The Philippine Statistician, 2003
Vol. 52, Nos. 1-4, pp. 39-57

Visualizing Regional and Provincial Poverty Structures Via the Self-Organizing Map[1]

Jose Ramon G. Albert, Lilia V. Elloso, Eric B. Suan and Mary Ann C. Magtulis[2]

received January, 2003; revised August, 2003

ABSTRACT

In order to bring about a world free of poverty, researchers and policy-makers would like to have a geographically-disaggregated poverty indicator system available to them. Such a system would enable a better understanding of the spatial distribution of poverty, and consequently result in a more efficient targeting of the poor. Typically, setting up such a system entails the collection of a heterogeneous geographically referenced poverty data since poverty is a complex social phenomenon, as well as the use of a Geographic Information System (GIS) for integrating the poverty data. Here, we discuss the use of the Self-Organizing Map (SOM) for analyzing various comparable sub-national poverty and welfare indicators across several time periods.

Key words: Self-Organizing Map, poverty, neural network

1. INTRODUCTION

For the past several decades, national economic development plans have underscored poverty alleviation as part of its overall goal. (Jurado, 2002). At the global front, international organizations have also come to recognize the importance of battling poverty, aiming toward a significant reduction of poverty incidence within the next decade. This makes the measurement and analysis of poverty a vital concern for researchers. It may seem as though counting the poor and analyzing the root causes of poverty is a rather trivial task but poverty is a rather complex social phenomenon so that its analysis actually requires the simultaneous consideration of a number of statistical indicators, especially monetary indicators like consumption and income (see, e.g., Deaton, 1997).

A geographically disaggregated poverty indicator system provides for a visualization of the disparities of living standards in geographic areas. The poverty indicator system can be also used as a decision support system to come up with specific evidence-based interventions, programs and plans for identifying areas that ought to benefit from additional resources for the purpose of poverty reduction. In the Philippines, for instance, the National Anti-Poverty Commission has been planning to construct a poverty map as a policy tool for designing a sub-national poverty reduction strategy through the *Kapit-bisig Laban sa Kahirapan* (KALAHI) program.

The literature, e.g. Bigman, *et al.* (2000), typically suggests the use of a Geographic Information System (GIS) for integrating heterogeneous, geographically referenced poverty data. Here, we exhibit and discuss an alternative poverty mapping system based on the Self Organizing Map that visually displays the structures of poverty relations among comparable

---

[1] Presented at the 54[th] Session of the ISI in Berlin, Germany, and at the National Mapping Research and Information Authority GIS Conference (held December 2002) at the Galleria Suites, Pasig City

[2] Chief, Statistician IV, Statistician III and Statistician III, respectively of the Research Division, Statistical Research and Training Center, J & S Building, 104 Kalayaan Ave. Diliman, Quezon City. Email of first author: jrgalbert@srtc.gov.ph. The findings, interpretations, and conclusions in this paper are the authors' own.

geographic areas at the sub-national level (such as regions and provinces) based on a poverty database sourced from surveys conducted by the National Statistics Office, viz., the Family Income and Expenditure Survey (FIES) and the Annual Poverty Indicator Survey (APIS). The maps shown in this paper follow the work of Kashi and Kohonen (1996) on the construction of a Kohonen Map for obtaining a welfare comparison of countries based on selected statistical indicators.

2. THE SELF-ORGANIZING MAP

The Self Organizing Map (SOM) is a rather robust analytical visualization tool developed by Teuvo Kohonen in the early 1980s based on an artificial neural network design.[i] It essentially provides a topological mapping from a general input space to a cluster of objects (Kohonen, 2001). The flow diagram for constructing the SOM is shown in Figure .1. The critical portion of the flow diagram is the Processing part, which is described in more detail below.

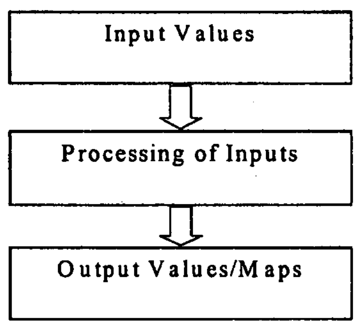

Figure 1: Flow Diagram for Constructing a Self-Organizing Map

As an artificial neural network, the SOM architecture is inspired from the operation of the human brain (see, e.g. Haykin, 1999). It is composed of neurons disposed on a grid that is usually two dimensional (cf. Figure 2), but sometimes the grid is one-dimensional and (rarely) three dimensional or higher. Unlike other neural network designs, the SOM is also a data compression tool that adaptively learns from a complex, high dimensional data set and projects the data set into a reduced data manifold, typically a regular two-dimensional array of visually decipherable clusters of nodes through a nonlinear projection of the probability density function of the input data.

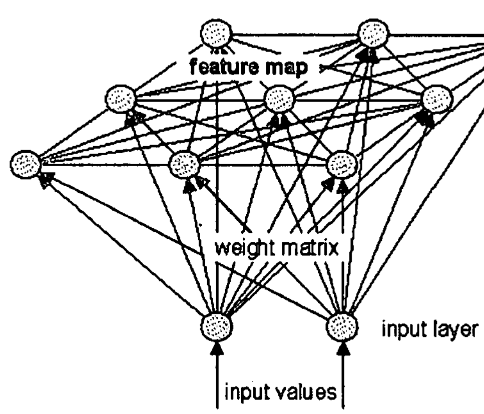

Figure 2: Representation of a Self-Organizing Map

As a projection method, the procedure for constructing the SOM is conceptually similar to that of multidimensional scaling (MDS) methods. (Ripley,1998). The SOM and MDS method, however, differ in that MDS tries to preserve the metric (or in the case of nonmetric MDS, the global ordering relations) of the original space, while the SOM tries to preserve the topology, i.e. the local neighborhood relations, of the data manifold.

The SOM may also be thought of as a cluster analytic tool since the Kohonen map attempts to represent all the available data with optimal accuracy in such a way that the units to be compared become ordered on the grid with similar units close to each other and dissimilar units far from each other. As a clustering tool, the SOM allows a data user to identify similar objects by their attributes and form clusters that are homogenous and different from other clusters. Unlike standard clustering tools such as the K-means clustering algorithm (see, e.g., Everitt, 2001), the SOM offers a visual contour map of the patterns in the input data; this contour map allows the decision maker to visually analyze and explain the relationships among the points in the map as well as the flexibility of choosing from multiple grouping alternatives. For a comprehensive description of algorithms for construction of the SOM, see Kohonen (2001) or Vesanto *et al.* (2000).

The basic architecture of the SOM is as follows. An adaptive parameter vector $\boldsymbol{m}_i$, called a model vector, is associated with every node (or neuron) $\boldsymbol{i}$ in an array of two dimensional map nodes. The model vectors have the same number of elements as the input vector $\boldsymbol{x}$. The values of the model vectors are changed iteratively either through a stochastic or batch-type algorithm.

