

一种新的基于全息映射的面板时空地理 加权回归模型方法^①

范巧 郭爱君

(兰州大学经济学院)

研究目标：架构适应面板数据分析、基于全息映射的时空地理加权回归模型分析范式。**研究方法：**基于近邻局部点对目标分析局部点的全息映射构建了适应面板数据空间计量局部分析的内生时空权重矩阵；在此基础上基于局部点参数估计与模型整体性质分析的分解，系统地架构了面板时空地理加权回归模型方法；还基于索洛余值法与面板时空地理加权回归模型方法的结合，对1990~2018年中国省级层面的全要素生产率及其增长率进行了重估。**研究发现：**基于全息映射的面板时空地理加权回归模型方法全面地解析了空间局部点之间影响效应的直接路径和间接路径；既考察了空间近邻局部点的同伴效应，也考虑了空间局部点自身的内生动力；同时，基于最优空间带宽和最优时间带宽纳入有效的近邻局部点，使得局部点空间依赖的规律性和异质性分析更为精准。**研究创新：**建立了新的基于全息映射的面板时空地理加权回归模型方法，编写了一整套基于MATLAB R2020a的标准化代码。**研究价值：**推动了面板时空地理加权回归模型方法的理论研究，并为面板数据空间计量局部分析的应用研究提供了标准化代码。

关键词 面板时空地理加权回归模型 时空权重矩阵 同伴效应 全息映射

中图分类号 F064.1 **文献标识码** A

DOI:10.13653/j.cnki.jqte.2021.04.007

引 言

地理加权回归模型是基于核密度估计和加权最小二乘法的一种局部回归方法，其与以通用嵌套空间计量模型及其各种退化模型为主的空间计量全局模型，共同构成了现代空间计量分析的主流方法。地理加权回归模型源于空间变参数回归（Foster和Gorr，1986），重在解释空间非平稳性和解释变量参数的漂移性（Brunsdon等，1996），其核心建模逻辑在于在局部点的分析中基于带宽纳入不同的近邻局部点从而解析空间异质性。时空地理加权回归模型是在空间溢出效应中添加时间影响因素的空间局部分析模型（Huang等，2010；Fotheringham等，2015），然而这些模型仅仅对面板数据进行大截面化处理，不适应面板数据的局部分析建模需要；这些模型还存在忽视样本地区信息向目标分析地区映射过程的间接路径、忽视样本地区空间溢出效应在时间上的转移和传导效应的问题；这些模型采用的自适应空间带

^① 本文获得2019年度重庆市社会科学规划项目（2019YBJJ045）和2019年度重庆市留学归国人员创新创业支持计划入选项目（cx2019112）的资助。

宽也具有其内在的本质缺陷,其随机扰动项方差估计及相关统计性质指标计算过程中也存在近邻局部点数据重复使用的问题,从而导致计算结果不够精准。

本文拟基于全息映射的理念,重点架构全面考虑样本地区信息向目标分析地区映射过程的直接路径、间接路径以及空间溢出效应在时间上的转移和传导效应的内生时空权重矩阵,以考察纳入分析的近邻局部点对目标分析地区的全方位的、时空维度的影响效应;在此基础上着力解决自适应带宽的内在缺陷及随机扰动项方差及相关统计性质指标估计过程的非精准性问题,架构适应于面板数据空间局部分析的面板时空地理加权回归模型的统一框架和分析范式;本文还将从索洛余值与空间局部分析模型互嵌的视角,以1990~2018年中国省级层面的全要素生产率核算问题为算例,探讨面板时空地理加权回归模型的应用问题。本文的完成,不仅实现了适应面板数据的时空地理加权回归模型的理论创新和范式架构,建立了面板数据空间计量局部分析的统一分析框架,基于本文理论框架所撰写的MATLAB标准化代码也将极大地推动面板数据空间计量局部分析在经济学研究中的应用,从而具有十分重要的理论价值、方法价值和应用价值。

一、近期文献回顾

1. 模型设定与衍生

地理加权回归模型族中,主要有GWR基本模型、半参数地理加权回归模型、多尺度地理加权回归模型、多尺度半参数地理加权回归模型等经典形式。其中,基本模型强调每个局部点给予不同的参数且模型中仅考虑局部解释变量;半参数地理加权回归模型强调在模型中既纳入局部解释变量,也纳入全局解释变量(Kang和Dall'Erba, 2016);多尺度地理加权回归模型强调不同的局部解释变量具有不同的空间带宽尺度(Fotheringham等, 2017);多尺度半参数地理加权回归模型既强调模型中纳入全局解释变量和局部解释变量,也强调局部解释变量具有不同的空间带宽尺度(Chen和Mei, 2020)。

地理加权回归模型的4种经典形式中,其被解释变量往往是连续型数据,随机扰动项也往往服从零均值、同方差的正态分布。对被解释变量数据属性和随机扰动项分布性质的延展,将衍生出各种各样的地理加权回归模型类型。当被解释变量为离散型数据时,GWR基本模型延展为地理加权逻辑斯蒂回归模型(Atkinson等, 2003)、地理加权判别式回归模型(Brunsdon等, 2007)等两类,其中第一类的被解释变量数据类型为贝努里离散选择数据,第二类的被解释变量数据类型为分类数据;当被解释变量为曲线、平面、三维图像等函数型数据时,GWR基本模型延展成为函数型数据地理加权回归模型(Romano等, 2020);当被解释变量为高频数据和实时更新数据时,GWR基本模型延展成为可逆近邻搜索地理加权回归模型(Reverse Nearest Neighbor GWR)(Tasyurek和Celik, 2020);当微观空间单元的数据缺乏时,GWR基本模型延展成为基于面域空间数据的微观空间单元地理加权回归模型(Area To Point GWR)(Murakami和Tsutsumi, 2015)。当随机扰动项的分布属性发生改变时,GWR基本模型延展成为地理加权泊松回归模型(Shariat-Mohaymany等, 2015)、地理加权泊松交互模型(Nakaya, 2001)、半参数地理加权泊松回归模型(Nakaya等, 2005)三类,其中,第一类模型随机扰动项服从泊松分布,第二类模型在此基础上纳入来源地、目的地及流量数据作为解释变量,第三类模型在第一类模型基础上纳入全局变量。

2. 空间权重矩阵设计与延展

空间权重矩阵是现代空间计量经济学区别于传统计量经济学的重要标志。GWR分析中

使用的空间权重矩阵,其元素维度与纳入分析的近邻局部点个数相同,主对角线元素为目标分析局部点与纳入分析的近邻局部点之间的空间影响关系强度,其余元素为0。局部点之间的空间影响关系强度一般基于核函数和空间距离而计算,主要的核函数包括 Gaussian 函数、Bisquare 函数等(Nakaya, 2001),空间距离一般依据局部点之间的经纬度,在欧氏距离或更为一般的 Minkowski 距离基础上加以计算,部分研究认为在空间距离计算中还应纳入局部点的经济规模或体量特征。在面板数据条件下,空间权重矩阵将延展成为时空权重矩阵,相应地,GWR 模型也会延展为时空地理加权回归模型(GTWR)。

GTWR 模型源于 Huang 等(2010)的研究,其主要思路是在将非平衡面板数据堆积为大截面数据、空间权重矩阵采用 Gaussian 函数和欧式距离的基础上,将时间间隔距离嵌入核函数中以构建时空权重矩阵,进而采用 GWR 一般分析范式进行建模。Huang 等(2010)构建的时空权重矩阵存在一些较为明显的缺陷,包括对面板数据的大截面化数据处理、核函数限于 Gaussian 函数且不可延展、时空权重矩阵设计中忽略了空间溢出效应在时间上的可变效应以及转移和传导效应。学界对这些缺陷进行了一定的修正,Harris 等(2013)将时空权重矩阵中的时间因素替换为情景变量,由此构建了情景地理加权回归模型;Wu 等(2014)在 GTWR 模型中嵌入了空间自回归模型,架构了时空地理加权自回归模型的分析范式;Fotheringham 等(2015)放宽了空间权重矩阵在时期上一致性的假定,并在固定时间带宽的基础上对各时期上的可变空间权重矩阵进行优选,并在此基础上基于不变的时间带宽对时间权重矩阵进行了优选,随后基于 Gaussian 函数将可变空间权重矩阵和不变的时间权重矩阵组合形成时空权重矩阵;Du 等(2018)将时空距离分解成了空间距离、季节性时间距离和非季节性时间距离,以此构建了季节性时空地理加权回归模型和季节性时间地理加权回归模型;Bidanset 等(2018)在时空权重矩阵中添加了特征因素,由此延伸了地理特征加权回归模型以及地理、时间和特征加权回归模型。值得指出的是,这些延展研究大多建立在大截面数据 GWR 分析基础上,其时空权重矩阵也仅仅是在传统的空间权重矩阵中添加了时间因素或者其他因素,并不能完全适应面板数据空间计量模型的建模需要;同时,这些延展研究多数建立在经典计量模型的局部分析延展框架下,未能实现地理加权回归模型与空间计量全局模型的融合研究。

3. 参数估计与模型遴选

参数估计是空间局部分析的关键环节,是地理加权回归模型建模过程的关键步骤。目前,主要存在四种参数估计方法来估计 GWR 模型族的相关参数:第一,针对地理加权回归模型的基本模型,主要采用加权最小二乘法,其核心思路在于以空间权重矩阵作为加权权重对局部点的参数进行估计;第二,针对半参数地理加权回归模型,主要采用两阶段估计方法,其核心思路在于按照一定的顺序,分别用最小二乘回归和地理加权回归处理全局变量和局部变量的参数估计问题,然后再将全局分析结果代入局部回归的参数估计式中(Kang 和 Dall'erba, 2016);第三,针对多尺度地理加权回归模型或多尺度半参数地理加权回归模型,主要采用基于线性可加模型和后向拟合法进行参数估计迭代,其主要做法是以平滑因子或者局部变量参数的地理加权回归估计值作为初始值,结合后向拟合法和 Gauss-Seidel 迭代或 Jacobi 迭代(Brunsdon 等, 1999),来实现对平滑因子或变量参数的估计;第四,针对被解释变量为离散选择数据的 GWR 模型,往往建立在极大似然估计的基础上,其主要做法在于基于随机扰动项的泊松分布或者其他分布设定,找准给定样本信息条件下参数出现的对数似然函数,并基于对数似然函数关于相关参数的一阶导数等于 0 来求解参数估计值。

