

Living Standards of Vietnamese Provinces: a Kohonen Map

Phong Nguyen

General Statistics Office, Vietnam

Dominique Haughton

Bentley University, USA and Toulouse School of Economics, France

Irene Hudson

University of South Australia, Australia

The measurement of living standards is widely recognized to be a multivariate challenge. In this paper, problems with the use of typical univariate indicators are outlined, and a novel approach is suggested which relies on Kohonen maps. The stability and accuracy of the map are evaluated via a bootstrap methodology. This paper presents an application of the technique of Kohonen maps in the context of a data set of indicators about Vietnamese provinces. It is suitable for readers with an intermediate level of statistics.

Keywords: *Kohonen maps, Self-organizing maps, Vietnam living standards*

1. Introduction

In Viet Nam, where provinces have been competing with each other in the area of economic development and fast poverty reduction, government leaders, policy makers and managers as well as researchers usually ask the question: "Which province is better off?" Efforts in answering this question lead to another question: "Which indicator should be used to rank provinces?". Traditionally, one used single indicators such as the Gross Domestic Product (GDP), the household income per capita, or the poverty rate in order to rank provinces. The report "National Human Development Report 2001: Doi Moi and Human Development in Viet Nam" (NHDR 2001) issued under the leadership and coordination of the National Center for Social Sciences and Humanities with the sup-

port of UNDP (United Nations Development Program) (National Political Publishing House, 2001) proposed a ranking of the 61 provinces of Viet Nam from most developed (1st position) to least developed (61st position) on the basis of the HDI – the Human Development Index.

The four indicators we have just mentioned, both single and composite, give different rankings of provinces. As an illustration, focusing on the GDP per capita in 1999, the household income per capita in 2002, the poverty rate in 2002, and the HDI in 1999, the rankings of four specific provinces are displayed in Table 1 below. Ha Noi is the capital city of the country and one of its most de-

veloped cities. Ho Chi Minh city is the largest and the most developed city in Viet Nam. Ba Ria Vung Tau is a province with natural oil and attractive beaches popular with tourists. Binh Duong is a new industrial province.

Table 1. Rankings of provinces according to different indicators

	GDP per capita 1999	Household Income per Capita 2002	Poverty Rate 2002	HDI 1999
Ha Noi	3	2	5	2
Ho Chi Minh City	2	1	1	3
Ba Ria Vung Tau	1	3	7	1
Binh Duong	4	6	2	6

In this case the question arises of which indicator should be used. In Viet Nam the HDI seems to be the preferred indicator in ranking provinces. Experts in government circles argue that because living standards are multidimensional, a composite indicator such as the HDI should be more suitable than any single indicator.

The HDI was developed and used by UNDP in ranking countries in terms of levels of human development. The HDI measures average achievements in a country in three basic dimensions of human development: (i) a long and healthy life, as measured by life expectancy at birth; (ii) knowledge, as measured by the adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight); and (iii) a decent standard of living, as measured by the GDP per capita (in PPP US \$).

HDI rankings in the report were felt in Viet Nam to be reasonable for Ha Noi (2nd position), HCM city (3rd position), and Binh Duong (6th position), but the first position allocated to Ba Ria Vung Tau by the HDI was considered by many to be very counter-intuitive. It was felt that the high GDP per-capita of Ba Ria Vung Tau, mostly from natural oil, dominated the HDI value for the province. As a result, this province was not mentioned in the analysis part of the report, leading among other things to dissatisfaction in several quarters.

In this paper, we propose to use the technique of Kohonen maps (Kohonen, 2001) as an alternative to this state of affairs. Kohonen maps (also known as Self Organizing Maps) have been used in the area of living standards and poverty analysis. Albert et al. (2003) used a Kohonen map with 15 poverty indicators to identify poor provinces in the Philippines for government poverty intervention. Kaski et Kohonen (1996) used 39 welfare indicators and a Kohonen map to compare the economic level and the

standard of living of different countries. Ponthieux and Cottrell (2001) used the Kohonen algorithm to combine different measures of living conditions and to classify households by their level of living conditions as well as differences within similar levels. We also refer the reader to Deichman et al. (2007) for a study of the international digital divide which relies on Kohonen maps and includes a fairly extensive discussion of the Kohonen methodology.

In this past work, however, no method is proposed or applied for validating the reliability of the Kohonen map. In this paper we will use a bootstrap approach and statistical tools to assess the reliability of self-organizing maps, following ideas in De Bodt et al (2002).

