▩ 응용논문

Model for the Spatial Time Series Data

Seongsik Lim
Dept. of Applied Statistics, Seo Kyeong University
Sinsup Cho
Dept. of Computer Science and Statistics, Seoul National University
Changsoo Lee
Dept. of Statistics, Soong Sil University

Abstract

We propose a model which is useful for the analysis of the spatial time series data. The proposed model utilizes the linear dependences across the spatial units as well as over time. Three stage model fitting procedures are suggested and the real data is analyzed.

1. Introduction

Many problems that arise in physical sciences require investigators to work with the data which is observed not only over time but also across the spatial units. For example, in the general marine fisheries contexts, one observes fisheries data at sensors that is distributed in space as well as time, Stoffer (1985). As the time series data is expected to be correlated over time, the data observed contiguously in space are also expected to be correlated across the space. Especially, the spatial time series data provide a rich environment for the statistical analysis because they contain the information necessary to deal with the intertemporal dynamics and the individuality of entries being investigated.

Several models for the analysis of the spatial data or the time series data have been proposed by many authors, see Cliff and Ord (1981) among others, but they are inappropriate for the analysis of the spatial time series data.

For this purpose Pfeifer and Deutsch (1980a, 1980b) suggested a space time ARMA model for the analysis of the Boston assault arrest data weighted by boundary between first-order neighbors which is the typical type of the spatial time series data. However, since too many parameters have to be estimated, an inappropriate model is often fitted and therefore it may result in inferior forecasts if the number of observations are small. Furthermore, the space time ARMA model is not applicable when the parameters are subject to change over time since it assumes the constancy of the parameters.

In this paper we propose a spatial time model which is useful not only for the small sample size but for the time-varying parameters.

2. The Spatial Time Model

For the analysis of acturial data that arise in the insuriance rate making problem Lee (1991) and Ledolter and Lee (1993) suggested a structural time series model which can handle the spatial time series data. But their approaches assume the independence among the spatial units. Hence their approaches are not appropriate when the spatial units are correlated. Adopting their approaches we propose a model which allows the correlation across the spatial units as follows :

$$
\begin{aligned}
y_t^{(i)} &= \mu_t + d_t^{(i)} + \varepsilon_t^{(i)}, \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \gamma_{t-1} + \eta_{t\alpha}, \\
\beta_t &= \beta_{t-1} + \gamma_{t-1} + \eta_{t\beta}, \\
\gamma_t &= \gamma_{t-1} + \eta_{t\gamma}, \\
d_t^{(i)} &= \rho \sum_{j=1}^{N} w_{ij} d_{t-1}^{(j)} + \eta_t^{(i)},
\end{aligned}
\tag{1}
$$

where w_{ij} is a weight denoting the interaction between the ith and jth spatial units but $w_{ii}=1$ and the scalar ρ determine the degree of correlation among the components of $d_t^{(i)}$. The distributions of all error terms, $\varepsilon_t^{(i)}$, η_{tj} and $\xi_t^{(i)}$, $i=1, \cdots, N$, $j=\alpha, \beta, \gamma$ are normally distributed with zero means and variances $r_t^{(i)}\sigma^2$, $q_{tj}\sigma^2$ and $q_{dt}^{(i)}\sigma^2$, respectively. Moreover, the distribution of the initial state vector $\underline{\alpha}_0=(\mu_0, \beta_0, \gamma_0, d_0^{(1)}, \cdots, d_0^{(N)})'$ at time $t=0$ is also assumed to be normally distributed with mean $\underline{\alpha}_0^*$ and variance $\mathbf{P}_{0|0}\sigma^2$. All error terms are assumed to be independent and independent of $\underline{\alpha}_0$.

In the above model (1), the state μ_t, β_t and γ_t represent only the time effects, whereas $d_t^{(i)}$'s represent the spatial effects as well as time effects. If $q_{t\gamma}=0$, the model in (1) is reduced to the local linear model. Also the model without the slope and quadratic components is called the local level model.

