© *The Journal of Risk and Insurance*, 2008, Vol. 75, No. 3, 617-642

Integrating Neural Networks for Risk-Adjustment Models

Shuofen Hsu
Chaohsin Lin
Yaling Yang

Abstract

This article demonstrates the possibility of an alternative approach for risk-adjustment models. In the proposed model the risk characteristics of the beneficiary's health within the same cohort classified by Self-Organizing Map network are highly homogeneous, whereas the numbers of individuals within each cohort remain sufficient to allow further investigation of the causal effect from clustered data. A comparison of different models by the 10-fold cross-validation reveals that the performance improvement in the proposed integration model is both significant and stable across the estimation and validation sampling.

Introduction

In 1995, Taiwan introduced a mandatory National Health Insurance (NHI) scheme. All Taiwanese residents are eligible for the NHI coverage, and all are legally required to participate in the NHI program. Since the NHI was implemented, medical expenditures have grown at a rate of 8.5 percent annually. This corresponds with the natural population growth, the aging demographic structure, and the advances in medical technology, but it exceeds the wage-tied annual premium growth rate of 3 percent (Bureau of NHI, 2003). To effectively monitor the increase in medical expenditures and prevent a continued financial deterioration of the health insurance programs, various measures such as co-payment and adjustment of reimbursement

Shuofen Hsu and Chaohsin Lin are professor and assistant professor, respectively, at the Department of Risk Management and Insurance, National Kaohsiung First University of Science and Technology. 2, Juoyue Rd., Nantz District, Kaohsiung 811, Taiwan. Yaling Yang is associate professor, Department of Aviation and Maritime Management, Chang Jung Christian University. 396 Chang Jung Rd., Sec.1, Kway Jen, Tainan 711, Taiwan. The authors can be contacted via e-mail: shuofen@ccms.nkfust.edu.tw, linchao@ccms.nkfust.edu.tw., and yly@mail.cjcu.edu.tw, respectively. This study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National Health Insurance, Department of Health, and managed by National Health Research Institutes. The interpretation and conclusions contained herein do not represent those of the Bureau of National Health Insurance, Department of Health, or National Health Research Institutes.

standards for drugs and hospitalization have been adopted. Besides controlling the demand side, the Bureau of NHI has introduced a system of prospective payments to create incentives for efficiency among health-care providers and to move toward a multi-carrier health insurance system involving competition among health carriers.

The prospective payment is a system of predetermined fees that the Bureau of NHI uses to reimburse hospitals for inpatient and outpatient services, as well as skilled nursing facilities, rehabilitation hospitals, and home health services (Bureau of NHI, 2003). Given competing health plans and premium regulation, when the system of prospective payments fails to reflect the health risks of the beneficiaries, cream skimming (preferred risk selection or cherry picking)[1] may occur because health plans may seek out only those patients for whom profits are expected to be high. Risk adjustment is the most effective strategy for reducing cream skimming. It refers to the use of information to determine the expected health expenditures of consumers over a fixed time interval (e.g., a month, quarter, or year) and then use that information to set subsidies to consumers or health plans to improve efficiency and equity (van de Ven and Ellis, 2000). The subsidies are the differences between the actual health expenditure an individual enrollee incurred and the premium payment contributed by that same enrollee.

Risk-adjustment model selection has received considerable attention from policymakers in different countries (Barros, 2003; van de Ven and Ellis, 2000). A crucial issue is whether the medical expenditures calculated by the risk-adjustment model reflects the health risks of the insured population. Most econometric work in health economics has focused on the problem of devising an appropriate stochastic model to fit the available data. Estimation of regression functions generally assumes that the regression function is linear and that the random error term is normally distributed. However, these assumptions frequently cannot be satisfied due to the unusual distributional properties of the medical expenditure data. In other words, there is an extreme skewness with a small proportion of the people accounting for a large proportion of expenditures, and a substantial proportion of people with no expenditures in a year (Pope et al., 1998).

Some of the solutions examined in the literature rely on logarithmic transformation to deal with heavily skewed dependent variables, such as household medical expenditures, or the decomposition of beneficiary responses into a series of estimation models to deal with specific parts of the distribution, for example, two-part or multipart models, as outlined in Duan et al. (1983). Both nonlinear transformations and multi-part models are used to improve the consistency of the ordinary least squares (OLS) model in situations with high heteroskedastic errors (van de Ven and Ellis, 2000). However, nonlinear transformations create the problem of retransforming to the original scale (e.g., dollars rather than log-dollars), to make relevant inferences for policy. Mullahy (1998) contended that, owing to nonlinearities and retransformations, the estimated parameters are insufficient for making inferences regarding important

[1] Cream skimming means selection by providers (or entities responsible for health-care provision) of those consumers expected to be profitable, given the system of risk-adjusted capitation payments (Barros, 2003).

policy parameters involving the level of medical expenditure. Furthermore, multi-part models suffer the disadvantages of being computationally burdensome and difficult to interpret (Pope et al., 1998).

This article demonstrates the possibility of an alternative approach for calculating individual medical expenditures. In particular, two neural network models, that is, Self-Organizing Maps (SOM, Kohonen, 1982, 1989, 1990) and Back Propagation Network (BPN, Rumelhart, Hinton, and Williams, 1986), are integrated to establish a risk-adjustment model. We first apply the SOM network for classifying the sample data and then employ the BPN for predicting the annual medical expenses of the beneficiaries. In this proposed model the risk characteristics of the beneficiary's health within the same cohort classified by SOM are highly homogeneous, whereas the numbers of individuals within each cohort remain sufficient to allow further investigation of the causal effect from clustered data. More specifically, the estimations of individual medical expenditures made using BPN following SOM classification will more closely approach the actual spending incurred and will thus alleviate the potential preferred risk selection by health-care providers.

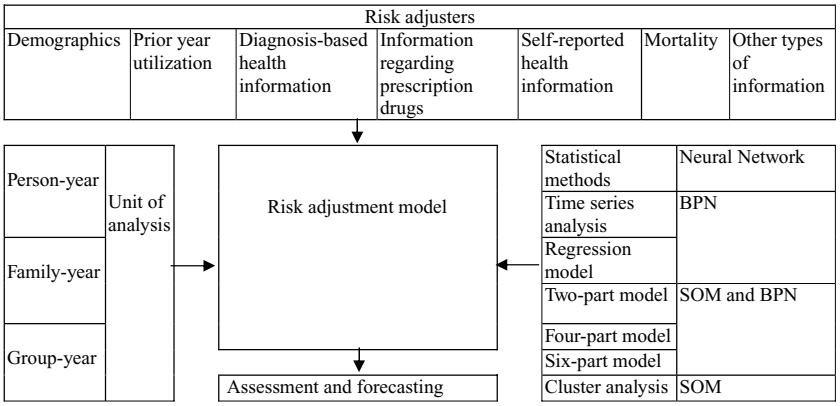
The remainder of this article is organized as follows. The next section provides a brief review of the literature concerning risk-adjustment models as well as the application of SOM and BPN. The section "Data, Variables, and Preliminary Analysis" describes the data used in this article and discusses associated preliminary analyses including the results of SOM risk classification and *K*-means clustering analysis. In the section "Benchmark Models" the estimated results of linear, log-linear, and two-part models are presented as benchmarks for comparison with the neural networks. The network design of the BPN is discussed in the section "Neural Network Models" and the predictive power of the alternative risk-adjustment models are compared in the section "Predictive Performance Assessment of Alternative Risk Adjustment Models." Finally, in the "Summary and Conclusions" section we draw our conclusions and provide some suggestion for future research.

LITERATURE REVIEW

Risk-adjustment models contain the following three main elements: adjuster selection, unit of analysis (which is linked to whether the data come from the individual or the group level), and functional form of the estimation model (Hsu, Lin, and Yang, 2006). Different combinations of these aspects yield alternative risk-adjustment models such as adjusted average per capita cost (AAPCC), ambulatory care group (ACG), diagnostic care groups (DCG), major diagnostic category (MDC), and the Robinson–Luft multi-equation model. Figure 1 shows the relationship among the three main elements.

