

Some problems in **SEM**

Hui Liu

10701838

School of Mathematical Sciences, Peking University

Oct. 19th, 2009

Abstract

In this report, we first talk about using **MLE** to estimate parameters in **SEM**, and we assume the structure just has three observed variables and three independent normal measurement-error variables. Specially we also assume some parameters stand for different type of structures. Secondly, we try to build confidence set by log-likelihood ratio for parameters talking above.

1. The MLE of SEM with assumed structure

the *sem* package in R is a tool to solve **SEM** with assumed structure. Internally, the *sem* function employs the reticular action model(**RAM**) formulation of model[1].

In the **RAM** model, the vector ν contains indicator variables, directly observed exogenous variables, and latent exogenous and endogenous variables; vector μ (which may overlap with ν) contains directly observed and latent exogenous variables, measurement-error variables, and structural-error variables. Not all classes of variables are present in every model.

The ν and μ vectors are related by the equation

$$\nu = A\nu + \mu$$

$$P = cov(\mu)$$

where the matrix A contains regression coefficients, the matrix P contains covariances among the elements of μ .

Let m represent the number of variables in ν , and let the first n entries of ν be the observed variables of the model. The the $m \times n$ selection matrix $J = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$ picks out the observed variables, where I_n is an order- n identity matrix and the 0 s are zero matrices of appropriate order. Covariance among the observed variables are therefore given by

$$C = E(J\nu\nu'J') = J(I_m - A)^{-1}P[(I_m - A)^{-1}]'J^{-1}$$

Let S denote the covariances among the observed variables computed directly from a sample of data. Estimating the parameters of the model — the unconstrained entries of A and P — entails picking values of the parameters that make C close in some sense to S . In particular, under the assumption that the latent variables and errors are multnormally distribution, maximum likelihood(**ML**)estimates

of the parameters minimize the fitting criterion

$$F(A, P) = \log(|C|) + \text{trace}(SC^{-1}) - n - \log(|S|) \quad (1)$$

The *sem* function minimizes the **ML** fitting criterion numerically using the *nlm* optimizer in R, which employs a Newton-type algorithm; *sem* by default uses an analytic gradient, but a numerical gradient may be optionally employed. The covariance matrix of the parameter estimates is based on the numerical Hessian return by *nlm*.

2. The MLE of SEM without structure assumption

Let the parameters are $\theta = (\gamma, \beta)$, where γ are parameters stand for different type of structures and β are parameters under fixed structure(which mean the rest parameters after fixed γ).

In this report, we just discuss models which contain three observed variables and three independent normal measurement-error variables. Some models in **SEM** could not be identified,such as models which have feedback relationships, so we use Recursive Rules as a sufficient condition of identifiable. Figure2.1 shows all models which are identifiable. Depending on the discussion above, there are 24 models we can choose, so $\beta \in \{1, 2, \dots, 24\}$.

Under the assumption that errors are multinormally distribution, maximum likelihood(**ML**)estimates of the parameters are

$$\begin{aligned} \arg \max_{\theta} P(\theta|x) &= \arg \max_{\gamma, \beta} P(\gamma, \beta|x) \\ &= \arg \max_{i=1,2,\dots,24} \left(\max_{\beta} P(\gamma_i, \beta|x) \right) \\ &= \arg \min_{i=1,2,\dots,24} \left(\min_{\beta} F(A_i, P_i) \right) \end{aligned} \quad (2)$$

where $F(A_i, P_i)$ is the the criterion (1) under *Model_i*.

The problem of MLE(2)

(1) The knee of scree plot between the value of criterion(1) and corresponding model number.