近期研究主要基于局部变量参数的线性分解或参数估计基本准则调整来延展相关参数估计方法。Wang 等 (2008) 在将局部变量参数经过泰勒级数展开为局部点参数及其在经度和纬度上的局部变动比率基础上, 架构了局部线性拟合的地理加权回归模型 (Local Linear Fitting-Based GWR, LLFBGWR) 的分析范式; Zhang 和 Mei (2011) 舍弃了最小二乘准则, 采用局部最小绝对偏差准则对 LLFBGWR 模型参数估计问题进行了深入探讨; Chen 等 (2012) 放弃了均值回归的经典做法, 对分位数以下和以上的数据各自赋予不同的权重, 结合参数估计准则的加权处理, 架构了地理加权分位数回归模型的分析范式; Comber 和 Harris (2018) 综合了 LASSO 回归和岭回归的建模思路, 在弹性网络回归基础上, 架构了地理加权网络弹性逻辑斯蒂回归模型的分析范式。

模型遴选是 GWR 模型参数估计之后的一个重要环节, 主要包括是否选择地理加权回归进行局部分析、在 GWR 分析中选择何种类型和何种规模的空间带宽、在 GWR 分析中选择何种模型进行局部分析等三个方面。对空间数据分析中是选择局部模型还是选择全局模型的问题, 主要通过参数在空间或时间上的一致性检验来完成 (Mei 等, 2004); 对地理加权回归分析中空间带宽的选择主要通过 CV 准则或 AICc、BIC、SBIC 等准则来进行; 对 GWR 分析中选择何种局部模型的分析主要通过假设检验来完成。事实上, 无论是对参数在空间或时间上一致性的假设检验, 还是对地理加权回归模型类型选择的假设检验, 都是要依据模型之间的转化条件, 设定原假设和备择假设, 并依据模型转化过程中的参数估计值或随机扰动项方差估计值等, 在构建一定的二次型统计量基础上, 结合二阶矩或者三阶矩卡方分布逼近, 构造一个卡方分布或者 F 分布来拒绝或接受原假设 (Brunsdon 等, 1999), 由此进行模型的遴选。近期研究中, 部分学者认为这种以二阶矩或三阶矩卡方分布逼近统计量分布的做法缺乏精准性, 主张以随机扰动项的 Bootstrap 抽样为基础, 结合迭代的思想, 以模型残差平方和变动比率来进行假设检验和模型遴选 (Mei 等, 2016)。同时, 在 GWR 模型族的假设检验中也容易遭受局部点联合检验时的显著性水平设定陷阱, 需要以样本容量、解释变量个数以及参数估计过程的帽子矩阵等因素来调整显著性水平 (Silva 和 Fotheringham, 2016)。

4. 建模过程中的特定问题处理

GWR 模型参数估计过程中, 往往会存在多重共线性, 这种多重共线性可能源自局部变量之间的相关关系, 也可能源自不同局部点分析中纳入相同的近邻点所形成变量之间的相关关系 (Wheeler 和 Tiefelsdorf, 2005)。GWR 分析中的多重共线性, 通常可以通过方差膨胀因子、条件数和局部相关系数来识别 (Fotheringham 和 Oshan, 2016), 当方差膨胀因子大于 10, 条件数大于 30, 或局部相关系数大于 0.9 时, 均意味着 GWR 模型中存在多重共线性。GWR 分析中的多重共线性主要通过局部解释变量的逐步回归法、基于贝叶斯的广义岭回归 (Bárcena 等, 2014)、LASSO 回归以及网络弹性回归 (Comber 和 Harris, 2018) 等方法加以消解, 其中, 第一种消解方法建立在局部解释变量的增减上, 第二至四种消解方法建立在对目标损失函数施加一定的约束条件基础上。

GWR 模型参数估计和模型遴选后, 需要进行稳健性评估。经典的稳健性评估方法包括删掉全局回归过程中残差值大于 3 的部分样本后重新进行 GWR 分析、对初始残差较大的部分样本进行向下加权后重新进行 GWR 分析等两种 (Fotheringham 等, 2002)。Harris 等 (2010) 认为, 这两种评估方法均存在一定的主观性, 即临界值数值大小直接决定删除样本量的多少, 而所选择的向下加权函数形式也会直接决定数据调整的幅度, 并主张以标

标准化后的全局残差和标准化后的局部残差作为样本删除的临界值，以降低稳健性分析中的主观性。预测是传统回归模型的一个重要功能，目前 GWR 分析的相关文献中少有对预测问题的专门探讨。Imran 等（2015）曾结合局部变量及其参数估计值，以及随机扰动项方差估计值的 KED 克里金插值，来对局部点进行预测。不过，这种预测未能考虑随机扰动项的分布特征，且基于点估计和随机扰动项估计值插值的预测方法精度不高，需要进一步改进。

二、面板时空地理加权回归模型的基本设定：基于同伴效应

面板时空地理加权回归模型是基于面板数据的地理加权回归模型，是从经典的面板数据计量经济学模型中衍生出来的，是包含了不同时间维度上不同地区截面数据的地理加权回归模型特定形式。与经典面板数据计量模型和地理加权回归模型有所区别，面板时空地理加权回归模型强调从目标分析地区的周边地区中借取一定数量的样本，从时空维度的样本地区信息中得到时空依赖关系的规律性及异质性，从而阐释目标分析地区被解释变量和解释变量之间的局部时空依赖关系，其基本模型设定如下式（1）和式（2）所示。

$$y_L = X_L\beta + \epsilon \tag{1}$$

$$y_l = X_l\beta_l + \epsilon_l \tag{2}$$

式（1）中， L 为在把同一时期上所有地区作为整体，按照时间上由近到远堆积排列后形成的地区集合，即 $L = \{1_T, 2_T, \dots, N_T, \dots, n_t, \dots, 1_1, 2_1, \dots, N_1\}'$ 。其中， n, t 分别为纳入分析的地区数和时期数， $n = 1, 2, \dots, N, t = 1, 2, \dots, T$ ；符号 $'$ 表示转置矩阵（全文同）。 y_L, X_L 为 L 地区对应的被解释变量和解释变量， β 为解释变量的外生参数， ϵ 为随机扰动项， $\epsilon \sim N(0, \sigma^2 I_{NT})$ ， I_{NT} 为 NT 阶单位矩阵。式（2）中， l 为目标分析地区， $l \in L$ ； l 的取值意味着不同时期上同一个目标地区将被作为单独的地区样本来对待。 y_l, X_l, ϵ_l 分别为目标分析地区的被解释变量、解释变量和随机扰动项， β_l 为目标分析地区解释变量的外生参数。一般地说， X_l 会包括多个解释变量，相应地 β_l 也可以包括多个外生参数。

面板时空地理加权回归模型是依据经济学中的“同伴效应”来进行模型架构的。同伴效应强调个体的行为会受到周边地区行为的影响，个体的行为将可以从一定数量的同伴行为中找到规律性（Bursztyn 等，2014）。式（3）显示了基于同伴效应的面板时空地理加权回归模型转化式。

$$STW_{\{\in l\}} y_{\{\in l\}} = STW_{\{\in l\}} X_{\{\in l\}} \beta_l + \epsilon_{\{\in l\}} \tag{3}$$

式（3）中， $\{\in l\}$ 表示在以 l 作为目标分析地区时，基于一定的遴选准则而确定纳入分析框架的不同时期的地区样本集合。 $STW_{\{\in l\}}$ 为时空权重矩阵，设 $\{\in l\}$ 中元素的个数为 $Num_{\{\in l\}}$ ，则 $STW_{\{\in l\}}$ 的维度为 $Num_{\{\in l\}} \times Num_{\{\in l\}}$ 。 $y_{\{\in l\}}, X_{\{\in l\}}$ 表示样本地区的被解释变量和解释变量矩阵，其维度分别为 $Num_{\{\in l\}} \times 1, Num_{\{\in l\}} \times (\kappa + 1)$ ， κ 为纳入分析的解释变量个数。 $\epsilon_{\{\in l\}}$ 为随机扰动项，满足零均值、同方差的正态分布假定。 β_l 为解释变量 $X_{\{\in l\}}$ 的外生参数矩阵， $\beta_l = [\beta_{l,0}, \beta_{l,1}, \beta_{l,2}, \dots, \beta_{l,\kappa}]'$ 。

式（3）相比较式（2）而言，解释变量参数未发生变化，均表示目标分析地区 l 的被解释变量和解释变量之间的局部空间依赖关系，但其被解释变量和解释变量则发生了变化，式

(2) 中的目标分析地区数据信息，在式 (3) 中变为了将样本地区作为“同伴”映射到样本地区后形成的新信息。这意味着面板时空地理加权回归模型建模过程，将严重依赖于将样本地区作为“同伴”映射到目标分析地区的过程，这一映射过程的精准性决定了面板时空地理加权回归模型的建模科学性。由此，在面板时空地理加权回归模型建模理论分析和实践中，必须将对目标分析地区有显著时空溢出效应的所有周边地区，作为“同伴”样本地区纳入分析框架。当然，作为目标分析地区自身，也应该被作为样本地区纳入分析过程，毕竟这些目标分析地区的内生影响，相比“同伴”样本地区的影响而言更为重要。图 1 显示了基于同伴效应的面板时空地理加权回归模型的基本建模理念。