2. Kohonen map methodology: a brief introduction

Kohonen maps, due to T. Kohonen and his research team in Finland, are a special case of a competitive neural network, and are also referred to as Self Organizing Maps (SOMs). A useful introduction to the methodology and a Matlab 6.0 toolbox (which was used in this paper) to build the maps can be found at the Kohonen map site <http://www.cis.hut.fi/projects/somtoolbox/>.

The basic algorithm for constructing Kohonen maps is as follows:

- i. Begin with a grid, typically 2-dimensional, with a vector $m_i(t)$ assigned to each grid position, initially typically randomly, of the same dimension as the number of variables.
- ii. For each data vector $x(t)$ find the best match c on the grid such as:
For every i , $|x(t) - m_c(t)| \leq |x(t) - m_i(t)|$
- iii. Update the vectors $m_i(t)$ as follows:
$$m_i(t+1) = m_i(t) + h_{c,i}(t)(x(t) - m_i(t)),$$

where $h_{ij}(t)$ is the neighborhood function, a function of t and of the geometric distance on the lattice between position i and position j . Typically $h_{ij} \rightarrow 0$ as the distance between i and j increases and as more iterations are performed.

- iv. Iterate this step over all available data vectors, and repeat until little change is observed in the $m_i(t)$.

The resulting map tends to organize the components of the estimated vectors, the $m_i(t)$, in a monotonic way (increasing or decreasing) as one moves on the map, hence the term Self Organizing Maps.

Table 2. Indicators used in the Kohonen map

Wealth of household	Education level of household	Health of household
Gdpindex99: GDP Index normalized between 0 and 1, and truncated (ppp 1999 values), UN index	Eduindex99: literacy rate (2/3) and combined enrollments rates (1/3), normalized between 0 and 1, UN composite index	Lifexpindex99: Life expectancy normalized between 0 and 1 (1999 values), UN index
Hpiindex99: UN composite poverty index (1999)	Adlircy: Adult literacy rate	Malnuunder598: Under five malnutrition rate
Gdppc99vnd: GDP per cap. in 99 VND	Percvocswc: % with vocational sec. school diplomas	Lifexpmale: Life expectancy (male)
Mincpcurpric: Avge income per cap. (current prices)	Percunivcoll: % with university and college diplomas	Lifexpfem: Life expectancy (female)
Percexpfood: % exp. allocated to food	Percmsphd: % with Masters and PhD degrees	Wgtforht: Weight for height malnutrition
Valdurgoods: Avge value of durable goods	Schlen3to5: School enrollment rate ages 3-5	Htforage: Height for age malnutrition
Percpermhouse: % permanent houses	Schlenlowsec: Lower secondary school enrollment rate	Wgtforage: Weight for age malnutrition
Povrate: Poverty rate	Schlenupse: Upper secondary school enrollment rate	Matdeathp1000: Number of pregnancy related deaths per 1000
		Infmort: Infant mortality rate

Hhsiz: Number of members of household

Source: NHDR 2001 (National Political Publishing House 2001), and Figures on Social Development, Doi Moi in Vietnam (Statistical Publishing House, 2000).

3. Kohonen maps and Vietnamese provinces

We decided to use the set of 25 variables listed in Table 2 in order to represent with the same number of variables (eight) each area covered by the HDI index, wealth, education and health, with an additional variable on household size.

The map is two-dimensional (8 rows by 5 columns, as chosen by default by the Kohonen Matlab toolkit software) and each of the 40 positions is associated with a 25-dimensional estimated component vector, obtained at convergence of the Kohonen algorithm.

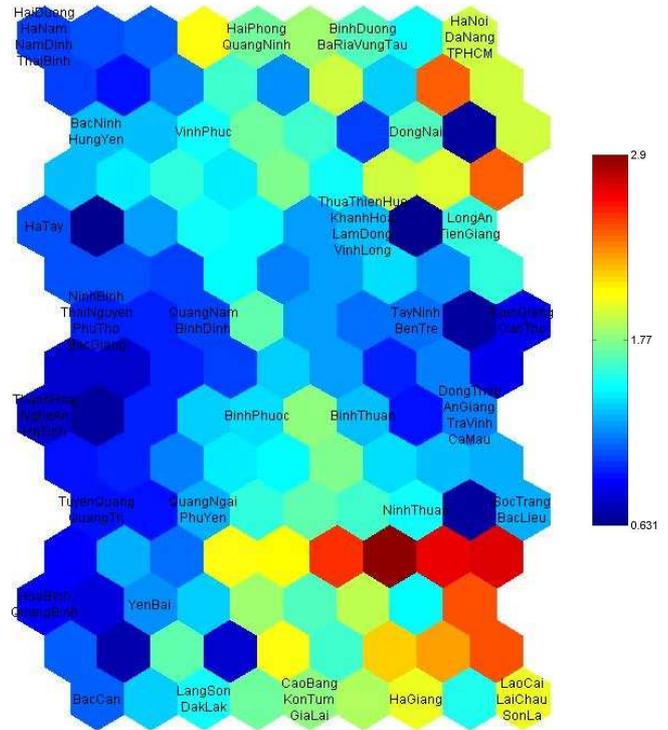


Figure 1. U-matrix of the Kohonen map of Vietnamese provinces on the basis of 25 indicators

