

# 博士学习期间研究构想

唐洁\*

(广西师范大学数学与统计学院, 广西桂林, 541004)

**1 拟解决问题：**高维情形下空间数据模型的经验似然推断.

**2 知识储备：**高等数学；线性代数；概率论与数理统计；极限理论等专业基础知识.

**3 创新点：**

首次考虑高维情形下的空间数据模型的经验似然推断；我们提出创新的经验似然方法并成功应用到高维情形的空间模型；模拟实验证明创新的经验似然方法有非常接近名义水平的覆盖率且计算速度非常快，无须增加惩罚项等复杂的过程.

**4 研究框架：**

拟研究问题的来源、研究的目的意义、国内外研究现状及水平.

(一) 拟研究问题的来源

学位论文选题来源于硕士导师主持的国家自然科学基金项目 (12061017, 12161009).

(二) 研究目的和意义

众所周知，各省之间的经济不是完全独立的，经济数据涉及到一定的空间位置关系，正如地理学第一定律所述，所有事物都与其他事物相关联，但较近的事物比较远的事物更关联，也就是说地理单元之间存在着交互效用（空间相依性）。然而，在同一个空间系统中，地理单元之间也存在差异（空间异质性）。同时，如同时间序列一样，过去的空间单元对现在的空间单元也存在着影响，产生了动态效应，如某时某地的 GDP、房价、病毒感染人数，等等。空间相依性、异质性和动态效应，推动了空间计量经济模型（简称空间模型）的建立。空间计量经济学的研究发展迅速，并在许多不同的科学领域中得到广泛的应用，比如区域经济学、人口统计学、传染病学、城市规划学、政治学和心理学等等。关于空间计量经济模型的估计方法大多采用（拟）极大似然（QML）方法、两阶段最小二乘（2SLS）方法和广义矩（GMM）方法等，且理论相对成熟。这些主流的估计手段集中在参数方法，而经验似然推断作为模型参数估计和区间估计的一种行之有效的非参数统计方法，在空间计量经济模型的理论研究和应用上还没有得到大力推广。因此，研究空间计量经济模型的经验似然推断的理论与应用具有现实意义。

(三) 国内外研究现状及水平

---

\*基本信息：唐洁，1995 年生，女，湖南人。报名号：1000297301。本科专业：数学与应用数学。研究生专业：统计学，研究方向：空间计量，经验似然。

自从 Cliff and Ord (1973) 首次考虑到空间效应之后, Anselin (1988) 从计量经济学角度处理空间效应, 将空间效应视为计量经济学模型中一般问题的特例, Cressie (1993) 对空间数据类型进行划分, 由此空间计量模型逐步由边缘步入主流. 此时的空间模型主要分为两类, 一是静态**截面**空间数据模型, 二是静态**面板**空间数据模型, 面板是截面空间数据模型的推广, 不仅包含同一空间系统里不同空间之间的交互效应和异质效应, 还考虑了时间因素. 第三类是由 Anselin (2001) 提出**动态**面板空间数据模型, 不仅考虑了被解释变量和解释变量在空间上的滞后, 还考虑了它们在时间上的滞后, 以及序列误差的自相关性.

关于第一类空间数据模型的相关研究有 Kelejian and Prucha (1998) 提出使用广义空间两段最小二乘法估计含空间误差的空间自回归截面数据模型的参数, Kelejian and Prucha (1999) 提出使用广义矩估计方法研究截面空间自回归模型, Kelejian et al. (2004) 推荐使用工具变量的迭代版本估计空间自回归截面数据模型的参数, Lee (2004) 证明了截面空间自回归模型的拟极大似然估计的渐近性质, Kelejian and Prucha (2006) 研究了在空间模型中方差-协方差矩阵的异方差和自相关系数的相合估计, Arraiz et al. (2010) 通过蒙特卡罗模拟表明, 在异方差情形下拟极大似然估计量不是相合估计, 同时证明在异方差情形下工具变量法导出的估计量为相合估计, 更多文献可参见 Cliff and Ord (1973), Anselin (1988), Cressie (1993), Kelejian and Prucha (1998, 1999, 2004, 2006), Liu et al. (2010) 等等. 这些空间截面数据模型下的研究情况, 其中的方法在空间面板数据模型下同样适用, 关于第二类空间数据模型的相关研究可参见 Anselin (1988), Elhorst (2003), Baltagi et al. (2003), Anselin et al. (2008), Parent and LeSage (2011), Baltagi et al. (2013), Lee and Yu (2016) 等等. 关于第三类空间数据模型的相关研究目前较少, 可参见 Yang et al. (2006), Yu et al. (2008), Lee and Yu (2010), Elhorst (2012), Su and Yang (2015), Qu et al. (2017) 等等. Anselin (2010) 对空间计量经济学领域在 2010 年之前的 30 年中的发展作了详细的阐述.

