

学校代号: 10731

学 号: 112070105013

密 级: 公开

兰州理工大学硕士学位论文

时空加权回归模型的参数估计 及其应用研究

学位申请人姓名: 李琪

导师姓名及职称: 玄海燕 副教授

培 养 单 位: 兰州理工大学

专 业 名 称: 运筹学与控制论

论文提交日期: 2014年5月4日

论文答辩日期: 2014年6月4日

答辩委员会主席: 王才士 教授



Parameter Estimation and Application Research of Geographically and
Temporally Weighted Regression Model

by

LI Qi

B.S.(Anhui University of Technology)2011

A thesis submitted in partial satisfaction of the

Requirements for the degree of

Master of Science

in

Operational Research and Cybernetics

in the

Graduate School

of

Lanzhou University of Technology

Supervisor

Associate Professor Xuan Haiyan

April, 2014

兰州理工大学 学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：李琪

日期：2014年6月10日

学位论文授权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权兰州理工大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密□，在_____年解密后适用本授权书。

2、不保密☒。

（请在以上相应方框内打“√”）

作者签名：李琪

日期：2014年6月10日

导师签名：李琪

日期：2014年6月10日

目 录

摘 要	I
Abstract	II
插图索引	IV
第 1 章 绪 论	1
1.1 空间变系数回归模型及其研究现状	1
1.2 时空加权回归模型及其研究现状	3
1.3 本文的主要研究内容	4
第 2 章 时空加权回归模型的 GTWR 拟合方法	6
2.1 GTWR 拟合方法	6
2.2 权函数的选择	8
第 3 章 时空加权回归模型的估计	11
3.1 局部线性估计及其性质	11
3.1.1 局部线性估计	11
3.1.2 局部线性估计的性质	14
3.2 二步估计及其性质	17
3.2.1 二步估计	17
3.2.2 二步估计的性质	20
第 4 章 时空加权回归模型的应用研究	25
4.1 数据来源及建立模型	26
4.1.1 数据来源与预分析	26
4.1.2 建立模型	26
4.2 利用局部线性估计模拟的结果与分析	28
4.3 利用二步估计模拟的结果与分析	30
结论与展望	32
参考文献	34
致谢	39
附录 A (攻读学位期间所发表的学术论文目录)	41

摘 要

近年来,随着海量数据的频繁出现和数据生成机制的日益复杂化,假定数据结构的某种一致性和回归函数具有某一特定的参数形式,已不能适应复杂数据的分析。非参数回归模型便是在这一背景下产生的具有高度灵活性和适应性的回归模型。在众多的非参数回归模型中,变系数回归模型作为经典线性回归模型的一个有用推广,受到普遍重视并得到广泛应用。地理加权回归模型作为空间变系数回归模型的一种,由于其不仅可以探测系数的变化而且还能对系数变化的显著性进行解释,因而地理加权回归模型的提出和应用具有较大的影响。时空加权回归模型是在地理加权回归模型的基础上推广得来的,模型中嵌入了时间因素,能够同时解决空间和时间的非平稳性问题。本文从时空加权回归模型的拟合方法、估计量的性质和应用方面做了研究。

介绍了时空加权回归模型基于最小二乘原理的拟合方法,给出了与之相关的权函数选取原则和确定窗宽参数的方法。结合局部线性估计和GTWR拟合方法给出了时空加权回归模型的局部线性估计方法及其估计的一个有限样本性质,并通过与GTWR方法的比较说明了局部线性估计方法在系数函数及误差方差估计方面的改进效果。基于时空加权回归模型的局部线性估计给出了时空加权回归模型的二步估计方法。通过对二步估计中的第二步估计的改进,给出了一个改进的二步估计方法,解决二步估计计算量大的缺点,并给出了系数函数局部线性估计的条件偏近似值及其估计和条件方差的近似值。

最后我们利用时空加权回归模型研究了气象学中的一个实例。基于气象数据,利用时空加权回归模型模拟了空气质量指数与降水量及温度之间的关系,考察了局部线性估计方法及二步估计方法对系数函数估计的效果,并分析了我国空气质量指数的时空分布特征。结果表明局部线性估计方法及二步估计方法对系数函数估计具有很好的效果,时空加权回归模型在实际应用中有着很好的效果。

关键词: 时空加权回归; 加权最小二乘估计; 局部线性估计; 二步估计; 应用研究

Abstract

In recent years, with the frequent appearance of massive data and the increasing complexity of data generation mechanism, assuming that the data structure has some consistency and the regression function possesses a particular parametric form has no longer adapted the analysis of complex data. In this context, non-parametric regression model with a high degree of flexibility and adaptability is generated. In the numerous non-parametric regression models, varying coefficient regression model as a useful extension of the classical linear regression model, has received wide attention and been extensively used. Geographically weighted regression (GWR) model as one of the spatial varying coefficient regression models, can not only probe the changes of coefficients but also interpret significant of its changes. Therefore the proposing and application of geographically weighted regression model has a great impact. By incorporating temporal effects into the geographically weighted regression model, an extended GWR model, geographically and temporally weighted regression (GTWR) model, has been developed to deal with both spatial and temporal nonstationarity simultaneously. The fitting method, finite sample properties of estimators and application of geographically and temporally weighted regression model have been researched in this article.

Based on the theory of weighted least squares estimate, the fitting method of geographically and temporally weighted regression model is given, the related select principle of weight function and cross-validation for fixing bandwidth parameters are also provided. Combining local linear estimation and GTWR fitting method, local linear estimation method and finite sample properties of geographically and temporally weighted regression model are given, and by comparing with GTWR method, we show that the local linear estimation method improves the estimated effect of the coefficient functions and the error variance. The two-step estimation procedure of geographically and temporally weighted regression model is given on the basis of local linear estimation. But the drawback of the two-step estimation procedure is large computation. In order to repair this drawback, an improved two-step estimation procedure is proposed and the approximation of the conditional bias and its estimation and the approximation of the conditional variance of local linear estimator of the coefficient functions are given respectively.

Finally, we use geographically and temporally weighted regression model to study an instance of Meteorology. In order to test the performance of the local linear

estimation method and two-step estimation method , based on the meteorological data, geographically and temporally weighted regression model is used to simulate the relationship among air quality index, temperature and precipitation. And then analyze the temporal and spatial distribution characteristic of air quality index. The results show that the local linear estimation method and two-step estimation method have good effects and the performance of geographically and temporally weighted regression model is favorable in the practical application.

Key words: geographically and temporally weighted regression; weighted least squares estimation; local linear estimation; two-step estimation; application research

插图索引

图 4.1	67个城市1月的月平均空气质量指数	27
图 4.2	北京市12个月的月平均空气质量指数	27
图 4.3	67个城市2月份的月降水量总和	27
图 4.4	杭州市12个月的月降水量总和	27
图 4.5	67个城市的2月份的月平均温度	27
图 4.6	兰州市12个月的月平均温度	27
图 4.7	67个城市2月份的 β_0 分布图	29
图 4.8	67个城市2月份的 β_1 分布图	29
图 4.9	67个城市2月份的 β_2 分布图	29
图 4.10	二步估计得到的67个城市2月份的 β_1 分布图	31

第1章 绪论

1.1 空间变系数回归模型及其研究现状

在地理学、经济学、环境学、气象学及流行病学等众多领域中,数据一般都来自于特定的地理位置及数据带有其地理位置的信息,这种具有空间位置属性的数据称之为空间数据。根据Tobler地理学第一定律(Tobler, 1970):任何事物与其他事物都是空间相关的,距离越近的事物之间的空间相关性越大。因此,空间数据与通常的截面数据不同,他们之间的空间相关性会导致回归关系的空间非平稳性或回归关系的空间异质性(spatial heterogeneity)。通常的线性回归模型由于忽略了数据的空间位置属性,只表征了平均意义下因变量和自变量的相关关系,而不能有效反映回归关系的空间非平稳性特征。因此如何将数据的空间位置属性纳入回归模型中以分析回归关系随空间位置而变化的特征在空间数据回归分析中有重要意义。

我们以某城市房屋价格 Y 和房屋面积 X 的关系为例来说明空间数据回归关系的空间非平稳性。显然,房屋价格不仅与房屋面积有关,而且还与房屋所在的地理位置 (u, v) (如经度和纬度或者某直角坐标系下的地理位置坐标)密切相关,因此相关数据属于空间数据。如果不考虑数据的空间位置信息而拟合 Y 与 X 的线性回归模型

$$Y = \beta X + \varepsilon$$

则回归系数 β 意味着该城市单位面积房屋的平均价格,而一个城市单位面积房屋的价格随房屋所在的地理位置不同有很大的差异,以上的线性回归关系则不能反映单位房价的这种地理位置上的差异性,不能有效挖掘 Y 和 X 之间回归关系的一些重要有用的“局部”特征。事实上,单位面积房屋的价格 β 应该是房屋所在的地理位置 (u, v) 的函数,以反映在不同地理位置处单位面积房屋价格的不同,因此一个更为合理的回归模型应该是如下的空间变系数回归模型

$$Y = \beta(u, v)X + \varepsilon$$

该模型则能反映 Y 与 X 之间回归关系的空间非平稳性或称之为空间异质性。若基于观察数据 $(Y_i; X_i, u_i, v_i)$,利用某种拟合方法得到回归系数函数 $\beta(u, v)$ 的估计,不但可以得到具有面积 X 的房屋在不同地理位置上的价格,而且可以根据 $\beta(u, v)$ 在各地地理位置处的估计值了解该城市单位面积房屋价格随空间位置的改变而变化的规律,分析回归关系的局部特征。

一般情况下,设 Y 为因变量, X_1, X_2, \dots, X_p 为自变量, $(Y_i; X_{i1}, X_{i2}, \dots, X_{ip})$ 为因变量和自变量在地理位置 (u_i, v_i) 处的观测值($i = 1, 2, \dots, n$),则空间变系数回归模

型的形式如下

$$Y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{i1} + \cdots + \beta_p(u_i, v_i)X_{ip} + \varepsilon_i, i = 1, 2, \cdots, n \quad (1.1)$$

其中 $\beta_j(j = 0, 1, 2, \cdots, p)$ 为空间地理位置 (u_i, v_i) 的未知函数; $\varepsilon_i(i = 1, 2, \cdots, n)$ 为独立同分布的误差项且 $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$ 。

空间变系数模型通过假定线性回归模型中的回归系数是观测点空间位置的函数。将数据的空间位置信息纳入到模型之中,利用回归系数函数在个空间位置处的估计值,分析回归关系的空间变化特征。

Brunsdon等^[1](1996)首次明确给出了如上的空间变系数模型并提出了以观测点之间的距离的函数为权值的非参数光滑估计方法,称之为地理加权回归方法(geographically weighted regression,简称GWR)。其后,他们又在不同领域的杂志上发表了数篇论文(例如,Fotheringham^[2,3]等,1997,1998;Brunsdon等^[4],1998),从不同的角度阐述了GWR方法在探索空间数据回归关系的非平稳性方面的作用,并通过一些实际数据的分析以及相关方法的比较说明了GWR方法的有效性,并于2002年出版了专著(Fotheringham等^[5],2002),系统阐述了GWR方法的原理、应用以及一些可能的推广方向。

虽然空间变系数模型及其GWR拟合方法本质上是变系数模型及其局部拟合方法在空间数据下的一个直接推广,但由于GWR方法十分明确的应用背景和拟合方法在概念上的简单性以及实现上的易操作性,该方法在短短的十多年中很快得到研究工作者,尤其是实际领域的数据分析工作者的重视。在理论基础方面,从有关统计推断与诊断、拟合方法的改进、模型的选取得到一定的研究。例如,Leung等^[6,7](2000a,2000b)给出了回归函数全局平稳性和回归系数随地理位置变化的显著性以及拟合残差的空间自相关性的近似检验方法;Páez等^[8,9](2002a,2002b)从误差方差空间异质性的角度提出了GWR的最大似然估计和统计推断框架;Wheeler和Tiefelsdorf^[10](2005)研究了自变量的共线性对回归系数GWR估计相关性的影响,发现自变量的共线性会使得回归系数估计之间的相关性增强,从而影响分析结果的合理解释,并建议用各系数估计的二元散点图和局部参数相关图考察自变量共线性对系数估计的影响;Wheeler^[11](2007)进一步将线性回归模型中的共线性诊断方法推广到GWR估计的情况,给出了系数函数地理加权岭估计以降低自变量共线性对回归系数估计的影响;Farber和Páez^[12](2007)通过系统的数值模拟实验考察了观测数据的异常值对GWR方法中窗宽参数值的影响并提出了改进措施;Wang等^[13](2008)将变系数模型的局部线性拟合方法推广到空间变系数模型,给出了空间变系数模型的局部线性GWR拟合方法以改进原GWR方法对系数函数估计的精确性,并通过数值模拟实验说明改进的拟合方法可以显著降低系数函数估计的偏和边界效应;Demsar等^[14](2008)用地理加权回归模型和geosual分析法研究了地理过程的时空动态,探讨了通过已知的不同的时空参数捕捉复杂的时空

过程的可行性。Wheeler^[15](2009)提出了地理加权回归的Lasso(Least absolute shrinkage and selection operator)方法,以实现模型的选择。Harris等^[16,17](2010,2011)联系和比较了地理加权回归模型与地质统计KED模型,得出了地理加权回归可以作为一个有价值的补充预测的结论,接着又将地理加权回归作为一个空间预测,模拟了多元线性回归(MLR)、传统地质统计模型和混合类统计模型的不同水平的空间异质性和空间自相关性。Wei等^[18](2012)对混合地理加权回归模型的估计和检验问题进行了研究。Harris和Brunsdon等^[19](2013)使用comap作为一种诊断工具对非平稳kriging模型进行了研究,在文中给出了两种地理回归变量在条件分布下的关系。