Each of the 40 positions is represented by a hexagon on a display referred to as the U-matrix, with additional hexagons added around each actual position hexagon. Hexagons surrounding an actual map position display different colors to represent the distance to other map positions; the color of a position hexagon corresponds to the average distance between this hexagon and its neighbors. For instance, the top right hexagon (HaNoi, DaNang and TPHCM) has an estimated vector which is not too different from that of the hexagon with BinhDuong and BaRiaVungTau (pale blue) and is moderately different from that of its neighbors (pale green). When the map is built, provinces are placed on it according to the estimated vector their data vector is closest to.

We can see from the map that Vung Tau, which is ranked first by HDI, now is clustered by the two-dimensional Kohonen Map with Binh Duong, which is ranked 6th by HDI; this is a more credible result if one takes into account area knowledge about these provinces.

We tried a one-dimensional Kohonen map with the same 25 variables. This map shows that Ba Ria Vung Tau has the third position.

Figure 2 displays the estimated values of each of the 25 components used to build the map. The self-organizing property of the map is clearly visible, and the diagonal directions seem to essentially represent a wealth and health axis, and an education axis. For example, estimated GDP per capita can be seen to decrease from the top-right to the bottom left part of the map, while school achievement variables tend to decrease from the top-left to the bottom-right part of the map. One feature to note is that the position of each of the 40 “map position” hexagons is the same on each component map and on the U-matrix.

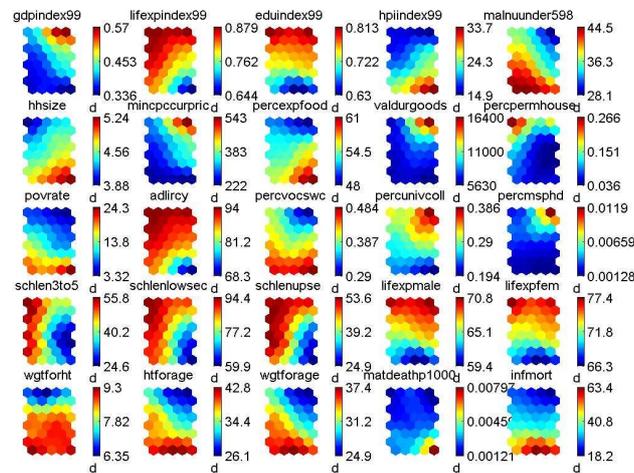


Figure 1. Components of the Kohonen map of Vietnamese provinces (25 indicators)

We will use a bootstrap approach and statistical tools to assess the reliability of self-organizing maps, following ideas proposed by De Bodt et al (2002).

We created 111 samples of size 61, by random selection with replacement from our initial sample of 61 provinces, and with the stipulation that Ba Ria Vung Tau and Binh Duong (the two provinces we will focus on for studying the neighborhood stability of the map) should be part of the samples. Each of the 111 samples is referred to as a bootstrap sample.

We first assess the stability of the quantization in the Kohonen map (which was estimated on the basis of standardized data) using

$$CV(SSIntra) = 100 \frac{\sigma_{SSIntra}}{\mu_{SSIntra}}$$

where SSIntra denotes the sum of squares of quantization errors – that is the squared distance between an observed

data vector x^i and its corresponding (nearest) unit at convergence on the Kohonen map – computed over all observations in each sample, $\mu_{SSIntra}$ denotes the mean of SSIntra errors of all bootstrapped samples, and $\sigma_{SSIntra}$ denotes the standard deviation and the SSIntra errors. A small value of CV implies that the SSIntra values are stable.

For our original Kohonen map, SSIntra (also often referred to as quantization error, as for example in Matlab) is 2.481. For bootstrapped maps $\mu_{SSIntra} = 2.290$, $\sigma_{SSIntra} = 0.125$ and $CV = 0.055$. This implies a good stability of the value of SSIntra. A histogram of the bootstrapped values of SSIntra is displayed in Figure 3.

Second, we assess the stability of the neighborhood relations in the Kohonen map. To do so we assess the stability and its significance of the neighborhood for Binh Duong and Ba Ria Vung Tau as a specific pair of observations.