图 1 的左图中，实心圆圈代表目标分析地区， T_0 表示目标分析地区对应的时期数，这个时间周期上的其他所有地区截面以圆圈表示。实线箭头表示对目标分析地区产生时空溢出效应的有效时间界限，以 T_e 表示；实线椭圆表示对目标分析地区产生时空溢出效应的有效空间界限。此时，在有效时间界限和有效空间界限之内的所有地区，构成对目标分析地区产生时空溢出效应的“同伴”样本地区（以下简称样本地区）。在面板时空地理加权回归的建模过程中，样本地区的行为信息将依据其对目标分析地区的时空溢出效应映射到目标分析地区上去，依据这些映射后的样本地区行为信息，可以进行建模并阐释目标分析地区的行为。对时空溢出效应的专门测度并将之纳入建模的全过程，是面板时空地理加权回归模型的独有特征，也是其与经典计量模型区别的根本标志。

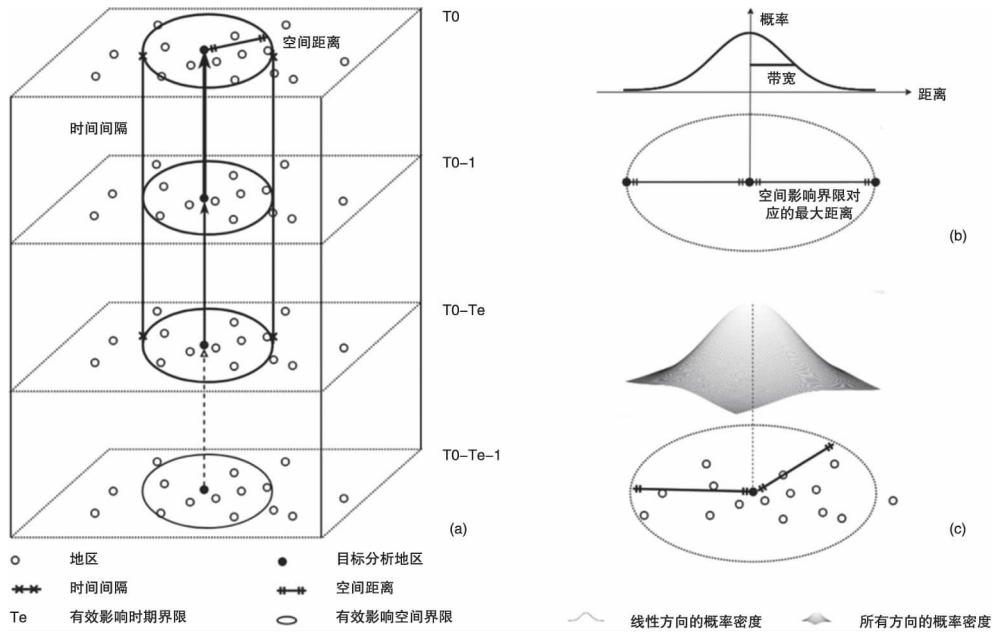


图 1 基于同伴效应的时空地理加权回归模型建模理念

三、面板时空地理加权回归模型的建模源起：基于 GTWR 模型建模过程的缺陷解析

面板时空地理加权回归模型的理论演进，建立在对时空地理加权回归 (GTWR) 模型传统建模方式固有缺陷的弥补基础上。Huang 等 (2010)、Fotheringham 等 (2015) 等对包含时间信息的空间数据的局部建模问题进行了初步研究，并架构了 GTWR 模型的研究范

式。这一范式多以截面数据地理加权回归模型为依据，以空间权重矩阵中纳入时间因素来替代时空权重矩阵的确定，并由此进行参数估计、带宽遴选和模型遴选等。GTWR 模型建模过程中，除了存在将时空数据进行大截面化处理从而不适应面板数据的建模需要外，还存在一些较为明显的缺陷，包括：

缺陷 1. GTWR 模型中对空间权重矩阵元素进行开平方处理将改变样本地区信息向目标分析地区映射过程中的空间影响关系强度。GTWR 模型的建模过程中，以矩阵 $STW_{(i,t)}$ 作为了加权最小二乘法估计中的加权权重，这意味着时空权重矩阵将由 $STW_{(i,t)}$ 矩阵中各元素开平方形成的新矩阵来替代，这实际上是在样本地区信息映射成目标分析地区信息过程中，对样本地区影响目标分析地区的时空溢出效应值进行了开平方处理。这种处理方式意味着地区之间的空间影响效应强度发生了明显改变，这直接影响着将样本地区映射成目标分析地区时的数据结构和数据信息。

缺陷 2. GTWR 模型中，时空权重矩阵计算过程忽视了样本地区信息向目标分析地区映射过程的间接路径。GTWR 模型建模中往往仅依据样本地区到目标分析地区的时空距离，结合空间带宽和核函数预设形式来确定时空权重矩阵的元素。这种设定方式，仅仅考虑了样本地区向目标分析地区映射的直接路径，忽略了样本地区通过周边地区向目标分析地区映射的间接路径，如下图 2 所示。

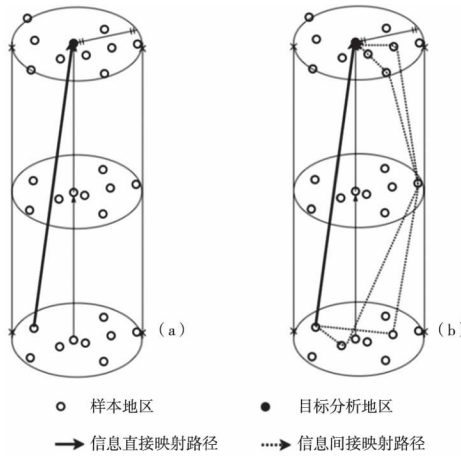


图 2 时空地理加权回归模型中直接映射与全息映射示意图

缺陷 3. GTWR 模型建模中未能准确捕捉样本地区空间溢出效应在时间上的转移和传导效应。在 GTWR 模型的建模过程中，通常会依据时空维度中不同样本地区的时空位置信息，结合三维空间的欧式距离公式或者其他距离公式以及一定的核函数和时间带宽等来确定时空权重矩阵中的元素。这种计算方式中的平方项将抹杀不同时期数之间出现的先后顺序，从而导致时期较近的样本地区会对时期久远的样本地区产生显著的空间影响效应，这与经济社会发展的现实情况不符。同时，直接以时间间隔来表示空间溢出效应在时间上的转移和传导效应的做法也并不完全科学，毕竟同一时间间隔之间空间溢出效应的转移和传导效应有可能不同。再者，将时空维度的样本地区以截面的形式置入抽样总体中并剥离其时间距离和空间距离的做法，容易丧失空间溢出效应在时间维度上的转移和传导效应信息。

缺陷 4. GTWR 模型中采用的自适应空间带宽具有其内在的缺陷。在 GTWR 模型的建

模过程中, 学者们倾向于在一定的核函数设定形式下, 基于样本地区和目标分析地区的空间距离和自适应空间带宽来确定时空权重矩阵的元素。然而, 自适应带宽有其固有的缺陷, 表现在对目标地区有显著空间影响效应的非样本地区未被纳入分析框架, 以及纳入分析框架的样本地区对目标分析地区并未产生显著的空间影响效应等方面。对于单个目标分析地区而言, 如果其近邻地区比较密集, 则在既定的数量阈值下, 部分非样本地区就有可能对该目标分析地区产生显著的空间影响效应; 反之, 如果单个目标分析地区的近邻地区比较稀疏, 则在既定的数量阈值下, 纳入分析框架的样本地区就有可能不会对目标分析地区产生显著的空间影响效应。同时, 单个目标分析地区对应的自适应空间带宽有所不同, 这意味着基于自适应空间带宽和样本地区到目标分析地区的空间距离而计算的空间溢出效应强度不具有可比性。

缺陷 5. GTWR 模型中随机扰动项方差值计算过程存在局部点数据的重复使用, 导致数据维度过大、存储不便和计算不够精准。在 GTWR 模型的建模过程中, 需要将所有局部点分析中纳入的近邻局部点全部纳入, 这会导致局部点数据的重复使用, 毕竟不同的局部点分析中有可能纳入同样的近邻局部点。在模型中截面数、时期数和空间带宽、时间带宽较大的条件下, 随机扰动项方差估计过程所使用的近邻局部点维度将非常之大, 从而计算中将对计算机的运算和存储能力要求变得特别高。同时, GTWR 模型随机扰动项方差估计式设计存在偏误, 由此计算的随机扰动项方差估计值将人为地变大, 甚至有为负值的风险, 将影响解释变量参数的方差—协方差矩阵、参数假设检验中的分布值、模型遴选准则值等的计算, 使得相关结果计算不够精准。

四、基于全息映射的面板时空地理加权回归模型方法逻辑

1. 基于全息映射的时空权重矩阵设计

全息, 是一种光学概念, 是一种全面收集、存储物体衍射光从而完整再现物体的技术(钟丽云等, 2004)。全息映射融合了这一概念, 旨在表征将样本地区通过所有路径映射到目标分析地区, 从而在目标分析地区上完整无损地再现样本地区信息的过程。本文设计的全息时空权重矩阵, 既包括从样本地区向目标分析地区映射的直接过程, 也包括从样本地区通过其他地区向目标分析地区映射的间接过程。由于全息时空权重矩阵涵盖了从样本地区向目标分析地区映射的全部信息, 从而相比时空地理加权回归模型的建模过程而言更为科学, 所使用的样本地区信息也更为完备, 其元素确定如下式(4)所示:

$$\Gamma = STW'_{\{\in l\}} STW_{\{\in l\}}$$

$$STW_{\{\in l\}} = STW_{l,direct} + [STW_{l,spillover} diag(STW_{l,direct})] \cdot * I_{Num_{\{\in l\}}} \quad (4)$$