在GWR方法的应用方面有更为丰富的工作,该方法目前已应用于如经济学、金融学、气象学、生态学、环境学、社会学等众多领域的空间数据分析问题。在此我们仅列出几篇有关应用方面的研究论文,其中包括Huang和Leung^[20](2002)分析了江苏省区域工业化水平的空间特征;Wang等^[21](2005)分析了我国森林生态系统的主要产出的空间特征;Yu^[22](2006)考察了北京及周边地区区域发展机制的空间非平稳性;Cho等^[23](2007)分析了我国农业生产的投入产出弹性在全国各地的变化情况;Kamarianakis等^[24](2008)用非线性混合模型和地理加权回归模型对遥测降雨量的估计进行了评估,文中探讨了地面降雨量的真实数据与卫星遥测得到的数据的关系,并用地理加权回归模型检验了降雨量的地面观测值与卫星遥测的估计降雨量之间的空间非平稳性。Luo和Wei^[25](2009)以南京为例分析了我国城市化发展模式的空间变化情况;Li等^[26](2010)分析了深圳市地表温度与环境因素的空间非平稳性;Zhao等^[27](2010)考察了海河流域天气和植被分布之间的空间变化特点。Brown等^[28](2012)用地理加权回归模型探讨了降雨量,土地覆盖与地表水域随时空变化的关系。Su等^[29](2012)结合地理加权回归模型对农业景观格局与城市化的空间变化关系进行了多尺度分析。Mayers等^[30](2012)使用地理加权回归方法分析研究了烟草批发商店密度与人口的关系。Fortheringham^[31](2012)通过分析爱尔兰饥荒对人口的影响,扩展了对地理加权回归模型中“地理”的理解。

1.2 时空加权回归模型及其研究现状

在很多情况下,数据的结构不仅与其空间地理位置有关,而且与时间因素有关。这些数据有以下特点:在一个时间断面上,它属于空间数据,具有空间地理位置上的相关性;而在特定的地理位置上,其观测值是一个时间序列,具有时间因素上的相关性,这种数据称为时空相关数据。由于拟合空间变系数回归模型的地理加权回归方法只考虑了数据的空间结构,而未考虑时间因素。因此该方法还不能有效地用于这类时空数据的分析。

为了将数据的时空特性嵌入到回归模型中予以分析,最近由Huang等^[32](2010)提出了时空加权回归模型 (geographically and temporally weighted regression model),形式如下:

$$y_i = \beta_0(u_i, \nu_i, t_i) + \sum_{k=1}^d \beta_k(u_i, \nu_i, t_i) x_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1.2)$$

其中, $(y_i; x_{i1}, x_{i2}, \dots, x_{id})$ 是因变量 Y 和自变量 X_1, X_2, \dots, X_d 在观测位置 (u_i, ν_i, t_i) 处的观测值, $\varepsilon_i, i = 1, 2, \dots, n$ 为独立的随机误差项, 均值为零, 方差为 σ^2 。 $\beta_k(u_i, \nu_i, t_i)$, $k = 0, 1, 2, \dots, d$ 是 $d+1$ 个未知的系数函数。

时空加权回归模型是在地理加权回归模型基础上推广的, 通过在变系数回归模型中假定回归系数是地理位置和观测时刻的任意函数, 将数据的时空特性嵌入到回归模型中, 为分析回归关系的时空特性创造了条件, 并为解决回归模型的空间非平稳性和时间非平稳性提供了可行性。Huang等^[32]将时空加权回归模型应用于房地产价格变化因素的分析上, 研究表明时空加权回归模型比传统的线性回归模型、地理加权回归模型有更好的拟合优度。张金牡^[33]对深圳市住宅价格进行时空加权统计回归分析, 研究结果表明, 时空加权回归模型不仅可以分析相关因素在空间的差异性上对住宅价格具有重要影响, 还可以揭示其在时间上的差异性, 得出相关因素对住宅价格的影响是随时间和空间变化的结论。

给定一个时空数据集, 利用时空加权回归模型进行拟合, 可以得到不同的地理位置、不同的观测时刻的参数估计值, 通过各参数在所研究区域及时间段内的变化来探索观测数据的全局平稳性、空间非平稳性和时间非平稳性。只通过不同观测点的估计值大小分析数据的空间非平稳性和时间非平稳性是不恰当的, 需用统计检验方法进行非平稳性检验。对于空间变系数回归模型, Leung等^[6,7]给出了检验回归关系的空间非平稳性及误差项的空间自相关性的统计检验方法, 然而, 此方法并不适用于时空加权回归模型, 因此, 对于模型 (1.2), 检验模型的空间非平稳性和时间非平稳性是一个非常重要的问题。Yu等^[34](2011)用时空加权回归模型探讨了当地的犯罪事件机制的非平稳性。Xuan和Li等^[35,36]对时空加权回归模型的统计推断和影响分析进行了研究。

1.3 本文的主要研究内容

基于以上的背景及分析, 本文的主要研究内容包括以下四个方面:

(1) 阐述本课题的研究背景和意义, 并且对国内外时空加权回归模型的研究现状进行综述。

(2) 介绍了时空加权回归模型基于最小二乘原理的GTWR拟合方法, 给出了与之相关的权函数选取原则和确定窗宽参数的方法。

(3) 结合局部线性估计和GTWR拟合方法给出了时空加权回归模型的局部线性估计方法及其估计的性质, 并基于时空加权回归模型的局部线性估计给出时空加权回归模型的一种新的二步估计方法。

(4) 基于气象学的一个实例研究, 使用局部线性估计方法和二步估计方法对时空加权回归模型进行拟合, 并基于有关的统计检验结果分析了我国空气质量指数(AQI)与降水量和温度关系的时空分布特征。

本文的具体内容安排如下: 第2章介绍了时空加权回归模型基于最小二乘原理的拟合方法; 第3章给出了时空加权回归模型的局部估计和二步估计; 第4章给出了时空加权回归模型的实例研究; 最后是本文的总结。

第2章 时空加权回归模型的GTWR拟合方法

在变系数回归模型中,当回归系数是地理位置和观测时刻的函数时,模型变为时空加权回归模型。给定 n 组观测值 $(y_i; x_{i1}, x_{i2}, \dots, x_{id}), i = 1, 2, \dots, n$, 时空加权回归模型的形式为(1.2)式, 其中 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 相互独立且满足 $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$ 。

本章介绍了时空加权回归模型的GTWR拟合方法, 给出与之相关的权函数选取原则和确定窗宽参数的交叉确认法。

2.1 GTWR拟合方法

考虑时空加权回归模型(1.2), 采用局部加权最小二乘方法估计回归系数。即对于给定的一个观测点, 对该点处的参数进行估计时, 靠近该点的观测应赋予较大的权值, 远离该点的观测应赋予较小的权值, 通过使观测值与拟合值的差的平方的加权和达到最小, 可求得该点处参数的估计值。

设 (u_0, ν_0, t_0) 为所研究区域内任一点, 在该点处指定一组权

$$\omega_1(u_0, \nu_0, t_0), \omega_2(u_0, \nu_0, t_0), \dots, \omega_n(u_0, \nu_0, t_0)$$

选择该观测点处的回归系数

$$\beta_0(u_0, \nu_0, t_0), \beta_1(u_0, \nu_0, t_0), \dots, \beta_d(u_0, \nu_0, t_0)$$

使得

$$\sum_{i=1}^n [y_i - \beta_0(u_0, \nu_0, t_0) - \beta_1(u_0, \nu_0, t_0)x_{i1} - \dots - \beta_d(u_0, \nu_0, t_0)x_{id}]^2 \omega_i(u_0, \nu_0, t_0)$$

达到最小。令

$$f(\beta) = \sum_{i=1}^n [y_i - \beta_0(u_0, \nu_0, t_0) - \beta_1(u_0, \nu_0, t_0)x_{i1} - \dots - \beta_d(u_0, \nu_0, t_0)x_{id}]^2 \omega_i(u_0, \nu_0, t_0)$$

对 $\beta_k(u_0, \nu_0, t_0), k = 0, 1, \dots, d$, 求偏导数, 并令其为零, 有如下方程组

$$\begin{cases} \frac{\partial f(\beta)}{\partial \beta_0(u_0, \nu_0, t_0)} = -2 \sum_{i=1}^n [y_i - \beta_0(u_0, \nu_0, t_0) - \sum_{k=1}^d \beta_k(u_0, \nu_0, t_0)x_{ik}] \omega_i(u_0, \nu_0, t_0) = 0 \\ \frac{\partial f(\beta)}{\partial \beta_1(u_0, \nu_0, t_0)} = -2 \sum_{i=1}^n [y_i - \beta_0(u_0, \nu_0, t_0) - \sum_{k=1}^d \beta_k(u_0, \nu_0, t_0)x_{ik}] \omega_i(u_0, \nu_0, t_0)x_{i1} = 0 \\ \vdots \\ \frac{\partial f(\beta)}{\partial \beta_d(u_0, \nu_0, t_0)} = -2 \sum_{i=1}^n [y_i - \beta_0(u_0, \nu_0, t_0) - \sum_{k=1}^d \beta_k(u_0, \nu_0, t_0)x_{ik}] \omega_i(u_0, \nu_0, t_0)x_{id} = 0 \end{cases}$$

解方程组得 $\beta_0(u_0, \nu_0, t_0), \beta_1(u_0, \nu_0, t_0), \dots, \beta_d(u_0, \nu_0, t_0)$ 的估计值。令

$$\beta(u_0, \nu_0, t_0) = (\beta_0(u_0, \nu_0, t_0) \ \beta_1(u_0, \nu_0, t_0) \ \dots \ \beta_d(u_0, \nu_0, t_0))^T \quad (2.1)$$

$$\hat{\beta}(u_0, \nu_0, t_0) = (\hat{\beta}_0(u_0, \nu_0, t_0) \ \hat{\beta}_1(u_0, \nu_0, t_0) \ \dots \ \hat{\beta}_d(u_0, \nu_0, t_0))^T \quad (2.2)$$

$$W(u_0, \nu_0, t_0) = \text{diag}[\omega_1(u_0, \nu_0, t_0), \omega_2(u_0, \nu_0, t_0), \dots, \omega_n(u_0, \nu_0, t_0)] \quad (2.3)$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1d} \\ 1 & x_{21} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nd} \end{pmatrix},$$

$$\beta(u_i, \nu_i, t_i) = \begin{pmatrix} \beta_0(u_i, \nu_i, t_i) \\ \beta_1(u_i, \nu_i, t_i) \\ \vdots \\ \beta_d(u_i, \nu_i, t_i) \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

上述方程组用矩阵表示为

$$[X^T W(u_0, \nu_0, t_0) X] \beta(u_0, \nu_0, t_0) = X^T W(u_0, \nu_0, t_0) Y \quad (2.4)$$

从而可求得 (u_0, ν_0, t_0) 处的参数估计

$$\hat{\beta}(u_0, \nu_0, t_0) = [X^T W(u_0, \nu_0, t_0) X]^{-1} X^T W(u_0, \nu_0, t_0) Y \quad (2.5)$$

因变量 Y 在 (u_0, ν_0, t_0) 处的拟合值为

$$\begin{aligned} \hat{y}_0 &= (1 \ x_{01} \ x_{02} \ \dots \ x_{0d}) \hat{\beta}(u_0, \nu_0, t_0) \\ &= \hat{\beta}_0(u_0, \nu_0, t_0) + \hat{\beta}_1(u_0, \nu_0, t_0) x_{01} + \dots + \hat{\beta}_d(u_0, \nu_0, t_0) x_{0d} \end{aligned} \quad (2.6)$$

分别取 $(u_0, \nu_0, t_0) = (u_i, \nu_i, t_i)$, $i = 1, 2, \dots, n$, 则回归系数在各观测点 (u_i, ν_i, t_i) 处的估计值为

$$\begin{aligned} \hat{\beta}(u_i, \nu_i, t_i) &= (\hat{\beta}_0(u_i, \nu_i, t_i) \ \hat{\beta}_1(u_i, \nu_i, t_i) \ \dots \ \hat{\beta}_d(u_i, \nu_i, t_i))^T \\ &= [X^T W(u_i, \nu_i, t_i) X]^{-1} X^T W(u_i, \nu_i, t_i) Y \end{aligned} \quad (2.7)$$

因变量 Y 在 (u_i, ν_i, t_i) 处的拟合值为

$$\begin{aligned} \hat{y}_i &= (1 \ x_{i1} \ x_{i2} \ \dots \ x_{id}) \hat{\beta}(u_i, \nu_i, t_i) \\ &= \hat{\beta}_0(u_i, \nu_i, t_i) + \hat{\beta}_1(u_i, \nu_i, t_i) x_{i1} + \dots + \hat{\beta}_d(u_i, \nu_i, t_i) x_{id} \end{aligned} \quad (2.8)$$

若以 $x_i^T = (1 \ x_{i1} \ x_{i2} \ \cdots \ x_{id})$ 记为矩阵 X 的第 i 行, 则 y_i 的拟合值为

$$\begin{aligned}\hat{y}_i &= x_i^T \hat{\beta}(u_i, \nu_i, t_i) \\ &= x_i^T [X^T W(u_i, \nu_i, t_i) X]^{-1} X^T W(u_i, \nu_i, t_i) Y\end{aligned}\quad (2.9)$$

设 $\hat{Y} = (\hat{y}_1 \ \hat{y}_2 \ \cdots \ \hat{y}_n)^T$ 为 y_1, y_2, \dots, y_n 的拟合值所组成的向量, $\hat{\varepsilon} = (\hat{\varepsilon}_1 \ \hat{\varepsilon}_2 \ \cdots \ \hat{\varepsilon}_n)^T$ 为相应的残差向量, 则

$$\hat{Y} = SY \quad (2.10)$$

$$\hat{\varepsilon} = Y - \hat{Y} = (I - S)Y \quad (2.11)$$

残差平方和为

$$RSS = \hat{\varepsilon}^T \hat{\varepsilon} = Y^T (I - S)^T (I - S) Y \quad (2.12)$$

其中

$$S = \begin{pmatrix} x_1^T [X^T W(u_1, \nu_1, t_1) X]^{-1} X^T W(u_1, \nu_1, t_1) \\ x_2^T [X^T W(u_2, \nu_2, t_2) X]^{-1} X^T W(u_2, \nu_2, t_2) \\ \vdots \\ x_n^T [X^T W(u_n, \nu_n, t_n) X]^{-1} X^T W(u_n, \nu_n, t_n) \end{pmatrix} \quad (2.13)$$

