Spatio-temporal Analysis for Small Area Data of Suicide in Japan

Kubota, Takafumi
The Institute of Statistical Mathematics, Risk Analysis Research Center
10-3 Midori-cho, Tachikawa
Tokyo 190-8562, Japan
E-mail: tkubota@ism.ac.jp

Ishioka, Fumio
Okayama University

Tomita, Makoto
Tokyo Medical and Dental University

Fujita, Toshiharu
The Institute of Statistical Mathematics

1. Introduction

Recently a number of suicide case in Japan becomes higher level, the number becomes more than 30,000 per year. To improve this situation for reducing the number of suicide, some results of analysis are needed for identifying where the higher areas are or when the higher time periods are, or what the reasons for suicide are.

Tomita, et al. (2010) and Ishioka, et al. (2010) applied spatial scan statistics for Japanese suicide data (respectively second medical districts (of both male and female) and Kanto districts of male) to investigate spatio-temporal clusters.

Little attention has been given to judge that there are spatial / spatio-temporal auto regressions for these suicide data. Therefore first objective of our study is to find especially spatial auto regression from small areal data of suicide in Japan to identify cumulated area or time. In addition it is said that suicide in Japan is caused from deprivation related to high unemployment rate or loneliness related to high widow rate. For our second objective we also treat census data in Japan to deal in social factors such as marital relation and relation of work by generalized linear model. Finally we discuss spatial-temporal and social relation to suicide data in Japan by results of these analyses.

We introduce the data we use in section 2, and then we show two types of spatial regressions to find spatial relationship between suicide and background social reasons in section 3. We extend spatio-temporal regression in section 4. Finally we summarize the results and future studies in section 5.

2. Data

We use two kinds of data; suicide data and census data. Fujita (2009) has updated the Ministry of Health, Labour and Welfare demographic survey of death "local statistics about suicide" in 2009, and used the second medical district in 2008. From these data we use 354 second medical district in Japan with 3 time periods from 1988 - 1992, 1993 - 1997 and 1998 - 2002 (we denote these time periods as 90, 95 and 00, respectively) for suicide data, because we focus on spatial / spatio-temporal regression. We use the ratio by the number of suicide persons to population (of 10,000) in each area as "suicide rate".

The national census is conducted once every five years in 1990, 1995 and 2000 in the time period from 1988 - 2002. We regard these time periods corresponds of 3 time periods of suicide from 1988 -

1992, 1993 - 1997 and 1998 - 2002, respectively (we also denote these time periods as 90, 95 and 00, respectively). We also use the ratio by the number of widowed persons to population (of 10,000) in each area as "widowed rate", the ratio by the number of unemployed persons to working population (of 10,000) in each area as "unemployed rate".

For both data we use both female and male cases (we denote f and m respectively).

Table 1 shows the example of data (the upper is male case, the lower is female case). In table 1, the variable name consists of first 3 characters of variable, time periods and sex. (For example, "sui00m" means the suicide rate in 1998 to 2002 of male cases.) The second left column is the code number of the second medical district.

Table 1: The example of data

No.	code	sui90m	wid90m	une90m	sui95m	wid95m	une95m	sui00m	wid00m	une00m
1	101	26.30	2707.80	5611.90	27.80	2914.30	6108.80	48.10	3066.50	6539.80
...	...	...	...	...	...	...	...	...	...	...
354	4705	37.60	2563.30	4811.60	38.20	2755.60	5611.40	38.10	2875.70	7490.50
No.	code	sui90f	wid90f	une90f	sui95f	wid95f	une95f	sui00f	wid00f	une00f
1	101	14.90	14337.90	4087.10	12.30	15115.50	5118.70	14.50	15740.80	5752.00
...	...	...	...	...	...	...	...	...	...	...
354	4705	5.20	12757.60	2668.10	6.00	13281.30	3878.70	13.20	13246.00	4820.10

3. Spatial regression

We apply two kinds of neighbor definitions of lags to spatial regression. One is the governmental boundary and the other is the distance between the central points of population. The former is the application of spatial Durbin model which Anselin (1998) proposed. The latter is the extension of spatial autocorreration model which Dormann et al. (2007).

3.1. Spatial Durbin model (SDM)

We apply spatial Durbin model to suicide and census data. The equation (1) shows that $\boldsymbol{y}$ is suicide rate, $\boldsymbol{X}$ consists of widowed rate and unemployed rate and $\boldsymbol{W}$ is spatial weighted matrix that consists of the element of w_{ij} in equation (2).

