

分类号 \_\_\_\_\_  
UDC \_\_\_\_\_

密级 \_\_\_\_\_  
编号 \_\_\_\_\_

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

# 博士学位论文

论文题目 基于密度幂散度族的稳健小域估计

研究生姓名: 王朝旭

指导教师姓名、职称: 庞智强 教授

学科、专业名称: 统计学、统计学

研究方向: 小域估计

提交日期: 2023年5月



Robust small area estimation with  
density power divergences

A Economics submitted to the  
Lanzhou University of Finance and Economics  
in partial fulfillment of the requirement  
for the degree of  
Doctor of Statistics  
in Statistics

By

Zhaoxu Wang

Supervisor: Professor Zhiqiang Pang

Lanzhou University of Finance and Economics

May, 2023



# 独创性声明

本人声明所提交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 王朝旭 签字日期： 2023.5.30

导师签名： 张碧强 签字日期： 2023.5.30

# 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 王朝旭 签字日期： 2023.5.30

导师签名： 张碧强 签字日期： 2023.5.30



## 摘要

在统计推断中,如何利用样本数据对总体目标变量进行有效的估计是非常普遍的研究问题,这类问题在实践应用中也很常见,其估计方法具有广泛的应用需求。基于抽样设计的直接估计方法是解决这类问题最直接的思路。然而,当估计总体区域由众多的小域构成时,估计小域上的目标变量时会面临小样本乃至无样本的情形。在这种情形下,利用样本的直接估计方法可能会产生较大的误差或无法得到有效估计。而小域估计方法则是解决这类问题的有效方法之一。相较于传统抽样估计方法,小域估计方法借助于辅助变量的信息能够得到小区域上目标变量的有效估计,能够解决小样本和无样本情形下的估计问题。近年来,小域估计方法在人口统计、生物统计、农业统计和政府统计等领域应用非常广泛,相关的学术研究也较为丰富,使小域估计的理论得到了系统发展。

基于模型的小域估计方法作为小域估计的主要方法,是小域估计的核心内容。基于模型的估计方法能够很好地将辅助变量应用于估计模型,从而达到“借力”的作用,以此来解决小样本和无样本问题。在小域估计的模型中,通常假设区域随机效应和模型随机误差均服从正态分布。但实践证明,当存在异常观测值时,基于正态分布假设的小域模型假设失效,这会导致参数估计和目标变量估计产生较大的偏差。因此,需要探究对异常观测值不敏感的稳健估计方法。在小域稳健估计中,目前应用较为广泛的方法有两类。一类方法为假设模型误差为有偏分布的估计方法,例如假设模型误差服从  $t$  分布或柯西分布,通过有偏分布建模来减小异常观测值对估计量的影响;另一类方法为利用 Huber  $\phi$  函数对经验线性无偏估计量进行稳健化处理,通过 Huber  $\phi$  函数的优良性质来达到稳健性的目的。虽然这两类方法在大多数情况下能有效降低异常观测值对估计量的影响,但当异常值过大时,其估计效果将受到限制,估计结果仍然会存在较大的偏差。因此,在实际应用中需要针对具体情况选择恰当的小域稳健估计方法进行估计。

研究稳健小域估计方法,是当前小域估计中非常现实的问题。由于非正态性观测数据的普遍存在以及异常值的出现,对小域估计方法提出了新的挑战。为了解决该情形下估计量不稳定以及预测值的大偏差问题,稳健小域估计方法被众多学者所关注到。本研究考虑到密度幂散度族在稳健估计中的重要特性,将其应用到小域估计中,提出基于密度幂散度族的稳健估计方法,以弥补现有稳健小域估计的不足。通过将密度幂散度族应用于小域估计中,探究非正态以及具有异常观

测情形下小域模型系数和目标变量的估计问题。在本研究中,旨在构建小域模型系数、目标变量的稳健估计量,并给出参数的置信区间以及估计量的均方误差。

首先,为了解决区域水平模型中的稳健估计问题,探究了密度幂散度和  $\gamma$  散度在 FH 模型中的应用。通过将密度幂散度应用于 FH 模型,得到了模型系数的稳健估计和渐进分布。在此基础上,讨论了目标变量的稳健估计量,并给出了估计量的均方误差。为了得到小域估计量的可靠估计,本研究同时给出了其目标变量的置信区间。通过对模拟数据和实际数据建立小域模型,结合本研究提出的稳健估计方法,进行稳健估计,并和现有的稳健估计方法进行了比较。通过比较发现,提出的估计方法能够通过调整参数控制估计的有效性和稳健性之间的平衡。当观测数据中不存在异常值时,本研究提出的估计方法通过使用较小的调整参数和现有的最优线性无偏估计方法得到的稳健估计效果相差不大。在观测数据中存在异常值时,提出的稳健估计方法比起现有的估计方法具有更小的均方误差,说明本研究提出的估计方法是有效的。

其次,研究了基于密度幂散度、 $\gamma$  散度的单元水平模型的稳健估计问题。将这两类散度应用于 NER 模型,对模型的系数进行了稳健估计,得到了模型系数的稳健估计及其渐进分布。在单元水平模型下,讨论了区域上关于目标变量函数形式的稳健估计量以及有限总体的区域均值的稳健估计。由于单元水平模型中关于目标变量的估计中会涉及多重积分的计算,本研究利用 MCMC 的方法给出了关于目标变量的函数形式的估计值,并结合 Bootstrap 方法给出了估计量的 MSE。同样地,将本研究提出的估计方法和现有的稳健估计方法进行了比较。通过对模拟数据和实际数据的应用发现,本研究提出的估计方法能够提高更加稳健的估计结果。无论是模型系数的估计,还是目标变量的小域估计量,本研究得到的结果均具有更小的偏差和均方误差。为了动态展示本研究提出的估计方法在混合正态分布中的表现,本研究比较了几类估计结果随着污染分布的方差以及污染分布的比例变化时估计的 MSE 的动态变化图,通过比较图形发现,无论是模型系数还是区域均值,其 MSE 的变化均不太显著,而现有的稳健估计方法却表现一般,受污染比例和污染方法方差的波动较大。

最后,在本研究中提出了用密度幂散度进行稳健估计的参数选择算法。对两类小域估计模型进行稳健估计时,引入的估计方法中存在一个调整参数,该参数能够根据观测数据的特征调节模型估计的有效性和稳健性。一般地,当模型中异常值较小时,可以选取较小的调节参数,而存在较多的异常值时,通过较大的调节参数来达到稳健估计的目的。本研究在调整参数的选取中,引入了一类迭代估

计算法, 该算法能够根据数据的特点自动选取使其估计 MSE 最小的调整参数, 文中给出了调整参数的选择算法。

本研究结合密度幂散度族提出了针对基本小域模型的稳健估计方法。在本研究中, 给出了模型参数、目标变量的估计表达式以及区间估计等。通过模拟和实际数据的验证, 发现本研究中提出估计方法优于现有的稳健小域估计方法, 并对非正态数据和异常观测值均具有较为理想的估计结果, 能够解决不满足基本假设的小域估计问题。在实际应用中, 本研究中提出的方法也具有较高的可操作性与估计效果, 并通过中国家庭调查数据进行了佐证。本研究中提出的方法能够适用于更加广泛的小域估计模型, 能够为决策者提供更加可靠的小域估计量。

**关键词:** 小域估计 FH 模型 NER 模型 稳健估计 密度幂散度  $\gamma$  散度



## Abstract

The effective estimation of population target variables using sample data is a common research problem in statistical inference, and the estimation methods have a wide range of practical applications. The direct estimation method based on sampling design is the most direct approach to solving such problems. However, when the estimated population area is composed of numerous small domains, estimating the target variables in small domains may face situations of small or even no samples. In such cases, the direct estimation method using samples may result in large errors or ineffective estimates. The small area estimation is one of the effective methods to solve such problems. Compared with traditional sampling estimation methods, the small area estimation can obtain effective estimates of target variables in small areas by utilizing auxiliary variable information, which can solve estimation problems in situations of small or no samples. In recent years, the small area estimation method has been widely used in fields such as population statistics, biostatistics, agricultural statistics, and government statistics, and related academic research is also relatively abundant, which has led to the systematic development of small area estimation theory.

As the main method of small area estimation, model-based small area estimation is the core content of small area estimation. Because the model-based estimation method can well apply auxiliary variables to the estimation model, so as to achieve the role of "leverage", so as to solve the problem of small sample and no sample. In small area estimation models, regional random effects and model random errors are usually assumed, and both of them are assumed to follow normal distribution. A large number of practical studies have shown that the assumption of normal distribution is not valid when there are abnormal observations in the model. This directly leads to the failure of the assumption of the basic small area model, which will make the parameter estimation based on the normal assumption and the estimation of the target variable have a large deviation, so it is necessary to further explore the robust estimation method which is insensitive to the abnormal observations. There are two kinds of widely used methods

in small area robust estimation. One kind of methods is the estimation method assuming that the model error is biased distribution, for example, assuming that the model error follows the  $t$ -distribution or Cauchy distribution. The biased distribution is constructed to reduce the influence of abnormal observations on the estimator. Another kind of methods is to use the Huber- $\phi$  function to robust the empirical linear unbiased estimator, through the properties of Huber- $\phi$  function to achieve the purpose of robustness. However, when the outliers are too large, the estimation effects of the above two estimation methods are limited, and in some cases, the estimation results will still produce large deviations.

Studying robust small area estimation methods is a current practical issue in small area estimation. The widespread existence of non-normal observable data and the occurrence of outliers have presented new challenges to small area estimation methods. In order to solve the problem of unstable estimates and large prediction biases in such situations, the robust small area estimation method has garnered attention from many scholars. In this thesis, we consider the important characteristics of density power divergence families in robust estimation and apply them to small area estimation, proposing a robust estimation method based on density power divergence families to address the shortcomings of existing robust small area estimation methods. By applying density power divergence families to small area estimation, we investigate the estimation problems of small area model coefficients and target variables under non-normal and outlier observation scenarios. The aim of this thesis is to construct robust estimates of small area model coefficients and target variables, and to provide confidence intervals for the parameters and the mean squared error of the estimates.

Firstly, the robust estimation of area level models based on density power divergence and  $\gamma$  divergence is studied in this thesis. By applying the density-power divergence to the FH model, the robust estimates and asymptotic distributions of the model coefficients are obtained. On this basis, the robust estimator of the target variable is discussed, and the mean square error of the estimator is given. In order to obtain the reliable estimation of the small area estimator, the confidence intervals of the target variables are also given. Through the establishment of small area model of simulated data and actual data, combined with the robust

estimation method proposed in this thesis, the robust estimation is carried out, and compared with the existing robust estimation methods. Through comparison, it is found that the proposed estimation method can control the balance between validity and robustness by tuning parameters. When there are no outliers in the observed data, the robust estimation results obtained by the proposed estimation method are similar to those obtained by the existing optimal linear unbiased estimation methods by using small tuning parameters. When there are outliers in the observed data, the proposed robust estimation method has a smaller mean square error than the existing estimation methods, which indicates that the proposed estimation method is effective.

Secondly, the robust estimation of unit level models based on density power divergence and  $\gamma$  divergence is studied. These two kinds of divergences are applied to the NER model, and the coefficients of the model are estimated robustly, and their asymptotic distributions are obtained. Under the unit level model, The robust estimator of the functional form of the target variable and the robust estimator of the area mean of the finite population are studied. Since the estimation of the target variable in the unit level model involves the calculation of multiple integrals, the MCMC method is used to give the estimate of the functional form of the target variable, and combine the bootstrap method to give the MSE of the estimator. Similarly, the proposed estimation method is compared with the existing robust estimation methods. Through the application of simulated data and real data, it is found that the estimation method proposed in this thesis can improve the more robust estimation results. Both the estimation of model coefficients and the small area estimator of the target variable, the results obtained in this thesis have smaller bias and mean square error. In order to dynamically display the performance of the estimation method of the mixed normal distribution proposed in this thesis, this thesis compared the dynamic changes of MSE estimated by several types of estimation results with the variance of pollution distribution and the proportion of pollution distribution. Through comparing the graphs, it was found that no matter the model coefficient or the regional mean value, the change of MSE was not significant. However, the existing robust estimation methods perform poorly, and the proportion of contamination and the variance of pollution

methods fluctuate greatly.

In this thesis, a parameter selection algorithm using density power divergence for robust estimation is proposed. When robustly estimating two types of small area estimation models, a tuning parameter is introduced into the estimation method, which can adjust the effectiveness and robustness of the model estimation according to the characteristics of the observed data. Generally, when there are fewer outliers in the model, a smaller tuning parameter can be selected, while when there are more outliers, a larger tuning parameter can be used to achieve the purpose of robust estimation. In the selection of tuning parameters, this thesis introduces an iterative estimation algorithm, which can automatically select the tuning parameter that minimizes the estimated mean square error (MSE) according to the characteristics of the data. The algorithm for selecting the parameter is presented in this thesis.

This thesis proposes a robust estimation method for the basic small area model based on the density power divergence family. The estimation expressions and interval estimations of the model parameters and target variables are provided in this thesis. Through simulations and validation with real data, it is found that the proposed estimation method in this thesis performs better than existing robust small area estimation methods, and has relatively ideal estimation results for non-normal data and outlier observations, which can solve small area estimation problems that do not meet the basic assumptions. In practical applications, the method proposed in this thesis has high operability and estimation effectiveness, and has been demonstrated by the China Family Panel Studies data. This method proposed in this thesis can be applied to a wider range of small area estimation models, and can provide more reliable small area estimates for decision-makers.

**Keywords:** Small area estimation, FH model, NER model, Robust estimation, Density power divergence,  $\gamma$  divergence

## 目 录

第 1 章 引言 .....	1
1.1 研究背景和研究意义 .....	1
1.1.1 研究背景 .....	1
1.1.2 研究意义 .....	6
1.2 研究现状 .....	7
1.2.1 小域估计研究综述 .....	7
1.2.2 稳健小域估计研究综述 .....	12
1.3 研究思路和研究内容 .....	14
1.3.1 研究思路 .....	14
1.3.2 研究内容 .....	14
1.3.3 研究目的 .....	18
1.4 创新点 .....	19
1.5 文章结构 .....	19
第 2 章 小域估计的基本方法及理论 .....	23
2.1 小域估计的模型 .....	23
2.1.1 区域水平模型 .....	23
2.1.2 单元水平模型 .....	24
2.1.3 线性混合模型 .....	26
2.1.4 广义线性混合模型 .....	26
2.2 经验最佳线性无偏估计 .....	27
2.2.1 小域模型的参数估计 .....	27
2.2.2 区域水平模型的 EBLUP .....	29
2.2.3 单元水平模型的 EBLUP .....	30
2.2.4 线性混合模型的 BLUP .....	31
2.3 估计量的均方误差 .....	32
2.3.1 基本模型的均方误差 .....	33
2.3.2 混合模型的均方误差 .....	36
2.4 稳健小域估计方法 .....	37
2.4.1 小域中的稳健估计 .....	39
2.4.2 密度幂散度族及其比较 .....	45

第 3 章 密度幂散度在区域水平模型下的稳健估计 .....	55
3.1 基于密度幂散度的 FH 模型的稳健估计方法 .....	56
3.2 FH 模型的参数与区域均值的稳健估计 .....	56
3.3 调整参数的选取 .....	58
3.4 稳健估计量的性质 .....	59
3.5 数值模拟 .....	61
第 4 章 基于 $\gamma$ 散度的区域水平模型的稳健估计 .....	69
4.1 基于 $\gamma$ 散度的 FH 模型的稳健估计方法 .....	69
4.2 稳健估计量 .....	70
4.3 FH 模型的参数估计 .....	71
4.4 调整参数的选取 .....	71
4.5 稳健估计量的性质 .....	73
4.6 稳健估计量的 MSE .....	73
4.7 稳健估计量的置信区间 .....	75
4.8 应用举例 .....	76
4.8.1 数值模拟 .....	76
4.8.2 实际应用 .....	83
第 5 章 基于密度幂散度的单元水平模型的稳健估计 .....	87
5.1 基于密度幂散度的 NER 模型的稳健估计方法 .....	87
5.2 DPD 方法在 NER 模型中的应用和模型的参数估计 .....	89
5.3 模型参数的渐近分布 .....	91
5.4 有限总体的稳健 EB 估计量 .....	93
5.5 区域均值的 EB 估计量 .....	95
5.6 调整参数的选取 .....	96
5.7 稳健估计量的 MSE .....	97
5.8 应用举例 .....	98
5.9 小结 .....	118
第 6 章 基于 $\gamma$ 散度的单元水平模型的稳健估计 .....	119
6.1 基于 $\gamma$ 散度的 NER 模型的稳健估计方法 .....	119
6.2 $\gamma$ 散度在 NER 模型中的应用和模型的参数估计 .....	119
6.3 模型参数的渐近分布 .....	121
6.4 调整参数的选取 .....	123
6.5 稳健 EB 估计量 .....	124
6.6 区域均值的 EB 估计量 .....	125

6.7 稳健估计量的 MSE .....	125
6.8 应用举例 .....	126
6.8.1 数值模拟 .....	126
6.8.2 实际数据 .....	142
6.9 小结 .....	144
6.10 应用研究 .....	144
6.10.1 数据来源 .....	144
6.10.2 小域模型及其估计 .....	145
<b>第 7 章 结论与展望 .....</b>	<b>151</b>
7.1 研究总结 .....	151
7.2 应用场景 .....	152
7.3 研究不足与展望 .....	153
7.3.1 不足 .....	153
7.3.2 展望 .....	154
<b>参考文献 .....</b>	<b>157</b>
<b>附 录 .....</b>	<b>169</b>
A.1 第四章中的定理证明 .....	169
B.2 5.1 中的证明和推导 .....	172
B.2.1 渐近分布协方差均值的计算 .....	175
B.2.2 模型中矩阵 $J^{(i)}$ 的计算 .....	176
B.2.3 矩阵 $J^{(i)}$ 的计算 .....	180
B.2.4 相关积分的期望 .....	183
C.3 6.1 中的证明和推导 .....	185
C.3.1 矩阵 $J^{(i)}$ 的计算 .....	187
<b>攻读学位期间发表的学术论文与研究成果 .....</b>	<b>193</b>
<b>致 谢 .....</b>	<b>195</b>



## 图形列表

1.1	目标变量与辅助变量的层次分布 .....	3
1.2	技术线路图 .....	15
1.3	主要研究内容与结构安排 .....	17
2.1	$g, f, h$ 之间的勾股定理关系 .....	53
3.1	情形 1 中各方法估计 MSE 的箱线图 .....	65
3.2	情形 2 中各方法估计 MSE 的箱线图 .....	65
3.3	情形 3 中各方法估计 MSE 的箱线图 .....	65
4.1	情形 2 中几类估计 MSE 的箱线图 .....	78
4.2	情形 3 中几类估计 MSE 的箱线图 .....	79
5.1	模型系数的 MSE 随 $e_{ij}$ 的污染比例变化图 .....	103
5.2	随机误差方差 MSE 随 $e_{ij}$ 的污染比例变化图 .....	103
5.3	模型系数的 MSE 随 $v_i$ 的污染比例变化图 .....	104
5.4	随机误差方差 MSE 随 $v_i$ 的污染比例变化图 .....	105
5.5	模型系数的 MSE 随 $v_i$ 和 $e_{ij}$ 的污染比例变化图 .....	105
5.6	随机误差方差 MSE 随 $v_i$ 和 $e_{ij}$ 的污染比例变化图 .....	106
5.7	模型系数的 MSE 随 $e_{ij}$ 的污染比例变化图 .....	107
5.8	随机误差方差 MSE 随 $e_{ij}$ 的污染比例变化图 .....	107
5.9	模型系数的 MSE 随 $v_i$ 的污染方差变化图 .....	108
5.10	随机误差方差 MSE 随 $v_i$ 的污染方差变化图 .....	108
5.11	模型系数的 MSE 随 $v_i$ 和 $e_{ij}$ 的污染方差变化图 .....	109
5.12	随机误差方差 MSE 随 $v_i$ 和 $e_{ij}$ 的污染方差变化图 .....	110
5.13	稳健估计的 MSE 随污染方差变化图 .....	111
5.14	稳健估计量的 MSE 随 $v_i$ 和 $e_{ij}$ 的污染方差变化图 .....	112
5.15	稳健估计量的 MSE 随污染比例变化图 .....	113
5.16	稳健估计量的 MSE 随 $v_i$ 和 $e_{ij}$ 的污染比例变化图 .....	114
6.1	模型系数的 MSE 随 $v_i$ 的污染比例变化图 .....	130
6.2	随机误差方差 MSE 随 $v_i$ 的污染比例变化图 .....	131
6.3	模型系数的 MSE 随 $e_{ij}$ 污染比例变化图 .....	132
6.4	估计方差的 MSE 随 $e_{ij}$ 污染比例变化图 .....	132
6.5	模型系数的 MSE 随污染比例变化图 .....	133

6.6	估计方差 MSE 随污染比例变化图	134
6.7	模型系数的 MSE 随污染分布方差变化图	135
6.8	估计方差 MSE 随污染分布方差变化图	135
6.9	模型系数的 MSE 随污染分布方差变化图	136
6.10	估计方差 MSE 随污染分布方差变化图	136
6.11	模型系数的 MSE 随污染分布方差变化图	137
6.12	估计方差 MSE 随污染分布方差变化图	138
6.13	稳健 EB 估计的 MSE 随污染分布方差变化图	139
6.14	稳健 EB 估计的 MSE 随污染分布比例变化图	140
6.15	稳健 EB 估计的 MSE 随污染分布变化图	141
6.16	家庭纯收入与消费性支出之间的散点图	146
6.17	家庭纯收入的 QQ 分布图	146
6.18	估计残差 QQ 图	147
7.1	老师修改的部分论文截图	195

## 表格列表

3.1 FH 模型参数估计的偏差和 MSE .....	63
3.2 目标变量 $\theta$ 估计的 MSE .....	64
3.3 DEB 方法的 MSE 与估计 MSE 的比较 .....	66
4.1 几类稳健方法所得的每组平均 MSE 和蒙特卡洛误差 .....	77
4.2 均方误差估计量的相对偏差和相对 MSE .....	80
4.3 置信度为 0.95 的置信区间的覆盖率 (平均长度) .....	81
4.4 有偏分布下稳健估计的 MSE 和蒙特卡洛误差 .....	82
4.5 模型参数的估计及其标准误 .....	84
4.6 目标变量的估计值及其 MSE .....	85
5.1 模型参数估计的偏差及其 MSE(a) .....	99
5.2 模型参数估计的偏差及其 MSE(b) .....	100
5.3 小域均值的估计偏差和 MSE .....	101
5.4 有限总体区域均值的估计偏差和 MSE .....	115
5.5 模型参数的估计及其标准误 .....	116
5.6 各县区玉米种植面积的预测值及其 Bootstrap MSPE .....	117
6.1 四种污染情形下模型参数估计的偏差和 MSE .....	127
6.2 四种污染情形下均值估计的偏差及其 MSE .....	129
6.3 不同污染情形下有限总体均值估计的偏差及其 MSE .....	141
6.4 模型参数的估计及其标准误 .....	143
6.5 区域玉米种植面积的预测及其 Bootstrap MSE .....	143
6.6 各省份的样本量 .....	145
6.7 模型系数的估计 .....	147
6.8 各省份家庭平均收入的小域估计 .....	148
6.9 各省份家庭平均收入估计的 MSE .....	149



## 符号列表

## 字符

符号	含义	单位
$N$	总体容量	1
$n$	样本容量	1
$m$	小域的个数	1
$p$	辅助变量的维数	1
$U$	研究总体	1
$s$	样本	1
$Y$	目标研究变量	1
$y_s$	抽样得到的目标变量构成的向量	1
$y_r$	未抽样的目标变量构成的向量	1
$\bar{Y}$	目标变量总体均值	1
$X$	$N \times p$ 的辅助信息矩阵	1
$e_i$	模型随机效应、抽样误差	1
$v_i$	区域随机效应、区域误差	1
$\alpha$	密度幂散度的调整参数	1
$\gamma$	$\gamma$ 散度的调整参数	1

## 算子

符号	含义
$\ln$	自然对数
$\min$	最小化
$\max$	最大化
$\frac{\partial f(x)}{\partial x}$	$f(x)$ 关于 $x$ 的偏导数

## 缩写

缩写	英文	中文
BLUP	Best Linear Unbiased Predictor	最佳线性无偏预测量
DPD	Density Power Divergence	密度幂散度
EB	Empirical Bayes	经验贝叶斯
EBLUP	Empirical Best Linear Unbiased Predictor	经验最佳线性无偏预测量
EBP	Empirical Best Predictor	经验最佳估计量
FH	Fay-Herriot	FH 模型
HT	Horvitz-Thompson	H-T 估计量
KL	Kullback-Leible Divergence	KL 散度
MDPD	Minimum Density Power Divergence	最小化密度幂
MSE	Measure Square Error	均方误差
NER	Nest Error Regression	误差嵌套回归
REB	Robust Empirical Bayes	稳健经验贝叶斯
SAE	Small Area Estimation	小域估计

## 第 1 章 引言

本章主要介绍研究背景和研究意义,小域的定义,相关的研究综述,有关稳健小域估计的研究现状,指出现存研究方法的不足,提出研究视角,给出本研究的主要研究内容和研究思路。

### 1.1 研究背景和研究意义

#### 1.1.1 研究背景

在统计应用研究中,通常要了解研究对象的总体特征,以此为基础对总体进行推断以及决策。这类问题构成了统计推断的基础问题之一,也受到了众多统计学家的关注。而为了掌握研究对象的数字特征和分布特征,越来越多的决策者致力于得到总体的可靠估计。这一问题也成为统计研究的热点问题,众多的数学家、统计学、管理者家做了大量的相关研究,形成了统计推断中不可或缺的理论。

实践中,决策者需要根据掌握的信息对某些区域或者子群的统计特征进行估计。例如估计某些地区的人均收入、某些地理区域上农作物种植面积等。在抽样调查的背景下,需要考虑如何估计区域内的目标参数,针对区域内目标参数的估计问题可以称为广义上的区域估计问题。区域估计问题普遍存在于我们的生活中,并在解决实际问题方面需求非常广泛。早在 11 世纪的英国,人口学家开始了小域估计的最早探索,在此之后,人口统计学家一直致力于人口统计中区域估计问题。与此同时,关于劳动力市场的失业率的区域估计、疾病发病率的区域估计等受到学者的关注。在这方面,抽样调查长期以来一直被当作最有效的数据收集手段,在区域估计方法的发展中扮演着不可或缺的角色。直到上世纪 30 年代,现代抽样理论才有了较为完善的体系。

上世纪 30 年代之后,基于抽样理论的直接估计成为了区域估计中主要的研究方法。这种方法是比较直观而且应用较为广泛的一类方法,主要是通过合理的调查设计,抽取区域上的样本数据,然后根据样本数据对目标变量进行直接估计的方法。如果目标变量的估计是通过抽样技术得到的样本结合抽样概率的方式加权得到,在估计中只涉及区域相关的样本,而不借助与其它相关的变量,则称这种估计为直接估计。统计学家将这类方法应用于人口统计调查、社区健康调查、住户调查、劳动力市场调查等领域。这类方法的优点是直观,当样本量较大时估计具有无偏性、正态性等特征。其不足也比较突出,一方面是区域较多时,对每

个区域进行大量的抽样调查需要较大的成本投入；另一方面是当我们获取的样本量有限时，直接估计的精度会受到影响，导致误差增加，对于没有样本的区域，直接估计方法就无法进行有效的估计。

区域估计发展的初期，大部分研究重要集中于主要区域（“大区域”）的估计，研究者只关注国家层面、主要地理区域和大型领域的估计。随着统计工作的实际需求，政策制定者以及决策者所面对的问题越来越复杂，要求越来越多元化，研究者也越来越关注更复杂的调查方法和更广泛的调查范围，同时计算能力和软件的发展也促进了抽样数据多元化和复杂化，这也导致了抽样估计方法的发展。而目前越来越多的决策者需要对用于决策的小域进行估计，不同于过去对主要区域上的估计，需要对市、县、乡或者其它形式的更小域的估计。这一背景便成为小域估计 (Small Area Estimation, SAE) 的最基本需求，也为小域估计方法的提出和完善提供了最根本的现实依据。

随着研究范围和研究目的的不断细化，抽样调查所设计的研究问题、调查方法以及调查范围都在进一步扩大。在早期的应用研究中，可能只关注国家层面或者主要地理区域的重点估计。由于数据计算能力、计算机软件、大数据的发展，为了给研究对象精准画像，人们越来越多的关注某些小区域、小范围的估计问题。在政府统计和商业决策等领域，对小域估计的需求也在与日俱增。区域估计中，如果一个区域的样本量不能支撑其得到精确的直接估计，则这类区域称为小域 (Small Area, Rao et al.<sup>[1]</sup>)。这里的区域可以是地理区域，也可以是其它某些属性划分的类别。例如根据学历、性别划分的人群可以认为是根据不同的区域进行划分的样本。由小域的定义不难看出，基于抽样理论的直接估计不能得有效的小域估计量。为了能够得到小域上目标变量的准确估计，研究者开始寻求借助于其它辅助信息的间接估计方法。

受限于小域的样本数量，间接估计方法成为主要的小域估计方法。小域估计为了保证足够的精度水平，常常需要使用“间接”估计，即通过从相关领域（时间）“借力”从而增加有效样本大小，对目标变量的值进行估计。在辅助变量的选择中，通常使用与目标变量有关的补充资料建立相关区域（时间）的模型，通过模型将这些目标值纳入估计过程，例如最近的人口普查数据和当前的行政记录。通常辅助信息的来源渠道有：统计局、具有较大或全面调查数据的单位、以前类似的调查、统计年鉴、各种类型的普查数据等。针对实际问题的调查设计可能很难涉及所要估计的全部领域，而通常政府统计数据、或者普查数据包含估计的所有领域，基于这一特点，间接估计就是利用收集到的数据，同时结合辅助变量数

据拟合一个合适的模型，然后生成每个区域上目标变量的估计。常见的间接估计方式有：域间接估计、时间间接估计、域与时间间接估计。

间接估计方法中，有两大类方法受到了研究人员的重视。第一种是基于隐式连接模型的传统间接估计。包括综合估计、复合估计和 James-Stein(JS) 收缩估计方法。这些估计方法通常是基于设计的，它们的设计方差通常相对于直接估计值的设计方差较小。然而，间接估计通常会有设计偏差。并且设计偏差不会随着总体样本量的增加而减小。如果隐式连接模型近似为真，那么设计偏差可能很小，导致与直接估计器的均方误差 (Mean Square Error, MSE) 相比，间接估计量的 MSE 显著较小。因此，降低 MSE 是使用间接估计的主要原因。另一种是基于线性混合模型的小域估计方法。基于模型的小域估计方法能充分结合辅助变量的信息，为我们提供可靠的估计。因其在假定的模型下，可以推导出“最优”估计量，且不同于传统的间接估计值所使用的全域估计值 (在小面积上平均)，区域特定的可变性测度可以与每个估计量相关联。另外，模型可以从样本数据中验证，根据响应变量的性质和数据结构 (如空间相关性和时间序列结构) 的复杂性，可以采用多种模型。由于这些优点，基于模型的方法成为了小域估计中的常用方法。

为了更好的表示和理解区域估计问题，将区域估计问题统计化，并做如下相关的记号。一般地，记一个调查总体为  $U$ ，其中包含  $N$  个区域，记作  $k = 1, 2, \dots, N$ ，将目标变量记作  $y$ ，则第  $i$  个区域上的目标变量为  $y_i$ ，第  $i$  个区域上包含  $N_i$  个单元。相应地，在区域  $i$  中，与目标变量  $y_i$  相关的  $p$  维辅助变量为  $\mathbf{x}_i$ ，与单元级目标变量  $y_{ij}$  相关的辅助变量为  $\mathbf{x}_{ij}$ 。目标变量与辅助变量的层次分布具体可参见图形 1.1。

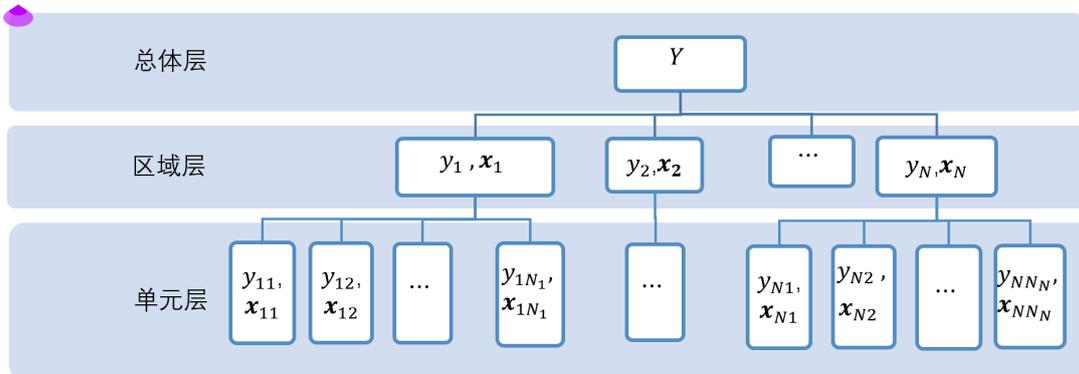


图 1.1 目标变量与辅助变量的层次分布

目标变量  $y$  的总和应该表示为  $Y = \sum_U y_i$ ，其平均应表示为  $\bar{Y} = \sum_U y_i / N$ ，

然而在实际应用中, 这些  $y_i$  并不全部能够获得。一般是从总体  $U$  中以抽样概率  $p(s)$  获得样本  $s$ , 其中抽样概率取决于抽样设计。抽样得到的样本数据为  $y_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, n_i$ , 即抽取的样本数据来自于  $n(n \leq N)$  个区域, 第  $i$  个区域上抽取了  $n_i(n_i \leq N_i)$  个单元。

在实际应用中, 通常关注某些区域的某些统计指标, 例如均值、分位数等。针对这一问题的解决方法就是基于抽样调查的直接估计。然而, 当划分的区域较多时, 难免出现样本量较小或者没有样本量的情形。在这种情况下, 利用传统的抽样调查方法进行的直接估计就会出现较大的波动或者没办法直接估计的问题。在前述的抽样框架中, 该问题表现在两个方面: 其一, 区域层中共有  $N$  个区域, 然而抽样区域只有  $n$  个, 即当  $n \leq N$  时, 有一部分区域是没有目标变量  $y$  的样本, 不能用直接估计的方法完成估计; 其二, 在单元层上, 第  $i(i = 1, 2, \dots, n)$  个区域的抽样单元数量为  $n_i$ , 当  $n_i$  较小时, 用小样本的直接估计量具有较大的偏差和方差, 其表现并不理想。因此, 借助于辅助信息的间接估计方法应用而生。若辅助变量  $x$  与目标变量  $y$  之间具有显著的线性关系或者函数关系, 虽然在某些区域上目标变量的样本较少甚至没有, 但是只要能够获取该区域上的辅助信息, 则根据两者之间的关系便可以得到目标变量的估计。小域估计中将这种依赖于辅助信息的估计思想称之为“借力”, 即通过借助于辅助信息来达到小域估计的目的。随着我们对估计区域的精细化要求以及政府统计中对区域估计需求的不断增加, 小域估计问题的研究越来越受到学者们的重视。而且小域估计问题在政府统计、农业调查、经济调查、医学调查中都有应用。

在辅助信息的加持下, 直接估计的效果得到了显著的改善, 间接估计方法受到了学者的重点关注。虽然间接估计方法的估计效果得到了显著改善, 随之而来的挑战不可忽视。首先, 无论目标变量是否被抽中, 都要获取相应的辅助变量。即区域层的辅助变量是  $N$  个, 第  $i$  个区域上的辅助变量应该有  $N_i$  个。故辅助信息的获取成为小域估计的第一大挑战。其次, 辅助变量与目标变量之间的关系难以准确界定, 且区域间的差异难以体现。在实际应用中经常遇到的情况便是某些区域上  $y_i$  与  $x_i$  之间具有显著的关系, 而另外的区域上这种关系不显著。这两个挑战使得简介估计的应用条件变得非常严格, 应用门槛随之升高。于是, 为了更好地刻画区域间地差异, 学者们提出了基于模型的小域估计方法, 该方法主要借助于区域内的目标变量和其它辅助变量的模型来达到估计目标变量的目的。基于模型的小域估计方法在模型中将与目标变量相关的辅助变量和区域间的随机效应均纳入模型, 以此来反映区域间的差异。而且目标变量与辅助变量之间的

关系通过统计模型来反映,变得更加精确,对辅助信息的要求也有所减弱。其中这些辅助变量可能是来自于普查或者其它调查的变量,基于这一优势,近十几年来小域估计在这方面的应用变得十分广泛。

小域模型中,根据目标变量和辅助变量的特征,小域模型可以分为区域水平模型和单元水平模型。如果目标变量和辅助变量均是描述区域特征的变量,则可以建立区域水平模型。若目标变量和辅助变量在研究区域内的单元上可用,则可以建立单元水平模型。例如需要调查某市的家庭收入水平,将该市的每个县作为抽样区域,则以每个县的家庭平均收入为目标变量,每个县的 GDP、固定资产投资等变量为辅助变量的模型为区域水平的小域估计模型。若考虑每个县内的每户家庭,以家庭收入为目标变量,家庭的人口、教育程度等变量为辅助变量的模型为单元水平的小域估计模型。从这两种模型比较中不难发现,单元水平模型的辅助信息更加难以获取。因此,单元水平模型估计的更加精确,但是辅助信息的获取成为难题。

在基于模型的小域估计方法中,适用最为普遍的是由 Fay-Herriot 提出的标准区域水平模型 Fay III et al.<sup>[2]</sup> 和 Battese 等提出的单元水平模型 Battese et al.<sup>[3]</sup>。这两类模型可以认为是特殊的混合效应模型,在该类模型中,通过区域随机效应来刻画区域间的差异,利用随机误差项来捕捉不能用辅助变量解释的部分。这样以来,模型既能够联系目标变量和辅助变量又能体现出区域间的差异性特征,满足小域估计的最基本需求,也能到达估计的目的。小域估计要解决的核心是目标变量的估计,在小域模型中用  $y_i(i = 1, 2, \dots, n)$  表示目标变量的样本数据,用  $\theta_i$  表示其真实数据,则  $\theta_i$  的估计值  $\hat{\theta}_i$  是需要通过模型得到的。在小域模型中,根据平方损失函数最小,利用不同的方法得到最优线性无偏估计量(Best Linear Unbiased Predictor,BLUP)、经验最优线性无偏估计量(Empirical Best Linear Unbiased Predictor,EBLUP)等。估计标准的判别方面,通常使用均方误差来衡量估计的参数的性质。在标准的 FH 模型中,EBLUP 是 MSE 最小的线性无偏估计,受到学者的广泛引用。

混合效应模型很好地捕获了小域估计的基本特性,自然也成为了小域估计的最常用模型之一。然而,在混合效应模型中,需要假设区域随机效应和随机误差都是服从正态分布的。在模型的假设条件下可以得到观测数据  $y_i$  也应该服从正态分布。但是很难保证抽样得到的数据是服从正态分布的,尤其是在商业调查数据中这一点尤为突出。当观测数据中存在异常观测时,模型的正态性假定就很难得到满足,这些问题将直接影响估计量的准确性。Ronchetti et al.<sup>[4]</sup> 指出这样

的问题可能发生在统计的每个学科领域,因此,研究这种背景下稳健小域估计是必然选择。

结合上述背景发现,小域估计问题在今天的应用中依旧存在<sup>[5-26]</sup>。政府、企业的决策者需要根据掌握的有限信息进行小域上的估计并做出相应的决策,而提供精确的估计是关键。目前的小域估计方法中,EBLUP 应用较多且比较完善,但是该方法针对具有离群观测的数据时估计不准确,且估计量的 MSE 较大,必须采用稳健估计方法。关于稳健估计方法的讨论,Ghosh et al.<sup>[27]</sup>,Sinha et al.<sup>[28]</sup> 基于 Huber' s 函数构建稳健估计量,目前大量的稳健讨论均基于这两种方法,且这两种方法的稳健估计容易出现过度收缩的情形、模型中的未知参数的估计易受离群观测的影响,有必要改善目前的稳健估计方法。本研究重点讨论观测数据存在异常值时小域估计问题,通过引入其他的统计估计方法来改进现有的稳健估计。提出对异常值不敏感,不会出现过度收缩的稳健估计,同时还聚焦于稳健估计的 MSE 的变化。另一方面,有时候参数的区间估计也是有必要的,因此在此背景下目标参数的区间估计也是我们关注的一个方面。

### 1.1.2 研究意义

研究小域稳健估计,对具有异常观测的数据进行可靠的估计。正态分布假设下的 EBLUP 估计量对于异常观测值较为敏感,而且估计量的 MSE 较大,因此有必要考虑稳健估计量。目前关于小域稳健估计量在较大离群观测和较大抽样误差时的稳健估计量存在两个方面的弊端,其一,当抽样误差较大时,通常会会出现过度收缩的现象;其二,目前的稳健估计量的偏差较大,其估计量的 MSE 较大。由于这两个不足,研究更加稳健的小域估计量是本文的主要工作,通过本研究,将大大改善上述两个弊端的发生。研究不同模型下小域稳健估计量,为存在异常观测数据的小域估计提供更多的选择方法,改进现有估计方法。通过本文的研究,弥补小域估计中稳健性不足的缺陷,使得小域估计量更加准确、稳定,从而为决策者提供更加可靠、稳定的小域估计量。

研究不同模型下的小域估计问题,解决目标变量在不同水平下的估计问题。在实际应用中,经常会遇到目标变量为单元水平变量、离散型变量的问题,同时还会遇到目标变量中存在大量异常观测的情形,针对这一问题,本研究通过区域水平模型、单元水平模型下的稳健小域估计方法,用来解决目标变量有异常观测情形的小域估计问题。进一步丰富了小域估计的方法,为这类小域估计问题提供了更加可靠的理论基础。

构造基于稳健估计量的置信区间,为决策者提供可靠的区间估计。由于异常观测的存在,对于置信区间的估计就显得很难,因此针对这类情形提供可靠的区间估计能为决策者提供依据。本文构建基于稳健估计量的 Bootstrap 置信区间,并推导置信区间的渐近误差。

应用小域估计方法分析住户调查数据,为政策制定和城市发展规划提供依据。本研究将应用小域估计方法,分析中国家庭追踪调查数据。由于中国家庭调查数据的样本量相比较于全国的居民数量就显得很少,如果采用调查数据进行直接估计,那么得到的数据肯定具有很大的偏差,而且样本分布比较随机,有些区域的样本量很小,甚至有些区域没有样本,要进行区域估计就变得很难。小域估计方法能很好地解决这一问题,借助于其它辅助变量信息,来对关注的变量进行区域估计,这不仅能改善原来估计方法的不足,还能得到区域上的估计值,便于相互区域之间的比较,进一步为制定区域差异的政策提供参考依据。