 x_3

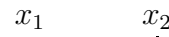
M1


 x_3

M2


 x_3

M3


 x_3

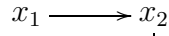
M4


 x_3

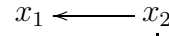
M5


 x_3

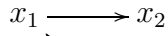
M6


 x_3

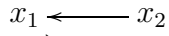
M7


 x_3

M8


 x_3

M9


 x_3

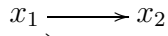
M10


 x_3

M11


 x_3

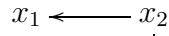
M12


 x_3

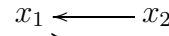
M13


 x_3

M14


 x_3

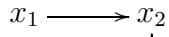
M15


 x_3

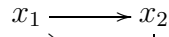
M16


 x_3

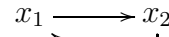
M17


 x_3

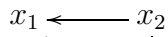
M18


 x_3

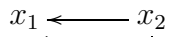
M19


 x_3

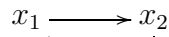
M20


 x_3

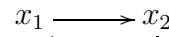
M21


 x_3

M22


 x_3

M23


 x_3

M24

1 Figure2.1 all models with three variables

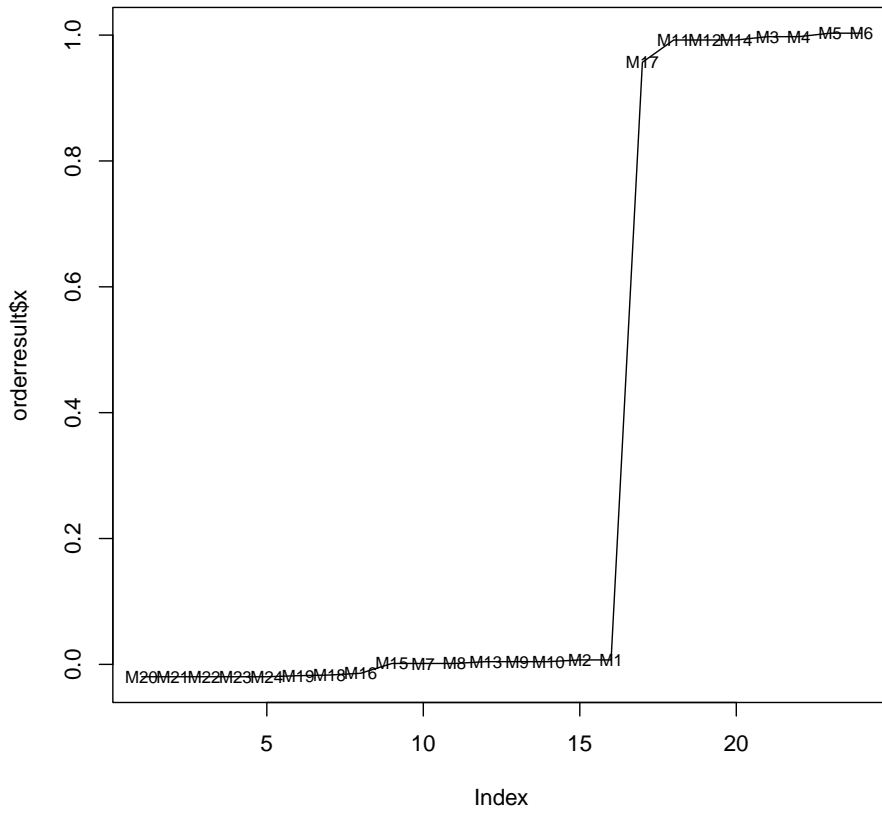


Figure 2.2 Scree plot of $Model_1$

Figure 2.2 has a obviously knee point between M1 and M17, and models before this knee point all contain the true model M1. Figure 2.2 also shows that models are ordered by their complexity.

So we guess that the true model is near the knee point for sparse model.