The stochastic algorithm performs a competitive learning process through a nonparametric recursive regression. Initial values of the components of the model vectors are arbitrary, and may even be chosen at random. In practice, the initial values are selected in some orderly fashion[ii] so that no prior order is imposed on the feature map. All data vectors x are projected to the map by finding the best match of the data with the model vectors. This matching "competition" is done by determining the distance[iii] in some metric (usually the normal Euclidean distance) between the vectors $\boldsymbol{m}_i$ and $\boldsymbol{x}$ for all $\boldsymbol{i}$, and choosing the model vector that is closest. That is, we determine the index $\boldsymbol{c}$ that satisfies the condition

$$\boldsymbol{c} = \arg\min_i \|\boldsymbol{x} - \boldsymbol{m}_i\|$$

The particular node $\boldsymbol{c}$ is called the winning node for the input vector $\boldsymbol{x}$. Note that minimizing the distance between $\boldsymbol{m}_i$ and $\boldsymbol{x}$ for all $\boldsymbol{i}$ is mathematically equivalent[iv] to choosing the largest inner product $\boldsymbol{m}_i^T\boldsymbol{x}$ for each of the nodes.

The winning node locates the center of a topological neighborhood of cooperating nodes. An unsupervised learning principle leads to gradual and adaptive changes in the values of the model vectors. At step $\boldsymbol{t}$ of the random sequence of the given $X(t)$ values, the values of the $\boldsymbol{m}_i$ are adapted according to one of the following equations:

(a) $\boldsymbol{m}_i(\boldsymbol{t}+1) \leftarrow \boldsymbol{m}_i(\boldsymbol{t}) + \boldsymbol{h}_{ci}(\boldsymbol{t})[\boldsymbol{x}(\boldsymbol{t}) - \boldsymbol{m}_i(\boldsymbol{t})]$ for each node $\boldsymbol{i}$ in the neighborhood around the winner

(b) $\boldsymbol{m}_i(\boldsymbol{t}+1) \leftarrow \boldsymbol{m}_i(\boldsymbol{t})$

where

$$h_{ij}(t) = h(\|r_i - r_j\|; t)$$

is some neighborhood kernel function, r_i and r_j are the geometric locations of the nodes i and j, respectively, for which $h_{ij} \to 0$ as either $\|r_i - r_j\| \to \infty$, or $t \to \infty$. This regression is iterated over the available data. Widely applied kernel functions are monotonically decreasing functions of the time t, or kernels that take the form of a Gaussian function, i.e.

$$h_{ij}(t) = \alpha(t)\exp\left(-\frac{\|r_i - r_j\|^2}{2\sigma^2(t)}\right)$$

where $\alpha(t)$ and $\sigma(t)$ are monotonically decreasing functions of t, respectively representing a learning rate and a kernel width. The learning rate can be linear, exponential or inversely proportional to the time t. The selection of the learning rate becomes important if one has to construct very large maps.

The basic idea in the SOM learning process is that the further a neighbor is from the "winner" of the competition, the smaller its weight change. Also, while at the beginning of the learning process, the radius of the neighborhood may be chosen to be fairly large, however, as the training goes on, the neighborhood gradually shrinks so that local corrections of the model vectors will be more specific. At the end of training, the neighborhoods have shrunk to zero size. During the learning process, individual changes may actually be contradictory, but the net outcome is that ordered values for the model vectors emerge over the array.

Typically, the training process is performed in two phases. In the first phase, the initial values $\alpha(0)$ and $\sigma(0)$ of the learning rate and neighborhood radius are set to be rather large. In the second phase, the initial values are already small. This essentially corresponds to firstly tuning the Kohonen map to the same space as the input data and then fine tuning the map.

The batch algorithm is also an iterative algorithm just like the stochastic algorithm, but instead of using a single data vector at a time, the whole data set is presented to the map before any adjustments are made. For a comprehensive description of the batch algorithm and its implementation details, see Kohonen (2001) or Vesanto *et al.* (2000).

Early applications of the Kohonen map were in the area of speech recognition but they have since expanded into visualizations of complex data sets in other engineering problems, specifically image data compression and image or character recognition, and process control; applications have also been developed for economic analysis, medical diagnosis as well as diagnostic studies in industry. (Kohonen, 2001) Much work has been done on perturbing the algorithm for constructing SOMs (see, e.g., Obermayer and Sejnowski, 2001). Details about the selection of the parameters, variants of the map, and many other aspects about constructing such maps are covered in Kohonen (2001). The aim of this paper is not to expound on the mathematical and statistical properties of the SOM, but to regard the SOM as an alternative approach for poverty mapping. Self-organizing maps may be constructed with the aid of the SOM-PAK program (Kohonen *et al.*, 1996) available at the website

http://www.cis.hut.fi/research/som_pak/

This package is a public domain software developed at the Neural Networks Research Center of the Helsinki University of Technology. It is written in C language for UNIX and PC environments and has also been ported for use in commercial packages such as Data Mining Solutions, e.g., SPSS Clementine and SAS Enterprise Miner (Wielanga, Lucas and George, 1999), and in other commercial software such as MATLAB (see Vesanto, *et al.* 2000 for details) through its add-on SOM Toolbox.

3. SUB-NATIONAL POVERTY DATABASE

The regional and provincial level poverty database used for constructing the SOMs in the next section are sourced from a wealth of monetary and non-monetary information that can be generated from the 1997 FIES, the 1998 APIS, the 1999 APIS and the 2000 FIES.

Official poverty measurement in the Philippines from 1988 onwards is based on the FIES, which is conducted once every three years. Since 1998, the NSO conducted the APIS during non-FIES years. The conduct of the APIS is in response for the need of policy makers and various stakeholders on poverty issues to have more frequent and reliable information on poverty, especially on non-income based poverty correlates.

The FIES and APIS are designed to have rather reliable estimates of poverty statistics at the provincial level. Strictly speaking, however, the FIES and APIS are actually not comparable surveys. The consumption module of the 1997 FIES is much more robust and detailed (going up to more than 20 pages of more than 400 expenditure lines) than the 1998 APIS (2 page) module (which consists of 27 expenditure lines). In addition, the FIES has a full calendar year reference period (January to December) while the APIS data is limited to the second and third quarters of the year.

Bearing in mind the incomparability of data sources, we constructed a regional and provincial poverty database for years 1997-2000 from information obtained in the FIES and APIS. This database was then used for constructing the Kohonen maps in the next section. It contains the following poverty indicators:

(a) poverty headcount measure[v]
(b) poverty gap measure[vi]
(c) poverty severity measure[vii]
(d) food poor[viii]-headcount measure
(e) food poor-gap measure
(f) food poor-severity measure
(g) percentage of population belonging to families:
    a. with no access to potable water
    b. with family sizes greater than 6
    c. with makeshift/light housing materials
    d. with no access to electricity
    e. with heads who have no jobs[ix]
    f. whose heads' highest schooling is at most elementary[x]

g. with sanitary toilets
h. whose heads engage in agricultural/fishing businesses

These data describe the poverty profile of the entire population for the different regions and provinces in the country.

4. DISCUSSION

A visualization of the SOM based on the poverty database generated from the 1997 FIES is displayed in Figure 3. Here, we have an illustration of the component indicators of the database and a graphical display of the U-matrix[xi] which shows "distances" between neighboring regions. The component planes offer a convenient way to visualize all components simultaneously and hunt for correlations among the variables. From the component planes, we readily observe variation in the indicators as well as a correlation among practically all the variables. The indicator for access to potable water appears, however, to be the sole variable that is not strongly correlated with the rest of the variables. The diagram of the U-matrix helps visualize the cluster structure of the map and describes the standard of living among the regions. The rows and columns between the units represent the magnitude of the gradient. High values on the U-matrix (black being very high) signify large distances between neighboring map units or regions, and thus indicate cluster borders. The gray-level at each unit symbolizes the median of the surrounding gradients. Clusters are typically areas of low values. The gray-shade colorbar provides a psedocolor reference scale for determining what are high and low values.