(1)
$$\begin{aligned} &\boldsymbol{y} = \rho_d \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta}_{1d} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}_{2d} + \boldsymbol{\varepsilon} \\ &\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n) \end{aligned}$$

where $\boldsymbol{\beta}_{1d}$ is regression parameters of the data set of widowed and unemployed rate (of spatial durbin model), $\boldsymbol{\beta}_{1d}$ is spatial lag parameter of the data set of widowed and unemployed rate and ρ_d is spatial lag parameter of the suicide rate. The error term $\boldsymbol{\varepsilon}$ is distributed $N(0, \sigma^2 \boldsymbol{I}_n)$ where $\boldsymbol{I}_n$ is the identity matrix of n x n.

(2)
$$w_{ij} = \begin{cases} 1 & (\text{if } i \text{ is adjacent to } j) \\ 0 & (\text{otherwise}) \end{cases}$$

3.2. Spatial autocorrelation model (SCM)

We extend spatial autocorrelation model to add in not only the spatial lag of dependent variable (suicide rate) but also spatial lags of independent variables (widowed and unemployed rate) as the equation (3).

(3)
$$\boldsymbol{y} = \rho_c \boldsymbol{a} + \boldsymbol{X}\boldsymbol{\beta}_{1c} + \boldsymbol{B}\boldsymbol{\beta}_{2c} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\beta}_{1c}$ is regression parameter of the data set of widowed and unemployed rate (of spatial autocorrelation model), $\boldsymbol{\beta}_{1d}$ is spatial lag parameter of the data set of widowed and unemployed rate and ρ_c is spatial lag parameter of the suicide rate. The elements of autocorrelation $\boldsymbol{a}$ is as follows,

$$a_i = \frac{\sum_{j \in k_i} u_{ij} y_j}{\sum_{j \in k_i} u_{ij}} \tag{4}$$

where k_i is subset of neighborhood at the small area i and u_{ij} is weight term calculated by the distance between the central points of population at the small areas. In this study we chose 400 kilometers to define k_i through trial and error; our target area is approximately 1000 kilometers square in both longitude and latitude direction, and u_{ij} as squared inverse of the distances.

3.3. Results

First of all we apply two models to suicide and census data in the time period 00. Figure 1 shows the choropleth map that the area is colored by the predicted value of the SCM in male cases; the darker green, the higher suicide predicted values (from 28.13 to 68.44).