The model (1) can be represented as the multivariate state space form

$$
\begin{aligned}
\underline{y}_t &= \mathbf{H}\,\underline{\alpha}_t + \underline{\varepsilon}_t, \\
\underline{\alpha}_t &= \mathbf{T}\,\underline{\alpha}_{t-1} + \underline{\delta}_t,
\end{aligned} \tag{2}
$$

where the $(N\times(N+3))$ matrix $\mathbf{H}=(1\ \underline{0}\ \underline{0} \vdots \mathbf{I})$ is a observation matrix and the $(N+3)\times(N+3)$ transition matrix is

$$
\mathbf{T} = \begin{bmatrix} 1 & 1 & 1 & \underline{0}' \\ 0 & 1 & 1 & \underline{0}' \\ 0 & 0 & 1 & \underline{0}' \\ \underline{0} & \underline{0} & \underline{0} & \rho\mathbf{W} \end{bmatrix}
$$

The sysytem noise vector $\underline{\delta}_t=(\eta_{t\alpha},\ \eta_{t\beta},\ \eta_{t\gamma},\ \xi_t^{(1)},\ \cdots,\ \xi_t^{(N)})'$ is multivariate normally distributed with zero mean vector and covariance matrices $\mathbf{Q}_t$.

We can estimate the parameters of the model (2) through the Kalman filter recursions, Kalman(1960) and Kalman and Bucy (1961). The Kalman filter is a recursive procedure used to estimate the vector of unknown parameters $\underline{\alpha}_t$ in each period of time t given all observations up to and including time t in (2). Let $\underline{\alpha}_{t-1|t-1}$ in (2) denote the minimum mean square linear estimate (MMSLE) of $\underline{\alpha}_{t-1}$ based on the observation up to and including $\underline{y}_{t-1}$ and let $\mathbf{P}_{t-1|t-1}$ denote the $N\times N$ covariance matrix of the estimation error, $\underline{\alpha}_{t-1}-\underline{\alpha}_{t-1|t-1}$. Given $\underline{\alpha}_{t-1|t-1}$ and $\mathbf{P}_{t-1|t-1}$, the optimal estimate of $\underline{\alpha}_t$ is given by

$$
\begin{aligned}
\underline{\alpha}_{t|t-1} &= \mathbf{T}\,\underline{\alpha}_{t-1|t-1}, \\
\mathbf{P}_{t|t-1} &= \mathbf{T}\mathbf{P}_{t-1|t-1}\,\mathbf{T}' + \mathbf{Q}_t.
\end{aligned} \tag{3}
$$

Once the new observation, $\underline{y}_t$, becomes available, the estimate of $\underline{\alpha}_t$ can be updated. The appropriate equations are

$$
\begin{aligned}
\underline{\alpha}_t &= \underline{\alpha}_{t-1} + \mathbf{P}_{t|t-1}\,\mathbf{H}'\mathbf{F}_t^{-1}(\underline{y}_t - \mathbf{H}\,\underline{\alpha}_{t|t-1}), \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\,\mathbf{H}'\mathbf{F}_t^{-1}\,\mathbf{H}\,\mathbf{P}_{t|t-1},
\end{aligned} \tag{4}
$$

where

$$
\mathbf{F}_t = \mathbf{H}\,\mathbf{P}_{t|t-1}\,\mathbf{H}' + \mathbf{R}_t
$$

The equations in (3) are known as the *prediction equations* whereas those in (4) are the *updating equations*. Thus the l-step-ahead forecast of the observation vector $\underline{y}_{t+l}$ is given by

$$\underline{y}_{t+l|t} = \mathbf{H}\,\underline{\alpha}_{t+l|t} \tag{5}$$

and covariance matrix is obtained by

$$\mathbf{F}_{t+l} = \mathbf{H}\,\mathbf{P}_{t+l|t}\mathbf{H}' + \mathbf{R}_{t+l},$$

where $\mathbf{P}_{t+l|t} = \mathbf{T}\,\mathbf{P}_{t+l-1|t}\,\mathbf{T}' + \mathbf{Q}_{t+l}$