2.2 权函数的选择

在本质上, 有两种加权方式可被使用: 固定核和自适应核。对于固定核来说, 距离是恒定的, 但最近邻元素的数目不同。对于自适应核而言, 距离是变化的, 但最近邻元素的数目是维持恒定的。最常用的核为高斯核函数

$$\omega_{ij} = \exp\left(-\frac{d_{ij}^2}{h^2}\right) \quad (2.14)$$

其中 h 为窗宽参数, h 是非负的。 d_{ij} 为位置 i 与 j 之间的距离, 使用点坐标 (x_i, y_i) 和 (x_j, y_j) , 这种距离通常被定义为欧式距离。

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2.15)$$

根据公式(2.14)和(2.15)可知, 如果 i 与 j 相同, 则观测值的权为1, 且当 i 与 j 之间的距离增加时, 由高斯曲线可知, 其它数据的权会减少。

为了避免(1.2)式夸大数据稀疏地区的非平稳性或(2.7)式屏蔽掉数据密集地区的不明显的空间非平稳性, 自适应加权函数通过改变核的大小以适应局部的观测格局。数据点稀疏的分布, 核有着较大的窗宽, 而在数据密集处有较小的窗宽。通过调整窗

宽的大小，相同数目的非零权被用于分析中的每一个回归点 i 。例如，自适应双平方加权函数如下

$$\omega_{ij} = \begin{cases} [1 - (\frac{d_{ij}}{h_i})^2]^2, & \text{若 } d_{ij} < h_i \\ 0, & \text{其他} \end{cases}$$

其中 h_i 代表位置点 i 处的特定窗宽。

GTWR的权矩阵代表的是在时空位置 i 处，被用来估计参数的数据集中，每一次独立观测的不同的严重程度。一般地，越靠近观测点 i ，权值越大。因而，每一个估计点 i 有唯一的权矩阵。

对于GTWR模型，假定在时空坐标系中距离点 i 近的观测数据对 $\beta_k(u_i, v_i, t_i)$ 估计的影响比距离点 i 远的数据点大。在这种情形下，“近”的定义可以由两个变量合并而成：时间封闭性和空间封闭性。因此，这种所谓的封闭性的定义和度量就成了GTWR模型中的关键问题。

假定观测数据都落在一个三维时空坐标系中且这些点都靠近位置 i 。例如，如果这个时间坐标系在距离上有相同的尺度效应，我们可以以特定的回归点 i 为中心，以 r 为半径画出一个球体，并且使用普通最小二乘方法仅用球体内的观测值来校准回归模型。 $\beta_k(u_i, v_i, t_i)$ 可以被看作 i 及其周围变量结合的估计。

然而，地理位置和时间通常使用不同的单位（如在本文的实例中，地理位置用千米，时间用月），因此它们有不同的尺度效应。使用椭球坐标系来度量回归点与其周围观测点的“近”更加合适。模拟时间距离的一种简单和直接方法就是将其与空间距离整合到时空距离函数中。

给一个空间距离 d^S 和一个时间距离 d^T ，结合它们建立一个时空距离 d^{ST} ，如

$$d^{ST} = d^S \otimes d^T \quad (2.16)$$

其中 \otimes 可以代表不同的操作。如果“+”运算被用来度量整个时空距离 d^{ST} ，则它可以被表示为 d^S 与 d^T 的一个线性组合

$$d^{ST} = \lambda d^S + \mu d^T \quad (2.17)$$

其中 λ 和 μ 为平衡各自度量系统中被用来度量其空间距离和时间距离的尺度因素。因而，如果适当地调整参数 λ 和 μ ， d^{ST} 可以被用来度量时空空间中扩展的“近”。需要指出的是，尺度因素是必须的，否则，如果 d^S 远大于 d^T ，则 d^{ST} 将被 d^S 主导，这样将降低时间的影响，反之亦然。

由公式(2.17)，如果用欧式距离和高斯核函数来构造时空加权矩阵，则有 (u_0, v_0, t_0) 与观测点 (u_i, v_i, t_i) ， $i = 1, 2, \dots, n$ 的时空距离为：

$$d_{0i} = \sqrt{\lambda[(u_0 - u_i)^2 + (v_0 - v_i)^2] + \mu(t_0 - t_i)^2} \quad (2.18)$$

其中 t_0 和 t_i 是在地理位置 i 和 j 处的观测时间。

$$\begin{aligned}
 \omega_i(u_0, \nu_0, t_0) &= \exp\left(-\frac{d_{0i}^2}{h^2}\right) \\
 &= \exp\left\{-\frac{\lambda[(u_0 - u_i)^2 + (\nu_0 - \nu_i)^2] + \mu(t_0 - t_i)^2}{h^2}\right\} \\
 &= \exp\left\{-\frac{(u_0 - u_i)^2 + (\nu_0 - \nu_i)^2}{h_1^2}\right\} \exp\left\{-\frac{(t_0 - t_i)^2}{h_2^2}\right\} \\
 &= \omega_i(u_0, \nu_0) \cdot \omega_i(t_0)
 \end{aligned} \tag{2.19}$$

其中 h 为时空窗宽参数, $h_1 = \sqrt{\frac{h^2}{\lambda}}$ 和 $h_2 = \sqrt{\frac{h^2}{\mu}}$ 分别为空间和时间窗宽参数。

依据公式(2.18),计算出位置 i 和其他观测值之间的距离后,其权函数已可以被构造出。理论上,如果观测数据中没有时间变量,则参数 μ 可以被设为0,进而模型降为传统的GWR模型。另一方面,如果参数 λ 被设为0,只需考虑时间距离和时间非平稳性,这会导致模型降为TWR模型。然而,在大多数情况下, μ 和 λ 均不为0,且空间和时间距离均可用公式(2.18)计算。

从上述可以看出要确定权值,需确定窗宽参数 h_1 和 h_2 。在此,利用交叉确认法^[37-39]来确定前面所给的权函数中的窗宽参数 h_1 和 h_2 的值。令

$$CV(h_1, h_2) = \sum_{i=1}^n [y_i - \hat{y}_{(i)}(h_1, h_2)]^2 \tag{2.20}$$

其中 y_i 为因变量 Y 在观测点 (u_i, ν_i, t_i) 处的观测值, $\hat{y}_{(i)}(h_1, h_2)$ 为在给定 h_1 和 h_2 之下,去掉在 (u_i, ν_i, t_i) 处的观测值 $(y_i; x_{i1}, x_{i2}, \dots, x_{id})$ 后,利用上述拟合方法求得的因变量 Y 在观测点 (u_i, ν_i, t_i) 处的拟合值(即在上述过程中取 $(u_0, \nu_0, t_0) = (u_i, \nu_i, t_i)$)。选择 h_{10} 和 h_{20} 使得

$$CV(h_{10}, h_{20}) = \min_{h_1 > 0, h_2 > 0} CV(h_1, h_2) \tag{2.21}$$

则 h_{10} 和 h_{20} 称之为由交叉确认法确定的窗宽参数值。

在实际计算中,可以在 h_1 和 h_2 的定义域内选择一系列的实数对,分别计算 $CV(h_1, h_2)$ 的值,找到 $CV(h_1, h_2)$ 最小值所对应的 h_{10} 和 h_{20} 。

第 3 章 时空加权回归模型的估计

3.1 局部线性估计及其性质

第二章中时空加权回归模型的估计方法,其本质上属于Nadaraya-Watson核估计方法。我们知道Nadaraya-Watson核估计方法存在边界效应。而对于时空加权回归模型,其系数函数是时空坐标的三元函数,因此可以想象GTWR方法的边界效应会更加严重。这种严重的边界效应会使得系数函数在边界区域的估计严重失真,从而导致不真实的分析结果。考虑到局部线性估计的方差与Nadaraya-Watson核估计的一样,但其偏差比Nadaraya-Watson核估计小,且局部线性估计具有自动更正边界效应等诸多优点。因此,本节结合局部线性估计和GTWR方法给出了时空加权回归模型的局部估计方法。即将系数函数局部展开为时空坐标的线性函数,再利用前述的GTWR拟合方法得到系数函数的估计。

3.1.1 局部线性估计

为表示方便,我们将时空加权回归模型(1.2)改写成如下形式:

$$y_i = \sum_{j=0}^p \beta_j(u_i, v_i, t_i) x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.1)$$

只需假定 $x_{i0} \equiv 1$,即可使模型(3.1)包含时空变截距项。 (u_i, v_i, t_i) 为椭球坐标系下的任一点的时空坐标,设模型(3.1)中的每个系数函数 $\beta_j(u, v, t) (j = 0, 1, \dots, p)$ 关于空间位置坐标 u, v 和时间坐标 t 均有连续的偏导数。 (u_0, v_0, t_0) 为研究区域内给定的任一点,对每个 $j = 0, 1, \dots, p$,由Taylor公式知,在 (u_0, v_0, t_0) 的邻域内可用 u, v 和 t 的线性函数逼近,即有

$$\begin{aligned} \beta_j(u, v, t) \approx & \beta_j(u_0, v_0, t_0) + \beta_j^{(u)}(u_0, v_0, t_0)(u - u_0) + \beta_j^{(v)}(u_0, v_0, t_0)(v - v_0) \\ & + \beta_j^{(t)}(u_0, v_0, t_0)(t - t_0) \end{aligned}$$

其中 $\beta_j^{(u)}(u_0, v_0, t_0), \beta_j^{(v)}(u_0, v_0, t_0)$ 和 $\beta_j^{(t)}(u_0, v_0, t_0)$ 分别表示 $\beta_j(u, v, t)$ 关于 u, v 和 t 的偏导数在 (u_0, v_0, t_0) 处的值。根据变系数模型的局部线性拟合及GTWR方法,极小化以下式子

$$\begin{aligned} \sum_{i=1}^n \{y_i - \sum_{j=0}^p [\beta_j(u_0, v_0, t_0) + \beta_j^{(u)}(u_0, v_0, t_0)(u_i - u_0) + \beta_j^{(v)}(u_0, v_0, t_0)(v_i - v_0) \\ + \beta_j^{(t)}(u_0, v_0, t_0)(t_i - t_0)] X_{ij}\}^2 \omega_i(u_0, v_0, t_0) \end{aligned} \quad (3.2)$$

其中的符号涵义与第二章中相同。若令

$$X_1(u_0, v_0, t_0) = \begin{bmatrix} x_{10} & \cdots & x_{1p} & x_{10}(u_1 - u_0) & \cdots & x_{1p}(u_1 - u_0) \\ x_{20} & \cdots & x_{2p} & x_{20}(u_2 - u_0) & \cdots & x_{2p}(u_2 - u_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n0} & \cdots & x_{np} & x_{n0}(u_n - u_0) & \cdots & x_{np}(u_n - u_0) \end{bmatrix}$$

$$X_2(u_0, v_0, t_0) = \begin{bmatrix} x_{10}(v_1 - v_0) & \cdots & x_{1p}(v_1 - v_0) & x_{10}(t_1 - t_0) & \cdots & x_{1p}(t_1 - t_0) \\ x_{20}(v_2 - v_0) & \cdots & x_{2p}(v_2 - v_0) & x_{20}(t_2 - t_0) & \cdots & x_{2p}(t_2 - t_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n0}(v_n - v_0) & \cdots & x_{np}(v_n - v_0) & x_{n0}(t_n - t_0) & \cdots & x_{np}(t_n - t_0) \end{bmatrix}$$

则

$$X(u_0, v_0, t_0) = (X_1(u_0, v_0, t_0), X_2(u_0, v_0, t_0))$$

$$P(u_0, v_0, t_0) = (\beta_0(u_0, v_0, t_0), \cdots, \beta_p(u_0, v_0, t_0), \beta_0^{(u)}(u_0, v_0, t_0), \cdots, \beta_p^{(u)}(u_0, v_0, t_0), \\ \beta_0^{(v)}(u_0, v_0, t_0), \cdots, \beta_p^{(v)}(u_0, v_0, t_0), \beta_0^{(t)}(u_0, v_0, t_0), \cdots, \beta_p^{(t)}(u_0, v_0, t_0))^T$$

则上述最小二乘问题的解可用矩阵表示为:

$$\begin{aligned} \hat{P}(u_0, v_0, t_0) &= (\hat{\beta}_0(u_0, v_0, t_0), \cdots, \hat{\beta}_p(u_0, v_0, t_0), \hat{\beta}_0^{(u)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(u)}(u_0, v_0, t_0), \\ &\hat{\beta}_0^{(v)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(v)}(u_0, v_0, t_0), \hat{\beta}_0^{(t)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(t)}(u_0, v_0, t_0))^T \\ &= [X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1}X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \end{aligned} \quad (3.3)$$

其中 $W(u_0, v_0, t_0)$ 与 Y 分别与第二章中的相应量相同. 记

$$\begin{aligned} \hat{\beta}(u_0, v_0, t_0) &= (\hat{\beta}_0(u_0, v_0, t_0), \hat{\beta}_1(u_0, v_0, t_0), \cdots, \hat{\beta}_p(u_0, v_0, t_0))^T \\ \hat{\beta}^{(u)}(u_0, v_0, t_0) &= (\hat{\beta}_0^{(u)}(u_0, v_0, t_0), \hat{\beta}_1^{(u)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(u)}(u_0, v_0, t_0))^T \\ \hat{\beta}^{(v)}(u_0, v_0, t_0) &= (\hat{\beta}_0^{(v)}(u_0, v_0, t_0), \hat{\beta}_1^{(v)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(v)}(u_0, v_0, t_0))^T \\ \hat{\beta}^{(t)}(u_0, v_0, t_0) &= (\hat{\beta}_0^{(t)}(u_0, v_0, t_0), \hat{\beta}_1^{(t)}(u_0, v_0, t_0), \cdots, \hat{\beta}_p^{(t)}(u_0, v_0, t_0))^T \end{aligned}$$