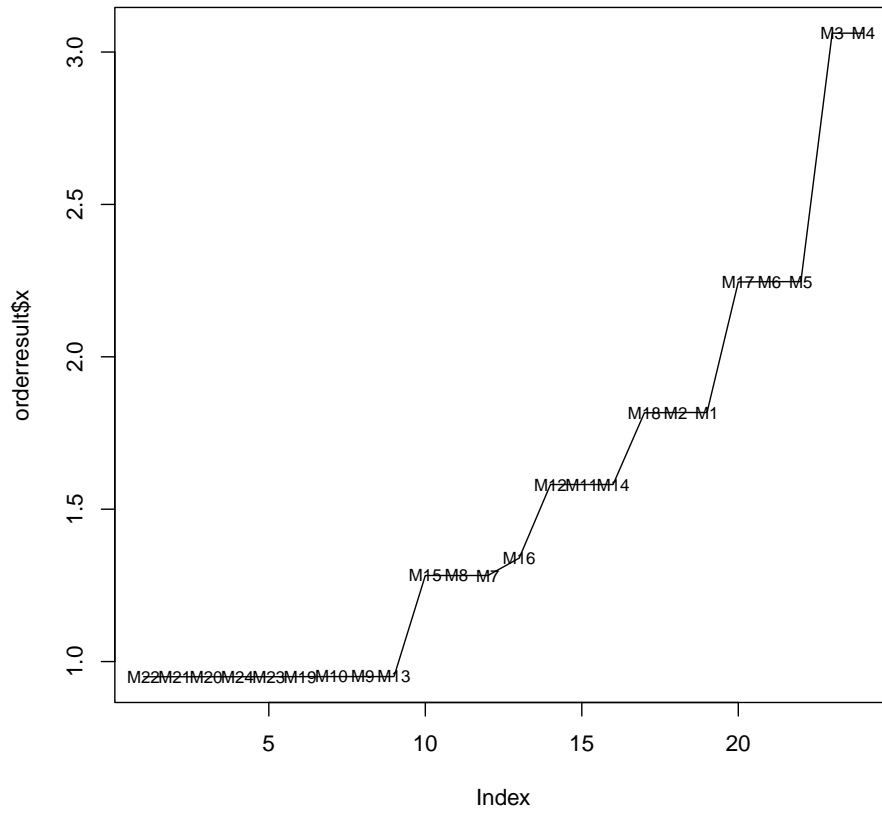


Figure 2.3 Scree plot of $Model_{10}$

Figure 2.3 has a obviously knee point between M13 and M15, and models before this knee point all contain the true model M10.

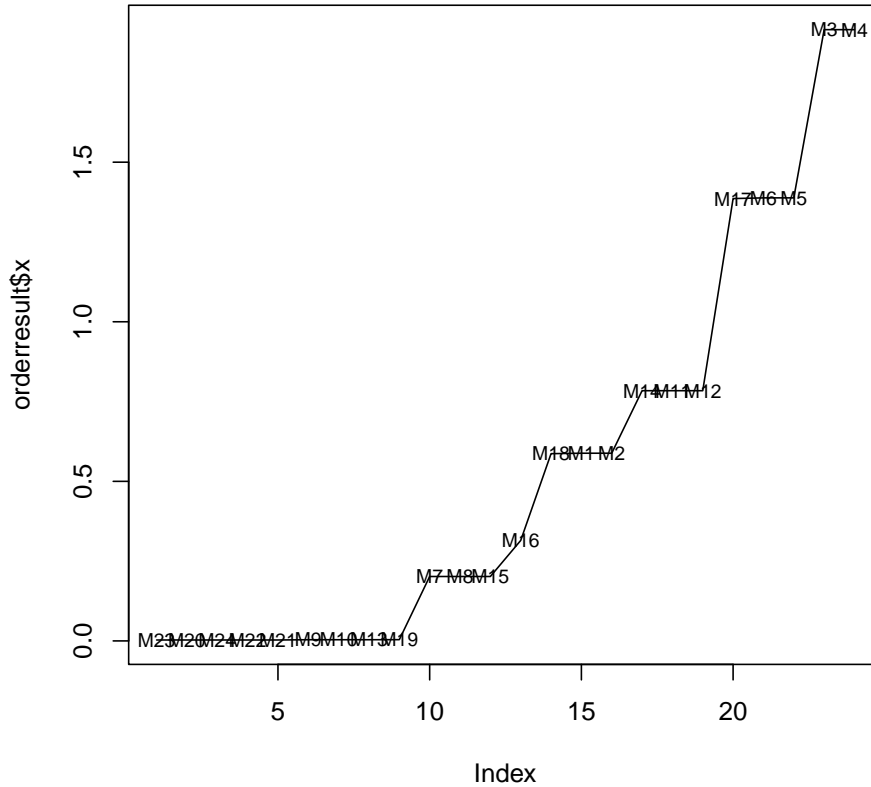


Figure 2.4 Scree plot of $Model_{20}$

Figure 2.3 has a obviously knee point between M19 and M7, but some models before this knee point don't contain the true model M10, such as M9 M10 M13. And these three models just miss one relationship in true model.

(2) **MLE** (2) can not find the true model exactly, it favours complex models.

The models which contain all relationships in the true model and some other relationships more than the true model always have lower value than the true model in the criterion(1). So the most complex model which contain the true model would be the **MLE**(2). for example, in Figure 2.2 ,the true model is $Model_1$, but the MLE is $Model_{20}$ which contain $Model_1$.