式(4)中, Γ 表示面板时空地理加权回归模型参数估计时所使用的加权重矩阵; $STW_{\{\in l\}}$ 表示将样本地区 $\{\in l\}$ 全息映射到目标分析地区 l 时所采用的全息时空权重矩阵, 矩阵中仅主对角线有元素, 非主对角线元素全为 0。 $STW_{\{\in l\}}$ 由两部分空间影响效应值相加而成, 一部分来自于样本地区向目标分析地区直接映射的空间影响效应, 另一部分来自于样本地区向目标分析地区间接映射的空间影响效应。 $STW_{l,direct}$ 指表征样本地区 $\{\in l\}$ 对目标分析地区 l 产生直接时空影响效应的时空权重矩阵; $STW_{l,spillover}$ 指表征样本地区 $\{\in l\}$ 两两之间时空溢出效应关系的时空权重矩阵; $diag(\cdot)$ 为提取括号内矩阵主对角线元素所形成新向量; $I_{Num_{\{\in l\}}}$ 为 $Num_{\{\in l\}}$ 阶单位矩阵, 设纳入分析的地区截面数和时间周期数分别为

$n_{\{ \in l \}} = 1, 2, \dots, N_{\{ \in l \}}$ 和 $t_{\{ \in l \}} = 1, 2, \dots, T_{\{ \in l \}}$, 则 $Num_{\{ \in l \}} = N_{\{ \in l \}} \times T_{\{ \in l \}}$; 符号 $\cdot *$ 表示矩阵之间的点乘, 这实际上是把矩阵 $STW_{l,spillover} \text{diag} (STW_{l,direct})$ 中的元素, 分别放到 $I_{Num_{\{ \in l \}}}$ 主对角线元素对应位置上去, 从而形成新的矩阵。

表征样本地区 $\{ \in l \}$ 两两之间时空溢出效应的时空权重矩阵 $STW_{l,spillover}$ 将按照范巧和 Hudson (2018) 的方法确定, 如式 (5) 所示。式 (5) 中, $STW_{l,spillover}$ 、 $TW_{l,spillover}$ 、 $SW_{l,spillover}$ 分别表示样本地区标准化后的时空权重矩阵、时间权重矩阵和空间权重矩阵, \otimes 为克罗内克积符号。 $\widetilde{s\omega}_{l,spillover}$ 表示样本地区初始空间权重矩阵的元素值, 这些元素值经过标准化处理后形成空间权重矩阵 $SW_{l,spillover}$; $f(\cdot)$ 表示核函数; n_o 、 n_d 分别表示样本地区空间溢出效应分析中的起点地区和终点地区, 这些地区必须包括在纳入分析框架的样本地区范畴之内, 即 $n_o \in n_{\{ \in l \}}$, $n_d \in n_{\{ \in l \}}$; $d_{n_o \rightarrow n_d}$ 表示起点地区 n_o 到终点地区 n_d 的空间距离; h_d 表示终点地区 n_d 所对应的空间带宽。出于分析便利和结果可比较考虑, 后文的核函数设计统一采用高斯核函数形式, 空间带宽采用自适应空间带宽形式, 距离公式采用经纬度距离。 $\widetilde{t\omega}_{l,spillover}$ 表示样本地区初始时间权重矩阵的元素值, 这些元素值经过标准化处理后形成时间权重矩阵 $TW_{l,spillover}$; t_o 、 t_d 分别表示样本地区中起点地区对应的时期数和终点地区对应的时期数, t_o 、 t_d 的时间间隔应在时空溢出效应的有效界限之内, 即 $|t_d - t_o| \in t_{\{ \in l \}}$; MI_{t_o} 、 MI_{t_d} 分别表示起点地区和终点地区对应的时期数上依据所有地区截面而计算的全局 Moran 指数。

$$STW_{l,spillover} = TW_{l,spillover} \otimes SW_{l,spillover}$$

$$\widetilde{s\omega}_{l,spillover} = \begin{cases} f(d_{n_o \rightarrow n_d}, h_d), n_o \neq n_d \\ 0, n_o = n_d \end{cases} \quad \widetilde{t\omega}_{l,spillover} = \begin{cases} MI_{t_d}/MI_{t_o}, t_d - t_o \geq 0 \\ 0, t_d - t_o < 0 \end{cases} \quad (5)$$

表征样本地区 $\{ \in l \}$ 对目标分析地区 l 产生直接时空影响效应的时空权重矩阵 $STW_{l,direct}$, 将依据下式 (6) 确定。

$$STW_{l,direct} = \{ \text{diag} (TW_{l,spillover}) \cdot * I_{T_{\{ \in l \}}} \} \otimes SW_{l,direct}$$

$$\widetilde{s\omega}_{l,direct} = f(d_{n_{\{ \in l \}} \rightarrow l}, h_l) \quad (6)$$

式 (6) 中, $TW_{l,spillover}$ 为式 (5) 中所确定的标准化后的时间权重矩阵; $\text{diag}(\cdot)$ 、 $I_{T_{\{ \in l \}}}$ 和符号 \otimes 、 $\cdot *$ 与前文定义一致。 $SW_{l,direct}$ 指表征样本地区 $\{ \in l \}$ 按照所属地区截面依据时间周期排列时的所有地区截面 $n_{\{ \in l \}}$ 对目标分析地区 l 产生直接空间影响效应的空间权重矩阵, 其主对角线元素为 $\widetilde{s\omega}_{l,direct}$; $f(\cdot)$ 仍为核函数, h_l 为目标分析地区 l 所对应的自适应空间带宽, $d_{n_{\{ \in l \}} \rightarrow l}$ 为地区截面 $n_{\{ \in l \}}$ 到目标分析地区 l 的空间距离。值得注意的是, 与式 (5) 中 $SW_{l,spillover}$ 需要标准化的情况不同, $SW_{l,direct}$ 不再需要标准化, 它表示样本地区 $\{ \in l \}$ 所对应的所有截面地区 $n_{\{ \in l \}}$ 对目标分析地区 l 产生直接空间影响效应的水平情况。

依据式 (5) 和式 (6) 确定空间权重矩阵元素时, 为了保证所得到的空间效应程度值具有可比性, 需要对自适应带宽 h_d 和 h_l 进行调整, 即利用 $h = \text{Max} (h_d, h_l)$ 来替代自适应带宽的初始值 h_d 和 h_l , $\text{Max}(\cdot)$ 为取最大值。这种替代过程既可以保证每个目标地区纳入分析框架的样本地区总数相同, 也可保证在统一的尺度下来进行空间影响效应或者空间溢出效应计算。具体程序设计中, h 将可以依据如下步骤计算: 首先, 基于纳入分析框架的样本地区数量阈值, 确定每个目标分析地区应该纳入的样本地区数量, 以及每个样本地区应该纳

入的近邻地区数量。其次, 计算样本地区到目标分析地区的所有距离 $d_{n_{i \in l} \rightarrow l}$, 以及每个样本地区到其纳入分析的近邻地区的所有距离 $d_{n_o \rightarrow n_d}$ 。最后, 依据式 (7), 计算调整后的自适应空间带宽 h 。

$$h = \text{Max}(d_{i \in l \rightarrow l}, d_{n_o \rightarrow n_d}) / \sqrt{-\frac{1}{\eta} \ln(\text{sevc})} \quad (7)$$

式 (7) 中, η 为高斯核函数中的经验常数, 本文取值为 0.5。sevc 为空间影响边界上可以容忍的空间影响效应临界值, 考虑到空间影响效应的定义域 (0, 1) 以及显著性水平临界值的设计方式, 本文设 $\text{sevc} = 0.05$, 这意味着当地区之间的空间溢出效应小于 0.05 时空间溢出效应不存在。

2. 基于全息时空权重矩阵的面板时空地理加权回归模型建模逻辑

全息时空权重矩阵的设计, 需要建立在一定的自适应空间带宽、时间带宽基础上, 这些自适应的带宽一般依据纳入的空间点个数和时期数的阈值而确定。在给定空间带宽和时间带宽条件下, 单个局部点分析中纳入的近邻局部点将形成一个新的面板, 这个新面板是包含所有空间点和所有时期数的初始面板的一个子集, 此时, 可以依据面板数据计量经济学的传统建模逻辑来确定单个局部点的所有参数及统计性质。同时, 由于局部点参数及统计性质的计算依赖于自适应空间带宽、时间带宽的确定, 不同的自适应带宽将意味着不同的最适宜模型, 本文将基于模型的整体统计性质来确定自适应带宽的优选和模型遴选问题, 这些整体统计性质计算过程中的相关指标主要基于模型平均的理念而设计。

(1) 面板时空地理加权回归模型的参数估计和显著性评估逻辑。基于式 (4) 中设计的全息时空权重矩阵, 结合对时空地理加权回归模型建模过程的改进, 可以得到面板时空地理加权回归模型的参数估计和显著性评估的一般逻辑。其中参数及其方差估计值分别为 $\tilde{\beta}_l = \tilde{S} y_{l \in l}$ 和 $\text{Var}(\tilde{\beta}_l) = \text{Diag}_l[\tilde{\Omega}(\tilde{\beta}_l)]$; $\tilde{\Omega}(\tilde{\beta}_l) = \tilde{S} \Gamma^{-1} \tilde{S}' \tilde{\sigma}_l^2$, $\tilde{S} = (X'_{l \in l} \Gamma X_{l \in l})^{-1} X'_{l \in l} \Gamma$; $\text{Diag}_l(\cdot)$ 为提取括号内矩阵的主对角线元素组成的向量中第 l 个元素; 随机扰动项方差估计值将依据式 (8) 计算:

$$\begin{aligned} \tilde{\sigma}_l^2 &= (H_{l \in l} - \tilde{H}_{l \in l})' (H_{l \in l} - \tilde{H}_{l \in l}) / (\tilde{v}_{0,l} - 2\tilde{v}_{1,l} + \tilde{v}_{2,l}) \\ \tilde{\sigma}^2 &= (H - \tilde{H})' (H - \tilde{H}) / (\tilde{V}_0 - 2\tilde{V}_1 + \tilde{V}_2) \end{aligned} \quad (8)$$

