Learning from Correlation

Li Zhao

Shenzhen University

2016 10 15

outline

Problem description

main sample set

$$\mathscr{X}^* = \{(\mathbf{x}_i^*, \mathbf{t}_i^*)\}_{i=1}^N \subset \mathbf{R}^n \times \mathbf{R}^1,$$

auxiliary sample set

$$\mathscr{X} = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \subset \mathbf{R}^m \times \mathbf{R}^1,$$

relation

$$cov(T^*, T) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

conditional distribution of normal distribution

Suppose that ξ is a p-dimensional random variable, and η is a q-dimensional random variable, the joint distribution of (ξ, η) is $N_{(p+q)}((\beta, \mathbf{0})', \ \mathbf{\Sigma})$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

Then
(i) the conditional distribution of ξ given $\eta = z_2$ is

$$\xi | \eta = z_2 \sim N_p(\beta + \Sigma_{12}\Sigma_{22}^{-1} z_2, \ \Sigma_{11.2}), \tag{1}$$

where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.
(ii) $\xi - \Sigma_{12}\Sigma_{22}^{-1}\eta$ is independent of η.

$$\xi - \Sigma_{12}\Sigma_{22}^{-1}\eta \sim N_p(\beta, \ \ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

conditional expectation improved estimator

If $\Sigma_{12}\Sigma_{22}^{-1}$ is known, $\xi - \Sigma_{12}\Sigma_{22}^{-1}\eta$ is called as the conditional expectation improved estimator (CEIE) of β.

property

1. The variance of $\xi - \Sigma_{12}\Sigma_{22}^{-1}\eta$ is always smaller than ξ if Σ is known.
2. The variance of $\xi - \Sigma_{12}\Sigma_{22}^{-1}\eta$ is smaller than ξ when $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$ is big if Σ is unknown

generalized canonical correlation variables pair

Suppose Z_1 is an n_1-dimensional random vector, with $E(Z_1)=\beta$, and Z_2 is an n_2-dimensional random vector, with $E(Z_2)=0$. The covariance matrix of $(Z_1^T, Z_2^T)^T$ is

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11} \geq 0$, $\Sigma_{22} \geq 0$, and $\Sigma_{12} \neq 0$. The singular value decomposition of Σ_{12} is

$$\Sigma_{12} = P \Lambda Q^T,$$

where $P=(p_1,\cdots,p_m)$, $Q=(q_1,\cdots,q_m)$, $P^TP=Q^TQ=I_m$, and $\Lambda = diag(\lambda_1,\cdots,\lambda_m), \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$. Then, $(p_i^T Z_1,\ q_i^T Z_2)$ is defined as the ith generalized canonical correlation variables pair and λ_i is defined as the ith generalized canonical covariance, for $i=1,2\cdots m$.

generalized canonical correlation variables improved estimator

Under the same assumptions as those in Definition 1, assume $P_0 = (P, P_c)$ is an orthogonal matrix of order n_1, and Z_1 is an unbiased estimator of β. Note that

$$Z_1 = I_{n_1} Z_1 = P_0 P_0^T Z_1 = P_0 (P_0^T Z_1) = P_0 \left(p_1^T Z_1, \ldots, p_m^T Z_1, (P_c^T Z_1)^T \right)^T .$$

When Σ is known, replace $p_i^T Z_1$ by $p_i^T Z_1 - \frac{\lambda_i}{q_i^T \Sigma_{22} q_i} q_i^T Z_2$, for $i = 1, \cdots, m$. Then,

$$\begin{aligned}\hat{\beta}_{GC} =& P_0 \left(p_1^T Z_1 - \frac{\lambda_1}{q_1^T \Sigma_{22} q_1} q_1^T Z_2 \ldots p_m^T Z_1 - \frac{\lambda_m}{q_m^T \Sigma_{22} q_m} q_m^T Z_2, (P_c^T Z_1)^T \right)^T \\ =& Z_1 - P \left(\frac{\lambda_1}{q_1^T \Sigma_{22} q_1} q_1 \ldots \frac{\lambda_m}{q_m^T \Sigma_{22} q_m} q_m \right)^T Z_2\end{aligned}$$

is the generalized canonical correlation variables improved estimator (GCCVIE) of β.

property

Under the same assumptions as those in Definition, if $\Sigma_{22} = \sigma^2 \mathbf{I}$,1 then the GCCVIE $\hat{\beta}_{GC}$ is a BLUE of β, based on (Z_1, Z_2).

Two stage estimator

When Σ is unknown Zellner(1962) replaced Σ with its consistent estimator $\hat{\Sigma}$ and obtained a feasible estimator of β, According to restricted residuals, the estimator of Σ is,

$$\hat{\Sigma} = (\hat{\sigma}_{ij})_{(N\times N)}, \quad \hat{\sigma}_{ij} = \frac{1}{T}\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j \tag{2}$$

where

$$\hat{\mathbf{u}}_j = \mathbf{Y}_j - \mathbf{X}_j\hat{\beta}_{jOLS} \triangleq \mathbf{N}_j\mathbf{Y}_j,$$

It can be represented as

$$\hat{\Sigma}_1 = \frac{1}{T}\hat{\mathbf{U}}'\hat{\mathbf{U}},$$

where $\hat{\mathbf{U}} = (\hat{\mathbf{u}}_1,\ \hat{\mathbf{u}}_2, \dots,\ \hat{\mathbf{u}}_N)$.

Table: The MSE of the CRWNN TCRWNN and the single RWNN with different ρ in testing set.

	training set			testing set		
ρ	RWNN	CRWNN	TCRWNN	RWNN	CRWNN	TCRWNN
0.0	0.8091	0.8091	0.8093	1.4225	1.4225	1.4279
0.1	0.7734	0.7739	0.7753	1.3025	1.2757	1.2891
0.2	1.0999	1.1016	1.1005	0.8734	0.8698	0.8704
0.3	1.0535	1.0610	1.0547	0.9001	0.8738	0.8869
0.4	1.0733	1.0797	1.0739	1.2536	1.1188	1.2078
0.5	0.9712	0.9800	0.9784	1.2010	1.1874	1.1876
0.6	0.8396	0.8743	0.8659	1.0620	1.0109	1.0159
0.7	1.0577	1.0925	1.0598	1.5172	1.1632	1.4189
0.8	0.8685	0.8918	0.8896	1.3744	1.1871	1.1926
0.9	0.9649	1.0058	1.0026	0.9911	0.9401	0.9408