## 1.2 研究现状

### 1.2.1 小域估计研究综述

关于区域估计的讨论,最早可以追溯到 11 世纪的英国和 17 世纪的加拿大,就已经有人口统计学家在人口普查和行政记录中用到区域数据。人口统计学家一直在使用各种方法对人口进行小区域估计,并对人口普查后的年份进行其它感兴趣的特征估计。这些方法利用了来自行政登记的当前数据以及来自最近人口普查的相关数据,例如 Ghosh et al.<sup>[29, 30]</sup>, Rao et al.<sup>[31]</sup> 等工作就涉及这些方面。随后一段时间,关于区域估计的研究一直集中在抽样设计的基础上,这也促进了抽样调查学科的发展,直至上世纪 30 年代,形成了较为完整的抽样调查体系 (Cochran<sup>[32]</sup>)。

在抽样调查的基础上,目标变量的直接估计是最有效的估计方法。在这方面的研究中, Cochran<sup>[32]</sup> 讨论了在简单随机抽样下的 Horvitz-Thompson (H-T) 估计量, Rao<sup>[33]</sup> 详细讨论了基于设计的无偏估计量及其估计量的方差。直接估计量的研究是在不同的抽样设计下形如  $\hat{Y} = \sum_s w_j y_j$  的估计量,其中  $w_j$  为与抽样设计有关的权重。Cochran<sup>[32]</sup>, Deville et al.<sup>[34]</sup> 等研究了简单随机抽样、分层多阶段抽样等抽样设计下的直接估计以及估计量的方差。在过去 60 年左右的时间里,在较大区域使用样本对总量和平均值的“最优”直接估计受到了广泛的关注。尤其是在调查设计领域中,涉及如何分层、层数的选取、样本的分配以及抽样概率等,这些理论在 Cochran<sup>[32]</sup> 的专著中得到系统描述。由于操作上的限制

等因素, 这样的“最优”估计在实践中很难实现。在实际应用中, 调查得到的样本可能没有覆盖所有的区域, 有时样本数据能够支撑较大区域上的直接估计, 但不能得到每个小域上的直接估计。在基于设计的直接估计中, 关于总和的直接估计为  $\hat{Y} = \sum_s w_j y_j$ , 其中  $w_j(s)$  表示第  $j$  个区域在抽样设计  $s$  中的权重, 如果  $w_i = 1/\pi_i$ , 其中  $\pi_i$  表示第  $i$  个区域的样本在抽样设计  $s$  中的入样概率, 此时的直接估计量就成了著名的 Horvitz-Thompson (H-T) 估计量。从直接估计量的表达形式可以看出, 基于设计的直接估计仅仅包含了样本  $y_i$  的信息, 对于样本量较少, 乃至无样本的区域得到的估计不可靠, 或者得不到估计。由于这些不足, 借助于辅助信息的小域间接估计得到广泛应用。

由于直接估计受限于样本的数量, 借助于其它辅助信息的间接估计量受到了大量的关注。间接估计的核心思想就是借助于辅助信息来“增加”样本量的方式, 降低直接估计的误差。常见的间接估计量有综合估计量、复合估计量、以及 James-Stein(JS) 收缩估计量。Hansen et al.<sup>[35]</sup> 描述了综合回归估计器在无线电监听调查中的首次应用。Levy<sup>[36]</sup> 提出了一种回归调整的综合估计量, 试图通过结合特定区域的协变量来解释局部变化。Gonzalez et al.<sup>[37]</sup> 给出了综合估计量的定义。复合估计量为了平衡综合估计量的潜在偏差和直接估计量的不稳定性, 将二者进行了加权平均。Schaible<sup>[38]</sup> 讨论了最优权重的选取, Purcell et al.<sup>[39]</sup> 提出复合估计的另一种方法, 对所有区域使用一个共同的权重, 然后最小化总 MSE, 这确保了一些小域的整体效率, 其中关于最小化 MSE 的过程中用到了 JS 方法。这些传统的间接估计方法是基于隐式模型的, 通过补充数据的方式和小域估计相联系。为了反映区域之间的随机效应, 基于模型的方法受到了学者的欢迎。

基于模型的小域估计成为了当下最受欢迎的小域估计方法<sup>[1,3,11,24,29-31,40,41,41-50]</sup>。这主要是依赖于模型估计的几大优点: 一、模型诊断可以用来找到与数据很吻合的适当模型。模型诊断包括残差分析, 以检测偏离假设模型, 为模型选择辅助变量, 以及病例删除诊断, 以检测有影响的观察。二、特定区域精度测量可与每个小域估计相联系, 而不同于综合估计经常使用的平均测量。三、基于模型的估计可以建立线性混合模型以及非线性模型, 如 logistic 回归模型和具有随机面积效应的广义线性模型。对于复杂的数据结构, 如空间依赖和时间序列, 也可以基于模型的方法处理。四、随机效应模型可以用来实现精确的小域推断。基于以上优点, 目前小域估计的研究主要集中于模型的估计。

近 20 年来, 基于模型的小域估计方法成为了该领域研究的主流, 在理论和应用中均涌现了大量的研究成果, Ghosh, Rao, Jiang, Lahiri, Datta, Pfeiffermann

等人<sup>[29-31,40-44]</sup> 在小域理论发展中做了大量的工作。Rao et al.<sup>[1]</sup> 的专著《小域估计》系统介绍了小域研究的基础理论和方法,成为该领域较为全面系统的研究文献。在小域估计中,根据观测数据的不同形式,主要包含两类模型的研究,即区域水平模型和单元水平模型。

在区域水平模型的研究方面。Fay III et al.<sup>[2]</sup> 在根据调查数据研究美国人口少于 1000 的区域上平均收入时,引入了基本的区域水平 Fay-Herriot(FH) 模型,在此后关于区域水平变量的小域估计中,FH 模型成了标准的区域水平模型。Ericksen et al.<sup>[51]</sup> 利用了 FH 模型估计美国十年一次的人口普查的不足数。Dick<sup>[52]</sup> 等研究了 FH 模型在人口统计中的应用。当所研究的区域变量不止一个时,Fay<sup>[46]</sup>, Ghosh<sup>[53]</sup> 提出了多元 FH 模型,并证明它可以得到对小域平均的更有效的估计,因为它利用了区域估计值之间的相关性,而不是单变量模型。Ghosh et al.<sup>[30]</sup> 采用多元模型估计了美国各州(小域)四人家庭的当前收入中位数。Rao et al.<sup>[31]</sup> 推广了基本 FH 模型用来处理时间序列和截面数据,Ghosh et al.<sup>[30]</sup> 提出了一种用于小面积估算的不同时间序列横截面模型。众多的学者将 FH 模型应用于收入估计、失业率估计等方面<sup>([2,54-56])</sup>, Burgard et al.<sup>[57]</sup> 建立了双变量的 FH 模型,用于估计西班牙居住条件调查中的贫困率。

在单元水平模型研究方面。Battese et al.<sup>[3]</sup> 利用样本调查数据结合卫星信息估计县农作物面积时,提出了误差嵌套模型(Nested Error Regression, NER)。作为基本的单元水平模型成为研究单元数据可用时小域估计的有力工具。当关注变量大于 1 个时,多元情形的 NER 模型被广泛使用,在上述估计农作物面积的应用中,考虑大豆和玉米的种植面积的估计即为二元 NER 模型。Stukel et al.<sup>[58]</sup> 通过双嵌套误差回归模型(Two-Fold NER)分析了人口结构。Datta et al.<sup>[41]</sup> 推广了 NER 模型,提出了广义线性混合模型。在广义线性混合模型中,不仅 NER 模型作为其中的一种特殊情形,还包含了逻辑混合模型, Poisson 混合模型等。并将这些模型应用于死亡率估计、劳动力调查分析。Ghosh et al.<sup>[59]</sup> 提出了关注变量来自于指数分布族时的,具有随机面积效应的广义线性模型。Esteban et al.<sup>[7]</sup>, Arias-Salazar et al.<sup>[8]</sup> 提出了针对组合数据的单元水平估计模型,并利用该模型估计了消费结构、贫困率指标的小域估计问题。

经典的区域水平模型由 Fay III et al.<sup>[2]</sup> 提出, FH 模型表述为:

$$y_i = x_i^\top \beta + v_i + e_i, i = 1, \dots, m$$

其中  $m$  为小域的个数,  $y_i$  为第  $i$  个小域上目标变量的直接调查估计量,  $x_i$  为协变量构成的向量,  $\beta$  为未知的回归系数,  $v_i$  为不能由预测因子解释的区域随机效应,

$e_i$  为抽样误差。一般假设  $v_i$  与  $e_i$  相互独立, 且  $v_i \sim N(0, A), e_i \sim N(0, D_i)$ , 其中  $A$  是未知的方差,  $D_i, i = 1, \dots, m$  是已知的抽样方差, 若其未知, 一般可通过抽样获取较为精确的估计。区域模型一般用于表述区域整体水平的变量估计中, 若是估计的区域水平变量不止一个, 还可以建立多维情形的区域水平模型、对于具有空间相关结构的区域水平变量, 可以考虑空间结构的区域水平模型、对于时间序列的区域水平变量, 还可以建立时空相关的区域水平模型。

经典的单元水平模型是由 Battese et al.<sup>[3]</sup> 提出的误差嵌套模型 (Nested Error Regression, NER)。单元水平的 NER 模型为:

$$y_{ij} = x_{ij}^T \beta + v_i + e_{ij}, i = 1, \dots, m, j = 1, \dots, n_i$$

其中  $m$  为小域的个数,  $n_i$  为第  $i$  个小域上样本量,  $y_{ij}$  为第  $i$  个小域上第  $j$  个单元的直接调查估计量,  $x_{ij}$  为相应协变量构成的向量,  $\beta$  为未知的回归系数,  $v_i$  为不能由预测因子解释的区域随机效应,  $e_{ij}$  为随机误差。且  $v_i \sim N(0, \sigma_v^2), e_{ij} \sim N(0, \sigma_e^2)$ , 方差  $\sigma_v^2, \sigma_e^2$  未知。根据单元水平变量的形式, 同样可以建立多维单元水平模型 Esteban et al.<sup>[7]</sup>, Arias-Salazar et al.<sup>[8]</sup> 以及单元水平的广义可加模型 Mori et al.<sup>[60]</sup>。

小域估计中最核心的方法就是如何估计目标参数, 并给出估计参数的分布等性质。估计方法方面, 最优线性无偏估计量 (EBLUP) 是线性无偏估计中的均方误差最小, 并且不依赖于随机效应的正态性, 但是它们依赖于随机效应的方差。针对于一般的线性混合模型  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}$ , 其中  $\mathbf{y}$  是  $n \times 1$  的观测向量,  $\mathbf{X}, \mathbf{Z}$  是辅助信息,  $\mathbf{v}, \mathbf{e}$  是独立的均值为  $\mathbf{0}$ , 方差为  $\mathbf{G}, \mathbf{R}$  的随机效应, 由此可得  $\mathbf{y}$  的  $\tilde{\boldsymbol{\mu}}^H = t(\boldsymbol{\delta}, \mathbf{y}) = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \tilde{\mathbf{v}} = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \mathbf{GZ}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$ , 其中  $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$ ,  $\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\boldsymbol{\delta}) = \mathbf{GZ}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$ 。由于 BLUP 估计量依赖于实际应用中未知的方差参数  $\boldsymbol{\delta}$ 。考虑用  $\boldsymbol{\delta}$  的估计量替换  $\boldsymbol{\delta}$ , 则就是经验最优线性无偏估计量 (Empirical Best Linear Unbiased Prediction, EBLUP)。在 EBLUP 的表达式中, 包含有未知的参数  $\beta, A$ , 如果利用参数的 ML 或者 REML 估计量替换 EBLUP 表达式中的参数, 则可以得到经验贝叶斯 (Empirical Bayes, EB) 估计量。EBLUP 估计量主要是针对连续型观测变量采用的方法, 当观测变量为计数型或逻辑型变量时, 经验贝叶斯和分层贝叶斯 (Hierarchical Bayes, HB) 方法能够很好地应用于这类数据<sup>[61,62]</sup>。EB 方法可以通过广义线性混合模型处理二值数据和计数数据, 也可以通过单元级模型处理一般小面积参数。EB 和 HB 方法的主要区别就在于是否对每个超参数都有一个先验分布。

学者们在基于正则性条件下关于 EBLUP 的研究已经取得了相对成熟的理论<sup>[1,3,46,63]</sup>。针对标准 FH 模型的 EBLUP 估计量为:  $\hat{\theta}_i = (1 - \hat{\rho}_i) y_i + \hat{\rho}_i x_i^\top \hat{\beta}$ , 其中  $\hat{\rho}_i = D_i / (\hat{A} + D_i)$ ,  $\hat{\beta}, \hat{A}$  是  $\beta, A$  的估计值。从 EBLUP 估计量的表达式可以看出, EBLUP 估计量为直接估计量  $y_i$  和回归估计量  $x_i^\top \hat{\beta}$  的加权平均。换句话说, EBLUP 可以看作是用收缩因子  $\hat{\rho}_i$  将直接估计量向回归估计量收缩。如果  $A$  相对于  $D_i$  较大, 即区域间差异较大, 此时收缩因子  $\hat{\rho}_i$  较小, 则直接估计量通过回归估计量从其他区域借鉴的强度不大, 极端情形下, 收缩系数将接近于零, 意味着没有收缩。反之, 若  $A$  相对于  $D_i$  较小, 则区域间差异较小, 此时收缩因子  $\hat{\rho}_i$  变大, 估计量向回归估计收缩。然而, 当观测数据存在离群观测值时, EBLUP 估计量易受极端  $y_i$  的影响, 而达不到收缩的效果, 同时会导致 MSE 较大, 估计的准确性受到影响。同时由于异常观测值在很多情形下都会出现, 这导致了有必要更进一步研究稳健估计量。当  $y_i$  可以用辅助信息  $x_i$  很好地解释时, 经验贝叶斯估计是有用的。然而, 在实际应用中,  $x_i$  并不一定是  $y_i$  在所有领域的良好辅助信息; 换句话说,  $y_i$  在某些区域可能离  $x_i^\top \hat{\beta}'$  很远, 我们称之为离群观测。对于这些观测, 相应的  $\theta_i$  可以由不同于假设的正态分布中产生; 也就是说,  $\theta_i$  的假设分布可能被错误指定。这将导致过度收缩, 以及估计量容易受离群观测的影响。

标准模型下的 EB 估计量是小域估计的理想选择, 但是, 在传统的小域估计模型中, 区域随机效应和模型随机效应部分的误差均假定服从正态分布, 但在实际应用中, 观测数据中离群值很常见, 使用传统的 EBLUP 方法会产生较大的估计偏差。在研究中, 我们重点关注随机效应分布非正态及模型观测存在离群值的基本模型, 并提出针对这些模型的稳健估计方法。

离群观测是任何调查中都不可避免的事实 Chambers et al.<sup>[64]</sup>, 因离群值普遍存在, 且会对估计量产生较大影响, 因此众多研究专门探讨减小离群观测对调查估计影响的方法。常见的有三种: 第一种, 删除离群观测值, 用剩余观测值进行估计和预测。很显然, 这一方法当样本量非常大时, 删除少数的离群观测可能是可行的。但当样本量本来就比较小时, 这种方法不仅会带来信息的丢失, 同时还会导致估计量偏离真实值。第二种, 用非离群观测值替代离群观测值进行估计, 并在非抽样总体中采用稳健投影的方式进行稳健预测。然正如 Chambers<sup>[65]</sup> 所言, 若观测值是准确获取的, 不能认为每个观测值是唯一的, 且没有理由说明在未抽样总体中不包含离群观测值。Hosseini et al.<sup>[9]</sup> 等用半参数的模型研究了保险的小域估计问题。第三种方法, 通过构建对于离群观测值不敏感的稳健估计量, 来减小离群值对估计结果影响, 这类方法也是学者们关注的主要方法, 本文

也采用了这类稳健估计方法。

### 1.2.2 稳健小域估计研究综述

关于稳健估计量的研究方面,早期对 SAE 稳健估计的研究可以追溯到 Ghosh et al.<sup>[66]</sup> 提出的层次均值稳健估计中将小域均值作为一种特殊情况。Datta et al.<sup>[43]</sup> 讨论了异常值是如何影响收缩估计的,即使是一个单一的异常值也可能导致所有小域估计都收缩成相应的直接估计。Datta et al.<sup>[43]</sup> 使用了一个分层贝叶斯 (HB) 框架研究  $v_i$  中异常值的影响,  $v_i$  的分布被假定为正态分布和一般混合分布的比例混合。这种混合族诱导长尾分布,特别是  $t$  分布和柯西分布。基于假设分布族的 HB 估计比基于正态分布的估计对离群  $y_i$  更有稳健性。Bell et al.<sup>[67]</sup> 对  $\theta_i$  的 HB 估计量如何变化进行了实证研究,当使用区域效应  $v_i$  上自由度为  $k$  (较小) 的  $t$  分布来代替正态分布,他们的结果表明,使用  $k$  值较小的  $t$  分布可以减少离群值的影响。因为与  $v_i$  在正态性假定下相比较,利用混合分布时,对正常点直接估计量  $\theta_i$  给与了更大的权重,若  $y_i$  是异常值,那么假定抽样误差  $e_i$  服从  $t$  分布,导致对  $y_i$  给予更小的权重。Ghosh et al.<sup>[68]</sup> 研究了区域水平模型下  $\theta_i$  的稳健估计,使用 Huber' s  $\phi$  函数对 EBLUP 估计量进行稳健化,同时在 Molina et al.<sup>[69]</sup> 中提出了稳健小域估计的 R 包 sae。然而,当正态性假定不满足的时候,其经验贝叶斯估计量表现并不好。Sinha et al.<sup>[28]</sup> 考虑了在正态假设下使用 Huber' s  $\phi$  函数改变似然方差,通过给异常点相关的项更小的权重的方式来获得稳健估计量。目前,在小域估计稳健估计中,这两种方法是应用较为普遍的两种方法。但是,当异常值过大时,估计量的渐近性表现一般。另外关于小域稳健估计的方法还有 Chambers et al.<sup>[70]</sup>, Bianchi et al.<sup>[71]</sup>, Fabrizi et al.<sup>[72]</sup> 提出的基于分位数的小域估计方法, Opsomer et al.<sup>[73]</sup> 利用 P 样条的非参数模型讨论了目标变量的稳健估计。<sup>[74-76]</sup> 等人利用混合效应模型,探究了稳健小域估计问题。

综合当前关于小域稳健估计的研究不难发现,虽然已有不少的文献用于研究稳健小域估计,但是大多都是考虑在特殊情形下的稳健估计,在估计的效果上,现有的稳健估计方法仍会出现对异常值敏感以及过度收缩等问题,导致估计量的方差增大和估计量的 MSE 有偏。为了研究非正态总体以及离群观测存在时的小域估计问题,探究普适的稳健小域估计变得极为重要。Sugasawa<sup>[77]</sup> 首次提出了利用密度幂散度研究 FH 模型的稳健估计方法,并给出了稳健估计量。通过引入密度幂散度,借助于调整参数,得到了依赖于调整参数的稳健估计量,可以根据数据的分布特征,选取适当的调整参数,得到估计量及其估计量的 MSE。基

于这一优点,本研究在此基础上,探索基于  $\gamma$  散度的小域稳健估计和单元水平下的稳健估计方法。Jiang et al.<sup>[63]</sup> 对小域稳健估计的发展以及可能的研究方向进行了综述,并指出测量误差是影响模型假设的主要原因,且稳健估计方法对模型错误假定时是非常重要的,因为对于大多不习惯使用复杂模型的人员更加适用。庞智强等<sup>[78]</sup> 利用  $\gamma$  散度研究了单元水平模型的稳健小域估计问题,给出了对异常值不敏感的稳健估计方法。

我国对小域估计方法的研究相比国外来说起步较晚,理论文献也不多。金勇进等<sup>[79]</sup> 最早详细介绍了抽样调查中常出现的域估计问题,也算是小域相关的问题;李莉莉等(2003)研究了采用样本追加策略与事后分层技术下域估计的理论与方法;冯士雍<sup>[80]</sup> 强调小域估计是我国抽样调查领域需要研究的重点问题,并对经验贝叶斯和分层贝叶斯估计方法进行了简单介绍;吕萍<sup>[81]</sup> 将分位数回归模型引入小域估计中,处理了实际中存在异常值或是重尾现象的小域估计问题;刘乐平等<sup>[26]</sup> 对小域估计中的经验贝叶斯方法和分层贝叶斯方法进行了介绍,并结合实际数据验证了分层贝叶斯方法的稳健性。周巍等<sup>[19]</sup> 利用遥感辅助信息,构建了单元层次的多响应变量的多元回归形式的小域模型,研究了黑龙江省县域水平的农作物小域估计问题。贺建风等<sup>[82]</sup> 分层贝叶斯的广义线性模型的小域估计问题;李腾等<sup>[83]</sup> 讨论了基于混合地理加权 Fay-Herriot 模型的小域估计。巩红禹等<sup>[20]</sup> 将小域估计方法应用到了贫困测度中。刘小惠等<sup>[84]</sup> 探究了模型存在偏误时小域估计的评价问题,并给出了相应的评价方法。上述国内研究大多数是基于小域估计的应用研究,很少有人关注小域稳健估计。

基于密度的最小距离推断是参数统计推断的重要组成部分。在最小距离推断中,通过最小化数据与假设参数模型之间的偏差的适当度量来估计目标参数,而差异是由两个密度函数之间的散度来度量。常见的方法有 KL 散度、密度幂散度(Density Power Divergence,DPD)方法、 $\gamma$  散度等。对于 KL 散度对假设分布的敏感性,Basu et al.<sup>[85]</sup> 等引入了密度幂发散族作为两个概率密度函数之间差异的度量,并使用该族在通常的独立同分布数据设置下估计模型参数。在 DPD 的定义中,通过调整参数  $\gamma$  来平衡估计效率和稳健性。并证明了当调整参数  $\gamma = 0$  时,DPD 散度退化成 KL 散度。同时,由最小化 DPD 方法构建的似然函数在  $\gamma = 0$  时就会变成正态假设下的似然函数。因此,学者们将 DPD 方法作为一种推广应用到稳健估计中,并取得了不错的估计效果,如 Ghosh et al.<sup>[42, 86]</sup>, Riani et al.<sup>[87]</sup> 等讨论了 DPD 在稳健回归中的应用。

Jones et al.<sup>[88]</sup> 引入了另一个具有类似性质的  $\gamma$  散度,Fujisawa et al.<sup>[89]</sup> 利

用  $\gamma$  散度研究了重尾情形下参数的稳健估计, Nakagawa et al.<sup>[90]</sup> 利用  $\gamma$  散度提出了稳健 Bayes 估计方法。 $\gamma$  散度作为另一种统计距离, 在稳健估计中发挥着重要的作用, 在某些情形下能够得到比 DPD 更加稳健的估计效果。Kawashima et al.<sup>[91]</sup> 将  $\gamma$  散度应用于稳健回归模型, 讨论了一般回归模型的稳健性和稀疏性问题。本研究中, 将讨论  $\gamma$  散度在稳健小域估计中的应用, 探讨小域估计模型不满足正态性假设条件下的估计问题。

这两种密度幂散度的主要思想是对与离群值相关的项给予较小的权重, 使得参数估计对离群值变得稳健。在这两种情况下, 调整参数控制稳健性和效率之间的权衡。调整参数的值越小, 模型的效率越高, 调整参数的值越高, 离群值的稳定性越好。在这两种情况下, 对应于调整参数为 0 的最小散度估计量就是最大似然估计量。有关 DPD 散度和  $\gamma$  散度的定义及其比较研究, 将在第三章进行详细的讨论。

在基于散度的稳健估计方法中, 由一个单一的调优参数控制对异常值的稳健性。如果调优参数被设置为比必要值更小的值, 得到的估计量仍可能受到异常值的影响。另一方面, 若选取调优参数的值选择过大, 则会导致统计效率的损失 Basu et al.<sup>[85]</sup>。尽管基于散度的稳健统计方法的性质及其理论分析已得到证明, 但根据观测数据选取调整参数经常被忽视, 只有少数文献 (Warwick et al.<sup>[92]</sup>, Basak et al.<sup>[93]</sup>) 提出了频率论方法, 通过渐近均方误差选择最优调整参数值。Sugasawa et al.<sup>[94]</sup> 提出了最小化 Hyvarinen 得分的参数选择标准。在本研究中, 密度幂散度方法将应用于小域估计模型, 并在调整参数的选择中使用上述两种估计方法, 探索基于数据的自适应调整参数选择方法。详细的算法将在第三章呈现。

### 1.3 研究思路和研究内容

#### 1.3.1 研究思路

结合研究背景和目前的研究现状, 本研究仍然着眼于小域估计问题, 只是本研究的侧重点在于离群观测下的稳健小域估计方法。详细的研究技术线路图如1.2所示。

#### 1.3.2 研究内容

根据小域估计的研究现状和稳健估计的必要性, 本文主要研究内容为小域稳健估计和混合模型中的小域估计的研究。即在标准的小域模型中, 由于较大的

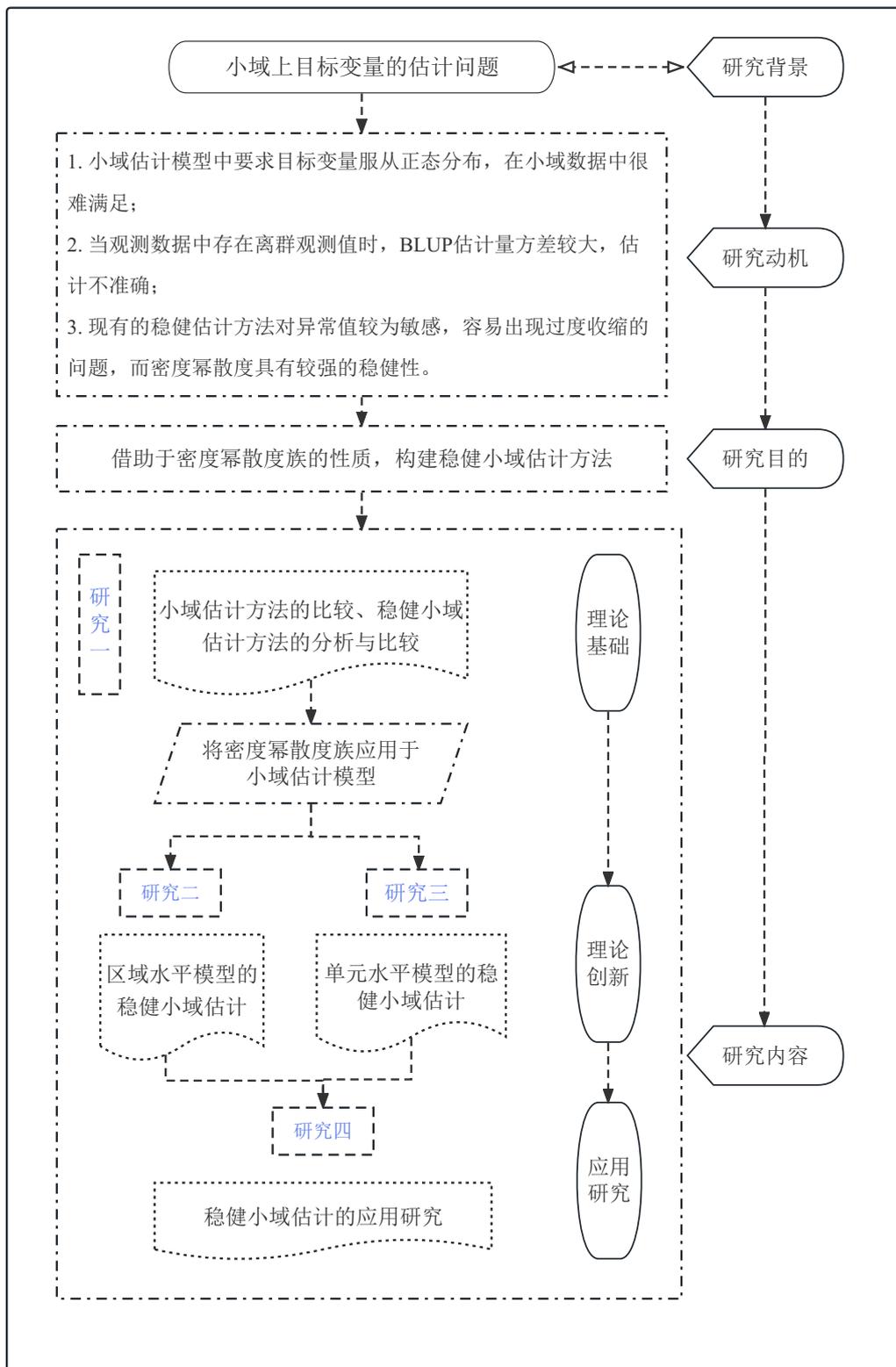


图 1.2 技术线路图

抽样误差、异常观测的出现,导致模型的正态性假设条件不能满足。在这种背景下,研究小域估计的稳健估计量,一方面提高模型对离群观测点的稳健性,另外讨论稳健估计量的渐近性质。主要的研究内容分为如下几个方面:

1) 小域估计理论方法的比较研究。本部分的研究内容主要从两个方面展开,第一,小域估计基本理论方法的比较研究。区域水平模型和单元水平模型作为小域估计的两个基本模型,在正态性假设条件下的估计理论已经比较完善。为探索基本假设条件不满足时的小域估计量,比较 EBLUP 估计在不同的模型条件下的表现,探究其 MSE 的变化情况。第二,总结现有的稳健小域估计量,比较现有稳健估计量和 EBLUP 估计的区别,并总结现有稳健小域估计量的不足。

2) 区域水平模型的稳健估计量研究。在区域水平的  $FH$  模型中,由于传统的 EB 估计量  $\hat{\theta}_i = (1 - \hat{\rho}_i)y_i + \hat{\rho}_i x_i^\top \hat{\beta}$  的收缩因子  $\hat{\rho}_i = D_i / (\hat{A} + D_i)$  只和  $A$  有关,对异常观测值较为敏感。如果用  $\hat{\eta}$  表示估计量,则当  $|y_i - x_i^\top \beta| \rightarrow \infty$  时,  $|\hat{\eta}_i - x_i^\top \beta|$  的表现直接反映了估计量对于异常观测的影响。而 EB 估计量  $|\hat{\eta}_i - x_i^\top \beta| \rightarrow \infty$ ,由此可见,EB 估计量对离群观测比较敏感。因此,研究区域水平模型的稳健估计量是本研究的重点内容之一。本研究将考虑基于 DPD 散度和  $\gamma$  散度的稳健小域估计方法,给出基于两种散度的稳健估计量和估计量的 MSE。

3) 单元水平模型的稳健估计研究。本研究的另一重点是考虑基于 DPD 方法和  $\gamma$  散度的单元水平模型稳健性研究,进一步改善传统稳健估计中过度收缩和估计量方差较大的情形。通过散度族的方式构建似然函数,并在此基础上提出稳健估计量,推导估计量的 MSE。通过实验模拟来说明我们提出统计量的性质,并和现有估计量进行比较。推广基于密度的稳健估计方法,研究在广义线性模型下的稳健估计量。由于受观测数据分布的影响,广义线性模型在小域估计中也较为普遍。异常值同样会影响 EB 估计量,广义线性模型下估计量的稳健估计显得尤为重要。

4) 研究基于稳健估计量的置信区间。由前面的分析可知,在离群观测数据中,传统的置信区间的渐近误差收敛速度较慢,同时由于异常观测值的存在往往会出现置信区间较长的情形。如果用传统的正态方法构造区间估计,可能会导致区间长度太长的情况的发生,因此,在稳健估计量的基础上构建 Bootstrap 置信区间是本研究的重点任务。因此研究存在异常观测时小域目标变量的稳健置信区间的构造也是内容之一。

5) 探究基于观测数据的参数选择算法。本研究将密度幂散度族应用于小域估计,得到对模型假设不敏感以及对异常观测具有稳健性的估计量。为了达到对

观测数据的稳健性，该类方法中引入了一个平衡稳健性和模型效率的调整参数。选择不同的调整参数会导致不同的稳健效果，因此选择最优的调整参数是本研究的关键环节。本研究中将探索基于观测数据的自适应选择算法，根据数据特点选择出最优的调整参数，并将其应用于稳健小域估计。

6) 基于中国家庭追踪调查的应用研究。本研究将在前述理论研究的基础上，探究中国家庭追踪调查中平均收入的小域估计。基于北京大学完成的中国家庭追踪调查数据，分析中国家庭平均收入的小域估计问题。无论是家庭调查还是通常的住户调查，都会涉及到小域估计的问题，同时这类数据中经常会遇到异常观测的问题，因此本文将结合理论研究的结果，分析小域上家庭平均收入的变化情况。

本研究的主要内容和框架可见图1.3。

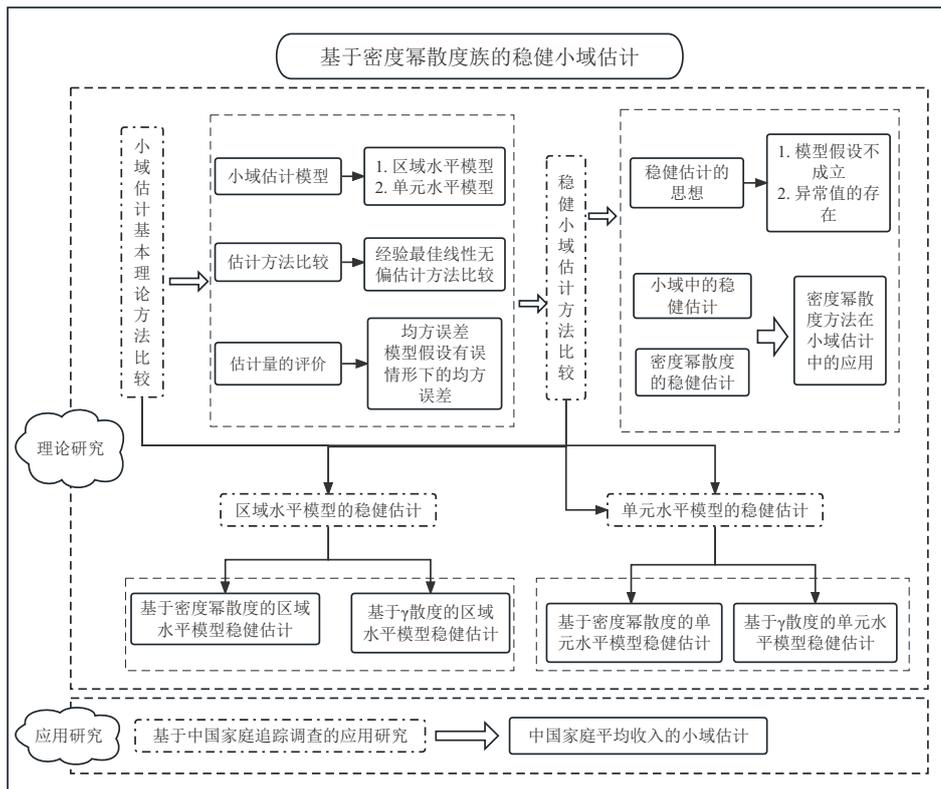


图 1.3 主要研究内容与结构安排

### 1.3.3 研究目的

本研究主要研究稳健小域估计问题，即观测数据存在离群点时小域估计问题。通过结合最小化密度幂散度的方法，改进传统估计方法中估计误差以及估计的均方误差较大的情况，得到在给定精度下的稳健估计。本研究的主要目的分为如下几个方面：

1) 总结小域估计研究的基本理论，探究现有稳健估计方法的不足。虽然小域估计理论具有久远的研究历史，且已经发展的相对成熟，但是在区域随机效应不服从正态分布时，传统的估计理论得到的结果表现不好。通过比较现有的估计理论、稳健估计方法，探索估计量变化的规律，以及估计结果不佳的原因。进一步尝试密度幂散度族的稳健估计方法，并尝试将该类方法应用于小域估计。

2) 改进观测数据中有离群观测时 FH 模型的稳健估计。作为标准的区域水平模型，在实践应用中有非常多的情景会涉及到区域水平的研究，然而，标准的区域水平模型的假设条件在很多情形下难以满足，尤其是在具有离群观测和较大的抽样误差时，EBLUP 估计量难以得到准确的估计。目前的涉及区域水平的稳健方法相对较少，而且容易出现过度收缩，或者导致估计的 MSE 很大等缺陷。本研究借助于密度幂散度族的性质，构建似然函数的方式得到参数的稳健估计量，进而得出小域稳健估计量。

3) 研究小域单元水平模型的稳健估计问题。对于单元级别的数据而言，离群观测存在时同样会影响其 EBLUP 估计量，而且目前关于单元水平模型稳健估计方法的研究较少。在求单元水平模型的小域参数时，似然函数的高维积分问题是制约其稳健估计的因素之一。本文主要借助于密度幂散度、条件分布以及拉普拉斯近似等方式，研究单元水平下稳健小域估计。

4) 研究基于稳健估计量的置信区间。区间估计作为统计推断中一种估计方式，在参数变化范围的估计中引用非常广泛，也是参数估计中不可或缺的一类方法。而小域中目标变量的区间估计对决策者提供变量可靠的变化范围显得至关重要。目前关于小域置信区间的研究相对较少，普遍引用的是基于正态分布构造的置信区间，而这种置信区间的近似误差为  $O(n^{-1})$ 。另一种是利用 Bootstrap 方法构建的置信区间，这种区间的渐近误差为  $O(n^{-3/2})$ ，相比传统的置信区间  $O(n^{-1})$  有了很大的改进。很少有学者关注观测数据在异常情况下的置信区间的构造，本研究将在前述稳健估计量的基础上，考虑置信区间的构造，并推导其渐近误差。

5) 住户调查数据的应用研究。统计部门关于住户调查是每个季度都跟踪统计的调查，然而住户调查的样本量却非常有限。如果用直接估计的方式分析收入

和消费水平,经常会出现估计不准确、偏差较大的情况。本研究将结合小域估计的特点,利用北京大学实施的关于中国家庭追踪调查数据,分析小区域上家庭消费和收入水平,分析区域之间的差别,对政策的制定和区域发展规划提供可行的建议。

#### 1.4 创新点

本研究旨在利用密度幂散度族构建稳健小域估计量,提出稳健小域估计方法。结合前述有关文献梳理和本文研究思路,本文的主要特色和创新之处在于:

1) 针对区域水平模型,提出了基于密度幂散度和  $\gamma$  散度的稳健小域估计方法。在区域水平模型中,利用 ML 方法得到的 BLUP 对异常值较为敏感,本研究中通过比较 BLUP 的基本形式,构建基于密度幂散度和  $\gamma$  散度的稳健估计量。并给出估计参数的性质和渐近分布,给出估计量的置信区间的构造方法和估计量 MSE 的估计形式。

2) 提出了密度幂散度族在单元水平模型下的稳健小域估计方法。不同于区域水平模型,单元水平模型中需要估计区域中每个目标单元的值,但是小域估计问题中难免会有未抽样单元,本研究中给出了借助于密度幂散度族的稳健小域估计方法,通过条件分布的方式给出了估计量的蒙特卡洛估计值,同时给出了对于有限总体中目标变量的估计和区域均值的估计。

3) 本研究中给出了密度幂散度族中的调节参数的选择算法。不同于经验选择方法,本研究给出的调整参数的选择算法能够结合数据的分布特征点选择调整参数,使得估计量具有最小的 MSE。

4) 给出了小域目标估计量 MSE 的 Bootstrap 估计程序,并在此基础上构建小域估计量的置信区间。相比于现有的有关小域估计的研究结论,本研究中给出的小域估计量的置信区间具有更准确的覆盖率和更小的区间长度。

#### 1.5 文章结构

本文主要研究基本的小域模型中基本假设不满足以及观测数据中存在异常观测情形下的小域稳健估计问题。围绕这一背景,分别研究了区域水平模型和单元水平模型下的稳健小域估计问题,提出了稳健估计量,给出了估计量的渐近性质, MSE 的估计等。论文全文共有六章,具体的行文安排如下。

第一章为引言部分,首先介绍了本文的研究背景,回顾了小域估计问题的提出和发展背景。研究范围上,从最初的主要区域估计问题到精细化的小域估计问

题；估计方法上，从基于抽样的直接估计到基于模型的小域估计；在模型的研究中，从标准的小域估计模型到模型基本假定不满足情形下的稳健估计。在此背景下，本研究梳理了小域估计方法、小域估计模型应用、稳健小域估计等方面的文献。在总结梳理现有文献的基础上，提出了当前研究中的不足，提出了本文的研究问题和内容，指出了研究的意义和目的。

第二章为小域估计理论方法的比较研究。在前述文献综述的基础上，整理了目前小域估计的基本理论和方法，并对现有小域估计方法的研究做了比较，指出了现有小域估计方法的不足。尽管有大量的文献用于研究基本的小域估计模型，以及结合基本的小域模型的应用。但是模型中假定区域随机效应和模型误差均服从正态分布，这一假设条件在很多小域估计问题中难以满足，本文指出了在模型不满足基本假定条件下传统小域估计方法的弊端，以及该情形下小域估计结果的 MSE 的变化情况。

第三章为稳健估计方法的比较研究。首先指出了稳健估计研究的必要性，比较了常用的稳健估计方法。通过梳理比较发现，近年来关于密度幂散度的稳健估计方法的研究受到了众多学者的推崇，并指出在特定条件下具有良好的稳健性，且其参数估计具有显式表达。在稳健小域估计领域，虽然稳健小域估计已经引起了部分学者的关注，并提出了相应的稳健估计方法，然而目前的稳健小域估计方法对离群观测仍然较为敏感，容易受随机误差项的方差的影响。考虑到密度幂散度在稳健统计中的优良性质，本文引入了基于密度幂散度的稳健估计方法，并将密度幂散度稳健估计方法应用于基本的小域模型。

第四章为区域水平模型的稳健估计研究。本章给出了密度幂散度、 $\gamma$  散度在区域水平模型中的应用研究。通过将这两类散度应用于区域水平模型，得到了带有调整参数的稳健估计量，从估计量的表达式中不难发现，基于散度的稳健估计方法依赖于随机误差方差的变化，且具有较强的稳健性。本章中给出了区域水平模型稳健估计量的性质，同时给出了调整参数的选择算法，通过选取最优调整算法，使得该模型的 MSE 最小。最后通过模拟比较发现，在随机效应是混合正态分布等有偏分布的假设条件下，基于散度的稳健小域估计方法比现有的稳健估计方法具有更小的 MSE，表现出更好的稳健性。