分别为各回归系数函数 $\beta_j(u, v, t)$ ($j = 0, 1, \cdots, p$)在 (u_0, v_0, t_0) 处的估计值所构成的列向量以及关于 u, v 和 t 的偏导数的估计值所构成的列向量, 则由(3.3)式可得

$$\begin{aligned} \hat{\beta}(u_0, v_0, t_0) &= (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ &\quad X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \end{aligned} \quad (3.4)$$

$$\hat{\beta}^{(u)}(u_0, v_0, t_0) = (0_{p+1}, I_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \quad (3.5)$$

$$\hat{\beta}^{(v)}(u_0, v_0, t_0) = (0_{p+1}, 0_{p+1}, I_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \quad (3.6)$$

$$\hat{\beta}^{(t)}(u_0, v_0, t_0) = (0_{p+1}, 0_{p+1}, 0_{p+1}, I_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \quad (3.7)$$

其中 I_{p+1} 和 0_{p+1} 分别表示 $p+1$ 阶单位阵和 $p+1$ 阶零矩阵, 我们称上述方法为时空加权回归模型的局部线性估计方法。此估计方法不仅可以得到各系数函数的估计值, 同时还能得到系数函数在空间位置坐标 u, v 和时间坐标 t 的偏导数的估计值。

分别取 $(u_0, v_0, t_0) = (u_i, v_i, t_i) (i = 1, 2, \dots, n)$, 则由(3.4)式容易得到系数函数在各观测位置处的估计值

$$\hat{\beta}(u_i, v_i, t_i) = (\hat{\beta}_0(u_i, v_i, t_i), \hat{\beta}_1(u_i, v_i, t_i), \dots, \hat{\beta}_p(u_i, v_i, t_i))^T \\ = (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_i, v_i, t_i)W(u_i, v_i, t_i)X(u_i, v_i, t_i)]^{-1} \\ X^T(u_i, v_i, t_i)W(u_i, v_i, t_i)Y, \quad i = 1, 2, \dots, n. \quad (3.8)$$

因此, 可得因变量在各观测位置 (u_i, v_i, t_i) 处的拟合值为

$$\hat{Y}_i = X_i \hat{\beta}(u_i, v_i, t_i) \\ = (X_i, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y, \quad i = 1, 2, \dots, n. \quad (3.9)$$

其中 $X_i = (1, x_{i1}, \dots, x_{ip})^T$ 为自变量 X_1, X_2, \dots, X_p 的第 i 组观测值及 X_0 所构成的列向量。从而有因变量 Y 在个观测位置处的拟合值

$$\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)^T = LY \quad (3.10)$$

其中

$$L = \begin{bmatrix} (X_1^T, 0_{p+1}, 0_{p+1}, 0_{p+1})(X^T(u_1, v_1, t_1)W(u_1, v_1, t_1)X(u_1, v_1, t_1))^{-1}X^T(u_1, v_1, t_1)W(u_1, v_1, t_1) \\ (X_2^T, 0_{p+1}, 0_{p+1}, 0_{p+1})(X^T(u_2, v_2, t_2)W(u_2, v_2, t_2)X(u_2, v_2, t_2))^{-1}X^T(u_2, v_2, t_2)W(u_2, v_2, t_2) \\ \vdots \\ (X_n^T, 0_{p+1}, 0_{p+1}, 0_{p+1})(X^T(u_n, v_n, t_n)W(u_n, v_n, t_n)X(u_n, v_n, t_n))^{-1}X^T(u_n, v_n, t_n)W(u_n, v_n, t_n) \end{bmatrix}$$

由(3.5)式可知, 局部线性估计是以 L 为光滑矩阵的线性估计, 即因变量的拟合值为其 n 个观测值的线性组合。进一步可得局部线性估计的残差向量为

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n)^T = Y - \hat{Y} = (I - L)Y \quad (3.11)$$

残差平方和为

$$RSS = \hat{\varepsilon}^T \hat{\varepsilon} = Y^T (I - L)^T (I - L) Y$$

由此可得模型(3.1)的误差方差 $Var(\varepsilon_i) = \sigma^2$ 的估计为

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{tr((I - L)^T (I - L))} \\ &= \frac{Y^T (I - L)^T (I - L) Y}{tr((I - L)^T (I - L))} \end{aligned} \quad (3.12)$$

由于局部线性估计对窗宽的敏感性很强。窗宽太大，会引起大的估计偏差；窗宽太小，会引起大的估计方差。所以取一个合适的窗宽是相当重要的。前述的GTWR方法中关于窗宽的选择方法亦可以应用于本章节中窗宽的选择，即窗宽亦可由交叉确认法来确定。

3.1.2 局部线性估计的性质

本小节给出时空加权回归模型的局部线性估计的一个有限样本性质，并通过与本文第二章中的最小二乘方法比较说明局部线性估计在系数函数及误差方差方面的改进效果。

定理 3.1 在时空加权回归模型(3.1)中，如果所有系数函数 $\beta_j(u, v, t) (j = 0, 1, \dots, p)$ 均为时空坐标 (u, v, t) 的线性函数，则对于区域中的任一点 (u_0, v_0, t_0) 和任何窗宽参数值 h ，有

$$E(\hat{P}(u_0, v_0, t_0)) = P(u_0, v_0, t_0), \quad E(\hat{\sigma}^2) = \sigma^2$$

其中 $P(u_0, v_0, t_0)$ 和 $\hat{P}(u_0, v_0, t_0)$ 与前述一致。

证明 先证 $E(\hat{P}(u_0, v_0, t_0)) = P(u_0, v_0, t_0)$ 。设

$$\beta_j(u, v, t) = a_j + b_j u + c_j v + m_j t, \quad (j = 0, 1, \dots, p)$$

其中 a_j, b_j, c_j 和 m_j 均为常数。此时，对研究区域内任一点 (u_0, v_0, t_0) ，有

$$\begin{aligned} \beta_j(u, v, t) &= \beta_j(u_0, v_0, t_0) + \beta_j^{(u)}(u_0, v_0, t_0)(u - u_0) + \beta_j^{(v)}(u_0, v_0, t_0)(v - v_0) \\ &\quad + \beta_j^{(t)}(u_0, v_0, t_0)(t - t_0), \quad (j = 0, 1, \dots, p) \end{aligned}$$

其中

$$\begin{aligned} \beta_j(u_0, v_0, t_0) &= a_j + b_j u_0 + c_j v_0 + m_j t_0, \quad \beta_j^{(u)}(u_0, v_0, t_0) = b_j, \\ \beta_j^{(v)}(u_0, v_0, t_0) &= c_j, \quad \beta_j^{(t)}(u_0, v_0, t_0) = m_j, \quad (j = 0, 1, \dots, p) \end{aligned}$$

故对于每个 $i = 1, \dots, n$, 有

$$\begin{aligned} y_i &= \sum_{j=0}^p \beta_j(u_i, v_i, t_i) x_{ij} + \varepsilon_i \\ &= \sum_{j=0}^p [\beta_j(u_0, v_0, t_0) + \beta_j^{(u)}(u_0, v_0, t_0)(u_i - u_0) + \beta_j^{(v)}(u_0, v_0, t_0)(v_i - v_0) \\ &\quad + \beta_j^{(t)}(u_0, v_0, t_0)(t_i - t_0)] x_{ij} + \varepsilon_i \end{aligned}$$

从而

$$Y = (Y_1, Y_2, \dots, Y_n)^T = X(u_0, v_0, t_0)P(u_0, v_0, t_0) + \varepsilon$$

其中 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ 为误差向量, 满足 $E(\varepsilon) = 0$. 故有

$$E(Y) = X(u_0, v_0, t_0)P(u_0, v_0, t_0)$$

因此, 对于任意的窗宽参数 h , 由(3.3)式可得

$$\begin{aligned} &E(\hat{P}(u_0, v_0, t_0)) \\ &= [X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)E(Y) \\ &= P(u_0, v_0, t_0) \end{aligned}$$

再证 $E(\hat{\sigma}^2) = \sigma^2$. 由 $\hat{P}(u_0, v_0, t_0)$ 和 $P(u_0, v_0, t_0)$ 的定义及 $E(\hat{P}(u_0, v_0, t_0)) = P(u_0, v_0, t_0)$ 可知, 对每个 $i = 1, \dots, n$, 由于

$$\hat{Y}_i = \sum_{j=0}^p \hat{\beta}_j(u_i, v_i, t_i) X_{ij}$$

故

$$\begin{aligned} E(\hat{Y}_i) &= \sum_{j=0}^p E(\hat{\beta}_j(u_i, v_i, t_i)) X_{ij} \\ &= \sum_{j=0}^p \beta_j(u_i, v_i, t_i) X_{ij} = E(Y_i) \end{aligned}$$

从而有

$$E(\hat{Y}_i) = E(Y_i)$$

这时

$$\begin{aligned} \hat{\varepsilon} &= Y - \hat{Y} \\ &= Y - E(Y) - (\hat{Y} - E(\hat{Y})) \\ &= Y - E(Y) - (LY - E(LY)) \\ &= (I - L)(Y - E(Y)) \\ &= (I - L)\varepsilon \end{aligned}$$

故

$$Y^T(I-L)^T(I-L)Y = \varepsilon^T \hat{\varepsilon} = \varepsilon^T(I-L)^T(I-L)\varepsilon$$

由此可得

$$\begin{aligned} E(Y^T(I-L)^T(I-L)Y) \\ &= E(\varepsilon^T(I-L)^T(I-L)\varepsilon) \\ &= \text{tr}((I-L)^T(I-L)E(\varepsilon^T\varepsilon)) \\ &= \sigma^2 \text{tr}((I-L)^T(I-L)) \end{aligned}$$

因此

$$E(\hat{\sigma}^2) = \frac{E(Y^T(I-L)^T(I-L)Y)}{\text{tr}((I-L)^T(I-L))} = \sigma^2$$

此定理说明, 当时空加权回归模型的系数函数均为时空坐标的线性函数时, 对任意的窗宽参数值 h , 系数函数及其偏导数在区域内任一点的局部线性估计均为相应量的无偏估计, 同时(3.12)式所给出的模型误差项方差的估计也是它的一个无偏估计。为方便比较, 下面给出最小二乘估计的有关结果。

定理 3.2 在时空加权回归模型(3.1)中, 若系数函数 $\beta_j(u, v, t)(j = 0, 1, \dots, p)$ 为常数, 即 $\beta_j(u, v, t) = \beta_j(j = 0, 1, \dots, p)$, 则对于区域中的任一点 (u_0, v_0, t_0) 和任何窗宽参数值 h , 系数及误差的加权最小二乘估计满足

$$E(\hat{\beta}(u_0, v_0, t_0)) = \beta = (\beta_1, \beta_2, \dots, \beta_p)^T, \quad E(\hat{\sigma}^2) = \sigma^2$$

其中 $\hat{\beta}(u_0, v_0, t_0)$ 和 $\hat{\sigma}^2$ 与前述第二章中符号一致。

证明 先证明 $E(\hat{\beta}(u_0, v_0, t_0)) = \beta$ 。事实上, 当 $\beta_j(u, v, t) = \beta_j(j = 0, 1, \dots, p)$ 时, 时空加权回归模型退化为通常的线性模型

$$Y = X\beta + \varepsilon$$

故有 $E(Y) = X\beta$ 。对于任一点 (u_0, v_0, t_0) 和窗宽参数 h , 由(2.7)式得

$$E(\hat{\beta}(u_0, v_0, t_0)) = [X^T W(u_0, v_0, t_0) X]^{-1} X^T W(u_0, v_0, t_0) E(Y) = \beta$$

即系数函数在任一点 (u_0, v_0, t_0) 处的加权最小二乘估计是无偏估计。

再证 $E(\hat{\sigma}^2) = \sigma^2$ 。此时有 $E(\hat{\beta}(u_i, v_i, t_i)) = \beta(u_i, v_i, t_i) = (\beta_1, \beta_2, \dots, \beta_p)^T$, 故由

$$\hat{Y}_i = \sum_{j=0}^p \hat{\beta}_j(u_i, v_i, t_i) X_{ij}, \quad i = 1, \dots, n$$

立得

$$E(\hat{Y}_i) = \sum_{j=0}^p \beta_j x_{ij} = E(Y_i), \quad i = 1, \dots, n$$

即

$$E(\hat{Y}) = E(Y)$$

这时

$$\begin{aligned}\hat{\varepsilon} &= Y - \hat{Y} = Y - E(Y) - (\hat{Y} - E(\hat{Y})) \\ &= (I - S)(Y - E(Y)) = (I - S)\varepsilon\end{aligned}$$

同定理3.1的证明可得

$$E(\hat{\sigma}^2) = \frac{E(\varepsilon^T(I - S)^T(I - S)\varepsilon)}{\text{tr}((I - S)^T(I - S))} = \sigma^2$$

由于常值函数只是线性函数的特例, 因此由定理3.1和定理3.2知, 时空加权回归模型的局部线性估计和本文第二章中所给出的最小二乘方法相比, 在更广泛的系数函数类型下, 可以得到系数函数和误差方差的无偏估计。从这个意义上讲, 局部线性估计改进了原有的最小二乘方法, 在更广泛的系数函数类型下, 可以得到系数函数本身和误差方差的无偏估计。

3.2 二步估计及其性质

从非参数回归模型的核光滑方法的拟合中我们可以看到, 当回归函数的变化较为平缓时, 即其光滑度较大时, 相应的窗宽参数值会较大。相反, 当回归函数变化较为剧烈时, 即其光滑度较小时, 相应的窗宽参数值会较小。同样的, 在时空加权回归模型中, 若各系数的光滑度较为一致, 则可选择一个全局窗宽参数值进行模型拟合。而当时空加权回归模型各系数函数的光滑度有较大差异时, 直观上, 较为光滑的系数函数需要一个较大的窗宽参数, 而光滑度低的系数函数需要一个较小的窗宽参数, 因此用一个全局的光滑参数值就不可能同时得到所有系数函数的满意的估计, 上述基于单一窗宽参数的局部线性估计方法就不能同时得到各系数函数的最优估计, 需要进一步的优化改进。

3.2.1 二步估计

假设在模型(3.1)中, $\beta_p(u, v, t)$ 比其他系数函数具有更高的光滑度, 且具有连续的三阶导数, 而其他系数函数 $\beta_j(u, v, t)$ ($j = 0, 1, \dots, p-1$)具有大致相同的光滑度, 且均有连续的一阶导数, 二步估计的目的是给出 $\beta_p(u, v, t)$ 的更优估计^[41]。