To check whether the model is appropriate, we use the following test statistic suggested by Hosking (1980)

$$\begin{aligned} Q_P &= T\sum_{h=1}^{P}[vec\,C_v(h)]'\,C_v(0)^{-1}\otimes C_v(0)^{-1}[vec\,C_v(h)] \\ &= T\sum_{h=1}^{P} tr\,[C_v(h)'C_v(0)^{-1}C_v(0)^{-1}C_v(h)C_v(0)], \end{aligned} \tag{6}$$

where $C_v(h) = \mathrm{T}^{-1}\sum_{t=1}^{T-h}\underline{v}_t\underline{v}'_{t+h}$ the autocovariance matrix at lag h of the residual vector $\underline{v}_t$.

3. Model Building Procedure

For the application of the spatial time model, we suggest the following three stage model building procedures, the identification of the model, the estimation of parameters and the diagnostic checking.

Stage 1. Identification

We have to determine first whether there are dependences across spatial units using the spatial autocorrelation measures due to Cliff and Ord (1981). If the spatial autocorrelation exists, we use the spatial time model(ST model) with non zero off diagonal weight matrix, otherwise, the spatial time model with the identity weight matrix(UST model), i.e., $\mathbf{W}=\mathbf{I}$, is used.

The next step in the model identification stage is to identify the most appropriate form of the model among the class of models we propose. The essential tool is the time series plot of the spatial time series data, since it provides the basis of the model identification. If the plots of the spatial time series data do not reveal any tendency of trend, we use the locally level model with time varying components. Otherwise, we use the appropriate trend model depending on the type of the trends, linear or quadratic.

Stage 2. Estimation

The Kalman filter requires an exact knowledge of the observational noise variance matrix $\mathbf{R}_t$ and the system noise variance matrix $\mathbf{Q}_t$. But in practice, the variance matrices are seldom known and therefore if the model operates with fixed estimates of these variances, the results may cause inferior forecasts. Estimates of variances are obtained from the Kalman filter via the prediction error decomposition, see Schweppe (1965) and Harvey (1990).

Stage 3. Diagnostic Checking

After parameters are estimated, adequacy of the model is assessed by checking whether the model assumptions are satisfied. The model diagnostic checking is in general accomplished through a careful analysis of the residual series. The main diagnostics are based on the residuals obtained by running the Kalman filter initialized with diffuse prior.

The vector residuals may be used to construct a multivariate portmanteau test statistic. The test statistics are proposed by several authors, Hosking (1980), Li and McLeod (1981) and Reinsel (1993). The multivariate Q-statistic in (6) is asymptotically distributed as χ^2 with (N^2P-n^*) degrees of freedom where n^* is the loss of degrees of freedom. If the adequacy of the model is not questioned, we use the forecasts obtained by the selected model.