第五章为单元水平模型的稳健估计研究。本章节将密度幂散度、 $\gamma$  散度应用于单元水平模型，提出了基于这两类散度的稳健小域估计方法。给出了目标变量在单元水平上的稳健估计量，给出了稳健估计量的表达式，研究了估计参数的渐近性质。本章研究了在有限总体情形下的稳健估计量，指出了区域均值的稳健估

计形式。在调整参数的选取中,提出了依赖于观测数据的调整参数选择算法,以及在选定调整参数的情形下的 MSE 的估计。由于单元水平模型中数据的多元特性,很难得到 MSE 的显式表达,本章中给出了稳健估计量 MSE 的 Bootstrap 估计算法。最后通过模拟数据比较了基于散度的稳健估计方法、ML 估计和现有稳健估计方法,通过模拟发现,在模型不满足正态分布时,本章中提出的稳健估计方法无论是参数的估计还是目标变量的估计中,均具有较小的 MSE。说明提出的稳健小域估计方法是有效的,并在实际数据中进行了验证。

第六章是应用研究以及未来研究展望。本章节中,将应用前述章节中有关稳健小域估计的理论方法,研究中国家庭平均收入的小域估计。通过小域估计模型,探究家庭收入的小域分布情况,并探索研究影响区域家庭收入的主要因素。最后,结合本研究的主要结论,指出本研究的不足指出以及需要改善的地方。另外分析说明未来可能研究的方向和内容。



## 第 2 章 小域估计的基本方法及理论

本章主要介绍小域估计的主要发展理论, 常用的小域估计模型, 小域估计的常用方法及其比较研究。基于模型的小域估计中, 最常用的两类模型就是区域水平模型和单元水平模型。本章主要介绍这两类模型的基本形式, 估计理论, 以及这两类模型存在的不足, 这些将为的研究奠定相关的理论基础。

不同于基于抽样数据的直接估计, 基于模型的估计可以针对数据提供更加可靠和适合的模型。另外, 特定区域的精确度量可以与每个小域的估计相关联, 而不像传统综合估计一样的全局度量。同时针对一些复杂的数据, 可以模拟线性混合模型、非线性混合模型、广义线性混合模型等。

在小域估计的框架下, 可从直接估计量的表达形式看出, 基于设计的直接估计仅仅包含了样本  $y_i$  的信息, 对于样本量较少, 乃至无样本的区域得到的估计不可靠, 或者得不到估计。而基于模型的小域估计, 作为感兴趣变量的间接估计, 较好地解决了这一问题。因此需要重点关注基于模型的小域估计方法。

### 2.1 小域估计的模型

#### 2.1.1 区域水平模型

为了对区域上目标变量进行估计, 小域模型成了最有力的工具。其中区域水平模型是最先应用于小域估计问题的模型之一, 而且应用范围也非常广泛。经典的区域水平模型由 Fay III et al.<sup>[2]</sup> 提出, 假设  $\mu_i$  为第  $i$  个区域上观测变量的均值 (或总面积), 设  $y_i$  为  $\theta$  的直接估计量。FH 模型定义为:

$$y_i = \mu_i + e_i \quad \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i, i = 1, \dots, n. \quad (2.1)$$

其中  $n$  为小域的个数,  $y_i$  为第  $i$  个小域上感兴趣变量的直接估计量,  $\mathbf{x}_i$  为协变量构成的向量,  $\boldsymbol{\beta}$  为未知的回归系数,  $v_i$  为不能由预测因子解释的区域随机效应,  $e_i$  为抽样误差。一般假设  $v_i$  与  $e_i$  相互独立, 且  $v_i \sim N(0, A)$ ,  $e_i \sim N(0, D_i)$ , 其中  $A$  是未知的方差,  $D_i, i = 1, \dots, n$  是已知的抽样方差。一般将模型 (2.1) 的未知参数记作  $\boldsymbol{\theta} = (\boldsymbol{\beta}, A)^T$ 。因此, 在模型 (2.1) 的假设条件下有,  $y_i | \mathbf{x}_i \sim N(\mathbf{x}_i \boldsymbol{\beta}, A + D_i)$ 。

FH 模型作为最基本的区域水平模型, 广泛应用于经济统计和人口统计的领域。如 Fay III et al.<sup>[2]</sup> 在研究美国的那些人口总数不超过 1000 的区域的人均收入时引入了 FH 模型, 并被应用于其他统计估计的领域。美国统计局利用

区域水平模型用来估计每个县区、州每年的贫困率和贫困学龄儿童的数量 (Rao et al.<sup>[1]</sup>)。

当需要研究每个区域上多个特征的估计问题时, 多维 FH 模型给提供了可用的框架。该模型是由 Fay<sup>[46]</sup> 通过推广 FH 模型提出。一般地, 假设要估计  $r \times 1$  维的特征构成的向量  $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ir})^T$ , 其中  $\theta_{ij} = g_j(\bar{Y}_{ij})$  和  $\bar{Y}_{ij}$  分别是第  $i$  个区域的第  $j$  个特征的目标值和平均值,  $j = 1, \dots, r$ ,  $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \dots, \hat{\mu}_{ir})^T$  是估计向量。则多元 Fay-Herriot 模型为

$$\hat{\boldsymbol{\mu}}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{v}_i + \mathbf{e}_i, \quad i = 1, \dots, n, \quad (2.2)$$

其中抽样误差  $\mathbf{e}_i = (e_{i1}, \dots, e_{ir})^T \sim N_r(\mathbf{0}, \boldsymbol{\Psi}_i)$ , 其中  $\boldsymbol{\Psi}_i$  是  $\mathbf{e}_i$  在  $\boldsymbol{\mu}_i$  的条件下的条件协方差矩阵。域随机效应  $\mathbf{v}_i \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_v)$ ,  $\mathbf{X}_i$  是一个  $r \times rp$  的辅助信息构成的矩阵, 其第  $j$  行为  $(\mathbf{0}^T, \dots, \mathbf{0}^T, \mathbf{z}_{ij}^T, \mathbf{0}^T, \dots, \mathbf{0}^T)$ ,  $\boldsymbol{\beta}$  是  $rp$  维回归系数向量。

### 2.1.2 单元水平模型

作为小域估计的基础模型之一, 单元水平模型能够处理小域中每个单元级别的目标变量的估计, 并且能通过单元数据计算相应的区域水平值。相比较于区域水平模型, 单元水平模型能够得到更加小的估计区域上的估计、以及个体的估计。因其能够获取更加精细的估计值, 因而与区域水平模型相比, 在区域水平上的估计具有更小的偏差。受数据的收集、辅助变量的获取、以及模型计算等方面的限制, 单元水平模型并不像区域水平那样受学者的关注。如能够获取单元级别的观测数据和辅助信息, 建立单元水平模型则是小域估计中更好的选择。

误差嵌套模型 (Nested Error Regression model, NER) 是最基本的单元水平模型, 由 Battese et al.<sup>[3]</sup> 提出。其具体模型如下:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad j = 1, \dots, N_i, \quad i = 1, \dots, m, \quad (2.3)$$

其中  $y_{ij}$  表示目标变量在第  $i$  个区域上的第  $j$  个单元的取值,  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$  是其对应的单元级别的辅助变量。 $\boldsymbol{\beta}$  是未知的  $p$  维回归系数向量,  $v_i$  是区域随机效应,  $e_{ij}$  是模型随机误差。一般假设  $v_i \stackrel{\text{iid}}{\sim} N(0, \sigma_v^2)$ , 单元随机误差满足  $e_{ij} = k_{ij} \tilde{e}_{ij}$ , 其中  $k_{ij}$  是已知的常数, 且  $\tilde{e}_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$ ,  $v_i$  与  $e_{ij}$  也是相互独立的。

Battese et al.<sup>[3]</sup> 利用 NER 模型根据抽样数据和卫星数据估计县级水平农作物面积。Datta et al.<sup>[41]</sup> 推广了 NER 模型, 利用分层贝叶斯方法讨论了广义线性混合模型下的估计量。有关单元水平模型的相关研究可参见文献 Rao et al.<sup>[1]</sup>, Battese et al.<sup>[3]</sup>, Datta et al.<sup>[41]</sup>, Chakraborty et al.<sup>[48]</sup>。

一般地, 对于包含  $N_i$  个单元的区域  $i, i = 1, 2, \dots, m$ , 假设从中抽取容量为  $n_i$  的样本  $s_i$ , 且抽取的样本同样服从模型 (2.3)。用  $\mathbf{y}_i^P$  记作目标变量在区域  $i$  中所有单元构成的向量, 假设它是随机的, 具有给定的联合概率分布。用  $\mathbf{y}_i$  记作由抽样单元构成的  $\mathbf{y}_i^P$  的子向量,  $\mathbf{y}_{ir}$  为未抽样单元构成的子向量, 而且不失一般性地假设  $\mathbf{y}_i^P$  的前  $n_i$  个单元为抽样单元。则结合抽样数据的单元水平模型可写成

$$\mathbf{y}_i = \mathbf{x}_i \boldsymbol{\beta} + v_i \mathbf{1}_i + \mathbf{e}_i, \quad i = 1, \dots, m, \quad (2.4)$$

其中  $\mathbf{x}_i$  是一个  $n_i \times p$  维的辅助信息矩阵,  $\mathbf{y}_i, \mathbf{e}_i$  均为  $n_i \times 1$  维的向量,  $\mathbf{1}_i$  是  $n_i \times 1$  维的所有元素为 1 的向量。设  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  是模型的未知参数向量。在正态性假定下,  $\mathbf{y}_i | \mathbf{x}_i$  服从正态分布, 其密度函数为

$$f_{\boldsymbol{\theta}}(\mathbf{y}_i | \mathbf{x}_i) = (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta}) \right\}, \quad (2.5)$$

其中

$$\mathbf{V}_i = \sigma_e^2 \text{diag}_{1 \leq j \leq n_i} (k_{ij}^2) + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T.$$

利用下述的 Sherman-Morrison 公式, 可以计算  $\mathbf{V}_i$  的逆矩阵。

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{u}\mathbf{v}^T \mathbf{A}^{-1} / (1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u}),$$

为了简化公式记  $a_{ij} = k_{ij}^{-2}$ ,  $a_{i\cdot} = \sum_{j=1}^{n_i} a_{ij}$ ,  $\mathbf{a}_i = (a_{i1}, \dots, a_{in_i})^T$ , 则有

$$\mathbf{V}_i^{-1} = \frac{1}{\sigma_e^2} \left[ \text{diag}_{1 \leq j \leq n_i} (a_{ij}) - \frac{\rho_i}{a_{i\cdot}} \mathbf{a}_i \mathbf{a}_i^T \right], \quad |\mathbf{V}_i| = (\sigma_e^2 + a_{i\cdot} \sigma_v^2) \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2. \quad (2.6)$$

其中  $\rho_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / a_{i\cdot})$ 。

同样地, 区域  $i$  上的全部  $N_i$  个单元满足模型 (2.3) 的假设, 由全部单元构成的模型可写成

$$\mathbf{y}_i^P = \mathbf{x}_i^P \boldsymbol{\beta} + v_i \mathbf{1}_i^P + \mathbf{e}_i^P, \quad i = 1, \dots, m, \quad (2.7)$$

其中  $\mathbf{x}_i^P$  是一个  $N_i \times p$  维的辅助信息矩阵,  $\mathbf{y}_i^P, \mathbf{e}_i^P$  均为  $N_i \times 1$  维的向量,  $\mathbf{1}_i^P$  是  $N_i \times 1$  维的所有元素为 1 的向量。将模型 (2.7) 按照抽样单元和非抽样单元进行划分,

$$\mathbf{y}_i^P = \begin{bmatrix} \mathbf{y}_{is} \\ \mathbf{y}_{ir} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{is} \\ \mathbf{X}_{ir} \end{bmatrix} \boldsymbol{\beta} + v_i \begin{bmatrix} \mathbf{1}_{is} \\ \mathbf{1}_{ir} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{is} \\ \mathbf{e}_{ir} \end{bmatrix}$$

其中带有下标为  $r$  的均表示由非抽样单元构成的向量。在对未知参数  $\boldsymbol{\theta}$  的估计中, 利用样本数据的条件密度进行估计。

### 2.1.3 线性混合模型

在线性模型中, 假设样本观测是相互独立的, 因此其不能反映数据之间的相关结构。在实际应用中, 线性模型的观测数据通常来源于同一总体, 为了刻画具有更复杂数据结构, 以及来源于不同总体 (群类) 的观测之间的关系, Datta et al.<sup>[41]</sup> 线性混合模型。混合模型可以描述多水平或者分层结构。针对于来源于同一类 (组) 的观测, 要求其是组内相关的, 而来源于不同类的观测, 又是组间相互独立的。混合模型能够很好地捕捉这种数据间的相互关系, 并具有很高的灵活性和适用性。一般的线性混合模型表述为

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \quad (2.8)$$

其中  $\mathbf{y}_{n \times 1}$  是观测值构成的向量,  $\boldsymbol{\beta}_{p \times 1}$  是固定效应向量,  $\mathbf{u}_{q \times 1}$  是随机效应向量,  $\mathbf{X}_{n \times p}$  和  $\mathbf{Z}_{n \times q}$  是已知设计矩阵,  $\mathbf{e}_{n \times 1}$  是模型的误差向量。假设抽样误差和随机误差是相互独立的具有零均值和已知方差矩阵的正态分布随机向量, 即  $\mathbf{v} \sim N(\mathbf{0}, \mathbf{V}_v), \mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ 。其中

$$\text{var}(\mathbf{v}) = E[\mathbf{v}\mathbf{v}^T] = \mathbf{V}_v \quad \text{和} \quad \text{var}(\mathbf{e}) = E[\mathbf{e}\mathbf{e}^T] = \mathbf{V}_e$$

均为依赖于未知参数  $\sigma$  的正定矩阵。结合 (2.8) 可知  $\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{V}_v\mathbf{Z}^T + \mathbf{V}_e$ , 且  $\mathbf{V}$  是非奇异的矩阵。在上述模型中, 如果取  $\mathbf{Z} = \mathbf{I}$ , 同时, 对应的  $\mathbf{y}, \mathbf{x}$  均为区域水平的观测值时, 该模型就变成了基本的区域水平模型。当  $\mathbf{Z} = \mathbf{1}$ , 同时, 对应的  $\mathbf{y}, \mathbf{x}$  均为单元水平的观测时, 该模型就变成了基本的单元水平模型。因此可以认为线性混合模型是上述基本模型的一般化。

线性混合模型建立了目标变量与解释变量之间的线性关系。同时, 线性混合模型可以处理观测数据不独立的情形, 且线性混合模型可以对误差有相关关系的情形建模。线性混合模型作为线性模型的推广, 可以更好地用于对因变量的分析。混合模型可以包含随机效应、层次效应、重复测量、空间和时间相关性等。线性混合模型在小域估计中具有广泛的适用性, 其中结合不同信息来源和解释不同误差来源的灵活性为小域估计提供了应用背景。混合模型通过连接样本的所有观测值, 增加了估计过程中的有效信息, 同时它们可以反映区域间的变化。

### 2.1.4 广义线性混合模型

广义线性混合模型是针对目标变量  $y$  是分类变量或计数型变量时的建模。更一般地, 假设在随机效应  $\mathbf{v}_i$  的条件下, 目标变量  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$  来自于指

数分布族, 其密度函数为

$$f_{y_{ij}|v_i}(y_{ij}|\mathbf{v}_i, \boldsymbol{\beta}, \phi) = \exp \left\{ \frac{y_{ij}\eta_{ij} - b(\eta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\},$$

其中  $a(\cdot), b(\cdot), c(\cdot, \cdot)$  为已知的实值函数。设随机效应  $\mathbf{v}_1, \dots, \mathbf{v}_m$  服从均值为  $\mathbf{0}$  协方差矩阵为  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  的多元正态分布, 其中  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  依赖于未知参数  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^T$ 。一般地, 假设随机效应在区域之间是相互独立的, 在区域之内是相关的, 因此, 协方差矩阵  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  是一个块对角矩阵的形式。广义线性模型通过一个连接函数将观测变量的条件均值和辅助变量联系起来, 其一般形式为:

$$\begin{aligned} \mu_{ij} &= E(y_{ij}|\mathbf{v}_i), \\ g(\mu_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{v}_i. \end{aligned} \quad (2.9)$$

其中  $\mathbf{x}_{ij}, \mathbf{z}_{ij}$  分别是与固定效应和随机效应相关的设计矩阵。

常见的广义线性混合模型中, 当  $y$  服从伯努利分布时, 其对应的模型就成为逻辑混合模型, 当  $y$  服从二项分布时, 可以建立二项混合模型。当然, 根据观测数据的形式, 可以建立区域水平的广义线性混合模型和单元水平的广义线性混合模型。上述模型 (2.9) 可以认为是单元水平的广义线性混合模型的一般形式。

## 2.2 经验最佳线性无偏估计

在本节中, 重点介绍小域模型的参数估计方法和相关的经验最佳线性无偏预测 (Empirical Best Linear Unbiased Prediction, EBLUP)。由上一节关于小域模型的介绍可知, 基本的单元水平模型和区域水平模型可以作为混合模型的特殊情形, 因此, 在这里只讨论关于更一般情形的混合模型的参数估计和 EBLUP。

### 2.2.1 小域模型的参数估计

在小域模型中, 未知参数包含模型的回归系数  $\boldsymbol{\beta}$  和与随机效应协方差矩阵相关的未知参数  $\boldsymbol{\sigma}$ 。针对小域模型的参数估计的常用方法有最小二乘法 (Least Squares, LS) 和极大似然 (Maximum Likelihood, ML) 估计方法, 下面针对两种方法进行详细的阐述。

在模型 (2.8) 中, 观测变量  $\mathbf{y} \sim N(\mathbf{0}, \mathbf{V})$ 。首先对其进行标准化, 令  $\mathbf{y}^* = \mathbf{V}^{-1/2} \mathbf{y}$ ,  $\mathbf{e}^* = \mathbf{V}^{-1/2}(\mathbf{Z}\mathbf{v} + \mathbf{e})$ , 和  $\mathbf{X}^* = \mathbf{V}^{-1/2} \mathbf{X}$ , 则模型 (2.8) 可以写成

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{e}^*,$$

且  $\text{var}(\mathbf{e}^*) = \mathbf{V}^{-1/2} \text{var}(\mathbf{Z}\mathbf{v} + \mathbf{e})\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{I}_n$ 。因此可以使用 LS 方法, 即

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} \{\mathbf{e}^{*T}\mathbf{e}^*\}.$$

因为

$$\begin{aligned} \mathbf{e}^{*T}\mathbf{e}^* &= (\mathbf{V}^{-1/2}\mathbf{y} - \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta})^T (\mathbf{V}^{-1/2}\mathbf{y} - \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^T \mathbf{V}^{-1} \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta} \end{aligned}$$

对上式关于  $\boldsymbol{\beta}$  求偏导, 可得

$$\frac{\partial \mathbf{e}^{*T}\mathbf{e}^*}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} + 2\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta}$$

因此, 当  $\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}$  和  $\mathbf{V}$  可逆时, 可得参数  $\boldsymbol{\beta}$  的估计

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

下面讨论用极大似然估计方法求解未知参数。

Henderson 等<sup>[95]</sup> 利用  $\mathbf{v}, \mathbf{e}$  的正态性假设, 通过最大化  $\mathbf{y}$  和  $\mathbf{v}$  的联合密度求未知参数, 即最大化

$$\phi = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v})^T \mathbf{V}_e^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v}) - \frac{1}{2}\mathbf{v}^T \mathbf{V}_v^{-1} \mathbf{v}$$

在上式中, 关于  $\boldsymbol{\beta}, \mathbf{v}$  求偏导数, 并令其为零。可得估计方程

$$\begin{bmatrix} \mathbf{X}^T \mathbf{V}_e^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{V}_e^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{V}_e^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{V}_e^{-1} \mathbf{Z} + \mathbf{V}_v^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{v}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{V}_e^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{V}_e^{-1} \mathbf{y} \end{bmatrix}$$

在解上述估计方程中, 结合公式

$$\mathbf{V}_e^{-1} - \mathbf{V}_e^{-1} \mathbf{Z} (\mathbf{Z}^T \mathbf{V}_e^{-1} \mathbf{Z} + \mathbf{V}_v^{-1})^{-1} \mathbf{Z}^T \mathbf{V}_e^{-1} = \mathbf{V}^{-1}$$

和

$$(\mathbf{Z}^T \mathbf{V}_e^{-1} \mathbf{Z} + \mathbf{V}_v^{-1})^{-1} \mathbf{Z}^T \mathbf{V}_e^{-1} = \mathbf{V}_v \mathbf{Z}^T \mathbf{V}^{-1}$$

可得  $\boldsymbol{\beta}, \mathbf{v}$  的 BLUP 估计

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}, \quad (2.10)$$

和

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\boldsymbol{\sigma}) = \mathbf{V}_v \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \quad (2.11)$$

结合上述参数  $\boldsymbol{\beta}, \mathbf{v}$  的估计, 针对感兴趣的小域上的变量  $\mu = \mathbf{1}^T \boldsymbol{\beta} + \mathbf{m}^T \mathbf{v}$  的估计为

$$\tilde{\mu} = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \tilde{\mathbf{v}} = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \mathbf{V}_v \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \quad (2.12)$$

### 2.2.2 区域水平模型的 EBLUP

对满足 (2.1) 的模型, 和线性混合模型进行对比, 相当于在线性混合模型中  $\mathbf{Z}_i = \mathbf{1}$ , 协方差矩阵  $\mathbf{V}_{vi} = A$ ,  $\mathbf{V}_{ei} = D_i$ ,  $\mathbf{V}_i = A + D_i$ 。则结合上述关于混合模型的 BLUP 估计量。在区域水平模型中, 重点关注  $\mu_i$  的估计值  $\hat{\mu}_i$ 。因为  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$ , 这相当于在混合模型的估计中取  $\mathbf{1}_i^T = \mathbf{x}_i^T$ ,  $m_i = 1$ , 因此, 基本的 FH 模型的区域均值的 BLUP 估计量为

$$\begin{aligned}\hat{\mu}_i &= \mathbf{1}_i^T \hat{\boldsymbol{\beta}} + \mathbf{m}_i^T \hat{\mathbf{v}}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{v}_i \\ &= \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \frac{A}{A + D_i} (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \\ &= \rho_i y_i + (1 - \rho_i) \mathbf{x}_i^T \hat{\boldsymbol{\beta}}.\end{aligned}\quad (2.13)$$

其中  $\rho_i = \frac{A}{A + D_i}$ ,  $\boldsymbol{\beta}$  的估计量

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} = \left[ \sum_{i=1}^m \frac{\mathbf{x}_i \mathbf{x}_i^T}{A + D_i} \right]^{-1} \left[ \sum_{i=1}^m \frac{\mathbf{x}_i^T y_i}{A + D_i} \right]. \quad (2.14)$$

公式 (2.13) 中给出了区域水平模型中区域均值  $\mu_i$  的最佳线性无偏估计, 可以理解为区域均值的估计为直接估计和回归估计的加权平均, 对应的权重分别为  $\rho_i, 1 - \rho_i$ 。但是, 在实际应用中, 可以结合抽样设计获得抽样方差  $D_i$  的值, 但是模型随机误差的方差  $A$  是未知的, 因此需要对  $A$  进行估计。

目前关于方差  $A$  的估计方法主要有两种, 一种是矩估计方法, 包括 Fay III et al.<sup>[2]</sup> 给出的基于最小二乘的矩估计和 Prasad et al.<sup>[96]</sup> 给出的基于加权最小二乘的矩估计方法; 另一种是迭代估计方法, 通过结合参数  $\boldsymbol{\beta}$  的 ML 估计给出  $A$  的迭代估计。

在正态分布的假定下, 满足 FH 模型的变量  $y_i - \mathbf{x}_i^T \boldsymbol{\beta} \sim N(0, A + D_i)$ , 因此有  $\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{A + D_i} \sim \chi^2(m - p)$ , 其中  $p$  为辅助变量的维数,  $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(A)$  为系数的估计量。根据  $E\left(\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{A + D_i}\right) = m - p$ , 那么  $A$  的估计可以通过解

$$\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{A + D_i} = m - p$$

得到。Fay 和 Herriot<sup>[2]</sup> 提出利用迭代算法求解  $A$ :

$$A^{k+1} = A^k + \frac{1}{f'(A^k)} (m - p - f(A^k))$$

其中,  $f(A) = \sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{A + D_i}$ ,  $f'$  表示  $f(\cdot)$  关于  $A$  的导数。

还有一种思路求  $A$  的估计, 利用 ML 估计方法, 结合似然函数, 利用 Fisher 得分算法, 迭代求解得到  $\hat{A}$ 。

在得到  $\beta$  和  $A$  的估计后, 代入公式 (2.13), 便可得到区域均值的 EBLUP 估计。

### 2.2.3 单元水平模型的 EBLUP

将满足模型 (2.3) 的单元水平模型和模型 (2.8) 的线性混合模型进行对比, 相当于在线性混合模型中, 取  $\mathbf{Z}_i = \mathbf{1}_{n_i}$ , 对应的方差矩阵为  $\mathbf{V}_{vi} = \sigma_v^2, \mathbf{V}_{ei} = \sigma_e^2 \text{diag}_{1 \leq j \leq n_i} (k_{ij}^2)$ , 所以

$$\mathbf{V}_i = \sigma_e^2 \text{diag}_{1 \leq j \leq n_i} (k_{ij}^2) + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T.$$

其中矩阵  $\mathbf{V}_i$  的逆在公式 (2.6) 所示。

对于目标参数  $\mu_i = \bar{\mathbf{X}}_i^T \beta + v_i$ , 相当于在公式 (2.12) 中取  $\mathbf{1} = \bar{\mathbf{X}}, \mathbf{m} = \mathbf{1}$ 。则可以得到  $\mu_i$  的 BLUP 估计量为:

$$\hat{\mu}_i = \bar{\mathbf{X}}_i^T \hat{\beta} + \hat{v}_i \quad (2.15)$$

其中,  $\bar{\mathbf{x}}_i^T$  是辅助信息在每个区域上的均值构成的矩阵, 即  $\bar{\mathbf{X}}_i^T = (\bar{\mathbf{X}}_{i1}, \dots, \bar{\mathbf{X}}_{in_i})^T$ ,  $\hat{v}_i = \gamma_i(\bar{y}_{ia} - \bar{\mathbf{x}}_{ia}^T \hat{\beta})$ , 表达式中的  $\bar{y}_{ia}$  和  $\bar{\mathbf{x}}_{ia}^T$  本别表示抽样数据的加权平均, 通过下式给出:

$$\bar{y}_{ia} = \sum_{j=1}^{n_i} a_{ij} y_{ij} / a_{i.}, \bar{\mathbf{x}}_{ia} = \sum_{j=1}^{n_i} a_{ij} \mathbf{x}_{ij} / a_{i.}.$$

不难发现, 公式 (2.15) 中的公式可以理解为“调查回归”估计量  $\bar{y}_{ia} + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_{ia})^T \hat{\beta}$  和回归综合估计量  $\bar{\mathbf{X}}_i^T \hat{\beta}$  的加权平均:

$$\hat{\mu}_i = \rho_i [\bar{y}_{ia} + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_{ia})^T \hat{\beta}] + (1 - \rho_i) \bar{\mathbf{X}}_i^T \hat{\beta}, \quad (2.16)$$

其中权重  $\rho_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / a_{i.}) \in (0, 1)$  为区域之间不可解释的波动  $\sigma_v^2$  与总变动  $\sigma_v^2 + \sigma_e^2 / a_{i.}$  之间的比例。如果区域间不可解释变动  $\sigma_v^2$  相对较小, 则  $\rho_i$  比较小, 这时综合估计部分具有更大的权重。相反, 当  $a_{i.}$  变大时, 调查回归估计部分具有更大的权重。由于  $a_{i.}$  是与  $n_i$  同阶的, 即  $O(n_i)$ , 并且当  $k_{ij} = 1$  时  $a_{i.}$  等于  $n_i$ , 因此, 当  $n_i$  越大, 调查回归部分的权重越大。

当  $k_{ij} = 1$  时, 调查回归估计量趋近于  $n_i$  较大时简单随机抽样下  $\mu_i$  的设计无偏估计量。在  $k_{ij}$  的一般情形下, 当区域随机效应  $\hat{v}_i$  和  $\hat{\beta}$  是无偏估计的条件下,  $\mu_i$  也是关于模型条件无偏估计。因此, 由公式 (2.16) 得到的 BLUP 估计量是渐近无偏的估计量。

## 2.2.4 线性混合模型的 BLUP

针对形如 (2.8) 的线性混合模型, 一般假设目标变量为  $\mu = \mathbf{l}^T \boldsymbol{\beta} + \mathbf{m}^T \mathbf{v}$ , 则  $\mu$  的 BLUP 估计为

$$\hat{\mu} = t(\boldsymbol{\delta}, \mathbf{y}) = \mathbf{l}^T \hat{\boldsymbol{\beta}} + \mathbf{m}^T \hat{\mathbf{v}} = \mathbf{l}^T \hat{\boldsymbol{\beta}} + \mathbf{m}^T \mathbf{V}_v \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}), \quad (2.17)$$

其中

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

是  $\boldsymbol{\beta}$  的 BLUE, 且

$$\hat{\mathbf{v}} = \hat{\mathbf{v}}(\boldsymbol{\delta}) = \mathbf{V}_v \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}),$$

其中  $\mathbf{V} = \mathbf{v}(\boldsymbol{\delta}) = \mathbf{V}_e + \mathbf{Z} \mathbf{V}_v \mathbf{Z}^T$ .

在一般线性混合模型 (2.8) 的条件下, 将模型写成分块对角矩阵结构, 具体形式如下:

$$\begin{aligned} \mathbf{y} &= \text{col}_{1 \leq i \leq m}(\mathbf{y}_i) = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T, \quad \mathbf{X} = \text{col}_{1 \leq i \leq m}(\mathbf{X}_i) \\ \mathbf{Z} &= \text{diag}_{1 \leq i \leq m}(\mathbf{Z}_i), \quad \mathbf{v} = \text{col}_{1 \leq i \leq m}(\mathbf{v}_i), \quad \mathbf{e} = \text{col}_{1 \leq i \leq m}(\mathbf{e}_i), \end{aligned}$$

其中  $\mathbf{X}_i$  是  $n_i \times p$  的矩阵,  $\mathbf{Z}_i$  是  $n_i \times h_i$  的矩阵,  $\mathbf{y}_i$  是  $n_i \times 1$  的向量,  $\sum_{i=1}^m n_i = n$  和  $\sum_{i=1}^m h_i = h$  成立, 另外有

$$\mathbf{R} = \mathbf{V}_e = \text{diag}_{1 \leq i \leq m}(\mathbf{R}_i), \quad \mathbf{G} = \mathbf{V}_v = \text{diag}_{1 \leq i \leq m}(\mathbf{G}_i),$$

因而  $\mathbf{V}$  同样具有块对角结构,  $\mathbf{V} = \text{diag}_{1 \leq i \leq m}(\mathbf{V}_i)$ , 其中  $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^T, i = 1, \dots, m$ . 至此模型 (2.8) 可以写成如下的形式

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i + \mathbf{e}_i, \quad i = 1, \dots, m \quad (2.18)$$

模型 (2.18) 作为一般的线性混合模型, 包含了很多小域估计的模型。一般感兴趣的是形如  $\mu_i = \mathbf{l}_i^T \boldsymbol{\beta} + \mathbf{m}_i^T \mathbf{v}_i, i = 1, \dots, m$  的估计。结合 (2.17) 中的结论,  $\mu_i$  的 BLUP 估计量为

$$\hat{\mu}_i = t_i(\boldsymbol{\delta}, \mathbf{y}) = \mathbf{l}_i^T \hat{\boldsymbol{\beta}} + \mathbf{m}_i^T \hat{\mathbf{v}}_i,$$

其中

$$\hat{\mathbf{v}}_i = \hat{\mathbf{v}}_i(\boldsymbol{\delta}) = \mathbf{G}_i \mathbf{Z}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}),$$

和

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\delta}) = \left( \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i$$

下面给出针对有限总体的一般预测性结论, 同样将总体划分为抽样单元和非抽样单元。特别地, 若估计的目标参数为目标变量的线性组合形式, 即  $\mu = \gamma^T \mathbf{y}$ , 其中  $\gamma$  是  $N \times 1$  的向量。将其进行分解可以写成

$$\mu = \gamma^T \mathbf{y} = (\gamma_s^T, \gamma_r^T)(\mathbf{y}_s^T, \mathbf{y}_r^T)^T = \gamma_s^T \mathbf{y}_s + \gamma_r^T \mathbf{y}_r.$$

在上述分解条件下, 假设  $\hat{\mu} = \mathbf{g}_s^T \mathbf{y}_s$  是  $\mu$  的无偏估计, 则  $\mu$  的最优线性无偏估计具有如下的结论。

用样本  $\mathbf{y}_s$  估计的  $\mu$  的最优线性无偏估计的一般形式为

$$\hat{\mu}_{opt} = \gamma_s^T \mathbf{y}_s + \gamma_r^T \left[ \mathbf{X}_r \hat{\boldsymbol{\beta}} + \mathbf{V}_{rs} \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}) \right], \quad (2.19)$$

其中  $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{y}_s$ . 且估计量  $\hat{\mu}_{opt}$  的方差为

$$\begin{aligned} \text{var}_M(\hat{\mu}_{opt} - \mu) &= \gamma_r^T (\mathbf{V}_r - \mathbf{V}_{rs} \mathbf{V}_s^{-1} \mathbf{V}_{sr}) \gamma_r \\ &\quad + \gamma_r^T (\mathbf{X}_r - \mathbf{V}_{rs} \mathbf{V}_s^{-1} \mathbf{X}_s) (\mathbf{X}_s^T \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_r - \mathbf{V}_{rs} \mathbf{V}_s^{-1} \mathbf{X}_s)^T \gamma_r \end{aligned}$$

### 2.3 估计量的均方误差

为了评价估计量的表现, 需要进一步探究估计量的 MSE 用于衡量估计量的波动程度。前面内容中出了总体线性参数的 BLUP 估计量, 由于 BLUP 依赖于一些未知参数, 通常情况下估计量的方差是未知的。因此, 需要对这些未知的参数用适当的估计量进行替代, 结果估计量就是 EBLUP。然而, EBLUP 的精确均方误差并没有得到解析的推导。研究者提出了众多的研究方法用来研究 MSE 的近似估计。

关于小域估计量的 MSE 的估计的研究大致可以分为两类方法。第一类方法在正态性假定的基础上, 在正则性假设条件下, 利用 ML 方法或 REML 方法得到参数的估计, 在此基础上利用正则性条件得到 MSE 的近似估计。在模型的误差和随机效应满足正态性的条件下, Kackar et al.<sup>[97]</sup> 得到了 MSE 的简化表达式。在此基础上, Kackar et al.<sup>[98]</sup> 提出了 MSE 的近似表达, 并提出了估计量。Prasad et al.<sup>[96]</sup> 给出了块对角协方差矩阵模型的一种新的近似。在对模型和方差分量的正则性假设条件下, 研究表明当块的个数  $m$  趋于无穷时, 它们的近似是  $o(m^{-1})$  阶的。他们提出了 MSE 的一个估计量, 并给出了 F-H 模型、NER 模型和随机系数模型的估计量的具体表达式。Datta et al.<sup>[99]</sup> 提供了具有块对角协方差矩阵的一般模型的模拟 MSE 估计, 其中方差分量由 ML 或 REML

方法估计得到。Das et al.<sup>[100]</sup> 研究了一类更广泛模型的 MSE 近似, 包括方差分析和纵向随机效应模型, 同样 ML 或 REML 估计用于其方差部分的估计。另一类估计 MSE 的方法是当传统的估计方法无法得到估计量时, 利用重抽样技术估计 MSE 的方法。当然, 即便有时传统估计方法可用, 通过使用 Bootstrap 或刀切法 (jackknife) 方法仍可以得到更准确的估计。Jiang et al.<sup>[40]</sup> 提出了利用刀切法得到一般线性混合模型的最佳预测。Pfeffermann et al.<sup>[101]</sup> 提出了利用参数或非参数 Bootstrap 方法求状态空间模型下 MSE 的估计。Hall et al.<sup>[102, 103]</sup> 介绍了参数和非参数的 Bootstrap 算法, 双重 Bootstrap 算法用于估计 MSE。González-Manteiga et al.<sup>[104]</sup> 利用参数 Bootstrap 方法估计量线性混合模型的 MSE。

综述现有小域估计 MSE 的研究发现, 当模型满足正态性假设条件时, 能够得到小域估计量的 MSE 的显式表达, 并有 MSE 的估计方法。然而, 当模型正态性假定不能满足时, 小域估计量的 MSE 很难得到显式表达, 对其估计也具有一定的挑战。对于本研究中提到的模型不满足正态分布和存在异常观测情形下的小域估计问题, 目前鲜有文献讨论其稳健小域估计量及其 MSE, 而且仅有的关于 MSE 估计的文献也仅仅是使用 Bootstrap 方式给出特殊条件下 MSE 的估计。在本文的后续研究中, 将提出依赖于密度幂散度族的稳健小域估计方法, 该类方法不仅能够得到小域目标变量的估计, 还能借助于密度幂散度的优点给出区域水平模型中 MSE 的显式表达。对于单元水平模型, 由于涉及未知的高维积分问题, 本研究中也给出了其估计量的 MSE 的 Bootstrap 估计方法。下面先讨论基本模型 MSE 的估计方法的比较。

### 2.3.1 基本模型的均方误差

FH 模型的 BLUP 估计量为公式 (2.13) 所示, 文献 Rao et al.<sup>[1]</sup>, Kackar et al.<sup>[97]</sup> 给出了 BLUP 估计量的 MSE

$$\text{MSE}(\hat{\mu}_i) = E(\hat{\mu}_i - \mu_i)^2 = g_{1i}(A) + g_{2i}(A), \quad (2.20)$$

其中

$$g_{1i}(A) = AD_i / (A + D_i) = \rho_i D_i,$$

以及

$$g_{2i}(A) = (1 - \rho_i)^2 \mathbf{x}_i^T \left[ \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T / (A + D_i) \right]^{-1} \mathbf{x}_i,$$

且在正则性假设条件下, 公式 (2.20) 中的第一项  $g_{1i}(A)$  是  $O(1)$  的, 第二项  $g_{2i}(A)$  是  $O(m^{-1})$  的对于较大的  $m$ 。

由于 BLUP 估计量中的参数  $A$  是未知的, 因此需要对  $A$  进行估计, 并将其代入 BLUP 估计从而得到目标参数的 EBLUP 为  $\hat{\mu}_i^E = \hat{\rho}_i y_i + (1 - \hat{\rho}_i) \mathbf{x}_i \hat{\boldsymbol{\beta}}$ , 此时其 MSE 为

$$\text{MSE}(\hat{\mu}_i^E) = g_{1i}(A) + g_{2i}(A) + g_{3i}(A), \quad (2.21)$$

其中  $g_{1i}(A), g_{2i}(A)$  和公式 (2.20) 中的一样, 且

$$g_{3i}(A) = D_i^2(A + D_i)^{-3} \bar{V}(A),$$

其中  $\bar{V}(A)$  是估计量  $\hat{A}$  的渐近方差。有关 EBLUP 估计量的 MSE 的推导可参见文献 Rao et al.<sup>[1]</sup>, Rao<sup>[105]</sup>。

对于形如公式 (2.16) 的单元水平模型的 BLUP, 若设未知参数为  $\boldsymbol{\delta} = (\sigma_v^2, \sigma_e^2)^T$ , 则其 MSE 为

$$\text{MSE}(\hat{\mu}_i) = E(\hat{\mu}_i - \mu_i)^2 = g_{1i}(\sigma_v^2, \sigma_e^2) + g_{2i}(\sigma_v^2, \sigma_e^2), \quad (2.22)$$

其中

$$g_{1i}(\sigma_v^2, \sigma_e^2) = \rho_i(\sigma_e^2/a_i),$$

和

$$g_{2i}(\sigma_v^2, \sigma_e^2) = (\bar{\mathbf{X}}_i - \rho_i \bar{\mathbf{x}}_{ia})^T \left( \sum_{i=1}^m \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} (\bar{\mathbf{X}}_i - \rho_i \bar{\mathbf{x}}_{ia}),$$

上式中, 第一项用于预测区域随机效应  $v_i$ , 且其是  $O(1)$  的, 第二项  $g_{2i}(\sigma_v^2, \sigma_e^2)$  在正则性条件和较大的  $m$  下是  $O(m^{-1})$  的。

由于单元水平模型 BLUP 估计量中的参数  $\boldsymbol{\delta}$  是未知的, 因此需要对  $\boldsymbol{\delta}$  进行估计, 并将其代入 BLUP 估计从而得到目标参数的 EBLUP 为  $\hat{\mu}_i^E = \hat{\rho}_i [\bar{y}_{ia} + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_{ia})^T \hat{\boldsymbol{\beta}}] + (1 - \hat{\rho}_i) \bar{X}_i^T \hat{\boldsymbol{\beta}}$ , 此时其 MSE 为

$$\text{MSE}(\hat{\mu}_i^E) \approx g_{1i}(\sigma_v^2, \sigma_e^2) + g_{2i}(\sigma_v^2, \sigma_e^2) + g_{3i}(\sigma_v^2, \sigma_e^2), \quad (2.23)$$

其中  $g_{1i}(\sigma_v^2, \sigma_e^2)$  and  $g_{2i}(\sigma_v^2, \sigma_e^2)$  和公式 (2.22) 相同, 且

$$g_{3i}(\sigma_v^2, \sigma_e^2) = a_i^{-2} (\sigma_v^2 + \sigma_e^2/a_i)^{-3} h(\sigma_v^2, \sigma_e^2),$$

有

$$h(\sigma_v^2, \sigma_e^2) = \sigma_e^4 \bar{V}_{vv}(\boldsymbol{\delta}) + \sigma_v^4 \bar{V}_{ee}(\boldsymbol{\delta}) - 2\sigma_e^2 \sigma_v^2 \bar{V}_{ve}(\boldsymbol{\delta}),$$

这里  $\boldsymbol{\delta} = (\sigma_v^2, \sigma_e^2)^T$ ,  $\bar{V}_{ee}(\boldsymbol{\delta})$ , 和  $\bar{V}_{vv}(\boldsymbol{\delta})$  分别为  $\hat{\sigma}_e^2$  和  $\hat{\sigma}_v^2$  的渐近方差估计量,  $\bar{V}_{ve}(\boldsymbol{\delta})$  是  $\hat{\sigma}_v^2$  和  $\hat{\sigma}_e^2$  的渐近协方差估计量。有关详细的证明和推导过程可参见文献 Rao et al.<sup>[1]</sup>, Morales et al.<sup>[106]</sup>。