As indicated in Figure 1, risk adjusters can be summarized into seven types based on the kind of data used for the prediction according to van de Ven and Ellis (2000). The unit of analysis of medical spending can be person-year, family-year, or group-year. The functional forms for estimating models include the time-series model, regression model, multi-part models, and the clustering models. The simple linear model is the most widespread method in the literature and has the advantage of not suffering from any retransformation problems. Another reason for the simple linear model being extremely suitable for practical use is that it stays as close as possible to the

FIGURE 1
Alternative Risk-Adjustment Models



cell-based approach, the calculation of the average expenditure per risk group, which is mainly used by governmental regulators for risk adjustment and by health insurers for premium rating. The log-linear model uses nonlinear transformations of the dependent variable to generate unbiased estimates. However, as Mullahy (1998) observed, it is important that the error structure strictly satisfies the homoskedastic error assumption, otherwise a nonlinear smearing correction can produce seriously biased estimates. Most multi-part models are applications of the two-part model. They are usually estimated using a logit or probit model to determine the probability of a positive value being observed for the dependent variable, that is, medical expenditures. This is combined with an OLS being conducted on the subsample of positive observations in which sample subdivision can further be carried out, based on the specific upper limits on medical spending. Another approach different from the multi-part model for dealing with sample division is the clustering analysis. In it the data with similar characteristics are clustered, and estimation is performed according to each cluster, respectively.

However, similar to the limitations imposed on the estimation of regression models, it is well recognized that the two most important problems in cluster analysis are the assumption of normality in the underlying distributions, and the difficulty in identifying an appropriate function for the distributions (Back, Sere, and Vanharanta, 1998). In addition, in cluster analysis one of the groups may have just one or very few vectors whereas another may have 99 percent of the vectors. Thus, further investigation of the causal effect from clustered data is impossible in cases where the clustered data sample is small (Brockett, Xia, and Derrig, 1998). Moreover, with few data in the cluster, health insurance carriers would not be able to pool their underwriting risk for some type of peril using the Law of Large Numbers. Another limitation of conventional cluster analysis is that identifying the groups based on the nature of the observations for each group is rather difficult, including observations such as which group should be considered a "catastrophic risk" and which group should only be considered a "low risk" for a particular case (Brockett, Xia, and Derrig, 1998). Furthermore, the

analytical results are difficult to visualize in situations involving several explanatory variables (Back, Sere, and Vanharanta, 1998; Vermuelen, Spronk, and Der Wijst, 1994).

As will be demonstrated later on in this article, neural networks cannot only be used for the functional form of risk-adjustment models to solve the estimation bias caused by the highly skewed medical spending (BPN) but they can also deal with the above main problems encountered with cluster analysis (SOM). In addition, the clustering results of SOM can be visualized in a comprehensive way. As a feed-forward neural network, SOM uses not only an unsupervised training algorithm, but also a process called self-organization to configure output units into a topological representation of the original data. SOM belongs to a general class of neural network methods, that is, nonlinear regression approaches that can be trained to learn or find relationships between inputs and outputs, or to organize data to identify unknown patterns or structures. The SOM network model has been applied in over 5,384 applications in numerous different areas (Oja, Kaski, and Kohonen, 2003). Kaski, Sinkkonen, and Peltonen (2001) and Charalambous, Charitou, and Kaourou (2000) used SOM to predict bankruptcies, and Lewis, Ware, and Jenkins (1997) used it for property valuation. Moreover, Serrano-Cinca (1996, 1997) used SOM for financial diagnosis and the classification of financial information, respectively. SOM has also proven to be a valuable tool for data mining and knowledge discovery and has applications in financial data analysis (Länsiluoto et al., 2004).

In contrast to the unsupervised learning method of SOM, BPN uses a supervised learning technique for training neural networks. A number of finance-related studies have implemented BPN in the area of bankruptcy predictions (Brockett et al., 1994; Huang, Dorsey, and Boose, 1994; Tam and Kiang, 1992), mortgages (Grudintski, Do, and Shilling, 1995), property valuation (Do and Grundintski, 1992; Worzala, Lenk, and Silva, 1995), investment analysis (George and Yang, 1992), and stock and futures price prediction (Grudnitski and Osburn, 1993; Mirmirani and Li, 2004; Narain and Narain, 2002). Among these literatures, however, few have integrated SOM and BPN in the applications for solving financial or economic problems. Brockett, Xia, and Derrig (1998) applied SOM to classify automobile injury claims, and subsequently used BPN to examine the validity of the SOM classification approach. This article diverges from Brockett, Xia, and Derrig in that we first apply SOM for classifying the sample data to ensure the similar risk characteristics within the same cohort and then employ BPN for predicting the annual individual medical expense rather than for validating the results of SOM.

DATA, VARIABLES, AND PRELIMINARY ANALYSIS

Data

Data were taken from the National Health Research Institute (NHRI) in Taiwan for the period January 1, 1999 to December 31, 2001. To avoid any potential geographical bias created by the six regional branches of the NHI, we only focused on one of the six branches when constructing the risk-adjusted capitation formula since the same approach could be applied to the remaining five branches. In addition, since Taiwan has a population of about 23 million, even a 0.1 percent random sample from each of the six branches will undoubtedly increase the complexity of the data processing. At the same time, the potential validity margin improvement from the

larger sample set is most likely minimal. Therefore, the data used in this article consist of a random sample of 5,557 individuals, which constitutes about 0.1 percent of the specific branch population being examined. Apart from random sampling, we also totaled the individual's medical expenses according to their ID and date of birth in order to obtain the total medical expense of each beneficiary during each year. This was done because the records of a beneficiary within a specific branch may be scattered among different contracted medical care institutions and may be located outside the specific branch being examined.

Variables

Among the above seven types of risk adjusters mentioned in the literature, age and sex are the two most widely used. In addition to these two demographic risk adjusters, this article also employs prior year expenditures, including frequency of inpatient visits, total outpatient and inpatient medical expenses, in the risk-adjusted capitation formula. Whether a patient has a major illness identity is used as a proxy variable for diagnosis-based information. In addition, the six beneficiary categories that are based on different occupations of the insured are used as proxy variables to indicate possible income effects on health-care expenditure. Table 1 gives the descriptive statistics for the sample data. The 75th percentile of the total medical expenditure was less than the average total expenditure in the same year. This implies a positively skewed distribution of medical expenses as indicated in the skewness coefficient.

SOM Risk Classification

SOM Architecture. A typical Kononen's SOM with a two-layer network is employed. The input layer contains the input training vector X that comprises the vector of risk adjusters discussed above and is the same as the explanatory variables of the regression and BPN models. The output layer consists of the resulting network output, and is expressed as O. The SOM network architecture is shown in Figure 2.

Each neuron i of the SOM is represented by a seven-dimensional weight vector denoted as $W_i = [w_{i1},\ldots, w_{i7}]$, since there are seven input vectors in this article. The neurons are connected to adjacent neurons by a neighborhood relationship that dictates the topology or structure of the map. Typically, a rectangular or hexagonal neighborhood is used. We choose a network topology that was hexagonal with 20*20 neurons.

SOM training is accomplished by presenting one input pattern X at a time in random sequence, and then comparing, in parallel, this pattern with all the reference vectors. The best match unit (BMU), which can be calculated using the Euclidean metric, represents the weight vector with the greatest similarity with that input pattern.

Denoting the winner neuron by O^*, the BMU can be formally defined as the neuron for which:

$$\| X - W_{o^*} \| = \min_i \{ \| X - W_i \| \}, \tag{1}$$

where $\| \cdot \|$ denotes the Euclidean distance measure. Here, X is a vector of the risk adjusters that includes two demographics (x_d, $d = 1, 2$ represents age, and gender, respectively), three factors of prior medical utilization (x_p, $p = 3, 4, 5$ represents

TABLE 1
Descriptive Statistics of the Sample Data

	Explanatory Variables												Dependent Variable	
	Age		Ratio of the Male		Outpatient Expenses (NT$/Person/Year)		Inpatient Visits		Inpatient Expenses (NT$/Person/Year)		Ratio of Major Illness/Injury		Individual's Medical Expenses (NT$/Year)	
Year	1999	2000	1999	2000	1999	2000	1999	2000	1999	2000	1999	2000	2000	2001
Mean	33.68	34.68	0.507	0.507	9,674	8,252	0.12	0.12	4,596	4,908	0.014	0.014	14,991	13,496
Standard deviation	20.10	20.10	–	–	26,974	28,407	0.54	0.55	39,522	39,289	–	–	51,834	53,678
Kurtosis	−0.53	−0.53	–	–	468	337	145	134	505	310	–	–	155	238
Skewness	0.42	0.42	–	–	18.81	16.45	9.64	9.45	19.58	15.21	–	–	11.06	13.08
Minimum	1	2	–	–	0	0	0	0	0	0	–	–	0	0
25th percentile	18	19	–	–	1,523	833	0	0	0	0	–	–	1,630	896
50th percentile	32	33	–	–	4,610	2,995	0	0	0	0	–	–	5,000	3,096
75th percentile	47	48	–	–	10,554	7,663	0	0	0	0	–	–	11,752	8,544
95th percentile	71	72	–	–	33,027	29,788	1	2	15,400	15,880	–	–	46,418	47,844
99th percentile	81	82	–	–	66,614	66,086	2	2	102,504	104,203	–	–	198,249	198,733
Maximum	98	99	–	–	853,069	731,892	12	12	1,419,307	1,174,594	–	–	1,215,931	1,534,576