式 (8) 中, $\tilde{\sigma}_l^2$ 为局部点 l 的随机扰动项方差估计值, $l = 1, 2, \dots, NT$; $H_{l \in l} = \text{STW}_{l \in l} y_{l \in l}$; $\tilde{H}_{l \in l} = \tilde{h}_l y_{l \in l}$, $\tilde{h}_l = \text{STW}_{l \in l} X_{l \in l} S$; $\tilde{v}_{0,l} = \text{tr}(\text{STW}_{l \in l})$, $\tilde{v}_{1,l} = \text{tr}(\tilde{h}_l)$, $\tilde{v}_{2,l} = \text{tr}(\tilde{h}_l' \tilde{h}_l)$ 。 $\tilde{\sigma}^2$ 为模型整体的随机扰动项方差估计值; $H = [H_1; H_2; \dots; H_l; \dots; H_{NT}]$, $H_l = y_l$; $\tilde{H} = [\tilde{H}_1; \tilde{H}_2; \dots; \tilde{H}_l; \dots; \tilde{H}_{NT}]$, $\tilde{H}_l = A \tilde{H}_{l \in l}$; A 为局部点 l 分析中所有产生有效影响的近邻局部点对局部点 l 的空间影响效应份额, $A = \text{tr}^{-1}\{\text{STW}_{l \in l}\} \text{Diag}\{\text{STW}_{l \in l}\}$, $\text{Diag}(\cdot)$ 表示提取括号内矩阵的主对角线元素形成的新向量, $\text{tr}^{-1}(\cdot)$ 为括号内矩阵的迹统计量的倒数 (全文同); $\tilde{V}_i = \sum_l \tilde{v}_{i,l} / NT$, $i = 0, 1, 2$ 。

在确定随机扰动项方差估计值的基础上, 容易确定各局部点参数估计的 T 统计量及其分布, 其中 $T_{l,k} = \tilde{\beta}_{l,k} / \text{Se}(\tilde{\beta}_{l,k}) \sim T(\tilde{v}_{0,l} - 2\tilde{v}_{1,l} + \tilde{v}_{2,l})$, $\tilde{\beta}_{l,k}$ 为向量 $\tilde{\beta}_l$ 的第 k 个元素, $\text{Se}(\tilde{\beta}_{l,k})$ 为向量 $\text{Var}(\tilde{\beta}_l)$ 中第 k 个元素的开平方。此时, 由于局部点参数估计过程仍按照传统的面板数据计量经济学方法进行, 则其概率值及概率临界值仍需按照传统算法进行计算, 相关统计性质指标的确定也仍然按照传统算法进行设计 (包括拟合优度、AICc 准则值、

F分布值、F分布的概率值等)。由于模型整体分析中单个解释变量参数变得无足轻重,则不再需要对整体模型的解释变量参数及假设检验过程做出单独分析。

(2) 面板时空地理加权回归模型中的整体统计性质与模型遴选逻辑。面板时空地理加权回归模型中,目标分析地区行为将可以通过样本地区行为及其在所有地区对目标分析地区时空溢出效应总水平中所占份额的乘积来解释或推断、预测;基于样本地区行为,对目标分析地区行为的推断或预测精度,将可以决定最适宜模型,并由此决定最优的空间带宽。

基于这一逻辑,目标分析地区的被解释变量 y_l ,既可以由 $\vec{y}_l = ASTW_{(\in l)} y_{(\in l)}$ 解释,也可以由 $\tilde{y}_l = ASTW_{(\in l)} X_{(\in l)} \tilde{\beta}_l$ 解释。A的定义与前文相同。设地区集合L的被解释变量值为 $y_L = [y_{1_T}, y_{2_T}, \dots, y_{N_T}, \dots, y_l, \dots, y_{1_1}, y_{2_1}, \dots, y_{N_1}]'$,则基于样本地区数据信息而得到的地区集合L的被解释变量推断值,以及基于面板时空地理加权回归模型而得到的地区集合L的被解释变量估计值将可以分别由如式(9)表示:

$$\begin{aligned} \vec{y}_L &= [\vec{y}_{1_T}, \vec{y}_{2_T}, \dots, \vec{y}_{N_T}, \dots, \vec{y}_l, \dots, \vec{y}_{1_1}, \vec{y}_{2_1}, \dots, \vec{y}_{N_1}]' \\ \tilde{y}_L &= [\tilde{y}_{1_T}, \tilde{y}_{2_T}, \dots, \tilde{y}_{N_T}, \dots, \tilde{y}_l, \dots, \tilde{y}_{1_1}, \tilde{y}_{2_1}, \dots, \tilde{y}_{N_1}]' \end{aligned} \tag{9}$$

基于式(9),可以计算面板时空地理加权回归模型中的映射误差和估计误差。映射误差表示基于样本地区信息映射得到的目标分析地区被解释变量的推断值 \vec{y}_L ,偏离目标分析地区被解释变量真值 y_L 的程度,即 $RSS_1 = (\vec{y}_L - y_L)' (\vec{y}_L - y_L)$ 。估计误差表示基于面板时空地理加权回归模型得到的目标分析地区被解释变量的估计值 \tilde{y}_L 偏离推断值 \vec{y}_L 的程度,即 $RSS_2 = (\tilde{y}_L - \vec{y}_L)' (\tilde{y}_L - \vec{y}_L)$ 。这两部分构成了面板时空地理加权回归模型的预测误差,即 $RSS = RSS_1 + RSS_2$ 。在面板时空地理加权回归模型建模过程中,不同的自适应带宽数量阈值,意味着不同的自适应空间带宽设定值。自适应空间带宽值的不同设定,将导致不同的模型参数估计结果和模型性质,面板时空地理加权回归模型建模过程中的带宽优选,实际上可以从自适应空间带宽的不同设定中找到一个最适宜的数量阈值,以保证面板时空地理加权回归模型中对目标分析地区被解释变量的预测误差最小,即 $\text{Min}(RSS)$ 。

面板时空地理加权回归模型的带宽优选也可以按照传统的CV准则、GCV准则和AICc准则等来完成,只不过其计算公式需要稍作调整,如式(10)所示。式(10)中,NT、k分别表示局部点的总数和局部模型分析中纳入的解释变量个数; \tilde{V}_0 、 \tilde{V}_1 、 $\tilde{\sigma}^2$ 如式(8)所示, π 为圆周率。面板时空地理加权回归模型将依据CV准则、GCV准则和AICc准则的取值最小来决定最优的自适应空间带宽和时间带宽。当不同的准则取值指向不同的最优带宽时,可以对这些不同最优带宽下的模型全部进行试算,并依据模型优选结果来确定最优带宽的遴选准则及遴选结果。

$$\begin{aligned} CV &= (\tilde{y}_L - y_L)' (\tilde{y}_L - y_L) \\ GCV &= CV / (NT - k - 1)^2 \\ AICc &= \tilde{V}_0 [\ln(\tilde{\sigma}^2) + \ln(2\pi) + (\tilde{V}_0 + \tilde{V}_1) / (\tilde{V}_0 - 2 - \tilde{V}_1)] \end{aligned} \tag{10}$$

面板时空地理加权回归模型中不仅要带宽做出优选,还应对不同解释变量纳入以及不同模型效应下模型整体的优劣做出诊断。本文拟基于式(11)的相关统计性质指标来对模型整体做出综合选择。

$$Rate_Sig = 1 - \frac{\sum_l I\{Prob(T_{l,k}, \tilde{v}_{0,l} - 2\tilde{v}_{1,l} + \tilde{v}_{2,l}) > \alpha\}}{NT \times (k + 1)}$$

$$\begin{aligned} \tilde{R}^2 &= 1 - \frac{(\tilde{y}_L - y_L)'(\tilde{y}_L - y_L)/(\tilde{V}_0 - 2\tilde{V}_1 + \tilde{V}_2)}{(\tilde{y}_L - \bar{y}_L)'(\tilde{y}_L - \bar{y}_L)/(\tilde{V}_0 - 1)} & F &= \frac{(\tilde{y}_L - \bar{y}_L)'(\tilde{y}_L - \bar{y}_L)/(2\tilde{V}_1 - \tilde{V}_2)}{(\tilde{y}_L - y_L)'(\tilde{y}_L - y_L)/(\tilde{V}_0 - 2\tilde{V}_1 + \tilde{V}_2)} \\ \tilde{\alpha} &= \alpha(k+1)/(2\tilde{V}_1 - \tilde{V}_2) & \ln L &= -\frac{\tilde{V}_0}{2} [\ln(\tilde{\sigma}^2) + \ln(2\pi) + 1] \end{aligned} \quad (11)$$