A larger perspective of the U-matrix plane in Figure 3 is shown in Figure 4 (a). Labels of the regions for Figures 3 and 4 are listed in Table 1. The visualization in Figure 4 allows us to compare and infer welfare disparities[xii] among the regions from the order of the regions in the display.

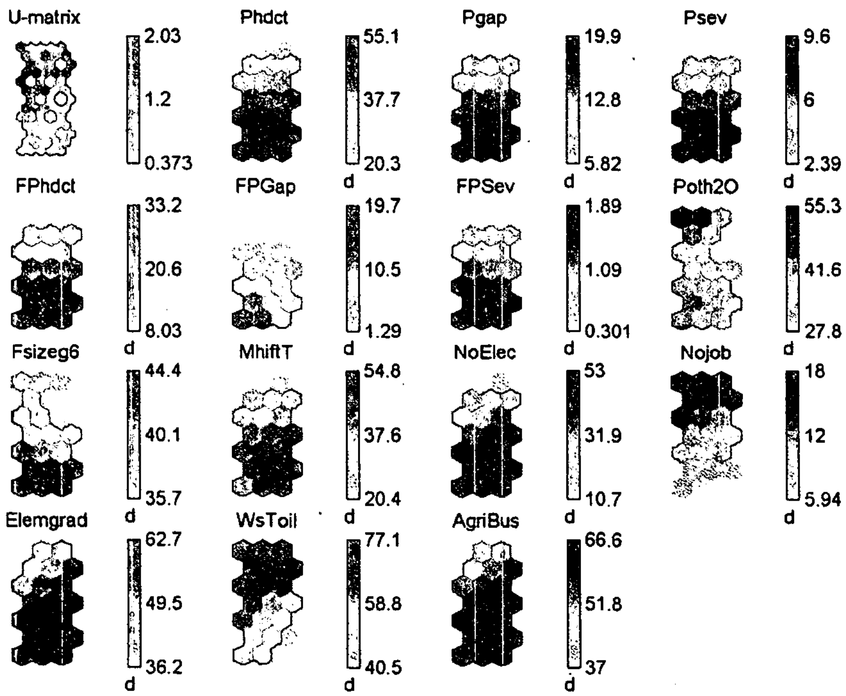

Figure 3: A map display of the poverty database of 14 indicators sourced from the 1997 FIES. The diagram was constructed using the SOM algorithm and describes the 1997 standard of living among the regions.

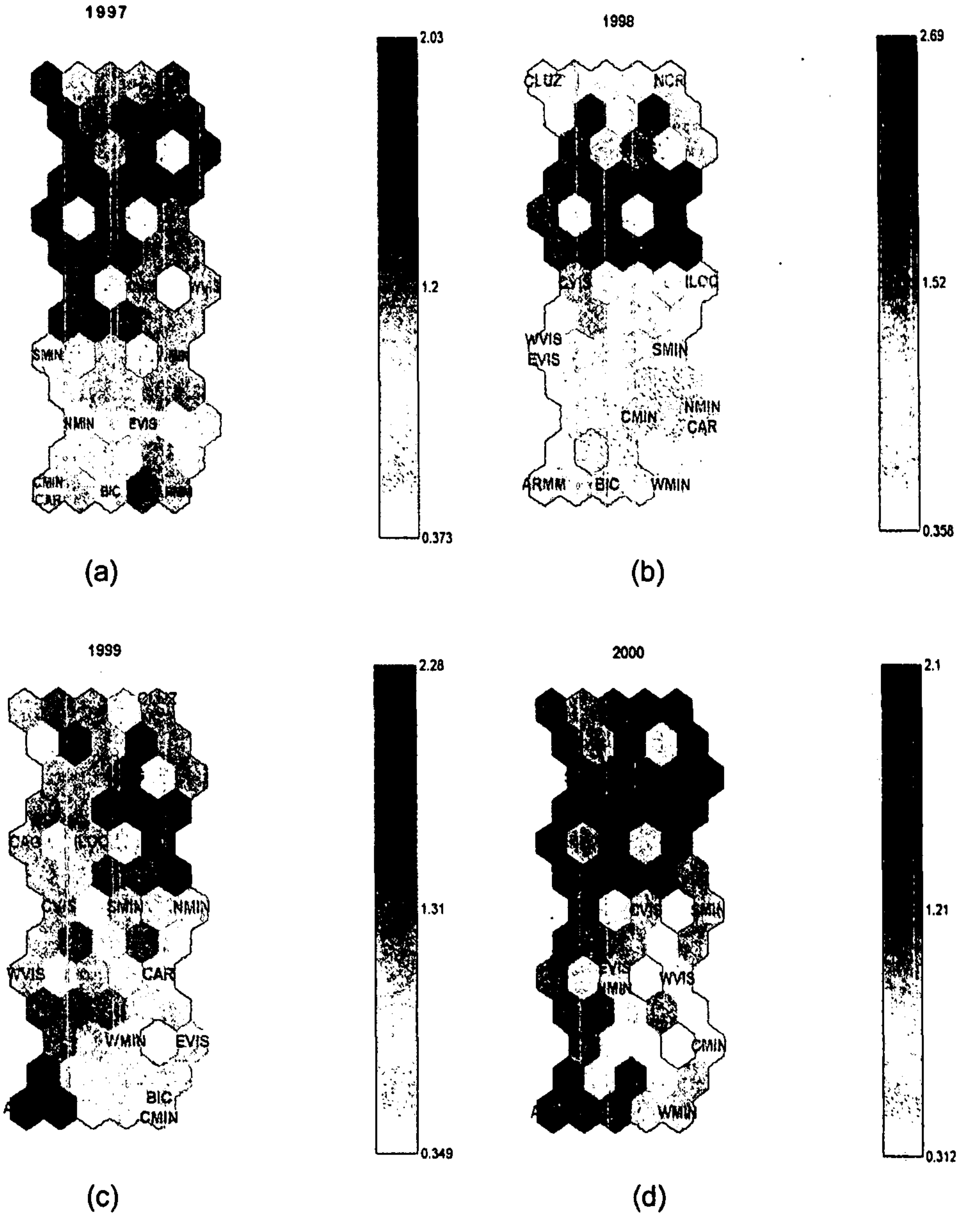

Figure 4:Regional poverty map for (a) 1997 (b) 1998 (c) 1999 and (d) 2000

Table 1. Legend of symbols used in Figures 2 and 3.

ARMM	Autonomous Region of Muslim Mindanao	**ILOC**	Ilocos
BIC	Bicol	**NCR**	National Capital Region
CAG	Cagayan	**NMIN**	Northern Mindanao
CAR	Cordillera Administrative Region	**SMIN**	Southern Mindanao
CLUZ	Central Luzon	**STAG**	Southern Tagalog
CMIN	Central Mindanao	**WMIN**	Western Mindanao
CVIS	Central Visayas	**WVIS**	Western Visayas
EVIS	Eastern Visayas		

In particular, it is readily noted that there are about two regional clusters, with Cagayan Valley along the threshold separating the two clusters. Interestingly, the spatial ordering in the U-matrix map seems to somewhat reflect a geographic ordering (even if actually no geographic information was presented when constructing the map). In particular, the cluster in the lower portion of the map generally refers to Mindanao and Visayas, while the upper edge of the map are by and large the regions in Luzon. This suggests a North-South divide with the cluster consisting of Luzon (without Bicol and the Cordillera Administrative Region) being the areas with the best welfare status, and the other grouping being the areas where government ought to focus its poverty alleviation efforts. The map clearly suggests that government has to improve on the regional welfare disparities, particularly, by focusing attention to Mindanao, CAR and Bicol. These places happen to be where insurgencies and terrorism are concentrated.