FIGURE 2
SOM Network Architecture

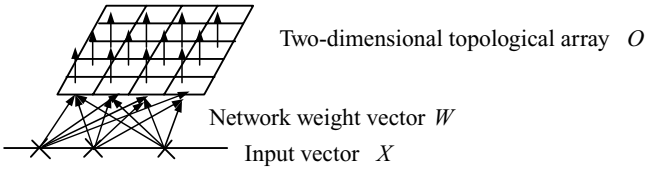
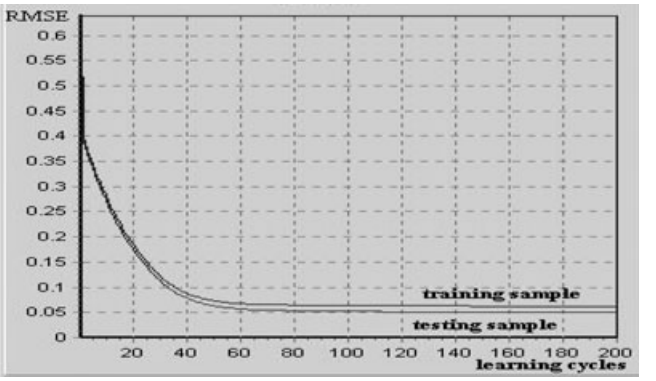


FIGURE 3
SOM Network Learning Convergence Diagram for the Estimation Sample



outpatient expenses, inpatient visits, and inpatient expenses, respectively), and dummy variables x_6 and x_7 that represent beneficiaries with major illness identities if $x_6 = 1$ otherwise $x_6 = 0$ and for six beneficiary occupation categories.

The SOM update rule for the weight vector is as follows:

$$W_i(t+1) = W_i(t) + h_{O^*i}(t)[X(t) - W_i(t)], \tag{2}$$

where t is the number of iterations, $X(t)$ is the risk adjusters randomly drawn from the risk adjusters set at time t, and $h_{o^*i}(t)$ is the neighborhood kernel around the winner unit O^*at t. This last term is a nonincreasing function of time and of the distance of unit i from BMU and is usually formed by two components: the learning rate function $\alpha\ (t)$ and the neighborhood function $h(d, t)$ as follows:

$$h_{o^*i}(t) = \alpha(t)h(\|r_{O^*} - r_i\|, t), \tag{3}$$

where r_i denotes the location of unit i on the map grid.

The stopping criterion for a training iteration of SOM is the root mean squared error (RMSE) of the Euclidean distances between each input vector and its BMU in the SOM. Figure 3 shows that after 200 epochs of learning, the RMSE of training sample and testing sample converge to 0.069 and 0.053, respectively.

FIGURE 4
U-Matrix for the Estimation Sample

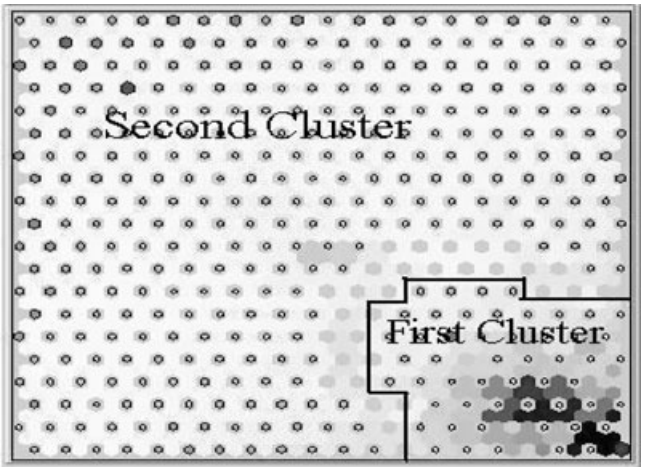


FIGURE 5
Network Topology for the Estimation Sample

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	21	15	12	26	33	41	32	44	27	39	13	30	34	20	28	29	33	35	5	11
1	18	55	42	35	27	42	22	25	18	13	21	15	20	14	13	12	12	9	13	16
2	44	29	43	18	15	23	24	29	19	12	4	21	18	22	13	11	14	11	15	20
3	22	31	23	65	20	12	21	19	10	16	13	16	12	14	20	21	11	17	16	20
4	25	30	26	10	14	5	8	9	8	10	10	13	11	8	13	19	13	12	18	20
5	34	31	17	21	15	5	11	10	7	5	6	7	11	9	13	16	11	16	25	22
6	34	17	23	15	12	11	6	13	8	7	8	10	10	12	12	13	15	17	18	31
7	32	24	20	11	15	10	6	6	17	5	6	9	6	8	7	12	18	12	16	17
8	22	17	18	19	16	11	8	8	6	9	2	4	4	5	11	12	14	11	19	12
9	40	19	18	19	16	11	8	8	6	9	2	4	4	5	11	12	14	11	19	12
10	28	31	26	8	12	10	6	7	2	4	4	8	0	0	0	0	0	14	17	19
11	24	23	17	12	15	13	11	8	4	4	6	0	0	0	0	0	0	0	8	8
12	19	27	18	11	15	8	7	7	3	4	0	0	0	13	20	19	17	0	0	0
13	33	13	16	15	12	7	8	11	9	9	0	0	12	10	6	16	9	10	14	5
14	30	14	15	14	14	19	14	15	5	10	0	0	16	5	6	1	3	4	4	1
15	21	21	11	16	20	10	11	11	12	9	0	0	15	12	6	1	6	4	4	3
16	20	15	16	15	17	15	19	9	12	11	0	0	16	10	3	1	5	5	3	0
17	26	11	7	14	13	12	19	20	20	17	14	0	0	4	6	4	2	8	5	3
18	5	12	14	15	13	26	27	14	17	23	14	10	0	15	4	4	0	6	11	4
19	14	19	10	15	17	29	28	24	20	12	16	5	0	4	2	1	0	6	2	83

Significance of Difference Among Clusters. The structured data of the estimation sample after 200 epochs of learning are visualized using the U-matrix method (Ultsch and Siemon, 1990) as shown in Figure 4. It is evident that the light color contains a large uniform area in which beneficiaries correspond to homogeneous characteristics. At the same time the lower right-hand corner comprises a clearly separated area in which beneficiaries possess different characteristics from the preceding cluster. The network topology shown in Figure 5 shows exactly how many beneficiaries are clustered within each neuron. For example, the grid of (19, 19) in Figure 5 is 83, which indicates this neuron contains 83 beneficiaries.

TABLE 2
Test of Between-Groups Effect

Variable	Variance Source	Sum of Squares	Degree of Freedom	Mean of Square	F	p-Value
Age	Between	33,900	1	33,900	85	$\cong 0$
	Within	2,210,999	5,555	398		
	Total	2,244,900	5,556			
Inpatient expenses	Between	1.16E + 12	1	1.16E + 12	860	$\cong 0$
	Within	7.51E + 12	5,555	1.35E + 09		
	Total	8.68E + 12	5,556			
Inpatient visits	Between	752	1	751.7104	4,858	$\cong 0$
	Within	859	5,555	0.154724		
	Total	1,611	5,556			
Outpatient expenses	Between	2.86E + 11	1	2.86E + 11	422	$\cong 0$
	Within	3.76E + 12	5,555	6.76E + 08		
	Total	4.04E + 12	5,556			
Whether patient has major illness identity	Between	29	1	29	1,678	$\cong 0$
	Within	95	5,555	0.017158		
	Total	124	5,556			

To examine whether there are significant differences between the risk groups clustered by SOM, this article conducts a one-way analysis of variance (one-way ANOVA) to confirm the following null hypothesis H_0:

H_0: *The two risk groups do not differ significantly*.

Table 2 lists the result of the hypothesis test. The null hypotheses is not supported, demonstrating that age, outpatient and inpatient expenses, inpatient visits, and whether or not the patient has a major illness identity display a strong significant difference (all p-values < 0.01). Two risk cohorts clustered by SOM are thus confirmed by one-way ANOVA.

Apart from the estimation sample, the validation sample is also clustered by SOM in the same way in order to validate the forecast. This is done because risk characteristics may change with time causing subgroup numbers to change, or result in different subgroup numbers of the estimation and validation sample. If the SOM subgroup numbers of the estimation and validation sample are inconsistent, which implies a change over time in risk classification, then the individual data cannot be simultaneously validated by SOM classification and BPN prediction. Under this circumstance, the data in the estimation and validation sample can be combined into the SOM instead of performing the SOM twice, as is done in the current approach, in order to obtain a consolidated pattern of changing risk classification. This is known as the "dynamic SOM" and will be discussed later (Kiviluoto and Bergius, 1998; Kasslin, Kangas, and Simula, 1992; Tryba, Metzen, and Goser, 1989).