下面我们首先介绍本项目感兴趣的空间计量经济模型, 经验似然方法的研究进展, 并提出本项目研究的可行性和必要性.

### (1) 空间数据模型的定义

一般的空间数据模型如下所示, 该模型也称为带固定效应的动态含空间自回归误差的空间自回归面板数据模型:

$$\begin{cases} y_t = \lambda y_{t-1} + \rho_1 W_n y_t + x_t \beta + z \gamma + u_t, \\ u_t = \rho_2 M_n u_t + \varepsilon_t, \end{cases} \quad t = 1, 2, \dots, T, \quad (1)$$

其中,  $n$  和  $T$  分别表示是空间单元数和时间单位数,  $y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$  是  $n \times 1$  维

被解释变量的观测值,  $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$  是  $n \times p$  维解释变量随时间变化的样本资料矩阵,  $z = (z_1, z_2, \dots, z_n)'$  是  $n \times q$  维解释变量不随时间变化的样本资料矩阵, 标量  $\lambda$  ( $|\lambda| < 1$ ) 是动态效应因子,  $\rho_j$  ( $|\rho_j| < 1$ ),  $j = 1, 2$  是空间自相关系数,  $\beta$  是  $p \times 1$  维  $x_t$  的回归系数向量,  $\gamma$  是  $q \times 1$  维  $z$  的回归系数向量,  $W_n$  是解释变量  $y_t$  的空间邻接权重矩阵,  $M_n$  是扰动项  $u_t$  的空间邻接权重矩阵, 它们都是预先给定的  $n \times n$  空间权重矩阵, 二者可以相等,  $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})'$  是  $n \times 1$  维误差向量,  $\varepsilon_{it}$  为独立同分布的随机误差项, 且满足  $E(\varepsilon_{it}) = 0$ ,  $Var(\varepsilon_{it}^2) = \sigma^2$ , 其中  $0 < \sigma^2 < \infty$ .

模型 (1) 可写成如下的矩阵形式:

$$\begin{pmatrix} A_n & 0 & \cdots & 0 \\ 0 & A_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \lambda \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \beta + \begin{pmatrix} z \\ z \\ \vdots \\ z \end{pmatrix} \gamma + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

和

$$\begin{pmatrix} B_n & 0 & \cdots & 0 \\ 0 & B_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_n \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

或者

$$\begin{cases} (I_T \otimes A_n)Y = \lambda Y_{-1} + X\beta + Z\gamma + u, \\ (I_T \otimes B_n)u = \varepsilon, \end{cases}$$

其中  $Y = (y'_1, y'_2, \dots, y'_T)'$ ,  $Y_{-1} = (y'_0, y'_1, \dots, y'_{T-1})'$ ,  $X = (x'_1, x'_2, \dots, x'_T)'$ ,  $Z = \mathbf{1}_T \otimes z$ ,  $u = (u'_1, u'_2, \dots, u'_T)'$ ,  $\varepsilon = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_T)'$ ,  $A_n = A_n(\lambda) = I_n - \rho_1 W_n$  以及  $B_n = B_n(\lambda) = I_n - \rho_2 M_n$ .

