \hat{\alpha}_t - \hat{\beta}_t x_{tj})(y_{tj} - \hat{\alpha}_t - \hat{\beta}_t x_{tj})^{'}$$

Of which $\bar{y}_t = \frac{1}{n_t} \sum_{j=1}^{n_t} y_{tj}$, $M_{y_t x_t^{'}} = \frac{1}{n_t} \sum_{j=1}^{n_t} (y_{tj} - \bar{y}_t)(x_{tj} - \bar{x}_t)^{'}$; The form of $\bar{x}_t$ and $M_{x_t x_t^{'}}$ is shown in lemma3.2.

Proof: First, we get the score function of α_t:

$$\frac{\partial(-2\ln L)}{\partial \alpha_t} = \sum_{j=1}^{n_t} \frac{\partial}{\partial \alpha_t} (2x_{tj}^{'} \beta_t^{'} V_t^{-1} \alpha_t - 2y_{tj}^{'} V_t^{-1} \alpha_t + \alpha_t^{'} V_t^{-1} \alpha_t)$$

$$= \sum_{j=1}^{n_t} [\frac{\partial}{\partial \alpha_t} 2tr(x_{tj}^{'} \beta_t^{'} V_t^{-1} \alpha_t) - 2\frac{\partial}{\partial \alpha_t} tr(y_{tj}^{'} V_t^{-1} \alpha_t) + \frac{\partial}{\partial \alpha_t} tr(\alpha_t^{'} V_t^{-1} \alpha_t)]$$

$$= \sum_{j=1}^{n_t} (2V_t^{-1} \beta_t x_{tj} - 2V_t^{-1} \alpha_t) = 0$$

We can get $\alpha_t = \bar{y}_t - \beta_t \bar{x}_t$. And the score function of β_t is as follows:

$$\frac{\partial(-2\ln L)}{\partial \beta_t} = \sum_{j=1}^{n_t} (2V_t^{-1} \beta_t x_{tj} x_{tj}^{'} - 2V_t^{-1} y_{tj} x_{tj}^{'} + 2V_t^{-1} \alpha_t x_{tj}^{'}) = 0$$

Simplified as $$\sum_{j=1}^{n_t} (\beta_t x_{tj} x_{tj}^{'} - y_{tj} x_{tj}^{'} + \alpha_t x_{tj}^{'}) = 0$$

Fit β_t into above formula of α_t, we set $\hat{\alpha}_t = \bar{y}_t - M_{y_t x_t^{'}} (M_{x_t x_t^{'}})^{-1} \bar{x}_t$.

In the same way, we can get the maximum likelihood estimates of the other three parameters.

Theorem3.1 is proved!

In fact, as for the multidimensional structural regression model with the response variants as shown in model (2), we can see from the above demonstration that the maximum likelihood estimates of parameter α_t and β_t do not depend on the selection of the covariant matrix and random error matrix. Moreover, we can also see that if we want to estimate the parameter that measure relationship between the covariant and the j^{th} （*j=1,……, n_t*）response variable, we should only note the j^{th} response variable. i.e. the model(2) can be divided into p one dimensional models, which is evident in Jinhua's paper(2000).

3.3 Population Average Causal Effect

Definition3.1 provides the covariant X, if the assumption of Strongly Ignorable condition (1) is fulfilled, then treatment E=1 to E=0 average causal effect ATE is

$$ATE = E(Y_1) - E(Y_0) = E_x(E(Y_1) - E(Y_0) | X = x)$$

$$=E_x(E(Y_1|E=1,X=x)-E(Y_0|E=0,X=x))$$
$$=E_x(\alpha_1+\beta_1 x-\alpha_0-\beta_0 x)=(\alpha_1-\alpha_0)+(\beta_1-\beta_0)\mu_x$$

Theorem3.2 suppose model (2) is valid, and treatment assignment variant E is strongly ignorable, then

(1) The maximum likelihood estimator of population average treatment effect ATE is

$$\widehat{ATE}=\overline{Y}_1-\overline{Y}_0-\frac{n_0\hat{\beta}_1+n_1\hat{\beta}_0}{n_0+n_1}(\overline{X}_1-\overline{X}_0)$$

(2) $\widehat{ATE}$ is consistent unbiased estimator of ATE.

(3) With the case of the large sample, $A\hat{T}E$ is close to normal distribution N（ATE, Γ）.

Of which $$\Gamma\triangleq\frac{V_0}{n_0}+\frac{V_1}{n_1}+\frac{(\beta_1-\beta_0)\Sigma_x(\beta_1-\beta_0)'}{n_0+n_1}$$

Proof: (1) with model (2), if treatment assignment variant E is strongly ignorable, then the population average treatment effect ATE= $(\alpha_1-\alpha_0)+(\beta_1-\beta_0)\mu_x$

Based on theorem 3.1 and the invariance of maximum likelihood estimator, we can know

$$\widehat{ATE}=(\hat{\alpha}_1-\hat{\alpha}_0)+(\hat{\beta}_1-\hat{\beta}_0)\hat{\mu}_x\quad=\overline{Y}_1-\overline{Y}_0-\frac{n_0\hat{\beta}_1+n_1\hat{\beta}_0}{n_0+n_1}(\overline{X}_1-\overline{X}_0)$$

(2) Consistency: we know based on the model (2)

$$\overline{Y}_t=\alpha_t+\beta_t\overline{X}_t+\overline{e}_t,\ t=0,1$$

Again based on theorem3.1, we get

$$\hat{\beta}_t=M_{Y_tX_t'}(M_{X_tX_t'})^{-1}=\beta_t+(M_{e_tX_t'})(M_{X_tX_t'})^{-1}$$

Based on the law of large numbers and the nature of the random variant order converging by probability, we can know

$$M_{X_tX_t'}\xrightarrow{P}\Sigma_x,M_{e_tX_t'}\xrightarrow{P}0$$

Then $\hat{\beta}_t\xrightarrow{P}\beta_t$

In the same way, $\overline{X}_t\xrightarrow{P}\mu_x,\overline{Y}_t\xrightarrow{P}\alpha_t+\beta_t\mu_x$

So far, again based on the nature of the random variant order converging by probability, we get $\widehat{ATE}\xrightarrow{P}ATE$, i.e. $\widehat{ATE}$ is consistent estimate of the population average treatment effect ATE.

