Supervised Learning Vector Quantization for Projecting Missing Weights of Hierarchical Neural Networks

CIN-RU CHEN[1], LIANG-TING TSAI[2], & CHIH-CHIEN YANG[3]
[1]Department of Information Management
Ta Hwa Institute of Technology
1 Dahua Rd, Qionglin Shiang, Hsinchu
TAIWAN
cchen111@ms45.hinet.net

[2]Cognitive NeuroMetrics Laboratory
Graduate Institute of Educational Measurement & Statistics
National Taichung University
TAIWAN
tlt@ms3.ntcu.edu.tw

[3]Cognitive NeuroMetrics Laboratory
Graduate Institute of Educational Measurement & Statistics
National Taichung University
TAIWAN
noahyang@ntcu.edu.tw

Abstract: A supervised learning vector quantization (LVQ) method is proposed in this paper to project stratified random samples to infer hierarchical neural networks. Comparing with two traditional methods, i.e., list-wise deletion (LWD), and non-amplified (NA), the supervised LVQ shows satisfying efficiencies and accuracies in simulation studies. The accomplishments of proposed LVQ method can be significant for sociological and psychological surveys in properly inferring the targeted populations with hierarchical neural network structure. In the numerical simulation study, successes of LVQ in projecting samples to infer the original population are further examined by experimental factors of sampling sizes, missing rates, and disproportion rates. The experimental design is to reflect practical research and under these conditions it shows the neural network approach is more accurate and reliable than its competitors.

Key-Words: Neural Network, Learning Vector Quantization, Missing Weights, Stratified Structure, Simulation Study, Large Scale Data.

1 Introduction

Random samples drawn from a stratified super population (e.g., Progress in International Reading Literacy Study, PIRLS[1], Trends in International Mathematics and Science Study, TIMSS[2], Programme for International Student Assessment, PISA[3], National Longitudinal Survey of Youth, NLSY[4], The National Assessment of Educational Progress, NAEP[5], Early Childhood Longitudinal Study, Kindergarten Class of 1998-99[6], etc.) are often extensively studied by researchers in order to infer the most representative properties of their origins. While the application of stratified sampling is the core of study design in sociological and psychometrical survey research, to properly amplify these small portions of the entire population just cannot be over emphasized, e.g., [7][8][9][10][11][12]. The process of amplification needs some critical information, e.g., gender, racial, or some personal background to restore original sampling proportions due to the stratification effects[13][14]. Yet the information can be sensitive and respondents may refuse to report due to the rise of self-consciousness. Without the critical information, the amplification process is likely to be failed.

This critical information—that is, group membership and sampling weights--can not be neglected when inferring the characteristics of a population. In other words, the proper application of group membership and sampling weights to amplify random samples to stratified characteristics of a population was fairly important [15][16][17][18][19][20][21]. Without a proper method to impute this information to substitute the traditional methods (such as list-wise deletion and non-amplified), seriously biased conclusions are possible.

To counter this effect, we proposed a neural network approach, learning vector quantization (LVQ), to estimate the possible information of the missing data. The principal function of LVQ is to classify information and make predictions on missing information [22][23][24][25]. In order to ensure the accuracy and consistency of the proposed method, this study chose the CFA (confirmatory factor analysis) model as its experimental model, and simulated practical survey research. Otherwise, this study contains various sampling sizes, missing rates, and disproportion rates to assess the accuracy and consistency of performances of LVQ.

2 LVQ Algorithm

To solve the problem of missing information in inferring the original population, this study proposed utilizing LVQ to infer the missing group memberships and weights of the study subjects. Firstly, LVQ established relationships between group memberships and the dependent variables by subjects with complete group membership and dependent variable data. After these relationships have been successfully established, we were then able to utilize the relationships to estimate the losing information. Finally, the researcher must simply impute the information for the subjects when inferring the original psychometrical population.

The LVQ algorithm is one of the supervised learning neural networks first presented by Kohonen and it was especially designed to accomplish pattern classification assignments [26][27][28][29]. The LVQ network was displayed in Fig. 1. The LVQ structure consisted of input, hidden, and output layers. The input layer conveyed input data to the network; a hidden layer dealt with the actual data information, and an output layer generated a particular class or category from the input layer. The input layer was associated with the hidden layer through reference vectors, which were updated through the learning. The hidden layer to output layer was associated and the weights were fixed at 1 [17].

The algorithm of an LVQ included learning and recalling processes. In the learning process, in order to achieve accurate classification in the output layer, the Euclidean distance (D_i) was utilized as a basic rule of competition to find the winner [30][31][32]. The simplest LVQ learning process is as follow [10][29][32][33][34].

Step 1: Initialize the reference vector Z_i of neuron i .

Step 2: Input a training paradigm vector Y and a corresponding category to the network.