当  $t$  为固定时间点且  $\lambda = 0$  时, 模型 (1) 即为带固定效应的含空间自回归误差的空间自回归截面数据模型:

$$\begin{cases} y_t = \rho_1 W_n y_t + x_t \beta + z \gamma + u_t, \\ u_t = \rho_2 M_n u_t + \varepsilon_t. \end{cases}$$

当  $t$  为固定时间点且  $\lambda = 0, \gamma = 0$  时, 模型 (1) 即为不带固定效应的含空间自回归误差的空间自回归截面数据模型:

$$\begin{cases} y_t = \rho_1 W_n y_t + x_t \beta + u_t, \\ u_t = \rho_2 M_n u_t + \varepsilon_t. \end{cases} \quad (2)$$

若  $\rho_1 = \rho_2 = \gamma = 0$ , 模型 (1) 即为一般线性模型:

$$\begin{cases} y_t = x_t \beta + u_t, \\ u_t = \varepsilon_t. \end{cases}$$

当  $t$  为时间序列且  $\rho_1 = 0$  时, 模型 (1) 即为带固定效应的动态含空间误差的面板数据模型:

$$\begin{cases} y_t = \lambda y_{t-1} + x_t \beta + z \gamma + u_t, \\ u_t = \rho_2 M_n u_t + \varepsilon_t, \end{cases} \quad t = 1, 2, \dots, T,$$

## (2) 拟研究的空间数据模型

拟研究问题从不带固定效应的空间截面数据模型 (2) 开始, 并简写为 SARAR 模型, 用新的符号记为:

$$Y_n = \rho_1 W_n Y_n + X_n \beta + u_{(n)}, u_{(n)} = \rho_2 M_n u_{(n)} + \epsilon_{(n)},$$

其中,  $n$  是空间单元数量,  $\rho_j, j = 1, 2$  是空间自回归系数且  $|\rho_j| < 1, j = 1, 2$ ,  $X_n = (x_1, x_2, \dots, x_n)'$  是  $n \times p$  维解释变量的样本资料矩阵,  $\beta$  是  $p \times 1$  维  $X_n$  的回归系数向量,  $Y_n = (y_1, y_2, \dots, y_n)'$  是  $n \times 1$  维响应变量,  $W_n$  是解释变量  $Y_n$  的空间邻接权重矩阵,  $M_n$  是扰动项  $u_{(n)}$  的空间邻接权重矩阵, 它们都是已知的  $n \times n$  空间邻接权重矩阵 (非随机), 二者可以相等,  $\epsilon_{(n)}$  是  $n \times 1$  维空间误差向量, 满足  $E\epsilon_{(n)} = 0, Var(\epsilon_{(n)}) = \sigma^2 I_n$ .

SARAR 模型对存在空间依赖性的数据有较好的解释作用, 无论是滞后项存在空间依赖性还是扰动项存在空间依赖性, 空间依赖性由  $\rho_j$  刻画,  $\rho_1$  度量空间滞后项  $W_n Y_n$  对解释变量  $Y_n$  的影响, 也就是  $Y_n$  之间的空间依赖性, 即相邻空间之间可能存在扩散、溢出等效应,  $\rho_2$  度量空间误差滞后项  $M_n u_{(n)}$  对扰动项  $u_{(n)}$  的影响, 也就是不包含在  $X_n$  中且对  $Y_n$  有影响的遗漏变量的空间依赖性. SARAR 模型是更一般的空间计量截面数据模型, 当  $\rho_1 = \rho_2 = 0$  时, 该模型即为一般线性模型; 当  $\beta = 0, \rho_2 = 0$  时, 该模型为最简单的空间截面数据模型; 当  $\rho_1 = 0$  时, 该模型为空间误差模型, 简写为 SEM 模型; 当

$\rho_2 = 0$  时, 该模型为空间自回归模型, 简称为 SAR 模型.

我们拟研究空间解释变量数据维数  $p \rightarrow \infty$  的情形, 即高维情形.

## (2) 经验似然方法研究进展

经验似然方法是 Owen (1988) 提出的在完全样本下的一种非参数统计推断方法, 在一定约束条件下可以将参数似然比极大化, 具有类似于 Bootstrap 的抽样特性. 经验似然方法比传统的或现代的统计方法具有很多突出的优点, 比如用经验似然构造置信区间具有域保持性、变换不变性、置信区间的形状由数据本身自行决定, 还有 Bartlett 纠偏性和无需构造轴统计量等优点. 经验似然的这些优良特性深受许多统计学者的欢迎, Qin and Lawless (1994) 在 Owen 基础上给出了一般情况下的经验似然估计方法, 从此, 经验似然得到了广泛地应用. 更多经验似然的参考文献可参见 Chen and Qin (1993), Qin and Lawless (1994), Zhong and Rao (2000), Owen (1988, 1990, 1991, 2001) and Wu (2004) 等.