式 (11) 中, $Rate_Sig$ 为整体模型中所有局部点参数估计值显著的比例, NT 、 k 定义如式 (10), α 为预设的显著性水平 (0.01、0.05、0.1); $I\{\cdot\}$ 为示性函数, 当括号内算式为真时取值为 1, 否则取值为 0; $Prob(T_{l,k}, \tilde{v}_{0,l} - 2\tilde{v}_{1,l} + \tilde{v}_{2,l})$ 为在自由度为 $\tilde{v}_{0,l} - 2\tilde{v}_{1,l} + \tilde{v}_{2,l}$ 时, 统计量 $T_{l,k}$ 对应的概率值。 \tilde{R}^2 为整体模型中修正的拟合优度值, \tilde{y}_L 、 y_L 、 \tilde{V}_0 、 \tilde{V}_1 、 \tilde{V}_2 定义与前文相同, \bar{y}_L 为 y_L 的均值。 F 为模型整体显著性的 F 统计量, 其分布形式为 $F(2\tilde{V}_1 - \tilde{V}_2, \tilde{V}_0 - 2\tilde{V}_1 + \tilde{V}_2)$, 由此可以计算 F 统计量的值及其对应的概率水平; 模型整体的概率临界值必须依据预设的显著性水平 α 重新核算; $\ln L$ 为模型整体的对数似然值, \tilde{V}_0 、 $\tilde{\sigma}^2$ 、 π 与前文定义一致。一般地说, $Rate_Sig$ 取值越大, \tilde{R}^2 取值越大, F 越能通过假设检验, 则对应的模型越可能是最优模型。

五、算例：基于面板时空地理加权回归模型的中国省级层面全要素生产率及其增长率重估

全要素生产率核算是经济增长研究中一个十分重要的议题, 传统的核算方法主要包括索洛余值法、随机前沿分析和 DEA 非参数核算法等。在索洛余值法框架下进行全要素生产率的核算, 最重要的是要对经验生产函数做出有效的估计, 不过, 在索洛余值法的传统分析中, 经验生产函数估计过程往往会忽略空间溢出效应, 从而导致要素投入份额之和大于 1, 从而导致全要素生产率核算结果的非精准性。这意味着在索洛余值法框架下进行全要素生产率的核算, 需要嵌入空间计量分析模型, 以确保经验生产函数估计中能够科学地解析经济增长及其投入要素的空间溢出效应关系, 从而得到有效的要素投入份额结果。

近期研究中, 部分学者曾基于索洛余值法与空间计量全局模型的嵌套对全要素生产率的核算问题进行了探讨 (Tientao 等, 2016; 范巧和郭爱君, 2019), 然而空间计量全局模型需要纳入所有地区所有时期的数据来得到经济增长及其投入要素之间空间依赖关系的规律性, 这会忽视了单个局部点的全要素生产率核算中经济增长及其投入要素之间空间依赖关系的异质性。有鉴于此, 本文将在作者的前期相关研究基础上 (范巧和郭爱君, 2019), 重点考虑在索洛余值法框架下纳入面板时空地理加权回归模型, 对中国省级层面的全要素生产率及其增长率进行重新核算, 其经验生产函数设定如式 (12) 所示。

$$\begin{aligned} \ln\left(\frac{Y}{L}\right)(u_l, v_l, t_l) &= \gamma_0(u_l, v_l, t_l) + \gamma_1(u_l, v_l, t_l) \ln\left(\frac{K}{L}\right) \\ &+ \gamma_2(u_l, v_l, t_l) \ln\left(\frac{E}{L}\right) + \mu(u_l, v_l, t_l) \end{aligned} \quad (12)$$

式 (12) 中, Y 、 K 、 L 、 E 分别表示实际 GDP (万元)、资本存量 (万元)、年末总人口数 (万人) 和能源消费总额 (吨标准煤); (u_l, v_l, t_l) 表示局部点, u_l 、 v_l 、 t_l 分别表示局部点 l 的经度、纬度和时间; $\gamma_i(u_l, v_l, t_l)$ 表示局部点 l 的外生参数, $i=0, 1, 2$ 项; $0 < \gamma_1(u_l, v_l, t_l) < 1$, $0 < \gamma_2(u_l, v_l, t_l) < 1$, $\gamma_2(u_l, v_l, t_l) + \gamma_2(u_l, v_l, t_l) \leq 1$; $\mu(u_l, v_l, t_l)$ 表示局部点的随机扰动项。

本文拟基于 1990~2018 年的相关数据, 估计式 (12) 的经验生产函数, 其主体数据来源于范巧和郭爱君 (2019) 的文章, 并依据《中国统计年鉴》(2019) 更新 GDP 和年末总人

口数的 2017~2018 年数据, 2017~2018 年资本存量数据仍依据永续盘存法计算。其余数据更新按照如下方式进行: ①2017 年各地区能源消费总量数据来源于《中国能源统计年鉴 2018》; 2018 年各地区能源消费总量基于 2017 年各地区能源消费总量与增速的乘积核算, 增速数据来源于国家统计局官网公布的“2018 年分省(自治区、直辖市)万元地区生产总值能耗降低率等指标公报”; 2017 年、2018 年西藏能源消费总量源于全国能源消费总量乘以西藏 GDP 占全国的比重; 2018 年全国能源消费总量依据《2018 年国民经济和社会发展统计公报》获取(46.4 亿吨标准煤), 并将单位转换为吨标准煤。②2017 年资本形成总额数据依据资本形成总额(当年价)和固定资产投资价格指数折算。2018 年的资本形成总额数据, 除北京外多数地区均未公布, 按照如下方式进行补缺: 首先, 依据 2018 年中国资本形成总额(当年价格)与 2016~2017 年各地区资本形成总额的和在所有地区总和中的比例, 计算 2018 年各地区资本形成总额初始值; 其次, 依据计算得到的 2018 年北京市资本形成总额与《北京统计年鉴 2019》提供的 2018 年北京市资本形成总额值, 确定调整系数; 最后, 依据各地区资本形成总额初始值和调整系数, 来确定 2018 年各地区资本形成总额数据。③2017 年和 2018 年固定资产投资价格指数依据上一年固定资产投资价格指数(1990 年不变价)与固定资产投资价格指数(上年=100)计算; 2017 年和 2018 年 GDP 平减指数依据上一年 GDP 平减指数(1990 年不变价)与 GDP 平减指数(上年=100)计算, GDP 平减指数依据各年名义地区生产总值与国内生产总值指数(上年=100)平减计算。④中国省级层面的经纬度数据, 以百度地图显示的省级政府驻地的经纬度数据替代, 数据获取时间为 2020 年 9 月 14 日 16 点 16 分。

结合面板时空地理加权回归模型方法, 利用基于 MATLAB R2020a 的自编标准化程序, 对式(12)的经验生产函数进行试算, 基于 AICc 准则的最优空间带宽和最优时间带宽分别为 31 和 17, 基于 GCV 准则和 RSS 准则的最优空间带宽和最优时间带宽均分别为 31 和 14。由于基于 GCV 准则的最优带宽与基于 CV 准则的最优带宽遴选结果基本等价, 本文未考虑基于 CV 准则来优选最优空间或时间带宽。鉴于 AICc 准则、GCV 准则和 RSS 准则指向不同的最优空间带宽和最优时间带宽的事实, 本文分别试算了这两种最优空间、时间带宽下面板时空地理加权回归模型的整体统计性质, 如表 1 所示。鉴于不同带宽维度下纳入的有效近邻局部点构成新的面板数据, 本文还在两种最优带宽维度下分别试算了混合效应、个体固定效应、时期固定效应和个体一时期双固定效应面板时空地理加权回归模型的整体统计性质。由于本文讨论的是中国省级层面的全要素生产率核算问题, 事实上纳入所有空间单元, 无须考虑随机效应下面板时空地理加权回归模型。

从表 1 对各种可能模型的整体统计性质试算结果来看, 在最优空间带宽为 31、最优时间带宽为 17 的条件下, 个体一时期双固定效应面板时空地理加权回归模型体现出了比较好的统计性质, 其中局部系数的估计值显著性比率达到了 100%, 修正的拟合优度达到了 0.9999, F 统计量也能够通过显著性水平为 0.01 的假设检验。同时, 在最优空间带宽为 31、最优时间带宽为 14 的条件下, 个体一时期双固定效应面板时空地理加权回归模型同样体现出了良好的整体统计性质。相比较而言, 最优空间带宽为 31、最优时间带宽为 17 下, 模型整体的随机扰动项方差估计值更小, CV 准则值、GCV 准则值和 AICc 准则值更小, 体现出更为优良的整体统计性质。有鉴于此, 本文认为在阐释中国省级层面的经验生产函数时, 最优的空间带宽和时间带宽分别为 31 与 17。在选定的最优空间、时间带宽条件下, 可以估计所有局部点的相关参数及其统计性质, 还可以对局部点参数的有效性进行诊断, 如图 3 所示。

表 1 两种带宽维度、四种效应下算例模型的整体统计性质

	AICc 准则 (最优空间带宽=31、最优时间带宽=17)				GCV\RSS 准则 (最优空间带宽=31、最优时间带宽=14)			
	混合效应	个体固定效应	时期固定效应	个体-时期双固定效应	混合效应	个体固定效应	时期固定效应	个体-时期双固定效应
	局部系数估计值的显著比率	60.88%	89.15%	70.24%	100%	59.55%	90.66%	73.41%
样本容量	899	899	899	899	899	899	899	899
自由度	172	175	173	175	164	167	164	167
随机扰动项方差估计值	16.9832	4.6699	1.0720	23.6814	17.4407	3.7314	1.1222	25.9125
CV 准则值	2921.1	817.2	185.5	4144.3	2860.3	623.1	184.0	4327.4
GCV 准则值	0.0036	0.0010	0.0002	0.0052	0.0036	0.0008	0.0002	0.0054
AICc 准则值	5106.6	3943.5	2620.7	5403.3	5130.5	3741.8	2661.9	5484.3
修正的拟合优度	0.9829	0.7258	-0.2999	0.9999	0.9878	0.8083	-0.1994	0.9999
F 统计量值	5623.2	4135.6	1118.1	2124.4	7701.0	5795.5	1171.1	3014.1
F 统计量的概率	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
修正的概率临界值 ($\alpha=0.01, 0.05, 0.1$)	0.0319	0.0454	0.0527	0.0527	0.0313	0.0451	0.0516	0.0552
	0.1593	0.2268	0.2636	0.2636	0.1563	0.2255	0.2581	0.2758
	0.3187	0.4536	0.5273	0.5273	0.3126	0.4510	0.5162	0.5515
对数似然值	-2548.7	-1968.4	-1306.9	-2698.2	-2560.7	-1867.5	-1327.4	-2738.6