Results of the SOM clustering for the estimation and validation samples indicate that cluster patterns are stable and exhibit no changes within two data divisions. It

TABLE 3
Descriptive Statistics of Sample Data After SOM Classification

Year	1999		2000	
Cluster	1	2	1	2
Observations	509	5,048	479	5,078
Average age	41.45	32.89	45.90	33.62
Ratio of male	0.503	0.508	0.499	0.508
The number of major illness/injury	128	–	128	–
Average outpatient expenses (NT$/person/year)	32,247	7,396	35,111	5,718
Average inpatient visits (NT$/person/year)	1.28	–	1.37	–
Average inpatient expenses (NT$/person/year)	50,166	–	56,929	–
Predicted-year average total annual expenses (NT$/person/year)	61,797	10,271	53,139	9,757

is worth noting that although clustering patterns (numbers of clusters) may remain unchanged, beneficiaries within each cluster may differ, as discussed later. The risk characteristics of a specific beneficiary within one cluster tend change over time, whereas risk-type patterns within the entire population tend to remain unchanged.

Characteristics of SOM Clusters. The descriptive statistics and characteristics of each SOM cluster in the estimation and validation sample are summarized in Table 3. The first cluster for the estimation sample contains 509 beneficiaries. Individuals in this cluster have a relatively high inpatient visit frequency as well as inpatient and outpatient medical expenses. Average numbers of inpatient visits, inpatient and outpatient expenses are 1.28 times, NT$50,166 and NT$32,247, respectively. Some of the beneficiaries in this cluster have major illness/injury identities. Moreover, the predicted-year average medical expenses of these individuals are NT$61,797. We define this cluster as the high-risk cohort. More specifically, the first risk group is characterized by individuals who have hospitalization expenses, or have major illness/injury identities. The other estimation sample cluster contains 5,048 beneficiaries who either only use outpatient medical resources or use no medical resources at all. The average outpatient expenses of this cluster were NT$7,396 in 1999 and their total average medical expenses were NT$10,271 in 2000. We define this cluster as the low-risk cohort. The same characteristics of the high- and low-risk cohorts can also be found in the validation sample.

It should be noted that there is a decrease of 30 beneficiaries in the high-risk cohort of the validation sample. This decrease can be attributed to the changes in individual risk during 1999 and 2000. Specifically, only 168 beneficiaries remained in the high-risk cohort during these 2 years. Among the high-risk cohort during 1999, 341 beneficiaries were assigned to the low-risk cohort in 2000, whereas 311 individuals who had been in the low-risk cohort in 1999 were assigned to the high-risk cohort in 2000. As mentioned before, even though there are changes in the individual's risk characteristics, the clustering numbers remain the same for the two cohorts during 1999 and 2000.

TABLE 4
Descriptive Statistics of *K*-Means Classification

Cluster	1	2
Observations	12	5,545
Average age	42.5	33.7
Ratio of male	0.667	0.507
The number of major illness/injury	8	120
Average outpatient expenses (NT$/person/year)	110,806	9,453
Average inpatient visits (NT$/person/year)	3.83	0.109
Average inpatient expenses (NT$/person/year)	696,343	3,098
Predicted-year average total annual expenses (NT$/person/year)	620,627	13,680

Clustering Analysis by *K*-Means

We employed *K*-means to investigate if the cluster analysis suffers from normality distribution assumption and function identifying difficulties as mentioned earlier. The *K*-means clustering algorithm can be presented as follows:

$$\min \sum_{j=1}^{k} (X_j - C_j)'(X_j - C_j), \tag{4}$$

where k denotes the number of clusters that is randomly chosen in advance, and C denotes the vector of the jth cluster center. To allow direct comparison with SOM clustering, we define $k = 2$. The results of *K*-means clustering are shown in Table 4. One of the *K*-means clusters contains 12 beneficiaries, among which 8 individuals have major illness identities with average medical expenses of NT$620,627. It is obvious that the health-care expenditures of this cohort are much higher than those of the first/high-risk cohort clustered by SOM. The average inpatient expenses of this *K*-means cohort are 10 times more than that of the SOM high-risk cohort. The results of *K*-means show that further investigation by regression or BPN analysis is impossible in that one of the *K*-means clusters constitutes only 0.2 percent of the sample population, and the biased situation is even worse than mentioned by Brockett, Xia, and Derrig (1998). The problems associated with cluster analysis as discussed before do exist in the current data set of this article. Therefore, we solely rely on the classification results of SOM to proceed on individual health-care forecasting.

BENCHMARK MODELS

Linear, log-linear, and two-part models are described as Equations (5)–(7).

$$Y = \beta' X + \varepsilon \tag{5}$$

$$\ln(Y + 1) = \beta' X + \varepsilon \tag{6}$$

$$E(Y \,|\, X) = P(Y > 0 \,|\, X) E(Y \,|\, Y > 0, X)\,, \tag{7}$$

where Y denotes the beneficiary medical expenditure. $X = \{x_1, x_2, \ldots, x_7\}$ is a vector of the risk adjusters and comprise seven variables as defined before. Here, β' is the transpose vector of parameters estimated by the regression model and ε denotes the error term and is assumed to be of independent identical distribution.

The two-part model is estimated by a logit for $P(Y > 0|X)$, and least squares on Y. We follow Mullahy (1998) for considering two alternative estimators to make a full correction for heteroskedasticity in error terms. First, given that $E(Y|Y > 0, X)$ must be positive, an exponential conditional mean specification $E(Y|Y > 0, X) = \exp(X\beta_2)$ is used. Combining this with a logistic specification for $P(Y > 0|X)$, the model gives

$$\begin{aligned} E(Y \mid X) &= P(Y > 0 \mid X) E(Y \mid Y > 0, X) \\ &= [\exp(X\beta_1)/(1 + \exp(X\beta_1)] \exp(X\beta_2) \\ &= \exp(X(\beta_1 + \beta_2))/(1 + \exp(X\beta_1)). \end{aligned} \tag{8}$$

The model can be estimated by a two-step estimator, using logit for β_1 and nonlinear least squares for the positive observations. Alternatively, it can be estimated in one step, using the full sample to estimate the above equation by nonlinear least squares. The advantages of Mullahy's (1998) specification are that it is straightforward to use instrumental variables for dealing with problems of unobservable heterogeneity in the model, and that the elasticities, $\partial E(Y)/\partial X$, are simple to compute and interpret. The price of using this simpler specification is that it does not allow separate inferences for $P(Y > 0|X)$ and $E(Y|Y > 0, X)$.

The estimation results are shown in Tables 5 and 6. Compared to most empirical literature in which demographics of age and sex are most widely used yielding adjusted R^2 values ranging from at most 0.59 to 0.001 (van de Ven and Ellis, 2000), the overall results of these benchmark models suggest a rather satisfactory explanatory power of the chosen risk adjusters in explaining the variation of individual medical expenses. The adjusted R^2 values of the linear regression and the two-part model are 0.38 and 0.39, respectively, whereas the value of the log-linear model is 0.084, which indicates a less convincing explanatory power in model specification. Therefore, from here on we will focus on the discussion of linear and two-part models.

Except for the parameters of gender and the dummy variables for beneficiary occupation categories, the rest are of statistical significance in both linear and two-part models. In particular, age, outpatient and inpatient expenses, together with inpatient visits have a positive influence in the following year's medical expenses. This indicates that the higher the spending of medical resources is in 1999, the higher the expected average medical expense will be in 2000. Average medical expenses of the insured with a major illness identity in the next year are more than that without a major illness identity.