For assessing the stability of the neighborhood for Binh Duong and Ba Ria Vung Tau, we will calculate:

$$STAB_{ij}(r) = \frac{\sum_{b=1}^B NEIGH_{ij}^b(r)}{B}$$

where $NEIGH_{ij}^b(r) = \begin{cases} 0 & \text{if } x^i \text{ and } x^j \text{ are not neighbors within radius } r \\ 1 & \text{if } x^i \text{ and } x^j \text{ are neighbors within radius } r \end{cases}$

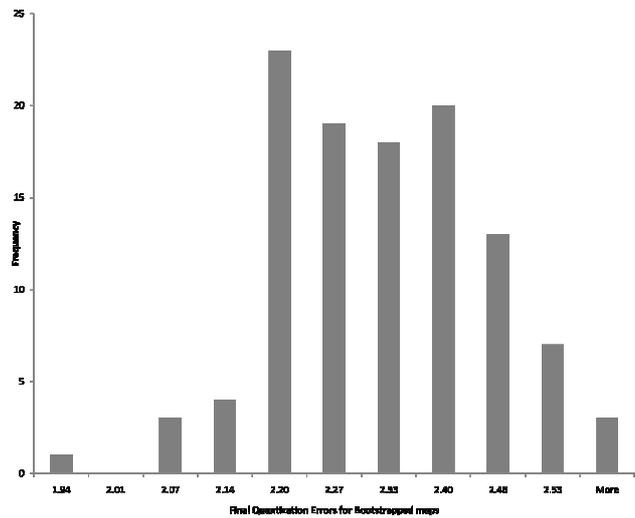


Figure 2. Histogram of quantization errors (SSIntra values) for bootstrapped maps

Being neighbors within radius r means that the two observations are projected on two centroids on the map such that the distance between these centroids is smaller

than or equal to r . If r equals 0, the two observations are projected on the same centroid; if $r = 1$, the two observations are projected on the same centroid or on immediately neighboring centroids. The superscript b denotes bootstrap sample b .

For the pair of Ba Ria Vung Tau and Binh Duong and for $r = 0$, $STAB_{BaRiaVungTauBinhDuong}^{(0)} = 0.4324$. This implies that 43% of bootstrap samples have Ba Ria Vung Tau and Binh Duong projected on the same centroid. For $r = 1$, $STAB_{BaRiaVungTauBinhDuong}^{(1)} = 0.8829$. This implies that 88% of bootstrap samples have Ba Ria Vung Tau and Binh Duong projected on the same centroid or on immediately neighboring centroids.

The significance of the neighborhood for a pair can be evaluated by the standard deviation of the proportions .4324 and .8829 (respectively .0470 and .0305).

4. Conclusion

Our paper used a study of living standards in 61 Vietnamese provinces to demonstrate that the Kohonen map methodology can serve as a better tool to rank provinces, compared to measures such as the Human Development Index (HDI). It makes it possible to take into account a wide range of indicators in the ranking, and to give more sensible results. In general, the technique is also useful for mapping for example geographical areas like provinces into similar clusters onto a two (or one)-dimensional map on the basis of a larger number of socio-economic variables.

Correspondence: dhaughton@bentley.edu

REFERENCES

- Albert, J.R., Elloso, L., Suan, E. & Magtulis, M.A. (2003). Visualizing Regional and Provincial Poverty Structures via the Self-Organizing Map. *The Philippine Statistician*, 52 (1-4), 39-57.
- De Bodt, E., Cottrell, M. & Verleysen, M. (2002). Statistical Tools to Assess the Reliability of Self-organizing Maps. *Neural Networks*, 15, 967-978.
- Deichman, J., Eshghi, A., Houghton, D. Sayek, S. and Woolford, S. (2007). Measuring the international digital divide: an application of Kohonen self-organising maps. *International Journal of Knowledge and Learning*, 3(6), 552-575.
- Kaski, S. & Kohonen, T. (1996). Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World. *Neural Networks in Financial Engineering*, 498-507. Singapore: World Scientific.
- Kohonen, T. (2001). *Self-Organizing Maps*, 3rd edition. Berlin: Springer-Verlag.
- National Political Publishing House (2001). Doi Moi and Human Development in Vietnam. Available at http://hdr.undp.org/docs/reports/national/VIE_Vietnam/Vietnam_en.pdf.
- Ponthieux, S. & M. Cottrell. (2001). Living Conditions: Classification of Households Using the Kohonen Algorithm. *European Journal of Economic and Social Systems*, 15(2), 69-84.
- Statistical Publishing House (2000). Figures on Social Development: Doi Moi Period in Vietnam.