经验似然方法虽然具有上述提到的许多优点, 在实际应用中并没有被更多的计量经济学家和实证研究者关注, 经验似然方法在空间数据模型中的应用方面的已有成果较少, 原因是经验似然应用于空间模型时由拟似然函数构造的估计方程是误差的线性-二次型形式, 这为得分函数的构造带来困难. 我们经过研究发现, 对于多数空间计量经济模型, 这样的情形下可以构造一个鞅差序列, 把线性-二次型转化为鞅差序列的线性形式, 且不需要对数据进行分组, 可以直接利用经验似然方法, 这为非参数方法在空间模型上的应用迈进一大步. 此结论由 Jin and Lee 以及本人导师秦永松教授独立发现, 并成功应用到含空间自回归误差的空间自回归截面数据模型的研究, Jin and Lee (2019) 利用广义经验似然方法研究 SARAR 模型的估计和检验, 证明其广义经验似然估计与广义矩估计方法具有相同的渐近分布, Qin (2021) 利用经验似然方法构造 SARAR 模型的经验似然比统计量, 证明其经验似然比统计量是渐近卡方分布, 由此关于空间模型的经验似然统计推断的研究蓬勃发展, Qin and Lei (2022) 详细介绍了空间计量模型的经验似然研究进展, 更多文献可见 Qin and Lei (2021), Li and Qin (2022), Rong et al. (2021) 等等.

理论上, 经验似然方法能在空间计量模型上得到应用, 然而实际中仍存在一些问题仍有许多问题值得研究. 在分析高维数据过程中面临的一大挑战是维数的膨胀, 也就是通常所说的“维数灾难”. 有研究表明, 当维数越来越大时, 分析和处理数据的成本于复杂度成指数级增长. 在对高维数据进行分析时, 所需的样本量会随着维数的增加而呈指数增长. 处理这类问题的非参数方法主要依赖大样本理论, 但会出现样本量相比数据维数显得较少的情况, 从而大样本理论处理高维数据方法失效. 另外, 传统的经验似然方法在处理高维数据时不能满足稳健性要求; 高维导致样本量变少使得理论上的渐近性难以实现; 维数的增加导致数据的计算量迅速上升等等.

最早由 Dempster 发现当维数大于样本容量时, 传统 Hotelling 检验没用定义, 因为



样本协方差矩阵不可逆, 并发表 Dempster (1958) 和 Dempster (1960) 两篇论文提出一种非精确检验 (non-exacttest) 来解决高维数据的两样本均值问题, 关于高维数据的两样本均值检验问题更多文献可见 Bai and Saranadasa (1996), Chen and Qin (2010), Wang and Xu (2022) 等等. 随着高维数据的出现越来越频繁, 变量的维数  $p$  的增加使得传统统计推断的精度变低, 日常数据分析亟需在之前的基础上寻找或改进为新的统计方法来处理高维数据, 研究高维数据成为当下统计学的热点问题.

在经验似然领域中, 处理高维数据的常见手段分两种, 一类以修正原来的经验似然比统计量为基础, 另一类采用降维的思想. 第一类中, Shi (2007) 指出, 当协变量维数以合理的速度趋于无穷大时, 仍可利用经验似然方法构造高维线性模型参数的置信域, 不过修正的经验似然统计量的渐近分布为正态分布而非卡方. Hjort et al. (2009) 从理论上拓展了经验似然方法的适用范围, 即当  $p = o_p(n^{1/3}) \rightarrow \infty$  时, 在一定条件下, 经验似然比统计量渐近分布为正态分布. Chen (2009) 研究了均值模型下数据维数对经验似然的影响, 并证明当  $p = o_p(n^{1/2}) \rightarrow \infty$  时, 比 Hjort 更少的约束条件下, 经验似然比统计量渐近分布为正态分布. Li et al (2012) 研究了高维变系数部分线性模型的经验似然推断, 提出纠偏经验似然方法, 并证明其统计量渐近分布为正态. 更多文献可见 Tang et al. (2013), Liu et al. (2013), Fang et al. (2017) 等等. 这类方法虽然使得统计量不发散, 但都以  $p < n$  为前提在应用中有所局限.