Most empirical researches have confirmed the empirical results of Newhouse (1977) concerning the income elasticity of health expenditure and the high explanatory power of the relationship. However, the statistical significances of the six beneficiary occupation categories, which are the proxy variables for income status, are diversified between these two models. The same result as that of Newhouse (1977) can be

TABLE 5
Estimating Results of Linear and Log-Linear Regression Models

	Linear			Log-Linear		
Risk Adjuster	Coefficient	SE	*t*-Test	Coefficient	SE	*t*-Test
Age	209	28.26	7.40***	−0.7283	0.002	−9.78***
Gender						
Male	465	1,097.07	0.42	0.0097	0.074	5.048***
Female (reference group)						
Outpatient expenses	0.79	0.02	35.64***	2.1537E-05	1.51E-06	14.25***
Inpatient expenses	0.13	0.02	7.13***	−2.4395E-06	1.24E-06	−1.97**
Inpatient visits	136,059	1,338.19	10.17***	0.4330	0.0909	4.77***
Whether patient has major illness identity	56,055	4,076.69	13.75***	0.3649	0.2768	1.32
Insurance category						
Insurance category 1	−4,400	1,699.97	−2.59***	0.7240	0.1154	6.27***
Insurance category 2	−4,092	1,992.46	−2.05**	0.6782	0.1353	5.01***
Insurance category 3	−3,387	2,072.57	−1.63	0.7752	0.1407	5.51***
Insurance category 4	−7,659	8,831.29	−0.87	0.9986	0.5996	1.67*
Insurance category 5	−5,707	7,604.61	−0.75	−0.0899	0.5163	−0.17
Insurance category 6 (reference group)						
Intercept	213	1,926.80	0.11	6.9604	0.1308	53.20***
Adjusted R^2	0.38		0.084			
Observations	5,557					

Note: Response variable: annual medical expenses in 2000 (NT$/person/year).
*, **, *** indicates statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

found in the linear regression in that the next year's average medical expenses of the beneficiaries in categories 1 and 2 are significantly lower than those of category 6. Since individuals who belong to categories 1 and 2 are most often civil servants or salaried employees, the average income is relatively higher and more stable compared to that of category 6 that contains mostly veterans or their survivors. Nonetheless, no statistical significance was found in these six occupation categories in the two-part model.

NEURAL NETWORK MODELS

BPN Architecture

The BPN is constructed using the same input variables as in the benchmark models in order to allow a direct comparison between them. The network architecture used for the present medical expenditure forecasting problem contains two hidden layers, as shown in Figure 6.

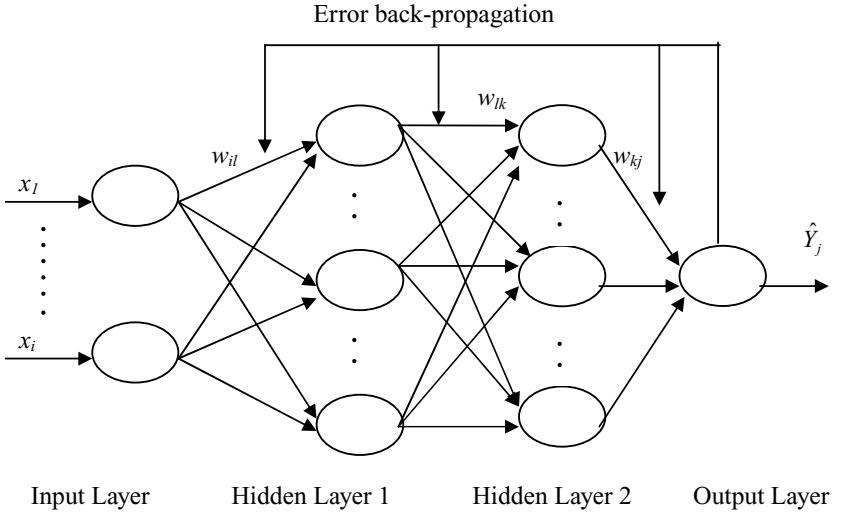
Apart from the number of input variables and hidden layers, there are many other parameters that are not known, such as the number of hidden nodes, the type of activation functions in the hidden and output layers, the value of the learning and the momentum rate, and the amount of training. The main problem with BPN is that

TABLE 6
Estimating Results of Two-Part Model

		Logit Parameter Estimates				OLS Parameter Estimates			
Risk Adjuster		Coefficient	SE	ChiSquare	Prob > ChiSq	Coefficient	SE	*t*-Test	Prob > |*t*|
Age	Biased	0.11	510.42	0.00	0.9998	217.53	30.18	7.21***	< 0.0001
Gender									
Male	Biased	1.69	13,428.59	0.00	0.9999	−196.00	592.94	−0.33***	< 0.0001
Female (reference group)									
Outpatient expenses	Biased	0.27	78.692	0.00	0.9973	0.81	0.02	34.99***	< 0.0001
Inpatient expenses	Biased	0.001	3.01	0.00	0.9998	0.14	0.02	7.20***	< 0.0001
Inpatient visits	Biased	−67.76	313,659.70	0.00	0.9998	13,821.18	1,376.02	10.04***	< 0.0001
Whether patient has major illness identity	Unstable	−125.64	412,297.08	0.00	0.9998	−24,209.86	2,113.32	−11.4***	< 0.0001
Insurance category									
Insurance category 1	Biased	−36.49	28,183.14	0.00	0.9990	−359.49	2,231.36	−0.16	0.8720
Insurance category 2	Biased	−46.76	21,274.81	0.00	0.9982	−69.14	2,409.48	−0.03	0.9771
Insurance category 3	Biased	−42.98	31,077.13	0.00	0.9989	335.80	2,468.60	0.14	0.8918
Insurance category 4	Zeroed	88.58	0	99,999***	0.0000	−3,909.15	7,932.96	−0.49	0.6222
Insurance category 5	Zeroed	73.07	0	99,999***	0.0000	70.92	7,042.61	0.01	0.9920
Insurance category 6 (reference group)									
Intercept	Biased	193.31	414,160.07	0.00	0.9996	20,356.98	3,158.63	6.44***	< 0.0001
Adjusted R^2	–						0.388066		
Observations			5,557				5,000		

*** indicates statistical significance at the 0.01 level.

FIGURE 6
BPN Model



there are no established rules to help with choosing the appropriate values of these parameters, and it is necessary to resort to trial and error to obtain their appropriate values (Binner et al., 2005). However, as Hoptroff (1993) suggested, 10 nodes in a hidden layer are usually sufficient for most forecasting problems although more nodes can be used but usually result in slower learning without an improvement in result. We considered 100 possibilities of networks with the current two hidden layers each containing 1 to 10 possible hidden nodes. It is worth noting that according to the results of SOM clustering, the number of hidden nodes associated with each BPN model whose data set obtained from SOM clustering may be different. The number of hidden nodes associated with these two BPNs following the two SOM clusters should be determined separately since each cluster may possess different risk characteristics. As a result, the functional relationships between input and output layers described by the hidden nodes may not be the same. We follow the usual way in determining the remaining parameters; that is, the learning and momentum rates are 0.01 and 0.5, respectively, and focus only on the discussion of activation function mainly because the SOM clustering before the BPN prediction may affect the BPN learning process that is heavily influenced by the type of activation.

The stopping criterion for a training iteration of BPN is the RMSE of the sum of the difference between the actual and the forecasted individual medical expenditure. We chose 1,000 times of training since the network always converges to stable when the amount of training reaches 1,000. In addition, across-channel normalization is used to rescale the data in the range [0, 1] to obtain the stability of the neural networks.

Activation Function—Advantages of Integrating SOM and BPN

The goal of the BPN learning process is to determine a set of weights through the activation functions in such a way that the desired individual medical expenditure produced by the network will be as close as possible to the actual ones. The gradient

descent method is the most commonly used method for calculating the necessary adjustments of the connection weights for minimizing the error term as follows:

$$E = \frac{1}{2}\sum(Y - \hat{Y})^2 = \frac{1}{2}\sum\left(Y - f\left(\sum_k w_{kj} h\left(\sum_l w_{lk} g\left(\sum_i w_{ik} x_i\right)\right)\right)\right)^2, \quad (9)$$

where g, h, and f denote the activation functions used in the first, the second hidden layer, and the output layer and are connected by the connection weights of w_{ik}, w_{lk}, and w_{kj}, respectively.

The formula to adjust the connection weights in the activation function is as follows:

$$w(t+1) = w(t) + \Delta w(t+1), \quad (10)$$

where t is the number of iterations.

The above formula can further be rewritten according to the definition of the gradient descent method as follows:

$$w(t+1) = w(t) + \Delta w(t+1) = w(t) - \eta \frac{\partial E(f(w(t))}{\partial w(t)}, \quad (11)$$

where η is the learning rate, and $-\eta \frac{\partial E(f(w(t))}{\partial w(t)}$ is the adjustment to the connection weights.