Figure 4(a) also manifests that poverty is a rural phenomenon since the country's best welfare conditions are in the urban center Metro Manila and its neighboring regions in Luzon. Poverty statistics for previous years have likewise pointed this out (see, e.g, Intal, 1994). Concrete plans thus have to be drawn up to improve areas outside Metro Manila. Otherwise, we can expect to have a continuous influx toward Metro Manila, leading to overcrowding of the urban center, worsening urban poverty, and consequently the welfare status of Metro Manila over time. The dark shades in the Northern cluster suggest that the members of this cluster are not quite close to each other so that these better off areas are not equally better off (in comparison with the poor areas of the Southern cluster which have light shades of gray indicative that these regions are very similar in their poverty and welfare status).

With a regional database of poverty indicators from 1998 to 2000 sourced from the 1998 APIS, the 1999 APIS and the 2000 FIES, we can also readily generate the poverty maps in Figures 4(b) to 4(d).Comparing Figure 4(a) to (d), we see a steady rise in the position of CAR from 1997 to 2000. This region is known to have experienced high growth and low inflation during this period. ARMM, which in 1997, was relatively no different from other poor regions, appears to be getting more unusual from the other poor regions (as evidenced from its light shade in 1997 and its dark shade in 2000). Also, we readily observe changes in the positions of Ilocos, Southern Tagalog and Eastern Visayas from 1997 to 1998. This is in the wake of the Asian financial crisis and El Niño. Such changes were also noted in a panel study analyzing the effects of the financial crisis and El Niño on poverty (e.g., Tabunda and Albert, 2002). A few movements in the regional positions can also be observed going into 1999 and also into 2000. However, by and large (albeit with the special exception of CAR), the dynamics in regional poverty and welfare structures appear to be rather minor. The figures suggest that poverty structures and welfare relations among the regions do not considerably change over a period of four years. This may be a surprise to policy planners. David (2000) has, in fact, suggested that poverty monitoring of detailed poverty indicators at sub-national levels be done only infrequently especially since the standard errors for estimates at levels of disaggregation lower than regional levels tend to be rather high. Given the cost of conducting surveys, there may be indeed be more sense in doing annual poverty monitoring of only some indicators at the national level. If there is a need to do frequent poverty monitoring of areas where special poverty alleviation programs are being implemented to assess program impact, only these areas ought to be monitored together with a few areas to serve as controls. There may not be much value added in monitoring all areas, as is currently done.

With a provincial[xiii] database of poverty indicators from 1997 to 2000 sourced from the 1997 FIES, the 1998 APIS, the 1999 APIS and the 2000 FIES, we can also readily generate the poverty maps in Figures 5-8. The labels for the provinces in these figures are listed in Table 2.