针对单元水平模型的 EBLUP 估计量的 MSE 的估计中, 由于涉及到未知方差以及未抽样样本的数据, 有关上述渐近方差的估计比较困难, 另外当目标变量是  $\mathbf{y}_i$  的非线性函数时, 对 MSE 的估计同样造成了不小的挑战。因而提出基于重抽样技术的 MSE 估计程序。下面分别介绍 Bootstrap 算法和刀切法的一般步骤。

参数 Bootstrap 被用于估计 EBLUP 估计量的 MSE, 这里给出基本的 NER 单元水平模型的 Bootstrap 步骤, 其它模型的步骤与此类似。其具体步骤如下:

1. 首先用 ML 或 REML 方法估计模型中的未知参数  $\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2$  和  $\hat{\sigma}_e^2$ ;
2. 生成 Bootstrap 区域随机效应  $v_i^* \stackrel{\text{iid}}{\sim} N(0, \hat{\sigma}_v^2), i = 1, \dots, m$ ; 生成独立于  $v_1^*, \dots, v_m^*$  的误差  $e_{ij}^* \stackrel{\text{iid}}{\sim} N(0, \hat{\sigma}_e^2 k_{ij}^2), j = 1, \dots, N_i, i = 1, \dots, m$ .
3. 根据模型生成 Bootstrap 总体

$$\mathbf{y}_{ij}^* = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + v_i^* + e_{ij}^*, \quad j = 1, \dots, N_i, \quad i = 1, \dots, m.$$

4. 令  $\mathbf{y}_i^{P*} = (y_{i1}^*, \dots, y_{iN_i}^*)^T$  记作区域  $i$  上生成的 Bootstrap 响应变量, 并计算该区域上的关于目标变量的函数值  $\tau_i^* = h(\mathbf{y}_i^{P*}), i = 1, \dots, m$ . 其中  $\tau(\cdot)$  是关于目标变量的函数。
5. 从 Bootstrap 总体  $\mathbf{y}_i^{P*}$  中随机抽取样本, 在区域  $i$  上的容量为  $n_i$  的样本  $\{(y_{ij}^*, \mathbf{x}_{ij}); j = 1, \dots, n_i, i = 1, \dots, m\}$ 。并用参数估计方法 (ML, REML) 估计模型参数, 记作  $\hat{\sigma}_v^{2*}, \hat{\sigma}_e^{2*}, \hat{\boldsymbol{\beta}}^*$ 。
6. 利用估计的参数得到目标函数值  $\tau_i$  的 EBLUP 值并记作  $\hat{\tau}_i^{E*}, i = 1, \dots, m$ 。
7. 重复步骤 (2)-(6)  $B$  次。对于  $b = 1, \dots, B$ , 令  $\tau_i^*(b)$  是第  $b$  次通过步骤 4 得到的真实值,  $\hat{\tau}_i^{E*}(b)$  为利用步骤 5 得到的估计值。
8.  $\hat{\tau}_i^E$  的 Bootstrap MSE 估计为

$$\text{MSE}_B(\hat{\tau}_i^E) = B^{-1} \sum_{b=1}^B [\hat{\tau}_i^{E*}(b) - \tau_i^*(b)]^2.$$

### 2.3.2 混合模型的均方误差

对于形如 (2.17) 的线性混合模型的 BLUP, 将其分解为

$$t(\boldsymbol{\delta}, \boldsymbol{y}) = t^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) + \boldsymbol{d}^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

上式中,  $t^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y})$  是假设  $\boldsymbol{\beta}$  已知时的 BLUP 估计量, 即

$$t^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) = \boldsymbol{l}^T \boldsymbol{\beta} + \boldsymbol{b}^T (\boldsymbol{y} - \boldsymbol{X} \boldsymbol{\beta}),$$

其中  $\boldsymbol{b}^T = \boldsymbol{m}^T \boldsymbol{G} \boldsymbol{Z}^T \boldsymbol{V}^{-1}$  和  $\boldsymbol{d}^T = \boldsymbol{l}^T - \boldsymbol{b}^T \boldsymbol{X}$ 。根据

$$E[(\boldsymbol{b}^T (\boldsymbol{Z} \boldsymbol{v} + \boldsymbol{e}) - \boldsymbol{m}^T \boldsymbol{v}) (\boldsymbol{v}^T \boldsymbol{Z}^T + \boldsymbol{e}^T) \boldsymbol{V}^{-1}] = \mathbf{0},$$

针对  $[t(\boldsymbol{\delta}, \boldsymbol{y}) - \mu]^2$  求期望, 可得线性混合模型 BLUP 的 MSE 为

$$\text{MSE}[t(\boldsymbol{\delta}, \boldsymbol{y})] = \text{MSE}[t^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y})] + V[\boldsymbol{d}^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})] = g_1(\boldsymbol{\delta}) + g_2(\boldsymbol{\delta}), \quad (2.24)$$

其中

$$g_1(\boldsymbol{\delta}) = V[t^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) - \mu] = \boldsymbol{m}^T (\boldsymbol{G} - \boldsymbol{G} \boldsymbol{Z}^T \boldsymbol{V}^{-1} \boldsymbol{Z} \boldsymbol{G}) \boldsymbol{m}$$

和

$$g_2(\boldsymbol{\delta}) = \boldsymbol{d}^T (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{d}.$$

公式 (2.24) 中的第二项  $g_2(\boldsymbol{\delta})$  反映了  $\boldsymbol{\beta}$  的估计量  $\hat{\boldsymbol{\beta}}$  的波动程度。

由于 BLUP 估计中包含未知参数, 需要进一步对参数进行估计, 并从中得到 EBLUP 估计量, 接下来讨论关于 EBLUP 的 MSE。EBLUP 估计量  $t(\hat{\boldsymbol{\delta}})$  的误差可以分解为

$$t(\hat{\boldsymbol{\delta}}) - \mu = [t(\boldsymbol{\delta}) - \mu] + [t(\hat{\boldsymbol{\delta}}) - t(\boldsymbol{\delta})].$$

对上述误差平方后取期望, 可得

$$\text{MSE}[t(\hat{\boldsymbol{\delta}})] = \text{MSE}[t(\boldsymbol{\delta})] + E[t(\hat{\boldsymbol{\delta}}) - t(\boldsymbol{\delta})]^2 + 2E[t(\boldsymbol{\delta}) - \mu][t(\hat{\boldsymbol{\delta}}) - t(\boldsymbol{\delta})]. \quad (2.25)$$

在随机效应  $\boldsymbol{v}$  和误差  $\boldsymbol{e}$  服从正态分布的条件下, 根据 Kackar et al.<sup>[98]</sup> 的结论, 公式 (2.25) 中的交叉项为 0, 因此

$$\text{MSE}[t(\hat{\boldsymbol{\delta}})] = \text{MSE}[t(\boldsymbol{\delta})] + E[t(\hat{\boldsymbol{\delta}}) - t(\boldsymbol{\delta})]^2. \quad (2.26)$$

从公式 (2.26) 中容易得知在正态性假定条件下, EBLUP 的 MSE 总是大于 BLUP 的 MSE。因此通过  $\text{MSE}[t(\boldsymbol{\delta})]$  去近似  $\text{MSE}[t(\hat{\boldsymbol{\delta}})]$  的做法可能会导致较大

偏差, 使得  $\text{MSE}[t(\hat{\delta})]$  被低估, 尤其是在  $t(\delta)$  随  $\delta$  有较大的变化以及  $\hat{\delta}$  的变动性较大时。

通常情况下, 公式 (2.26) 中最后一项的计算是很困难的。因此需要得到该项的近似表达。Das et al.<sup>[100]</sup> 得到了一般线性混合模型中  $E[t(\hat{\delta}) - t(\delta)]^2$  的二阶近似, 详细内容可参见文献 Rao et al.<sup>[1]</sup>, Kackar et al.<sup>[98]</sup>, Das et al.<sup>[100]</sup>。对公式 (2.26) 中最后一项进行一阶泰勒展开, 若记  $\mathbf{d}(\delta) = \partial t(\delta)/\partial \delta$ , 则有

$$t(\hat{\delta}) - t(\delta) \approx \mathbf{d}(\delta)^T(\hat{\delta} - \delta),$$

其中  $\approx$  用来表示近似值中的剩余项为低阶 (二阶近似值), 且假定包含有  $\hat{\delta} - \delta$  的高次幂项比  $\mathbf{d}(\delta)^T(\hat{\delta} - \delta)$  的阶更低。此外, 在正态性假设下, 包含  $\tilde{\beta} - \beta$  关于  $\delta$  的导数的项具有更低的阶数, 所以

$$\mathbf{d}(\delta) \approx \partial t^*(\delta, \beta)/\partial \delta = (\partial \mathbf{b}^T/\partial \delta)(\mathbf{y} - \mathbf{X}\beta) =: \mathbf{d}^*(\delta), \quad (2.27)$$

进而

$$E \left[ \mathbf{d}(\delta)^T(\hat{\delta} - \delta) \right]^2 \approx E \left[ \mathbf{d}^*(\delta)^T(\hat{\delta} - \delta) \right]^2. \quad (2.28)$$

此外,

$$\begin{aligned} E \left[ \mathbf{d}^*(\delta)^T(\hat{\delta} - \delta) \right]^2 &\approx \text{tr} \left[ E(\mathbf{d}^*(\delta)\mathbf{d}^*(\delta)^T) \bar{V}(\hat{\delta}) \right] \\ &= \text{tr} \left[ (\partial \mathbf{b}^T/\partial \delta) \mathbf{V} (\partial \mathbf{b}^T/\partial \delta)^T \bar{V}(\hat{\delta}) \right] =: g_3(\delta). \end{aligned} \quad (2.29)$$

其中  $\bar{V}(\hat{\delta})$  是  $\hat{\delta}$  的渐近方差矩阵。结合公式 (2.27), (2.28), 和 (2.29) 有

$$E[t(\hat{\delta}) - t(\delta)]^2 \approx g_3(\delta). \quad (2.30)$$

将 (2.30) 式代入 (2.26), 可以得到  $t(\hat{\delta})$  的 MSE 的一个二阶近似为

$$\text{MSE}[t(\hat{\delta})] \approx g_1(\delta) + g_2(\delta) + g_3(\delta). \quad (2.31)$$

上述表达式中, 由于要估计  $\beta$  和  $\delta$ , 比起主要部分  $g_1(\delta)$ , 项  $g_2(\delta)$  和  $g_3(\delta)$ , 的阶相对较低。

## 2.4 稳健小域估计方法

在统计推断和预测中, 大量的统计方法都会依赖于若干假设条件。通过这些假设条件, 可以将所面临的待解决实际问题具体化, 同时能够给出显式的表达形

式且便于计算和处理。然而, 所得到的形式模型是对现实的简化, 其有效性至多是近似接近现实的。例如, 正态性假设条件成了许多模型和方法中的最基本假设条件。由于中心极限定理的存在, 这个假设在统计学中被广泛应用, 并且已经成为回归、方差分析和多元分析中所有经典方法的框架。

在经典统计方法中, 通过正态分布的假设, 能够给出许多真实数据集的近似表示, 同时在理论上可以导出最佳统计方法的显式公式, 如最大似然和似然比检验。然而, 理论和计算的便利性并不是统计实践的目的, 更多的时候是在能达到统计实践目的的条件下再考虑理论和计算的便利。在实践中, 通过正态分布模型可以近似地描述大多数观测结果, 例如具有正态误差的线性回归模型。当模型的随机性分配给观测误差时, 虽然许多数据集的行为看起来相当正常, 但这只是大致如此, 有可能一小部分观察结果由于与大部分数据相距甚远而是非典型的。这种类型的行为在整个数据分析和统计建模应用过程中都很常见。这种非典型数据被称为异常值或离群值, 即使是一个离群值也会对经典统计方法产生很大的扭曲影响。而经典统计方法在正态性或线性假设下是最优的, 这显然与现实情况有所区别。当观测数据的分布中心区域呈正态分布, 但尾部较重或比正态分布的“胖”时, 有时可以用正态分布去近似这种分布。然而, 即使这种近似正态性成立, 使用正态分布理论的结果却不会近似成立。如果假设数据是正态分布, 但实际分布有重尾, 那么, 基于最大似然原则的估计不仅不再是“最好的”, 而且如果尾部是对称的, 可能会有较低的统计效率(不必要的大方差), 如果尾部是不对称的, 可能会有很大的偏差。

统计建模和数据分析的稳健方法旨在推导出可靠参数估计、相关检验和置信区间。这类方法不仅适应于数据完全遵循给定分布的情况, 而且适用于数据仅在上述意义上近似地遵循给定分布的情况。稳健方法以数据为导向的特征是它很好地拟合了大量数据: 如果数据中不包含异常值, 稳健方法得到的结果与经典方法大致相同, 而如果有少量异常值, 稳健方法得到的结果与经典方法应用于“典型”数据的结果大致相同。由于能够很好地拟合大量数据, 稳健方法提供了一种非常可靠的检测异常值的方法, 即使在高维多元情况下也是如此。

Tukey<sup>[107]</sup> 关于混合正态分布的研究标志着现代稳健方法的开始, 产生了处理稳健问题的数学方法。随后, 有较多的统计学家开始关注稳健性统计问题, Huber<sup>[108]</sup> 和 Hampel<sup>[109]</sup> 提出了早期稳健性研究理论。Huber<sup>[108]</sup> 包含了一种新的稳健估计理论, 它详细讨论了受污染正态分布位置参数估计的渐近理论, 并展示了在所有平移不变估计中渐近稳健的估计量。Cressie et al.<sup>[110]</sup> 讨论了在中心

区域为正态分布而尾部大于正态分布时变异函数的稳健估计。Scott<sup>[111]</sup> 研究了综合平方误差, 并以  $L_2$  距离为工具估计多参数统计模型。近几十年来, 越来越多的学者开始关注稳健性问题, 形成了大量有关稳健统计的专著和文章, 同时受益于现代计算机技术的发展和计算能力的提升, 使得稳健估计变得极为普遍且高效, 相关的研究如 Hampel<sup>[112, 113]</sup>, Maronna et al.<sup>[114]</sup>, Huber<sup>[115]</sup>, Fellner<sup>[116]</sup>, 金蛟等<sup>[117]</sup>。

在本节中, 主要研究稳健小域估的研究现状, 小域估计的基本思想, 系统阐述稳健小域估计的基本理论和不足。介绍基于密度幂的稳健估计方法方法。

### 2.4.1 小域中的稳健估计

小域估计的目的是获取小域上尽可能准确可靠的目标变量的统计量, 如均值、分位数、总和等。基于设计的直接估计和基于模型的小域估计是应用最为广泛的方法, 尤其是小域估计方法因其能够处理小样本情形乃至无样本情形而倍受关注。作为基于模型的估计, 这两类方法都严重依赖于模型假设, 即对模型基层分布假设有较强的要求。面对复杂多样的现实问题和数据, 我们很难保证抽样得到的数据是服从指定分布的, 即使有时可以通过数据去拟合分布, 当也难免会出现离群观测导致模型假设不成立的情形发生。因此探究稳健小域估计是获取更加可靠估计的必然选择。

传统的小域估计模型中, 假设模型的随机误差和区域随机效应均服从正态分布。在这一假设条件下, 观测变量也是服从正态分布的。然而在实际应用数据中, 抽样得到的数据可能是非正态分布的, 或者观测数据中存在离群值, 这些问题都会导致小域模型基本假设的失效, 这种情形下模型就是有误的, 需要对模型进行修正或者研究其他的参数估计方法。而这些问题正好是稳健统计中所关注的问题, 需要利用稳健统计的方法解决这些问题。

针对上述问题带来的挑战, 目前在小域估计中的解决方案可以分为如下几类。第一类方法为放宽模型假设的条件, 增加模型的适用范围。这类方法中, 对于不满足模型假设以及传统模型失效的情形, 通过提出更加弱的假设条件使得模型更加切合实际数据, 同时增加模型的应用范围。第二类方法为非参数或半参数方法, 这类方法中, 绕开了极大似然估计等类似方法中有关正态性假定的要求, 从而实现稳健估计。第三类方法统称为稳健小域估计方法。这类方法通过应用稳健统计中的估计方法, 针对上述问题, 提出改进策略和方法以达到稳健性的目的。下面将分别讨论每个方面的方法。

在第一类方法的研究中, 放宽假设条件来实现估计的目的。Ghosh et al.<sup>[66]</sup> 最早讨论了稳健小域估计问题, 提出了分层均值的稳健估计量。他们在后验线性假设条件下推导了经验贝叶斯估计量, 并指出在贝叶斯框架下小域均值的后验估计  $\mu_i$  是抽样观测  $y_i$  的线性函数。Rao et al.<sup>[1]</sup> 的 9.9.2 节中也提到了这种估计方法, 并指出这种假设条件下, 区域均值  $\mu_i$  具有更加广泛的分布。也指出这种方法相当于没有指定观测值  $y_{ij}$  的分布假设。意味着这种方法放宽了原来关于正态性假设的条件。Matloff<sup>[118]</sup> 提出了一种具有更弱假设条件的回归平均 (Regression Average, RA) 的方法用于稳健小域估计。众所周知,  $t$  分布相比较于正态分布具有厚尾的特征, 这种假设可以视为比正态分布更弱的假设, Bell et al.<sup>[67]</sup> 在  $t$  分布的假设条件下研究了具有离群观测的区域水平模型的贝叶斯稳健估计。当模型中存在离群观测值时, 正态性假定显得有些苛刻, 而混合分布可以作为正态分布的一种拓展。为了更好地拟合这种情形, Gershunskaya et al.<sup>[75]</sup> 利用混合分布来模拟离群过程的单元水平模型的小域估计问题, 针对离群区域进行离群值检验, 并根据结果进行稳健估计。通过放宽模型误差同方差的假设条件, 考虑模型误差为异方差下小域估计的研究也受到了广泛的推崇。除了放宽 NER 模型中的常数方差假设外, Jiang et al.<sup>[119]</sup> 考虑了区域内方差完全未知的异方差嵌套误差回归 (HNER) 模型。该文章中提到 HNER 模型比 NER 模型有更弱的假设条件, 通过模拟研究, 作者表明, 一方面, 即使 NER 模型确实成立, 只要样本量  $n_i$  不是太小, 基于 HNER 的 EBLUP 与基于 NER 的 EBLUP 的性能没有太大差异; 另一方面, 如果 NER 模型不成立, 而 HNER 模型成立, 则两者之间存在显著差异, 基于 HNER 的 EBLUP 表现更优。关于模型误差为异方差的详细的讨论可参见文献 Rao et al.<sup>[1]</sup> 中 7.6.1 节的内容。

实现稳健性的另一种方法是降低底层模型的限制性。最直接的方法就是建立非线性的模型, 可以包括高阶项, 如协变量的二次或三次函数。高阶模型将线性模型作为一种特殊情况 (高阶项系数为零), 因此比线性模型的限制性要小, 即使线性模型失效, 高阶模型仍然可能有效。更一般地, 可以用非参数或半参数的方法来模拟平均函数。针对非线性的小域估计模型, Opsomer et al.<sup>[73]</sup> 通过使用惩罚样条作为非参数趋势的表示, 将非参数小域估计问题表示为混合效应模型的回归。通过使用现有的模型拟合方法, 如限制最大似然, 得到模型参数的估计。Rao et al.<sup>[120]</sup> 考虑了半参数混合模型下的小域估计问题, 通过使用惩罚样条函数逼近半参数混合模型, 得到了小域均值的 EBLUP 和 MSPE 的 Bootstrapstrap 估计, 扩展了 Sinha et al.<sup>[28]</sup> 的工作。Lombardía et al.<sup>[121]</sup> 研究了线性部分包含随

机效应的广义部分线性模型,提出了一个将混合效应模型的似然方法与核方法相结合的估计量,并将其应用到了小域估计领域。Datta et al.<sup>[122]</sup> 考虑了结构测量误差模型和基于 Fay-Herriot 模型的半参数方法,为感兴趣的小区域特征产生可靠的预测区间。

除了上述两种方法用于稳健小域估计之外,另外一种方法就是借助于小域估计模型,应用稳健统计中用到的估计方法,来实现估计量的稳健性。Jiang et al.<sup>[63]</sup> 中提到,小域估计方法在某种程度上更稳健。通过前面的论述我们已经了解了如下事实:首先,在统计调查中,离群观测是不可避免的,且根据调查数据,并不能判定已有的离群值是唯一的,即不能确定未抽样数据中是否包含有离群值。其次,统计模型只能根据观测数据的特点,尽可能的去拟合真实数据。通过改善假设条件使统计模型更加贴合真实数据,有可能出现模型错误的情况。针对这两个问题,在小域估计中,已有相关的文献关注到这两方面的研究。当模型中存在离群值时,最直观的方法当然就是选取对离群值不敏感的估计方法。Chambers et al.<sup>[70]</sup> 提出了基于 M 分位数回归的小域估计方法,利用分位数的优点来消除异常值带来的影响。在高斯混合模型下,EBLUP 估计是有效的,而且具有更好的统计性质,但是该方法对离群值异常敏感,而离群值的不可避免就可能导高斯混合分布的假设不可能完全成立。在利用稳健统计方法解决小域估计稳健性的研究中,Sinha et al.<sup>[28]</sup> 为代表性的工作,其利用 Huber $\phi$  函数来改变似然估计方程中的具有异常值的项的权重,使具有异常值的项具有更小的权重,以减少异常值带来的影响。该论文中还提出了稳健估计量的渐近正态性,利用 Bootstrap 方法得到了 MSE 的估计。Sugasawa<sup>[77]</sup> 将密度幂散度引入小域估计中,给出区域水平平均值的稳健估计。当模型失效时,基于模型的估计方法就不太可靠,而基于设计的方法不受模型假设的影响,但是基于设计的方法是建立在大样本基础之上的,因此,产生了将基于模型的方法和基于设计的方法相结合的思想。Jiang et al.<sup>[123]</sup> 提出了一种模型辅助的 EBP 方法,使用假设的混合效应模型(可能是线性或非线性的)来推导 EBP。同时证明了无论假设模型是否正确,提出的 EBP 方法在大样本时是接近基于设计的估计,这说明了至少在大样本区域上提出的 EBP 方法是有效的。

结合目前稳健小域估计研究的现状,尽管有大量的文献在关注稳健小域估计,但是大多数方法只能解决特定条件下的稳健性,同时对于估计量的性质和估计参数的性质的研究有待进一步完善。因此,我们尝试给出稳健小域估计的新的方法,能够尽可能完善目前的研究。在本研究中,主要聚焦于基于密度幂散度的

稳健小域估计方法。详细内容将会在后续的章节中进行阐述说明。

#### 2.4.1.1 稳健估计的主要思想

通过前面的论述我们已经知道，当模型中存在离群观测或者模型假设存在错误时，传统的估计方法产生的估计量不太可靠，需要探索更加稳健的估计量。这里所说的模型假设错误可能更容易理解，主要包含两个方面。其一，假设的模型和数据真实的结构相差甚远。例如数据变量之间本来是指数函数的关系，而我们假设成了线性关系，这就会导致模型假设错误的发生。其二，模型的基本形式准确，但是模型的误差或变量的提取不准确。例如，当选取的解释变量和被解释变量之间并没有关系时，将其纳入模型，就会导致模型的错误，另外模型的误差分布通常假设是正态分布，这一假设也未必成立，这也会导致模型假设错误。因此，我们为了保证模型假设正确，在建立模型之前就要做大量的基础分析工作，才能保证模型的准确性。

离群观测作为抽样中不可避免的情况，对估计结果造成不可估量的影响。关于离群值的界定，主要形成了这样的几个观点。Hawkins<sup>[124]</sup>中提到了关于离群值的定义：“异常值是一种观测结果，它与其他观测结果相差甚远，以至于人们怀疑它是由不同的机制产生的”。离群值是与其它数据显著不同的数据，一般来说，一个离群值主要表现为：偏离数据的正常行为（已知）、观测值与期望值相差甚远、与其它任何数据不相连或相似等。在有些文献上，将离群值 (Outlier) 称为异常值。在本文中，因为我不去能确定这种观测是否正确，为了更加直观和分析需求，故而将其称为离群值。有关离群值的判别方面，常见的做法有如下几类：和常用的标准进行比较判别，若观测值与标准值相差很大，则判定为离群值；利用  $3\sigma$  原则，超出  $3\sigma$  范围的观察认为是离群值。通常可以通过基本的描述统计、箱线图、散点图等工具来判定观测值是否为离群值。有关离群值的详细定义，诊断以及相关研究可以参见文献 Hawkins<sup>[124]</sup>, Aggarwal<sup>[125]</sup>。

根据以上论述，需要建立针对模型假设和离群值稳健的估计方法。那么，怎样的统计方法才能算是稳健的统计方法？关于稳健方法的描述，Huber<sup>[126]</sup>提出了两个原则，

1. **定性的稳健性**。基础分布的微小变化只会导致统计方法性能的微小变化。
2. **定量的稳健性**。稍微偏离核心模型不会导致灾难。

上述原则中，定性的稳健性要求说明一个稳健的估计方法应该不会因基础分布的微小变化改变估计的效率。定量的稳健性强调数量上的变化，当模型潜在的分

布发生变化时, 估计量的分布发生变化的多少。Huber<sup>[126]</sup> 中针对上述两条原则给出了详细的说明, 并给出了数学推导。

小域估计中, 目标变量的 EBLUP 可以表示成“直接估计量”和“回归估计量”的加权平均的形式, 例如 (2.12)、(2.13)、(2.16) 均是这种形式。这里以 (2.13) 为例, 说明稳健估计的主要思想。在 (2.13) 中,  $D_i$  为已知的抽样方差,  $A$  为未知的模型误差。若  $y_i$  为离群观测, 则  $A$  的估计会变大, 这导致  $\rho_i$  的估计变大, 通过 (2.13) 不难发现, BLUP 估计量给予离群观测更大的权重, 因此会导致估计量向离群值靠近, 与真实情况相悖, 因为我们希望离群值尽可能小的影响估计结果。而稳健的估计量则通过给离群值赋予更小的权重, 来减小其对估计结果的影响。这一点在后面的讨论中会详细阐述。

#### 2.4.1.2 常见的几种稳健小域估计方法

目前关于小域稳健估计方法中, 大量的研究方法主要集中于针对特殊问题的小域估计问题, 通过前述论述可知, 通过不同的途径来达到稳健估计的目的。应用广泛的理论方法有 Ghosh et al.<sup>[27]</sup> 中提到的针对区域水平模型的稳健估计方法, Sinha et al.<sup>[28]</sup> 提出的针对单元水平模型的小域稳健估计方法, 本节将针对这两种方法进行简单的回顾。

众所周知, 当区域观测中出现离群值时, 传统的小域估计量具有较大的偏差和均方误差, 其 EBLUP 估计表现易受离群观测影响。为了探究区域水平模型的小域稳健估计, Ghosh et al.<sup>[27]</sup> 通过测量有离群观测和没有这种观测的回归系数的后验密度函数之间的散度, 找到对应于每个区域级观测的影响函数。然后, 在对这些影响函数进行适当标准化的基础上, 提出了一些新的稳健经验贝叶斯小域估计。针对形如 (2.1) 的 FH 模型, Ghosh et al.<sup>[27]</sup> 通过比较两中估计参数的散度, 结合其影响函数, 给出了区域目标变量  $\theta_i$  的经验贝叶斯分层估计量:

$$\tilde{\theta}_i^G = y_i - \frac{D_i v_i(A)^{1/2}}{A + D_i} \psi_K \left\{ \frac{y_i - x_i^T \hat{\beta}(A)}{v_i(A)^{1/2}} \right\}, \quad (2.32)$$

其中

$$\hat{\beta}(A) = \left( \sum_{i=1}^m \frac{x_i x_i^T}{A + D_i} \right)^{-1} \left( \sum_{i=1}^m \frac{x_i y_i}{A + D_i} \right), v_i(A) = A + D_i - x_i^T \left( \sum_{i=1}^m \frac{x_i x_i^T}{A + D_i} \right)^{-1} x_i,$$

$\psi_K(u) = u \min(1, K/|u|)$  是调整参数为  $K > 0$  的 Huber  $\psi$  函数。

在第2章介绍了线性混合模型的基本理论, 包括参数的估计和 EBLUP。针对

形如 (2.8) 的线性混合模型，利用极大似然估计方法的混合模型方程为

$$\begin{aligned} \mathbf{X}^T \mathbf{V}_e^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v}) &= 0, \\ \mathbf{Z}^T \mathbf{V}_e^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v}) - \mathbf{V}_v^{-1}\mathbf{v} &= 0, \end{aligned} \quad (2.33)$$

上述方程组的解为 (2.10) 和 (2.11) 中所示的估计参数，同时目标变量的 BLUP 和 EBLUP 均是根据上述参数的估计而得到，具体表达式可见第2章。

Fellner<sup>[116]</sup> 通过调整估计方程 (2.33) 来限制离群观测的影响，得到了线性混合模型参数  $\boldsymbol{\beta}$  和  $\mathbf{u}$  的估计。该文献中提出使用影响函数来限制 (2.33) 中残差的影响，其具体公式为：

$$\mathbf{X}^T \mathbf{V}_e^{-\frac{1}{2}} \psi \left( \mathbf{V}_e^{-\frac{1}{2}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) \right) = 0, \quad (2.34)$$

和

$$\mathbf{Z}^T \mathbf{V}_e^{-\frac{1}{2}} \psi \left( \mathbf{V}_e^{-\frac{1}{2}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) \right) - \mathbf{V}_u^{-\frac{1}{2}} \psi \left( \mathbf{V}_u^{-\frac{1}{2}}\mathbf{u} \right) = 0. \quad (2.35)$$

其中  $\psi(\cdot)$  是有界单调函数，Fellner<sup>[116]</sup> 建议可以利用 Huber<sup>[108]</sup> 中提到的 Huber 影响函数  $\psi_b(x) = x \min \left( 1, \frac{b}{|x|} \right)$ ，这里  $b$  是一个调整参数，通常选取  $b = 1.345$ 。基于这一结果，Sinha et al.<sup>[28]</sup> 提出了一种类似的方法用于估计单元水平模型的稳健估计。与 Fellner<sup>[116]</sup> 的方法不同，其估计方程是由 (2.8) 的边际模型导出的，即对  $\mathbf{y}$  的边际对数似然函数对  $\boldsymbol{\beta}$  和  $\boldsymbol{\delta}$  求一阶导数，结果为：

$$\mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0,$$

和

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_l} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_l} \right) = 0,$$

对  $l = 1, \dots, Q$  其中  $Q$  表示方差参数  $\boldsymbol{\delta}$  中未知参数的个数。利用类似于 (2.35) 的方法修正上述估计方程，以限制离群观测的影响，因为残差具有有界的影响，使用的影响函数为  $\psi(\cdot)$ ：

$$\begin{aligned} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) &= 0, \\ \psi(\mathbf{r})^T \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \text{tr} \left( \mathbf{K} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_l} \right) &= 0, \end{aligned} \quad (2.36)$$

其中  $\mathbf{K} = \mathbb{E}(\psi_b^2(z)) \mathbf{I}_n$  为与  $\mathbf{V}$  具有相同结构的对角矩阵，另  $\mathbf{r} = \mathbf{r}(\boldsymbol{\beta}) = \mathbf{U}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  表示标准化残差的向量， $\mathbf{U} = \mathbf{U}(\boldsymbol{\delta})$  是包含  $\mathbf{V}$  的对角线元素的矩阵。

有了混合模型估计方程, 需要对上述方程求解, 从而得到参数  $\beta$  和  $\delta$  的估计。Sinha et al.<sup>[28]</sup> 建议使用牛顿拉夫逊 (Newton-Raphson) 算法得到参数  $\beta$  和  $\delta$  的稳健估计。Rao et al.<sup>[1]</sup> 中 7.4.1 节中指出, 当参数  $\beta$  和  $\delta$  的初始值与其真实值相差较大时, 该算法不稳定。作为补充, Chatrchi<sup>[127]</sup> 提出利用不动点算法求解上述方程组得到参数  $\beta$  和  $\delta$  的估计。Schoch<sup>[128]</sup> 提出利用迭代重加权最小二乘 (IRWLS) 求解上述方程。有关算法的详细推导可参见参考文献 Rao et al.<sup>[1]</sup>, Sinha et al.<sup>[28]</sup>, Chatrchi<sup>[127]</sup>, Schoch<sup>[128]</sup>。

在给定参数  $\beta$  和  $\nu$  的条件下, Sinha et al.<sup>[28]</sup> 提出直接将参数估计值代入 (2.12) 的 EBLUP 估计式, 得到稳健 EBLUP (REBLUP)。从而得到区域均值的 REBLUP,

$$\hat{y}_i^\psi = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \left( \mathbf{x}_{ij}^T \hat{\beta}^\psi + \hat{\nu}_i^\psi \right) \right\}, \quad (2.37)$$

其中, 上标  $\psi$  表示估计  $\hat{\beta}^\psi$  和  $\hat{\nu}^\psi$  依赖于影响函数  $\psi$ 。

#### 2.4.2 密度幂散度族及其比较

参数最小距离推断中最重要的思想是通过适当的类距离测度, 将样本数据和参数模型之间的接近程度量化为一组未知参数的函数。因此, 参数的估计是通过最小化参数空间上的这个“距离”来获得的。需要说明的是这类距离并不是严格意义上的数学距离, 它们的参数可能不对称, 也可能不满足三角形不等式。对这些度量要求的唯一性质是非负性, 当且仅当参数完全相等时, 距离应该等于零。在统计推断中将这类距离称为“散度 (divergence)”, 在相关的文献研究中将这种方法称为“最小距离”或“最小散度”。基于密度的散度形成一类特殊的统计距离, 其中几个最小距离估计量具有很高的模型效率。特别地, 最大似然估计量 (MLE) 也属于基于密度的最小距离估计量, 是似然函数的最小值, 这是 KL(Kullback-Leibler) 散度下的最小距离估计量。但是 MLE 的主要缺点之一是它不具有稳健性, 即使是很小比例的离群观测也会导致毫无意义的推断。事实上, 正是经典方法的失效, 如最大似然处理异常值和模型的轻微偏差, 导致了稳健性领域的出现, 更详细的表述可参见 Ronchetti et al.<sup>[4]</sup>, Maronna et al.<sup>[114]</sup>, Rousseeuw et al.<sup>[129]</sup>, Pardo<sup>[130]</sup>。

由于最小化统计距离方法能够兼顾较强的稳健性和较高的模型估计效率, 基于不同散度族的最小距离推断开始不断地涌现。如 Rousseeuw et al.<sup>[129]</sup>, Pardo<sup>[130]</sup>, Bregman<sup>[131]</sup> 等文献中提到了基于  $\phi$  散度族的最小距离测度、基于 Bregman 散度族的最小距离估计量。其中 Bregman 散度 Bregman<sup>[131]</sup> 是一种点间距离的度量,

已经在数学和信息论有广泛的应用。当点用概率分布表示时,对应的 Bregman 散度是一个统计距离,详见 Pardo<sup>[130]</sup>, Stummer et al.<sup>[132]</sup> 等关于 Bregman 散度的一些理论和相关应用示例。目前统计文献中 Bregman 散度估计的主要代表是基于 Basu et al.<sup>[85]</sup> 的密度幂散度 (DPD) 类的最小密度幂散度估计量 (MDPDE)。在过去的二十年中,这类散度提供了便于应用的方法来平衡参数估计、假设检验和相关推断中的稳健性和效率。基于 DPD 的最小散度估计量已被证明在模型不规范的情况下提供了高度的稳定性,且在模型效率方面损失最小。

统计距离推断作为统计推断的重要内容,受到了众多学者的关注,如 Pardo<sup>[130]</sup>, Basu et al.<sup>[133]</sup> 系统描述了距离推断的定义和重要的应用。在统计推断的基础方法中,似然方法作为应用较多的方法之一,其本身就属于最小距离推断方法。Lindsay<sup>[134]</sup> 就指出,最大似然估计方法是最小化 Kullback-Leibler(KL) 散度的情形。在不同的最小散度方法中,基于密度的散度最小化方法尤为重要,由于在稳健性和有效性等方面的优点,广泛应用于稳健估计领域。基于密度函数的最小距离估计值具有很强的稳健性,这使得它在模型假设错误或数据集包含与模型假设不一致的离群值的情况下特别有用。密度幂散度 (Density power divergence,DPD) 作为统计距离中的一类,主要用于度量两个概率密度函数之间的差异。该方法被 Basu et al.<sup>[85]</sup> 提出以来,被广泛应用与稳健估计领域。Basu et al.<sup>[85]</sup> 将 DPD 方法应用于独立同分布的数据得到了稳健估计量。Ghosh et al.<sup>[135]</sup> 将 DPD 方法推广到了独立但非同分布的数据的稳健估计。使用 DPD 方法能够很好地达到稳健估计的目的,其主要原理是通过与观测数据的密度有关的权重来降低离群值对估计量的影响,对离群观测值赋予更小的权重达到稳健估计的目的。这种估计方法的一个特别吸引人的特点是在不使用任何非参数密度估计技术(如核密度估计或基于样条的方法)的情况下构造要最小化的目标函数。Basu et al.<sup>[133]</sup> 指出,该方法在理论和实践上都比最小 Hellinger 距离估计及相关方法简单。而且该方法应用广泛,不仅在统计领域,而且在机器学习、数学信息理论和许多其他科学分支中都有重要的应用。

关于 DPD 方法的研究中,Ghosh et al.<sup>[86]</sup> 将 DPD 方法引入贝叶斯分析框架,提出了以  $R^\alpha$  为后验密度的贝叶斯估计方法。Ghosh et al.<sup>[42]</sup> 研究了 DPD 方法在广义线性模型中的应用,提出了针对计数变量的 Poisson 回归和针对分类变量的逻辑回归模型的稳健估计方法。Patra et al.<sup>[136]</sup> 研究了  $\phi$  散度族和密度幂散度族之间的数学关系。Basu et al.<sup>[137]</sup> 考虑了基于 DPD 的参数假设检验问题,并给出了检验统计量的渐近零分布,探讨了检验统计量的渐近幂函数。

Roy et al.<sup>[138]</sup> 在密度幂的基础上提出了一类包含 DPD 的新的 Bregman 散度。Saraceno et al.<sup>[139]</sup> 讨论了基于最小化 DPD 方法的混合线性模型的稳健估计, 并讨论了估计量的渐近性质。

在密度类散度中, Jones et al.<sup>[88]</sup> 将 DPD 方法中的函数替换成对数函数之后, 提出了一种新的密度幂散度, 并将这种散度和 DPD 进行了比较。Fujisawa et al.<sup>[140]</sup> 将这种散度称为  $\gamma$  散度, 并表明相应的估计量在某些条件下对严重污染数据具有较小的潜在偏差。密度幂发散等其他稳健方法不能达到上述性质, 且估计量会受到离群值比率的影响。Kawashima et al.<sup>[91]</sup> 将  $\gamma$  散度扩展到回归问题, 研究了基于  $\gamma$  散度的稳健稀疏回归, 并表明它在严重污染下具有较强的稳健性, 即使离群点是异质的。Nakagawa et al.<sup>[90]</sup> 将  $\gamma$  散度应用于贝叶斯估计中, 基于  $\gamma$  散度提出了针对离群值稳健的后验分布, 并展示了所提估计的渐近性质。

从上面的综述不难发现, DPD 和  $\gamma$  散度已经广泛应用于统计推断的各个领域, 足以表明这两个散度在稳健估计中具有很好的性质。但在小域估计领域, 鲜有学者将其应用于小域稳健估计, 本文中, 将这两类散度距离应用于小域估计中, 研究小域稳健估计问题, 着重解决传统的小域估计模型中模型正态性假定不满足以及观测数据中存在离群值的问题, 通过引入 DPD 和  $\gamma$  散度来得到小域稳健估计量。本文重点研究区域水平模型、单元水平模型以及线性混合模型的小域稳健估计问题。相关的研究内容将在本文后续的章节展开研究分析。

#### 2.4.2.1 密度幂散度族的定义

在统计推断中, 我们假设样本  $X_1, X_2, \dots, X_n$  是独立同分布的, 是来自于真实分布为  $G$  (密度函数为  $g$ ) 的样本。考虑模型的一个参数族  $\mathcal{F} = \{f_\theta\}$ , 其下标  $\theta \in \Theta \subset R^p$  表示参数空间中不同的参数, 密度函数为  $f_\theta$ 。我们的目标是用参数类密度  $\mathcal{F}$  来模拟未知的真密度  $g$ 。

根据 Basu et al.<sup>[85]</sup> 的结论, 定义密度函数  $g$  和  $f$  之间的 DPD 散度为

$$d_\alpha(g, f) = \int \left\{ f^{1+\alpha}(z) - \left(1 + \frac{1}{\alpha}\right) g(z)f^\alpha(z) + \frac{1}{\alpha} g^{1+\alpha}(z) \right\} dz, \quad (\alpha > 0) \quad (2.38)$$

当  $\alpha = 0$  时, 式 (2.38) 的被积表达式是没有定义的, 定义此时的散度  $d_0(g, f)$  为

$$d_0(g, f) = \lim_{\alpha \rightarrow 0} d_\alpha(g, f) = \int g(z) \log\{g(z)/f(z)\} dz$$

其中  $\alpha$  是用来控制密度幂估计量的效率和稳健性的平衡参数。特别地, 当调整参数  $\alpha = 0$  时, 上述 DPD 测度退化成 Kullback-Leibler (KL) 散度, 且有极限形式, 当  $\alpha \rightarrow 0^+$  时,  $d_0(g, f)$  为 KL 散度。

在基于 DPD 建模的方法中, 式 (2.38) 中的  $g$  为估计总体未知的真实密度函数,  $f$  为模型参数族中的  $f_\theta$ 。解决上述模型的思想就是通过最小化 (2.38) 给出的 DPD 函数  $d_\alpha(g, f_\theta)$  来获得参数空间  $\Theta$  上的最优参数  $\theta_D$ , 把这种方法称为最小化密度幂散度 (MDPD) 方法。因而有  $d_\alpha(g, f_{\theta_D}) = \min_{\theta \in \Theta} d_\alpha(g, f_\theta)$ , 并称利用 MDPD 方法得到的估计量  $\theta_D$  为最小化密度幂估计量 (MDPDE)。由于在传统的统计估计中, 最小化 KL 散度等价于最大化对数似然函数, 及最小化 KL 散度得到的估计量为最大似然估计量 (MLE)。因此, 此处的 MDPD 方法可以认为是最大似然估计的推广, 当调整参数  $\alpha = 0$  时, MDPDE 就退化成 MLE。或者可以将 MLE 认为时 MDPDE 在  $\alpha = 0$  时的特殊情况。

在公式 (2.38) 的基础上利用 MDPD 方法时, 将  $d_\alpha(g, f_\theta)$  写成如下的表达式,

$$d_\alpha(g, f_\theta) = \int f_\theta^{1+\alpha}(x)dx - \left(1 + \frac{1}{\alpha}\right) \int f_\theta^\alpha(x)dG(x) + K, \quad (2.39)$$