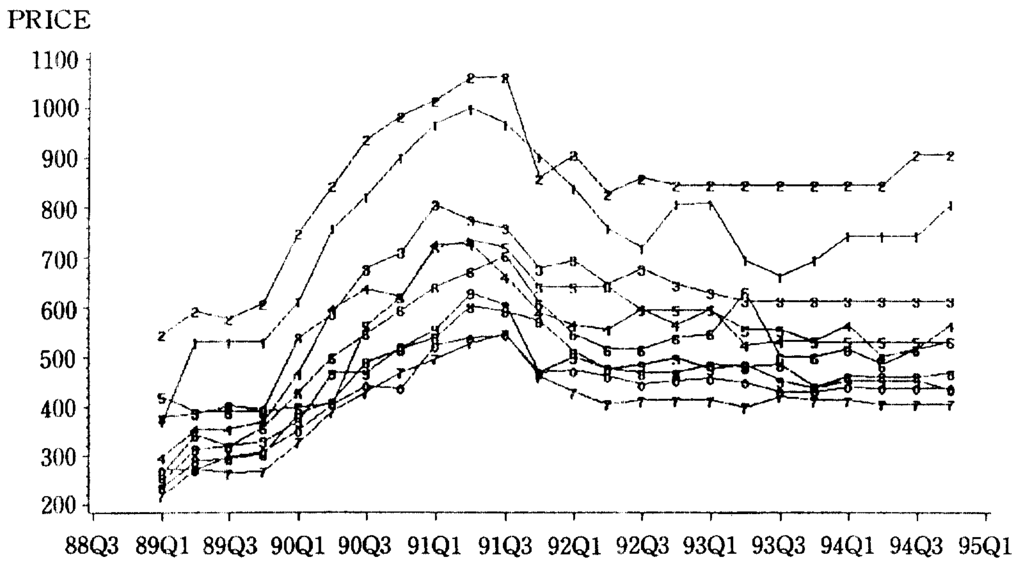
4. Example

To investigate the prediction performances of the proposed model we apply the ST model and the UST model to the real estate data given in the appendix, the quarterly apartment prices per 3.3 square meter in 10 districts of Seoul city from March 1990 to December 1994 published in the Real Estate Bank (1993, 1994). Comparison of the prediction performances is based on MSE's of the one-step ahead predictions for each model. MSE's of one-step ahead predictions are obtained using the Kalman filter recursions. Since the sample size is small and the parameters are time varying, the space time ARMA model by Peifer and Deutsch is not analyzed.

Time series plots of the data are given in ⟨Figure 1⟩ We see that correlations exist among the cross-sections, i.e., the time series plots show the same up and down patterns across time, especially from 4th quarter of 1989 to 2nd quarter of 1992. It should be noted that 4th quarter of 1989 and 3rd quarter of 1993 are associated with the build-up and fade-out phases and 2nd quarter of 1991 is the peak of the apartment price. Around this time, we see very active real estate transactions.

To test the adequacy of the model, we performed χ^2- test with (N^2P-n^*) degrees of freedom. The Portmanteau statistics of UST and ST models when $P=2$ are also given in ⟨Table 1⟩ where $n^*=22$. Since p-values of models are 0.0657 and 0.2604 with $df=178$, both ST and UST models fit the data well. The results of one-setp ahead prediction from quarterly apartment prices are summarized in ⟨Table 1⟩. In ⟨Table 1⟩, we observe that the performance of ST model is superior to UST model in the accuracy of prediction. It is clear from this comparison that when the spatial autocorrelation exists, the use of the model under the assumption of independence is inadequate. Therefore, if spatial autocorrelation exists among contiguous boroughs, it is reasonable to use the ST model with the appropriate weights.

Model useful for the analysis of the spatial time series data that are observed not only over time but also over spatial units was provided in this paper. The proposed model is capable of applying irrespective of the sample size and/or time varying parameters.



⟨ Figure 1 ⟩ Time Series Plots of the Quarterly Apartment Prices in the Seoul Boroughs

〈 Table 1 〉 MSE of One-Step Ahead Prediction of the Spatial Time Model

No	Borough	UST model	ST model
1	KangNam	3388.36	2317.26
2	SeoCho	5956.81	4440.46
3	SongPa	3180.08	1967.05
4	KangDong	4307.68	3463.04
5	DongJak	2068.58	1119.15
6	SeongDong	2176.14	1472.44
7	Dongdaemoon	1024.38	411.16
8	JoongRyang	1123.36	463.94
9	NoWeon	1778.43	936.14
10	DoBong	1145.77	479.18
	mean	2615.01	1706.98
	Q_p	216.98	189.70
	p-value	0.0657	0.2604

References

[1] 아파트 시세 바이블(1993), 서울 : (주)정보성.

[2] 부동산 뱅크(1993, 1994), 서울 : (주)정보성.

[3] Brockwell, P. J. and Davis, R. A. (1981), *Time Series Theory and Method*, Springer-Verlag, New York.