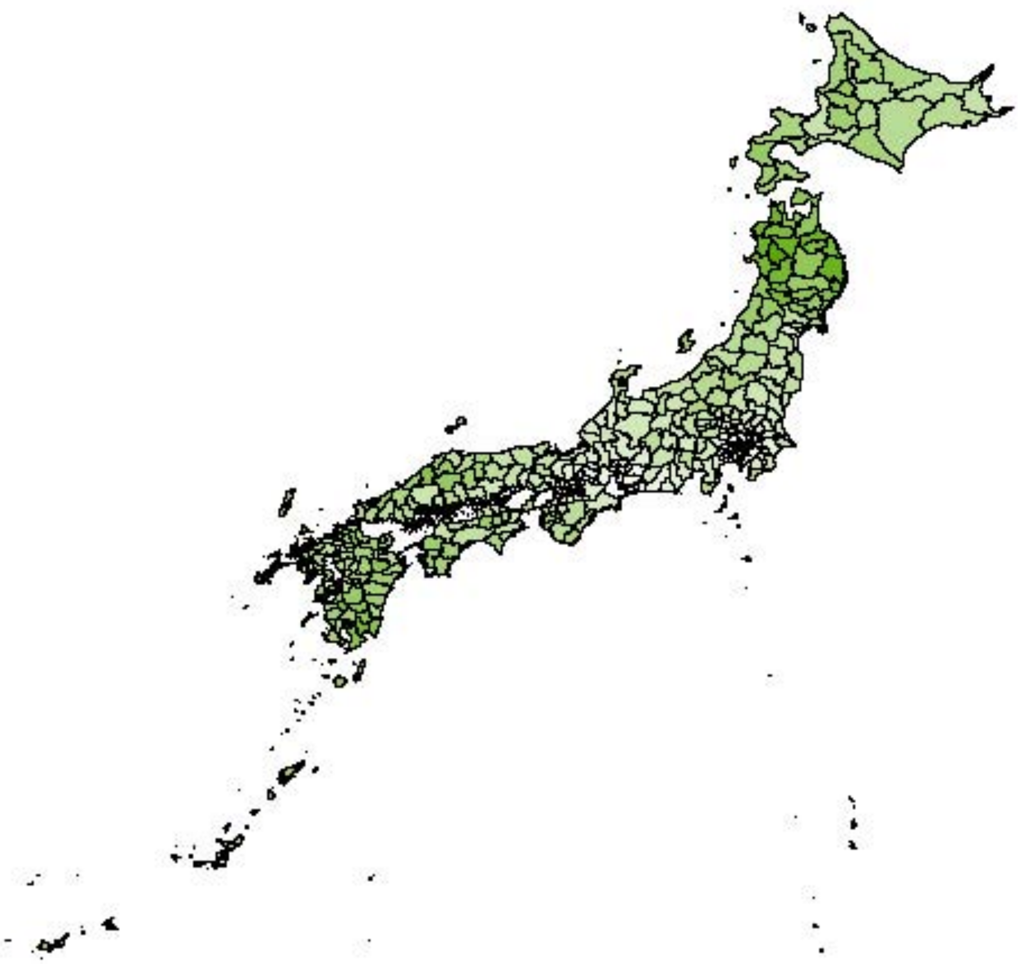

Figure 1: the predicted value of the SCM in male cases at the time period 00

We focus on the Japan Sea and the Seto Inland Sea of the choropleth map of residual values (observed suicide rate - predicted suicide rate); the stronger red shows the higher positive residuals, and vice versa (blue) in Figure 2. The west side of Japan corresponds to the Japan Sea, and the left lower sea area which is cumulated blue area is the Seto Inland Sea.

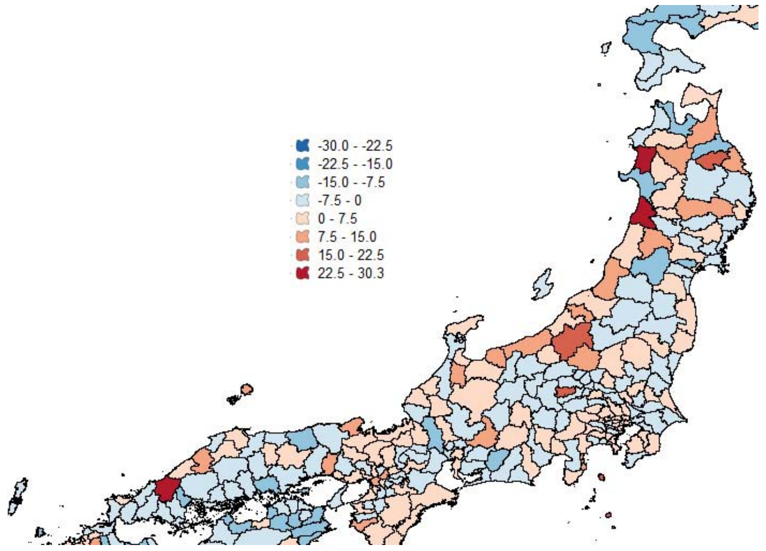

Figure 2: residual values focused on the Japan Sea and the Seto Inland Sea

Then we compare SDM and SCM in the time terms of 90, 95 and 00 in the male cases. Table 2 shows the result of regressions; standardized coefficients of widowed rate (widowed), unemployed rate (unemployed), the spatial lag of suicide rate (lag.suicide), the spatial lag of widowed rate (lag.widowed), the spatial lag of unemployed rate (lag.unemployed) and the coefficient of determinations (adjusted R2). The column names consists of model names (SDM or SCM) and time periods (90, 95 or 00). In the table 2 SCM is little better than SDM in the aspect of adjusted R2. That indicates the

Table 2: Standardized coefficients SDM and SCM of 1990, 1995 and 2000 (Male)

	SDM90	SCM90	SDM95	SCM95	SDM00	SCM00
widowed	0.59	0.56	0.58	0.54	0.50	0.48
unemployed	0.02	-0.02	0.02	-0.02	0.10	0.04
lag.suicide	0.53	1.15	0.61	1.13	0.67	1.16
lag.widowed	-0.31	-0.61	-0.35	-0.60	-0.28	-0.57
lag.unemployed	0.09	-0.03	0.02	-0.06	-0.05	-0.11
adjusted R2	0.58	0.60	0.59	0.58	0.63	0.64

neighbor definition of the weighted distance of central population points is litter better than that of the governmental boundary. Therefore we use only SCM to compare between male and female cases, and spatio-temporal regression hereafter. Table 3 shows the comparison between male and female cases. The row names are as same as table 2, but the column names is added the sex (f or m). The mark of * means the coefficients are statistically significant (p-values is less than 0.05).