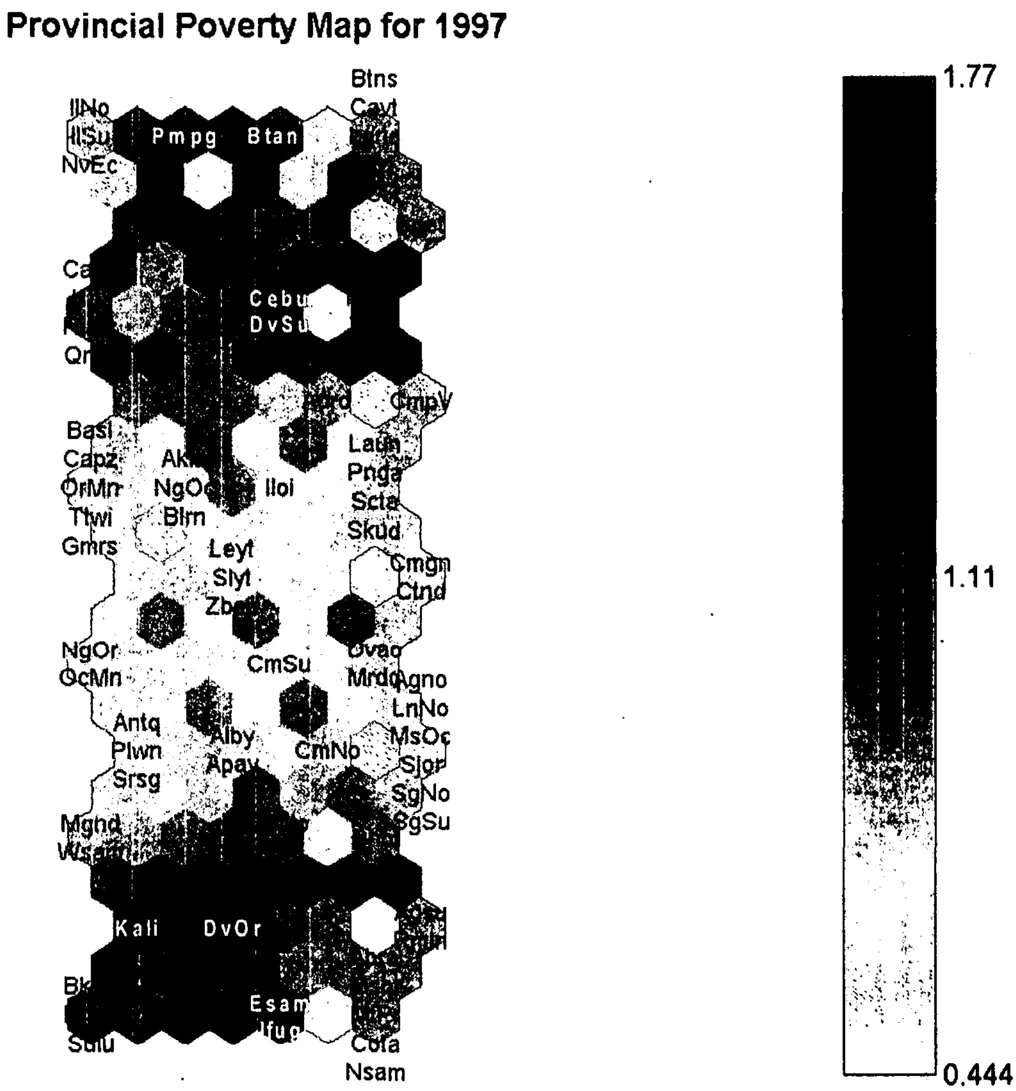

Figure 5:Provincial poverty map for 1997

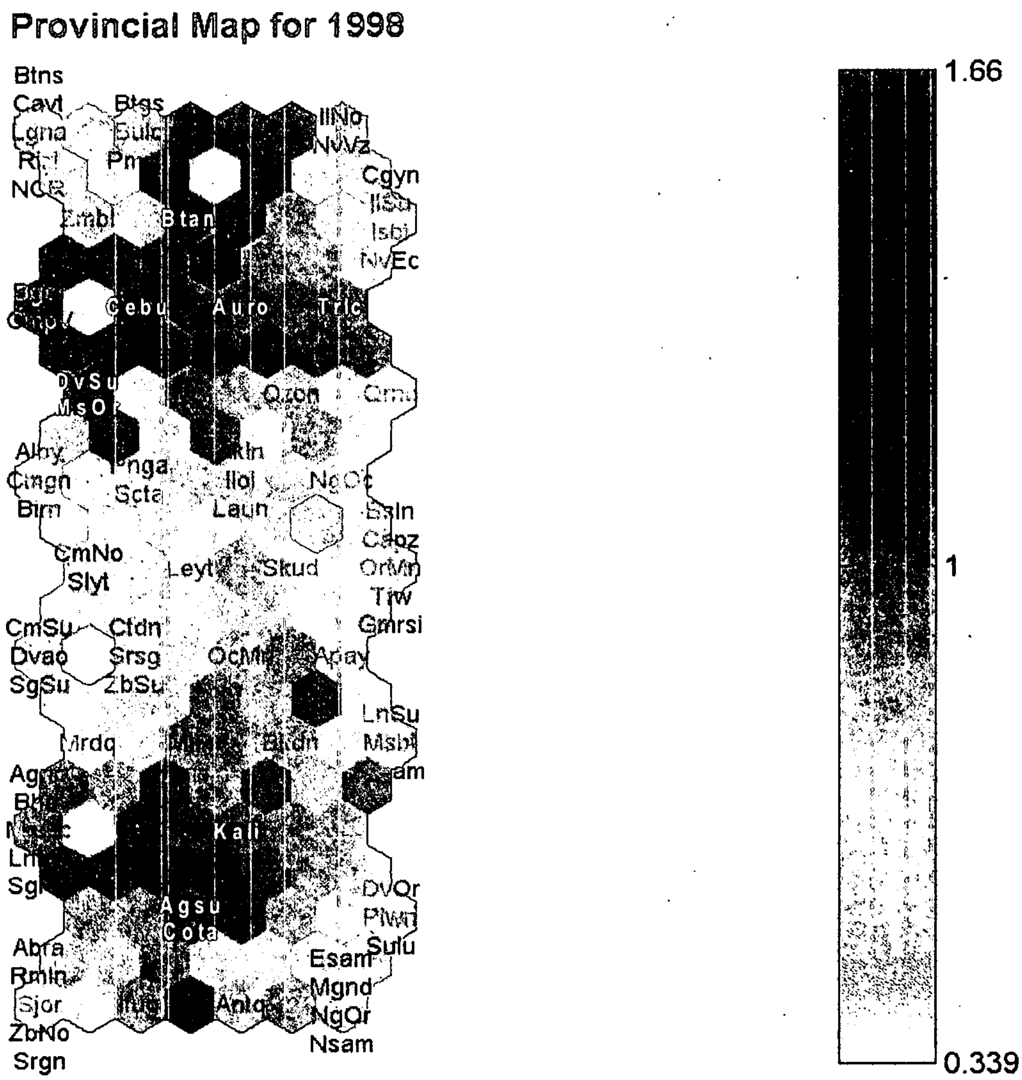

Figure 6:Provincial poverty map for 1998

Provincial Poverty Map for 1999

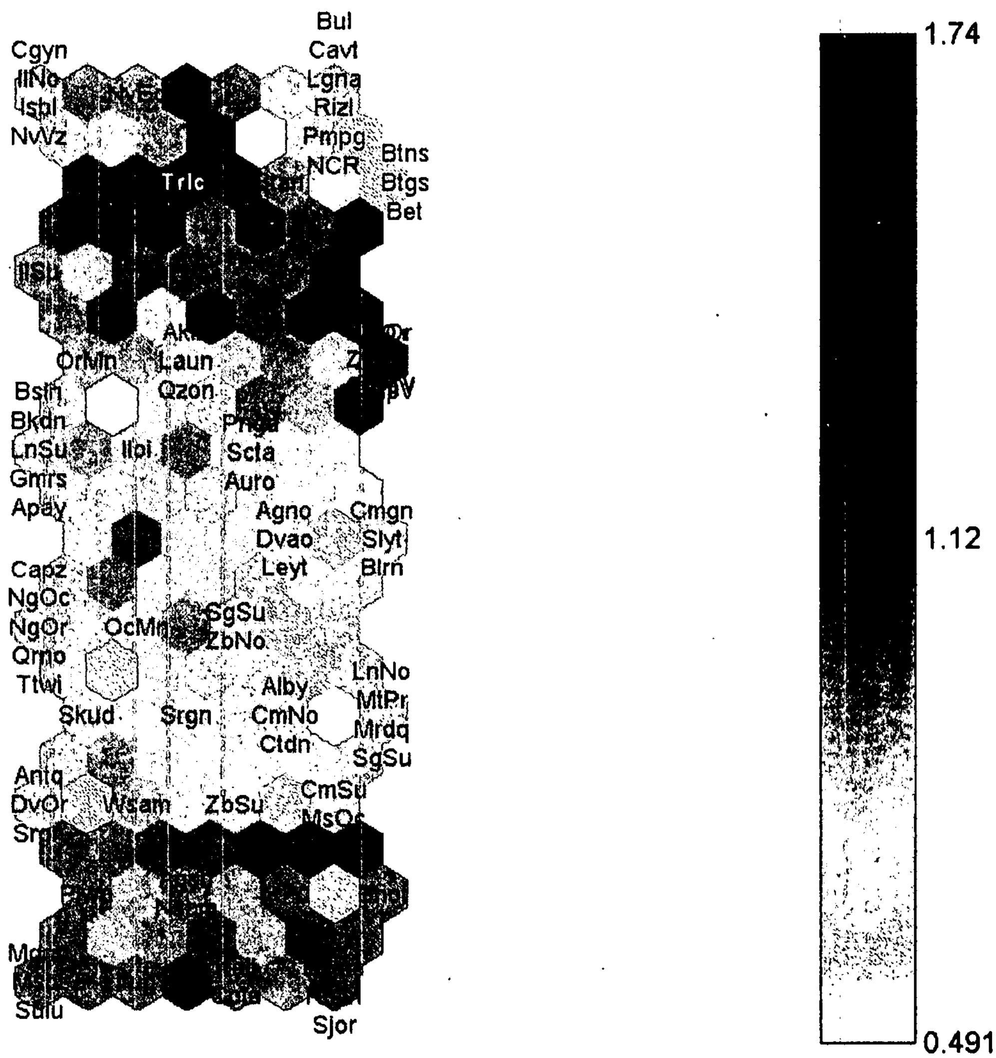

Figure 7:Provincial poverty map for 1999

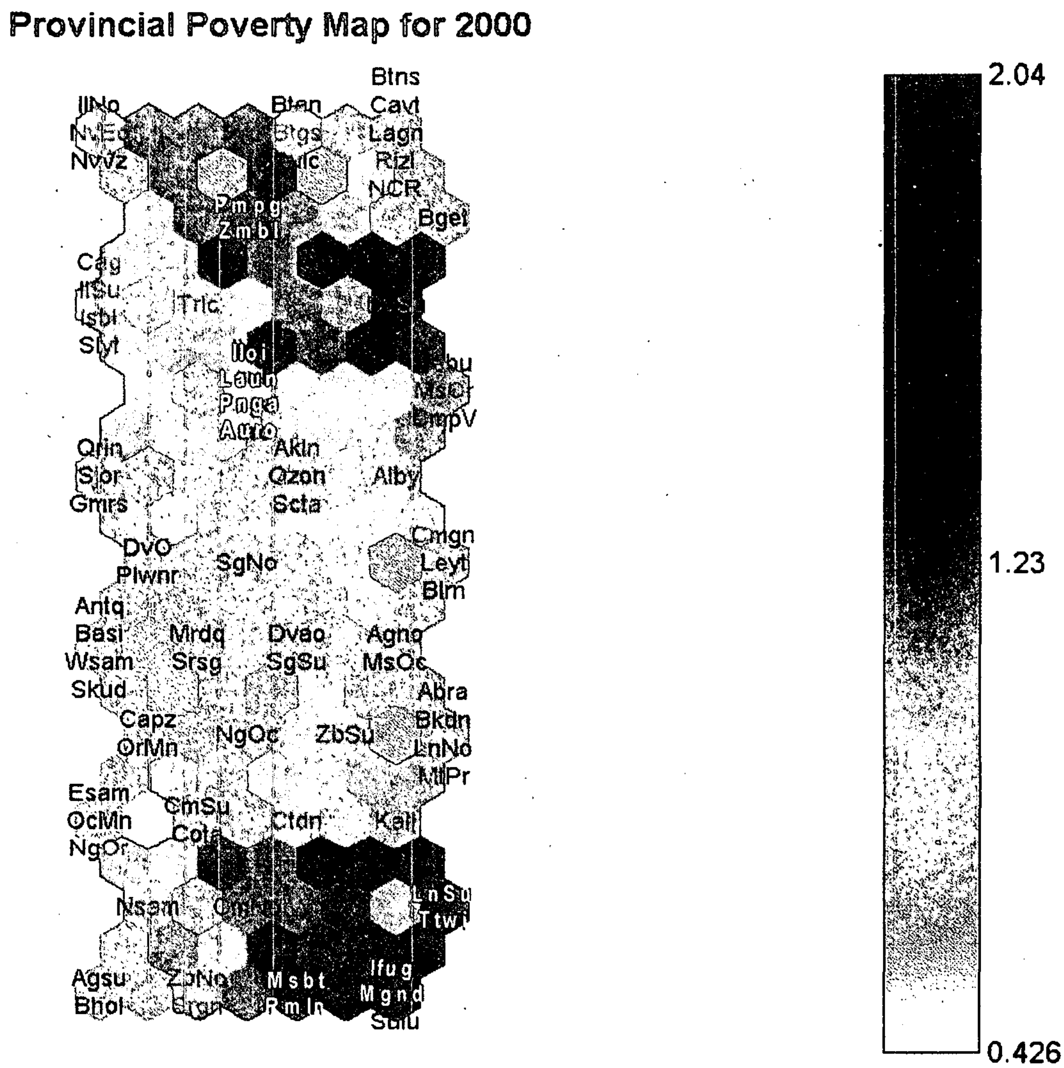

Figure 8:Provincial poverty map for 2000

Table 2. Legend of symbols used in Figures 5.

Abra	Abra	**Ctdn**	Catanduanes	**Nsam**	Northern Samar
Agno	Agusan del Norte	**DvOr**	Davao Oriental	**NvEc**	Nueva Ecija
Agsu	Agusan del Sur	**DvSu**	Davao del Sur	**NvVz**	Nueva Vizcaya
Akln	Aklan	**Dvao**	Davao	**OcMn**	Occidental Mindoro
Alby	Albay	**Esam**	Eastern Samar	**OrMn**	Oriental Mindoro
Antq	Antique	**Gmrs**	Guimaras	**Plwn**	Palawan
Apay	Apayao	**Ifug**	Ifugao	**Pmpg**	Pampanga
Auro	Aurora	**IlNo**	Ilocos Norte	**Pnga**	Pangasinan
Basi	Basilan	**IlSu**	Ilocos Sur	**Qrno**	Quirino
Bget	Benguet	**Iloi**	Iloilo	**Qzon**	Quezon
Bhol	Bohol	**Isbl**	Isabela	**Rizl**	Rizal
Bkdn	Bukidnon	**Kali**	Kalinga	**Rmln**	Romblon
Blrn	Biliran	**Laun**	La Union	**Scta**	South Cotabato
Btan	Bataan	**Leyt**	Leyte	**SgNo**	Surigao del Norte
Btgs	Batangas	**Lgna**	Laguna	**SgSu**	Surigao del Sur
Btns	Batanes	**LnNo**	Lanao del Norte	**Sjor**	Siquijor
Bulc	Bulacan	**LnSu**	Lanao del Sur	**Skud**	Sultan Kudarat
Cagy	Cagayan	**NCR**	Metro Manila	**Slyt**	Southern Leyte
Capz	Capiz	**Mgnd**	Maguindanao	**Srsg**	Sorsogon
Cavt	Cavite	**Mrdq**	Marinduque	**Srgn**	Saranggani
Cebu	Cebu	**MsOc**	Misamis Occidental	**Sulu**	Sulu
CmNo	Camarines Norte	**MsOr**	Misamis Oriental	**Trlc**	Tarlac
CmSu	Camarines Sur	**Msbt**	Masbate	**Ttwi**	Tawi tawi
Cmgn	Camiguin	**MtPr**	Mountain Province	**Wsam**	Western Samar
CmpV	Compostella Valley	**NgOC**	Negros Occidental	**ZbNo**	Zamboanga del Norte
Cota	Cotabato	**NgOr**	Negros Oriental	**ZbSu**	Zamboanga del Sur