(3) **MLE**(2) can not distinguish the direction of relationships.

The change of directions almost has no impact to the value of the criterion(1).

Even when variances of measurement-error are different, **MLE** still can not distinguish the direction.

(4) The model modification tests(likelihood ratio test page 138) of **MLE** works good.

MLE(2) always a models which contain the true model, but more complex than the true model. Fortunately, model modification tests can help to test out relationships that don't exist.

For example, sample x comes from $Model_1$, but the **MLE**(2) is $Model_{20}$. The result of model modification tests is in Table.1

Talbe.1 The result of model modification tests

	Estimate	<i>Std - Error</i>	<i>z - value</i>	$Pr(z)$	
β_{21}	1.052173	0.073526	14.31031	0.00000	$x_2 \leftarrow x_1$
β_{13}	0.047524	0.068924	0.68952	0.49050	$x_1 \leftarrow x_3$
β_{23}	-0.152060	0.069013	-2.20335	0.02757	$x_2 \leftarrow x_3$

3. The confidence set of **SEM**

Section two shows that the **MLE** of $\theta = (\gamma, \beta)$, which contains parameters that represent the different between structures(models), is not convergent to the true value exactly. Such as in the example above, $\gamma \in \{1, 2, \dots, 24\}$ stands for 24 models, and estimating γ is the selection of structures. But **MLE** is not a good estimator of γ , it could not found out the true structure. So we try to built a confidence set of **SEM** by log-likelihood radio.

Fang[2] gave a general way to built confidence set, we choose log-likelihood radio as design function

$$h(x, \theta) = \log P(\hat{\theta}|x) - \log P(\theta|x)$$

where $\hat{\theta}$ is the **MLE** of θ . Set

$$G(x, \theta) = P_{\theta}(h(X, \theta) \geq h(x, \theta))$$

where $x \in R^3$ and $\theta \in \{1, 2, \dots, 24\} \times R^p$. For given $\alpha \in (0, 1)$ and any $x \in R^3$, define the confidence set

$$S(x) = \{\theta : \theta \in \{1, 2, \dots, 24\} \times R^p \text{ and } G(x, \theta) > \alpha\}$$

For every $\theta \in \{1, 2, \dots, 24\} \times R^p$, the distribution of $h(x, \theta)$ shows by numerical results.

Algorithm 1:

- (1) Fix $\theta_0 \in \{1, 2, \dots, 24\} \times R^p$ and $K=2000$ (iterations).
- (2) Generate $x^{(i)}$ from $P(x|\theta_0)$, $i = 1, 2, \dots, K$.
- (3) Calculate $h^{(i)} = \log(|C_{\hat{\theta}}|) + \text{trace}(SC_{\hat{\theta}}^{-1}) - \log(|C_{\theta_0}|) - \text{trace}(SC_{\theta_0}^{-1})$, where $\hat{\theta}$ is the **MLE** of θ , $i = 1, 2, \dots, K$.
- (4) Plot the histogram of $h(x, \theta_0)$ by sample $h^{(i)}$, $i = 1, 2, \dots, K$