对研究区域内的任一给定的点 (u_0, v_0, t_0) , 在一较小的窗宽参数 h_0 下, 利用局部线性拟合方法得到 $\beta(u_0, v_0, t_0) = (\beta_0(u_0, v_0, t_0), \dots, \beta_p(u_0, v_0, t_0))^T$ 的初始估计如(3.4)式

所示, 在这里记为

$$\begin{aligned}\tilde{\beta}(u_0, v_0, t_0) &= (\hat{\beta}_0(u_0, v_0, t_0), \hat{\beta}_1(u_0, v_0, t_0), \dots, \hat{\beta}_p(u_0, v_0, t_0))^T \\ &= (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ &\quad X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)Y\end{aligned}\quad (3.13)$$

其中 $W_{h_0}(u_0, v_0, t_0)$ 为窗宽参数是 h_0 的权矩阵。其他记号与前述一致。

将 $\tilde{\beta}_j(u, v, t)(j = 0, 1, \dots, p-1)$ 代入时空加权回归模型(3.1)中, 得到

$$\tilde{y}_i = y_i - \sum_{j=0}^{p-1} \tilde{\beta}_j(u_i, v_i, t_i)x_{ij} = \beta_p(u_i, v_i, t_i)x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

由于 $\beta_p(u, v, t)$ 有连续的三阶导数, 因此由Taylor公式, 在 (u_0, v_0, t_0) 的邻域内有

$$\begin{aligned}\beta_p(u, v, t) &\approx \beta_p(u_0, v_0, t_0) + \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(u)}(u_0, v_0, t_0)(u - u_0)^k \\ &\quad + \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(v)}(u_0, v_0, t_0)(v - v_0)^k + \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(t)}(u_0, v_0, t_0)(t - t_0)^k\end{aligned}\quad (3.14)$$

$$\begin{aligned}&\sum_{i=1}^n \{ \tilde{x}_i - [\beta_p(u_0, v_0, t_0) - \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(u)}(u_0, v_0, t_0)(u_i - u_0)^k - \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(v)}(u_0, v_0, t_0) \\ &\quad \times (v_i - v_0)^k - \sum_{k=1}^3 \frac{1}{k!} \beta_p^{(k)(t)}(u_0, v_0, t_0)(t_i - t_0)^k] x_{ip} \}^2 \omega_i(u_0, v_0, t_0)\end{aligned}$$

解此最小二乘问题, 得到 $\beta_p(u_0, v_0, t_0)$ 的估计值。则 $\beta_p(u, v, t)$ 在点 (u_0, v_0, t_0) 的二步估计为

$$\begin{aligned}\hat{\beta}_p(u_0, v_0, t_0) &= e_{1,10}^T [G^T(u_0, v_0, t_0)W_{h_1}(u_0, v_0, t_0)G(u_0, v_0, t_0)]^{-1} \\ &\quad G^T(u_0, v_0, t_0)W_{h_1}(u_0, v_0, t_0)\tilde{Y}\end{aligned}\quad (3.15)$$

其中 $W_{h_1}(u_0, v_0, t_0)$ 为窗口参数是 h_1 的权矩阵, $e_{1,10}$ 表示第1个元素为1其余元素为0的10维列向量。

$$\tilde{Y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)^T, \quad \tilde{y}_i = y_i - \sum_{j=0}^{p-1} \tilde{\beta}_j(u_i, v_i, t_i)x_{ij}$$

$$G_1 = \begin{bmatrix} x_{1p} & x_{1p}(u_1 - u_0) & \cdots & x_{1p}(u_1 - u_0)^3 \\ x_{2p} & x_{2p}(u_2 - u_0) & \cdots & x_{2p}(u_2 - u_0)^3 \\ \vdots & \vdots & \ddots & \vdots \\ x_{np} & x_{np}(u_n - u_0) & \cdots & x_{np}(u_n - u_0)^3 \end{bmatrix}$$

$$G_2 = \begin{bmatrix} x_{1p}(v_1 - v_0) & \cdots & x_{1p}(v_1 - v_0)^3 & x_{1p}(t_1 - t_0) & \cdots & x_{1p}(t_1 - t_0)^3 \\ x_{2p}(v_2 - v_0) & \cdots & x_{2p}(v_2 - v_0)^3 & x_{2p}(t_2 - t_0) & \cdots & x_{2p}(t_2 - t_0)^3 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{np}(v_n - v_0) & \cdots & x_{np}(v_n - v_0)^3 & x_{np}(t_n - t_0) & \cdots & x_{np}(t_n - t_0)^3 \end{bmatrix}$$

则

$$G(u_0, v_0, t_0) = (G_1, G_2)$$

上述二步估计虽然比一步局部线性估计具有更好的效果,但其有着拟合过程中计算量大的缺点,故为减少计算量,我们在二步估计的第二步中也采取局部线性估计拟合,这样得到的系数的二步估计与利用局部三次多项式拟合的估计具有几乎同样好的效果。第一步的估计方法与前述的二步估计中的第一步完全相同,在第二步估计中,对 $\beta_p(u, v, t)$,在 (u_0, v_0, t_0) 的邻域内用 u, v 和 t 的线性函数逼近,即有

$$\begin{aligned} \beta_p(u, v, t) \approx & \beta_p(u_0, v_0, t_0) + \beta_p^{(u)}(u_0, v_0, t_0)(u - u_0) + \beta_p^{(v)}(u_0, v_0, t_0)(v - v_0) \\ & + \beta_p^{(t)}(u_0, v_0, t_0)(t - t_0) \end{aligned}$$

其中 $\beta_p^{(u)}(u_0, v_0, t_0), \beta_p^{(v)}(u_0, v_0, t_0)$ 和 $\beta_p^{(t)}(u_0, v_0, t_0)$ 分别表示 $\beta_p(u, v, t)$ 关于 u, v 和 t 的偏导数在 (u_0, v_0, t_0) 处的值。根据变系数模型的局部线性拟合及本文第二章中给出的拟合方法,极小化

$$\begin{aligned} \sum_{i=1}^n \{ & \tilde{y}_i - [\beta_p(u_0, v_0, t_0) + \beta_p^{(u)}(u_0, v_0, t_0)(u_i - u_0) + \beta_p^{(v)}(u_0, v_0, t_0)(v_i - v_0) \\ & + \beta_p^{(t)}(u_0, v_0, t_0)(t_i - t_0)] x_{ij} \}^2 \omega_i(u_0, v_0, t_0) \end{aligned}$$

解此最小二乘问题,由局部线性估计可知

$$\begin{aligned} \hat{\beta}_p(u_0, v_0, t_0) = & e_{1,10}^T [X^T(u_0, v_0, t_0) W_{h_2}(u_0, v_0, t_0) X(u_0, v_0, t_0)]^{-1} \\ & X^T(u_0, v_0, t_0) W_{h_2}(u_0, v_0, t_0) \bar{Y} \end{aligned} \quad (3.16)$$

其中 $W_{h_2}(u_0, v_0, t_0)$ 为窗宽参数是 h_2 的权矩阵,其余几号与前述相同。

3.2.2 二步估计的性质

定理 3.3 在时空加权回归模型(3.1)中, 如果所有系数函数 $\beta_j(u, v, t) (j = 0, 1, \dots, p)$ 在其定义域内均有连续的三阶导数, 则对于区域中的任一点 (u_0, v_0, t_0) , 局部线性拟合的系数函数的条件偏近似为

$$\begin{aligned} & \text{Bias}(\hat{\beta}(u_0, v_0, t_0) \mid \Phi) \\ & \approx (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)\tau(u_0, v_0, t_0) \end{aligned}$$

且可得系数函数的条件偏的估计为

$$\begin{aligned} & \hat{\text{Bias}}(\hat{\beta}(u_0, v_0, t_0) \mid \Phi) \\ & = (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)\hat{\tau}(u_0, v_0, t_0) \end{aligned}$$

其中 Φ 如证明中(3.17)式所示, $\tau(u_0, v_0, t_0)$ 如(3.22)式所示, $\hat{\tau}(u_0, v_0, t_0)$ 为 $\tau(u_0, v_0, t_0)$ 的估计。

证明 由于各系数函数 $\beta_j(u, v, t) (j = 0, 1, \dots, p)$ 在其定义域内有连续的三阶导数, 则由Taylor公式, 对每个 $j = 0, 1, \dots, p$, 在 (u_0, v_0, t_0) 的邻域内有

$$\begin{aligned} \beta_p(u, v, t) & \approx \beta_p(u_0, v_0, t_0) + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(u)}(u_0, v_0, t_0)(u - u_0)^k \\ & \quad + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(v)}(u_0, v_0, t_0)(v - v_0)^k + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(t)}(u_0, v_0, t_0)(t - t_0)^k \end{aligned}$$

用

$$\Phi = \{(u_1, v_1, t_1), \dots, (u_n, v_n, t_n), x_{01}, \dots, x_{0n}, \dots, x_{p1}, \dots, x_{pn}\} \quad (3.17)$$

表示自变量观测值集合。由(3.4)式可知系数函数向量 $\beta(u, v, t) = (\beta_0(u, v, t), \beta_1(u, v, t), \dots, \beta_p(u, v, t))^T$ 在 (u_0, v_0, t_0) 处的局部线性估计为

$$\begin{aligned} \hat{\beta}(u_0, v_0, t_0) & = (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)Y \end{aligned}$$

从而

$$\begin{aligned} E(\hat{\beta}(u_0, v_0, t_0) \mid \Phi) & = (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)E(Y \mid \Phi) \end{aligned} \quad (3.18)$$

再注意到 $E(\varepsilon_i | \Phi) = 0 (i = 1, 2, \dots, n)$, 故

$$E(Y | \Phi) = \left(\sum_{j=0}^p \beta_j(u_1, v_1, t_1) x_{1j}, \dots, \sum_{j=0}^p \beta_j(u_n, v_n, t_n) x_{nj} \right)^T \quad (3.19)$$

而对每一个 $i = 1, 2, \dots, n$,

$$\begin{aligned} & \sum_{j=0}^p \beta_j(u_i, v_i, t_i) x_{ij} \\ & \approx \sum_{j=0}^p [\beta_j(u_0, v_0, t_0) + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(u)}(u_0, v_0, t_0) (u_i - u_0)^k + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(v)}(u_0, v_0, t_0) \\ & \quad (v_i - v_0)^k + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(t)}(u_0, v_0, t_0) (t_i - t_0)^k] x_{ij} \end{aligned}$$

将以上各式代入(3.19)式, $X_i = (1, x_{i1}, \dots, x_{ip})^T, (i = 1, 2, \dots, n)$, 则(3.19)式可近似表达为

$$E(Y | \Phi) \approx X(u_0, v_0, t_0)P(u_0, v_0, t_0) + \tau(u_0, v_0, t_0) \quad (3.20)$$

其中

$$\begin{aligned} P(u_0, v_0, t_0) = & (\beta_0(u_0, v_0, t_0), \dots, \beta_p(u_0, v_0, t_0), \beta_0^{(1)(u)}(u_0, v_0, t_0), \dots, \beta_p^{(1)(u)}(u_0, v_0, t_0), \\ & \beta_0^{(1)(v)}(u_0, v_0, t_0), \dots, \beta_p^{(1)(v)}(u_0, v_0, t_0), \\ & \beta_0^{(1)(t)}(u_0, v_0, t_0), \dots, \beta_p^{(1)(t)}(u_0, v_0, t_0))^T \end{aligned} \quad (3.21)$$

若令

$$\begin{aligned} \beta^{(u)} &= \left(\sum_{k=2}^3 \frac{1}{k!} \beta_1^{(k)(u)}(u_0, v_0, t_0), \dots, \sum_{k=2}^3 \frac{1}{k!} \beta_n^{(k)(u)}(u_0, v_0, t_0) \right) \\ \beta^{(v)} &= \left(\sum_{k=2}^3 \frac{1}{k!} \beta_1^{(k)(v)}(u_0, v_0, t_0), \dots, \sum_{k=2}^3 \frac{1}{k!} \beta_n^{(k)(v)}(u_0, v_0, t_0) \right) \\ \beta^{(t)} &= \left(\sum_{k=2}^3 \frac{1}{k!} \beta_1^{(k)(t)}(u_0, v_0, t_0), \dots, \sum_{k=2}^3 \frac{1}{k!} \beta_n^{(k)(t)}(u_0, v_0, t_0) \right) \\ j &= 0, 1, \dots, p, \quad k = 2, 3. \end{aligned}$$

则

$$\tau(u_0, v_0, t_0) = \begin{bmatrix} X_1^T (\beta^{(u)}(u_1 - u_0) + \beta^{(v)}(v_1 - v_0) + \beta^{(t)}(t_1 - t_0)) \\ X_2^T (\beta^{(u)}(u_2 - u_0) + \beta^{(v)}(v_2 - v_0) + \beta^{(t)}(t_2 - t_0)) \\ \vdots \\ X_n^T (\beta^{(u)}(u_n - u_0) + \beta^{(v)}(v_n - v_0) + \beta^{(t)}(t_n - t_0)) \end{bmatrix} \quad (3.22)$$

将(3.21)式代入(3.18)式中, 则可得 $\hat{\beta}(u_0, v_0, t_0)$ 的条件渐进偏近似为

$$\begin{aligned} & \text{Bias}(\hat{\beta}(u_0, v_0, t_0) \mid \Phi) \\ &= E\left(\hat{\beta}(u_0, v_0, t_0) \mid \Phi\right) - \hat{\beta}(u_0, v_0, t_0) \\ &\approx (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ &\quad X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)\tau(u_0, v_0, t_0) \end{aligned} \quad (3.23)$$