A scrutiny of Figures 5-8 shows again the discrepancy between mostly the provinces of Luzon as against provinces of Mindanao, the Visayas, and the Bicol region. In particular, for 1997, we find two concentrations of dark shades in the upper portion and the lower portion that forms at least three clusters of provinces. The upper portion of the map consisting of NCR, Rizal, Laguna, Cavite, Batanes, Bataan, Batangas, Pampanga, Ilocos Norte, Ilocos Sur, Nueva Ecija, Tarlac, Benguet, Cagayan, Isabela, Nueva Vizcaya, Quirino, Cebu, Davao Sur, Misamis Oriental and Zambales, which mostly are provinces of Luzon, is the group with the best welfare conditions. The provinces in the bottom portion of the map, viz., Sulu, Lanao del Sur, Bukidnon, Kalinga, Davao Oriental, Eastern Samar, Ifugao, Abra, Masbate, Mountain Province, Cotabato, Northern Samar, Agusan del Sur, Romblon, Maguindanao, Zamboanga del Norte, Sorsogon, Bohol, is the group with the worst welfare conditions. These provinces are mostly in Mindanao, with a few from Visayas, Bicol and the CAR.

Comparing maps from one year to another, we see a few provinces, e.g., Tawi-tawi, moving into a worse poverty status from 1997 to 2000. Most of the worst off provinces in 1997 continue to be in the worst status (or within a neighborhood of the worst) in 2000.

Provinces with the best welfare status in 1997 by and large also continue to have the best welfare status in 2000.

Year-to-year variations in the Self-Organizing provincial poverty maps appear to be a combination of the impact and recovery from the twin macro-economic crises – the Asian financial crisis and the El Nino phenomenon -- that struck the Philippines between middle 1997 to early 1998, -- and sampling variation. As was earlier mentioned, since the two surveys are, strictly speaking, not comparable, crisis effects (and subsequent recovery) and sampling errors are also confounded with measurement issues.

Clearly, the numerous poverty measures and indicators that can be sourced from the FIES and APIS have, however, been effectively summarized by the resulting Self-Organizing poverty maps both at the regional and provincial levels of disaggregation. The constructed .poverty maps are a powerful tool for displaying information that can be meaningful and comprehensible, and the provide a means of comparison, especially as far as clusters, patterns and spatial trends are concerned.