Step 3: Calculate the Euclidean distance between Y and Z_i , where Y_j and Z_{ij} are the j th elements of Y and Z_i , respectively.

$$D_i = \|Z_i - Y\| = \sqrt{\sum_j (Z_{ij} - Y_j)^2} \qquad (1)$$

Step 4: Update the reference vector Z_i that was closest to the input paradigm vector. The neuron was called the winner when it had the minimum distance (i.e. the reference vector Z_c with the smallest Euclidean distance with regard to the input paradigm vector Y).This sole winner was allowed to correct the reference vector by using the following formulas. If the winner Z_c and Y belong to the same category (the classification has been correct), then

$$Z_i(t+1) = Z_i(t) + \alpha(t) \times \big(Y(t) - Z_i(t)\big) \qquad (2)$$

However, if the winner and learning vector belong to different category (the classification has been incorrect), then

$$Z_i(t+1) = Z_i(t) - \alpha(t) \times \big(Y(t) - Z_i(t)\big) \qquad (3)$$

where $\alpha(t)$ was the learning rate. The rate is a monotonically decreasing function of time t which controlled how quickly the reference vector is allows to change. In this study, the initial identification of rate was 0.2 and it continued decreasing by multiplying a constant 0.9.

Step 5: The researcher should return to the Step 2, input a new learning vector and repeat the process until the neural network was stabilized (i.e. the difference of $Z_i^{(t+1)} - Z_i^t$

converges to the stopping criterion) or number of iterations had been carried out and the optimum reference vector obtained.

After completing the learning process, the next procedure is the recalling process. The researcher could then use the optimum reference vector from the learning process to predict the subjects to which the missing group memberships belong. When the CFA model was estimated, the researcher simply had to estimate group weights and imputed them to the samples.

To obtain a better classification performance when using the LVQ algorithm to estimate missing group memberships, many parameters must be identified, including the learning coefficient, number of hidden layers, number of learning paradigms, etc. The improper identification of these parameters will negatively influence the classification accuracy and learning time. This study adopted the same parameter definitions as Tsai and Yang [10][34]. The number of hidden layers was 2. The learning coefficient was 0.2, and the constant k was 0.9. The definition of the number of learning paradigms is adopting all the subjects with known group membership and repeating the learning process until all paradigms are used.

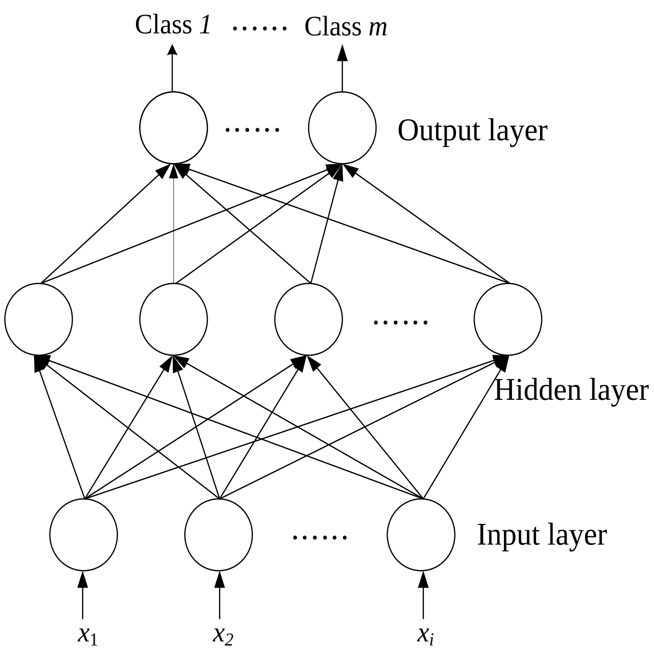

Fig 1. LVQ network model

3 Research Design

The primary purpose of this study was to evaluate the stability and accuracy of applying the LVQ in amplifying random samples to infer their original psychometrical population. A total population of 100,000 was combined with two groups of 80,000 and 20,000 respectively. The both groups all followed a basic CFA model with only one latent variable (η) estimated by five continuous observations ($y_1,\ldots,y_5$), but differed in terms of factor loadings. The CFA model was as follows:

$$y_i = \lambda_j \eta + \varepsilon_i \tag{4}$$

where λ was the factor loading, η was the latent variable, and ε was the measure errors. The path diagram (Fig. 2) was drawn in to illustrate the model structure.

The difference between the two groups was a second factor loading (λ_2) variance of 0.4; all the other parameters were designed to be equal between the two groups. The factor loadings were 1, 0.8, 0.8, 0.8, and 0.8 for group 1 and 1, 0.4, 0.8, 0.8, and 0.8 for group 2, respectively (see Table 1). The latent variable and measurement errors for both groups were generated from standard normal distribution in all conditions.