其中 $\tau(u_0, v_0, t_0)$ 如(3.22)所示。

如果能给出 $\beta_j^{(k)(u)}(u_0, v_0, t_0)$, $\beta_j^{(k)(v)}(u_0, v_0, t_0)$ 和 $\beta_j^{(k)(t)}(u_0, v_0, t_0)$ ($k=2, 3$) ($j=0, 1, \dots, p$)的估计, 则可得 $\tau(u_0, v_0, t_0)$ 的估计, 进而由(3.23)式得到 $\text{Bias}(\hat{\beta}(u_0, v_0, t_0) \mid \Phi)$ 的估计。为此对时空加权回归模型(3.1)式利用局部三次多项式拟合, 即在 h_T 下, 求解下列局部加权最小二乘问题

$$\begin{aligned} & \sum_{i=1}^n \left\{ y_i - \sum_{j=0}^p [\beta_j(u_0, v_0, t_0) + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(u)}(u_0, v_0, t_0)(u_i - u_0)^k + \sum_{k=1}^3 \frac{1}{k!} \right. \\ & \quad \left. \beta_j^{(k)(v)}(u_0, v_0, t_0)(v_i - v_0)^k + \sum_{k=1}^3 \frac{1}{k!} \beta_j^{(k)(t)}(u_0, v_0, t_0)(t_i - t_0)^k] x_{ij} \right\}^2 \omega_i(u_0, v_0, t_0) \end{aligned}$$

可得 $\beta_j^{(k)(u)}(u_0, v_0, t_0)$, $\beta_j^{(k)(v)}(u_0, v_0, t_0)$ 和 $\beta_j^{(k)(t)}(u_0, v_0, t_0)$ ($k=2, 3$)的估计为

$$\begin{aligned} \beta_j^{(2)(u)}(u_0, v_0, t_0) &= (0_{(p+1) \times 4(p+1)}, I_{p+1}, 0_{(p+1) \times 5(p+1)})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0) \\ &\quad \times \tilde{X}(u_0, v_0, t_0)]^{-1} \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

$$\begin{aligned} \beta_j^{(3)(u)}(u_0, v_0, t_0) &= (0_{(p+1) \times 7(p+1)}, I_{p+1}, 0_{(p+1) \times 2(p+1)})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0) \\ &\quad \times \tilde{X}(u_0, v_0, t_0)]^{-1} \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

$$\begin{aligned} \beta_j^{(2)(v)}(u_0, v_0, t_0) &= (0_{(p+1) \times 5(p+1)}, I_{p+1}, 0_{(p+1) \times 4(p+1)})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0) \\ &\quad \times \tilde{X}(u_0, v_0, t_0)]^{-1} \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

$$\begin{aligned} \beta_j^{(3)(v)}(u_0, v_0, t_0) &= (0_{(p+1) \times 8(p+1)}, I_{p+1}, 0_{p+1})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)\tilde{X}(u_0, v_0, t_0)]^{-1} \\ &\quad \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

$$\begin{aligned} \beta_j^{(2)(t)}(u_0, v_0, t_0) &= (0_{(p+1) \times 6(p+1)}, I_{p+1}, 0_{(p+1) \times 3(p+1)})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0) \\ &\quad \times \tilde{X}(u_0, v_0, t_0)]^{-1} \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

$$\begin{aligned} \beta_j^{(3)(t)}(u_0, v_0, t_0) &= (0_{(p+1) \times 9(p+1)}, I_{p+1})[\tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)\tilde{X}(u_0, v_0, t_0)]^{-1} \\ &\quad \tilde{X}^T(u_0, v_0, t_0)W_{h_T}(u_0, v_0, t_0)Y \end{aligned}$$

令

$$\tilde{X}_1 = \begin{bmatrix} x_{10}(u_1-u_0)^2 & \dots & x_{1p}(u_1-u_0)^2 & x_{10}(v_1-v_0)^2 & \dots & x_{1p}(v_1-v_0)^2 & x_{10}(t_1-t_0)^2 & \dots & x_{1p}(t_1-t_0)^2 \\ x_{20}(u_2-u_0)^2 & \dots & x_{2p}(u_2-u_0)^2 & x_{20}(v_2-v_0)^2 & \dots & x_{2p}(v_2-v_0)^2 & x_{20}(t_2-t_0)^2 & \dots & x_{2p}(t_2-t_0)^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n0}(u_n-u_0)^2 & \dots & x_{np}(u_n-u_0)^2 & x_{n0}(v_n-v_0)^2 & \dots & x_{np}(v_n-v_0)^2 & x_{n0}(t_n-t_0)^2 & \dots & x_{np}(t_n-t_0)^2 \end{bmatrix}$$

$$\tilde{X}_2 = \begin{bmatrix} x_{10}(u_1-u_0)^3 & \dots & x_{1p}(u_1-u_0)^3 & x_{10}(v_1-v_0)^3 & \dots & x_{1p}(v_1-v_0)^3 & x_{10}(t_1-t_0)^3 & \dots & x_{1p}(t_1-t_0)^3 \\ x_{20}(u_2-u_0)^3 & \dots & x_{2p}(u_2-u_0)^3 & x_{20}(v_2-v_0)^3 & \dots & x_{2p}(v_2-v_0)^3 & x_{20}(t_2-t_0)^3 & \dots & x_{2p}(t_2-t_0)^3 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n0}(u_n-u_0)^3 & \dots & x_{np}(u_n-u_0)^3 & x_{n0}(v_n-v_0)^3 & \dots & x_{np}(v_n-v_0)^3 & x_{n0}(t_n-t_0)^3 & \dots & x_{np}(t_n-t_0)^3 \end{bmatrix}$$

则

$$\tilde{X}(u_0, v_0, t_0) = (X(u_0, v_0, t_0), \tilde{X}_1, \tilde{X}_2)$$

$W_{h_T}(u_0, v_0, t_0)$ 是窗宽参数为 h_T 的权矩阵, 这里 h_T 亦可由交叉确认法确定。

将 $\beta_j^{(k)(u)}(u_0, v_0, t_0)$, $\beta_j^{(k)(v)}(u_0, v_0, t_0)$ 和 $\beta_j^{(k)(t)}(u_0, v_0, t_0)$ ($k = 2, 3$)的估计 $\hat{\beta}_j^{(k)(u)}(u_0, v_0, t_0)$, $\hat{\beta}_j^{(k)(v)}(u_0, v_0, t_0)$ 和 $\hat{\beta}_j^{(k)(t)}(u_0, v_0, t_0)$ ($k = 2, 3$)代入(3.22)式中得 $\tau(u_0, v_0, t_0)$ 的估计 $\hat{\tau}(u_0, v_0, t_0)$, 从而可得系数函数的局部线性拟合的条件偏的估计为

$$\begin{aligned} & Bias(\hat{\beta}(u_0, v_0, t_0) | \Phi) \\ &= (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W_{h_0}(u_0, v_0, t_0)\hat{\tau}(u_0, v_0, t_0) \end{aligned} \quad (3.24)$$

定理 3.4 在时空加权回归模型(3.1)中, 如果所有系数函数 $\beta_j(u, v, t)$ ($j = 0, 1, \dots, p$)在其定义域内均有连续的三阶导数, 则对于区域中的任一点 (u_0, v_0, t_0) , 局部线性拟合的系数函数的条件方差近似为

$$\begin{aligned} & Var(\hat{\beta}(u_0, v_0, t_0) | \Phi) \\ & \approx (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad X^T(u_0, v_0, t_0)W^2(u_0, v_0, t_0)X(u_0, v_0, t_0)(X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0))^{-1} \\ & \quad (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})^T \sigma^2 \end{aligned}$$

其中 $W(u_0, v_0, t_0)$ 和 σ^2 均同第二章中符号一致, Φ 如式(3.17)。

证明 当给定 Φ 时, ε_i 与 ε_j 条件独立, 且 $E(\varepsilon_i | \Phi) = 0$ ($i = 1, 2, \dots, n$), 故 $E(\varepsilon_i \varepsilon_j | \Phi) = 0$ ($i \neq j$)。而 $E(\varepsilon_i^2 | \Phi) = \sigma^2(u_i, v_i, t_i)$ ($i = 1, 2, \dots, n$), 故由(3.4)式得

$$\begin{aligned} & Var(\hat{\beta}(u_0, v_0, t_0) | \Phi) \\ & \approx (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})[X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & \quad \times X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)\Sigma W(u_0, v_0, t_0)X(u_0, v_0, t_0) \\ & \quad \times [X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1}(I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})^T \end{aligned}$$

其中 $\Sigma = \text{diag}(\sigma^2(u_1, v_1, t_1), \sigma^2(u_2, v_2, t_2), \dots, \sigma^2(u_n, v_n, t_n))$ 。基于局部齐性逼近, 近似有

$$W(u_0, v_0, t_0)\Sigma W(u_0, v_0, t_0) \approx \sigma^2(u_0, v_0, t_0)W^2(u_0, v_0, t_0)$$

从而可得 $\hat{\beta}(u_0, v_0, t_0)$ 的条件方差近似为

$$\begin{aligned} & \text{Var}(\hat{\beta}(u_0, v_0, t_0) \mid \Phi) \\ & \approx (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1}) [X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & X^T(u_0, v_0, t_0)W^2(u_0, v_0, t_0)X(u_0, v_0, t_0) [X^T(u_0, v_0, t_0)W(u_0, v_0, t_0)X(u_0, v_0, t_0)]^{-1} \\ & (I_{p+1}, 0_{p+1}, 0_{p+1}, 0_{p+1})^T \sigma^2(u_0, v_0, t_0) \end{aligned} \quad (3.25)$$

第4章 时空加权回归模型的应用研究

随着中国经济的发展、人民生活水平的提高,环境污染的问题引起越来越多的社会关注。近段时间,雾霾问题最受全国人民的关注,据《中国环境状况公报》显示,我国城市空气质量处在较重的污染水平,北方城市重于南方城市。二氧化硫年均值浓度在3~248微克/米³范围之间,全国年均值为66微克/米³。一半以上的北方城市和三分之一以上的南方城市年均值超过国家二级标准(60微克/米³)。北方城市年均值为72微克/米³;南方城市年均值为60微克/米³。近两年来,空气污染越来越严重再加上新闻媒体对环境监测结果的曝光,让更多的人对环境污染问题加以关注。

多年调查研究表明,城市空气质量好坏与季节以及气象条件的关系十分密切。在冬季采暖期,北京市和北方许多城市的大气污染元凶是燃煤烟雾,其次是汽车尾气,两者的共同作用使空气污染更加严重;而在非采暖期,则以大量的机动车尾气和悬浮颗粒物污染为主。相对于每周或每天而言,当污染源排放量没有大的变化情况下,风、雨、气压、温度等气象条件直接影响着空气质量的好坏,使空气污染指数会有很大的差别。例如1998年9月下旬,北京市不少行人感到眼睛干,嗓子疼,街面空气呛人。这与当时静风天气闷热有密切关系。

空气质量与很多因素有关。首先,大气逆温现象直接影响着大气污染物的扩散。逆温是空气温度随高度增加而增高的大气垂直层结现象。一般来说,冬季逆温层较强较厚,维持时间较长;夏季则相对偏弱。通常在晴朗微风的夜间有逆温现象存在,使低层大气比较稳定,非常不利于污染物扩散。太阳出来后,随着地表温度的升高,使逆温层逐渐消失,大气湍流混合和垂直对流加强,有助于污染物质的扩散。其次,与是否有雨雪有关。自然降雨、降雪对空气污染物能起着清除和冲刷作用。在雨雪作用下,大气中的一些污染气体能够溶解在水中,降低空气中污染气体的浓度,较大的雨雪对空气中污染物粉尘颗粒也起着有效的清除作用。但是需要指出的是,当前空气中的雨水已经不很干净。降水与空气中的二氧化硫等气体混合溶解会形成酸雨,则是大气质量差的另一种表现形式。再从季节角度来说,由于冬季北方降水较少,气候干燥,刮风天气较少,光照较弱,日照时间短,逆温层最厚,且温度较低,大气对流不活跃等不利于空气中污染物质扩散的因素较多。夏季由于太阳辐射很强,大气对流活动旺盛,逆温层的生成存在时间缩短,且降雨天气较多,降雨量很大,对污染物质清除作用明显,使空气污染程度相对减轻。从以上分析可知空气质量与降水量和温度具有一定的相关性。而衡量空气污染程度的标准为空气质量指数,空气质量指数(Air Quality Index,简称AQI)是定量描述空气质量状况的无量纲指数。了解我国空气质量指数的分布特征对于环境的治理和空气质量的提高都有重要的意义。

本章我们分别使用局部线性估计方法和二步估计方法对时空加权回归模型进行

拟合,并基于有关的统计检验结果分析我国空气质量指数(AQI)与降水量和温度关系的时空分布特征。在此需要说明的是,由于本章的主要目的是说明时空加权回归模型在实际的时空数据分析中的应用,因此我们只考虑了降水量和温度对空气质量指数(AQI)的影响,更全面的分析当然还应该考虑风力、气压等其他重要因素。

4.1 数据来源及建立模型

4.1.1 数据来源与预分析

影响空气质量指数的可能因素有很多,比如降水量、日照、温度、气压、风力、相对湿度和污染源排放量等等。一般来说,当污染源排放量没有大的变化情况下,当温度、降水量、气压、风力、相对湿度增加时,空气质量指数会减少。本章只考虑月降水量总和、月平均温度与月平均空气质量指数的关系。

我们选取了我国67个环保重点城市作为站点,收集了各城市2013年共12个月的有关数据,主要包括:月平均空气质量指数、月降水量总和(单位:mm)、月平均温度(单位:°C)以及经纬度。月平均空气质量指数的数据来源于中国环境监测总站,月降水量总和与月平均温度的数据均来自于中国气象科学数据共享服务网。利用谷歌地图查询得到各城市的经纬度值,以表示该城市的地理位置。

如图4.1至图4.6所示,我们给出了我国67个环保重点城市1月份的月平均空气质量指数分布图、北京市12个月的月平均空气质量指数柱形图、67个环保重点城市2月份的月降水量总和分布图、杭州市12个月的月降水量总和柱形图、67个环保重点城市2月份的月平均温度分布图、兰州市12个月的月平均温度柱形图。由图4.1可知,我国67个环保重点城市中位于东南沿海及南方的大部分城市1月份的月平均空气质量指数低,空气质量优良,而我国中东部及华北地区的城市月平均空气质量指数比较高,空气质量状况令人担忧;由图4.2可以看出,相比较于其他月份,北京市1月份的月平均空气质量指数比较高,空气质量状况较差;由图4.3可知,我国67个环保重点城市位于东南沿海及南方的大部分城市2月份的月降水量比较大,而内陆城市的降水量相对较少;由图4.4可以看出杭州市6月份及8月份的降水量较多,其他月份相对较少;由图4.5可知我国67个环保重点城市的月平均温度从南到北、从沿海地区到内陆地区由高到低。