4. Spatio-temporal regression

We focus on the spatio-temporal prediction of the time period 00 to use the suicide rate of 95 or to use the autocorrelation of suicide 95. The equation (5) (STRM1) shows the former that is added the one previous term $\gamma_{t-1}\boldsymbol{y}_{t-1}$ ($\boldsymbol{y}_{t-1}$ is the one previous suicide rate and γ_{t-1} is regression parameter of $\boldsymbol{y}_{t-1}$) from the equation (1). The equation (6) (STCM1) shows the latter that is added the one previous spatial autocorrelation $\rho_{t-1}\boldsymbol{a}_{t-1}$ ($\boldsymbol{a}_{t-1}$ is the one previous autocorrelation of suicide (which elements is defined in the equation (4)) and ρ_{t-1} is parameter of $\boldsymbol{a}_{t-1}$) from the equation (3).

Table 3: Standardized coefficients of SCM of 1990, 1995 and 2000 (Male and Female)

	SCM90m	SCM95m	SCM00m	SCM90f	SCM95f	SCM00f
widowed	0.56*	0.54*	0.48*	0.43*	0.35*	0.37*
unemployed	-0.02	-0.02	0.04	-0.21*	-0.24*	-0.10
lag.suicide	1.15*	1.13*	1.16*	1.18*	1.21*	1.22*
lag.widowed	-0.61*	-0.60*	-0.57*	-0.49*	-0.44*	-0.44*
lag.unemployed	-0.03	-0.06	-0.11	0.20*	0.23*	0.09
Adjusted R2	0.60	0.58	0.64	0.52	0.47	0.40

(5) $$\boldsymbol{y}_t = \rho_t \boldsymbol{a}_t + \boldsymbol{X}_t \boldsymbol{\beta}_{1ct} + \boldsymbol{B}_t \boldsymbol{\beta}_{2ct} + \gamma_{t-1} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}$$

(6) $$\boldsymbol{y}_t = \rho_t \boldsymbol{a}_t + \boldsymbol{X}_t \boldsymbol{\beta}_{1ct} + \boldsymbol{B}_t \boldsymbol{\beta}_{2ct} + \rho_{t-1} \boldsymbol{a}_{t-1} + \boldsymbol{\varepsilon}$$

The equations (7) (STRM2) and (8) (STCM2) are added two previous time terms form (5) and (6), respectively.

(7) $$\boldsymbol{y}_t = \rho_t \boldsymbol{a}_t + \boldsymbol{X}_t \boldsymbol{\beta}_{1ct} + \boldsymbol{B}_t \boldsymbol{\beta}_{2ct} + \gamma_{t-1} \boldsymbol{y}_{t-1} + \gamma_{t-2} \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}$$

(8) $$\boldsymbol{y}_t = \rho_t \boldsymbol{a}_t + \boldsymbol{X}_t \boldsymbol{\beta}_{1ct} + \boldsymbol{B}_t \boldsymbol{\beta}_{2ct} + \rho_{t-1} \boldsymbol{a}_{t-1} + \rho_{t-2} \boldsymbol{a}_{t-2} + \boldsymbol{\varepsilon}$$

Table 4 shows the result of spatio-temporal regressions; Standardized coefficients of widowed rate (wid), unemployed rate (une), the spatial lag of suicide rate (lag.sui), the spatial lag of widowed rate (lag.wid), the spatial lag of unemployed rate (lag.une), the one previous time lag of suicide rate (stlag.95), the two previous time lag of suicide rate (stlag.90), and the coefficient of determinations (adj.R2). The column names consists of model names (STRM1, STCM1, STRM2 or STCM2) and the sex (f or m). The mark of * means as same as table 3.