Table 1 Factor loadings used to generate two groups artificial subjects

	Group 1	Group 2
λ_1	1	1
λ_2	0.8	0.4
λ_3	0.8	0.8
λ_4	0.8	0.8
λ_5	0.8	0.8

For the sake of approximation to actual sociological and psychometrical survey research, this study designed a variety of sampling sizes (200, 600, and 1000), missing rates (5%, 10%, and 15%), and disproportion rates (R=8, 4, 2, and 1). These three various sampling sizes were chosen to represent small, medium, and large sample sizes. To assess the practicability of LVQ, all of the three proportions of subjects with missing information occurred completely occur in group 2, it was to have them as examples for missing at random (MAR)[35][36][37][38]. Otherwise, four levels of the disproportionate sampling of groups were defined by the ratio of sampling weights of the two groups. For instance, the under conditions were R=4 and 600 sampling sizes, for group 1, with 300($n_1 = 300$) subjects drawn at random whose sampling weight was 266.667 ($w_1 = \frac{80{,}000}{300} = 266.667$). For group 2, 300($n_2 = 300$) subjects were drawn at random, and the sampling weight was 66.667 ($w_2 = \frac{20{,}000}{300} = 66.667$). In this example, the sampling for group 1 and group 2 implied under and over-sampling, respectively. However, missing data has become fairly common for respondents to refuse to take questionnaires, or to leave certain sensitive questions blank. Therefore, in the example aforementioned, 5% of the whole sampling sizes(n) with missing group memberships and sampling weights, namely, 30 subjects, were in group 2; the remaining subjects with complete information was 270. Similarly, the remaining complete subjects with 10% and 15% missing data rates were respectively 240 and 210. The designs of the missing information rates were adopted and revised from the studies of Ender and Bandalos[35] and Ender and Peugh[36]. Besides, the ratio was equal to 1 to represent the proportionate sampling of the two groups. Other specific sampling designs were as outlined in Table 2 below.

In this study, all the missing data were designed to completely occur in group 2; it was to have them as examples for MAR. When MAR holds, according to the methodologies of the missing data analysis, researchers may be able to interpolate or evaluate the proper missing data with complete data to ensure correct statistical inferences for various population characteristics [35][36]. These theories state that it is possible to use the completed data to evaluate or infer the missing group membership of the subjects.

The processes involved in this practical simulation study consisted of three steps. First, the subjects of both groups were generated by Mplus 4,12[39] software following the structure of CFA model and the factor loading in Fig. 2. The authors wrote a Matlab 7.7.0 computer program for the total sampling sizes which were sampled randomly from the two individual groups, and whose subjects with missing information were removed according to the missing proportions in Table 2 after data generation had been completed. Second, one of three methods (LWD, non-amplified, and LVQ) was utilized to handle the data with missing information. The first method was list-wise deletion (LWD), which deleted the subjects with missing information, and used the remaining complete subjects to infer the original psychometrical population. The second method, non-amplified(NA), directly ignored the group membership and sampling weights. The third method used to interpolate the missing information was LVQ. In the example aforementioned, LVQ was utilized to classify the 30 subjects with missing information under 5% missing data rate. The weights recorded 266.667, when LVQ inferred the subjects to group 1. To the contrary, when LVQ inferred the subjects to group 2, the weights recorded 66.667. Finally, the datasets were estimated using the software Mplus 4.21[39] with the estimator of pseudo maximum likelihood (MLR) estimation algorithm. The MLR estimate was obtained by maximizing the weighted log-likelihood [15][18][39].

$$\log(L) = \sum_i w_i \log(L_i) \tag{5}$$

where the i was sampling weight for subjects i . To confirm the stability of the numerical simulation, 200 replications were performed for each condition.