4.1.2 建立模型

以 Y 表示月平均空气质量指数(AQI), X_1 表示月降水量总和(单位:mm), X_2 表示月平均温度(单位:°C),利用67个站点12个月的观测数据建立如下的时空加权回归模

型:

$$Y_i = \beta_0(u_i, v_i, t_i) + \beta_1(u_i, v_i, t_i)X_{i1} + \beta_2(u_i, v_i, t_i)X_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, 67. \quad (4.1)$$

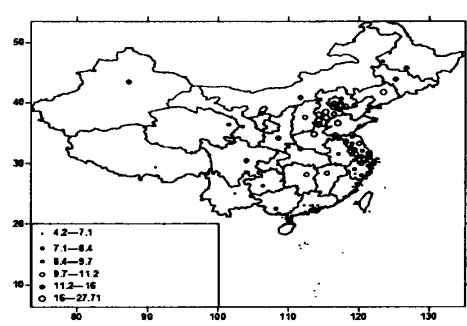


图 4.1 67个城市1月的月平均空气质量指数

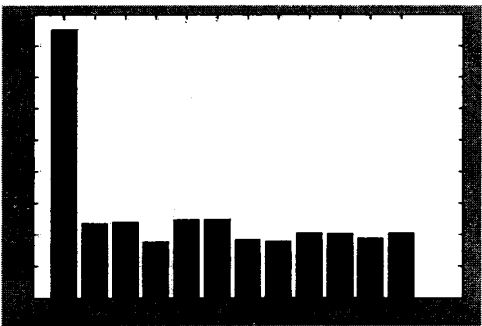


图 4.2 北京市12个月的月平均空气质量指数

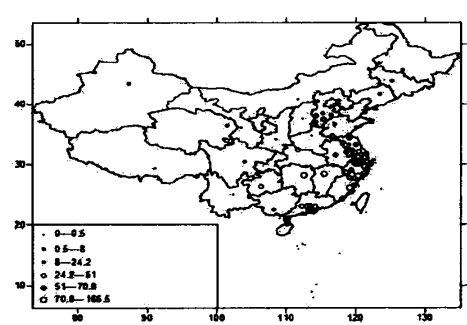


图 4.3 67个城市2月份的月降水量总和

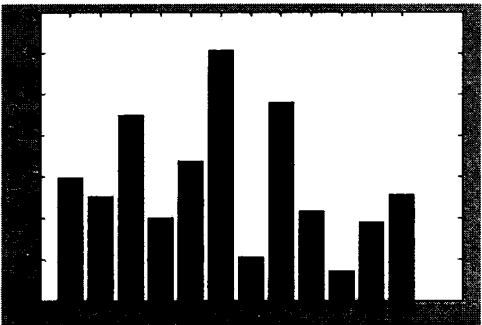


图 4.4 杭州市12个月的月降水量总和

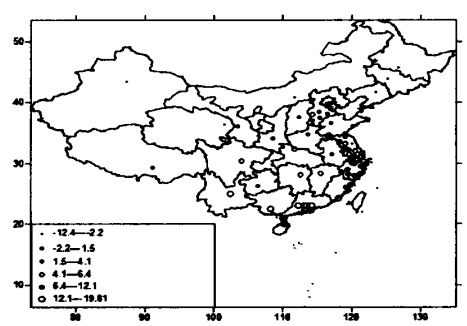


图 4.5 67个城市的2月份的月平均温度

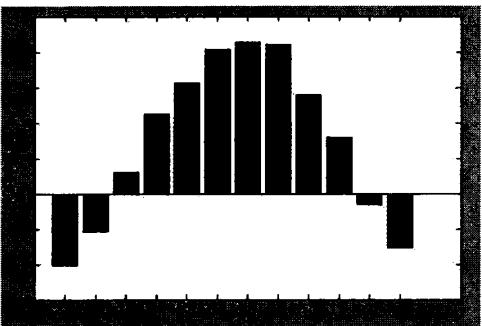


图 4.6 兰州市12个月的月平均温度

其中, $\beta_0(u_i, v_i, t_i)$ 为基准月平均空气质量指数, $\beta_1(u_i, v_i, t_i)$ 表示各站点的月平均空气质量指数随月降水量总和的变化率, 而 $\beta_2(u_i, v_i, t_i)$ 则表示各站点的月平均空气质量指数随月平均温度的变化率。

4.2 利用局部线性估计模拟的结果与分析

根据玄海燕等在文献[35]和梅长林等在文献[37]所给出的显著性检验方法, 构造检验统计量, 使用三阶矩 χ^2 逼近分别求得模型4.1的全局非平稳性检验及各系数函数变化显著性检验的检验 p 值, 见表4.1。其中 p 为模型回归关系的全局非平稳性检验 p 值, p_0 、 p_1 和 p_2 分别为使用局部线性估计方法得到的回归系数函数 β_0 、 β_1 和 β_2 的变化显著性检验的检验 p 值。

表 4.1 局部线性拟合的窗宽参数及 p 值

h_S	h_T	p	p_0	p_1	p_2
67.73km	0.6801月	1.31E-307	0.0000958	0.0002781	0.000233

本章中核函数均为Guass核, 窗宽参数均由交叉确认法确定。表4.1中的显著性检验结果表明, 模型的全局非平稳性检验及各系数函数变化显著性检验的检验 p 值均很小(趋于0)。说明在2013年期间, 67个环保重点城市的月平均空气质量指数与月降水量总和及月平均温度之间存在明显的显著性差异, 即月降水量总和及月平均温度对月平均空气质量指数有着显著的影响。

根据收集得来的时空数据集, 使用SAS进行编程计算, 然后根据结果再使用Sufer软件画图处理, 我们得到了67个城市2月份的 β_0 、 β_1 和 β_2 分布图, 如图4.7-图4.9。

如图4.7所示, 为我国67个城市2月份的 β_0 分布特征图。可以看出各城市 β_0 的值的变化范围为(3.535-9.972), 而其中 β_0 的值在(6.458-9.972)之间的城市大多位于我国华北地区, 尤其以河北地区最为集中, 11个点中有5个处于河北省地区。从图4.7, 我们还可以看出, β_0 的值小于6的城市主要集中在我国南方、东南沿海地区, 这些城市普遍为降水充足, 平均温度较高或森林覆盖率大的城市。由此可见, 我国基准月平均空气质量指数 β_0 的分布特征为南方沿海、东南沿海及西南边疆的最低, 向内陆延伸逐渐增高, 到华北地区 β_0 的值达到最高。

由图4.8可知, 67个城市2月份的 β_1 的值大多为负, 只有南宁、呼和浩特、兰州等6个城市的 β_1 的值为正, 这是因为这6个城市的森林覆盖率很高(如南宁)或风力对该城市的影响较大(如兰州), 而本章中未考虑这些影响因素对空气质量指数的影响, 这就导致2月份的降水量对这6个城市的平均空气质量指数的影响较为反常, 出现了 β_1 的值为正的情况。容易看出, 降水量对东北地区城市、华北地区城市及江苏省和浙江省

的城市的影响最大,而对长江中下游的城市及南方的大部分城市的影响较小,这主要是因为我国降水量对东北地区城市、华北地区城市工业相对发达而且降水量相对较少且不均匀,而江苏省和浙江省的城市工业发达降水量较为充足但季节性变化较为明显,因此受降水量影响较大。长江中下游的城市及南方的大部分城市主要发展的是第三产业或高新技术产业,空气质量较好而且降水量每月均十分充足,因此降水量对这些城市的影响较小。

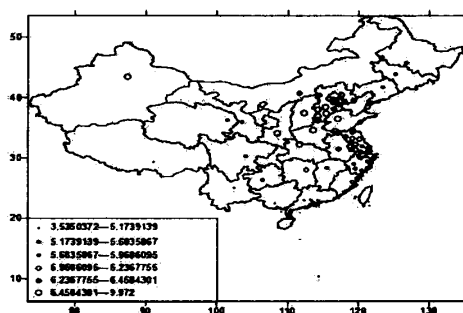


图 4.7 67个城市2月份的 β_0 分布图

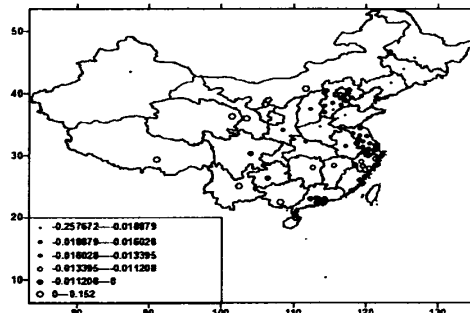


图 4.8 67个城市2月份的 β_1 分布图

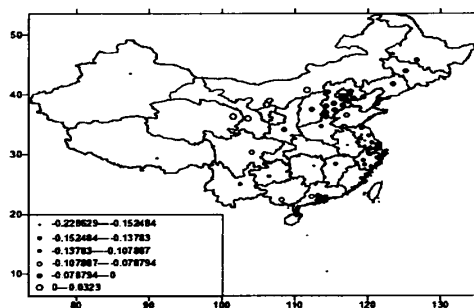


图 4.9 67个城市2月份的 β_2 分布图

从图4.9我国67个城市2月份的 β_2 分布图可以看出,月平均温度对月空气质量指数影响最大的城市位于我国中东部地区、长江中游地区及西藏和乌鲁木齐地区(这些地区的温度变化均较为明显)其次是东部沿海及东南沿海地区。而在南方地区的城市及四川省,平均温度的变化影响较小,这主要是因为这些地区的温度变化不是很明显。华北地区及东北地区的城市主要是受降水量的影响较大而平均温度对这些城市的月平均空气质量指数的影响很小(大于-0.079)但仍为负值。而兰州、呼和浩特等4个城市的 β_2 的值为正,这说明在这4个城市月平均温度不是影响月空气质量指数的主要因

素, 结合图4.8, 我们可知, 月降水量总和及月平均温度对这4个城市的影响较小, 均不是主要因素, 主要因素可能是我们未考虑的风力等因素。

4.3 利用二步估计模拟的结果与分析

根据玄海燕等在文献[35]和梅长林等在文献[37]所给出的显著性检验方法, 构造检验统计量, 使用三阶矩 χ^2 逼近分别求得模型4.1的全局非平稳性检验及各系数函数变化显著性检验的检验 p 值, 见表4.2。其中 p 为模型回归关系的全局非平稳性检验 p 值, p_0 、 p_1 和 p_2 分别为使用局部线性估计方法得到的回归系数函数 β_0 、 β_1 和 β_2 的变化显著性检验的检验 p 值。

表 4.2 二步估计拟合的窗宽参数及 p 值

	h_S	h_T	p	p_0	p_1	p_2
第1步	67.73km	0.6801月	1.31E-307	0.0000958	0.0002781	0.000233
第2步	65.09km	0.6396月	0		6.702E-10	

从表4.2中给出的显著性结果可以看出, 在第一步使用局部线性估计的基础上, 第二步估计选择对光滑程度较高(实际生活经验)的变量月降水量总和 X_1 进行估计。得到的回归系数函数 β_1 的显著性检验的检验 p 值为6.702E-10, 显著优于局部线性估计所得到的 p_1 值(0.0002781)。进一步验证了月降水量总和对月平均空气质量指数的影响显著。

使用SAS对二步估计的第二步进行编程计算, 然后根据结果再使用Sufer软件画图处理, 我们得到了67个城市2月份的 β_1 的分布图, 如图4.10所示。

从图4.10可看出, 由二步估计所得的67个城市2月份的 β_1 的值均为负。其中乌鲁木齐市受降水量的影响最大, 变化率为-0.1798, 即在平均温度不变的情况下, 当月降水量总和增加一个单位时, 月平均空气质量指数减少0.1798。乌鲁木齐市在67个城市中2月份受降水量影响最大的原因是其是典型的煤烟型污染城市, 而且2月份的乌鲁木齐市逆温天气及大气稳定度以稳定型居多, 不利于空气污染物的扩散稀释^[42]。降水量对我国东北、华北、东部沿海地区、南部沿海大部分城市及西藏的空气质量指数的影响也较大, 变化率在-0.01931到-0.01041之间。而在我国西安、兰州和南宁等中西部及西南地区的城市的空气质量指数受降水量的影响相对较小, 原因有很多, 如南宁市则是因为其在2月份是受高压脊的影响较大^[43], 而兰州市主要是因为山谷地形条件下, 2月份的逆温层结稳定深厚, 风力因素对兰州的空气质量指数影响较大^[44]。

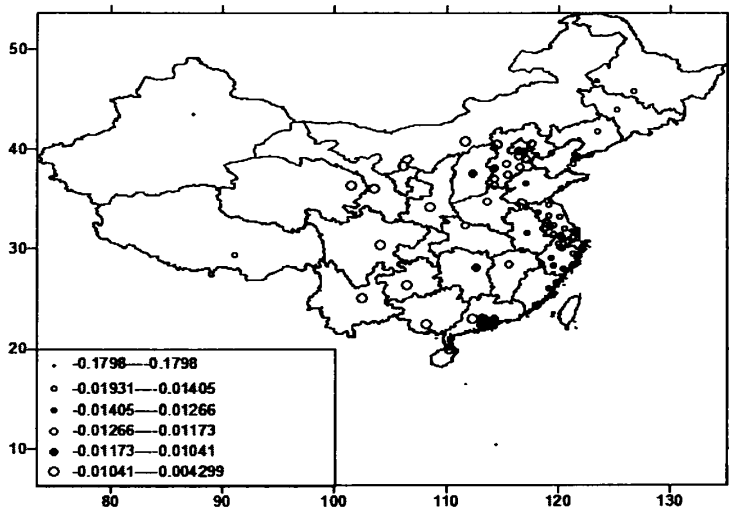


图 4.10 二步估计得到的67个城市2月份的 β_1 分布图

综上所述，我们可知在控制月降水量总和、月平均温度在各地的水平时，月平均空气质量指数随空间位置的变化明显，大致为华北地区的最高，东南沿海次之，我过南部地区最低。2013年我国华北地区的空气质量指数最高，全国空气质量最差十座城市，其中7个在河北地区，这与我们的模拟结果相符。 β_1 反映的是在月平均温度不变的情况下，各城市的月平均空气质量指数随月平均温度的变化。当月降水量总和增加一个单位时，华北地区城市、东部沿海城市、东北地区城市及乌鲁木齐的变化率最大，局部线性估计与二步估计的模拟结果基本保持一致。而如果控制月降水量不变，随着月平均温度的增加，安徽、浙江、江苏等省份的部分城市的月平均空气质量指数减少最多，月平均温度每升高1摄氏度，月平均空气质量指数减少0.152484-0.228629。由此可知，我们使用实际数据集模拟的结果与我们在实际生活中等到的结果和预分析的基本一致，因此，时空加权回归模型能够帮助我们很好的模拟分析我国空气质量指数的分布特征，时空加权回归模型是我们实际应用研究中一个很重要的工具，具有实际意义。