注：***、**、* 分别表示在 1%、5% 和 10% 的显著性水平上通过假设检验。

资料来源：作者整理。

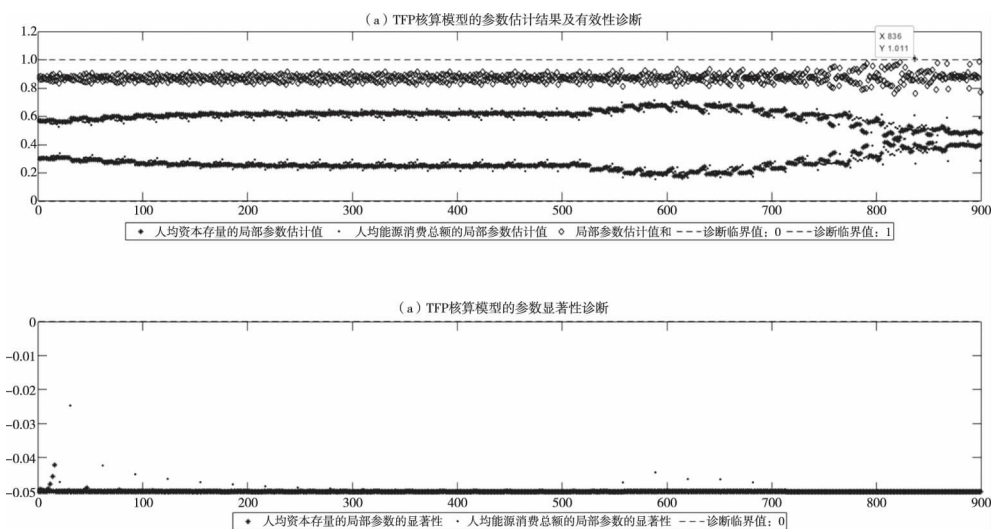


图 3 算例模型参数估计结果有效性与显著性诊断 ($N=31, T=17$)

从图 3 可以看出，所有局部点参数估计值均在 0 和 1 之间，且人均资本对数值和人均能源消费总额对数值的参数估计值和也均小于 1（2017 年宁夏回族自治区除外，如图 3 中第

836个局部点)。这说明,与在索洛余值法中嵌入空间计量全局模型效果类似,基于面板时空地理加权回归模型也可以对中国省级层面的生产函数进行有效的估计。当然,相比较基于空间计量全局模型的分析而言,面板时空地理加权回归模型给予每个局部点以单独的分析,且在各自的分析过程中仅纳入对局部点真正产生有效影响的近邻局部点,从而使得经验生产函数的估计更加精准,以此核算的全要素生产率也将更加精准。基于局部点参数估计结果,本文依据索洛余值法核算了1990~2018年中国省级层面的全要素生产率,并计算了全要素生产率指标及全要素生产率增长率,如下图4所示。

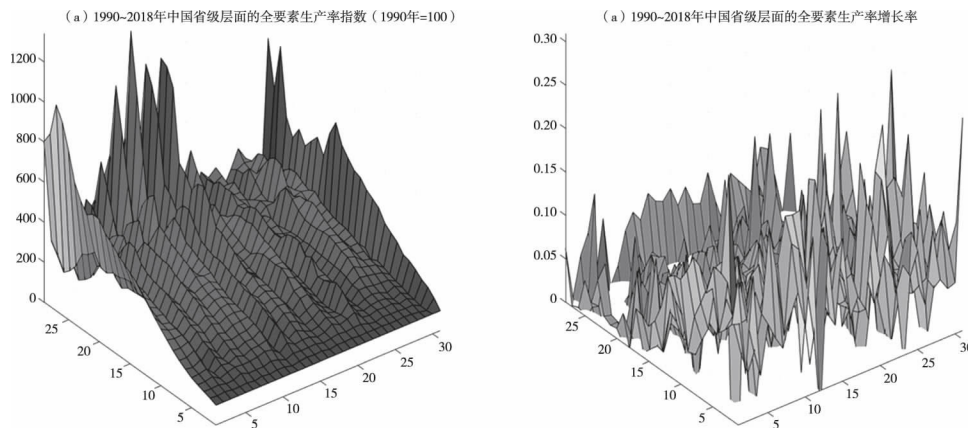


图4 中国省级层面1990~2018年TFP指数及增长率核算结果

六、结论及进一步的研究方向

本文在对地理加权回归模型方法与进展的综合评述基础上,结合对时空地理加权回归模型建模逻辑及建模缺陷的系统梳理,依据全息映射理念构建了包含有效近邻局部点样本地区对目标分析地区直接映射和间接映射的内生时空权重矩阵,并在此基础上架构了适应于面板数据分析的空间计量局部分析模型——面板时空地理加权回归模型的理论框架与分析范式,还以1990~2018年中国省级层面全要素生产率核算问题为算例,探讨了面板时空地理加权回归模型的应用问题,得到了一些十分重要的结论,整理如下:

第一,时空地理加权回归模型建模过程存在将面板数据大截面化、空间权重矩阵设计忽略了间接映射且未能考虑空间溢出效应在时间上的转移和传导效应、所采用的自适应空间带宽在计算空间影响效应存在固有的缺陷、随机扰动项的方差估计值及相关统计性质计算中存在对近邻局部点重复计算且样本地区数量确定不够准确导致计算结果非精准性等问题。

第二,面板时空地理加权回归模型是专门针对面板数据的空间计量局部分析模型。这一模型不仅纳入近邻局部点对目标分析地区的直接映射和间接映射来架构内生时空权重矩阵,还依据最优空间带宽和最优时间带宽的遴选来确定单个局部点分析中需要纳入的有效近邻局部点,从而形成新的面板数据子集,并在这一新的面板数据子集基础上结合经典的面板数据计量经济学分析方法来解析单个局部点的参数及统计性质,随后还基于随机扰动项方差估计值的局部和整体分解,探讨了局部点参数显著性诊断问题,还重新设计了表征面板时空地理加权回归模型的整体统计性质的系列指标。

第三, 本文还从索洛余值法与面板数据空间计量局部分分析方法互嵌的视角, 探讨了面板时空地理加权回归模型的应用问题, 对 1990~2018 年中国 31 个省份的全要素生产率及其增长率进行了重新估计。相比较索洛余值法与空间计量全局分析的互嵌方法而言, 在全要素生产率核算中引入了面板时空地理加权回归模型, 将摒弃从所有局部点中找全要素生产率核定关键系数规律性的做法, 也摒弃了传统面板数据分析中纳入所有截面和所有时期的做法, 在单个局部点分析中仅仅纳入对其产生有效影响的截面和时期局部点进行相对独立的分析, 这使得在同一模型中既考虑空间依赖性质也考虑空间异质性成为可能, 也为空间计量局部分析模型与全局分析模型的进一步融合提供了契机和可能。

本文可能的创新点主要体现在以下几个方面: 第一, 在有效近邻局部点的纳入过程中, 不仅考虑了对目标分析地区产生有效影响的截面局部点和时期局部点, 还纳入了目标分析地区本身, 使时空地理加权回归模型的建模基础理念不仅包含了同伴效应, 还更考虑了自身的内生动力, 使得分析结论更接近现实。第二, 在时空权重矩阵设计过程中, 不仅仅是纳入了空间位置因素和时间因素, 更是考虑了空间溢出效应及其在时期上的转移和传导效应; 不仅仅纳入了近邻局部点对目标分析地区的直接影响, 还考虑所有近邻局部点对目标分析地区可能的间接影响, 从而更加真实、精准地模拟了空间局部点之间的时空溢出效应。第三, 在面板时空地理加权回归模型建模过程中, 基于随机扰动项方差估计值的局部和整体分解, 将整个分析过程分解成了单个局部点分析和模型整体分析两部分。其中, 单个局部点分析中仍然沿袭了面板数据计量经济学方法的经典分析范式, 使得理论和方法具有延续性和可延展性; 模型整体分析中则采用了模型平均的分析范式, 这种分解使建模过程更加简单、便利, 对计算机运行和存储能力的要求也大为降低。第四, 在对时空地理加权回归模型理论和方法的创新推进基础上, 作者还基于 MATLAB R2020a 软件撰写了一系列标准化程序代码, 可以让使用者仅需输入应用研究中相关被解释变量和解释变量的面板数据, 就可以自动优选自适应的空间带宽和时间带宽, 并自动完成局部参数估计及统计性质分析、模型整体性质分析、相关假设检验的全过程, 对基于面板时空地理加权回归模型展开应用研究大有裨益。

对面板时空地理加权回归模型的可能延展包括以下几个方向: 第一, 在单个局部点分析中摒弃经典面板数据计量模型的分析框架, 加入半参数理念或多尺度理念, 架构半参数面板时空地理加权回归模型的分析范式或多尺度面板时空地理加权回归模型的分析范式, 并创新基于线性可加模型和后向拟合法的参数迭代估计方法; 第二, 在单个局部点分析中嵌入空间计量全局模型, 在面板时空地理加权回归模型中实现全局模型和局部模型的进一步融合; 第三, 建立所有延展模型在同一框架下进行优选的假设检验逻辑, 更加真实、更加有效地遴选与现实问题相吻合的空间计量模型; 第四, 对所有延展模型及其优选逻辑进行 MATLAB 标准化代码写作和平台化建设, 使空间计量经济学在经济学应用研究中发挥更大的作用。

参 考 文 献

- [1] Atkinson P. M., German S. E., Sear D. A., Clark M. J., 2003, *Exploring the Relations between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression* [J], *Geographical Analysis*, 35 (1), 58~82.