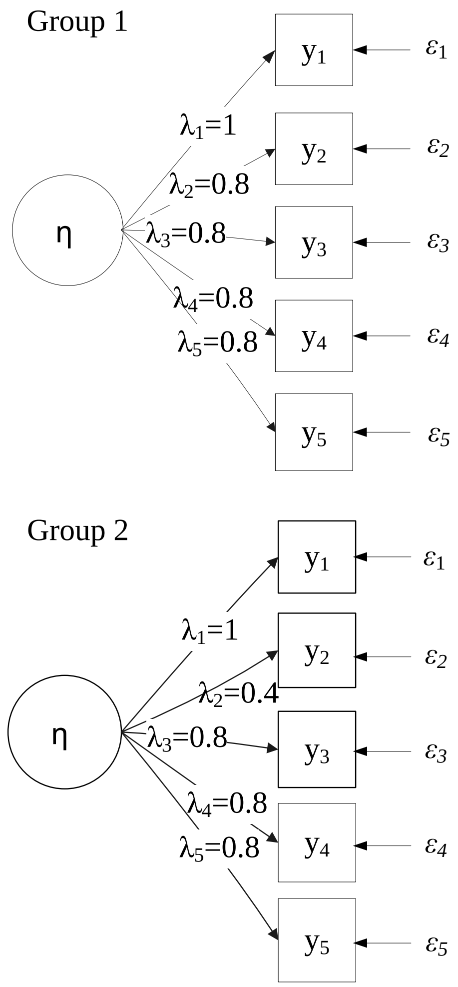

Fig. 2 Path diagram for CFA model

Table 2 Sampling design for the simulation study

Total Sampling Sizes (n)	R	Projected Group Sizes		% & Sizes of Unknown Groups		
				5%	10%	15%
		n_1	n_2	n_2	n_2	n_2
200	8	67	133	123	113	103
	4	100	100	90	80	70
	2	133	67	57	47	37
	1	160	40	30	20	10
600	8	200	400	370	340	310
	4	300	300	270	240	210
	2	400	200	170	140	110
	1	480	120	90	60	30
1000	8	333	667	617	567	517
	4	500	500	450	400	350
	2	667	333	283	233	183
	1	800	200	150	100	50

4 Results

This section will discuss the stability and accuracy of the three methods to deal with the missing information under different experimental factors of sampling sizes, missing rates, and disproportion rates while inferring the original population characteristics. The simulate results were summarized in Table 3. The last three columns, named LWD, NA, and LVQ, recorded the 95% confidence interval coverage rates of the second factor loading (λ_2) to provide comparisons of the three methods for dealing with the missing information. The missing rates, sampling sizes, and disproportion rates (R) were listed in the first three columns. The 95% confidence interval coverage rates recorded the 95% confidence interval covering the true population factor loadings. 95% coverage rates for any method need to be near the theoretical 95% value to demonstrate an acceptable accomplishment in terms of inferring the targeted population characteristics[9][40]. The following subsections discuss the results with respect to different experimental factors.

Table 3 95% coverage rates of CFA parameter

Missing Rates	Sampling Sizes	R	95% coverage rates		
			LWD	NA	LVQ
5%	200	8	0.545	0.515	0.935
		4	0.790	0.765	0.950
		2	0.925	0.910	0.950
		1	0.930	0.930	0.930
	600	8	0.175	0.155	0.940
		4	0.475	0.450	0.915
		2	0.875	0.840	0.925
		1	0.970	0.955	0.955
	1000	8	0.035	0.025	0.930
		4	0.345	0.300	0.945
		2	0.855	0.780	0.950
		1	0.935	0.945	0.945
10%	200	8	0.565	0.515	0.930
		4	0.825	0.765	0.950
		2	0.935	0.910	0.955
		1	0.950	0.930	0.930
	600	8	0.225	0.155	0.945
		4	0.560	0.450	0.895
		2	0.900	0.840	0.915
		1	0.970	0.955	0.955
	1000	8	0.060	0.025	0.930
		4	0.440	0.300	0.925
		2	0.910	0.780	0.950
		1	0.900	0.945	0.945
15%	200	8	0.605	0.515	0.925
		4	0.835	0.765	0.935
		2	0.960	0.910	0.950
		1	0.950	0.930	0.930
	600	8	0.265	0.155	0.930
		4	0.645	0.450	0.910
		2	0.920	0.840	0.925
		1	0.945	0.955	0.955
	1000	8	0.080	0.025	0.925
		4	0.615	0.300	0.910
		2	0.920	0.780	0.915
		1	0.910	0.945	0.945

4.1 Sampling size

Fig. 3 displayed the 95% coverage rates profiles under different sampling sizes. Numerical results were also recorded in Table 3. The left side of Fig. 3 showed the tendency of the coverage rates of LWD method. The central and right side displayed the tendency of the coverage rates of NA and LVQ method, respectively.

Many comparisons could be made from Table 3 and Fig. 3. The LVQ method could produce higher coverage rates for the CFA parameter than both LWD and non-amplified methods in almost all conditions. The coverage rates of the LWD and non-amplified methods decreased as the sampling sizes increased. For instance, under the conditions of 5% missing data rate and 8 disproportion rate, for sampling size 200, 600, and 1000, the LWD method had coverage rates of 0.545, 0.175, and 0.035 respectively. Under same condition, the non-amplified method had coverage rates of 0.515, 0.155, and 0.025 respectively. This meant that the proportion of successful inferences of the most representative properties of their targeted population decreased commensurately. The coverage rates of LVQ were not nearly affected by the sampling sizes; they remained above 91% and approached 95%. In the condition aforementioned, the LVQ method had coverage rates of 0.935, 0.940, and 0.930. Otherwise, the LWD and non-amplified methods were sensitive to sampling sizes. To sum up, LVQ was the most stable and accurate among the three methods for sociological and psychological surveys in properly inferring the targeted populations with missing information subjects.