By and large, the poverty maps, both at the regional and provincial levels provide us a broad sense of the appropriate public actions and policy orientations that need to be undertaken. We continue to see from the provincial poverty maps that poverty in the Philippines is largely a rural phenomenon and that Mindanao, particularly ARMM, needs special attention. Poverty reduction programs that aim to enhance income-generating capabilities must thus be targeted toward the rural areas. Rural industrialization, agricultural modernization and the creation of jobs in the countryside ought thus to be a top priority. Many government programs toward improving the plight of the poor, including perhaps packages for improving micro-credit access, improving access to social services and enhancing the capacity of families to have their desired family sizes would have to be directed to residents outside of Metro Manila and urban areas.

Since access of the poor to basic social services (e.g. education, nutrition and health) as well as to credit is rather limited, part of effectively targeting poverty alleviation program beneficiaries means identifying specific areas where the poor, especially the ultra poor, are located. Poverty maps could be used for geographic targeting in order to direct resources specifically geographic areas that are identified as poor (see Baker and Grosh, 1994). For such geographic targeting, policy planners are undoubtedly in need of information beyond the regional and provincial levels. A study done by Ravallion in Bangladesh and Indonesia, shows that gains from geographical targeting at the regional level are rather small (Bigman and Fofack, 2000). If data are thus available at very small areas, such as municipalities and even *barangays*, then a very detailed poverty map can be set up and resources can be channeled to the needy areas. However, one has to take note of the accuracy and comparability of such poverty data. Sample surveys are not designed to provide reliable poverty statistics for such levels of aggregation. Small area statistics would have to be generated in conjunction with data from censuses and/or reliable administrative records. One may want to establish benchmarks using small area estimation methods to identify the very poorest areas. Alternatively, one may want to go about identifying the poorest regions, then the poorest provinces within these poor regions, then the poorest towns within these poor provinces, then the poorest barangays within these poor towns. Monitoring could then be done only for these targeted areas and for a few control areas, i.e., poverty monitoring need not be actually done for all areas given the cost of generating information. Work relative to this idea of project targeting may, for instance, be performed on the KALAHI project where selected municipalities will be provided interventions, and control municipalities will be selected. Kohonen maps based on small area statistics may help in selecting the control municipalities that will not be provided interventions but are otherwise comparable to the targeted municipalities. The comparison between treatment versus control groups would then provide a protocol for establishing project impact.

The choice of poverty indicators may, of course, change the poverty maps generated especially if the indicators are not highly correlated with each other. Poverty indicators that could be selected for constructing poverty maps may have to be focused on the specific program intervention to be implemented to recipient regions/provinces/communities. That is, if the form of government intervention is in the area of health, then the statistical indicators that ought to be included in the poverty profiling should be health-related.

We wish to issue a caveat that any constructed poverty maps will naturally be sensitive to the quality of data generated That is, the values of the poverty indicators used for poverty maps ought to be of high reliability in order for the constructed maps to be truly meaningful and useful.

ACKNOWLEDGEMENT

The authors wish to thank the National Statistics Office for data support, the National Statistical Coordination Board and the Statistical Research and Training Center for research support, as well as the Asia Pacific Center for Research and SAS Institute, Philippines for allowing access to SPSS Clementine and SAS Enterprise Miner, respectively. Special thanks to the International Statistical Institute (ISI) and the Philippine Social Science Center for support in presenting this paper before the 54th Session of the ISI. Thanks also to Dr. Joselito Magadia, Ms. Ofelia Templo and the anonymous referee for several constructive and substantive suggestions for improving earlier drafts of this paper.

References

BALTAGI, B. H. (2001). *Econometric Analysis of Panel Data.* 2nd Edition. Chichester: John Wiley.

BAKER, J. L. and GROSH, M.E. (1994). "Measuring the Effects of Geographic Targeting on Poverty Reduction", World Bank Living Standards Measurement Study Working Paper 99.

BIGMAN, D. and FOFACK, H. (2000). "Geographical Targeting for Poverty Alleviation: An Introduction to the Special Issue", *The World Bank Economic Review* 14 (1): 129-145.

BIGMAN, D., DERCON, S., GUILLAUME, D. and LAMBOTTE, M.. (2000). "Community Targeting for Poverty Reduction in Burkina Faso", *The World Bank Economic Review* 14 (1):167-193.

BISHOP, C. (1995). *Neural Networks for Pattern Recognition.* Oxford: Oxford University Press.

EVERITT, B. (2001). *Cluster Analysis.* London: Edward Arnold.

DAVID, I. P. (2000). "Poverty Statistics and Indicators: How Often Should They be Measured?". Unpublished manuscript.

DEATON, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: John Hopkins University Press.

HAYKIN, S. (1999). *Neural Networks: A Comprehensive Foundation*, 2nd edition. New Jersey: Prentice-Hall.

INTAL, P. (1994). "The State of Poverty in the Philippines: An Overview". In *Understanding Poverty and Inequity in the Philippines* (Intal, P. S. and Bantilan, M. C. S, eds.) Pasig: National Economic and Development Authority.

JURADO, G. M. (2002). "What Have We Learned from Growth Models". Lecture Paper for the Philippine Institute for Development Studies-Philippine Economic Society Distinguished Speakers' Lecture Series.

KASKI S. AND KOHONEN, T. (1996). "Exploratory Data Analysis by the self-organizing map: Structures of welfare and poverty in the world". In A.-P. N. Referenes, Y. Abu-Mustafa, J. Moody and A. Weigend. (eds). *Neural Networks in Financial Engineering*, pp. 498-507. Singapore: World Scientific.

KOHONEN, T. (2001). *Self-organizing Maps*, 3rd edition. Berlin: Springer-Verlag.

KOHONEN, T., HYNNINEN, J., KANGAS, J., LAAKSONEN, J. (1996). SOM_PAK: "The self-organizing map program package". Report A31. Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland.

KRAAIJVELD, M. A., MAO, J., and JAIN, A. K. (1992). Proceedings of the International Conference on Pattern Recognition. IEEE Computing Society Press: Los Alamos, CA.

OBERMAYER, K. and SEJNOWSKI, T. (eds.) (2001). *Self Organizing Map Formation: Foundations of Neural Computation.* Massachussetts: MIT Press.

REYES, C. (2002). "The Poverty Fight: Have We Made an Impact?" Philippine Institute for Development Studies Symposium Series on Perspective Papers.

RIPLEY, B. D. (1998). *Pattern Recognition and Neural Networks.* New York: Cambridge University Press.

TABUNDA, A. L. and ALBERT, J. R. G. (2002). "Philippine Poverty in the Wake of the Asian Financial Crisis and El Nino" in *Impact of East Asian Financial Crisis Revisited.* Manila: World Bank Institute-Philippine Institute for Development Studies.

ULTSCH, A., SIEMON, H. (1989). Technical Report 329. Univ. of Dortmund, Dortmund, Germany.

VESANTO, J. HIMBERG, J. ALHONIEMI, E. PARHANKANAS, J. (2000). SOM Toolbox for Mathlab 5. Technical Report. Helsinki University of Technology.

WIELENGA, D., LUCAS, B., and GEORGE, J. (1999). Enterprise Miner: Applying Data Mining Techniques Course Notes. Cary, NC: SAS Institute Inc.

