Applied Statistical Analysis of Detailed Geographical Data with Emphasis on Spatial Effects(空間効果を考慮した詳細地理データの応用統計分析)

著者	Paez Robles Martinez Hector Antonio
号	2566
発行年	2000
URL	http://hdl.handle.net/10097/7839

	パエス・ロブレス・マルティネス・エクトル・アトニオ
氏名	Páez Robles Martínez Héctor Antonio
授与学位	博士（工学）
学位授与年月日	**平成12年9月13日**
学位授与の根拠法規	学位規則第4条第1項
研究科，専攻の名称	東北大学大学院工学研究科（博士課程）土木工学専攻
学位論文題目	Applied Statistical Analysis of Detailed Geographical Data with Emphasis on Spatial Effects (空間効果を考慮した詳細地理データの応用統計分析)
指導教官	東北大学教授　宮本和明
論文審査委員	主査　東北大学教授　宮本　和明　東北大学教授　澤本　正樹 東北大学教授　森杉　壽芳　東北大学教授　稲村　肇 東北大学助教授　内田　敬

論文内容要旨

Regional scientists and planners, environmental scientists, geographers and other specialists who routinely deal with matters spatial, have witnessed a notable increase of interest in their fields in recent years. Importantly fueled by the advent of cheap computing power, more powerful software and analysis techniques, and a wealth of statistical and remotely sensed data, a greater awareness of the importance of the spatial analysis of data has also been evident. At the same time, new developments in the field of spatial data analysis in general, and spatial statistics in particular, have opened the way to richer and more meaningful studies of typical as well as new problems that range from environmental to economic and planning issues. In this context, there is an increased need for empirical studies involving spatial data, and also for further development of statistical methods to describe and model spatial processes. Such is the topic of the present thesis.

Considering the wide field of spatial analysis, this study is concerned with a very specific aspect therein, namely that of modeling spatial phenomena often found in regional analysis and engineering. Regression analysis is the classic statistical tool employed to empirically estimate models in situations where physical or deterministic formulations of a process are too expensive to implement or too difficult to define. For some time now, however, it has been known that many spatial processes are inherently continuous, and that there are some seemingly unsurpassable limitations in data collection and coding. Spatial processes such as interaction, exchange and transfer, dispersion, and diffusion, as well as many data handling limitations, are often directly unobservable, but constitute the cause underlying spatial effects observable from the data. These effects include the issues of scale and rezoning, edge effects and the definition of a pattern of interactions. In this thesis we concentrate on three of the most important and tractable of them, namely spatial association (spatial patterns of similarity or dissimilarity), spatial heterogeneity (lack of uniformity or structural instability) and spatial nonstationarity (spatial parametric instability or variation). An important characteristic of these effects is that they invalidate many standard statistical results, most prominently because they violate crucial independence assumptions. As a direct consequence, ordinary regression models that ignore them can no longer be considered to be adequate tools when applied to spatial data.

Using empirical case studies, we investigate, in a progressive fashion, the three aforementioned effects, their consequences for modeling, alternative modeling techniques that explicitly consider them, and their interpretation. To accomplish this, a basic spatial econometric framework is adopted for this study, which we proceed to expand in a number of ways to suit our necessity for models of joint effects, and to address some relevant inferential questions about the models. To better demonstrate the use of the methods that we introduce, we undertake two relevant empirical examples, one taken from environmental science, and one from transportation planning. The first case study is concerned with the description and explanation of the urban heat island phenomenon, and helps us demonstrate a situation where spatial association is present. Three subsequent case studies thread on the topic of urban land price determination and modeling in the presence of spatial effects. All the case studies use Sendai City's data set, which turns out to be an ideal example, because it can be used, at every stage, to show the limitations of partial analysis frameworks, in other words those ignoring some, or all, spatial effects. The case studies also hold empirical relevance, not the least because they deal with important and actual issues in planning.

The structure of the study is as follows: In the first chapter we introduce the rationale of the study and briefly review the contents of the thesis to provide the reader with a basic overview of the structure of our work. Next, in chapter 2, we start preparing the ground to begin the study. In this chapter, we discuss some preliminary topics regarding the basic tools that serve as the starting line for our contribution. In this chapter, for instance, we review the literature and present the existing methods, models and statistics that have been so far proposed. We basically concentrate on two statistics of spatial association, and the econometric model of spatial association and heterogeneity. Some of these tools we use, but in general we try to expand and alter them in later chapters to satisfy our need for specialized analytical tools. We also set the scope of the thesis discussing spatial effects in general, the three effects that we cover here, and others important effects, that nonetheless lay beyond the scope of the present study for practical reasons.