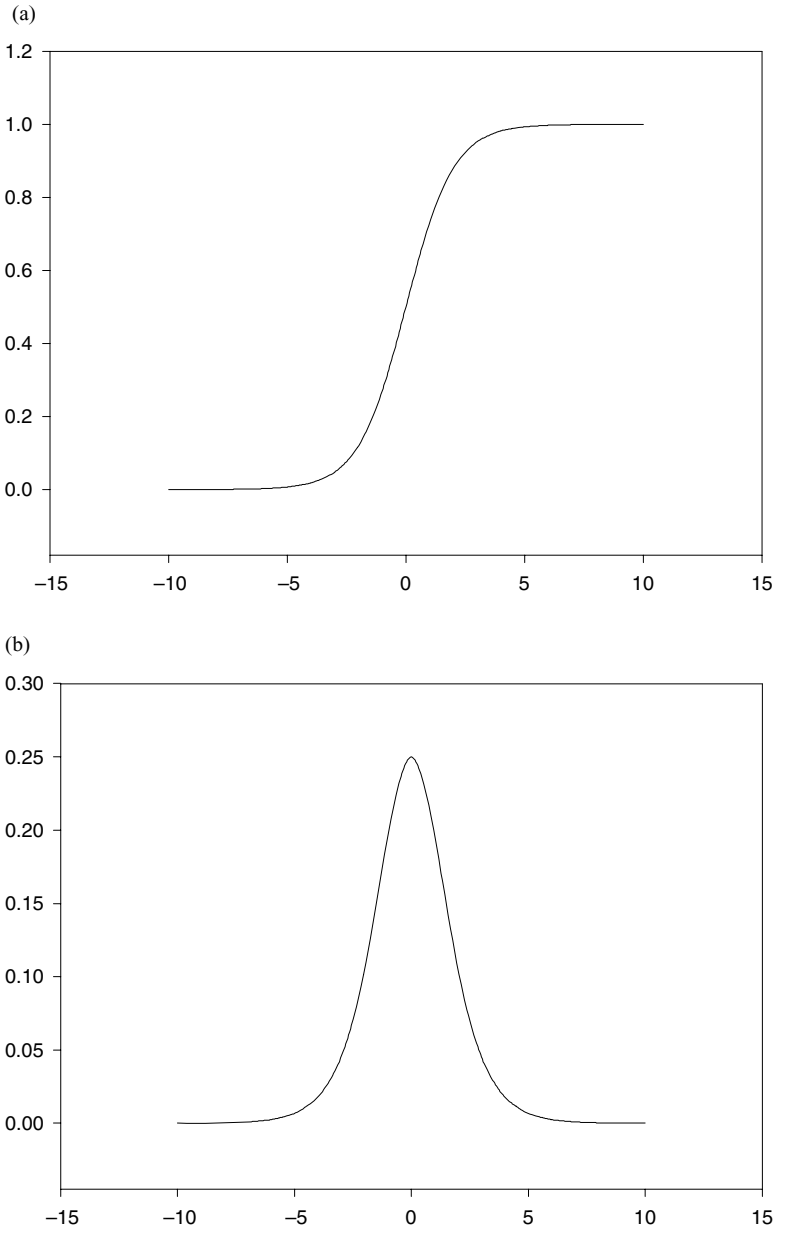
It can be seen that the larger the partial derivatives of the activation function with respect to the weights, that is, $\frac{\partial E(f(w(t))}{\partial w(t)}$, the larger the total weight adjustments will be, leading to the more efficient BPN learning.[2] In order to increase network learning efficiency, the logistic function with the form $S(b) = 1/(1 + e^{-b})$ is employed as the activation function for the current medical expenditure forecasting, where b is the input net of the hidden node, that is, the sum of weighted risk adjusters $\sum x_i w_i$.

Figure 7a and b show the logistic function and their differentials. It is worth noting that the logistic function converges to a constant at the extreme values of hidden nodes in the output layers, and consequently the partial derivative of the logistic function with respect to the connection weights converges to 0. This indicates the stagnation in adjusting the connection weights of the hidden layer nodes and disturbs the network learning.

The advantage of integrating SOM with BPN is that the insured with similar risk factors, which are the input vectors of SOM and BPN, are grouped together, and hence the output values will be within a certain range. More specifically, through the risk classification process by SOM these input vectors have similar characteristics and have few outliers. Thus, the range of the sum of risk factors (x_i) and their connection weights, that is, $\sum x_i w_i$, as well as the output values of hidden layer nodes will

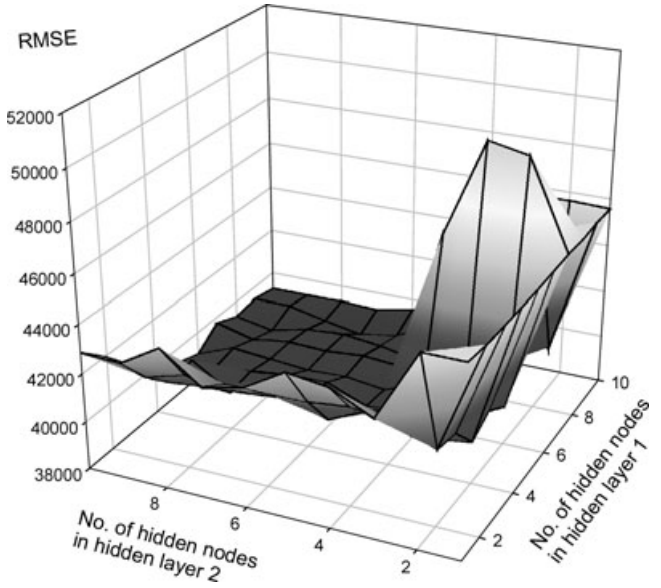
[2] See Haykin (1994) for a more detailed discussion of the network connection weights adjustments.

FIGURE 7
(a) Logistic Function (b) Derivative of the Logistic Function



be clustered within a certain range. In particular, the most effective adjusting range around the logistic function where partial derivatives are larger is actually being used to adjust the connection weights. This allows the error term between the estimated and the actual medical expenses to be quickly minimized, leading to efficient BPN network learning.

FIGURE 8
RMSE Value of the Estimation Sample for BPN Alone



Estimating (Training) Results of BPN With/Without SOM

All BPN estimates are transformed back to their original units before calculating the fitness assessment in order to allow comparison. The results of 100 times trials on the possible hidden nodes combinations in hidden layers 1 and 2 are shown in Figure 8. It shows the RMSE (Fair, 1986) values associated with different combinations of hidden nodes from the estimation sample with regard to the BPN model alone. It is obvious that the RMSE prediction error increases as the number of nodes in the hidden layer decreases. In Figure 8, for instance, when there are two and one hidden nodes on the first and second hidden layers, respectively, the RMSE value is 45,584, which is relatively high compared to other possibilities on the ordinate.

The same pattern of the RMSE values for the estimation samples is retained when BPN is integrated with SOM. As mentioned before, the number of hidden nodes should be selected with care, since too few hidden nodes will lead to poor predictive power. We select the number of hidden nodes for the estimating sample by finding the optimal model fitness based on the criteria that the RMSE is low. Table 7 shows the selected numbers of hidden nodes and the corresponding fitness assessments where BPN is, or is not, integrated with SOM.

PREDICTIVE PERFORMANCE ASSESSMENT OF ALTERNATIVE RISK-ADJUSTMENT MODELS

To evaluate the performance of the alternative risk-adjustment models, it is important to test a validation sample that differs from the estimation sample used to establish the model parameters. If a single sample is used for both estimation and validation, then generally speaking the explanatory power of the model will be overstated (Pope et al., 1998). Typically, a "split-sample" design is employed, where models are

TABLE 7
Hidden Nodes Used and Estimation Results for BPN With/Without SOM

No. of Hidden Nodes	BPN Alone	SOM High-Risk Cohort + BPN	SOM Low-Risk Cohort + BPN
Hidden layer 1	9	10	4
Hidden layer 2	7	6	8
RMSE	39,046	113,859	21,647

estimated based on a portion of a cross-sectional sample and then validated using the remainder of the sample. The relatively small sample sizes available for the NHRI in any given year render the cross-sectional split-sample design unattractive. Using this design may lead to highly unstable parameter estimates and validation results (Pope et al., 1998). Because the national health insurance budget is determined prospectively at the beginning of each year, we exploited the longitudinal nature of the NHRI by using 1999 risk adjusters to estimate the 2000 expenditures and then validating models using 2000 risk adjusters to predict 2001 expenditures. Under this approach, 2 years of data are required for both estimation and validation, because the purpose is to evaluate prospective risk-adjustment models that use previous risk adjusters to predict expenditures during the subsequent year. In addition, a 10-fold cross-validation is used to show the stable and reliable performance improvement in the proposed integration model of SOM and BPN. Specifically, first the estimation as well as the validation samples are each randomly divided into 10 data sets. One out of the 10 estimation subsamples is retained and all the data not in this one retained subsample are trained. The performance measures of the corresponding validation (test) subsample in the next year are calculated according to these 10 sample subdivisions.