[4] Cliff, A. D. and Ord, J. K. (1981), *Spatial Process: Models Applications*, Pion Limited, London.

[5] Harvey, A. C. (1990), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

[6] Hosking, J. R. M. (1980), "The Multivariate Portmanteau Statistic," *Journal of the American Statistical Association*, Vol. 75, pp. 602-608.

[7] Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transaction of ASME, Journal of Basic Engineerings*, 82D, pp. 35-45.

[8] Kalman, R. E. and Bucy, R. S. (1961), "New Results in Linears Filtering and Prediction Theory," *Transaction of ASME, Journal of Basic Engineerings*, 83D, pp. 95-108.

[9] Lee, C. S. (1991), *Time Series Models for the Credibility Estimation of Insurance Premiums*, Unpublished Pd.D. thesis, University of Iowa.

[10] Li, W. K. and McLeod, A. I. (1981), "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Model," *Journal of the Royal Statistical, Socity, Series B*, pp. 231-239.

[11] Pfeifer, P. E. and Deutsch, S. J. (1980a), "A Three-Stage Iterative Procedure for Space-Time Modeling," *Technometrics*, Vol. 22, pp. 35-47.

[12] Pfeifer, P. E. and Deutsch, S. J. (1980b), "Identification and Interpretation of First Order Space-Time ARMA Models," *Technometrics*, Vol. 22, pp. 397-408.

[13] Reinsel, G. C. (1993), *Elements of Multivariate Time Series Analysis* Springer-Verlag, New York

[14] Schweppe, F. (1965), "Evaluation of Likelihood Function for Gaussian Signals," *IEEE Transactions on Information Theory*, Vol. 11, pp. 61-70.

[15] Stoffer D. S. (1985), "Maximum Likelihood fitting of STARMAX Models to Incomplete Space-Time Series Data, *Time Series Analysis: Theory and Practice* 6, Anderson O.D., Ord J. K. and Robinson.

Appendix

Quarterly Apartment Prices per 3.3 square meter in the Seoul Broughs

(unit : 1000won)

Kang-Nam	SeoCho	SongPa	Kang-Dong	DongJak	Seong-Dong	Dongdae-moon	Joong-Ryang	NoWeon	DoBong
3710	5470	3790	2980	4220	2640	2190	2350	2500	2730
5320	5940	3870	3550	3910	3440	2730	2920	3150	2730
5320	5780	4030	3550	3910	3200	2660	2940	3210	2970
5320	6090	3950	3710	3910	3590	2690	3040	3310	3090
6130	7500	5400	4680	4000	4300	3280	3870	3710	3520
7580	8440	5370	5970	4060	5000	3910	4110	4680	4060
8230	9380	6770	6370	5630	5470	4300	4900	4680	4420
9030	9840	7100	6210	6210	5940	4690	5160	5190	4380
9680	10160	8060	7260	7190	6410	4970	5400	5560	5230
10000	10630	7740	7260	7340	6720	5280	6020	6290	5390
9680	10630	7580	6610	7190	7030	5470	5920	6050	5440
9030	8590	6770	5890	6410	6090	4610	5730	4680	4690
8390	9060	6940	5650	6410	5470	4300	5110	5000	4730
7580	8280	6450	5560	6410	5160	4060	4760	4760	4610
7180	8590	6770	5970	5940	5160	4140	4680	4840	4450
8060	8440	6450	5650	5940	5390	4140	4680	4980	4530
8070	8440	6290	5970	5940	5470	4140	4840	4760	4570
6940	8440	6130	5240	5550	6310	3990	4790	4840	4490
6610	8440	6130	5320	5550	5000	4220	4840	4520	4310
6940	8440	6130	5330	5310	5000	4140	4400	4360	4310
7420	8440	6130	5640	5310	5160	4140	4610	4520	4400
7420	8440	6130	5000	5310	4840	4060	4600	4520	4380
7420	9060	6130	5160	5310	5160	4060	4600	4520	4380
8060	9060	6130	5640	5310	5310	4060	4680	4350	4380