- [2] Bárcena M. J. , Menéndez P. , Palacios M. B. , 2014, *Alleviating the Effect of Collinearity in Geographically Weighted Regression* [J], *Journal of Geographical Systems*, 16 (4), 441~466.
- [3] Bidanset P. , McCord M. , Lombard J. R. , Davis P. , McCluskey W. , 2018, *Accounting for Locational, Temporal, and Physical Similarity of Residential Sales in Mass Appraisal Modeling: Introducing the Development and Application of Geographically, Temporally, and Characteristically Weighted Regression* [J], *Journal of Property Tax Assessment & Administration*, 14 (2), 4~12.
- [4] Brunsdon C. , Aitkin M. , Fotheringham S. , Charlton M. , 1999, *A Comparison of Random Coefficient Modelling and Geographically Weighted Regression for Spatial Non-Stationary Regression Problems* [J], *Geographical & Environmental Modelling*, 3 (1), 47~62.
- [5] Brunsdon C. , Fotheringham A. S. , Charlton M. E. , 1996, *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity* [J], *Geographical Analysis*, 28 (4), 281~298.
- [6] Brunsdon C. , Fotheringham A. S. , Charlton M. , 2007, *Geographically Weighted Discriminant Analysis* [J], *Geographical Analysis*, 39 (4), 376~396.
- [7] Bursztyn L. , Ederer F. , Ferman B. , Yuchtman N. , 2014, *Understanding Mechanisms Underlying Peer Effects: Evidence from a Field Experiment on Financial Decisions* [J], *Econometrica*, 82 (4), 1273~1301.
- [8] Chen F. , Mei C. , 2021, *Scale-Adaptive Estimation of Mixed Geographically Weighted Regression Models* [J], *Economic Modelling*, 94, 737~747.
- [9] Chen V. Y. , Deng W. , Yang T. , Matthews S. A. , 2012, *Geographically Weighted Quantile Regression (GWQR): An Application to U. S. Mortality Data* [J], *Geographical Analysis*, 44, 134~150.
- [10] Comber A. , Harris P. , 2018, *Geographically Weighted Elastic Net Logistic Regression* [J], *Journal of Geographical Systems*, 20 (4), 317~341.
- [11] Du Z. , Wu S. , Zhang F. , Liu R. , Zhou Y. , 2018, *Extending Geographically and Temporally Weighted Regression to Account for Both Spatiotemporal Heterogeneity and Seasonal Variations in Coastal Seas* [J], *Ecological Informatics*, 43, 185~199.
- [12] Foster S. A. , Gorr W. L. , 1986, *An Adaptive Filter for Estimating Spatially Varying Parameters: Application to Modeling Police Hours Spent in Response to Calls for Service* [J], *Management Science*, 32 (7), 878~889.
- [13] Fotheringham A. S. , Brunsdon C. , Charlton M. , 2002, *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships* [M], Hoboken: Wiley.
- [14] Fotheringham A. S. , Crespo R. , Yao J. , 2015, *Geographical and Temporal Weighted Regression (GTWR)* [J], *Geographical Analysis*, 47 (4), 431~452.
- [15] Fotheringham A. S. , Oshan T. M. , 2016, *Geographically Weighted Regression and Multicollinearity: Dispelling the Myth* [J], *Journal of Geographical Systems*, 18 (4), 303~329.
- [16] Fotheringham A. S. , Yang W. , Kang W. , 2017, *Multiscale Geographically Weighted Regression (MGWR)* [J], *Annals of American Association of Geographers*, 107 (6), 1247~1265.
- [17] Harris P. , Fotheringham A. S. , Juggins S. , 2010, *Robust Geographically Weighted Regression: A Technique for Quantifying Spatial Relationships between Freshwater Acidification Critical Loads and Catchment Attributes* [J], *Annals of the Association of American Geographers*, 100 (2), 286~306.
- [18] Harris R. , Dong G. , Zhang W. , 2013, *Using Contextualized Geographically Weighted Regression to Model the Spatial Heterogeneity of Land Prices in Beijing, China* [J], *Transactions in GIS*, 17 (6), 901~919.
- [19] Huang B. , Wu B. , Barry M. , 2010, *Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices* [J], *International Journal of Geographical Information Science*, 24 (3), 383~401.
- [20] Imran M. , Stein A. , Zurita-Milla R. , 2015, *Using Geographically Weighted Regression Kriging for Crop Yield Mapping in West Africa* [J], *International Journal of Geographical Information Science*, 29

(2), 234~257.

[21] Kang D. , Dall'erba S. , 2016, *Exploring the Spatially Varying Innovation Capacity of the US Counties in the Framework of Griliches' Knowledge Production Function: A Mixed GWR Approach* [J], *Journal of Geographical Systems*, 18 (2), 125~157.

[22] Mei C. , He S. Fang K. , 2004, *A Note on the Mixed Geographically Weighted Regression Model* [J], *Journal of Regional Science*, 44 (1), 143~157.

[23] Mei C. , Xu M. , Wang N. , 2016, *A Bootstrap Test for Constant Coefficients in Geographically Weighted Regression Models* [J], *International Journal of Geographical Information Science*, 30 (8), 1622~1643.

[24] Murakami D. , Tsutsumi M. , 2015, *Area-to-Point Parameter Estimation with Geographically Weighted Regression* [J], *Journal of Geographical system*, 17 (3), 207~225.

[25] Nakaya T. , 2001, *Local Spatial Interaction Modelling Based on the Geographically Weighted Regression Approach* [J], *GeoJournal*, 53, 347~358.

[26] Nakaya T. , Fotheringham A. S. , Brunson C. , 2005, *Geographically Weighted Poisson Regression for Disease Association Mapping* [J], *Statistics in Medicine*, 24 (17), 2695~2717.

[27] Romano E. , Mateu J. , Butzbach O. , 2020, *Heteroskedastic Geographically Weighted Regression Model for Functional Data* [J], *Spatial Statistics*, 38, 100444.

[28] Shariat-Mohaymany A. , Shahri M. , Mirbagheri B. , Matkan A. A. , 2015, *Exploring Spatial Non-Stationarity and Varying Relationships between Crash Data and Related Factors Using Geographically Weighted Poisson Regression* [J], *Transactions in Gis*, 19 (2), 321~337.

[29] Silva A. R. D. , Fotheringham A. S. , 2016, *The Multiple Testing Issue in Geographically Weighted Regression* [J], *Geographical Analysis*, 48 (3), 233~247.

[30] Tasyurek M. , Celik M. , 2020, *RNN-GWR: A Geographically Weighted Regression Approach for Frequently Updated Data* [J], *Neurocomputing*, 399, 258~270.

[31] Tientao A. , Legros D. , Pichery M. C. , 2016, *Technology Spillover and TFP Growth: A Spatial Durbin Model* [J], *International Economics*, 145, 21~31.

[32] Wang N. , Mei C. , Yan X. , 2008, *Local Linear Estimation of Spatially Varying Coefficient Models: An Improvement on the Geographically Weighted Regression Technique* [J], *Environment and Planning A*, 40 (4), 986~1005.

[33] Wheeler D. , Tiefelsdorf M. , 2005, *Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression* [J], *Journal of Geographical Systems*, 7 (2), 161~187.

[34] Wu B. , Li R. , Huang B. , 2014, *A Geographically and Temporally Weighted Autoregressive Model with Application to Housing Prices* [J], *International Journal of Geographical Information Science*, 28 (5), 1186~1204.

[35] Zhang H. , Mei C. , 2011, *Local Least Absolute Deviation Estimation of Spatially Varying Coefficient Models: Robust Geographically Weighted Regression Approaches* [J], *International Journal of Geographical Information Science*, 25 (9), 1467~1489.

[36] 范巧、Hudson Darren:《一种新的包含可变时间效应的内生时空权重矩阵构建方法》[J],《数量经济技术经济研究》2018年第1期。

[37] 范巧、郭爱君:《一种嵌入空间计量分析的全要素生产率核算改进方法》[J],《数量经济技术经济研究》2019年第8期。

[38] 钟丽云、张以谟、吕晓旭:《数字全息中的一些基本问题分析》[J],《光学学报》2004年第4期。

A New Geographically and Temporally Weighted Regression Model for Panel Data Based on Holographic Mapping

Fan Qiao Guo Aijun

(School of Economics, Lanzhou University)

Research Objectives: Construct a new geographically and temporally weighted regression model for panel data (PGTWR) based on the holographic mapping. **Research Methods:** This paper constructed an endogenous spatiotemporal weight matrix based on the holographic mapping from the adjacent points to the target points, and also constructed the PGTWR model by two steps, including parameters estimation of local points and the overall model property analysis. On this basis, this paper reassessed the total factor productivity and its growth rate at the provincial level in mainland China during 1990~2018 using PGTWR and the Solow Residual Method. **Research Findings:** The PGTWR model based on holographic mapping can more accurately analyze the spatial and temporal relationships among spatial local points. In the PGTWR analysis, each local point and its effective adjacent points are chosen by the optimal spatial and temporal bandwidth to study the laws of spatiotemporal dependence and its heterogeneity. Also, The PGTWR model not only considers the peer effect from the adjacent points but also the endogenous power from each local point itself. **Research Innovations:** Promotes a new PGTWR analysis based on holographic mapping, and provides a set of corresponding standardized code based on Matlab R2020a. **Research Value:** Makes important marginal contribution for theoretical researches of the PGTWR analysis, and provides standardized Matlab code for the application research of the spatial econometrical local analysis with panel data.

Key Words: Geographically and Temporally Weighted Regression Model for Panel Data (PGTWR); Spatiotemporal Weight Matrix; Peer Effects; Holographic Mapping

JEL Classification: C18; C51; R15

(责任编辑: 白延涛)