Two traditional performance measures, that is , RMSE and mean absolute error (MAE, Fair, 1986), are used to compare the predictive performance of the linear, log-linear regression, two-part, and BPN with/without SOM integrated models. In addition, we also employ the predictive ratio (PR, Ash et al., 1989) and correlation coefficient (CC, Fair, 1986) for assessing these alternative risk-adjustment models. The predictive ratio compares the prediction results for the validation sample with the actual value and is used to assess the accuracy of the model predictions of overall medical expenses. It is defined as follows:

$$\text{Predictive ratio} = \left(\sum_{j=1}^{n} \hat{Y}_j\right) \Big/ \left(\sum_{j=1}^{n} Y_j\right), \tag{12}$$

where n represents the observations in the validation sample, and Y_j and $\hat{Y}_j$ are the actual and predicted individual annual medical expenses of the validation sample, respectively. A PR value greater than 1 indicates groups for which the model will lead to overpayment, whereas a PR value of less than 1 reflects groups whose costs are higher than the model prediction. The best model will have all PR for the selection of subgroups quite close to 1.

TABLE 8
Comparison of Predictive Performance for Different Risk-Adjustment Models by 10-Fold Cross-Validation

		Linear	Log-Linear	Two-Part	BPN Alone	SOM + BPN
RMSE	Mean	40,534.739	7.857e + 10	40,299.339	39,921.659	33,310.381
	Std dev	12,966.206	2.482e + 11	13,048.963	13,440.687	9,112.0839
	Std err mean	4,100.2745	7.849e + 10	4,126.4445	4,250.3185	2,881.4939
	Upper 95% mean	49,810.205	2.561e + 11	49,634.005	49,536.547	39,828.773
	Lower 95% mean	31,259.274	−9.9e + 10	30,964.673	30,306.77	26,791.989
MAE	Mean	12,343.573	3.6624e + 9	11,718.201	11,920.374	11,302.245
	Std dev	1,537.9645	1.157e + 10	1,515.1158	1,347.1808	1,610.677
	Std err mean	486.34709	3.6581e + 9	479.12167	426.01598	509.34079
	Upper 95% mean	13,443.767	1.194e + 10	12,802.05	12,884.089	12,454.454
	Lower 95% mean	11,243.38	−4.613e + 9	10,634.353	10,956.658	10,150.036
PR	Mean	1.0383815	256,598.75	0.9754217	0.844	0.9157
	Std dev	0.1394987	810,563.21	0.1235288	0.1573086	0.1420425
	Std err mean	0.0441134	256,322.59	0.0390632	0.0497454	0.0449178
	Upper 95% mean	1.1381729	836,440.74	1.0637889	0.9565318	1.0173111
	Lower 95% mean	0.9385901	−323,243.2	0.8870545	0.7314682	0.8140889
CC	Mean	0.5903336	0.4709783	0.5953253	0.5934	0.5957
	Std dev	0.2124349	0.3141637	0.2123497	0.2066168	0.1635906
	Std err mean	0.0671778	0.0993473	0.0671509	0.065338	0.0517319
	Upper 95% mean	0.7423004	0.6957175	0.7472311	0.7412047	0.7127257
	Lower 95% mean	0.4383668	0.2462392	0.4434194	0.4455953	0.4786743
			$N = 10$			

Furthermore, the CC is used to show the correction between actual and predicted individual annual medical expenses of the validation sample. The closer the CC is to 1, the better the predictive performance is. The CC is formulized as follows:

$$\mathrm{CC} = \sum_{j=1}^{n}(Y_j - \bar{Y})(\hat{Y}_j - \bar{\hat{Y}}) \Bigg/ \sqrt{\sum_{j=1}^{n}(Y_j - \bar{Y})^2 * \sum_{j=1}^{n}(\hat{Y}_j - \bar{\hat{Y}})^2}, \tag{13}$$

where n, Y_j, and $\hat{Y}_j$ are defined as in Equation (12), and $\bar{Y}$ and $\bar{\hat{Y}}$ are the actual and the predicted mean of an individual's annual medical expenses of the validation sample, respectively.

Table 8 shows the 10-fold cross-validation results using these four assessment indexes. In particular, the mean, standard deviation, and standard error mean of each performance measure together with an interval specifying the upper and lower 95 percent mean of RMSE, MAE, PR, and CC are shown. The integrated model of SOM and BPN possesses lower mean values of MAE and RMSE but higher mean values of CC and PR than those derived from the benchmarks as well as from the BPN alone. This indicates that integrating SOM and BPN provides better forecasting power. In addition, the slight standard deviation together with a relatively smaller mean interval of the proposed model shows that the prediction improvement is stable across the

estimation/validation sampling. The unsatisfactory results of the log-linear model may be attributed to the retransformation problem mentioned before. All these models, except for the linear regression, yield underestimates. Nevertheless, the estimation bias from the integrated model of SOM and BPN is relatively small, suggesting that the risk classification by SOM can increase the predictive power of a risk-adjustment model.

The predictive power of the BPN alone and the linear regression models, both without SOM classification, are somewhat diversified. This is evident by the fact that although the PR values for the BPN exhibit less compelling results, the RMSE, MAE, and CC are nonetheless better than those from the linear regression. In addition, it should be noted that although the forecasting results of the two-part model are inferior to those of the integration model of SOM and BPN, its predictive performance is marginally better than those of all the other benchmarks and even comes close when compared to BPN alone. In particular, except for the RMSE, the remaining three assessment indexes give support to the two-part model rather than to the BPN alone. This suggests that the sample division by a logit or probit function yields more convincing forecasts than doing nothing.

SUMMARY AND CONCLUSIONS

The past decade has witnessed an increased use of neural networks in insurance-related applications. Shapiro (2002) proposed the possibility of merging neural networks, fuzzy logic, and genetic algorithms in order to capitalize on their strengths and compensate for their shortcomings. He proposed to adopt fuzzy inputs and/or fuzzy weights in neural networks to allow the use of neural networks as a universal approximator (Buckley and Hayashi, 1994; Feuring, Buckley, and Hayashi, 1998; Jiao et al., 1999; Shapiro, 2002; Shapiro and Jain, 2003), or to use a neural network to enhance the convergence of the genetic algorithm in the search for a global optimum (Javadi et al., 2005). This article integrated two neural networks to exploit the possible synergy effects of SOM and BPN to enhance the predictive power of the risk-adjustment model. Our main conclusion is that such a risk-adjusted capitation formula will reduce the incentives for cream skimming by decreasing estimation biases. The better model fitness of the integrated model of SOM and BPN may be due to the following two reasons:

1. Data availability: The problem of health spending having a thick upper tail may be dealt with by using extremely large samples and by correcting standard errors for heteroskedasticity. However, there is the significant practical problem of both the availability and how to acquire a large amount of data when estimating by means of the cluster analysis or statistical regression models. Neural networks relax this limitation on the sample size.
2. Fitness: The most widespread method for nonlinear transformation is logarithmic, such as the two-part or the multi-part model. However, estimators are biased if the residual still does not fit assumptions after transformation, and retransformation frequently results in seriously biased estimates. Although outliers can be eliminated, this may also lead to the loss of important information for accurately predicting the medical expenses of certain beneficiaries.

Besides predicting the medical expenditures of the beneficiaries in order to calculate the risk-adjusted capitation payments, this article identifies two risk groups for beneficiaries in Taiwan. Unlike two- or multi-part models that mainly rely on loss frequency for sample division, SOM clustering can be based on several risk adjusters. It should be noted that SOM possesses not only clustering ability but also the potential for analyzing changes in medical expenditures or the effects of any incentive mechanism on controlling medical demand and health insurance expenses. As noted in the previous section, the size of the high-risk cohort was reduced by 30 beneficiaries in 2000 owing to the net decrease of individuals shifting between the low- and high-risk groups. Several perspectives can be used to analyze the reasons for these shifting patterns. Of course, one perspective is the altered health status of an individual. However, another possibility with more interesting implications comes from the reform in the health-care system, such as the implementation of demand/or supply-side incentive mechanisms.

A two-stage SOM model can be applied for clustering changing patterns based on different periods separated by the timing of various policy reforms. During the first stage, data can be divided into several periods according to the implementation of incentive mechanisms or reforms. Patients can be classified using SOM based on the previously mentioned risk adjusters within each period. In the second stage, the "trajectory" of the SOM for each patient during different periods can be clustered again by SOM to derive the changing pattern of the medical expenditures of certain beneficiaries due to the policy reforms. From the trajectory of SOM, beneficiaries remaining in the same risk cohort are either "true" high/low risks or insensitive to policy changes. Further investigation is necessary to determine whether changes in individual risk cohort are attributable to the real changes in health status or if they are due to the effects of reforms made by the government. In addition, consideration of specific risk adjusters revealing diagnostic causes of patients, and the timing of the implementation of policy proposals may provide further implications for a payment system design that complements the risk-adjusted capitation formula.