第二类中, 在保证重要信息不损失的情形下对高维进行降维, 其中著名的方法就是惩罚经验似然. 惩罚的思想是在进行参数估计的同时, 利用惩罚函数将较小的系数估计值压缩为零, 而将系数估计值较大的保留, 在估计出系数的同时选择出重要变量, 这可以同时实现变量选择和系数估计两个目标. 惩罚变量选择普遍采用“损失函数 + 惩罚函数”的方式, 惩罚函数的选择有很多, 比如熵惩罚、桥回归 (Frank and Friedman, 1993)、LASSO 惩罚 (Tibshiani, 1996)、岭回归、硬门限惩罚 (Antoniadis, 1997) 等等, 一个关键问题是如何选择惩罚函数. Antoniadis and Fan (2001) 与 Fan and Li (2001) 对如何选择惩罚函数提出深刻的见解, 他们在 Fan and Li (2001) 一文中提出 SCAD 惩罚方法, 该方法得到的估计量具有无偏性、稀疏性、连续性. 类似地, 惩罚经验似然也使用“经验似然比函数 + 惩罚函数”方式. Tang and Leng (2010) 首次将惩罚经验似然方法于高维分析多变量的均值向量和线性模型的高维回归系数, 证明惩罚经验似然具有与一般经验似然的类似性质, 例如, 依靠数据确定置信区域的形状和取向, 无须估计共协方差, 相同的渐近分布等, 研究高维数据的非参数方法开辟新的道路. Leng and Tang (2012) 将高维惩罚经验似然方法应用于一般估计方程的参数估计和变量选择, 并证明惩罚经验似然方法具有 oracle 特征. Lahiri and Mukhopadhyay (2012) 推广高维中一种惩罚经验似然方法可适用于  $p > n$  的情形, 成功应用于总体均值模型. 至此, 高维经验似然方法备受学者们关注, 比如 Fang et al. (2017) 研究了半参数模型的高维惩罚经验似然, Yan

and Chen(2018) 研究了高维广义线性模型的惩罚拟似然 SCAD 估计, 更多文献可参见 Lahiri and Mukhopadhyay (2012), Peng et al. (2014), Yan and Chen (2018) 等等.

目前未见有当数据维数发散时关于空间截面数据模型的经验似然方法的相关研究报道, 因而拟研究问题属于科学前沿问题. 关于降维的方式, 我们拟采用主成分分析, 无须选择惩罚函数, 这样既可以保证信息损失较少又可以避免复杂的计算过程, 目前经验似然领域尚未有主成分分析思想的相关论文, 因此, 拟研究的问题丰富了高维情形的经验似然领域理论, 有值得研究的价值.

## 5 研究方法:

拟研究问题的主要内容、方案和准备采取的措施.

### (一) 研究主要内容

拟研究含空间数据模型的高维经验似然推断问题, 证明其高维经验似然比统计量的极限分布渐近服从卡方分布, 构造该模型参数的经验似然置信区间 (域), 并通过模拟实验验证这些置信区间 (域) 的优良性.

### (二) 拟解决的关键问题

本项目有以下几个关键问题需要解决:

(1) 经验似然方法的使用过程中, 得分方程或估计方程的确定是一个重要的前提, 相比通常的数据模型, 空间或时空数据模型不易得到得分方程, 即使找到了得分函数, 其形式往往也比较复杂, 不方便直接使用, 故在研究其经验似然推断时得分方程的确定是一个关键问题;

(2) 传统经验似然估计方程的维数发散, 如何进行有效的降维是一个难题;

(3) 在经验似然方法的研究中, 如何得到并证明高维经验似然比统计量的极限分布是一个需要解决的关键问题;

(4) 比较普通经验似然方法与要研究的估计方法的优劣也是一个需要解决的关键问题.

### (三) 拟采取的研究方法、手段或技术路线

(1) 采用拟极大似然 (QML) 方法或广义矩方法 (GMM) 得到的估计方程作为经验似然方法的初次估计方程, 如果初次估计方程是模型误差的线性-二次型, 则可以采用参考文献 Jin and Lee (2019) 和 Qin (2021) 中的方法, 构造一个鞅差序列, 将线性-二次型转化为鞅差序列的线性形式, 由此得到一般经验似然方法的最终估计方程.

(2) 对所得的高维估计方程进行选择, 将其由高维转化为低维, 如何选择是一个难点, 既要压缩信息又要充分利用信息, 拟采用手段是对既有的一般经验似然渐近协方差进行主成分分析.