4.2 Disproportion rates

Since the over- and under-sampling conditions generally appeared in the sociological and psychometrical survey research, the group membership and weights played an important role in accurately inferring the population characteristics. Fig. 4 displayed the 95% coverage rates profiles under different sampling rates. Numerical results were also recorded in Table 3. The left side of Fig. 4 showed the tendency of the coverage rates of sampling sizes 200. The central and right side displayed the tendency of the coverage rates of sampling sizes 600 and 1000, respectively.

Under disproportion sampling rates, the results showed that the successes of LVQ in amplifying samples to infer the original population were not affected by the disproportion sampling rates. For instance, under the condition of 10% missing data rate and sampling size 200, for disproportion rates 8, 4, 2, and 1, the LVQ method had coverage rates of 0.930, 0.950, 0.955, and 0.930, respectively. On the contrary, the coverage rates of LWD and non-amplified methods decreased as the disproportion sampling rates increased. Under same condition, the non-amplified method had coverage rates of 0.515, 0.765, 0.910, and 0.930 respectively. The LWD method had coverage rates of 0.565, 0.825, 0.935, and 0.950 under disproportion rates 8, 4, 2, and 1, respectively. The coverage rates of LVQ were not nearly affected by the sampling rates; they approached the theoretical 95% value to demonstrate an acceptable accomplishment contrary to LWD and non-amplified methods. The coverage rates of LWD and non-amplified methods decreased as the sampling rate increased.