其中最后一项  $K$  与估计参数  $\theta$  无关。

Basu et al.<sup>[85]</sup> 给出了当观测数据为独立同分布时的 MDPD 估计。设  $X_1, \dots, X_n$  来自于真实分布  $G$  的独立同分布的观测值,  $G$  的密度函数为  $g$ 。在 (2.39) 的计算中忽略与  $\theta$  无关的项  $K$ , 用经验分布  $G_n(x)$  替代未知分布  $G(x)$ , 通过最小化如下的目标函数获得 MDPDE  $\hat{\theta}$

$$\begin{aligned} H_n(\theta) &= \int f_\theta^{1+\alpha}(x)dx - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n f_\theta^\alpha(X_i) \\ &= n^{-1} \sum_{i=1}^n V_\theta(X_i), \end{aligned} \quad (2.40)$$

关于  $\theta \in \Theta$ , 其中

$$V_\theta(X_i) = \int f_\theta^{1+\alpha}(x)dx - (1 + 1/\alpha)f_\theta^\alpha(X_i)$$

Basu et al.<sup>[133]</sup> 指出在此处应用经验分布函数能够避免对未知密度的估计和使用一些非参数的光滑方法。因为对于大多数基于密度的最小距离方法, 默认需要在创建距离时对数据进行平滑处理。这也是估计过程中大量额外复杂性的来源, 而且这种处理方式是不可避免的。密度幂散度是一种特殊的散度, 它可以基于经验分布函数求其估计, 不需要数据平滑。

在模型可微的条件下, 参数  $\theta$  的 MDPDE 通过解下述的估计方程获得

$$\frac{1}{n} \sum_{i=1}^n u_\theta(Y_i) f_\theta(Y_i)^\alpha - \int u_\theta(y) f_\theta(y)^{1+\alpha} dy = 0, \quad (2.41)$$

其中  $u_\theta(y) = \nabla \ln f_\theta(y)$  是似然得分函数,  $\nabla$  是关于  $\theta$  的梯度算子。从式 (2.41) 的第一项可以看出, 每个似然得分值都是由模型的密度幂的形式来加权的, 因此模型中离群观测值具有较小的密度函数值, 从而离群观测值被自动降低权重, 这也是称之为密度幂的原因, 且这种方法的稳健性是比较直观的。权重降低的程度取决于参数  $\alpha$  的大小, 随着  $\alpha$  的增加, 降权的幅度随之增大。当  $\alpha = 0$  时, 表明没有降权, 此时的 MDPDE 就是  $\theta$  的 MLE。式 (2.41) 中的第二项就是使估计方程在模型下无偏所需要的调整项。由于其高稳健性和最小的效率损失, 以及即使在连续模型下也不使用非参数核平滑等性质, 使得 MDPDE 受到了广泛的关注。同时, MDPDE 也可以看作是属于 Huber<sup>[108]</sup> 中 M 估计类。

当观测数据不是独立同分布时, 使用 DPD 方法进行稳健估计时, 需要对上述方法适当的扩展, Ghosh et al.<sup>[135]</sup> 给出了非齐次的观测数据的 MDPDE。假设观测数据  $Y_1, \dots, Y_n$  是相互独立的, 但是对任意的  $i, Y_i \sim g_i$ , 其中  $g_1, \dots, g_n$  可能是具有一些共同控制参数的不同的密度函数。例如  $g_i$  是具有不同均值  $\mu_i$  和共同方差  $\sigma^2$  的正态分布的密度序列, 或者是具有不同的形状参数  $\alpha_i$  和共同的尺度参数  $\beta$  的  $\gamma$  分布的密度序列。在逻辑回归中,  $g_1, \dots, g_n$  可以是具有不同的成功概率  $p_i$  的伯努利分布概率序列, 其中  $p_i$  可能通过一个共同的参数  $\beta$  与解释变量相关联。根据这一特点, 我们利用具有共同参数  $\theta$  的密度族  $\mathcal{F}_{i,\theta} = \{f_i(\cdot; \theta) \mid \theta \in \Theta\}$   $i = 1, 2, \dots$  对  $g_i$  建模。Ghosh et al.<sup>[135]</sup> 提出最小化数据点和模型之间的平均散度的方法来估计未知参数  $\theta$ ,

$$\frac{1}{n} \sum_{i=1}^n d_\alpha(\hat{g}_i, f_i(\cdot; \theta)), \quad (2.42)$$

其中  $\hat{g}_i$  是密度函数  $g$  的估计, 若忽略散度表达式中的常数项, 上述表达式 (2.42) 可以表示为:

$$H_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left[ \int f_i(x; \theta)^{1+\alpha} dx - \left(1 + \frac{1}{\alpha}\right) f_i(Y_i; \theta)^\alpha \right] = \frac{1}{n} \sum_{i=1}^n V_i(Y_i; \theta), \quad (2.43)$$

其中  $V_i(\cdot; \theta)$  上述表达式中方括号内的项。同样地, 对上述方程关于  $\theta$  求偏导数, 可得参数估计方程, 通过求解参数估计方程可获得 MDPDE。

下面将给出  $\gamma$  散度的定义。正如 DPD 的定义一样,  $\gamma$  散度也用于度量两个概率密度函数之间的差异, 针对来自于真实密度函数为  $g$  的独立同分布的观测样本  $X_1, X_2, \dots, X_n$ 。考虑用参数模型  $f_\theta(x) = f(x; \theta) (\theta \in \Theta)$  作为备选模型用于估计真实的分布, 其中  $\Theta \subset \mathbb{R}^p$  是对应的参数空间。Jones et al.<sup>[88]</sup> 首次提出

了一类新的密度幂散度, Fujisawa et al.<sup>[140]</sup> 将这种散度称之为  $\gamma$  散度, 主要用于解决模型中存在离群观测, 或者模型分布被污染时的稳健估计问题, 其具体的定义如下。对于两个密度函数定义

$$\begin{aligned} d_\gamma(g, f) &= -\frac{1}{\gamma} \log \int g(x)f(x)^\gamma dx + \frac{1}{1+\gamma} \log \int f(x)^{1+\gamma} dx \quad (\gamma > 0) \\ &= -\log \left[ \left\{ \int g(x)f(x)^\gamma dx \right\}^{1/\gamma} / \left\{ \int f(x)^{1+\gamma} dx \right\}^{1/(1+\gamma)} \right], \end{aligned} \quad (2.44)$$

称为  $\gamma$  交叉熵, 其中  $\gamma$  是调整参数。结合观测数据的经验分布,  $\gamma$  交叉熵  $d_\gamma(g, f_\theta)$  可以经验地表述为

$$d_\gamma(\bar{g}, f_\theta) = -\frac{1}{\gamma} \log \left\{ \frac{1}{n} \sum_{i=1}^n f(x_i; \theta)^\gamma \right\} + \frac{1}{1+\gamma} \log \int f(x; \theta)^{1+\gamma} dx. \quad (2.45)$$

其中  $\bar{g}$  是经验概率密度函数。由交叉熵定义的  $\gamma$  散度为

$$\begin{aligned} D_\gamma(g, f) &= -d_\gamma(g, g) + d_\gamma(g, f) \\ &= \frac{1}{\gamma(1+\gamma)} \log \int g^{1+\gamma} dx - \frac{1}{\gamma} \log \int gf^\gamma dx + \frac{1}{1+\gamma} \log \int f^{1+\gamma} dx \\ &= \log \left[ \left\{ \int g^{1+\gamma} dx \right\}^{1/\gamma(1+\gamma)} \left\{ \int f^{1+\gamma} dx \right\}^{1/(1+\gamma)} / \left\{ \int gf^\gamma dx \right\}^{1/\gamma} \right]. \end{aligned} \quad (2.46)$$

根据最小化距离散度的思想, 需要对上述  $\gamma$  散度关于  $\theta$  求最小化。因此最小化  $\gamma$  散度的目标函数为

$$\min_{\theta \in \Theta} d_\gamma(g, f_\theta) = \min_{\theta \in \Theta} \left\{ -\frac{1}{\gamma} \log \int gf_\theta^\gamma dx + \frac{1}{\gamma+1} \log \int f_\theta^{1+\gamma} dx \right\}. \quad (2.47)$$

在上述目标函数中,  $g$  用其经验分布函数  $\bar{g}$  替代即可。

Nakagawa et al.<sup>[90]</sup> 对上述  $\gamma$  散度做了变换, 考虑如下形式的  $\gamma$  交叉熵

$$\tilde{d}_\gamma(g, f_\theta) = -\frac{1}{\gamma} \{ \exp(-\gamma d_\gamma(g, f_\theta)) - 1 \} = -\frac{\frac{1}{\gamma} \int gf_\theta^\gamma dx}{\left\{ \int f_\theta^{1+\gamma} dx \right\} \gamma / (1+\gamma)} + \frac{1}{\gamma}. \quad (2.48)$$

由于式 (2.48) 中的函数  $-\exp(-x-1)$  是单调递增的, 因此最小化 (2.47) 和 (2.48) 具有相同的结果。

#### 2.4.2.2 密度幂散度族的性质

为了方便后续内容的推导计算, 在本小节讨论两种散度的基本性质。Basu et al.<sup>[85]</sup> 中研究了 DPD 散度的基本性质和 MDPDE 的渐近性质。其主要性质如下:

**性质 2.1** (Basu et al.<sup>[85]</sup>)  $d_\alpha(g, f)$  对任意的  $f, g$  均有  $d_\alpha(g, f) \neq 0$ , 且  $d_\alpha(g, f) = 0$  当且仅当  $f = g$  几乎处处成立。

**性质 2.2** (Basu et al.<sup>[85]</sup>) 在正则性条件下, 存在  $\hat{\theta}$ , 当  $n \rightarrow \infty$  时使得  $\hat{\theta}$  一致收敛于  $\theta$ , 且  $n^{\frac{1}{2}}(\hat{\theta} - \theta)$  趋于均值 0, 协方差矩阵为  $J^{-1}KJ^{-1}$  的渐近正态分布, 其中  $J = J(\theta)$ ,  $K = K(\theta)$  满足

$$J = \int u_\theta(z)u_\theta^\top(z)f_\theta^{1+\alpha}(z)dz + \int \{i_\theta(z) - \alpha u_\theta(z)u_\theta^\top(z)\} \{g(z) - f_\theta(z)\} f_\theta^\alpha(z)dz,$$

$$K = \int u_\theta(z)u_\theta^\top(z)f_\theta^{2\alpha}(z)g(z)dz - \xi\xi^\top,$$

其中  $\xi = \int u_\theta(z)f_\theta^\alpha(z)g(z)dz$ .

当观测数据是独立非同分布情形下, 根据前述的公式 (2.43) 可以求解其参数估计方程, 获得参数的稳健估计。其估计量具有如下的性质,

**性质 2.3** (Basu et al.<sup>[133]</sup>) 假设由最小化公式 (2.43) 得到的 MDPDE 为  $\hat{\theta}_n$ , 则在正则性条件下, 有  $\Omega_n^{-1/2}\Psi_n \left[ n^{1/2}(\hat{\theta}_n - \theta_\alpha) \right]$  渐近服从均值为 0, 协方差矩阵为  $I_p$  的正态分布, 其中  $\theta_\alpha$  是最小化  $(1/n)\sum_{i=1}^n d_\alpha(g_i, f_i(\cdot; \theta))$  得到的理论估计值,  $u_i(y; \theta) = \nabla \ln f_i(y; \theta)$ ,

$$\begin{aligned} \Psi_n &= \frac{1}{n} \sum_{i=1}^n J^{(i)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{1+\alpha} E_{g_i} [\nabla_{kl} V_i(Y_i; \theta_\alpha)] \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \int u_i(y; \theta_\alpha) u_i^\top(y; \theta_\alpha) f_i^{1+\alpha}(y; \theta_\alpha) dy \right. \\ &\quad \left. - \int \{ \nabla u_i(y; \theta_\alpha) + \alpha u_i(y; \theta_\alpha) u_i^\top(y; \theta_\alpha) \} \{ g_i(y) - f_i(y; \theta_\alpha) \} f_i(y; \theta_\alpha)^\alpha dy \right], \end{aligned}$$

和

$$\begin{aligned} \Omega_n &= \frac{1}{n} \sum_{i=1}^n \text{Var}_{g_i} [\nabla V_i(Y_i; \theta_\alpha)] \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \int u_i(y; \theta_\alpha) u_i^\top(y; \theta_\alpha) f_i(y; \theta_\alpha)^{2\alpha} g_i(y) dy - \xi_i \xi_i^\top \right], \end{aligned}$$

其中

$$\xi_i = \int u_i(y; \theta_\alpha) f_i(y; \theta_\alpha)^\alpha g_i(y) dy.$$

结合 Fujisawa et al.<sup>[89]</sup>, Jones et al.<sup>[141]</sup>, Nakagawa et al.<sup>[142]</sup> 中有关  $\gamma$  散度的定义, 给出有关  $\gamma$  散度的性质。

**性质 2.4** (Fujisawa et al.<sup>[89]</sup>) 设  $g$  和  $f$  是两个正的函数,  $k_1, k_2$ , 和  $k$  是正的常数, 则  $\gamma$  交叉熵  $d_\gamma(g, f)$  和散度  $D_\gamma(g, f)$  具有如下的性质:

1.  $d_\gamma(k_1g, k_2f) = d_\gamma(g, f) - (1/\gamma) \log k_1$ .
2.  $D_\gamma(k_1g, k_2f) = D_\gamma(g, f)$ .
3.  $D_\gamma(g, f) \geq 0$ . 该等号成立当且仅当  $g = kf$ ; 特别地,  $g = f$  当  $g$  和  $f$  是概率密度函数时.
4.  $\lim_{\gamma \rightarrow 0} D_\gamma(g, f) = KL(\tilde{g}, \tilde{f}) = \int \tilde{g} \log(\tilde{g}/\tilde{f}) dx$ , 其中  $\tilde{g} = g / \int g dx$  和  $\tilde{f} = f / \int f dx$ .

在利用  $\gamma$  散度进行稳健估计时, 假设  $f(x)$  未知的真实密度函数,  $\delta(x)$  是与离群观测相关的污染概率密度函数。而观测数据是来源于被污染的分布, 假设其密度函数  $g(x)$  为  $g(x) = (1 - \varepsilon)f(x) + \varepsilon\delta(x)$ , 其中  $\varepsilon$  是污染比例。设  $f_\theta(x)$  (或  $f(x; \theta)$ ) 是含有未知参数的概率密度函数  $\hat{\theta}$  是基于来自于分布  $g$  的观测  $x_1, \dots, x_n$  得到的参数  $\theta$  的估计值。相比于一般的参数估计方法, 直观的目的就是使选取参数使得  $f_{\hat{\theta}}$  尽可能的接近样本来源分布  $g$ , 然而  $g$  并非是真实的分布密度,  $f$  才是真正的密度函数。而基于  $\gamma$  散度的稳健方法就是选取适当的参数分布  $f_{\hat{\theta}}$ , 使其充分靠近密度  $f$ , 即使在  $\varepsilon > 0$  也要成立。

在使用密度幂类型的稳健估计方法中, 假设  $x^*$  是离群值, 则假设密度值  $f(x^*)$  是充分小的。并假设:

$$v_f = \left\{ \int \delta(x)f(x)^{\gamma_0} dx \right\}^{1/\gamma_0} \text{ 是充分小的, 对适当的 } \gamma_0 > 0. \quad (2.49)$$

这个假设意味着污染密度  $\delta(x)$  主要影响真实密度  $f(x)$  的尾部。通常, 假设  $\gamma_0 = 1$ 。如果  $\delta(x)$  是离群值  $x^*$  处的狄拉克函数, 则公式 (2.49) 与传统的假设相同, 因为  $v_f = f(x^*)$ 。在基于密度幂的稳健方法中, 从不假设  $\varepsilon$  足够小, 换句话说, 关注的是严重污染的情况。

**性质 2.5** (Fujisawa et al.<sup>[89]</sup>) 假设正的函数  $h$  满足形如 (2.49) 的条件, 即将 (2.49) 中的函数  $f$  换成  $h$  时成立。设  $v = \max\{v_f, v_h\}$ , 那么  $g, f$ , 和  $h$  之间有近似的勾股定理关系式:

$$\Delta(g, f, h) = D_\gamma(g, h) - D_\gamma(g, f) - D_\gamma(f, h) = O(\varepsilon v^\gamma).$$

Fujisawa et al.<sup>[89]</sup> 指出上述勾股定理关系暗示了一个投影结构 (如图 2.1)。假设函数  $h$  与原始密度  $f$  具有相同的性质 (函数  $h$  满足相同的假设 (2.49))。那么,  $h$  与污染密度  $g$  的散度的最小值近似等于  $h$  与原始密度  $f$  的散度。如果我们把  $h$  看成  $f_\theta$ , 那么我们会明白为什么提出的稳健参数估计方法是有效的。

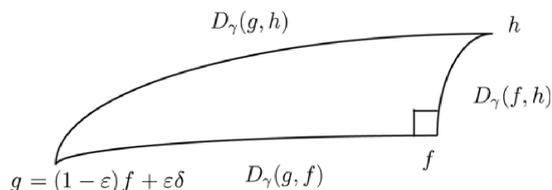


图 2.1  $g, f, h$  之间的勾股定理关系

上述关系式中的误差项是  $O(\epsilon v^\gamma)$ , 那么当  $\gamma$  相比  $v$  较小时,  $v^\gamma$  不会太小, 此时的稳健效果较差, 应该选取较大的  $\gamma$  使上述关系近似成立。当污染分布距离真实的分布较远时, 此时  $v$  会变得较小, 此时只需要较小的  $\gamma$  值也能达到稳健估计的目的。

假设  $\theta^*$  和  $\theta_\gamma^*$  是最小化  $d_\gamma(f, f_\theta)$  和  $d_\gamma(g, f_\theta)$  得到的估计参数。同时在正则性条件下, 假设对于足够小的  $v_\omega$  限制其参数空间为

$$\Omega_{v_\omega} = \{\theta : v_{f_\theta} \leq v_\omega\} \ni \theta^*, \theta_\gamma^*.$$

接下来, 假设  $\theta \in \Omega_{v_\omega}$  中, 这意味着参数密度  $f_\theta$  与潜在密度  $f$  具有相同的性质, 即污染密度  $\delta$  也位于  $f_\theta$  的尾部。  $d_{\theta_\gamma}$  是  $d_\gamma(\bar{g}, f_\theta)$  的最小值的新定义, 也就是说,  $d_\gamma(\bar{g}, f_\theta)$  的一个合适的局部最小值。在这些假设条件下,

$$\begin{aligned} \theta_\gamma^* &= \arg \min_{\theta} d_\gamma(g, f_\theta) \\ &= \arg \min_{\theta} \{d_\gamma(f, f_\theta) - (1/\gamma) \log(1 - \epsilon) + O(\epsilon v_{f_\theta}^\gamma)\} \\ &= \arg \min_{\theta} \{d_\gamma(f, f_\theta) - (1/\gamma) \log(1 - \epsilon) + O(\epsilon v_\omega^\gamma)\} \\ &= \theta^* + O(\epsilon v_\omega^\gamma). \end{aligned}$$

这一结果说明了基于  $\gamma$  交叉熵的方法的有效性。此外,  $g, f$  和  $f_\theta$  之间的毕达哥拉斯关系近似成立。

**性质 2.6** 假设  $\theta \in \Omega_{v_\omega}$ , 那么  $\theta_\gamma^* = \theta^* + O(\epsilon v_\omega^\gamma)$  和  $\Delta(g, f, f_\theta) = O(\epsilon v^\gamma)$  成立, 其中  $v = \max\{v_f, v_\omega\}$ .

不同于 KL 散度,  $\gamma$  散度具有关于变量的平移不变性。设  $D_\gamma^x(g, h)$  是关于随机变量  $x$  的两个密度之间的  $\gamma$  散度。对随机变量  $x$  进行平移变换  $y = Ax + b$ , 记  $D_\gamma^y(g, h)$  是关于随机变量  $y$  的两个密度之间的  $\gamma$  散度, 则这两个散度是相同的。

**性质 2.7 (平移不变性)** 假设  $D_\gamma^x(g(x), h(x))$  和  $D_\gamma^y(g(y), h(y))$  分别表示关于随机变量  $x$  和平移后随机变量  $y$  的密度函数之间的  $\gamma$  散度, 则

$$D_\gamma^x(g(x), h(x)) = D_\gamma^y(g(y), h(y)).$$

### 第 3 章 密度幂散度在区域水平模型下的稳健估计

在第 2 章介绍了 FH 模型及其相关的估计方法, 详细介绍了两类密度幂散度 (DPD), 及其相关的性质。本章主要介绍利用最小化密度发散方法 (MDPD) 对区域水平模型进行稳健估计。提出经验贝叶斯估计量, 推导估计量的 MSE, 并给出 MSE 的估计。并通过模拟实验和实际数据检验提出估计量的稳健性, 和常用的几类稳健估计方法进行比较。

设  $\theta_i$  和  $y_i$  是目标变量在区域  $i (i = 1, \dots, m)$  的真实值和观测值, 其中  $m$  是区域个数, 一个基础的两阶段 FH 模型为

$$\begin{aligned} y_i &= \theta_i + e_i; \quad \theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i \quad (i = 1, \dots, m), \\ v_i &\sim N(0, A); \quad e_i \sim N(0, D_i). \end{aligned} \quad (3.1)$$

其中  $\mathbf{x}_i$  是辅助信息矩阵,  $\boldsymbol{\beta}$  是未知的回归系数,  $e_i$  为估计量  $y_i$  的抽样误差,  $v_i$  为模型误差。假设  $e_i \sim N(0, D_i)$  是相互独立并且服从正态分布  $e_i \sim N(0, D_i)$ , 并且  $v_1, \dots, v_m$  与  $e_i \sim N(0, D_i)$  相互独立, 服从正态分布  $N(0, A)$ 。一般假定抽样方差  $D_i$  是已知的, 而  $A, \boldsymbol{\beta}$  均是未知的。在实践中, 抽样方差  $D_i$  未必能够精确获得, 但是可以通过抽样设计方法估计得到并具有较高的准确性<sup>[63]</sup>。

在参数估计过程中, 令  $\boldsymbol{\phi} = (\boldsymbol{\beta}^T, A)^T$  表示 (3.1) 中未知的参数向量。由于在上述模型的假定下  $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, A + D_i)$ , 那么对数似然函数可以表示为

$$\log f(y; \boldsymbol{\phi}) = -\frac{m}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^m \log(A + D_i) - \frac{1}{2} \sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{A + D_i}. \quad (3.2)$$

在样本观测数据  $y = (y_1, \dots, y_m)^T$  的基础上, 未知参数  $\boldsymbol{\phi}$  可以通过最大化 (3.2) 式获取其估计值。在平方损失函数的条件下  $\theta_i$  的贝叶斯预测量可以表示为:

$$\tilde{\theta}_i(y_i; \boldsymbol{\phi}) = y_i - \frac{D_i}{A + D_i} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) = (1 - \rho_i) y_i + \rho_i \mathbf{x}_i^T \boldsymbol{\beta}, \quad (3.3)$$

其中  $\rho_i = \frac{D_i}{A + D_i}$ ,  $\theta_i$  的经验贝叶斯预测量为  $\hat{\theta}_i = \tilde{\theta}_i(y_i; \hat{\boldsymbol{\phi}}_\alpha)$ 。

表达式 (3.3) 表明, EB 估计量是直接估计量  $y_i$  和间接回归估计量  $\mathbf{x}_i^T \boldsymbol{\beta}$  的加权平均值, 权重分别为  $1 - \rho_i$  和  $\rho_i$ 。另外, EB 估计量可以被看作是用收缩因子  $\rho_i$  将直接估计量缩小到回归估计量。然而, 当在一些小范围内存在离群观测时, EB 估计量有两个缺点。第一, 由于  $A$  的不确定性,  $\rho_i$  的值可能趋向于 0

或 1, 这意味着出现了过度收缩或收缩不足; 其次, 离群观测会导致模型正态性假设不满足, 使参数  $\hat{\phi}$  的估计会有较大的标准误差。有关更详细的讨论, 请参阅 [63]。本章重点考虑观测数据中存在离群观测的时, 传统的 EB 估计量表现不佳时, 如何利用密度幂散度来改善估计效果, 并提出  $\theta_i$  的稳健估计。

### 3.1 基于密度幂散度的 FH 模型的稳健估计方法

Sugasawa<sup>[77]</sup> 在研究形如 (2.1) 的区域水平模型的过程中, 结合 Efron<sup>[143]</sup> 发现针对区域水平模型的 BLUP 估计量 (2.13) 可以表示为:

$$\tilde{\theta}_i(y_i; \phi) = y_i + D_i \frac{\partial}{\partial y_i} \log f(y_i; \phi), \quad (3.4)$$

其中  $\tilde{y}_i$  表示区域  $i$  上目标变量的均值估计,  $\phi$  为未知的模型参数,  $\log f(y_i; \phi)$  表示目标变量的对数似然函数。

在第 2.4 章讨论过, 极大似然估计方法可以看成是最小化 KL 散度得到, 而 KL 散度又是密度幂散度的特殊情形。因此, 一个自然的想法就是将 KL 散度进行推广, 即考虑最小化密度幂散度的方法。基于公式 (3.4), Sugasawa<sup>[77]</sup> 考虑了结合密度幂散度提出的似然函数:

$$L_\alpha(y; \phi) = \frac{1}{\alpha} \sum_{i=1}^m f_i(y_i; \phi)^\alpha - \frac{1}{1+\alpha} \sum_{i=1}^m \int f_i(t; \phi)^{1+\alpha} dt, \quad \alpha \in (0, 1) \quad (3.5)$$

$L_\alpha(y; \phi)$  作为  $\log f(y_i; \phi)$  的推广, 直接将 (3.4) 中的  $\log f(y_i; \phi)$  进行推广, 从而得到了基于 DPD 的区域水平模型的小域稳健估计。

本研究中, 我们考虑从 DPD 的定义出发, 给出基于 DPD 的似然函数, 结合 MDPD 方法先给出模型参数的稳健估计量, 再进一步推导区域均值的稳健估计量。

### 3.2 FH 模型的参数与区域均值的稳健估计

根据 (2.40) 关于 MDPD 方法的定义, 其对应的目标函数为

$$H(\phi) = \frac{1}{m} \sum_{i=1}^m \int f_\theta^{1+\alpha}(y_i) dy_i - \left(1 + \frac{1}{\alpha}\right) m^{-1} \sum_{i=1}^m f_\theta^\alpha(y_i). \quad (3.6)$$

在上述目标函数中,  $f_\theta(y_i)$  表示观测变量  $y_i$  的密度函数。在单元水平模型中, 根据 (3.1) 中关于模型描述,  $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, A + D_i)$ , 则其密度函数为

$$f(y_i) = (2\pi)^{-\frac{1}{2}} (A + D_i)^{-\frac{1}{2}} \exp\left(-\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2(A + D_i)}\right). \quad (3.7)$$

将密度函数 (3.7) 代入目标函数 (3.6) 可以将目标函数化简成为:

$$H(\phi) = \frac{1}{m} \sum_{i=1}^m \left( V_i^{-\frac{\alpha}{2}} (1 + \alpha)^{-\frac{1}{2}} - \frac{1 + \alpha}{\alpha} V_i^{-\frac{\alpha}{2}} \exp \left( -\frac{\alpha(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2(A + D_i)} \right) \right). \quad (3.8)$$

其中  $V_i = 2\pi(A + D_i)$ 。在上述目标函数中, 当  $\alpha \rightarrow 0$  时, 有如下结论:

$$\lim_{\alpha \rightarrow 0} H(\phi) = \log f(y; \phi), \quad (3.9)$$

其中  $\log f(y; \phi)$  为目标变量的对数似然函数, 该公式说明这里的目标函数可以作为对数似然函数的推广, 进一步表明基于密度幂散度的方法是最大似然方法的推广, 且通过调整参数  $\alpha$  来控制稳健性和有效性之间的平衡, 当取  $\alpha$  充分接近于 0 时, 其估计效果与极大似然估计的结果相同。同时, 上述模型中的未知参数可以通过最小化上述目标函数迭代求得。

模型中未知参数的估计通过 (3.8) 求得, 对其关于  $A, \boldsymbol{\beta}$  求偏导数, 令其为零可得参数估计方程。

$$\begin{aligned} \frac{\partial H(\phi)}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^m (1 + \alpha) s_i(y_i; \phi) \frac{\mathbf{x}_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{A + D_i} = 0, \\ 2 \frac{\partial H(\phi)}{\partial A} &= \sum_{i=1}^m \left\{ \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 s_i(y_i; \phi)}{(A + D_i)^2} - \frac{s_i(y_i; \phi)}{A + D_i} + \frac{\alpha V_i^{-\alpha/2}}{(\alpha + 1)^{3/2} (A + D_i)} \right\} = 0. \end{aligned} \quad (3.10)$$

其中  $V_i = 2\pi(A + D_i)$ ,  $s_i(y_i; \phi) = V_i^{-\frac{\alpha}{2}} \exp \left( -\frac{\alpha(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2(A + D_i)} \right)$ 。利用牛顿拉夫逊 (Newton-Raphson) 算法可求解上述估计方程的解, 从而得到未知参数的估计值。在利用 N-R 算法的过程中, 可以将最大似然估计的结果作为初始值进行迭代。

结合公式 (3.4) 和 (3.9), 用  $H(\phi)$  替代目标变量的对数似然函数, 构建基于 DPD 方法的稳健小域估计量:

$$\hat{\theta}_i(y_i; \phi) = y_i + D_i \frac{\partial}{\partial y_i} H(\phi). \quad (3.11)$$

因

$$\frac{\partial}{\partial y_i} H(\phi) = \frac{\partial}{\partial y_i} \frac{(1 + \alpha) s_i(y; \phi)}{\alpha} = \frac{1 + \alpha}{A + D_i} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) s_i(y_i; \phi),$$

将其代入 (3.11) 中可得 MDPD 估计量

$$\hat{\theta}_i = y_i - \frac{D_i}{A + D_i} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) (1 + \alpha) s_i(y_i; \phi). \quad (3.12)$$

根据表达式, (3.12) 中估计量的收缩因子为  $(1 + \alpha) s_i(y_i; \phi) D_i / (A + D_i)$ 。与 (3.3) 中的经验贝叶斯估计量相比较, 不难发现, (3.3) 中的收缩因子与观测变

量  $y_i$  是无关系的, 而 (3.12) 中的收缩因子中包含观测变量  $y_i$ 。当  $y_i$  为离群观测时, 经验贝叶斯估计量容易受其影响, 而基于 MDPD 方法的估计量收缩因子会减小, 当  $y_i$  异常时, 会控制其向回归估计部分靠近而减小异常值带来的影响。另外, 当  $\alpha \rightarrow 0$  时,  $(1 + \alpha)s_i(y_i, \phi) \rightarrow 0$ , 这说明  $\hat{y}_i$  退化为  $\tilde{\theta}_i$  当  $\alpha \rightarrow 0$  时。此外, 当  $D_i \rightarrow 0$  时, 稳健贝叶斯预测量  $\hat{y}_i$  退化传统的直接估计量。

在运用公式 (3.12) 进行小域稳健估计时, 由于表达式中含有未知参数  $\phi = (A, \beta)$ , 因此需要先考虑未知参数的估计。根据参数估计方程 (3.10) 求得参数的估计值为  $\hat{\phi} = (\hat{\beta}^T, A)^T$ , 则利用 MDPD 方法得到的小域稳健估计量为:

$$\hat{\theta}_i = y_i - \frac{D_i}{\hat{A} + D_i} \left( y_i - \mathbf{x}_i^T \hat{\beta} \right) (1 + \alpha) s_i \left( y_i; \hat{\phi} \right). \quad (3.13)$$

### 3.3 调整参数的选取

稳健估计量 (3.13) 中包含未知的调整参数  $\alpha$ , 而  $\alpha$  的选取又与估计量的稳健性相关, 因此选取适当的调整参数也至关重要。结合文献 [27, 77] 中的思想, 选取使得稳健贝叶斯估计量 (3.12) 的均方误差最小的参数  $\alpha$ , 这样选取的目的是在保证稳健性的基础上, 尽可能使估计量具有更小的均方误差, 提高估计效率。稳健估计量的均方误差公式由下定理给出。

**定理 3.1** 在模型 (3.1) 的假设条件下, 稳健估计量的 MSE 为

$$E \left\{ \left( \hat{\theta}_i - \theta_i \right)^2 \right\} = g_{1i}(A) + g_{2i}(A),$$

其中

$$g_{1i}(A) = \frac{AD_i}{A + D_i}, \quad g_{2i}(A) = \frac{D_i^2}{A + D_i} \left\{ \frac{V_i^{2\alpha}}{(2\alpha + 1)^{3/2}} - \frac{2V_i^\alpha}{(\alpha + 1)^{3/2}} + 1 \right\}, \quad (3.14)$$

且  $g_{2i}(A)$  关于  $\alpha$  单调递增。

定理的证明可参见本章附录。从 Rao et al.<sup>[1]</sup> 中的结论可知, 使用最大似然估计方法得到的经验贝叶斯估计量 (3.3) 的 MSE 为  $g_{1i}(A)$ 。结合上述定理, 基于 MDPD 方法提出的估计量  $\hat{\theta}_i$  的 MSE(3.14) 比极大似然估计量的 MSE 大了  $g_{2i}(A)$ 。这里的  $g_{2i}(A)$  关于  $\alpha$  单调递增, 且当  $\alpha \rightarrow 0$  时  $g_{2i}(A) \rightarrow 0$ 。当  $\alpha$  增大时, 小域稳健估计量  $\hat{\theta}_i$  具有更强的稳健性, 但其估计量的 MSE 也随之增加, 即调整参数  $\alpha$  决定着稳健性和估计效率之间的平衡。

借助与 Sugawara<sup>[77]</sup> 的思想, 定义相对超量 MSE 的百分比为  $\text{Exc}(\alpha) = 100 \times \sum_{i=1}^m g_{2i}(\hat{A}_\alpha) / \sum_{i=1}^m g_{1i}(\hat{A}_\alpha)$ , 其中  $\hat{A}_\alpha$  是在固定  $\alpha$  时根据 (3.10) 得到

的  $A$  的估计值。在选取调整参数  $\alpha$  时, 根据需求, 先确定最大的相对百分比  $c\%$  使得  $\text{Exc}(\alpha)$  不超过这一界限。根据定理 3.1 的结论,  $\text{Exc}(\alpha)$  关于  $\alpha$  是单调递增的, 因此这样的  $\alpha$  是唯一存在的。最优调整参数  $\alpha$  的选取通过如下的算法实现。

---

#### 算法 1 参数 $\alpha$ 的选择算法

---

**输入:** 给定调整参数、模型参数的初值  $\alpha^{(0)}$ ,  $\phi^{(0)} = (\beta^{(0)T}, A^{(0)})$ , 并给出误差相对百分比的上限  $c$ 。

**重复:**

1: 根据公式 (3.14) 和  $\text{Exc}(\alpha)$  的定义, 利用二分法计算  $\text{Exc}(\alpha) = c$  的根, 并对  $\alpha$  进行迭代, 即

$$\alpha = \alpha^{(i+1)} \text{使得 } \text{Exc}(\alpha^{(i+1)}) = c.$$

2: 固定  $\alpha$ , 根据估计方程 (3.10), 利用 Newton-Raphson 算法估计得到参数  $\phi^{(i+1)}$ 。

3: 重复上述步骤 1 和 2, 直至  $\alpha$  收敛。即对于充分小的临界值  $\epsilon$  满足  $|\alpha^{(i+1)} - \alpha^{(i)}| < \epsilon$ 。

**输出:**  $\alpha^{(i+1)}$

---

### 3.4 稳健估计量的性质

为了考虑模型 (3.1) 下稳健估计的渐近性质, 给出以下的正则性假设条件:

1. 假设  $0 < D_* \leq \min_{1 \leq i \leq m} D_i \leq \max_{1 \leq i \leq m} D_i \leq D^* < \infty$  成立, 其中  $D_*$  和  $D^*$  与  $m$  无关。
2. 设  $\max_{1 \leq i \leq m} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i = O(m^{-1})$ , 其中  $X = (x_1, \dots, x_m)^T$ 。
3. 当  $m \rightarrow \infty$  时,  $X^T X / m$  收敛于一个正定矩阵。

上述正则条件中, 第一条为关于区域随机误差的约束, 这里  $D_i$  是区域随机效应的方差, 在小域估计中, 这一方差可通过抽样设计方法获得其直接估计, 因此在这里假定其是有界的。第二条和第三条是关于辅助信息矩阵的约束, 这两条保证获取的辅助信息之间是不重复的, 即获得的辅助信息矩阵是非奇异矩阵, 进一步保证模型系数估计的准确性。上述正则假设是小域估计中的最基本假设, 在通常的小域估计中均能够满足, 不会因这些假设而影响小域估计方法的应用。

在上述假设条件下, 先给出由 (3.10) 得到的估计参数的性质, 如下述定理。

**定理 3.2** 在上述假设条件下, 由 (3.10) 得到的估计参数  $\hat{\beta}_\alpha$  和  $\hat{A}_\alpha$  是相互独立的且渐近服从正态分布  $N(\beta, m^{-1}J_\beta^{-1}K_\beta J_\beta^{-1})$  和  $N(A, K_A/mJ_A^2)$ , 其中

$$J_\beta = \frac{1}{m(\alpha+1)^{3/2}} \sum_{i=1}^m \frac{V_i^\alpha \mathbf{x}_i \mathbf{x}_i^\top}{A+D_i}, \quad J_A = \frac{1}{2m} \sum_{i=1}^m \frac{V_i^\alpha (\alpha^2+2)}{(A+D_i)^2 (\alpha+1)^{5/2}},$$

$$K_\beta = \frac{1}{m(2\alpha+1)^{3/2}} \sum_{i=1}^m \frac{V_i^{2\alpha} \mathbf{x}_i \mathbf{x}_i^\top}{A+D_i}, \quad K_A = \frac{1}{m} \sum_{i=1}^m \frac{V_i^{2\alpha}}{(A+D_i)^2} \left\{ \frac{2(2\alpha^2+1)}{(2\alpha+1)^{5/2}} - \frac{\alpha^2}{(\alpha+1)^3} \right\}.$$

定理的详细证明可参见本章的附录。

**注 3.1** 根据 (3.9) 式, 当  $\alpha \rightarrow 0$  时,  $H(\phi)$  收敛于对数似然函数。因而当  $\alpha = 0$  时,  $\hat{\beta}_\alpha$  的渐近协方差矩阵和  $\hat{A}_\alpha$  的渐近方差与在极大似然估计方法下的渐近方差相同 [99]。

为了研究小域稳健估计量  $\hat{\theta}_i$  的估计效果, 接下来考虑估计量的 MSE  $M_i = E \left\{ \left( \hat{\theta}_i^R - \theta_i \right)^2 \right\}$ 。Rao et al. [1] 指出 MSE 可以看作是综合贝叶斯风险, 是小域估计中风险的标准度量。

由于  $\hat{\theta}_i$  依赖于估计值  $\hat{\phi}_\alpha$ , 因此在考虑  $\hat{\theta}_i$  的 MSE 时, 不仅包含 (3.14) 中的 MSE,  $M_i$  还要考虑由于  $\hat{\phi}_\alpha$  而产生的额外变化。用解析方法求解  $M_i$  是很困难的, 因此我们采用了  $M_i$  的二阶近似。其近似表达如下述定理所示。

**定理 3.3** 在假设条件 1-3 下, 稳健估计量  $\hat{\theta}_i$  的 MSE 为

$$M_i = g_{1i}(A) + g_{2i}(A) + \frac{g_{3i}(A)}{m} + \frac{g_{4i}(A)}{m} + \frac{2g_{5i}(A)}{m} + o(m^{-1}), \quad (3.15)$$

其中  $g_{1i}(A)$  和  $g_{2i}(A)$  如定理 3.1 所示, 另

$$g_{3i}(A) = \frac{D_i^2 V_i^{2\alpha}}{B_i^2 (2\alpha+1)^{3/2}} \mathbf{x}_i^\top J_\beta^{-1} K_\beta J_\beta^{-1} \mathbf{x}_i, \quad g_{4i}(A) = \frac{D_i^2 V_i^{2\alpha} K_A}{B_i^3 (2\alpha+1)^{7/2} J_A^2} \left( \alpha^4 - \frac{1}{2} \alpha^2 + 1 \right),$$

$$g_{5i}(A) = \frac{\alpha D_i^2 \mathbf{x}_i^\top J_\beta^{-1} K_\beta J_\beta^{-1} \mathbf{x}_i}{2B_i^4} (3B_i C_{11} - \alpha C_{21})$$

$$+ \frac{D_i^2 K_A}{24B_i^6 J_A^2} \{ 3\alpha B_i^2 C_{21} + (\alpha-2)(3\alpha+8)C_{11} \}$$

$$+ \frac{D_i^2 \mathbf{x}_i^\top J_\beta^{-1} \mathbf{x}_i}{B_i^4} (B_i C_{12} - \alpha C_{22}) + \frac{D_i^2}{2B_i^6 J_A} \{ \alpha C_{32} - 2B_i C_{22} + (2-\alpha)B_i^2 C_{12} \}$$

$$+ \frac{D_i^2}{2B_i^4} \left\{ b_A - \frac{\alpha V_i^\alpha}{(\alpha+1)^{3/2} B_i J_A} \right\} \{ (2-\alpha)B_i C_{11} - \alpha C_{21} \},$$

这里,  $B_i = A + D_i$ ,  $b_A = \lim_{m \rightarrow \infty} m E \left( \hat{A}_\alpha - A \right)$  是  $\hat{A}_\alpha$  的一阶偏差, 且

$$C_{jk} = (2j-1)!! B_i^j \left\{ V_i^{k\alpha} (k\alpha+1)^{-j-1/2} - V_i^{k\alpha+\alpha} (k\alpha+\alpha+1)^{-j-1/2} \right\},$$

其中  $(2j-1)!! = (2j-1)(2j-3)\cdots(1)$ 。

定理的证明步骤将在本章的附录中给出。这里需要说明, (3.15) 中的近似表达式是基于已知的调整参数  $\alpha$  而得到的, 因需要对  $\alpha$  进行估计, 因此可能会和 (3.15) 有所不同。同时, 当  $\alpha = 0$  时,  $C_{jk} = 0$  和  $g_{5i}(A) = 0$  成立, 此时的 MSE 等同于<sup>[99]</sup> 中的经典贝叶斯估计的 MSE。

由于定理 3 中给出的均方误差的近似值依赖于未知参数  $A$ , 因此无法在实际中使用。这里同样使用 Sugasawa<sup>[77]</sup> 中提到的利用二阶无偏近似上述定理中的 MSE, 且用 Bootstrap 方法对其 MSE 进行估计。其具体估计形式为

$$\hat{M}_i = 2g_{12i}(\hat{A}_\alpha) - \frac{1}{B} \sum_{b=1}^B g_{12i}(\hat{A}_\alpha^{(b)}) + \frac{g_{3i}(\hat{A}_\alpha)}{m} + \frac{g_{4i}(\hat{A}_\alpha)}{m} + \frac{2g_{5i}(\hat{A}_\alpha)}{m}. \quad (3.16)$$

其中  $g_{12i}(A) = g_{1i}(A) + g_{2i}(A)$ ,  $\hat{A}_\alpha^{(b)}$  是基于 Bootstrap 样本的 Bootstrap 估计量, Bootstrap 样本  $y_1^{(b)}, \dots, y_m^{(b)}$  根据模型的定义由下述表达式生成。

$$y_i^{(b)} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\alpha + v_i^{(b)} + \varepsilon_i^{(b)}, \quad v_i^{(b)} \sim N(0, \hat{A}_\alpha), \quad \varepsilon_i^{(b)} \sim N(0, D_i)$$

### 3.5 数值模拟

为了更好地拟合观测数据中出现离群值这一特征, 根据文献 [1, 27, 28] 中关于具有离群观测数据模拟的方法, 介绍如下的定义, 并在后续的章节中我们将已知沿用这种定义。

一般地, 对于正态分布而言, 称形如  $(1 - \eta)N(\mu_1, \sigma^2) + \eta N(\mu_2, \sigma_1^2)$  的分布为混合正态分布, 该分布用于刻画具有厚尾分布特征的数据。在本文中, 用混合正态分布  $(1 - \eta)N(0, \sigma^2) + \eta N(0, \sigma_1^2)$  来模拟正态性假定不满足、具有离群观测的情况, 其中  $N(0, \sigma^2)$  为真实分布,  $N(0, \sigma_1^2)$  称为污染分布,  $\eta$  称为污染比例。一般情况下  $\sigma_1^2 > \sigma^2$ , 当污染比例  $\eta = 0$  时, 表示真实分布未被污染。在本节中, 讨论模型误差和随机误差均是混合正态分布时的小域估计问题。

在数据模拟部分, 考虑如下的区域水平模型, 并在其基础上讨论几类稳健估计方法的效果。设生成数据的 FH 模型如下:

$$y_i = \theta_i + \varepsilon_i, \quad \theta_i = \beta_0 + \beta_1 \mathbf{x}_i + A^{1/2} u_i \quad (i = 1, \dots, m),$$

其中  $m = 50, \beta_0 = 0, \beta_1 = 2, A = 0.5$ , 抽样误差  $\varepsilon_i \sim N(0, D_i)$ 。辅助变量  $\mathbf{x}_i$  由  $(0, 1)$  上的均匀分布生成。并将  $m$  个区域平均划分成 5 个组, 每个组内的抽样误差具有相同的方差  $D_i$ , 且 5 个组抽样误差的方差分别为  $(0.2, 0.4, 0.6, 0.8, 1.0)$ 。

模型随机误差  $u_i$  由混合正态分布  $u_i \sim (1 - \xi)N(0, 1) + \xi N(0, 10^2)$  生成, 其中  $\xi$  是污染比例, 用于确定假设分布的污染程度。

考虑三种情形下的小域估计效果:

情形 (I):  $\xi = 0$ ;

情形 (II):  $\xi = 0.15$ ;

情形 (III):  $\xi = 0.30$ .