(3) 高维经验似然的最终估计方程是鞅差序列的线性函数, 在证明经验似然比统计量的渐近卡方分布时, 可利用鞅差序列的极限理论.

(4) 在高维情形下, 拟研究出主成分经验似然方法构造则高维经验似然比统计量并证明其统计量服从卡方分布, 并设计实验与传统经验似然方法对比.

#### 参考文献:

- Antoniadis, A., 1997. Wavelets in statistics: A review. *Journal Italian Statistics Association*, 6, 97-144.
- Antoniadis, A., Fan, J., 2001. Regularization of wavelets approximations. *Journal of the American Statistical Association*, 96, 939-967.
- Anselin, L., 1988. *Spatial econometrics: methods and models*. Berlin: Springer.
- Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. H. (Ed.), *A companion to theoretical econometrics*. Blackwell Publishers Ltd., Massachusetts, 310-330.
- Anselin, L., Le Gallo, J., Jayet, H., 2008. Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (Eds.), *The econometrics of panel data: fundamentals and recent developments in theory and practice*. SpringerVerlag, Berlin Heidelberg, 625-660.
- Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science*, 89, 3-25.
- Arraiz, I., Drukker, D. M., Kelejian, H. H., Prucha, I. R., 2010. A spatial cliff-ord-type model with hetero-skedastic innovations: small and large sample results. *Journal of Regional Science*, 50, 592-614.
- Baltagi, B. H., Song, S. H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117, 123-150.
- Baltagi, B. H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel data model with random effects. *Econometric Reviews*, 32, 650-685.
- Bai, Z. and Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6, 311-329.
- Chen, J., Qin, J., 1993. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80, 107-116.
- Chen, J., Variyath, A. M., Abraham, B., 2008. Adjusted empirical likelihood and its properties. *Journal of computational and graphical statistics*, 17, 426-443.
- Chen, S. X., Qin, Y. L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38, 808-835.
- Chen, S. X., Peng, L., Qin, Y. L., 2009. Effects of data dimension on empirical likelihood. *Biometrika*, 96, 711-722.
- Cliff, A. D., Ord, J. K., 1973. *Spatial autocorrelation*. London: Pion Ltd.



- Cressie, N., 1993. Statistics for spatial data. New York: Wiley.
- Dempster, A. P., 1958. A high dimensional two sample significance test. The Annals of Mathematical Statistics, 29, 995-1010.
- Dempster, A. P., 1960. A significance test for the separation of two highly multivariate small samples. Biometrics, 16, 41-50.
- Elhorst, J.P., 2003. Specification and estimation of spatial panel data models. International Regional Science Review, 26, 244-268.
- Elhorst, J. P., 2012. Dynamic spatial panels: models, methods, and inferences. *Journal of geographical systems*, 14, 5-28.
- Fang, J. L., Liu, W. R., Lu, X. W., 2017. Penalized empirical likelihood for semiparametric models with a diverging number of parameters. Journal of Statistical Planning and Inference, 186, 42-57.
- Frank, I. E., Friedman, J. H., 1993. A statistical view of some chemometrics regression tools (with discussion). Technometrics, 35, 109-148.
- Fan, J. Q., Li, R. Z., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348-1360.
- Hjort, N. L., Eague, I. W., Eilegom, I. V., 2009. Extending the scope of empirical likelihood. The Annals of Statistics., 37, 1079-1111.
- Jin, F., Lee, L. F., 2019. GEL estimation and tests of spatial autoregressive models. Journal of Econometrics, 208, 585-612.
- Kelejian, H. H., Prucha, L. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics, 17, 99-121.
- Kelejian, H. H., Prucha, L. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40, 509-33.
- Kelejian, H. H., Prucha, I. R., Yuzefovich, Y., 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results. Spatial and Spatiotemporal Econometrics, 18, 163-198.
- Kelejian, H. H., Prucha, I. R., 2006. HAC estimation in a spatial framework. Journal of Econometrics, 140, 131-154.
- Lahiri, S. N. and Mukhopadhyay, S., 2012. A penalized empirical likelihood method in high dimensions. The Annals of Statistics, 40, 2511-2540.
- Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial auto-regressive models. Econometrica, 72, 1899-1925.