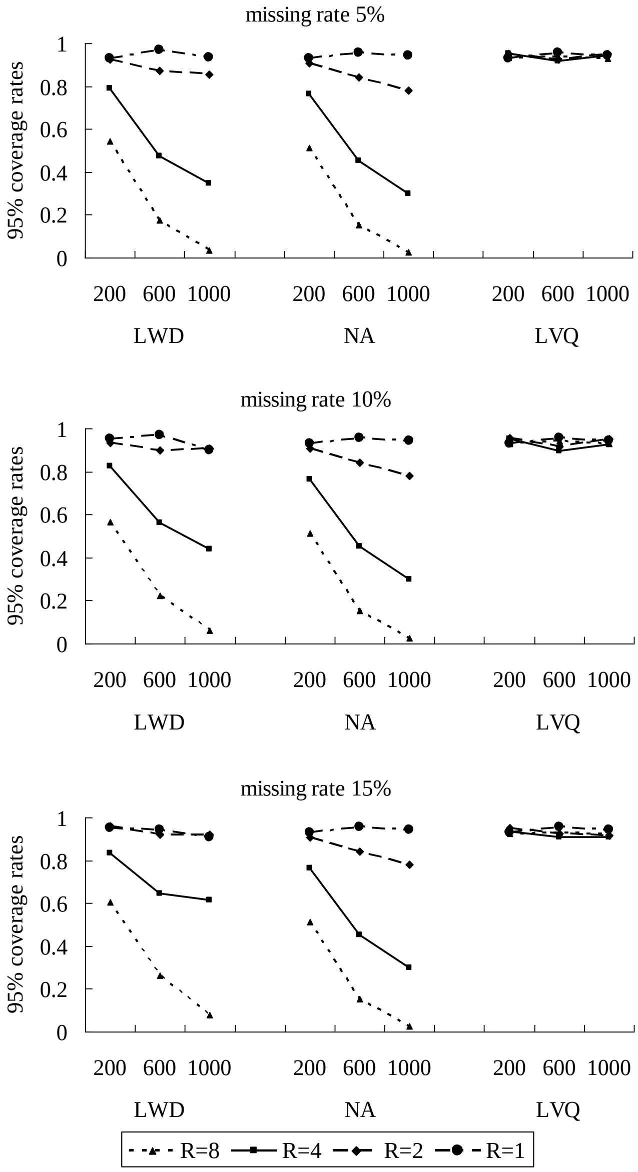

Fig. 3 95% coverage rates for different sampling sizes

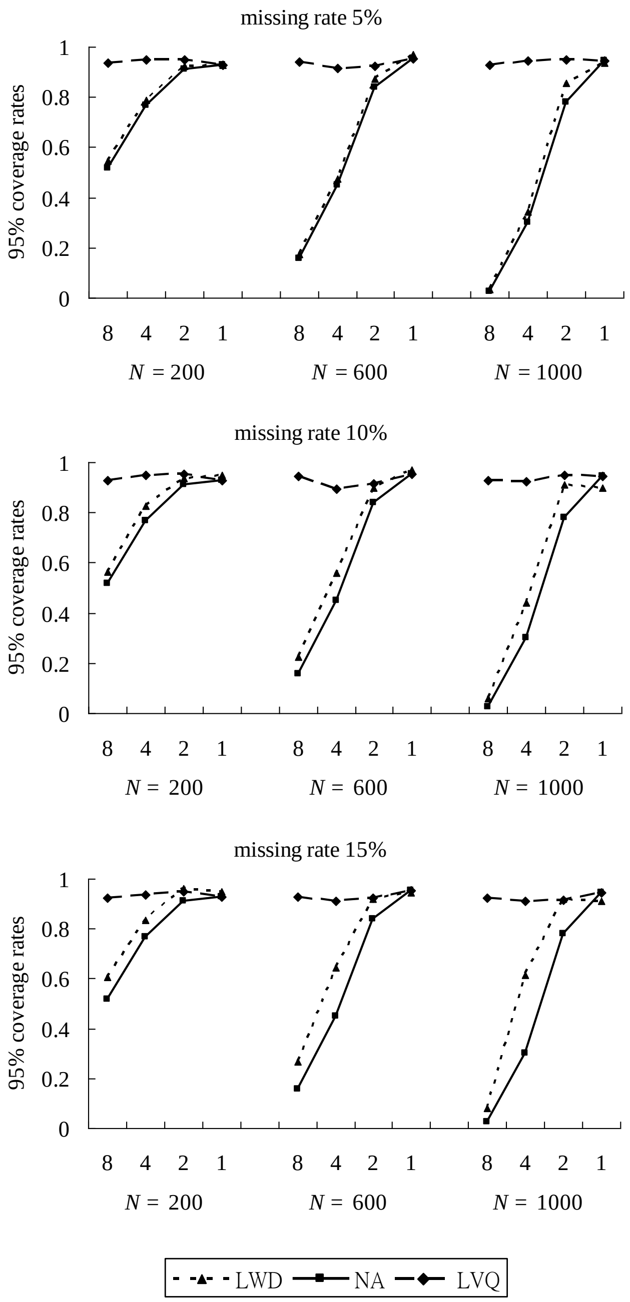

Fig. 4 95% coverage rates for different sampling rates

4.3 Missing rates

Fig. 5 displayed the 95% coverage rates profiles under different missing rates. Numerical results were also recorded in Table 3. The left side of Fig. 5 drew the tendency of the coverage rates of LWD method. The central and right side showed the tendency of the coverage rates of NA and LVQ method, respectively.

Many comparisons could be made from Table 3 and Fig. 5. The LVQ method could produce higher coverage rates for the CFA parameter than both LWD and non-amplified methods for almost all conditions. The coverage rates of the LWD method were slightly affected by the missing rates. For instance, under the condition of 8 disproportion rate and sampling size 200, for missing rates 5%, 10%, and 15%, the LWD method had coverage rates of 0.545, 0.565, and 0.605 respectively. Under the same condition, the non-amplified method had the same coverage rates of 0.515. This meant that the LWD and non-amplified methods could not successfully infer the properties of their targeted population. Oppositely, however, the coverage rates of LVQ were not nearly affected by the missing data rates; they approached the theoretical 95% value to demonstrate an acceptable accomplishment in terms of inferring the targeted population characteristics. In the condition aforementioned, the LVQ method had coverage rates of 0.935, 0.930, and 0.925. In conclusion, LVQ was the most stable and accurate among the three methods for sociological and psychological surveys in properly inferring the targeted populations with missing information subjects.

5 Conclusion

Results of the computerized numerical experiment showed that the Learning Vector Quantization (LVQ) method was more accurate, stable, and reliable than the LWD and non-amplified methods when inferring the targeted populations in a CFA model with missing information. For example, in different conditions, the LVQ method had an average success rate of 90% or above.

Since the disproportionate sampling generally appears in survey research, the weights play a great role in accurately inferring the population. If the researcher did not apply sampling weights, the accuracy of inferring population could possibly be biased. The empirical research that was still at its beginning stages will try its best to use proportional sampling to avoid the problem of weighting. However, it had become fairly common for respondents to leave certain questions blank or refuse to take questionnaires due to the rise in their self-consciousness. For those who refuse to answer certain questions, they generally share common characteristics such as being distant or egotistic. This study attempted to simulate this phenomenon by locating all the missing information in the same group. Both of these experiments show that no matter the size of the groups in the population, and whether there is a systemic information loss during the investigation procedures, without proper amplifying process, the inference of the targeted populations can turn out to be highly inaccurate.