这三种情形中, 情形 (I) 表示观测数据完全服从标准 FH 模型假设的正态分布, 不存在离群观测, 在情形 (II) 和 (III) 中, 表明观测数据收到一定程度的污染, 会出现离群观测, 且情形 (III) 更为突出。本文将重点比较在这几类情形下常见的稳健估计方法和提出的稳健估计方法之间的效率, 观察几类方法的优劣性。

在估计小域目标参数  $\theta_i$  时, 分别采用了提出的 MDPD 稳健估计量  $\theta_i^D$ 、传统的经验贝叶斯估计量 (EBLUP), Ghosh et al.<sup>[27]</sup> 提出的稳健估计量, Sinha et al.<sup>[28]</sup> 提出的稳健估计量。在基于密度幂的方法中, 选取调整参数  $\alpha$  时, 设定相对误差占比的上限  $c = 5, 10$  两种情形。<sup>[28]</sup> 中的稳健方法都涉及到了 Huber  $\phi$  函数, 设定其调整参数  $K = 1.345$ 。<sup>[27]</sup> 中稳健调整参数的选取同样根据相对误差比例来选取适当的调整参数  $K$ , 此处同样选取相对误差比例为  $c = 5, 10$  两种情形。

首先比较上述几种估计方法对模型参数的估计效果, 重复 1000 次上述模拟实验, 计算估计系数  $\beta$  和模型误差方差  $A$  的平均估计偏差和 MSE。这里重点比较提出的基于密度幂的估计方法 (DEB)、传统的经验贝叶斯估计方法 (EB)、Sinha et al.<sup>[28]</sup> 提出的稳健估计方法 (REB) 和 Ghosh et al.<sup>[27]</sup> 提出的稳健估计方法 (GEB) 之间的估计效果。估计结果参见表3.1所示, 表中给出了三种污染情形下模型参数的平均估计偏差, 括号内的数是估计的 MSE。表3.1中 DEB1 和 DEB2 分别表示  $c = 5, 10$  时的 MDPD 方法。

从表3.1中的结果可得出如下的结论。在情形 I 中, 观测数据没有被污染, 此时不存在离群观测值, FH 模型的基本假定成立。因此 EB 估计方法应该是最优的, 这一点从表中数据得到了进一步的验证。从表中还可以得出, 由于 GEB 方法在参数估计中应用了最大似然估计的方法, 因此 GEB 方法的结果和 EB 估计的结果接近。另外, 相比于 REB 估计方法, 提出的 DEB 估计具有更小的 MSE, 且 DEB1 与 EB 估计的 MSE 也较为接近, 表明即使在无离群值时, DEB 方法也可以得出较为理想的估计结果。在情形 II 中, 模型中存在一定比例的离群值, 在对三个参数的估计中, EB 方法和 GEB 方法的估计效果较差, 尤其对  $A$  的估

表 3.1 FH 模型参数估计的偏差和 MSE

情形	参数	误差	DEB 1	DEB 2	EB	REB	GEB
I	$\beta_0$	Bias	-0.0034	-0.0039	-0.0025	-0.0060	-0.0026
		MSE	0.0865	0.0875	<b>0.0852</b>	0.0904	0.0853
	$\beta_1$	Bias	-0.0036	-0.0033	-0.0040	-0.0003	-0.0042
		MSE	0.2599	0.2629	0.2561	0.2723	<b>0.2560</b>
	A	Bias	-0.0380	-0.0373	0.0006	-0.0386	0.0011
		MSE	<b>0.0408</b>	0.0414	0.0422	0.0601	0.0418
II	$\beta_0$	Bias	0.0018	0.0017	0.0148	0.0048	0.0148
		MSE	0.1585	<b>0.1363</b>	0.7132	0.1606	0.7132
	$\beta_1$	Bias	-0.0121	-0.0084	-0.0594	-0.0194	-0.0594
		MSE	0.4831	<b>0.4189</b>	2.1554	0.4979	2.1554
	A	Bias	1.0767	0.5865	7.4438	0.4991	7.4438
		MSE	1.7490	0.5622	78.5515	<b>0.4977</b>	78.5515
III	$\beta_0$	Bias	0.0157	0.0108	0.0059	0.0104	0.0059
		MSE	0.2979	<b>0.2245</b>	1.4476	0.3583	1.4476
	$\beta_1$	Bias	-0.0064	0.0031	-0.0257	-0.0081	-0.0257
		MSE	0.9027	<b>0.6848</b>	4.2630	1.0730	4.2630
	A	Bias	3.2924	1.8056	15.0654	2.0388	15.0654
		MSE	14.0173	<b>5.8441</b>	268.2934	7.0164	268.2934

计中，MSE 变得非常大，可见估计及其不稳定，其偏差也较大。REB 方法能够很好地改善离群值带来的影响，估计较为理想。与之相比，DEB 方法的估计效果显得更加明显，尤其是取得较大调整参数的 DEB2，其参数估计的 MSE 均相对较小。进一步，在情形 III 中，当污染比例变大时，此时有更多的离群观测，结合表中数据发现，EB 方法和 GEB 方法表现及其糟糕，REB 方法表现较好，能够得到较为稳健的估计，而与提出的 DEB2 相比，DEB 方法会表现更好。综合上述三种情形下模型参数估计的比较可知，提出的 DEB 方法不仅在没有离群观测时表现较为稳定，而且在离群观测存在时，估计效果显著，比现有的几类稳健估计方法都好。

为进一步验证提出方法的估计效果，下面我们检验在三种情形下，稳健估计

表 3.2 目标变量  $\theta$  估计的 MSE

情形	分组	DEB1	DEB2	EB	REB1	REB2	GEB1	GEB2
I	1	0.1505	0.1516	<b>0.1490</b>	0.4751	0.179	0.1506	0.1527
	2	0.2432	0.2478	<b>0.2388</b>	0.5073	0.2714	0.2441	0.2510
	3	0.3068	0.3164	<b>0.2974</b>	0.5135	0.3443	0.3040	0.3228
	4	0.3457	0.3605	<b>0.3324</b>	0.5259	0.3885	0.3501	0.3689
	5	0.3937	0.4179	<b>0.3697</b>	0.5327	0.4476	0.3998	0.4283
II	1	0.1763	<b>0.1727</b>	0.1927	8.3419	0.2148	0.1911	0.1924
	2	0.3277	<b>0.3168</b>	0.3851	8.4285	0.3472	0.3795	0.3848
	3	0.4369	<b>0.4187</b>	0.5371	7.9312	0.4503	0.5255	0.5392
	4	0.5669	<b>0.5435</b>	0.7143	7.4896	0.5579	0.7038	0.7212
	5	0.6492	<b>0.6217</b>	0.8541	8.0940	0.6237	0.8474	0.8669
III	1	0.1957	<b>0.1920</b>	0.2024	14.0467	0.2484	0.2020	0.2024
	2	0.3734	<b>0.3618</b>	0.3969	15.4921	0.4218	0.3977	0.3972
	3	0.5290	<b>0.5092</b>	0.5725	15.8427	0.5712	0.5733	0.5718
	4	0.6901	<b>0.6621</b>	0.7563	15.1586	0.6994	0.7567	0.7552
	5	0.8513	<b>0.8147</b>	0.9537	15.4203	0.8232	0.9551	0.9510

$\hat{\theta}$  的估计效果，其估计结果见表3.2。表3.2中，DEB1、DEB2 分别表示在相对误差比例不超过 0.05 和 0.1 时，对应的稳健估计量。EB 估计为传统的经验贝叶斯估计方法，REB1、REB2 分别表示 Sinha et al.<sup>[28]</sup> 提出的 EBLUP 方法和基于极大似然估计的方法，GEB1、GEB2 分别表示由 Ghosh et al.<sup>[27]</sup> 给出的方法中选取不同调整参数  $K$  时的稳健估计结果。

结合表中数据观察几类稳健估计的效果。总体而言，每种情形下，无论采用那种估计方法，随着抽样误差  $D_i$  的增加，组内平均 MSE 都在增加，另外，随着污染比例的增加，模型中的离群观测值增加时，估计的 MSE 也会随之增加。首先在无污染数据中，根据估计结果的 MSE 可以看出，EB 估计的效果最好。因为在这种情形下，模型的正态假设完全满足，因此 EB 估计方法最优。另外几类方法的 MSE 均较大，但提出的 DEB 方法中，DEB1 估计的 MSE 与 EB 估计的 MSE 最为接近，若在该方法中选取更小的调整参数  $\alpha$ ，则能够获得更加小的 MSE，足以说明在这种情形下提出的 DEB 方法仍具有适用性。在情形 II 中，存

在一定比例的离群观测值，此时模型的正态性假定不满足，结合几类估计方法的表现可知 DEB2 方法具有更小的 MSE，比其它几类稳健估计方法表现更好。在情形 III 中，提出的 DEB2 方法仍然表现最佳，且 DEB1 的估计效果会比其它两种稳健估计方法更好，进一步说明，在观测数据污染较为严重的情形下，提出的 DEB 仍然具有较好的稳健效果，且明显好于其它两类稳健估计方法。

为了直观反映上述几类方法的区别，根据上述模拟数据绘制各种方法估计 MSE 的箱线图。图3.1,3.2,3.3分别展示了三种情形下各组 MSE 的变化线型图。

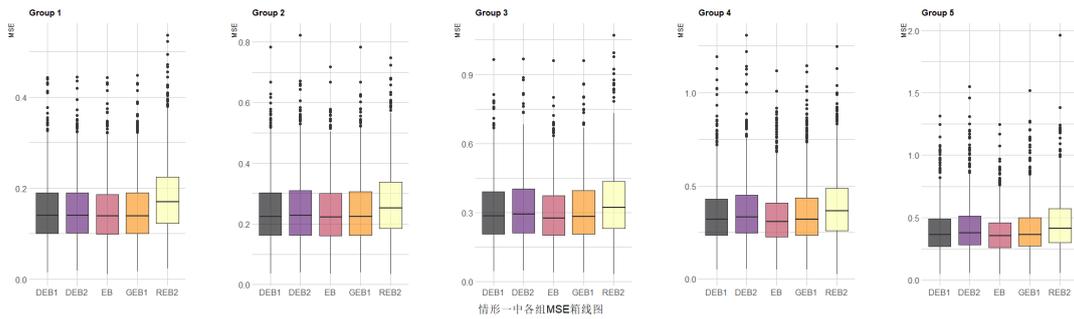


图 3.1 情形 1 中各方法估计 MSE 的箱线图

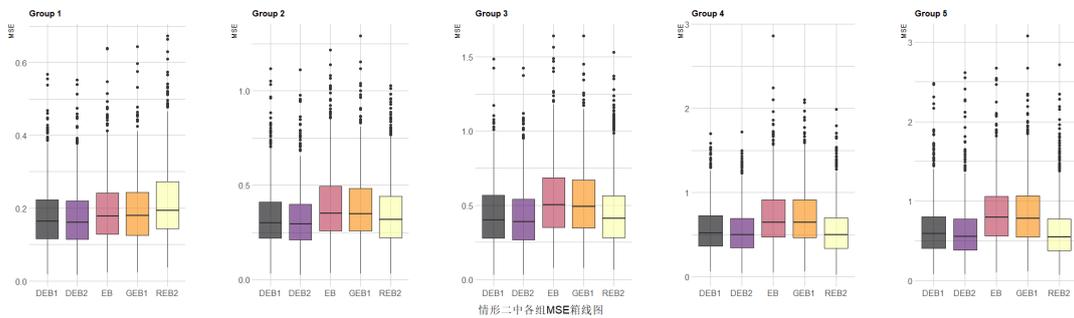


图 3.2 情形 2 中各方法估计 MSE 的箱线图

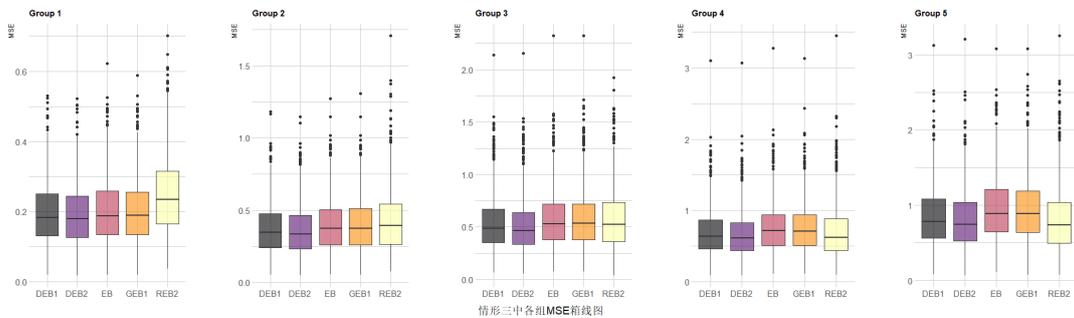


图 3.3 情形 3 中各方法估计 MSE 的箱线图

从上述图形可以直观得到和表3.2中相同的结论，在情形 I 中，EB 方法估计效果最好，但与 DEB1、GEB1 等方法的表现较为相近，差别不大。图形3.2、3.3反映了有污染的情形下几种估计方法的估计得到的每组 MSE 的箱线图。从图中不难发现，所有的箱线图中，DEB2 是每组中均值线最低，箱体最短的一个，说明 DEB2 方法的 MSE 估计均值最小，且方差不大，再次证明了 DEB 估计方法在有离群观测值时可得到较为稳健的估计，比其它稳健估计方法更好，这一结论和上述表格中得到的结论一致。

表 3.3 DEB 方法的 MSE 与估计 MSE 的比较

情形	分组	真实 MSE		估计 MSE	
		DEB1	DEB2	DEB1	DEB2
I	1	0.1525	0.1537	0.1537	0.1547
	2	0.2401	0.2450	0.2433	0.2486
	3	0.3099	0.3206	0.3043	0.3162
	4	0.3489	0.3651	0.3501	0.3703
	5	0.3847	0.4067	0.3873	0.4171
II	1	0.1756	0.1716	0.1795	0.1741
	2	0.3149	0.3043	0.3265	0.3097
	3	0.4384	0.4203	0.4515	0.4230
	4	0.5682	0.5417	0.5584	0.5181
	5	0.6652	0.6350	0.6554	0.6064
II	1	0.1920	0.1886	0.1923	0.1882
	2	0.3667	0.3554	0.3711	0.3575
	3	0.5439	0.5235	0.5387	0.5127
	4	0.6952	0.6664	0.6968	0.6571
	5	0.8175	0.7816	0.8488	0.7959

上述模拟实验是利用模拟生成的总体数据的值和区域上利用稳健估计方法得到的均值之间的比较，但在实际数据中，总体数据的均值是无法获取的，其 MSE 的计算只能利用 (3.16) 的公式进行估计。为了验证提出的 MSE 的估计式 (3.16) 的估计效果，同样利用上述估计模型估计真实的 MSE 和使用式 (3.16) 得出的 MSE 之间的差别。重复上述试验 1000 次，计算估计量的 MSE 在每组内的

均值，并用 (3.16) 的 Bootstrap 方法估计 MSE，每次 Bootstrap 中重复 2000 次实验得到其 MSE 的估计值，其估计结果见表 3.3。

表 3.3 中给出了三种情形下由两类 DPD 方法得出的真实 MSE 和 Bootstrap 方法得到的 MSE 的比较。从表中数据可知，两类方法的 Bootstrap 估计 MSE 和真实值均比较接近。进一步说明其 MSE 估计非常合理，能够充分接近与真实的 MSE。



## 第 4 章 基于 $\gamma$ 散度的区域水平模型的稳健估计

### 4.1 基于 $\gamma$ 散度的 FH 模型的稳健估计方法

在第三章中已经获知,  $\gamma$  散度用于衡量两个概率密度之间的差异, 且在稳健估计方法具有广泛的应用。一个自然的想法就是在式 (3.4) 中, 用 Nakagawa et al.<sup>[90]</sup> 提出的  $\gamma$  似然函数替换对数似然函数, 得到参数  $\theta_i$  的稳健估计。

$\gamma$  似然函数定义为:

$$L_\gamma(y; \phi) = -\frac{1}{\gamma} \sum_{i=1}^m \frac{f_i(y_i; \phi)^\gamma}{(\int f_i(t; \phi)^{1+\gamma} dt)^{\frac{\gamma}{\gamma+1}}} - \frac{n}{\gamma}, \quad \gamma \in (0, 1) \quad (4.1)$$

其中  $f_i(y_i; \phi)$  是  $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, A + D_i)$  的密度函数,  $\gamma$  同样是控制稳健性的调整参数。根据性质 2.4 的第 4 条可知, 有

$$\lim_{\gamma \rightarrow 0} L_\gamma(y; \phi) = \sum_{i=1}^m \log f_i(y_i; \phi) = \log f(y; \phi). \quad (4.2)$$

通过公式 (4.2) 可以看出, 当调整参数  $\gamma \rightarrow 0$  时, (4.1) 中的  $\gamma$  似然函数趋于对数似然函数, 足以说明,  $\gamma$  似然函数可以作为对数似然函数的推广, 对数似然可以作为其特殊情形。

对于形如 (2.1) 的 FH 模型, 其目标变量的观测值  $y_i$  满足  $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, A + D_i)$ , 因此结合密度函数的具体表达式, 将其代入公式 (4.1) 可得  $\gamma$  似然函数的具体表达式,

$$L_\gamma(y; \phi) = -\frac{1}{\gamma} \sum_{i=1}^m \left\{ (1 + \gamma)^{\frac{\gamma}{2(1+\gamma)}} (2\pi(A + D_i))^{-\frac{\gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2(A + D_i)}\right) \right\} - \frac{n}{\gamma}. \quad (4.3)$$

为了便于描述, 记  $V_i = \{2\pi(A + D_i)\}^{-1/2}$ ,  $\pi_\gamma = (1 + \gamma)^{\frac{\gamma}{2(1+\gamma)}}$ , 以及

$$s_i(y_i; \phi) = V_i^\gamma \exp\left\{-\frac{\gamma(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2(A + D_i)}\right\}; V_{i\gamma} = V_i^{-\frac{\gamma^2}{1+\gamma}},$$

那么  $L_\gamma(y; \phi)$  可以写成

$$L_\gamma(y; \phi) = -\frac{1}{\gamma} \sum_{i=1}^m \{\pi_\gamma V_{i\gamma} s_i(y_i; \phi)\} - \frac{n}{\gamma}. \quad (4.4)$$

因为  $\gamma$  散度在稳健估计方面的优良性质, 在本节中, 考虑基于  $\gamma$  散度的 FH 模型的小域稳健估计。用 (4.4) 替代式 (3.4) 中的对数似然函数  $\log f(y; \phi)$  构建小域目标变量的稳健估计。

## 4.2 稳健估计量

根据 (4.4) 中  $L_\gamma(y; \phi)$  的定义, 可得

$$\frac{\partial}{\partial y_i} L_\gamma(y; \phi) = -\frac{1}{\gamma} \pi_\gamma V_{i\gamma} \frac{\partial}{\partial y_i} s_i(y; \phi) = \frac{1}{A + D_i} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \pi_\gamma V_{i\gamma} s_i(y_i; \phi),$$

类似于 (3.4) 的形式, 定义 FH 模型的目标变量  $\theta_i$  的稳健贝叶斯估计量  $\tilde{\theta}_i^R$  为:

$$\tilde{\theta}_i^R = y_i - \frac{D_i}{A + D_i} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \pi_\gamma V_{i\gamma} s_i(y_i; \phi), \quad (4.5)$$

这里  $\gamma$  是用于控制稳健性的调整参数, 且满足  $\pi_\gamma V_{i\gamma} s_i(y_i; \phi) = 1$  当  $\gamma \rightarrow 0$  时。即说明当  $\gamma = 0$  时稳健估计量  $\tilde{\theta}_i^R$  退化成经验贝叶斯估计量 (EB)  $\tilde{\theta}_i$ 。

在小域模型中, 通过第2章的介绍可知, 观测值中难免会出现离群值。一般情况下, 离群观测都会远离回归估计  $\mathbf{x}_i^T \boldsymbol{\beta}$ , 在这种情况下, 辅助信息  $\mathbf{x}_i$  不能有效地通过模型 (2.1) 改善直接估计  $y_i$ 。因此, 对于这类观测点, 收缩率应越小越好, 理想的收缩率应为 0。关于这些性质的讨论可参考文献 [63, 77, 99]。

从 (4.5) 的表达式可以看出, 在传统 EB 估计量中的收缩因子为  $D_i / (A + D_i)$ , 其与观测值  $y_i$  无关, 那么当观测数据中存在异常值时, 该估计不会因为异常值的出现而发生任何改变, 即表明传统的 EB 估计对离群观测没有稳健的作用。反观式 (4.5) 中本文提出的稳健估计量, 其收缩因子为  $\pi_\gamma V_{i\gamma} s_i(y_i; \phi) D_i / (A + D_i)$ , 从表达式可知, 该收缩因子与观测值  $y_i$  有关, 当观测值  $y_i$  为离群观测时, 这时通过收缩影子将会增加增加在回归估计部分的权重, 从而在表达式 (4.5) 中, 估计量较 EB 估计量而言, 会向回归估计部分倾斜, 减小直接估计  $y_i$  带来的影响, 从而达到稳健估计的目的。另外, 当抽样方差  $D_i \rightarrow 0$  时, 稳健估计量  $\tilde{\theta}_i^R$  如同 EB 估计量一样, 退化为直接估计量  $y_i$ 。

下面让我们比较几种不同贝叶斯稳健估计量的收缩性质。根据参考文献 [99, 144, 145] 和 [77] 中的理论, 假设记  $\eta_i$  为贝叶斯估计量, 则需重点关注  $|y_i - \mathbf{x}_i^T \boldsymbol{\beta}| \rightarrow \infty$  时  $|\eta_i - y_i|$  的变化, 这里  $y_i$  为离群观测值。如果综合对比现有的几个小域稳健估计方法就会发现, 传统的经验贝叶斯 (EB) 方法、Ghosh et al.<sup>[27]</sup> 提出的稳健估计方法 (GEB)、Sinha et al.<sup>[28]</sup> 提出的稳健贝叶斯方法 (REB) 等, 这几类方法中随着  $|y_i - \mathbf{x}_i^T \boldsymbol{\beta}| \rightarrow \infty$  有  $|\eta_i - y_i| \rightarrow \infty$ , 这说明现有的这几类方法对离群观测的稳健性较差, 当出现极端的离群值时, 上述几类稳健估计方法也会呈现发散的情况, 即没有稳健效果。而本文中涉及的两类稳健估计方法: 基于密度幂的稳健估计 (DEB) 和基于  $\gamma$  散度的稳健估计 (GREB) 均满足  $|y_i - \mathbf{x}_i^T \boldsymbol{\beta}| \rightarrow \infty$  时

$|\eta_i - y_i| \rightarrow 0$  这一性质, 这说明本文中用到的两类估计方法对于极端离群值仍然具有较好的稳健性。

### 4.3 FH 模型的参数估计

根据上述讨论的关于模型 (2.1) 的稳健贝叶斯估计量 (4.5) 可知, 其稳健估计量中包含未知参数  $\phi = (\beta^T, A)^T$ , 因此本节中考虑未知参数的估计问题。令  $\phi$  的稳健估计为  $\hat{\phi}_\gamma$ , 则有  $\hat{\phi}_\gamma = \operatorname{argmax} L_\gamma(y; \phi)$ , 其中  $L_\gamma(y; \phi)$  是 (4.3) 给出的  $\gamma$  似然函数。因此, 稳健参数估计  $\hat{\phi}_\gamma$  满足如下估计方程:

$$\begin{aligned} \frac{\partial L_\gamma}{\partial \beta} &= \pi_\gamma \sum_{i=1}^m V_{i\gamma} \frac{\mathbf{x}_i s_i(y_i; \phi) (y_i - \mathbf{x}_i^T \beta)}{A + D_i} = 0, \\ \frac{\partial L_\gamma}{\partial A} &= \pi_\gamma \sum_{i=1}^m \left\{ V_{i\gamma} \frac{(y_i - \mathbf{x}_i^T \beta)^2 s_i(y_i; \phi)}{(A + D_i)^2} - \frac{V_{i\gamma} s_i(y_i; \phi)}{1 + \gamma} \frac{1}{A + D_i} \right\} = 0. \end{aligned} \quad (4.6)$$

上述估计方程的解可通过牛顿拉夫逊算法求得, 在初次迭代的过程中可利用极大似然估计的结果作为参数  $\phi$  的初始值进行迭代, 通过迭代, 直至估计参数收敛则可得模型参数的估计值。

通过上述算法求得未知参数的估计  $\hat{\phi}_\gamma$  之后, 将其代入表达式 (4.5), 则可得小域稳健经验贝叶斯估计量  $\hat{\theta}_i^R = \tilde{\theta}_i^R(y_i, \hat{\phi}_\alpha)$ 。

### 4.4 调整参数的选取

在目标变量  $\theta$  的 GREB 估计 (4.5) 中, 除了模型参数之外, 还包含有调整参数  $\gamma$ , 因而在实际应用中, 需要估计调整参数  $\gamma$  的大小, 同时, 选取不同的  $\gamma$  值将会导致不同的稳健估计结果。这里同样使用估计量的均方误差来推导小域估计调整参数的选取。接下来首先给出 GREB 估计量的 MSE 的推导及其相关的引理。

**引理 4.1** 当  $y_i \sim N(\mathbf{x}_i^T \beta, A + D_i)$  时, 有如下的结果

$$\begin{aligned} E \left\{ (y_i - \mathbf{x}_i^T \beta)^{2j-1} s_i^k \right\} &= 0, \quad j, k = 1, 2, \dots, \\ E \left\{ (y_i - \mathbf{x}_i^T \beta)^{2j} s_i^k \right\} &= V_i^{k\alpha} (k\alpha + 1)^{-j-1/2} (2j-1)!! (A + D_i)^j, \quad j, k = 0, 1, 2, \dots, \end{aligned}$$

其中  $(2j-1)!! = (2j-1)(2j-3)\cdots(1)$ 。

结合上述引理, 关于 GREB 估计量的 MSE 有如下的定理。

**定理 4.1** 在模型 (2.1) 的假设条下和正则性假设条件之下, 基于  $\gamma$  散度的稳健贝叶斯估计量  $\tilde{\theta}_i^R$  的均方误差如下:

$$E\left\{\left(\tilde{\theta}_i^R - \theta_i\right)^2\right\} = g_{1i}(A) + g_{2i}(A), \quad (4.7)$$

其中,

$$g_{1i}(A) = \frac{AD_i}{A + D_i}, \quad g_{2i}(A) = \frac{D_i^2}{A + D_i} \left\{ \frac{\pi_\gamma^2 V_i^{\frac{2\gamma}{1+\gamma}}}{(2\gamma + 1)^{3/2}} - \frac{2\pi_\gamma V_i^{\frac{\gamma}{1+\gamma}}}{(\gamma + 1)^{3/2}} + 1 \right\}.$$

无论是提出的稳健估计量 (4.5) 还是稳健估计量的 MSE(4.7) 中均含有未知的调整参数  $\gamma$ , 在这里采用类似于 MDPD 方法的参数选择程序。由定理4.1可知, GREB 估计量  $\tilde{\theta}_i^R$  的 MSE 比 EB 估计量的 MSE 大  $g_{2i}(A)$ 。为了平衡估计量的稳健性和有效性, 定义  $\text{Ex}(\gamma) = \sum_{i=1}^m g_{2i}(\hat{A}_\gamma) / \sum_{i=1}^m g_{1i}(\hat{A}_\gamma)$  为 MSE 的相对增量的百分比, 其中  $\hat{A}_\gamma$  是在给定  $\gamma$  的条件下, 由式 (4.6) 得出的参数  $A$  的估计值。本文根据  $\text{Ex}(\gamma)$  选择调优参数  $\gamma$ , 使其不超过给定的阈值  $c$  (如 0.05 和 0.1); 也就是说, 选取调整参数  $\gamma^*$  使其满足  $\text{Ex}(\gamma^*) = c$ 。根据定理4.1可知,  $\text{Ex}(\gamma)$  随着  $\gamma$  递增, 因此可以获取唯一的  $\gamma$  使得上式成立。在实践中, 我们首先给出一个固定的  $\gamma$  作为初始值, 然后用二分法计算出最优  $\gamma$ 。在后面的部分中, 我们假设  $\gamma$  已经是选择过程选择的最优参数。下面给出参数  $\gamma$  的选取算法。

---

#### 算法 2 参数 $\gamma$ 的选择算法

---

**输入:** 给定调整参数、模型参数的初值  $\gamma^{(0)}$ ,  $\phi^{(0)} = (\beta^{(0)T}, A^{(0)})$ , 并给出超量误差相对百分比的上限  $c$ 。

**重复:**

- 1: 根据公式 (4.7) 和  $\text{Exc}(\gamma)$  的定义, 利用二分法计算  $\text{Exc}(\gamma) = c$  的根, 并对  $\gamma$  进行迭代, 即

$$\gamma = \gamma^{(i+1)} \text{使得 } \text{Exc}(\gamma^{(i+1)}) = c.$$

- 2: 固定  $\gamma$ , 根据估计方程 (4.6), 利用 Newton-Raphson 算法估计得到参数  $\phi^{(i+1)}$ 。

- 3: 重复上述步骤1和2, 直至  $\gamma$  收敛。即对于充分小的临界值  $\epsilon$  满足  $|\gamma^{(i+1)} - \gamma^{(i)}| < \epsilon$ 。

**输出:**  $\gamma^{(i+1)}$

---

#### 4.5 稳健估计量的性质

在 3.4 中提到的正则性条件下, 本文研究了稳健估计参数的渐近性质, 并给出了其渐近分布, 具体的渐近分布参见如下的定理:

**定理 4.2** 在正则性条件下,  $\hat{\beta}_\gamma$  和  $\hat{A}_\gamma$  是渐近独立的, 并且服从的渐近分布为  $N(\beta, m^{-1}J_\beta^{-1}K_\beta J_\beta^{-1})$  和  $N(A, K_A/mJ_A^2)$ , 其中

$$J_\beta = -\frac{1}{m(\gamma+1)^{3/2}}\pi_\gamma \sum_{i=1}^m \frac{V_i^{\frac{\gamma}{1+\gamma}} \mathbf{x}_i \mathbf{x}_i^T}{A+D_i}, \quad J_A = \frac{1}{m(\gamma+1)^{5/2}}\pi_\gamma \sum_{i=1}^m \frac{V_i^{\frac{\gamma}{1+\gamma}}}{(A+D_i)^2},$$

$$K_\beta = \frac{\pi_\gamma^2}{m(2\gamma+1)^{3/2}} \sum_{i=1}^m \frac{V_i^{\frac{2\gamma}{1+\gamma}} \mathbf{x}_i \mathbf{x}_i^T}{A+D_i}, \quad K_A = \frac{\pi_\gamma^2(3\gamma^2+4\gamma+2)}{m(1+\gamma)^2(1+2\gamma)^{5/2}} \sum_{i=1}^m \frac{V_i^{\frac{2\gamma}{1+\gamma}}}{(A+D_i)^2}.$$

当  $\gamma=0$ ,  $\hat{\beta}_\gamma$  的渐近协方差矩阵和  $\hat{A}_\gamma$  的渐近协方差退化成  $\beta$  与  $A$  的极大似然估计量的渐近协方差矩阵和渐近协方差。这一点和  $\gamma=0$  时 (4.4) 退化成分数似然函数相一致。

#### 4.6 稳健估计量的 MSE

本节中, 重点讨论 GREB 估计量的 MSE。尽管定理 4.1 给出了贝叶斯估计量的 MSE 的表达式, 但是表达式 (4.7) 中仍然包含有待估的参数, 也就是其 MSE 可能由于模型参数的估计会变得更大。在实际应用中, 用估计值  $\hat{\theta}_i^R$  的均方误差  $MSE(\hat{\theta}_i^R)$  来作为估计的不确定度量。 $\hat{\theta}_i^R$  的 MSE 定义为  $MSE(\hat{\theta}_i^R) = M_i = E\left\{\left(\hat{\theta}_i^R - \theta_i\right)^2\right\}$ 。由于  $\hat{\theta}_i^R$  依赖于估计量  $\hat{\phi}_\gamma$ , 且  $M_i$  同样收到  $\hat{\phi}_\gamma$  的影响, 所以直接估计  $M_i$  的解析表达式是很难办到的, 因此有必要得到  $M_i$  的近似值。我们借助于 Rao et al.<sup>[1]</sup> 中二阶无偏估计的方法估计  $M_i$ , 可得如下的定理。

**定理 4.3** 在正则性假设条件之下, GREB 估计量的 MSE 的估计为

$$M_i = g_{1i}(A) + g_{2i}(A) + \frac{g_{3i}(A)}{m} + \frac{g_{4i}(A)}{m} + \frac{2g_{5i}(A)}{m} + o(m^{-1}), \quad (4.8)$$

其中  $g_{1i}(A)$  和  $g_{2i}(A)$  同定理4.2的定义, 另有

$$\begin{aligned}
g_{3i}(A) &= \frac{D_i^2 V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma^2 (7\gamma^2 + 4\gamma + 1)}{(A + D_i)^2 (2\gamma + 1)^{5/2}} \mathbf{x}_i^T J_\beta^{-1} K_\beta J_\beta^{-1} \mathbf{x}_i, \\
g_{4i}(A) &= \frac{D_i^2 V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma^2 (15\gamma^4 + 36\gamma^3 + 34\gamma^2 + 16\gamma + 4) K_A}{4(A + D_i)^3 (2\gamma + 1)^{5/2} (1 + \gamma)^2 J_A^2}, \\
g_{5i}(A) &= \frac{3\gamma D_i^2 V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma \mathbf{x}_i^T J_\beta^{-1} K_\beta J_\beta^{-1} \mathbf{x}_i}{2(A + D_i)^2} \left( V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma \frac{\gamma + 1}{(2\gamma + 1)^{5/2}} - \frac{1}{(\gamma + 1)^{5/2}} \right) \\
&+ \frac{D_i^2 V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma K_A}{4(A + D_i)^3 J_A^2} \left\{ V_i^{\frac{2\gamma}{1+\gamma}} \pi_\gamma \frac{15\gamma^4 + 40\gamma^3 + 48\gamma^2 + 30\gamma + 8}{2(2\gamma + 1)^{7/2} (1 + \gamma)^2} + \frac{\gamma - 4}{(1 + \gamma)^{7/2}} \right\} \\
&+ \frac{D_i^2 \pi_\gamma^2 V_i^2 \mathbf{x}_i^T J_\beta^{-1} \mathbf{x}_i}{(A + D_i)^2} \left( \frac{V_i^{2\gamma} (\gamma - 1)}{(2\gamma + 1)^{5/2}} + \frac{\pi_\gamma V_i^{\frac{2\gamma^2 + 3\gamma}{1+\gamma}}}{(3\gamma + 1)^{5/2}} \right) \\
&+ \frac{\pi_\gamma V_i^{\frac{2\gamma}{1+\gamma}} D_i^2}{2(A + D_i)^3 J_A} \left\{ \frac{1}{(2\gamma + 1)^{3/2}} - \frac{\pi_\gamma V_i^{\frac{\gamma}{1+\gamma}}}{(3\gamma + 1)^{3/2}} \right\} \left\{ \frac{9\gamma^2 + 6\gamma - 1}{(3\gamma + 1)(\gamma + 1)^2} - \frac{3\pi_\gamma (9\gamma^2 + 4\gamma - 3)}{(3\gamma + 1)^2 (\gamma + 1)} \right\}.
\end{aligned}$$

该定理详细的证明在本章附录中给出。根据 Datta et al.<sup>[99]</sup> 中的结论可知, 当  $\gamma = 0$  时, (4.8) 中的 MSE 估计退化为 EB 估计量的 MSE。从定理4.3可以看出, MSE 的近似估计公式与  $\gamma$  的近似估计公式密切相关, 并且这个估计公式的计算也是比较困难的。因此在实践中, 采用二阶近似的方法计算 MSE 的近似估计。由于计算公式 (4.8) 中的估计量需要大量的代数运算, 而且计算较为复杂。因此本文中采用 Bootstrap 估计方法估计估计量的 MSE。定义如下的 Bootstrap 估计量:

$$\hat{M}_i = 2g_{12i}(\hat{A}_\gamma) - \frac{1}{B} \sum_{b=1}^B g_{12i}(\hat{A}_\gamma^{(b)}) + \frac{g_{3i}(\hat{A}_\gamma)}{m} + \frac{g_{4i}(\hat{A}_\gamma)}{m} + \frac{2g_{5i}(\hat{A}_\gamma)}{m}, \quad (4.9)$$

其中  $g_{12i}(A) = g_{1i}(A) + g_{2i}(A)$  和  $\hat{A}_\gamma^{(b)}$  是基于参数 Bootstrap 样本  $y_1^{(b)}, \dots, y_m^{(b)}$  生成的 Bootstrap 估计量。

$$y_i^{(b)} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\gamma + v_i^{(b)} + \varepsilon_i^{(b)}, \quad v_i^{(b)} \sim N(0, \hat{A}_\gamma), \quad \varepsilon_i^{(b)} \sim N(0, D_i).$$

根据 Sugawara<sup>[77]</sup> 中的结论可知, 当  $B \rightarrow \infty$  有  $E(\hat{M}_i) - M_i = o(m^{-1})$  成立。而针对有限的 Bootstrap 样本而言, 其估计误差  $U_i$  的理想情形为  $U_i = O_p\{(mB)^{-1/2}\}$ 。这说明如果  $B = O(m^{1+\delta})$  对任意小的  $\delta > 0$  成立, 那么由有限的 Bootstrap 迭代得到的误差为  $o_p(m^{-1})$ , 这样以来, 估计量  $\hat{M}_i$  则会与理想的估计量具有相似的性质。这说明, 我们总可以找到有限次 Bootstrap 迭代来得到理想的近似结果。

根据定理 4.3, 由于  $g_{5i}(\hat{A}_\gamma)$  的表示结果较为复杂, 本文采用 Bootstrap 方法来估计  $g_{5i}(\hat{A}_\gamma)$ 。由定理 4.3 的证明可知,  $E\left\{\left(\hat{\theta}_i^R - \tilde{\theta}_i^R\right)\left(\hat{\theta}_i^R - \tilde{\theta}_i\right)\right\} = m^{-1}g_{5i}(A) + o(m^{-1})$ , 所以本文采用

$$B^{-1} \sum_{b=1}^B \left\{ \hat{\theta}_i^R \left( y_i^{(b)}, \hat{\phi}_\gamma^{(b)} \right) - \tilde{\theta}_i^R \left( y_i^{(b)}, \hat{\phi}_\gamma \right) \right\} \left\{ \left( \hat{\theta}_i^R \left( y_i^{(b)}, \hat{\phi}_\gamma \right) - \tilde{\theta}_i \left( y_i^{(b)}, \hat{\phi}_\gamma \right) \right) \right\},$$

来替代 (4.9) 中的  $m^{-1}g_{5i}(A)$  获取 MSE 的估计量, 其中  $\hat{\phi}_\gamma^{(b)}$  为参数 Bootstrap 估计量。