- Lee, L. F., Yu, J., 2016. Identification of Spatial Durbin Panel Models. *Applied Economics*, 31, 133-162.
- Lee, L. F., Yu, J., 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26, 564-597.
- Leng, C. L., Tang, X. Y., 2012. Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika*, 99, 706-716.
- Liu, Y., Zou, C., Wang, Z., 2013. Calibration of the empirical likelihood for high-dimensional data. *Annals of the Institute of Statistical Mathematics*, 69, 529-550.
- Li, Y. H., & Qin, Y. S., 2022. Empirical likelihood for spatial dynamic panel data models, *Journal of the Korean Statistical Society*, 51, 500-525.
- Li, G. R., Lin, L., Zhu, L. X., 2012. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. *Journal of Multivariate Analysis*, 105, 85-111.
- Liu, X., Lee, L. F., Bollinger, C. R., 2010. An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics*, 159, 303-319.
- Owen, A. B., 1988. Empirical Likelihood Ratio Confidence Intervals for a single functional. *Biometrika*, 75, 237-249.
- Owen, A. B., 1990. Empirical likelihood ratio confidence regions. *Ann. Statist.*, 18, 90-120.
- Owen, A. B., 1991. Empirical likelihood for linear models. *Ann. Statist.*, 19, 1725-1747.
- Owen, A. B., 2001. Empirical likelihood. London: Chapman & Hall.
- Parent, O., LeSage, J. P., 2011. A space-time filter for panel data models containing random effects. *Computational Statistics & Data Analysis*, 55, 475-490.
- Peng, L., Qi, Y. C., Wang, R. D., 2014. Empirical likelihood test for high dimensional linear models. *Statistics and Probability Letters*, 86, 85-90.
- Piribauer, P., Fischer, M. M., 2014. Model Uncertainty in Matrix Exponential Spatial Growth Regression Models. *Geographical Analysis*, 47, 240-261.
- Qu, X., Lee, L. F., Yu, J., 2017. QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Econometrics*, 197, 173-201.
- Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. *Annals of Statistics*, 22, 300-325.
- Qin, Y., 2021. Empirical likelihood for spatial autoregressive models with spatial autoregressive disturbances. *Sankhyā A: Indian J. Statist.*, 83, 1-25.

- Qin, Y. S., Lei, Q. Z., 2021. Empirical likelihood for mixed regressive, spatial autoregressive model based on GMM. *Sankhyā A: Indian J. Statist.*, 83, 353-378.
- Qin, Y. S., Lei, Q. Z., 2022. Review on empirical likelihood for spatial econometric models. *Journal of Guangxi Normal University ( Natural Science Edition)*, 40, 138-149.
- Rong, J. R., Liu, Y., Qin, Y. S., 2021. Empirical likelihood for spatial dynamic panel data models with spatial lags and spatial errors. *Communications in Statistics-Theory and Methods*. <https://doi.org/10.1080/03610926.2022.2032172>.
- Shi, J., 2007. Empirical likelihood for higher dimensional linear models. *J.Sys.Sci. & Math.Scis.*, 27, 124-133.
- Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. *Econometrics*, 185, 230-258.
- Tang, X. Y., Li, J. B., Lian, H., 2013. Empirical likelihood for partially linear proportional hazards models with growing dimensions . *Journal of Multivariate Analysis*, 121, 22-32.
- Tang, X. Y., Leng, C. L., 2010. Penalized high-dimensional empirical likelihood. *Biometrika*, 97, 905-920.
- Tibshiani, R. J., 1996, Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58, 267-288.
- Wang, R., Xu, W. L., 2022. An approximate randomization test for high-dimensional two-sample Behrens-Fisher problem under arbitrary covariances. *Biometrika*. <https://doi.org/10.1093/biomet/asac014>.
- Wu, C. B., 2004. Weighted empirical likelihood inference. *Statistics & Probability Letters*, 66, 67-79.
- Yan, L., Chen, X., 2018. Penalized quasi-Likelihood SCAD estimator in high-dimensional generalized linear models. *J. Wuhan Univ.*, 64, 533-539.
- Yang, Z., Li, C., Tse, Y.K., 2006. Functional form and spatial dependence in dynamic panels. *Economics Letters*, 91, 138-145.
- Yu, J., R, de Jong., Lee, L.F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics*, 146, 118-134.
- Zhong, B., Rao, J. N. K., 2000. Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, 87, 929-938.