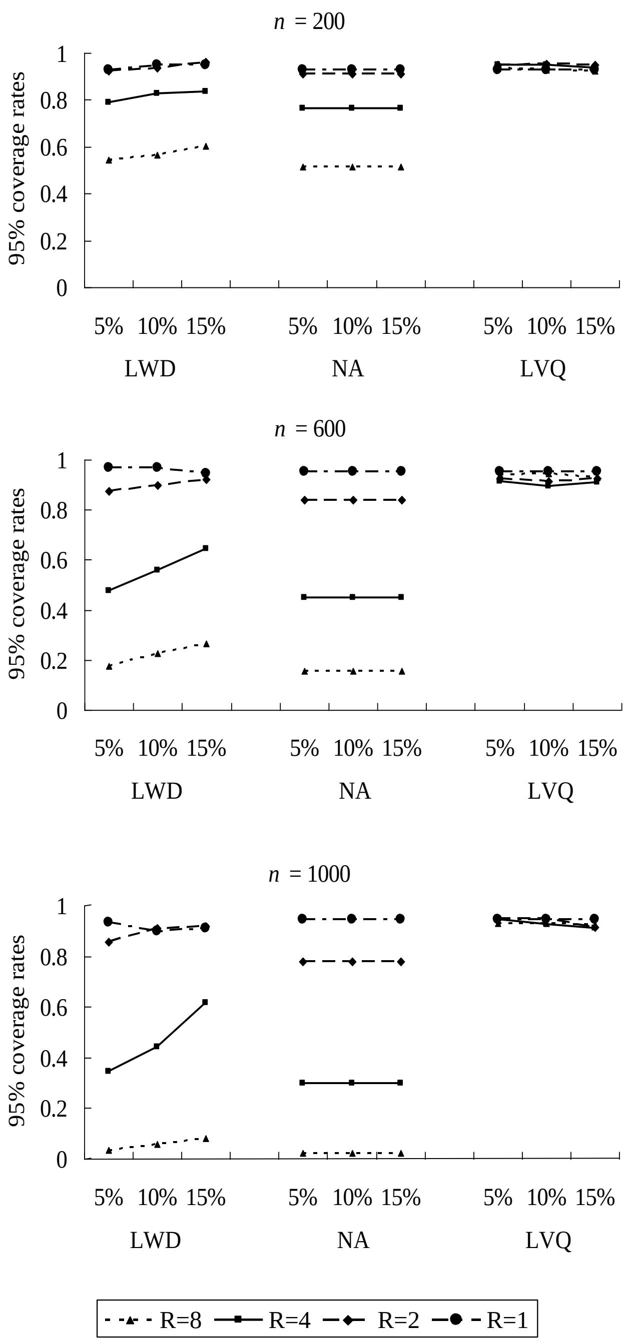

Fig. 5 95% coverage rates for different missing rates

Another phenomenon that was worth mentioning was that when the researcher processes the sampling, if he/she improperly administrates missing information, as the number of sampling size increases, the overall accuracy of estimating the population characteristics would decrease. If the researcher incorrectly uses method for dealing with missing information, the analytic errors would gradually be accumulated as the number of sampling size increases.

In practical application of the LVQ method, besides the self-written computer programs, other applications of computer packages of artificial neural network (e.g., NeuroSolution, PCNeuro, etc.,) can also process LVQ procedure. Therefore, the promotion and utilization of the LVQ method for amplifying random samples to infer their original psychometrical population should not to be difficult. Its practical value and influence can definitely be expected and confirmed.

References:

[1] Martin, M.O., Mullis, I.V.S., Kennedy, A.M. (eds), *PIRLS 2001 Technical Reports.* Publisher: International Study Center, Lynch School of Education, Boston College, 2003.

[2] Gonzalez, E.J., & Miles, J.A. (eds.), *TIMSS 1999 user guide for the International database: IEA's Repeat of the Third International mathematics and Science Study at the Eighth Grade.* Lynch School of Education, Boston College, 2001.

[3] Organization for Economic Co-Operation and Development (OECD), *The PISA 2003 Assessment Framework-Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* OECD, 2003.

[4] Ohio State University, *National longitudinal surveys,* Coiumbus, OH: Center for Human Resource Research, 1994.

[5] NCES. NAEP overview. Received June 30, 2009, from http://nces.ed.gov/nationsreport card/about. 2005

[6] U.S. Department of Education. *ECLS-K, Base year public-use data file, kindergarten class of 1998–99: Data files and electronic code book: (Child, teacher, school files)*. Washington, DC: National Center for Education Statistics, 2001

[7] Grilli, L., & Pratesi, M., Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs. *Survey methodology,* Vol.30 2004, pp. 4-14.

[8] Kaplan, D., & Ferguson, A. J., On the utilization of sample weights in latent variable models. *Structural Equation Modeling,* Vol.6 1999, pp. 305-321.

[9] Tsai, L.T., & Yang, C.C., Influence of sampling weights on SEM parameter estimation accuracy (in Chinese). *Journal of Education and Psychology,* Vol.31 2008, pp. 155-177.

[10] Tsai, L.T., & Yang, C.C., *Effects of LVQ-ESW method in interpolating missing sampling*

weights in a CFA model. 2009, Submitted for publication.

[11] Yang, C. C., & Tsai, L. T., A simulation study on computation and inference accuracy of factor loadings for large data mines. *Wseas Trans on Information Science and Application*, Vol.12 2006, pp. 2577-2579.