We start our study proper in chapter 3, addressing the topic of spatial association, defined as a degree of similarity or dissimilarity among spatially related observations. In this chapter we introduce the basic model of spatial association and apply it to the problem of the urban heat island in Sendai City. The urban heat island is caused by a heat exchange differential between the city and the surrounding countryside. As such, it has very distinctive spatial characteristics that, as we show, induce spatial association issues into the modeling process. We first estimate a regular non-spatial model, and then apply a spatially autoregressive specification, that is, a model of association of the objective variable, and show with this example the consequences of ignoring spatial association and effects in general. Given the number of recent applications of regression to study the heat island effect, the relevance of this issue in an empirical context is brought into the limelight.

In chapter 4, we approach the issue of spatial heterogeneity, which can be roughly defined as the existence of structural differences in the behavior of the process, meaning that it can be described differently at different locations within the study area. The switching regression modeling methodology used to model heterogeneity is introduced and expanded to obtain a specification that involves a spatial autoregressive element and structural parametric instability. Although this is not a standard form of the model, we derive from the basic framework the expressions (i.e. likelihood function, first and second order conditions) needed to estimate and test a model of land prices. In this way we effectively incorporate association and heterogeneity as part of the analysis. Moreover, we show how a local statistic of spatial association can be used to explore the issue of spatial heterogeneity in the data and to detect different spatial regimes. Taking as a case study land price determination in Sendai City, analysis of the data and preliminary model estimation show that, in addition to spatial association, the topic of chapter 3, there are strong indications of spatial heterogeneity. In this fashion we investigate the spatial structure of a city and calculate the implicit price of environmental characteristics, and show empirically how some measures frequently thought to neutralize spatial effects fail at times. In addition, drawing from the economic theory of externalities, we provide a justification for the use of a spatially autoregressive model to represent spatial spillovers in the land market.

In chapter 5 we explore the concept of spatial nonstationarity, defined as complex spatial variation in the description of a process, extending the example of spatial spillovers initiated in chapter 4. It is noted that the framework used therein is of limited help when it comes to the study of the geographically detailed characteristics of the model. In this section, we estimate a spatial econometric model of land prices, applying the technique of Geographically Weighted Regression (GWR), used to obtain spatially localized estimates of model parameters. The objective is to compare the spatial distribution of spillovers and other variable parameters to that of public transportation infrastructure provision, in particular rail systems. The case study shows that considerable parameter variation over space exists, sustaining the hypothesis of spatial nonstationarity. In an empirical vein, we show that spatial spillover effects are, on average, more favorable where transportation infrastructure exists, and reveal in a characteristic finding by this method, some interesting aspects relating the structure of the city, as shown by the rest of the variables' spatial distribution of the parametric. The existing methodological framework, however, does not provide the elements to conduct inference and to test the model against misspecification.

To overcome the above limitation, in chapter 6 we present a maximum likelihood-based framework (ML) for estimation and inference of a general Geographically Weighted Regression model. The framework generalizes the model of spatial processes, and covers simple linear models, by now well developed based on an ordinary least squares approach (OLS), as well as spatially autoregressive and spatial dependence in error terms models. Additionally, it extends the existing analytical formulation to enable statistical inference of the general model, and to test it against several forms of misspecification. In this way, any local model can be tested against spatial error autocorrelation and omission of a spatial lag (autoregressive structure), something that was not possible under the previously available modeling framework. Finally, in this chapter the fundamental relationship between spatial heterogeneity and nonstationarity is explored. We exemplify these new techniques in an application using data from Sendai City's CBD. The example confirms the correspondence between ML and OLS estimators and shows how the local models can be diagnosed for the presence of other spatial effects

Finally, in chapter 7 we summarize the study and offer some conclusions and recommendations for further study. It should be noted that the structure of the thesis reflects the chronological development of the research reported here. More importantly, we move from a basic modeling technique such as a spatial autoregression, to higher levels of complexity, involving heterogeneity and nonstationarity in models of joint effects. In this sense, we have taken a progressive approach that also mimics some recent developments in the field. In effect, the topic of spatial association (autocorrelation) was among the first to be noticed by spatial analysts, while spatial nonstationarity has very recently been the subject of renewed interest with the introduction of the Geographically Weighted Regression method.