结论与展望

近年来,随着海量数据的频繁出现以及观测数据生成机制的日益复杂化,经济学、环境学等领域内的时空数据统计分析已越来越受到人们的重视。时空加权回归模型是在地理加权回归模型的基础上推广的,通过在变系数回归模型中假定回归系数是地理位置和观测时刻的函数,将观测数据的时空特性嵌入到回归模型中,为分析回归关系的时空特性创造了条件,并为解决回归模型的空间非平稳性和时间非平稳性提供了可行性。对时空加权回归模型中的回归系数的估计目前多是使用最小二乘估计,而我们已知最小二乘估计存在着不足:1、不稳健,也就是当观测数据存在较大的偏离值时,参数的最小二乘估计不可靠,与其真值偏离太远;2、误差分布不是正态分布时,最小二乘估计不是最优估计。在实际应用中,时空加权回归模型对解决金融学、环境地理学及生态学等领域中的实际问题有着重要的作用,至今为止,使用时空加权回归模型对实际问题进行模拟研究的还很少。基于上述的考虑,本文从时空加权回归模型的拟合方法、估计量的性质及应用方面做了研究。总结如下:

(1) 基于局部加权最小二乘原理,系统阐述了时空加权回归模型的拟合方法。选取单调递减的Gauss函数来表示权值和时空距离之间的关系。通过设定两个参数为平衡空间距离和时间距离的比例因子,利用三维坐标计算观测点之间时空距离。利用交叉确认法得到最优窗宽参数。最终得到了回归系数的估计值、因变量拟合值以及残差平方和的表达式。

(2) 结合局部线性估计和GTWR拟合方法给出了时空加权回归模型的局部估计方法及其估计的一个有限样本性质,并通过与GTWR方法的比较说明了局部线性估计方法在系数函数及误差方差估计方面的改进效果。基于时空加权回归模型的局部线性估计给出了时空加权回归模型的二步估计方法。通过对二步估计中的第二步估计的改进,给出了一个改进的二步估计方法,解决二步估计计算量大的缺点,并给出了系数函数局部线性估计的条件偏差的近似值及其估计和条件方差的近似值。

(3) 通过气象学的一个实例研究,考察了时空加权回归模型在实际应用中的效果,通过使用局部线性估计方法和二步估计方法对实例进行模拟研究,分析了我国空气质量指数分布的时空特征。

但对于时空加权回归模型而言,仍有下面的问题有待进一步讨论:

(1) 对于时空加权回归模型的其空间非平稳性、时间非平稳性、检验统计量的构造等假设检验问题,都需要进一步的研究。

(2) 对于时空加权回归模型的局部线性估计,其系数函数估计值的条件均方误差的最优收敛速度问题,有待进一步深入研究。

(3) 本文对时空加权回归模型的二步估计的研究,未涉及到二步估计的渐近条

件偏差及其估计、渐近条件方差及其估计、最优收敛速度等问题的讨论，因此这些问题都有待进一步探讨。

上述问题的解决对时空加权回归模型的分析有重要的意义。在本文的工作基础上，期望在今后的研究中有更大的突破。

参 考 文 献

- [1] Brunson C, Fotheringham A S, Charlton M. Geographically weighted regression: a method for exploring spatial nonstationarity [J]. *Geographical Analysis*, 1996, 28(4): 281-298.
- [2] Fotheringham A S, Charlton M, Brunson C. Measuring spatial variations in relationships with geographically weighted regression [J]. In *Recent Developments in Spatial Analysis*, 1997, 60-82. London: Springer.
- [3] Fotheringham A S, Charlton M, Brunson C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis [J]. *Environment and Planning A*, 1998, 30(11): 1905-1927.
- [4] Brunson C, Fotheringham A S, Charlton M. Geographically weighted regression-modelling spatial non-stationarity[J]. *Journal of the Royal Statistical Society, Series D-The Statistician*, 1998, 47(3): 431-443.
- [5] Fotheringham A S, Brunson C, Charlton M. Geographically weighted regression: the analysis of spatial varying relationship[M]. Chichester: Wiley, 2002.
- [6] Leung Y, Mei C L, Zhang W X. Statistical tests for spatial non-stationarity based on the geographically weighted regression model[J]. *Environment and Planning A*, 2000a, 32(1): 9-32.
- [7] Leung Y, Mei C L, Zhang W X. Testing for spatial autocorrelation among the residuals of the geographically weighted regression[J]. *Environment and Planning A*, 2000b, 32(5): 871-890.
- [8] Páez A, Uchida T, Miyamoto K. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity[J]. *Environment and Planning A*, 2002, 34(4): 733-754.
- [9] Páez A, Uchida T, Miyamoto K. A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests[J]. *Environment and Planning A*, 2001, 34(5): 883-904.
- [10] Wheeler D C, Tiefelsdorf M. Multicollinearity and correlation among local regression coefficients in geographically weighted regression[J]. *Journal of Geographical Systems*, 2005, 7(2): 161-187.
- [11] Wheeler D C. Diagnostic tools and a remedial method for collinearity in geographically weighted regression[J]. *Environment and Planning A*, 2007, 39(10): 2464-2481.

- [12] Farber S, Páez A. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations[J]. *Journal of Geographical Systems*, 2007, 9(4): 371-396.
- [13] Wang N, Mei C L, Yan X D. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique[J], *Environment and Planning A*, 2008, 40(4): 986-1005.
- [14] Wheeler D C. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso[J]. *Environment and Planning A*, 2009, 41(3): 722-742.
- [15] Demsar U, Fotheringham A S, Charlton M. Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics[J]. *Table of Contents*, 2008, 7(3-4): 181-197.
- [16] Harris P, Fotheringham A S, Crespo R. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets[J]. *Mathematical Geosciences*, 2010, 42(6): 657-680.
- [17] Harris P, Brunsdon C, Fotheringham A S. Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor[J]. *Stochastic Environmental Research and Risk Assessment*, 2011, 25(2): 123-138.
- [18] Wei C H, Qi F. On the estimation and testing of mixed geographically weighted regression models[J]. *Economic Modelling*, 2012, 29(6): 2615-2620.
- [19] Harris P, Brunsdon C, Charlton M. The comap as a diagnostic tool for non-stationary kriging models[J]. *International Journal of Geographical Information Science*, 2012, 27(3): 511-541.
- [20] Huang Y, Leung Y. Analysing regional industrialisation in Jiangsu province using geographically weighted regression[J]. *Journal of Geographical Systems*, 2002, 4(2): 233-249.
- [21] Wang Q, Ni J, Tenhunen J. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems[J]. *Global Ecology and Biogeography*, 2005, 14(4): 379-393.
- [22] Yu D L. Spatially varying development mechanisms in the Greater Beijing Area: a geographically weighted regression investigation[J]. *Geographical and Environmental Modeling*, 2006, 40(1): 173-190.
- [23] Cho S H, Chen Z, Yen S T, English B C. Spatial variation of output-input elasticities: evidence from Chinese county-level agricultural production data[J]. *Papers in Regional Science*, 2007, 86(1): 139-157.

- [24] Kamarianakis Y, Feidas H, Kokolatos K. Evaluating remotely sensed rainfall estimates using nonlinear mixed models and geographically weighted regression[J]. *Environmental Modelling & Software*, 2008, 23(12): 1438-1447.
- [25] Luo J, Wei Y H D. Modeling spatial variations of urban growth patterns in Chinese cities: the case of Nanjing[J]. *Landscape and Urban Planning*, 2009, 91(2): 51-64.
- [26] Li S, Zhao Z, Xie M M, Wang Y L. Investigating spatial non-stationary and scale-dependent relationships between urban surface temperature and environmental factors using geographically weighted regression[J]. *Environmental Modelling & Software*, 2010, 25(12): 1789-1800.
- [27] Zhao N, Yang Y H, Zhaou X Y. Application of geographically weighted regression in estimating the effect of climate and site conditions on vegetation distribution in Haihe Catchment, China[J]. *Plant Ecology*, 2010, 209(2): 349-359.
- [28] Brown S, Vincent L. Assessment of spatiotemporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression[J]. *Environmental Modeling & Assessment*, 2012, 17(3): 241-254.
- [29] Su S L, Xiao R, Zhang Y. Multi-scale analysis of spatially varying relationships between agricultural landscape patterns and urbanization using geographically weighted regression[J]. *Applied Geography*, 2012, 32(2): 360-375.
- [30] Mayers R S, Wiggins L L, Fontaine H. Tobacco outlet density and demographics: A geographically weighted regression analysis[J]. *Prevention Science*, 2012, 13(5): 462-471.
- [31] Fortheringham A S, Kelly M H, Charlton M. The demographic impacts of the Irish famine: towards a greater geographical understanding[J]. *Transactions of the Institute of British Geographers*, 2013, 38(2): 221-237.
- [32] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for spatio-temporal modeling of house prices[J]. *International Journal of Geographical Information Science*, 2010, 24(3): 383-401.
- [33] 张金牡, 刘彪, 吴波等. 应用改进的时空地理加权模型分析城市住宅价格变化[J]. *东华理工大学学报*, 2010, 33(1): 53-59.
- [34] Yu P H, Lay J G. Exploring non-stationarity of local mechanism of crime events with spatial-temporal weighted regression[J]. *Spatial Data Mining and Geographical Knowledge Services*, 2011, 7-12.
- [35] 玄海燕, 李帅锋. 时空加权回归模型的非平稳性检验[J]. *甘肃科学学报*, 2012, 24(2): 1-4.

- [36] 玄海燕, 李帅锋, 张运虎. 时空加权回归模型的影响分析[J]. 兰州理工大学学报, 2013, 39(5): 135-138.
- [37] 梅长林, 王宁. 近代回归分析方法[M]. 北京: 科学出版社, 2012.
- [38] Vieu P. Nonparametric regression: optimal local bandwidth choice[J]. Journal of the Royal Statistical Society. Series B, 1991, 53(2): 453-464.
- [39] Fan J, Gijbels I, Hu T C, Huang L S. A study of variable bandwidth selection for local polynomial Regression[J]. Statistica Sinica, 1996, 6(1): 113-127.
- [40] Zhang W, Lee S Y. Variable bandwidth selection in varying-coefficient models[J]. Journal of Multivariate Analysis, 2000, 74(1): 116-134.
- [41] Fan J, Zhang W. Statistical estimation in varying coefficient models[J]. The Annals of Statistics, 1999, 27(5): 1491-1518.
- [42] 刘新春, 钟玉婷, 何清. 乌鲁木齐及周边城市空气质量变化特征及影响因素分析[J]. 沙漠与绿洲气象, 2010, 4(4): 12-17.
- [43] 谢宏斌. 影响南宁市空气质量的天气因素分析[J]. 云南环境科学, 2005, 24(Z1): 28-31.
- [44] 余晔, 夏敦胜, 陈雷华. 兰州市PM₁₀ 污染变化特征及其成因分析[J]. 环境科学, 2010, 31(1): 23-28.

此页不缺内容

致 谢

本文是在我的导师玄海燕副教授的悉心指导下完成的。值此毕业论文完成之际,我首先向我的导师玄海燕老师表示衷心的感谢!在研究生三年的学习和生活中,玄老师从课题研究到生活方面都给予我很大的关怀和帮助。玄老师一丝不苟的作风,严谨求实的态度,踏踏实实的精神,不仅授我以文,而且教我做人,虽历时三载,却给我终生受益无穷之道;感谢玄老师对我生活上的关心;更感谢玄老师严谨的治学态度、务实的作风,让我无以懈怠。在论文的创作过程中,从资料的搜集整理、论文的选题、多次修改直至最后的定稿,都倾注了玄老师大量的心血。在此学业完成之际,向玄老师致以真挚的感谢!

感谢理学院的老师们,感谢他们的传道授业解惑;感谢同门师兄以及同学们,感谢他们在学习过程中给我的帮助。

最后我要感谢我亲爱的父母,正是他们的关心、支持和鼓励,使我在求学的路上充满斗志。

李琪

2014 年 4月于兰州理工大学

此页不缺内容

附录A 攻读学位期间所发表的学术论文目录

- [1] 玄海燕, 李琪, 杨娜娜. 时空加权回归模型的变窗宽局部估计. 甘肃联合大学学报, 已发表, 2013年04期.
- [2] 玄海燕, 李琪, 张运虎. 基于地理加权回归的我国各市人口总数的空间特征分析. 生物数学学报, 已投稿.
- [3] Haiyan Xuan, Qi Li, Anqi Zhang, Youming Guo. Analyzing the Influence of Precipitation and Temperature on Air Quality with GTWR Approach. Water, Air, & Soil Pollution, Submitted.
- [4] Haiyan Xuan, Qi Li, Anqi Zhang, Youming Guo. Spatial and Temporal Characteristics Analysis of Air Quality in Chinese Cities Based on GTWR Approach. Applied Mathematical Modelling, Submitted.
- [5] 玄海燕, 李琪, 张安琪. 时空加权回归模型的局部线性估计的性质. 纯粹数学与应用数学, 已投稿.
- [6] 玄海燕, 李琪, 张安琪. 时空加权回归模型的二步估计的性质. 数学杂志, 已投稿.