[12] Yang, C.C., & Tsai, L.T., Inferring measurement equivalence between Likert-type questionnaires under effects of sampling weights (in Chinese). *Chinese Journal of Psychology*, Vol.50 2008, pp. 257-269.

[13] Lohr, S. L., *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press. 1999.

[14] Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B*, Vol.60 1998, pp. 23-56.

[15] Asparouhov, T., Sampling weights in latent variable modeling. *Structural Equation Modeling*, Vol.12 2005, pp. 411-434.

[16] Korn, E. L. & Graubard, B. I., Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician*, Vol.49 1995, pp.291-295.

[17] Scheaffer, R. L., Mendenhall Ⅲ, W., & Ott, R. L., *Elementary Survey Sampling*. Boston: Duxburt Press. 2006.

[18] Asparouhov, T., General multi-level modeling with sampling weights. *Communications in Statistics Theory and Methods*, Vol.35 2006, pp. 439-460.

[19] Stapleton, L., The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling*, Vol.9 2002, pp. 475-502.

[20] Stapleton, L., An assessment of practical solutions for structural equation modeling with complex sample data. *Structural Equation Modeling*, Vol.13 2006, pp. 28-58.

[21] Stapleton, L. Variance estimation using replication methods in structural equation modeling with complex sample data. *Structural Equation Modeling*, Vol.15 2008, pp. 183-210.

[22] Sato, A. S and Yamada, K., Generalized learning vector quantization. *Advances in Neural Information Processing Systems*, Tesauro, G., Touretzky, D.,and Leon, T., eds., Vol 7 1995, pp. 423-429.

[23] Hammer, B. and Villmann, T., Generalized relevance learning vector quantization. *Neural Networks*, Vol.15 2002, pp. 1059-1068.

[24] Vakil-Baghmisheh, M. T., and Pavesi, N., Premature clustering phenomenon and new training algorithms for LVQ. *Pattern Recognition*, Vol.36 2003, pp. 1901-1912.

[25] Filippi, A. M., and Jensen, J. R., effect of continuum removal on hyperspectral coastal vegetation classification using a fuzzy learning vector quantizer. *IEEE Transactions on geoscience and remote sensing*, Vol.45 2007, pp 1857-1869.

[26] Maritín-Valdivia, M. T., Ureña-López, L. A., & Gracía-Vega., The learning vector quantization algorithm applied to automatic text classification tasks. *Neural Networks*, Vol.20 2007, pp.748- 756.

[27] Brockett, P. L., Golden, L. L., Jang, J., & Yang, C., A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction. *The Journal of Risk and Insurance*, Vol.73 2006, pp. 397-419.

[28] Su, C. T. & Chou, C. J., A neural network-based approach for statistical probability distribution recognition. *Quality Engineering*, Vol.18 2006, pp.293-297.

[29] Yang, M. S. & Yang, J. H., A fuzzy-soft learning vector quantization for control chart pattern recognition. *International Journal of Production Research*, Vol.40 2002, pp.2721-2731.

[30] Doherty, K. A. J., Adams, R. G., & Davey, N., Unsupervised learning with normalised data and non-Euclidean norms. *Applied Soft Computing Journal*, Vol.7 2007, pp. 203-210.

[31] Guh, R. S., & Shiue, Y. R., Effective pattern recognition of control charts using a dynamically trained learning vector quantization network. *Journal of the Chinese Institute of Industrial Engineers*, Vol.25 2008, pp. 73-89.

[32] Pham, D. T. &, Sagiroglu, S., Neural network classification of defects in vector boards. *Journal of Process Mechanical Engineering Part B*, Vol.214 200, pp. 255-258.

[33] White, H., Some asymptotic results for learning in single layer feedforward network models. *Journal of the American Statistical Associate*, Vol.84 1989, pp.1003-1013.

[34] Tsai, L. T., & Yang, C. C., A simulation study on classified accuracy for learning vector quantization (in chinese). *Journal of Research on Measurement & Statistics*, Vol.12 2004, pp. 269-291.

[35] Ender, C. K. & Bandalos, D. L., The relative performance of full information maximum likelihood estimation for missing data in

structural equation models. *Structural Equation Modeling*, Vol.8 2001, pp. 430-457.

[36] Ender, C. K. & Peugh, J. L., Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling*, Vol.11 2004, pp. 1-19.

[37] Ender, C. K., A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, Vol.8 2001, pp. 128-141.

[38] Allison, P. D., Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, Vol.112 2003, pp. 545-557.

[39] Muthén, L. K., & Muthén, B. O.,. *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén. 1998-2006.

[40] Yang, C.C., Tsai, L.T., & Yang, C.C., Estimating missing sampling weights by using LVQ-ESW method in a CFA model (in chinese). *Chinese Journal of Psychology*, Vol.51 2009, pp.277-293.