Contributions by this study are methodological and empirical. In terms of methodology, the basic statistical framework is adapted and extended to fit the characteristics of the problems studied. In this way, a spatially switching regression, and a GWR version of the same model are obtained to tackle the problems raised by the particular characteristics of the case studies at hand. Also, the elements needed to conduct inference and misspecification tests of these models are derived. Moreover, the basic spatial statistical model that considers spatial association and heterogeneity is expanded, to obtain a general model that covers as special cases the models of association, heterogeneity and nonstationarity discussed above. In an empirical vein, we apply these tools to two relevant problems in planning and environmental studies, to show how richer insights can be thus obtained. At a time when spatial statistics is sometimes perceived as a set of techniques in search of a problem, we show how some prototypical problems can be analyzed to yield, at the very least, better empirical results. In some cases we even find a teaser for a conceptual reappraisal of the underlying theory given the new light offered by fresh evidence. This is the case, we believe, with the interpretation of a spatial autoregressive model as a representation of economic spillovers. Moreover, in terms of empirical applications, those concerning geographically detailed data and analysis that we present in chapters 5 and 6 are among the first of their kind to be advanced. In this way, we hope to have contributed to demonstrate the value and the relevance of the proposed techniques in the applied statistical analysis of geographically detailed data.

論文審査結果の要旨

空間分布データを取り扱う分野は多岐に及ぶが、空間データの特性、すなわち標本点の位置関係に基づく空間効果に対処する手法に関しては、従来は個別的な取り組みが主で、体系的には確立されていなかった。そこで、本研究においては、空間効果に関する既存の研究をふまえ、空間関連性（spatial association）、空間異質性（spatial heterogeneity）そして空間非定常性（spatial non-stationarity）に着目して、それらの効果を明示的に考慮することにより、従来の回帰分析におけるモデル構築の歪みを改善する手法を開発した。本論文は、この研究成果についてまとめたものであり、全文７章よりなる。

第１章は序論であり、本研究の背景および目的を述べている。

第２章では、既存研究について整理し、本研究の基礎としている。

第３章では、空間関連性、すなわち、空間分布データ間の位置に関連した類似性の度合いを取り扱っている。その方法としては、目的変数自体の周辺への影響を明示的に考慮した自己回帰モデルを構築し、仙台市のヒートアイランド現象への適用を通して、誤差の空間的な自己相関の改善を確認した。

第４章では、空間異質性、すなわち、回帰モデルパラメーターの位置に起因した不均一性について、空間関連性の分離をさらに発展させて、空間的多分自己回帰モデルを用いて空間関連性と異質性を同時に考えるモデルを構築している。そして、仙台市の地価分析に適用することにより、従来の回帰分析では無理であった、独立性の検定に合格するモデルがはじめて構築できることを示した。

第５章では、さらに空間非定常性を追加したモデルの構築を行った。空間非定常性は、回帰モデルパラメーターの地点ごとの変動性を指しており、異質な空間が無限に存在するという意味で空間異質性の一般化ともいえる。空間関連性、異質性、そして非定常性を表現するために、さらに、位置重み回帰分析モデル（GWR: Geographically Weighted Regression）を開発している。前章と同じく仙台市の地価分析に本モデルを適用したところ、近隣の地価の影響に関して交通条件がより良い所の方が大きいことなど、その特性が説明された。

第６章では、以上をまとめて、一般的な位置重み回帰式を構築し、ｔ検定方法を確立している。このモデルにより、従来は不可能であった、局所的なモデルの空間的自己相関や目的変数の空間的影響に関しての検定が可能となった。さらに、その有効性について、仙台市の地価に適用することにより確認した。

第７章は結論をまとめている。

以上要するに本論文は、空間分布の詳細なデータ処理において、従来は個別的に取り組まれてきた方法をもとに、統一的にその空間効果を明示的に考慮する一般的なモデルを理論的に構築し、また、その推定およびモデル同定に関する統計的検定方法に関しても整備した。さらに、実際の適用を通して、その統計的手法としての有効性のほか、モデル構造の解釈に関する指針を明示した。

よって、本論文は博士（工学）の学位論文として合格と認める。