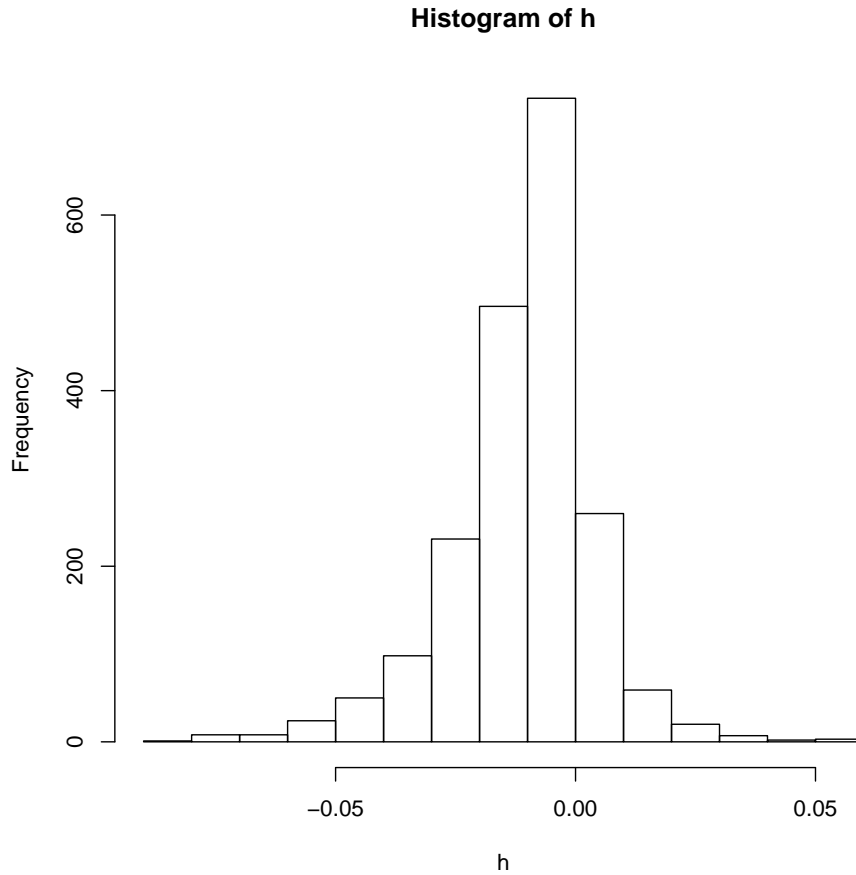


Figure 2.5 Histogram of h in $Model_1$

Figure 2.5 is the histogram of $h(x, \theta_1)$, where $\theta_1 = 1 \times \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

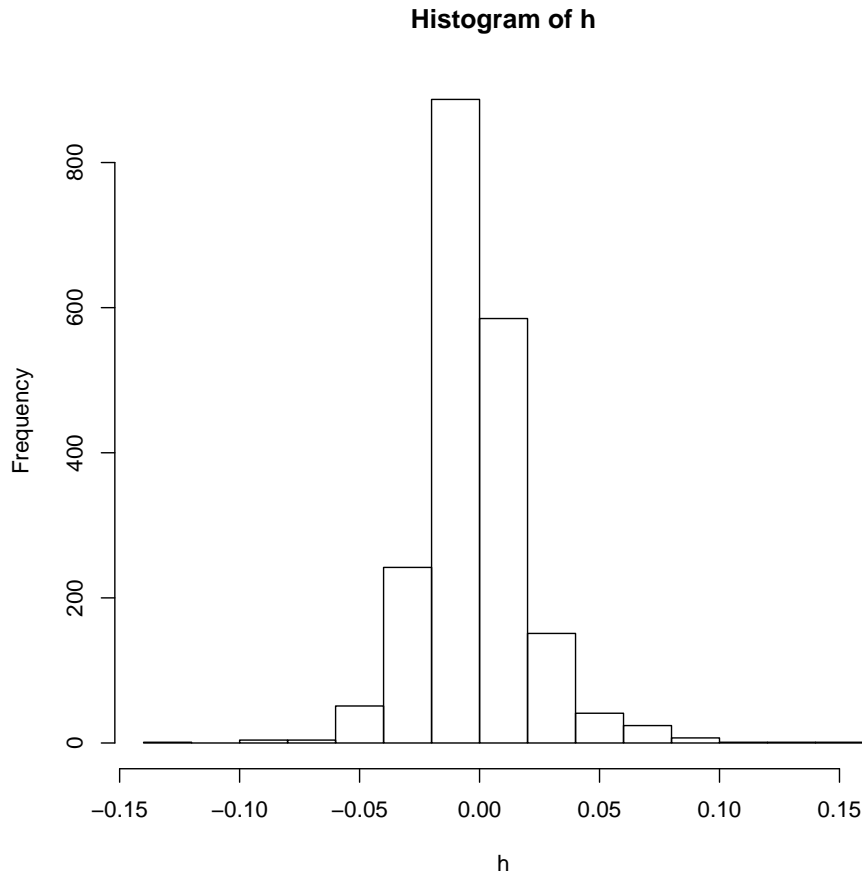


Figure 2.6 Histogram of h in $Model_{10}$

Figure 2.6 is the histogram of $h(x, \theta_{10})$, where $\theta_{10} = 10 \times \begin{pmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}$