REFERENCES

Ash, A., F. Porell, L. Gruenberg, E. Sawitz, and A. Beiser, 1989, Adjusting Medicare Capitation Payments Using Prior Hospitalization, *Health Care Financing Review*, 10(4): 17-29.

Back, B., K. Sere, and H. Vanharanta, 1998, Managing Complexity in Large Data Bases Using Self-Organizing Maps, *Accounting Management and Information Technologies*, 8: 191-210.

Barros, P. P., 2003, Cream-Skimming, Incentives for Efficiency and Payment System, *Journal of Health Economics*, 22: 419-413.

Binner, J. M., R. K. Bisondeeal, T. Elger, A. M. Gazely, and A. W. Mullineux, 2005, A Comparison of Linear Forecasting Models and Neural Networks: An Application to Euro Inflation and Euro Divisia, *Applied Economics*, 37(6): 665-680.

Brockett, P. L., W. W. Cooper, L. L. Golden, and U. Pitaktong, 1994, A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency, *Journal of Risk and Insurance*, 61(3): 402-424.

Brockett, P. L., X. Xia, and R. A. Derrig, 1998, Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *Journal of Risk and Insurance*, 65(2): 245-274.

Buckley, J. J., and Y. Hayashi, 1994, *Fuzzy Neural Networks* in: R. R. Yager and L. A. Zadeh, eds., (New York: Van Nostrand Reinhold), 233-249.

Bureau of National Health Insurance, 2003, *National Health Insurance Annual Statistical Report*. Department of Health, Executive Yuan, Republic of China: Bureau of National Health Insurance.

Charalambous, C., A. Charitou, and F. Kaourou, 2000, Comparative Analysis of Artificial Neural Network Models: Application in Bankruptcy Prediction, *Annals of Operations Research*, 99: 403-425.

Do, A. Q., and G. Grudintski, 1992, A Neural Network Approach to Residential Property Appraisal, *Real Estate Appraiser*, 58(3): 38-45.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, 1983, A Comparison of Alternative Models for the Demand for Medical Care, *Journal of Business and Economic Statistics*, 1(2): 115-126.

Fair, R. C., 1986, Evaluating the Predictive Accuracy of Models, in: Griliches, Z. and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 33, (New York: North Holland).

Feuring, T., J. Buckley, and J. Y. Hayashi, 1998, Adjusting Fuzzy Weights in Fuzzy Neural Nets, Proceedings of the Second International Conference on Knowledge-Based Intelligent Electronic System, 402-406.

George, S., and Y. Yang, 1992, Applying Artificial Neural Networks to Investment Analysis, *Financial Analysts Journal*, 48(5): 78-80.

Grudintski, G., A. Q. Do, and J. D. Shilling, 1995, A Neural Network Analysis of Mortgage Choice, *Intelligent Systems in Accounting Finance and Management*, 4: 127-135.

Grudnitski, G., and L. Osburn, 1993, Forecasting S&P and Gold Future Prices: An Application of Networks, *Journal of Futures Markets*, 13(6): 631-643.

Haykin, S., 1994, *Neural Networks: A Comprehensive Foundation* (New York: Macmillan).

Hoptroff, A. R., 1993, The Principles and Practice of Time Series Forecasting and Business Modelling Using Neural Nets, *Neural Computing and Applications*, 1(1): 59-66.

Hsu, S., C. Lin, and Y. Yang, 2006, Risk Classification and Prediction of Individual Health Expenses, *Management Review*, 25(4): 27-48.

Huang, C. S., R. E. Dorsey, and M. A. Boose, 1994, Life Insurer Financial Distress Prediction: A Neural Network Model, *Journal of Insurance Regulation*, 13(2): 131-167.

Javadi, S., S. Djajadiningrat-Laanen, H. Kooistra, A. M. van Dongen, G. Voorhout, F. J. van Sluijs, T. S. van den Ingh, W. H. Boer, and A. Rijnberk, 2005, Primary Hyperaldosteronism, *Mediator of Progressive Renal Disease in Cats*, 28: 85-104.

Jiao, K., S. A. Bullard, L. Salem, and E. Robert, 1999, Coordination of the Initiation of Recombination and the Reductional Division in Meiosis in Saccharomyces Cerevisiae, *Genetics*, 152: 117-128.

Kaski, S., J. Sinkkonen, and J. Peltonen, 2001, Bankruptcy Analysis With Self-Organizing Maps in Learning Metrics, *IEEE Transactions on Neural Networks*, 12(4): 936-947.

Kasslin, M., J. Kangas, and O. Simula, 1992, Process State Monitoring Using Self-Organizing Maps, *Proceedings of the 1992 International Conference (ICANN-92)*, 2(2): 1531-1534.

Kiviluoto, K., and P. Bergius, 1998, Two-Level Self-Organizing Maps for Analysis of Financial Statements, *Neural Networks Proceedings of IEEE World Congress on Computational Intelligence*, 1: 89-192.

Kohonen, T., 1982, Self-Organizing Formation of Topologically Correct Feature Maps, *Biological Cybernetics*, 43: 59-69.

Kohonen, T., 1989, *Self-Organizing and Associative Memory*, 3rd edition (New York: Spring-Verlag).

Kohonen, T., 1990, The Self-Organizing Map, *Proceedings of the IEEE*, 78(9): 1464-1480.

Länsiluoto, A., T. Eklund, B. Back, H. Vanharanta, and A. Visa, 2004, Industry-Specific Cycles and Companies' Financial Performance Comparison Using Self-Organizing Maps, *Benchmarking: An International Journal*, 11(3): 267-283.

Lewis, O. M., J. A. Ware, and D. Jenkins, 1997, A Novel Neural Network Technique for the Valuation of Residential Property, *Neural Computing & Applications*, 5(4): 224-229.

Mirmirani, S., and H. C. Li, 2004, Gold Price, Neural Networks and Genetic Algorithm, *Computational Economics*, 23(2): 193-200.

Mullahy, J., 1998, Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Economics, *Journal of Health Economics*, 17(3): 247-282.

Narain, L. S., and R. L. Narain, 2002, Stock Market Prediction: A Comparative Study of Multivariate Statistical and Artificial Neural Network Models, *Journal of Accounting and Finance Research*, 10(2): 85-94.

Newhouse, J. P., 1977, Medical Care Expenditure: A Cross-National Survey, *Journal of Human Resources*, 12: 115-125.

Oja, M., S. Kaski, and T. Kohonen, 2003, Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, *Neural Computing Surveys*, 3: 1-156.

Pope, G. C., K. W. Adamache, E. G. Walsh, and R. K. Khandker, 1998, Evaluating Alternative Adjusters for Medicare, *Health Care Financing Review*, 20(2): 109-129.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986, Learning Representations by Back-Propagating Errors, *Nature*, 323: 533-536.

Serrano-Cinca, C., 1996, Self Organizing Neural Networks for Financial Diagnosis, *Decision Support Systems*, 17: 227-238.

Serrano-Cinca, C., 1997, Feedforward Neural Networks in the Classification of Financial Information, *European Journal of Finance*, 3(3): 183-202.

Shapiro, A. F., 2002, The Merging of Neural Networks, Fuzzy Logic, and Genetic Algorithm, *Insurance Mathematics and Economics*, 31: 115-131.

Shapiro, A. F., and L. C. Jain, 2003, *Intelligent and Other Computational Techniques in Insurance* (London: World Scientific).

Tam, K. Y., and M. Y. Kiang, 1992, Managerial Applications of Neural Networks: The Case Bank Failure Predictions, *Management Science*, 38: 926-947.

Tryba, V., S. Metzen, and K. Goser, 1989, Designing Basic Integrated Circuits, in: Neuro-Nimes '89. International Workshop on Neural Networks and Their Applications, 225-235.

Ultsch, A., and H. P. Siemon, 1990, Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis, Proceedings of the International Neural Network Conference (INNC'90), Dordrecht, Netherlands, 305-308.

van de Ven, W., and R. P. Ellis, 2000, Risk Adjustment in Competitive Health Plan Markets, in: A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics* Vol. 1 (Amsterdam, the Netherlands: Elsevier Science).

Vermuelen, E. M., J. Spronk, and D. van Der Wijst, 1994, Visualizing Interfirm Comparison, *International Journal of Management Science*, 22: 237-249.

Worzala, E., M. Lenk, and A. Silva, 1995, An Exploration of Neural Networks and Its Application to Real Estate Evaluation, *Journal of Real Estate Research*, 10(2): 185-201.