#### 4.7 稳健估计量的置信区间

在本节中, 重点考虑上述 FH 模型的稳健估计量的置信区间, 以及置信区间的构造。

在模型 (3.1) 下, 置信水平为  $(1 - \alpha)$  的置信区间  $I_i(\alpha)$  定义为

$$P[\theta_i \in I_i(\alpha)] = 1 - \alpha. \quad (4.10)$$

目前关于 FH 模型中小域均值  $\theta$  的置信区间的构造主要有两类方法。第一种是利用估计量的正态分布性质, 构造基于正态分布的置信区间, 形如

$$I_i^{EB}(\alpha) = [\hat{\theta}_i^{EB} - z_{\alpha/2}s(\hat{\theta}_i^{EB}), \hat{\theta}_i^{EB} + z_{\alpha/2}s(\hat{\theta}_i^{EB})], \quad (4.11)$$

其中  $z_{\alpha/2}$  是标准正态分布  $N(0, 1)$  的上  $(\alpha/2)$  分位点,  $s^2(\hat{\theta}_i^{EB})$  是  $MSE(\hat{\theta}_i^{EB})$  的二阶无偏估计量。虽然这一类置信区间在置信水平  $1 - \alpha$  下是渐近收敛的, 但是在小域估计中, 由于小样本性以及估计量 MSE 不稳定性导致这类置信区间并不是非常有效。Diao et al.<sup>[146]</sup> 等指出这类置信区间在  $P[\theta_i \in I_i^{EB}(\alpha)] = 1 - \alpha + o(m^{-1})$  的意义下是一阶正确的。第二类置信区间是参数 Bootstrap 置信区间, 该类置信区间由 Chatterjee et al.<sup>[147]</sup> 提出。Chatterjee et al.<sup>[147]</sup> 指出在正则性假设条件下的参数 Bootstrap 置信区间  $I_i^{CLL}(\alpha)$  是二阶渐近收敛的, 且满足

$$P[\theta_i \in I_i^{CLL}(\alpha)] = 1 - \alpha + O(m^{-3/2}).$$

尽管在正态分布的假设下第一类置信区间更便于构造和更直观的结构, 但是在观测数据中有离群值时, EB 估计量的 MSE 就会很大, 这会导致基于 EB 估计量的置信区间的长度变大, 从而降低区间的有效性。因此在本小节中, 将通过参数 Bootstrap 的方式构造上述提出的稳健估计量的置信区间。

基于上述提出的稳健估计量  $\tilde{\theta}_i^R$  以及位置参数  $\hat{\beta}, \hat{A}$  的估计值, 通过如下的步骤构建 Bootstrap 置信区间:

1. 生成  $v_i^* \sim N(0, \hat{A})$  和  $e_i^* \sim N(0, D_i)$ , 根据模型 (3.1),  $y_i^* = \theta_i^* + e_i^*$ ;  $\theta_i^* = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + v_i^*$  生成数据  $(y_i^*, X_i)$ .
2. 根据估计方程 (4.6) 将数据  $(y_i^*, X_i)$  用于估计参数  $\hat{\boldsymbol{\beta}}^*$  和  $\hat{A}^*$ .
3. 根据公式 (4.7) 和定理4.1, 用  $y_i^*, \hat{\boldsymbol{\beta}}^*, \hat{A}^*$  替代  $y_i, \boldsymbol{\beta}, A$  得到 Bootstrap 估计量  $\tilde{\theta}_i^{R*}$  和  $\text{MSE } g_{12}^*(\hat{A}^*)$ .
4. 用 Bootstrap 数据计算分布  $(t^* = \theta_i^* - \tilde{\theta}_i^{R*})/[g_{12}^*(\hat{A}^*)]^{1/2}$  作为分布  $(t = \theta_i - \tilde{\theta}_i^{EB})/[g_{12}(\hat{A})]^{1/2}$  的渐近分布。
5. 重复上述步骤  $B$  次, 得到近似分布值  $t^*$ , 计算分位数  $q_1$  和  $q_2 (q_1 < q_2)$  使得  $P_*(q_1 \leq t^* \leq q_2) = 1 - \alpha$  成立, 其中  $P_*$  表示基于 Bootstrap 分布的概率。
6. 得到  $\theta_i$  的 Bootstrap 置信区间

$$I_i(\alpha) = \left[ \tilde{\theta}_i^R + q_1 g_{12}(\hat{A})^{1/2}, \tilde{\theta}_i^R + q_2 g_{12}(\hat{A})^{1/2} \right]. \quad (4.12)$$

如果模型满足正则性假设条件, 则基于稳健估计量的置信区间满足如下定理:

**定理 4.4** 设  $\alpha \in (0, 1)$ . 若  $(q_1, q_2)$  满足

$$P \left( q_1 \leq g_{12}^*(\hat{A}^*)^{-1/2} \left( \theta_i^* - \tilde{\theta}_i^{R*} \right) \leq q_2 \right) = 1 - \alpha.$$

则

$$P \left( \tilde{\theta}_i^R + q_1 g_{12}(\hat{A})^{1/2} \leq \theta_i \leq \tilde{\theta}_i^R + q_2 g_{12}(\hat{A})^{1/2} \right) = 1 - \alpha + O(m^{-3/2}).$$

## 4.8 应用举例

本节中, 将本文提出的 GREB 方法应用与小域 FH 模型的估计中, 并将其与现有的几个稳健估计方法进行对比, 检验本文提出的稳健估计量的效果。

### 4.8.1 数值模拟

首先我们用模拟数据验证几类稳健估计量的估计效果。为了便于和第3.5中的方法进行直接比较, 本节中采用和3.5提出的同样的模型和假设条件生成相关的模拟数据。即本接种同样采用 2 元标准 FH 模型, 并考虑观测数据来源

表 4.1 几类稳健方法所得的每组平均 MSE 和蒙特卡洛误差

情形	分组	GREB 1	GREB 2	DEB 1	DEB 2	EB	REB
I	1	153 (1.3)	154 (1.2)	153 (1.3)	154 (1.2)	151 (1.4)	162 (1.5)
	2	244 (2.8)	249 (2.8)	244 (2.8)	249 (2.8)	240 (2.7)	259 (2.5)
	3	306 (3.6)	316 (3.7)	306 (3.6)	316 (3.7)	297 (3.6)	316 (2.8)
	4	351 (3.1)	366 (3.2)	351 (3.1)	366 (3.3)	337 (3.0)	353 (4.1)
	5	384 (3.2)	406 (3.3)	384 (3.2)	406 (3.3)	363(3.6)	382 (4.1)
II	1	175 (1.9)	170 (1.9)	176 (1.9)	172 (1.9)	192 (2.1)	221 (13.7)
	2	317 (3.4)	306 (3.5)	321 (3.4)	311 (3.4)	374 (3.2)	631 (11.6)
	3	438 (3.1)	418 (3.1)	445 (3.2)	427 (3.1)	546 (6.3)	1438 (75.9)
	4	551 (6.1)	527 (5.7)	562 (6.2)	539 (5.8)	710 (8.2)	2502 (123.9)
	5	653 (8.3)	622 (7.9)	668 (8.4)	639 (8.0)	874 (7.6)	3546 (126.6)
III	1	189 (1.4)	184 (1.3)	190 (1.3)	187 (1.3)	197 (1.5)	191 (3.4)
	2	363 (4.3)	349 (4.2)	368 (4.4)	357 (4.3)	390 (4.5)	376 (6.5)
	3	523 (7.2)	498 (6.8)	534 (7.3)	514 (6.9)	577 (7.4)	593 (13.7)
	4	675 (10.3)	640 (9.9)	692 (10.4)	664 (9.9)	761 (10.7)	903 (45.2)
	5	821 (10.7)	776 (10.8)	844 (10.9)	809 (11.0)	938 (11.8)	1341 (52.5)

于同样的三种混合正态分布情形下的稳健估计。有关模拟数据的基本假设可参见3.5节的内容。

根据3.5节中的研究结果可知, DEB 估计方法具有显著的稳健估计效果, 且好于 GEB 和 REB 方法。本节中, 选取传统的经验贝叶斯估计方法 (EB)、Sinha et al.<sup>[28]</sup> 基于 Huber $\phi$  函数的稳健估计方法 (REB)、基于密度幂的稳健估计方法 (DEB) 和本节中提出的基于  $\gamma$  散度的稳健估计方法 (GREB) 用于比较之间的估计效果。在这几类估计方法中, REB 估计中选取调整参数  $K = 1.345$ , DEB 和 GREB 估计方法中, 选取调整参数使得其相对误差比例不超过 0.05 和 0.1, 即  $c = 0.1$  和  $c = 0.05$  两种情形。

模拟实验 20000 次重复下不同估计量的 MSE。表4.1给出了在每组内平均的 MSE 值, 以及估计的蒙特卡洛误差。为了便于比较, 表中数据均在估计的基础上乘以 1000 而成。表中 GREB1 和 GREB2 分别表示  $c$  分别取 0.05, 0.1, DEB1 和 DEB2 的参数表示同样的误差相对百分比下选取调整参数得到的估计方法。

结合表4.1的数据可以得出如下结论: 括号内的蒙特卡洛误差比 MSE 小得多, 说明稳健估计方法得到的 MSE 是相对稳定的。在情形 I, 由于随机误差  $u_i$  满足正态假设, 模型中没有异常值。理论上, 在这种情况下, EB 方法应该比其它方法有更小的 MSE。观察表4.1中的数据可以发现, 正如预期的结果一样, 该情形下 EB 方法表现最好, GREB 方法和 DEB 方法表现相当。在情形 II 和 III 中, 数据中有离群值, 本文提出的方法比其它方法有更小的 MSE, 特别在具有较大抽样方差的组中表现更为突出, 表明我们的方法在这种情况下是有效的。本文提出的 GREB 方法和 DEB 方法中,  $c$  的值代表相对超额误差比率, 值越大, 估计模型的稳健性越强。因此,  $c = 0.1$  时的 MSE 要小于  $c = 0.05$  时的 MSE。但是, 在相同的  $c$  下, 本节所提出的方法 GREB 估计的 MSE 比 DEB 方法要小, 并且 GEB1 估计的效果接近于 DEB2, 这表明 GREB 方法可以在较小的膨胀率下达到与 DEB2 相同的效果。在其它方法中, 由于异常值的存在, 稳健估计方法普遍优于 EB 方法。

可以看出, 当存在离群观测时, 基于密度幂散度和  $\gamma$ -散度的估计比其它稳健估计得到了更小的 MSE。为了进一步证明提出的  $\gamma$ -散度的性能, 绘制了在情形 II 和 III 下每组估计 MSE 的箱线图, 其中每组 MSE 的估计使用四种估计方法, 即 GREB1、GREB2、DEB1 和 DEB2。图4.1和图4.2分别描述了情形 II 和 III 中 MSE 的分布特征, 每个面板描绘了同一组中不同方法之间的 MSE 的箱线图。

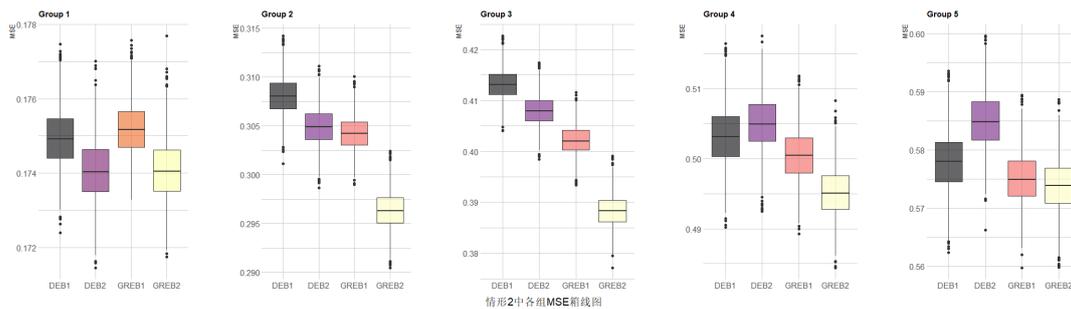


图 4.1 情形 2 中几类估计 MSE 的箱线图

从图 4.1和图4.2中我们可以得出以下结论: 首先, 在相同的条件下, 用 GREB 方法估计的 MSE 比用 DEB 方法估计的均值和方差更小。其次, 在异常值污染相同的情况下, 当抽样方差  $D_i$  增大时, GREB 方法将显著优于 DEB 方法, 在某些情况下, GREB1 方法将优于 DEB2 方法。最后, 当  $D_i$  相同时, GREB 方法在同一组内的性能优于 DEB 方法, 并且随着污染率的增加, GREB 方法估计

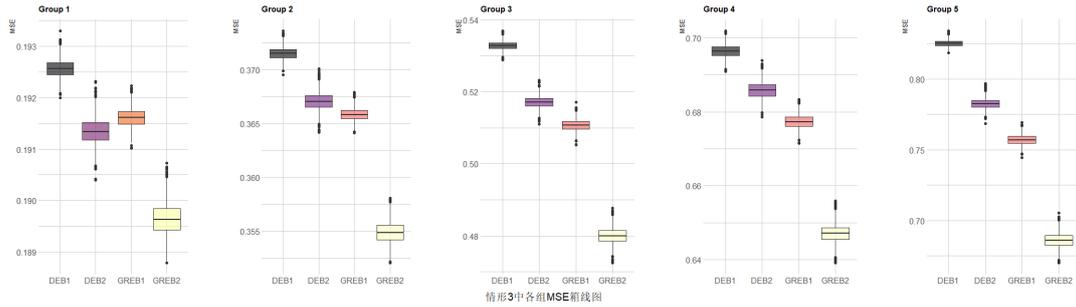


图 4.2 情形 3 中几类估计 MSE 的箱线图

的 MSE 方差更小。

接下来主要探究抽样为有限样本时，第4.6节中提到的 MSE 估计量  $\hat{M}_i$  的大小。这里使用与上述相同的模拟数据， $u_i$  满足上述三种情况的混合分布分布，其中抽样区域  $m = 50$ 。用  $\hat{A}_\gamma$  替换定理4.1和定理4.3中 MSE 公式中的  $A$  得到  $\hat{M}_i^{(n1)}$  和  $\hat{M}_i^{(n2)}$  的估计。注意  $\hat{M}_i^{(n1)}$  忽略模型参数估计的变化率， $\hat{M}_i^{(n2)}$  忽略  $g_{1i}(\hat{A}_\gamma) + g_{2i}(\hat{A}_\gamma)$  的偏差。考虑这些估计量和  $\hat{M}_i$  的目的是为了检验有限样本情形下二阶无偏性的重要性。基于 20000 次模拟数据，本文中估计了稳健经验贝叶斯估计器  $M_i$  的均方误差的真实值。估计量  $\hat{M}_i$  的相对偏差和相对均方误差的平方根为

$$\begin{aligned} \text{RBias}(\hat{M}_i) &= 100 \times E(\hat{M}_i - M_i) / M_i \\ \text{RRMSE}(\hat{M}_i) &= 100 \times E\left\{(\hat{M}_i - M_i)^2\right\} / M_i^2 \end{aligned}$$

这些值是基于 2000 次模拟运行的平均值，Bootstrap 样本容量为 1000；它们也在同一组内平均。表4.2给出了相对偏差和相对 MSE 的均方根  $\hat{M}_i$ 、 $\hat{M}_i^{(n1)}$  和  $\hat{M}_i^{(n2)}$ 。Bootstrap 估计量  $\hat{M}_i$  由于具有二阶无偏性会优于其它形式的估计器。所以粗略的估计在实践中是不可取的，因为它们会产生严重的负的偏差。由于估计量  $\hat{M}_i^{(n1)}$  和  $\hat{M}_i^{(n2)}$  存在严重的负偏差，而估计量  $\hat{M}_i$  具有更小的正偏差，因此本文所提到的 Bootstrap 估计量  $\hat{M}_i$  在实际应用中是可行的。

表 4.2 均方误差估计量的相对偏差和相对 MSE

情形	分组	RBias			RRMSE		
		$\hat{M}_i^{(n1)}$	$\hat{M}_i^{(n2)}$	$\hat{M}_i$	$\hat{M}_i^{(n1)}$	$\hat{M}_i^{(n2)}$	$\hat{M}_i$
I	1	-17.51	-11.32	-2.10	3.07	1.31	0.07
	2	-23.98	-17.38	-8.33	5.76	3.05	0.73
	3	-27.5	-20.0	-11.4	7.55	4.06	1.39
	4	-29.35	-23.13	-15.06	8.62	5.40	2.32
	5	-30.20	-24.70	-16.98	9.12	6.22	3.01
II	1	-4.28	-2.32	0.36	0.19	0.06	0.01
	2	-9.09	-7.74	-4.07	0.83	0.78	0.35
	3	-12.72	-12.15	-8.07	1.62	2.54	1.71
	4	-17.00	-13.39	-9.23	2.90	1.81	0.87
	5	-20.23	-16.13	-12.00	4.10	2.64	1.48
III	1	-0.04	0.77	1.48	0.005	0.01	0.03
	2	-1.23	-0.30	0.92	0.03	0.01	0.02
	3	-2.11	-3.84	-2.25	0.06	0.71	0.61
	4	-3.36	-1.48	0.39	0.13	0.06	0.04
	5	-4.71	-9.80	-7.74	0.24	2.67	2.31

接下来，使用上述模拟生成的数据进行置信区间的估计。使用 (4.11) 中提到的正态区间估计，Chatterjee et al.<sup>[147]</sup> 中提到的 Bootstrap 置信区间，以及本研究中提出的稳健置信区间 (4.12) 进行比较。在前两种置信区间的估计中参数  $A$  和  $\beta$  的估计使用最大似然估计方式得到，在本研究中提出的置信区间的构造中，参数是通过本文提到的稳健估计方法得到。在 Bootstrap 步骤中，设置  $B = 2000$ ，重复这一模拟 10000 次，在置信水平为 95% 时，计算每种情形下区域均值置信区间的覆盖率和平均长度，其结果呈现在表4.3中。其中 FH 表示使用 (4.11) 的置信区间，BI 表示用 Chatterjee et al.<sup>[147]</sup> 中提出的 Bootstrap 方法估计的置信区间，GBI 表示用本研究中提到的方法估计的置信区间。

从表4.3可以看出，在情形 I 中，Bootstrap 置信区间表现最好，具有更高的覆盖率和更短的长度。然而在具有离群观测的情形 II 和 III 中，本研究中提出的方法的置信区间的覆盖率更加接近于真实水平，且区间长度更小。这一点充分

表 4.3 置信度为 0.95 的置信区间的覆盖率 (平均长度)

情形	分组	FH	BI	GBI
I	1	0.9308(1.45)	0.9096(1.44)	0.9701(1.85)
	2	0.9260(1.79)	0.9816(2.48)	0.9714(2.25)
	3	0.9197(1.98)	0.9951(3.34)	0.9687(2.45)
	4	0.9201(2.11)	0.9990(4.03)	0.9658(2.56)
	5	0.9181(2.19)	0.9998(4.69)	0.9640(2.65)
II	1	0.9492(1.72)	0.6617(0.84)	0.9526(1.67)
	2	0.9487(2.39)	0.8187(1.64)	0.9549(2.25)
	3	0.9475(2.88)	0.8942(2.41)	0.9533(2.64)
	4	0.9459(3.27)	0.9341(3.14)	0.9513(2.93)
	5	0.9480(3.60)	0.9588(3.85)	0.9521(3.16)
II	1	0.9485(1.74)	0.5624(0.69)	0.9493(1.71)
	2	0.9498(2.44)	0.6594(1.18)	0.9505(2.36)
	3	0.9489(2.97)	0.7009(1.58)	0.9495(2.83)
	4	0.9486(3.40)	0.7197(1.88)	0.9486(3.21)
	5	0.9489(3.77)	0.7371(2.16)	0.9488(3.52)

说明本研究中给出的置信区间对存在离群值的观测数据具有更加稳健的估计效果。

在实际应用场景中, 有可能具有离群观测的数据并非来自于混合正态分布。接下来, 考虑当  $u_i$  服从其它有偏分布时, 使用上述用到的稳健方法对由模型 (2.1) 产生的数据进行稳健估计, 假设其中  $u_i$  的真实生成分布为以下情形:

$$\begin{aligned}
 (IV) u_i &\sim t_2, & (V) u_i &\sim \text{Ga}(0.5, 0.5) \\
 (VI) u_i &\sim \text{Ga}(2, 2), & (VII) u_i &\sim \text{ST}_3(2)/3^{1/2},
 \end{aligned}$$

这里  $t_n$  是自由度为  $n$  的  $t$  分布。  $\text{Ga}(a, b)$  是形状参数为  $a$  和尺度参数为  $b$  的伽马分布。  $\text{ST}_n(a)$  为有偏的自由度为  $n$  的  $t$  分布, 其中  $da$  表示偏度参数, 这一分布可以由 R 语言中的 “skewt” 包生成。

表 4.4 有偏分布下稳健估计的 MSE 和蒙特卡洛误差

情形	分组	GREB 1	GREB 2	DEB 1	DEB 2	EB	REB
IV	1	187(1.6)	187(1.6)	188(1.6)	187(1.6)	192 (1.6)	186(1.5)
	2	351(5.0)	350(5.1)	354(5.0)	352(5.1)	368(5.1)	353(14.6)
	3	505(8.3)	500(8.3)	510(8.5)	503(8.3)	533(8.6)	550(59.3)
	4	640(5.4)	631(5.1)	648(5.2)	636(5.1)	684(6.7)	734(29.6)
	5	770(6.3)	757(6.1)	782(6.3)	762(6.2)	830(6.7)	1263(388)
V	1	174 (1.3)	172(1.3)	174(1.3)	173(1.3)	181(1.5)	14618(2483)
	2	308(3.4)	304(3.4)	309(3.4)	307(3.4)	329(3.9)	13515(2326)
	3	419(5.9)	416(5.8)	422(6.0)	421(6.0)	456(6.7)	36968(81577)
	4	519(6.4)	516(5.9)	523(6.4)	523(6.0)	569(8.6)	14635(6286)
	5	607(7.4)	607(7.1)	611(7.3)	616(7.0)	669(8.8)	10934(1319)
VI	1	147(0.9)	148(1.0)	147(1.0)	148(1.0)	148(1.0)	167(2.6)
	2	236(2.4)	239(2.3)	237(2.4)	239(2.4)	238(2.7)	276(4.8)
	3	295(3.1)	301(3.3)	295(3.1)	301(3.3)	294(2.8)	336(4.1)
	4	338(4.4)	349(4.5)	338(4.4)	349(4.5)	334(4.7)	371(6.7)
	5	374(6.5)	391(7.1)	374(6.5)	391(7.1)	365(6.6)	398(8.6)
VII	1	174(2.1)	174(2.1)	175(2.1)	174(2.1)	179(2.4)	183(9.8)
	2	316(3.2)	311(3.2)	313(3.2)	312(3.2)	327(3.7)	364(16.5)
	3	428(4.2)	424(4.3)	431(4.2)	426(4.3)	454(5.1)	596(63.5)
	4	526(4.4)	518(4.2)	530(4.4)	520(4.2)	559(5.5)	897(154)
	5	618(8.8)	605(8.4)	624(8.8)	608(8.4)	661(8.6)	1165(146)

为了估计  $\theta_i$ , 本文采用采用相同的估计方法重复 2000 次模拟实验计算估计的 MSE。表 4.4给出了每组内平均的 MSE 值, 并在括号内给出了估计的蒙特卡洛误差。为了便于比较, 表中的数据均在其估计值基础上乘了 1000。

结合表4.4中的数据, 不难发现以下结论: 首先, 与 MSE 的估计值相比, 蒙特卡洛误差的估计值非常小, 这也表明这些稳健估计方法是相对稳定的。其次, 在情形 (IV) 和 (VII) 中, 由于  $t$  分布和有偏的  $t$  分布均具有厚尾特征, 因此产生的观测值中会包含有离群观测。在这一背景下, 从表中数据可以看出本节提出的 GREB 方法比其它方法估计的 MSE 更小。此时本节中提出的 GREB 方法更为

有效。在情形 (V) 和情形 (VI) 中, 虽然  $u_i$  的分布是有偏的, 但其生成的数据中未必一定包含有离群观测, 且情形 (V) 中的分布比情形 i (VI) 中的分布具有更大的偏度。结合表中数据可见, GREB 方法在情形 (V) 中的表现要好于其它估计方法。反之, 在情形 (VI) 中, 有些组内 EB 方法的估计效果要好于 GREB 方法, 但 GREB1 的估计结果和 EB 估计的结果非常相近, 且显著好于其它的稳健估计方法。总之, 无论是在重尾分布中还是普通的有偏分布中, GREB 方法均具有较好的稳健性质, 比其它的估计方法更值得推荐。

#### 4.8.2 实际应用

考虑一个来自美国劳工统计局新鲜牛奶支出数据的实际应用, 这一数据在 Arora et al.<sup>[148, 148]</sup> 中有所应用。数据集中有 43 个地区 1989 年鲜奶消费均值  $y_i$  的估计值, 以及抽样方差  $D_i$ 。根据 Arora et al.<sup>[148]</sup> 的研究, 建立形如 2.1 的 FH 模型。如果第  $i$  个区域属于第  $j$  个片区, 其 FH 模型可以表示为  $y_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_j (j = 1, \dots, 4)$  的形式, 其中片区共有四个, 其包含的区域为  $R_1 = \{1, \dots, 7\}, R_2 = \{8, \dots, 14\}, R_3 = \{15, \dots, 25\}$ , 和  $R_4 = \{26, \dots, 43\}$ 。根据文献 [77] 的结论, 片区  $R_1$  和  $R_2$  中包含有离群观测, 其离群值可通过如下的标准化残差来识别。本节中使用该数据来检验 GREB 估计方法的效果, 避免用传统的 EB 估计方法会导致估计偏差较大的问题,

$$r_i = \left( \hat{A} + D_i \right)^{-1/2} \left\{ y_i - \sum_{j=1}^4 \hat{\beta}_j I(i \in M_j) \right\} \quad (i = 1, \dots, m).$$

当 FH 模型 (2.1) 的假定是完全正确的, 那么  $r_i$  的分布是近似标准正态分布, 然而, 正如表 4.6 所示, 某些区域中  $r_i$  的绝对值非常大。

接下来使用上述提到的 REB、DEB、GREB 等方法估计模型的参数, 其估计结果可参见表 4.5。表中给出了几类方法下对于模型系数  $\beta_j$  和模型误差的方差  $A$  的估计及其标准误, 其中  $A$  的估计结果均乘了 100。

表 4.5 模型参数的估计及其标准误

	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$A$
EB	0.97 (0.07)	1.10 (0.07)	1.19 (0.06)	0.73 (0.04)	1.55 (0.68)
REB	1.01 (0.06)	1.18 (0.07)	1.19 (0.05)	0.73 (0.03)	0.80 (0.53)
DEB (1% )	0.97 (0.07)	1.12 (0.07)	1.19 (0.06)	0.73 (0.04)	1.50 (0.65)
DEB (5%)	0.98 (0.06)	1.15 (0.07)	1.19 (0.06)	0.73 (0.04)	1.35 (0.60)
GREB(1% )	0.97 (0.07)	1.12 (0.07)	1.19 (0.06)	0.73 (0.04)	1.50 (0.65)
GREB(5% )	0.98 (0.06)	1.15 (0.06)	1.19 (0.06)	0.73 (0.04)	1.35 (0.60)

从表中数据可以看出, 几种方法对  $\beta_3$  和  $\beta_4$  的估计相差不多, 这也是因为片区  $R_3$  和  $R_4$  片区中没有离群值。反之, 片区  $R_1$  和  $R_2$  中具有离群观测, 比较几类估计结果可见, GREB 方法和 DEB 方法得到了相同的估计结果, 且比另外两个稳健估计方法更加合理, 具有更小的偏差。

为了估计  $\theta_i$ , 本文采用 EB 方法、REB 方法、DEB 方法、GREB 方法估计得出对应目标变量的估计值和结合 (4.9) 式给估计量的 MSE。其中 DEB 方法和 GREB 方法中参数选择时使用相对误差比例上限  $c = 0.05$ , 利用 (4.9) 计算  $\hat{M}_i$  时采用的 Bootstrap 样本数  $B = 1000$ 。估计的目标值及其 MSE 均在表4.6中给出。

表4.6给出了各种估计方法下目标变量  $y_i$  的估计值  $\hat{\theta}_i$  及其  $MSE\hat{M}_i$ , 表中  $y_i$  表示目标变量的真实值,  $r_i$  表示标准化残差用于度量对应的目标变量是否为离群观测, 若  $r_i$  的绝对值越大, 则表明对应的观测值越有可能是离群观测。首先从表中的估计值比较而言, 在大多数的区域中, GREB 方法的估计值更加接近其真实值  $y_i$ , 这一点尤其是在  $r_i$  绝对值较大的区域显得更加突出。其次, 从表中 MSE 的估计结果来看, 几类方法的 MSE 均比较小, 说明估计结果较为稳定, 几类估计方法的 MSE 相差不大。

表 4.6 目标变量的估计值及其 MSE

区域	片区	$y_i$	$r_i$	$\hat{\theta}_i^{EB}$	$\hat{\theta}_i^{DEB}$	$\hat{\theta}_i^{GREB}$	$\hat{\theta}_i^{REB}$	$\hat{M}_i^{EB}$	$\hat{M}_i^{DEB}$	$\hat{M}_i^{GREB}$	$\hat{M}_i^{REB}$
1	1	1.099	0.640	1.016	1.017	1.017	1.027	0.0136	0.013	0.013	0.010
2	1	1.075	0.724	1.044	1.042	1.042	1.044	0.006	0.006	0.006	0.004
3	1	1.105	0.917	1.063	1.060	1.061	1.059	0.006	0.006	0.006	0.005
4	1	0.628	-2.053	0.775	0.763	0.762	0.907	0.009	0.009	0.009	0.006
5	1	0.753	-1.247	0.855	0.870	0.869	0.915	0.010	0.010	0.010	0.007
6	1	0.981	0.070	0.974	0.983	0.983	0.998	0.012	0.011	0.011	0.009
7	1	1.257	1.219	1.047	1.056	1.056	1.046	0.016	0.015	0.016	0.012
8	2	1.095	-0.004	1.095	1.129	1.129	1.153	0.011	0.010	0.010	0.008
9	2	1.405	1.479	1.205	1.235	1.236	1.231	0.014	0.014	0.014	0.011
10	2	1.356	1.198	1.181	1.209	1.209	1.216	0.015	0.014	0.015	0.011
11	2	0.615	-3.009	0.803	0.729	0.728	1.074	0.008	0.007	0.006	0.006
12	2	1.460	1.541	1.197	1.239	1.239	1.227	0.016	0.016	0.016	0.013
13	2	1.338	1.253	1.196	1.218	1.219	1.223	0.013	0.012	0.012	0.010
14	2	0.854	-1.274	0.991	1.021	1.020	1.106	0.012	0.012	0.012	0.009
15	3	1.176	-0.095	1.187	1.188	1.188	1.188	0.012	0.012	0.011	0.010
16	3	1.111	-0.437	1.159	1.166	1.166	1.170	0.012	0.011	0.011	0.009
17	3	1.257	0.340	1.223	1.217	1.217	1.212	0.011	0.011	0.010	0.008
18	3	1.430	1.109	1.276	1.268	1.268	1.243	0.014	0.013	0.013	0.010
19	3	1.278	0.451	1.232	1.224	1.225	1.218	0.011	0.011	0.010	0.008
20	3	1.292	0.475	1.230	1.222	1.222	1.215	0.013	0.013	0.012	0.010
21	3	1.002	-1.091	1.099	1.108	1.108	1.128	0.010	0.010	0.010	0.008
22	3	1.183	-0.042	1.192	1.193	1.193	1.191	0.017	0.017	0.016	0.016
23	3	1.044	-0.803	1.128	1.138	1.138	1.149	0.011	0.011	0.011	0.007
24	3	1.267	0.343	1.220	1.213	1.213	1.208	0.014	0.013	0.013	0.010
25	3	1.193	-0.009	1.194	1.194	1.194	1.193	0.008	0.008	0.008	0.006
26	4	0.791	0.379	0.759	0.755	0.755	0.750	0.009	0.009	0.009	0.007
27	4	0.795	0.402	0.761	0.756	0.757	0.751	0.009	0.009	0.009	0.007
28	4	0.759	0.118	0.732	0.732	0.732	0.731	0.016	0.016	0.016	0.017
29	4	0.796	0.433	0.766	0.761	0.762	0.756	0.008	0.008	0.008	0.006
30	4	0.565	-1.047	0.619	0.629	0.628	0.646	0.006	0.006	0.006	0.005

区域	片区	$y_i$	$r_i$	$\hat{\theta}_i^{EB}$	$\hat{\theta}_i^{DEB}$	$\hat{\theta}_i^{GREB}$	$\hat{\theta}_i^{REB}$	$\hat{M}_i^{EB}$	$\hat{M}_i^{DEB}$	$\hat{M}_i^{GREB}$	$\hat{M}_i^{REB}$
31	4	0.886	0.625	0.763	0.760	0.757	0.749	0.015	0.015	0.015	0.013
32	4	0.952	0.945	0.786	0.781	0.781	0.763	0.015	0.014	0.014	0.012
33	4	0.807	0.475	0.768	0.762	0.763	0.756	0.009	0.009	0.009	0.006
34	4	0.582	-1.013	0.614	0.621	0.621	0.634	0.004	0.004	0.004	0.003
35	4	0.684	-0.252	0.701	0.707	0.707	0.709	0.008	0.008	0.008	0.006
36	4	0.787	0.349	0.756	0.752	0.752	0.747	0.010	0.009	0.009	0.007
37	4	0.440	-1.842	0.541	0.538	0.538	0.611	0.007	0.006	0.006	0.005
38	4	0.759	0.186	0.741	0.740	0.740	0.737	0.010	0.010	0.010	0.007
39	4	0.770	0.280	0.752	0.750	0.750	0.746	0.007	0.007	0.007	0.006
40	4	0.800	0.445	0.766	0.761	0.761	0.755	0.009	0.008	0.008	0.006
41	4	0.756	0.206	0.747	0.745	0.745	0.743	0.006	0.006	0.005	0.005
42	4	0.865	0.805	0.797	0.788	0.789	0.776	0.009	0.009	0.009	0.007
43	4	0.640	-0.475	0.684	0.694	0.693	0.699	0.010	0.010	0.009	0.006

## 第 5 章 基于密度幂散度的单元水平模型的稳健估计

本章主要研究单元水平模型的稳健小域估计, 将密度幂散度应用于单元水平模型, 研究单元水平中小域目标单元的估计问题, 给出了参数估计程序, 估计量的表达式, 以及估计量的 MSE。并通过模拟数据和实际数据对本研究中提出的估计方法进行了验证, 证明本研究中所提出的方法具有最小的 MSE 和良好的稳健性。最后本章在前述理论研究的基础上, 考虑稳健小域估计量的应用研究。

类似于第3章的内容, 本章主要介绍利用最小化密度幂散度 (MDPD) 方法对单元水平模型进行稳健估计。提出经验贝叶斯估计量, 推导估计量的 MSE, 并给出 MSE 的估计。并通过模拟实验和实际数据检验提出估计量的稳健性, 和常用的几类稳健估计方法进行比较。

### 5.1 基于密度幂散度的 NER 模型的稳健估计方法

一般地, 记一个调查总体为  $U$ , 包含  $N$  个区域, 记作  $k = 1, 2, \dots, N$ , 将感兴趣的变量记作  $y$ , 第  $i$  个区域上包含  $N_i$  个单元, 则第  $i$  个区域上的第  $j$  个目标变量为  $y_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, N_i$ 。一般情形下, 重点关注目标变量的总和和均值, 目标变量  $y_{ij}$  的总和、平均值表示为  $Y = \sum_{i=1}^m \sum_{j=1}^{N_i} y_{ij}, \bar{Y} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{N_i} y_{ij}$ 。然而在实际应用中, 这些  $y_{ij}$  并不全部能够获得。一般是从总体  $U$  中以抽样概率  $p(s)$  获得样本  $s$ , 其中抽样概率取决于抽样设计。抽样得到的样本数据为  $y_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, n_i$ , 即抽取的样本数据来自于  $n$  个区域, 第  $i$  个区域上抽取了  $n_i$  个单元。假设单元水平上的辅助信息  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$  对于区域  $i$  上的第  $j(j = 1, \dots, N_i)$  个总体元素都是可用的, 即每个总体元素都能找到对应的辅助信息变量。那么标准的单元水平模型, NER 模型可描述为:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad j = 1, \dots, N_i, \quad i = 1, \dots, m. \quad (5.1)$$

其中  $\boldsymbol{\beta}$  是未知的  $p$  维的回归系数向量, 假设区域随机效应  $v_i$  是相互独立的且服从  $N(0, \sigma_v^2)$ , 单元误差  $e_{ij} = k_{ij} \tilde{e}_{ij}$ , 且  $k_{ij}$  是已知的常数,  $\tilde{e}_{ij}$  独立于  $v_i$  的独立同分布的随机误差服从分布为  $N(0, \sigma_e^2)$ 。

记从第  $i$  个区域的  $N_i$  个单元中抽取容量为  $n_i$  的样本为  $s_i$ , 则抽取的样本数据同样满足单元水平模型5.1。假设对于每个区域上的简单随机抽样或其它抽

样设计, 抽取的样本结合对应的辅助信息  $\mathbf{x}_{ij}$  均满足上述模型, 将抽样数据对应的单元水平模型写成矩阵形式, 可以表示为:

$$\mathbf{y}_i = \mathbf{x}_i \boldsymbol{\beta} + v_i \mathbf{1}_i + \mathbf{e}_i, \quad i = 1, \dots, m, \quad (5.2)$$

其中  $\mathbf{x}_i$  是  $n_i \times p$  的矩阵,  $\mathbf{y}_i$  和  $\mathbf{e}_i$  是  $n_i \times 1$  的向量,  $\mathbf{1}_i$  是  $n_i \times 1$  的所有元素为 1 的向量。

令  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  为模型 (5.1) 中未知参数构成的向量。在正态性假定下,  $\mathbf{y}_i | \mathbf{x}_i$  服从正态分布, 其条件概率密度函数为

$$f_{\boldsymbol{\theta}}(\mathbf{y}_i | \mathbf{x}_i) = (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta}) \right\}, \quad (5.3)$$

其中

$$\mathbf{V}_i = \sigma_e^2 \text{diag}_{1 \leq j \leq n_i} (k_{ij}^2) + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T.$$

矩阵  $\mathbf{V}_i$  可逆, 且根据 Sherman-Morrison 公式:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{u}\mathbf{v}^T \mathbf{A}^{-1} / (1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u}),$$

并记

$$a_{ij} = k_{ij}^{-2}, \quad \mathbf{a}_i = \sum_{j=1}^{n_i} a_{ij}, \quad \mathbf{a}_i = (a_{i1}, \dots, a_{in_i})^T,$$

可得其逆矩阵为

$$\mathbf{V}_i^{-1} = \frac{1}{\sigma_e^2} \left[ \text{diag}_{1 \leq j \leq n_i} (a_{ij}) - \frac{\rho_i}{a_i} \mathbf{a}_i \mathbf{a}_i^T \right], \quad |\mathbf{V}_i| = (\sigma_e^2 + a_i \sigma_v^2) \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2. \quad (5.4)$$

其中

$$\rho_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / a_i)$$

在本章中, 将通过使用 MDPD 方法估计模型 5.2 中的未知参数, 并对目标变量进行稳健估计。

在本节中, 基于最小密度幂散度提出了存在离群观测情形下的单元水平模型的稳健估计方法。首先通过引入最小密度幂函数, 给出了单元水平模型参数的估计方程, 并得到了参数稳健估计的渐近分布。考虑到稳健估计量中存在调整参数, 提出了最优参数选择的算法。其次, 有限总体中单元和区域均值的经验贝叶斯线性无偏估计量被给出, 还有, 提出的小区域稳健估计量的均方误差通过 Bootstrap 方法给出。最后通过模拟数据和实际数据验证了提出方法的优越表现。在模拟部分, 模拟了分布被污染时的稳健估计, 讨论了三种污染情形下几类稳健

估计方法的效果。还特别讨论了污染比例变化、以及污染分布的方差变化时几类估计方法的 MSE 的变化情况。同时模拟结果表明,提出的方法能更好地解决这种离群情况。实际数据中,用了一个小区域估计中很经典地数据来说明我们提出方法的有效性,通过比较,提出的方法能够很好地处理离群观测这一特殊情况。最后,通过模拟数据和实际数据验证了提出的估计量比 Sinha et al.<sup>[28]</sup> 中的稳健估计量具有更好的性质。

## 5.2 DPD 方法在 NER 模型中的应用和模型的参数估计

结合第2章的内容,模型密度  $f_{\theta}$  和真实密度  $g$  之间的密度幂散度 (DPD) 定义为

$$d_{\gamma}(f_{\theta}, g) = \begin{cases} \int_{\mathbf{y}} \left\{ f_{\theta}^{1+\gamma}(\mathbf{y}) - \left(1 + \frac{1}{\gamma}\right) f_{\theta}^{\gamma}(\mathbf{y})g(\mathbf{y}) + \frac{1}{\gamma}g^{1+\gamma}(\mathbf{y}) \right\} d\mathbf{y}, & \text{for } \gamma > 0 \\ \int_{\mathbf{y}} g(\mathbf{y}) \log \left( \frac{g(\mathbf{y})}{f_{\theta}(\mathbf{y})} \right) d\mathbf{y}, & \text{for } \gamma = 0 \end{cases}$$

其中  $\gamma$  是调整参数。上述模型中,  $G$  不必是模型分布族  $F_{\theta}$  中的一个分布,也可以进行估计。同时当  $\gamma = 0$ , DPD 通过近似化  $\gamma \rightarrow 0^+$  得到,这是 DPD 就是 Kullback-Leibler (KL) 散度。MDPD 方法的一般步骤就是针对参数模型,通过求 DPD 在参数空间  $\Theta$  中关于参数  $\theta$  的最小最小值而得到。当  $\gamma = 0$  时, MDPD 就相当于最大似然估计 (ML), 得到的参数就为最大似然估计量 (MLE)。

将 (5.3) 中的条件密度  $f_{\theta}(\mathbf{y}_i | \mathbf{X}_i)$  作为模型密度代入 DPD 的定义中, 得到基于小区域模型的 DPD 度量

$$\begin{aligned} d_{\gamma}(f_{\theta}, g) &= \int_{\mathbf{x}} \int_{\mathbf{y}} \left\{ f_{\theta}^{1+\gamma}(\mathbf{y} | \mathbf{x}) - \left(1 + \frac{1}{\gamma}\right) f_{\theta}^{\gamma}(\mathbf{y} | \mathbf{x})g(\mathbf{y} | \mathbf{x}) \right. \\ &\quad \left. + \frac{1}{\gamma}g^{1+\gamma}(\mathbf{y} | \mathbf{x}) \right\} h(\mathbf{x}) d\mathbf{x}d\mathbf{y}, \text{ for } \gamma > 0, \quad (5.5) \\ d_0(f_{\theta}, g) &= \int_{\mathbf{x}} \int_{\mathbf{y}} g(\mathbf{y} | \mathbf{x}) \log \left( \frac{g(\mathbf{y} | \mathbf{x})}{f_{\theta}(\mathbf{y} | \mathbf{x})} \right) h(\mathbf{x}) d\mathbf{x}d\mathbf{y}. \end{aligned}$$