Table 4: Standardized coefficients of spatio-temporal regressions (Male and Female)

	STRM1m	STCM1m	STRM2m	STCM2m	STRM1f	STCM1f	STRM2f	STCM2f
wid	0.19*	0.48*	0.15*	0.48*	0.12*	0.37*	0.04	0.36*
une	0.07	0.05	0.08*	0.06	0.04	-0.11	0.06	-0.11
lag.sui	0.66*	0.85*	0.62*	0.77*	0.44*	0.83*	0.34*	0.81*
lag.wid	-0.31	-0.64*	-0.32*	-0.76*	-0.18*	-0.48*	-0.11	-0.48*
lag.une	-0.11*	-0.15*	-0.14*	-0.20*	0.02	0.13	0.02	0.14
stlag.90	0.53*	0.25*	0.21*	0.26	0.66*	0.23	0.31*	0.07
stlag.95	-	-	0.40*	0.10	-	-	0.49*	0.19
adj.R2	0.76	0.65	0.77	0.65	0.63	0.41	0.67	0.41

5. Concluding remarks and future studies

Spatial / spatio-temporal autocorrelation existed in all models, time periods and sex (which are shown by third row (lag.suicide) of table 3). In the spatial autocorrelation model of male cases, the widowed rate in all time periods affected suicide rate (by left 3 columns in the first row of table 3) but the unemployed rate did not affect suicide rate (by left 3 columns in the second row of table 3). The choropleth map of residual values in figure 2 showed there is other factors which increase the suicide rate in the area near the Japan Sea and there is other factors which decrease it in the area near the Seto Inland Sea. In the female case, widowed rate affected suicide rate in all time periods (by the right 3 columns in the first row of table 3). The unemployed rate in 00 was not statistically significant but those in 90 and 95 negatively affected suicide rate (by the right 3 columns in the second row of table 3). We have to notice that unemployed rate is the ratio from working population; that does not include non-workers such as married woman, students and so on. The comparison of coefficient of determinations between female and male cases, the male cases had higher. Therefore that suggested the suicide rate of female have other background social or financial relationships.

In the spatio-temporal autocorrelation models, STRMs are better models than STDMs in both male and female cases (by the bottom rows between even and odd columns of table 4). That suggested the relations from just direct previous one or two time periods of suicide rate are effective than those of the correlation of previous time and space. The models using previous two time periods are better than one time period, which suggest two time periods model might explain hidden previous relationship. The result that the male cases are better than female cases suggested the same situation of spatial regression. The results that the spatio-temporal models are better than those of spatial models showed the advantage of using not only spatial but also temporal model simultaneously.

Our future studies are to apply other social or financial parameters for these spatio-temporal analyses. We will also discuss to use more detail data in the aspect of age groups and small areas. For complex models we will apply Bayesian approach that Congdon (2000) applied to suicide over London boroughs and time.

Acknowledgement

This is a part of funded research from National Institute of Mental Health, National Center of Neurology and Psychiatry and is also partially supported by KAKENHI 21700305 and KAKENHI 21700317. This research is also partially supported by ISM FY2011 travel support for an international paper/poster presentation by young researchers. (ISM stands for The Institute of Statistical Mathematics in Japan) .

REFERENCES (RÉFERENCES)

Anselin, L. (1998). Spatial Econometrics: Methods and Models, Dordrecht: Kluwer Academic Publishers.

Congdon, P. (2000). Monitoring Suicide Mortality: A Bayesian Approach , European Journal of Population, 16(3), pp.251-284.

Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., Davies, R., Hirzel, A., Jets, W., Kissling, W., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review, Ecography, Vol. 30, pp. 609-628.

Ishioka, F., Tomita, M. and Fujita, T. (2010). Detection of Spatial Cluster for Suicide Data using Echelon Analysis , COMPSTAT2010 Proceedings in Computational Statistics, pp.1159-1166.

Tomita, M., Ishioka, F. and Fujita, F. (2010). Space-time Data Analysis for Suicide data in Japan, Bulletin of the Computational Statistics of Japan, 23(1) (Japanese) pp.25 - 43.

RÉSUMÉ (ABSTRACT)

Nous appliquons Régression pour l'espace-temps pour les données du suicide japonaises. Notre premier but est trouver une corrnélation entre l'espace-temps de donnnées du suicide. Notre deuxiéme but est expliquer les relations de donnnées du suicide et données du recensement du national.