WOOLDRIDGE, J. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

End Notes

[i] Following Haykin (1999) and the signal-flow graph of a neuron illustrated in Figure App-1, an artificial neural network maybe defined as a "directed graph consisting of nodes with interconnecting synaptic and activation links, and is characterized by four properties:

a) Each neuron is represented by a set of linear synaptic links, an externally applied bias, an a possibly nonlinear activation link. The bias is represented by a synaptic link connected to an input fixed at +1.
b) The synaptic links of a neuron weight their respective input signals.
c) The weighted sum of a the input signals defines the induced local field of the neuron in question.
d) The activation link squashes the induced local field of the neuron to produce an output"

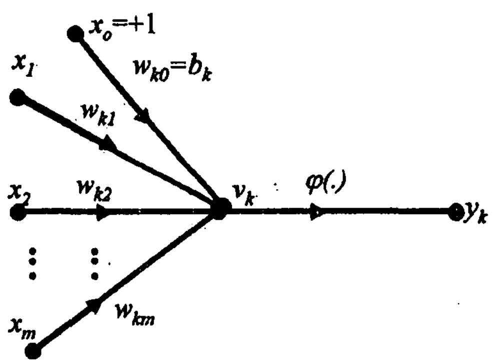

Figure App-1:Signal-Flow Graph of a Neuron

[ii] While random initialization of the model vectors shows that there exists a strong self-organizing tendency in the use of the Self Organizing Map algorithm, this need not be demonstrated every time. Typically, default software implementation as in the SOM toolbox add-on of Matlab, selects the model vectors as a regular array of vectorial values that lie on the subspace spanned by the eigenvectors corresponding to the two largest principal components of the input data. This has been shown to be increase considerably the speed of computation of the SOM.

[iii] Implementations of the SOM may include missing components of a database, which are simply excluded from the distance calculation. However, only a small portion of the data ought to be missing, in which case the remaining non-missing components contain information on the self organization. Weights may also be associated so that

$$\|\boldsymbol{m}-\boldsymbol{x}\|^2 = \sum_{k\in K} w_k (m_k - x_k)^2$$

[iv] $\min_i \|\boldsymbol{m}-\boldsymbol{x}_i\|^2 = \min_i \left(\|\boldsymbol{m}\|^2 + \|\boldsymbol{x}_i\|^2 - 2\boldsymbol{m}^T\boldsymbol{x}_i\right) = \max_i \left(\boldsymbol{m}^T\boldsymbol{x}_i\right)$ if $\|\boldsymbol{x}_i\|$ is constant for all i.

[v] The poverty headcount measure represents the percentage of poor persons in an area. A person is considered poor if he/she belongs to a poor family. A family is poor if its per capita income falls below the poverty line. In technical terms, if w_i and Y_i respectively represent the number of persons in and

per capita income of household i, and z represents a defined poverty line (below which a family is considered poor), the poverty headcount measure is:

$$\frac{1}{n}\sum_{i=1}^{n} w_i I(Y_i < z)$$

where $I(Y_i < z)$ takes the value 1 if the household per capita income Y_i is less than the poverty line z, and 0 otherwise. Poverty lines used here were those officially defined for urban/rural areas in each region by a Technical Working Group of the National Statistical Coordination Board.

[vi] A weakness of the poverty headcount measure is that it does take into account the intensity of poverty, i.e. it does not incorporate how much per capita income is needed for a poor household to exit out of poverty. A useful poverty statistic that adds up the extent to which persons fall below the poverty line (if they do so), is the poverty gap measure defined by

$$\frac{1}{n}\sum_{i=1}^{n} w_i G_i$$

where $G_i = \left(\frac{z - Y_i}{z}\right) I(Y_i < z)$ are the poverty gap ratios.

[vii] The severity of poverty, which is captured neither by the poverty headcount nor the poverty gap measure, is defined as:

$$\frac{1}{n}\sum_{i=1}^{n} w_i (G_i)^2$$

The poverty severity measure is also called the poverty squared gap measure.

[viii] The poverty line consists of both food and non-food requirements. Typically, to generate the poverty line, a food poverty line is firstly generated. If the food poverty line is used as the poverty line, we can generate subsistence or food-poor statistics, such as the food-poor headcount measure, the food-poor gap measure and the food-poor severity measure.

[ix] Not available in APIS 1998 and APIS 1999.

[x] Not available in APIS 1998.

[xi] The U-matrix is a widely used distance matrix technique that is used to show the cluster structure of the SOM. According to Kohonen (2001), this was developed by Ultsch and Siemon (1989) and Kraaijveld et al. (1992). The U-matrix shows distances between neighboring units, and it is thus closely related to single linkage clustering techniques. Note that the U-matrix visualization provides more hexagons than the component planes. This arises since distances between map units are shown in the U-matrix, and not only the distance values at the map units.

[xii] Even a linear projection onto the two dimensional linear subspace of the 1997 poverty database obtained through principal components analysis shown in Fig. App-2 and an application of classical Metric Multidimensional Scaling shown in Fig. App-3, show the regional disparities in the country.

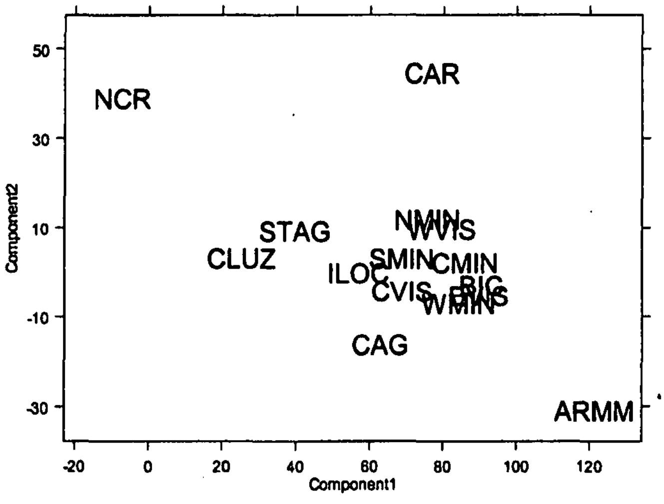

Figure App-2:First two principal components of 1997 poverty indicators

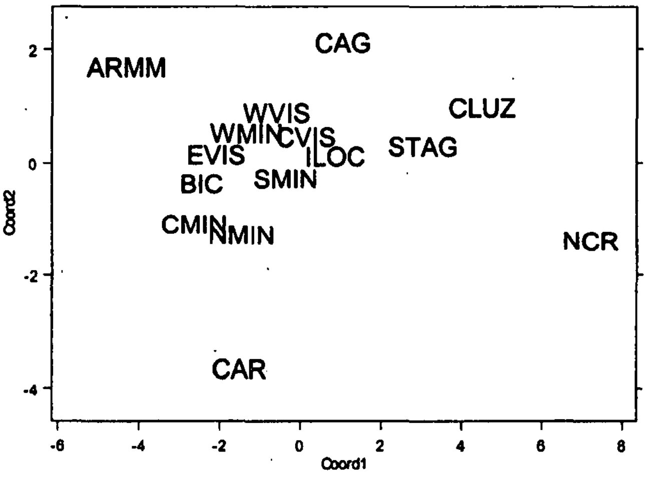

Figure App-3:Graph resulting from applying Multidimensional Scaling on 1997 poverty database

There are some slight differences in this Figure with the results generated from the Self-Organizing Map. Unlike in the Self Organizing Map, CAR and ARMM are shown in these figures as much more separate from the other regions. Also, Cagayan is not in the boundary of the "Northern" and "Southern" clusters as in the Self-Organizing Map.

[xiii] We included the National Capital Region (NCR) also as part of the provincial disaggregation (to serve as benchmark for the provinces) although technically, the NCR is a region, not a province.