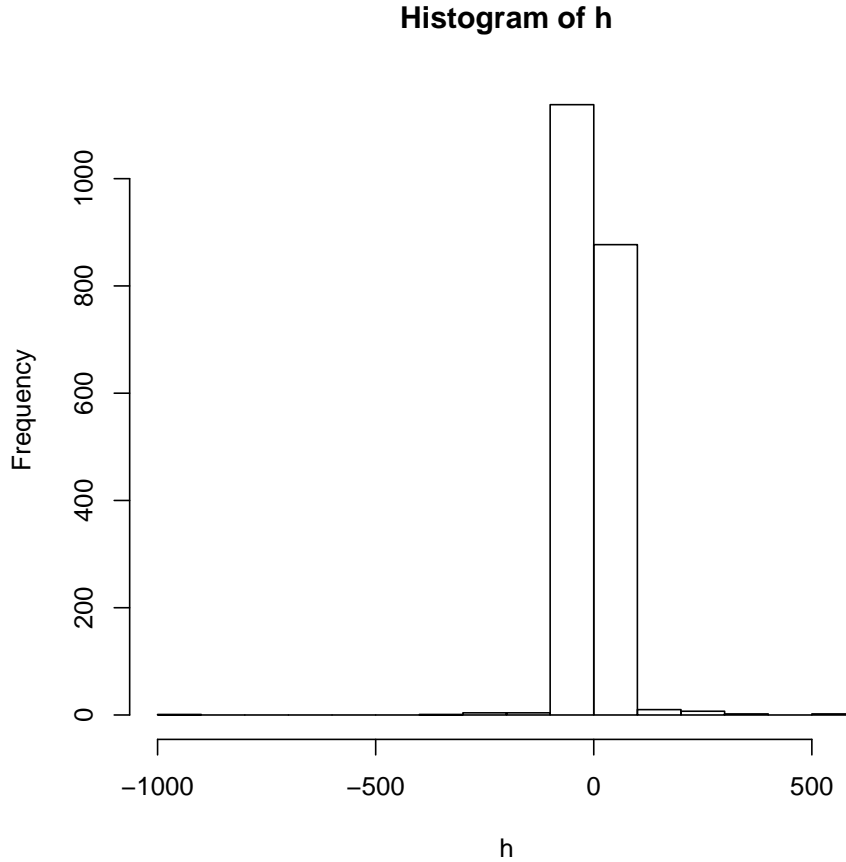


Figure 2.7 Histogram of h in $Model_{20}$

Figure 2.6 is the histogram of $h(x, \theta_{20})$, where $\theta_{20} = 20 \times \begin{pmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}$

Above three pictures show that with different θ the shapes of the distribution of $h(x, \theta)$ are different, and the distributions are not symmetrical. For examples, $Model_1$ is partial to left, $Model_{10}$ is partial to right, and the variance of $Model_{20}$ is smaller than the other two models'.

For any θ_0 , if $P_{\theta_0}(h(X, \theta_0) \geq h(x, \theta_0)) > \alpha$, then θ_0 belong to the confidence set that we built. Using the numerical results above, $P_{\theta_0}(\theta_0 \in S(x))$ can be estimated.

Algorithm2:

- (1) Fix $\theta_0 \in \{1, 2, \dots, 24\} \times R^p$, $k=200$ (iterations) and $\alpha = 0.05$.
- (2) Generate $x^{(j)}$ from $P(x|\theta_0)$, $j = 1, 2, \dots, k$.

(3) Calculate $I^{(j)} = I\left\{\frac{\#\{j:h^j \geq h(x^{(j)}, \hat{\theta})\}}{K} > \alpha\right\}$, where $K = 2000, j = 1, 2, \dots, k$.

(4) Calculate $\hat{P}_{\theta_0}(\theta_0 \in S(x)) = \frac{\sum I^{(j)}}{k}$.

$\hat{P}_{\theta_0}(\theta_0 \in S(x))$ and the numerical under quantile of the distribution of $h(x, \theta_0)$ are given in the Table.2.

Talbe.2 p-value and quantile

	<i>coveragerate</i>	<i>under quantile of $\alpha = 0.5$</i>
θ_1	0.885	0.009193197
θ_{10}	0.94	0.03515166
θ_{20}	0.975	0.03602372

References

[1] John Fox, Structural Equation Modeling With the sem Package in R ,STRUC-TURAL EQUATION MODELING,13(3),465-486.

[2] Xiangzhong Fang, Construction of Confidence Regions Without Pivot.