Unbiased：Through the demonstration of the consistency, we know

$$\hat{\beta}=\beta_t+[\frac{1}{n_t}\sum_{j=1}^{n_t}(e_{tj}-\overline{e}_t)(X_{tj}-\overline{X}_t)'](M_{X_tX_t'})^{-1}$$

Whereupon $$E(\hat{\beta})=E\{\beta_t+[\frac{1}{n_t}\sum_{j=1}^{n_t}(e_{tj}-\overline{e}_t)(X_{tj}-\overline{X}_t)'](M_{X_tX_t'})^{-1}\}$$
$$=\beta_t+E_x\{E[\frac{1}{n_t}\sum_{j=1}^{n_t}(e_{tj}-\overline{e}_t)(X_{tj}-\overline{X}_t)'](M_{X_tX_t'})^{-1}|X_t=x_t\}$$
$$=\beta_t+E_x(0|X_t=x_t)=\beta_t$$

Therefore $E(\overline{Y}_t)=\alpha_t+\beta_t\mu_x$, so based on lemma3.2, we set

$$E(\widehat{ATE}) = E(\overline{Y}_1 - \overline{Y}_0) - E(\frac{n_0\hat{\beta}_1 + n_1\hat{\beta}_0}{n_0 + n_1})E(\overline{X}_1 - \overline{X}_0)$$
$$= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\mu_x = ATE$$

(3) Asymptotically normality: Based on above demonstration, we know $E(\widehat{ATE}) = ATE$

Also because of $\widehat{ATE} \approx (\overline{Y}_1 - \overline{Y}_0) - \frac{n_0\beta_1 + n_1\beta_0}{n_0 + n_1}(\overline{X}_1 - \overline{X}_0)$

$$= [\alpha_1 + \frac{n_1(\beta_1 - \beta_0)}{n_0 + n_1}\overline{X}_1 + \overline{e}_1] - [\alpha_0 + \frac{n_0(\beta_1 - \beta_0)}{n_0 + n_1}\overline{X}_0 + \overline{e}_0]$$

Moreover $Var(\overline{X}_t) = \frac{1}{n_t}\Sigma_x$, $Var(\overline{e}_t) = \frac{1}{n_t}V_t$

Therefore

$$Var(\alpha_1 + \frac{n_1(\beta_1 - \beta_0)}{n_0 + n_1}\overline{X}_1 + \overline{e}_1) = \frac{n_1}{(n_0 + n_1)^2}(\beta_1 - \beta_0)\Sigma_x(\beta_1 - \beta_0)' + \frac{1}{n_1}V_1$$

$$Var(\alpha_0 + \frac{n_0(\beta_1 - \beta_0)}{n_0 + n_1}\overline{X}_0 + \overline{e}_0) = \frac{n_0}{(n_0 + n_1)^2}(\beta_1 - \beta_0)\Sigma_x(\beta_1 - \beta_0)' + \frac{1}{n_0}V_0$$

So $$Var(\widehat{ATE}) = \frac{1}{n_1}V_1 + \frac{1}{n_0}V_0 + \frac{1}{n_0 + n_1}(\beta_1 - \beta_0)\Sigma_x(\beta_1 - \beta_0)' \triangleq \Gamma$$

So based on the Central Limited Theorem, we know $\widehat{ATE} \sim AN(ATE, \Gamma)$.

Theorem 3.2 is proved!

References

[1] Rothman,K.J. Modern epidemiology[M]. Boston:Litte, Brown and Company, 1986: 56.

[2] Guo,J.H. Causal inference and confounding phenomenon. *Journal of Northeast Normal University*, 2002, 34: 25-41.

[3] Habtemariam,T. and Tameru,B. etc. Epidemiologic modeling of HIV/AIDS: Use of computational models to study the population dynamics of the disease to assess effective intervention strategies for decision-making. *Advances in Systems Science and Applications*, 2008, 8(1): 35-39.

[4] Tao,Q.S. and Li,L.M. Aetiology effect model based on dummy truth theory. *Chinese Journal of Epidemiology*, 2002, 23(1): 60-62.

[5] Rosenbaum,P. and Rubin,D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70: 41-55.

[6] Holland,P.W. Confounding in epidemiologic studies. *Biometrics*, 1989, 45: 1310-1316.

[7] Kong,F.L. and Wang,Y.F. Certain Martingale Methods in Parameter Estimation. *Advances in Systems Science and Applications*, 2004, 4(1): 1-6.

[8] Yu,J.W. and Cheng,D.H. etc. Factor analysis methods in the statistical literature of Chinese medicine diabetes research applications. *Advances in Systems Science and Applications*, 2007, 7(2): 161-166.

[9] Jin,H. and Fang,J.Q. Structural regression model for causal inference under strongly ignorable treatment assignment. *Journal of south china normal university*, 2000, 4: 7-12.

[10] Wang,S.G. etc. Theory of linear model[M]. Beijing: Science press, 2004: 44-47.