其中  $h(\mathbf{x})$  是  $\mathbf{X}$  的边缘密度函数,  $g(\mathbf{y} | \mathbf{x})$  是在给定  $\mathbf{X}$  的条件下  $\mathbf{Y}$  的真实的条件密度函数, 由于该真实的条件密度是未知的, 通常使用经验似然函数去近似该密度函数, 从而得到  $\gamma > 0$  的情形下的 DPD 度量为

$$\hat{d}_{\gamma}(f_{\theta}, g) = \frac{1}{m} \sum_{i=1}^m \int_{\mathbf{y}} f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y} - \frac{1+\gamma}{m\gamma} \sum_{i=1}^m f_{\theta}^{\gamma}(\mathbf{y}_i | \mathbf{x}_i) + c(\gamma). \quad (5.6)$$

其中公式 (5.6) 的最后一项  $c(\gamma) = \frac{1}{m\gamma} \sum_{i=1}^m \int_{\mathbf{y}} g_i^{1+\gamma}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y}$  与未知参数  $\theta$  是无关的, 因此最小化 (5.6) 中的最后一项可以省略。将 (5.3) 中的密度函数代入

(5.6) 中, 最小化目标函数变为

$$\hat{d}_\gamma(f_\theta, g) = \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \left[ (1+\gamma)^{-\frac{n_i}{2}} - \frac{1+\gamma}{\gamma} \exp\left[-\frac{\gamma}{2} B_i\right] \right]. \quad (5.7)$$

其中  $B_i = (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})$ , 结合公式 (5.4) 可得

$$B_i = \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})^2 - \frac{\rho_i}{\sigma_e^2 a_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}) \right\}^2$$

通过求  $\hat{d}_\gamma(f_\theta, g)$  关于  $\boldsymbol{\theta} \in \Theta$  的最小值, 可得  $\boldsymbol{\theta}$  的 MDPD 估计量 (MDPDE)。显然, 若第  $i$  个观测值为离群值, 那么条件密度  $f_\theta(\mathbf{y}_i | \mathbf{x}_i)$  的值比其它观测点的值要小。如此一来, 当  $\gamma > 0$  时式 (5.6) 中的第二项就会很小, 从而 MDPDE 对离群观测值比较稳健, 受离群值影响减小。 $\gamma$  作为调整参数, 控制 MDPDE 的有效性和稳健性之间的平衡。 $\gamma$  增加时, 稳健性增加, 有效性降低, 反之, 稳健性降低而有效性增加。另外, 当  $\gamma$  等于 0 时, 密度幂散度退化成 KL 散度, MDPDE 退化成了 MLE, 此时对于异常观测, 当  $f_\theta(\mathbf{y}_i | \mathbf{x}_i)$  趋于 0 时 KL 散度是发散的, 同时 MLE 方法也就失效了。因此, MLE 在异常值存在时失效, 因为异常值在损失函数占主导地位。

为了求解模型参数, 对方程 (5.7) 中的  $\hat{d}_\gamma(f_\theta, g)$  关于  $\boldsymbol{\theta}$  求偏导数, 得到如下的参数估计方程:

$$\begin{aligned} & \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \exp\left(-\frac{\gamma}{2} B_i\right) \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}) \mathbf{x}_{ij}^T \\ &= \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \exp\left(-\frac{\gamma}{2} B_i\right) \rho_i \mathbf{x}_i^T \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}), \end{aligned} \quad (5.8)$$

$$\begin{aligned} & \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}-1} \left( \gamma(1+\gamma)^{-\frac{n_i+2}{2}} - \exp\left(-\frac{\gamma}{2} B_i\right) \right) a_i \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2 \\ &= - \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \exp\left(-\frac{\gamma}{2} B_i\right) \frac{1}{(a_i \sigma_v^2 + \sigma_e^2)^2} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}) \right)^2, \end{aligned} \quad (5.9)$$

$$\begin{aligned} & \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}-1} \left( \gamma(1+\gamma)^{-\frac{n_i+2}{2}} - \exp\left(-\frac{\gamma}{2} B_i\right) \right) (n_i \sigma_e^{2(n_i-1)} + (n_i - 1) a_i \sigma_v^2 \sigma_e^{2(n_i-2)}) \prod_{j=1}^{n_i} k_{ij}^2 \\ &= \sum_{i=1}^m (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \exp\left(-\frac{\gamma}{2} B_i\right) \left\{ -\frac{1}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta})^2 \right. \\ & \quad \left. + \frac{2\sigma_e^2 \sigma_v^2 + a_i \sigma_v^4}{(a_i \sigma_v^2 + \sigma_e^2)^2 \sigma_e^4} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta}) \right)^2 \right\}. \end{aligned} \quad (5.10)$$

上述估计方程给出了模型未知参数满足的关系式，本研究利用迭代 N-R 算法，求出上述估计方程的解，得到参数的估计。在使用迭代算法的过程中，同样可以将 ML 估计作为初始值，进行迭代运算。

### 5.3 模型参数的渐近分布

在本节中，我们讨论模型参数稳健估计的渐近分布，其中观测数据的真实分布  $G(\mathbf{y} | \mathbf{x})$  不必在模型的假设分布族中。定义得分函数为  $\mathbf{u}_\theta(\mathbf{y}_i | \mathbf{x}_i) = \frac{\partial}{\partial \theta} \log f_\theta(\mathbf{y}_i | \mathbf{x}_i)$ ，则其可以表示成

$$\mathbf{u}_\theta(\mathbf{y}_i | \mathbf{x}_i) = \left( \mathbf{u}_\beta^T(\mathbf{y}_i | \mathbf{x}_i), \mathbf{u}_{\sigma_e^2}^T(\mathbf{y}_i | \mathbf{x}_i), \mathbf{u}_{\sigma_v^2}^T(\mathbf{y}_i | \mathbf{x}_i) \right)^T. \quad (5.11)$$

根据得分函数的定义和 5.3，可得

$$\begin{aligned} \mathbf{u}_\beta(\mathbf{y}_i | \mathbf{x}_i) &= \frac{\partial}{\partial \beta} \log f_\theta(\mathbf{y}_i | \mathbf{x}_i) = \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij} \beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta), \\ \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) &= \frac{\partial}{\partial \sigma_v^2} \log f_\theta(\mathbf{y}_i | \mathbf{x}_i) = -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta) \right\}^2, \\ \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) &= \frac{\partial}{\partial \sigma_e^2} \log f_\theta(\mathbf{y}_i | \mathbf{x}_i) = -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \\ &\quad + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta)^2 - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta) \right\}^2. \end{aligned} \quad (5.12)$$

为了便于书写和表述，对于  $i = 1, 2, \dots, N$ ，定义

$$\begin{aligned} \mathbf{J}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i) \mathbf{u}_\theta^T(\mathbf{y} | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y} \\ &\quad + \int_{\mathbf{y}} \{ I_\theta(\mathbf{y} | \mathbf{x}_i) - \gamma \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i) \mathbf{u}_\theta^T(\mathbf{y} | \mathbf{x}_i) \} \{ g(\mathbf{y} | \mathbf{x}_i) - f_\theta(\mathbf{y} | \mathbf{x}_i) \} f_\theta^\gamma(\mathbf{y} | \mathbf{x}_i) d\mathbf{y}, \\ \mathbf{K}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i) \mathbf{u}_\theta^T(\mathbf{y} | \mathbf{x}_i) f_\theta^{2\gamma}(\mathbf{y} | \mathbf{x}_i) g(\mathbf{y} | \mathbf{x}_i) d\mathbf{y} - \boldsymbol{\xi}^{(i)} \boldsymbol{\xi}^{(i)T}, \\ I_\theta(\mathbf{y} | \mathbf{x}_i) &= -\frac{\partial}{\partial \theta} \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i), \quad \boldsymbol{\xi}^{(i)} = \int_{\mathbf{y}} \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i) f_\theta^\gamma(\mathbf{y} | \mathbf{x}_i) g(\mathbf{y} | \mathbf{x}_i) d\mathbf{y}. \end{aligned}$$

进一步定义  $\mathbf{J} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbf{J}^{(i)}$ ,  $\mathbf{K} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbf{K}^{(i)}$ 。对于 MDPDE 的渐近分布，需要如下的假设条件：

1. 真实的密度函数  $g(y | x)$  在实数集  $\mathbb{R}$  上有支撑；
2. 存在开集  $\omega \in \Theta_0$  包含最优拟合函数  $\theta$ ，使得  $\mathbf{J}$  对任意的  $\theta \in \omega$  都是正定的；

3. 存在函数  $M_{jkl}(x, y)$  使得  $|\partial^3 \exp[(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T \mathbf{V}_i^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})] / \partial \theta_j \partial \theta_k \partial \theta_l| \leq M_{jkl}(x, y)$  对所有的  $\boldsymbol{\theta} \in \omega$ , 其中  $\int_x \int_y |M_{jkl}(x, y)| g(y | x) h(x) dy dx < \infty$  对所有的  $j, k$  和  $l$  成立。

在上述条件下, 模型参数满足如下的渐近分布。

**定理 5.1** 在正则条件 (1)-(3) 下, 当  $m \rightarrow \infty$  时, 以概率 1 存在  $\hat{\boldsymbol{\theta}}$ , 使得

1.  $\hat{\boldsymbol{\theta}}$  一致收敛于  $\boldsymbol{\theta}$ ;

2.  $\hat{\boldsymbol{\theta}}$  的渐近分布为

$$\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N_{p+2}(0, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1}).$$

定理的详细证明可参见附录中的证明5.

若真实的分布  $g(\mathbf{y} | \mathbf{x})$  属于模型分布族  $f_{\boldsymbol{\theta}}(\mathbf{y} | \mathbf{x})$  对  $\boldsymbol{\theta} \in \Theta$  成立, 那么

$$\begin{aligned} \mathbf{J}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y} | \mathbf{x}_i) \mathbf{u}_{\boldsymbol{\theta}}^T(\mathbf{y} | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y}, \\ \mathbf{K}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y} | \mathbf{x}_i) \mathbf{u}_{\boldsymbol{\theta}}^T(\mathbf{y} | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{2\gamma+1}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y} - \boldsymbol{\xi}^{(i)} \boldsymbol{\xi}^{(i)T}, \\ \boldsymbol{\xi}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{y} | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{\gamma+1}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y}. \end{aligned} \quad (5.13)$$

结合式 (5.12) 和 (5.13), 可推导出  $\mathbf{J}^{(i)}$  中的元素, 详细计算及结果可参见附录。对称矩阵  $\mathbf{J}^{(i)}$  可分解为

$$\mathbf{J}^{(i)} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}}^{(i)} & \mathbf{J}_{\boldsymbol{\beta}, \sigma_v^2}^{(i)} & \mathbf{J}_{\boldsymbol{\beta}, \sigma_e^2}^{(i)} \\ \cdot & \mathbf{J}_{\sigma_v^2}^{(i)} & \mathbf{J}_{\sigma_v^2, \sigma_e^2}^{(i)} \\ \cdot & \cdot & \mathbf{J}_{\sigma_e^2}^{(i)} \end{bmatrix}$$

结合方程 (5.12) 和 (5.13),  $\mathbf{J}^{(i)}$  中的元素可以通过积分、求期望等方式得到。这些详细的推导过程均在附录中给出。同样, 在附录中给出了  $\boldsymbol{\xi}^{(i)}$  的分量的推导公式。且  $\boldsymbol{\xi}^{(i)}$  可分解为  $\boldsymbol{\xi}^{(i)} = (\boldsymbol{\xi}_{\boldsymbol{\beta}}^{(i)T}, \boldsymbol{\xi}_{\sigma_v^2}^{(i)}, \boldsymbol{\xi}_{\sigma_e^2}^{(i)})^T$ 。

将矩阵  $\mathbf{J}^{(i)}$  看成是  $\gamma$  的函数, 即记作  $\mathbf{J}^{(i)} \equiv \mathbf{J}^{(i)}(\gamma)$ , 根据  $\mathbf{K}^{(i)}$  和  $\mathbf{J}^{(i)}$  的表达式 (5.13), 可得

$$\mathbf{K}^{(i)} = \mathbf{J}^{(i)}(2\gamma) - \boldsymbol{\xi}^{(i)} \boldsymbol{\xi}^{(i)T}.$$

因此,  $\mathbf{K}$  可以写成如下形式

$$\mathbf{K} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{J}^{(i)}(2\gamma) - \boldsymbol{\xi}^{(i)} \boldsymbol{\xi}^{(i)T}.$$

通过上述协方差矩阵的计算可知,  $\gamma$  增大时参数方差增大, 这说明 MDPDE 的效率随  $\gamma$  增大而减小。这一点进一步验证了优化参数  $\gamma$  用于控制 MDPDE 的效率和鲁棒性之间的权衡, 当  $\gamma$  增加时稳健性增强效率降低, 后面的模拟实验表明这种效率的损失并不大。

#### 5.4 有限总体的稳健 EB 估计量

本节讨论有限总体的目标变量函数的 EB 估计量。假设有限总体  $P$  包含  $N$  个单元, 一个来自  $P$  的容量为  $n$  的样本  $s$ 。用  $\mathbf{y}^P$  记作目标变量在总体中的单元值构成向量, 假设它是随机的, 具有给定的联合概率分布。用  $\mathbf{y}_s$  记作由抽样单元构成的  $\mathbf{y}^P$  的子向量,  $\mathbf{y}_r$  为未抽样单元构成的子向量, 而且不失一般性地假设  $\mathbf{y}^P$  的前  $n$  个单元为样本元素, 即  $\mathbf{y}^P = (\mathbf{y}_s^T, \mathbf{y}_r^T)^T$ 。

在 NER 模型中, 假设观测的第  $i$  个区域第  $j$  个单元的目标变量  $y_{ij}$  服从 (5.1) 中的 NER 模型。同时假设  $\mathbf{y}_i$  在辅助信息  $\mathbf{X}_i$  之下服从正态分布, 即  $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i)$ 。若将 (5.2) 划分为抽样部分和未抽样部分, 则 NER 模型可写成:

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{is} \\ \mathbf{y}_{ir} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{is} \\ \mathbf{X}_{ir} \end{bmatrix} \boldsymbol{\beta} + v_i \begin{bmatrix} \mathbf{1}_{is} \\ \mathbf{1}_{ir} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{is} \\ \mathbf{e}_{ir} \end{bmatrix}$$

其中下标带  $r$  的表示未抽样的单元构成的向量, 其协方差矩阵同样可以分解为:

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{V}_{is} & \mathbf{V}_{isr} \\ \mathbf{V}_{irs} & \mathbf{V}_{ir} \end{pmatrix}$$

其中

$$\mathbf{V}_{is} = \mathbf{R}_{is} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T = \text{diag}_{1 \leq j \leq n_i} (k_{ij}^2) \sigma_e^2 + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T,$$

$$\mathbf{V}_{ir} = \mathbf{R}_{ir} + \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}^T = \text{diag}_{n_i < j < N_i} (k_{ij}^2) \sigma_e^2 + \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}^T,$$

$$\mathbf{V}_{isr} = \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{N_i - n_i}^T, \quad \mathbf{V}_{irs} = \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{n_i}^T$$

未抽样单元构成的向量  $\mathbf{y}_{ir}$  服从服从的边缘模型可从总体模型 (5.1) 中推导得到, 即

$$\mathbf{y}_{ir} = \mathbf{X}_{ir}\boldsymbol{\beta} + \mathbf{1}_{N_i - n_i} v_i + \mathbf{e}_{ir}, \quad i = 1, \dots, m. \quad (5.14)$$

向量  $\mathbf{e}_{ir}$  是相互独立的且  $\mathbf{e}_{ir} \sim N(\mathbf{0}_{N_i - n_i}, \sigma_e^2 \mathbf{W}_{ir})$ , 其中  $\mathbf{W}_{ir} = \text{diag}_{n_i < j < N_i} (k_{ij}^2)$ 。因此向量  $\mathbf{y}_{ir}$  是相互独立服从正态分布

$$\mathbf{y}_{ir} \sim N(\boldsymbol{\mu}_{ir}, \mathbf{V}_{ir}),$$

其中  $\boldsymbol{\mu}_{ir} = \mathbf{X}_{ir}\boldsymbol{\beta}$ .

在给定样本数据  $\mathbf{y}_{is}$  的条件下, 未抽样单元向量  $\mathbf{y}_{ir}$  的分布为

$$\mathbf{y}_{ir} | \mathbf{y}_s \sim \mathbf{y}_{ir} | \mathbf{y}_{is} \sim N(\boldsymbol{\mu}_{ir|s}, \mathbf{V}_{ir|s}). \quad (5.15)$$

其条件均值为

$$\boldsymbol{\mu}_{ir|s} = \boldsymbol{\mu}_{ir} + \mathbf{V}_{irs}\mathbf{V}_{is}^{-1}(\mathbf{y}_{is} - \boldsymbol{\mu}_{is}) = \mathbf{X}_{ir}\boldsymbol{\beta} + \sigma_v^2 \mathbf{1}_{N_i-n_i} \mathbf{1}'_{n_i} \mathbf{V}_{is}^{-1}(\mathbf{y}_{is} - \mathbf{X}_{is}\boldsymbol{\beta}),$$

条件协方差矩阵为

$$\begin{aligned} \mathbf{V}_{ir|s} &= \mathbf{V}_{ir} - \mathbf{V}_{irs}\mathbf{V}_{is}^{-1}\mathbf{V}_{isr} = \sigma_v^2 \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} + \sigma_e^2 \mathbf{W}_{ir} \\ &\quad - \sigma_v^2 \mathbf{1}_{N_i-n_i} \mathbf{1}'_{n_i} \frac{1}{\sigma_e^2} \left( \mathbf{W}_{ir}^{-1} - \frac{\sigma_v^2}{\sigma_e^2 + a_{ir}\sigma_v^2} \mathbf{a}_{ir}\mathbf{a}_{ir}^T \right) \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}'_{N_i-n_i} \\ &= \sigma_v^2 \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} + \sigma_e^2 \mathbf{R}_{ir} - \frac{\sigma_v^4}{\sigma_e^2} \left( a_{ir} \cdot \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} - \frac{\sigma_v^2}{\sigma_e^2 + a_{ir}\sigma_v^2} a_{ir}^2 \cdot \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} \right) \\ &= \sigma_v^2 \left( 1 - \frac{\sigma_v^2}{\sigma_e^2} a_{ir} \cdot (1 - \rho_{ir}) \right) \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} + \sigma_e^2 \mathbf{R}_{ir} \\ &= \sigma_v^2 (1 - \rho_{ir}) \mathbf{1}_{N_i-n_i} \mathbf{1}'_{N_i-n_i} + \sigma_e^2 \mathbf{R}_{ir}, \end{aligned}$$

其中  $\mathbf{a}_{ir} = (a_{in_i+1}, \dots, a_{iN_i})^T$ ,  $a_{ir} = \sum_{j=n_i+1}^{N_i} a_{ij}$ ,  $\rho_{ir} = \frac{a_{ir}\sigma_v^2}{\sigma_e^2 + a_{ir}\sigma_v^2}$ .

若  $n_i \neq 0$ , 则对  $j \in U_i - s_i$  时的条件均值为

$$\begin{aligned} \mu_{ij|s} &= x_{ij}\boldsymbol{\beta} + \frac{\sigma_v^2}{\sigma_e^2} \mathbf{1}_{n_i}^T \left( \mathbf{W}_{is}^{-1} - \frac{\sigma_v^2}{\sigma_e^2 + a_{is}\sigma_v^2} \mathbf{a}_{is}\mathbf{a}_{is}^T \right) (\mathbf{y}_{is} - \mathbf{X}_{is}\boldsymbol{\beta}) \\ &= x_{ij}\boldsymbol{\beta} + \frac{\sigma_v^2}{\sigma_e^2} \left( 1 - \frac{a_{is}\sigma_v^2}{\sigma_e^2 + a_{is}\sigma_v^2} \right) \mathbf{a}_{is}^T (\mathbf{y}_{is} - \mathbf{X}_{is}\boldsymbol{\beta}) \\ &= x_{ij}\boldsymbol{\beta} + \rho_{is} (\bar{\mathbf{Y}}_{is}^w - \bar{\mathbf{X}}_{is}^w \boldsymbol{\beta}) \end{aligned}$$

其中  $\bar{\mathbf{Y}}_{is}^w = a_{is}^{-1} \mathbf{a}_{is}^T \mathbf{y}_{is} = a_{is}^{-1} \sum_{j=1}^{n_i} a_{ij} y_{ij}$  和  $\bar{\mathbf{X}}_{is}^w = a_{is}^{-1} \mathbf{a}_{is}^T \mathbf{X}_{is} = a_{is}^{-1} \sum_{j=1}^{n_i} a_{ij} \mathbf{x}_{ij}$ .

对任意  $j \in U_i - s_i$ , 有

$$\mu_{ij|s} = \begin{cases} x_{ij}\boldsymbol{\beta} + \rho_{is} (\bar{\mathbf{Y}}_{is}^w - \bar{\mathbf{X}}_{is}^w \boldsymbol{\beta}) & n_i \neq 0, \\ x_{ij}\boldsymbol{\beta} & n_i = 0, \end{cases} \quad \rho_{is} = \frac{a_{is}\sigma_v^2}{a_{is}\sigma_v^2 + \sigma_e^2}$$

对任意的  $j \in U_i - s_i$ , 条件方差为

$$v_{ij|s} = \begin{cases} \sigma_v^2 (1 - \rho_{is}) + k_{ij}^2 \sigma_e^2 & n_i \neq 0 \\ \sigma_v^2 + k_{ij}^2 \sigma_e^2 & n_d = 0 \end{cases}$$

一般情况下, 抽样调查的目标是利用现有的样本数据  $\mathbf{y}_s$  来估计关于总体向量  $\mathbf{y}^p$  的实值可测函数  $\tau = h(\mathbf{y}^p)$  的值。因此, 在给定  $\mathbf{y}_s$  的条件下未抽样单元

$\mathbf{y}_r$  的条件分布在计算总体目标的函数  $\tau = h(\mathbf{y})$  的最佳估计 (BP) 时是至关重要的。假设模型参数  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  是已知的, 在单元水平模型下, BP 估计量  $\hat{\tau} = \hat{h}(\mathbf{y}_s)$  是最小化 MSE 时关于  $\tau$  的无偏估计量, 即  $\hat{\tau}^{bp} = E_{\mathbf{y}_r}[\tau | \mathbf{y}_s]$ . 因此, 可以利用公式 (5.12) 得出  $\tau$  的 EB 估计量  $\hat{\tau} = \hat{h}(\mathbf{y}^P) = h(\mathbf{y}_s, \hat{\mathbf{y}}_r)$ , 其中,  $\hat{\mathbf{y}}_r = \boldsymbol{\mu}_{ir|s} + \hat{\mathbf{v}}_{ir|s}$ ,  $\hat{\mathbf{v}}_{ir|s} \sim N(\mathbf{0}, \mathbf{V}_{ir|s})$ .

在实际应用中, 模型参数  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  被其一致性估计量  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\sigma}_v^2, \hat{\sigma}_e^2)^T$  所替代, 同时变量  $\hat{\mathbf{y}}_r$  可由分布 (5.15) 生成。则  $\tau_i$  的经验贝叶斯估计 (EBP) 为:

$$\hat{\tau}_i^{ebp} = \frac{1}{N_i} \left\{ \sum_{j \in s_i} h(y_{ij}) + \sum_{j \in r_i} E_{\mathbf{y}_r} \left[ h(y_{ij}) | \mathbf{y}_s; \hat{\boldsymbol{\theta}} \right] \right\}. \quad (5.16)$$

一般来说, 函数  $h$  可能是不可解析表示的, 或者当  $n_i$  较小时, 上式的期望的计算可能会涉及多重积分, 因此上式直接求期望得到表达式通常是不现实的。通常使用蒙特卡洛 (Monte Carlo) 方法来计算  $\tau_i$  的 EBP 值。对于未抽样的样本  $y_{ij}, j \in r_i$ , 生成  $A$  个  $\tau_i^{(a)} = \hat{\mu}_{ij|s} + \hat{v}_{ij|s}, a \in 1, \dots, A$ , 其中  $\hat{\mu}_{ij|s}$  和  $\hat{v}_{ij|s}$  由式 (10) 中的分布生成。则  $\tau_i$  的 EBP 的蒙特卡洛近似估计为:

$$\hat{\tau}_i^{ebp} \approx \frac{1}{A} \sum_{a=1}^A \tau_i^{(a)}, \quad \tau_i^{(a)} = \frac{1}{N_i} \left\{ \sum_{j \in s_i} h(y_{ij}) + \sum_{j \in r_i} h(y_{ij}^{(a)}) \right\}. \quad (5.17)$$

## 5.5 区域均值的 EB 估计量

本节中给出区域均值  $\bar{Y}_i$  的稳健经验贝叶斯估计量 (REBP), 其中  $\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$ . 令  $\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2$  和  $\hat{\sigma}_e^2$  是模型参数  $\boldsymbol{\beta}, \sigma_v^2$  和  $\sigma_e^2$  的一致估计。在条件分布 (5.15) 下, 目标变量的估计值为

$$\hat{\mathbf{y}}_{is}^{ebp} = \mathbf{y}_{is}, \quad \hat{\mathbf{y}}_{ir}^{ebp} = \hat{\boldsymbol{\mu}}_{ir|s} = \mathbf{X}_{ir} \hat{\boldsymbol{\beta}} + \hat{\sigma}_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}'_{n_i} \hat{\mathbf{V}}_{is}^{-1} (\mathbf{y}_{is} - \mathbf{X}_{is} \hat{\boldsymbol{\beta}}). \quad (5.18)$$

或等价地写成

$$\hat{y}_{ij}^{ebp} = \begin{cases} y_{ij} & \text{if } j \in s_i. \\ \mathbf{x}_{ij} \hat{\boldsymbol{\beta}} + \rho_{is} (\bar{Y}_{is}^w - \bar{\mathbf{X}}_{is}^w \boldsymbol{\beta}) & \text{if } j \in r_i = U_i - s_i. \end{cases}$$

区域均值  $\bar{Y}_d$  的 EBP 为

$$\begin{aligned} \hat{\bar{Y}}_i^{ebp} &= \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{y}_{ij}^{ebp} = \frac{1}{N_i} \sum_{j \in s_i} y_{ij} + \frac{1}{N_i} \sum_{j \in r_i} \{ \mathbf{x}_{ij} \hat{\boldsymbol{\beta}} + \hat{v}_i \} \\ &= f_i \hat{\bar{Y}}_i + \frac{1}{N_i} \sum_{j \in U_i} \{ \mathbf{x}_{ij} \hat{\boldsymbol{\beta}} + \hat{v}_i \} - f_i \frac{1}{n_i} \sum_{j \in s_i} \{ \mathbf{x}_{ij} \hat{\boldsymbol{\beta}} + \hat{v}_i \} \end{aligned}$$

$$= (1 - f_i) [\bar{\mathbf{X}}_i \hat{\boldsymbol{\beta}} + \hat{v}_i] + f_i [\hat{Y}_i + (\bar{\mathbf{X}}_i - \hat{\bar{\mathbf{X}}}_i) \hat{\boldsymbol{\beta}}] \quad (5.19)$$

其中  $f_i = \frac{n_i}{N_i}$  为区域上样本抽样比例,  $\hat{Y}_i = \frac{1}{n_i} \sum_{j \in s_i} y_{ij}$ ,  $\hat{\bar{\mathbf{X}}}_i = \frac{1}{n_i} \sum_{j \in s_i} \mathbf{x}_{ij}$ , 和  $\bar{\mathbf{X}}_i = \frac{1}{N_i} \sum_{j \in U_i} \mathbf{x}_{ij}$ 。

## 5.6 调整参数的选取

从上述的结果中不难发现, 利用 DPD 方法迭代得到的参数  $\theta$  的估计表达式中, 含有未知的调整参数  $\gamma$ . 而  $\gamma$  的选取决定着稳健性和有效性的平衡, 当  $\gamma$  的取值越接近于 1, 则估计得到的参数具有更强的稳健性, 反之, 稳健性越弱, 有效性越强。因此选取适当的调整参数  $\gamma$ , 是稳健估计中很关键的因素。目前关于调整参数选取的方法主要有两个方面, 一个是根据有效性和稳健性的占比关系, 研究者根据自己的需求确定两者之间的占比, 再在此基础上, 选取最优的调整参数。例如 Sugawara et al.<sup>[144]</sup> 在利用 DPD 方法进行区域水平的估计时, 利用该方法选取调整参数。另一种方法是基于数据驱动的参数选择方法, Warwick et al.<sup>[92]</sup> 通过最小化估计参数的 MSE 得到最优的调整参数, 但是该方法依赖于初始值的选取, 不同的参数初始值可能会产生不同的调整参数。Basak et al.<sup>[93]</sup> 在 [92] 的基础上, 提出了不依赖于初始值的调整参数选择方法, 在本研究中, 我们将采用 Basak et al.<sup>[93]</sup> 中提到的参数选取方法, 选择最优的调整参数, 用于构建稳健小区域估计量。

对于未知参数  $\theta$  的真实值  $\theta^*$ , 最优的调整参数  $\gamma$  通过最小化 MDPD 估计量的 MSE 来获得, 即

$$E \left\{ \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^* \right)^T \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^* \right) \right\} = N^{-1} \text{tr} \left\{ J_\gamma^{-1}(\boldsymbol{\theta}_\gamma) K_\gamma(\boldsymbol{\theta}_\gamma) J_\gamma^{-1}(\boldsymbol{\theta}_\gamma) \right\} + \left( \boldsymbol{\theta}_\gamma - \boldsymbol{\theta}^* \right)^T \left( \boldsymbol{\theta}_\gamma - \boldsymbol{\theta}^* \right), \quad (5.20)$$

由公式 (5.20) 易知, 公式中含有未知参数  $\theta^*$ , 此时关于最优调整参数  $\gamma$  的选取通常有两种方法. 一种方法是认为估计  $\hat{\boldsymbol{\theta}}_\gamma$  就是参数  $\theta$  的真实值, 即只要求上式第一项的最小值即可得到最优的调整参数。这种方法易于操作, 但是由于  $\hat{\boldsymbol{\theta}}_\gamma$  是渐近趋向于  $\theta$  的, 因此直接替换会产生较大的误差。另一种方法是给定一个  $\theta^*$  的初始值  $\theta_p$ , 再最小化 5.20 求解最优的参数, Warwick et al.<sup>[92]</sup> 就是利用这种方法, 并给出了 Wick-Jones(WJ) 算法。然而, WJ 算法严重依赖于初始  $\theta_p$  的选取, 初始参数的选择直接决定了最优参数的选取。为了克服上述两个选取方法的不足, Basak et al.<sup>[93]</sup> 提出了迭代 WJ 算法 (IWJ), 用来计算最优的调整参数, 本研究中, 同样利用这种方法选取最优调整参数。具体的算法步骤如下

Warwick et al.<sup>[92]</sup> 利用一步 WJ 算法, 给出了最优调整参数的选择。具体的算法如下:

---

**算法 3** IWJ 算法
 

---

**输入:** 设置初始值  $\gamma, \boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \sigma_e^{2(0)}, \sigma_v^{2(0)})$

**重复:**

1: 利用式 (5.20) 中的 WJ 算法在区间  $I_\gamma$  通过如下关系更新  $\gamma$ ,

$$\gamma^{(i+1)} = \min_{\gamma \in I_\gamma} E \left\{ \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^{*(i)} \right)^T \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^{*(i)} \right) \right\}.$$

2: 固定  $\gamma^{(i+1)}$ , 将其代入 MDPD 迭代步骤, 得到参数  $\hat{\boldsymbol{\theta}}$  的估计, 并进行更新  $\boldsymbol{\theta}^{*(i+1)} \leftarrow \text{MDPD}(\boldsymbol{\theta}^{*(i+1)})$ .

3: 重复步骤 1,2, 直至  $|\gamma^{(i+1)} - \gamma^{(i)}| < \epsilon$  或  $|\boldsymbol{\theta}^{*(i+1)} - \boldsymbol{\theta}^{*(i)}| < \epsilon^*$ , 其中  $\epsilon$  和  $\epsilon^*$  表示参数估计中的要求精度。

**输出:**  $\boldsymbol{\theta}^{(m+1)}$

---

## 5.7 稳健估计量的 MSE

估计单元水平模型的均方预测误差通常是比较困难的, 导致这一困难的可能原因有二: 其一, 单元水平模型中误差项和未抽中样本单元所服从的真实分布并不知道, 从而无法获取密度函数进行 MSPE 的计算, 其二, 有时即便知道未抽样单元的分布, 由于单元水平模型涉及的是单元层级的数据, 在计算其期望的过程中会出现多重积分, 这对 MSPE 的计算造成了极大的挑战。在本研究中, 我们利用 Sinha et al.<sup>[28]</sup>, Hall et al.<sup>[103]</sup> 中提到的参数 Bootstrap 方法来估计 MSPE。具体的估计步骤如下:

1. 利用 MDPD 方法 (5.8,5.9,5.10) 得到参数  $\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2$  和  $\hat{\sigma}_e^2$  的稳健估计;
2. 根据  $v_i^* \stackrel{\text{iid}}{\sim} N(0, \hat{\sigma}_v^2), i = 1, \dots, m$  生成 Bootstrap 区域随机效应; 生成独立于  $v_1^*, \dots, v_m^*$  且服从分布  $e_{ij}^* \stackrel{\text{iid}}{\sim} N(0, \hat{\sigma}_e^2 k_{ij}^2), j = 1, \dots, N_i, i = 1, \dots, m$ . 的单元随机误差  $e_{ij}^*$ 。
3. 根据模型 (5.1) 生成 Bootstrap 总体目标变量

$$y_{ij}^* = \mathbf{x}_{ij}' \hat{\boldsymbol{\beta}} + v_i^* + e_{ij}^*, \quad j = 1, \dots, N_i, \quad i = 1, \dots, m.$$

4. 令  $\mathbf{y}_i^{P*} = (y_{i1}^*, \dots, y_{iN_i}^*)^T$  记作区域  $i$  上的 Bootstrap 目标变量。计算 Bootstrap 总体目标变量的函数  $\tau_i^* = h(\mathbf{y}_i^{P*}), i = 1, \dots, m$ .

5. 利用 Bootstrap 数据  $\{(y_{ij}^*, \mathbf{x}_{ij}); j = 1, \dots, N_i, i = 1, \dots, m\}$  拟合 NER 模型, 并根据 MDPD 方法得到 Bootstrap 拟合模型的参数估计量记为  $\hat{\sigma}_v^{2*}, \hat{\sigma}_e^{2*}$ , 和  $\hat{\beta}^*$ .
6. 利用 (5.17) 得到 Bootstrap EB 估计量  $\tau_i$  并记为  $\hat{\tau}_i^{\text{EB*}}, i = 1, \dots, m$ .
7. 重复步骤 (2)-(7)  $B$  次。令  $\tau_i^*(b)$  为真实的值,  $\hat{\tau}_i^{\text{EB*}}(b)$  为第  $bb = 1, \dots, B$  次重复中利用 Bootstrap 样本得到的 EB 估计量。
8.  $\hat{\tau}_i^{\text{EB}}$  的 Bootstrap MSE 的估计量通过下式给出

$$\text{mse}_B(\hat{\tau}_i^{\text{EB}}) = B^{-1} \sum_{b=1}^B [\hat{\tau}_i^{\text{EB*}}(b) - \tau_i^*(b)]^2.$$

## 5.8 应用举例

在本节中, 将通过模拟数据和实际数据的单元水平模型对 MDPD 方法进行检验, 验证上述估计形式和估计 MSE。和现有的经验最佳线性无偏估计量以及 Sinha et al.<sup>[28]</sup> 中的稳健估计方法进行比较, 进一步检验本研究提出的估计方法的稳健效果。接下来详细阐述数值模拟和实际数据的应用研究。

### 5.8.0.1 数值模拟

本研究中, 为了便于比较, 做和 Sinha et al.<sup>[28]</sup> 相同的模拟实验, 即具有单个辅助变量  $x$  的单元水平模型:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n$$

其中区域数  $k = 40$ , 每个区域上的抽样数  $n = 4$ , 辅助变量  $x_{ij} \sim N(1, 1)$ , 同时假设区域随机效应  $v_i$  和随机误差  $e_{ij}$  均来源于污染分布  $(1 - \eta)N(0, \sigma^2) + \eta N(0, \sigma_1^2)$ 。意味着有  $(1 - \eta)$  比例的误差项是来源于潜在的真实分布  $N(0, \sigma^2)$  和剩余  $\eta$  比例的误差是来源于不同于真实分布的污染分布  $N(0, \sigma_1^2)$ 。当  $\eta = 0$  时表明误差分布没有被污染。对于真实的分布, 假设  $\sigma_e^2 = \sigma_v^2 = 1$  并假设污染分布中  $\sigma_{e1}^2 = \sigma_{v1}^2 = 25$ , 污染比例为  $\eta_1 = \eta_2 = 0.10$ 。本研究考虑四种可能的污染情形  $\{(0, 0), (0, v), (0, e), (v, e)\}$ , 其中  $(0, 0)$  表示模型没有被污染, 其误差分布均为假设的正态分布;  $(0, v)$  表示仅在区域随机效应  $v_i$  中存在污染, 模型随机误差未被污染; 其它情形如此类推。

在模拟实验中，假定  $(\beta_0, \beta_1) = (1, 1)$ ，本研究中对上述四种情形的估计均进行 500 次模拟，并用 Bootstrap 方法计算平均 MSE 的估计值，在进行 Bootstrap 模拟时，分别设置 Bootstrap 样本数量为 100, 200, 和 500 来研究其估计量之间的 MSE 的波动情况。实验中，重点比较传统的极大似然方法 (ML)、Sinha et al.<sup>[28]</sup> 中提到的稳健估计方法 (RML)、以及本研究提出的取不同调整参数 ( $\gamma = 0.1, 0.2, 0.3$ ) 时的稳健最小化密度幂散度的方法 (RMD)。首先比较四种污染情形下的模型参数的估计。

表 5.1 模型参数估计的偏差及其 MSE(a)

参数	ML	RML	RMD1	RMD2	RMD3	RMD.Opt.
无污染		$K=2$	$\gamma=0.1$	$\gamma=0.2$	$\gamma=0.3$	$\bar{\gamma}=0.186$
$\beta_0 = 1$	0.0012	-0.0014	0.0022	0.0033	0.0045	0.0022
	<b>0.0320</b>	0.0325	0.0328	0.0350	0.0384	0.0345
$\beta_1 = 1$	-0.0008	-0.0035	-0.0004	-0.0001	0.0005	-0.0002
	<b>0.0076</b>	0.0079	0.0080	0.0086	0.0095	0.0085
$\sigma_v^2 = 1$	-0.0206	-0.0223	-0.0196	-0.0239	-0.0313	-0.0266
	<b>0.0757</b>	0.0790	0.0781	0.0853	0.0965	0.0760
$\sigma_e^2 = 1$	-0.0012	-0.0008	-0.0027	-0.0020	-0.0002	-0.0002
	<b>0.0160</b>	0.0170	0.0168	0.0184	0.0204	0.0181
$v$ 被污染						$\bar{\gamma}=0.262$
$\beta_0 = 1$	0.0360	0.0284	0.0207	0.0115	0.0088	0.0095
	0.0927	0.0739	0.0502	0.0418	0.0423	<b>0.0414</b>
$\beta_1 = 1$	0.0040	-0.0003	0.0025	0.0013	0.0006	0.0008
	<b>0.0092</b>	0.0094	<b>0.0092</b>	0.0100	0.0110	0.0105
$\sigma_v^2 = 1$	2.2996	1.6601	0.9475	0.4110	0.2555	0.2924
	8.4199	4.2792	1.4907	0.4066	0.2645	<b>0.2479</b>
$\sigma_e^2 = 1$	-0.0036	-0.0091	-0.0010	0.0037	0.0083	0.0049
	<b>0.0175</b>	0.0191	0.0185	0.0211	0.0243	0.0224

表5.1、5.2 中展示了几种稳健估计方法在不同污染情形下的参数估计的偏差和均方误差，其中每个参数对应的第一行表示估计的偏差，第二行表示相应的均方误差。

表 5.2 模型参数估计的偏差及其 MSE(b)

参数	ML	RML	RMD1	RMD2	RMD3	RMD.Opt.
$e$ 被污染						$\bar{\gamma}=0.239$
$\beta_0 = 1$	-0.0095	-0.0134	-0.0078	-0.0089	-0.0096	-0.0090
	0.0429	<b>0.0325</b>	0.0350	0.0368	0.0402	0.0371
$\beta_1 = 1$	0.0061	-0.0031	-0.0069	0.0060	0.0059	0.0058
	0.0280	0.0123	0.0139	0.0117	0.0118	<b>0.0115</b>
$\sigma_v^2 = 1$	-0.0665	-0.1722	-0.0389	-0.0221	-0.0032	-0.0072
	0.1764	0.1099	0.1078	0.1069	0.1176	<b>0.1041</b>
$\sigma_e^2 = 1$	2.4014	0.6680	0.9386	0.4676	0.3373	0.4496
	6.6502	0.5050	1.0269	0.2909	<b>0.1722</b>	0.2642
$(v, e)$ 被污染						$\bar{\gamma}=0.197$
$\beta_0 = 1$	0.0004	-0.0116	0.0009	0.0003	0.00001	0.0007
	0.1044	0.0768	0.0545	<b>0.0475</b>	0.0485	0.0502
$\beta_1 = 1$	-0.0012	-0.0101	-0.0006	0.0009	-0.0018	0.0004
	0.0282	0.0148	0.0146	<b>0.0120</b>	<b>0.0120</b>	0.0138
$\sigma_v^2 = 1$	2.2479	1.5532	1.0189	0.4796	0.3481	0.6205
	8.5019	4.6170	1.7546	0.6007	<b>0.4387</b>	0.7466
$\sigma_e^2 = 1$	2.4730	0.8036	1.0355	0.5038	0.3631	0.7501
	6.9244	0.7273	1.2498	0.3370	<b>0.1963</b>	0.9414

<sup>1</sup> ML, 最大似然估计方法; RML, Sinha et al.<sup>[28]</sup> 中的稳健 ML 方法; RMD1, RMD2, RMD3, 稳健最小密度幂散度方法, 分别表示对应的调整参数选取  $\gamma = 0.1, 0.2, 0.3$ ; RMD.Opt, 通过参数选择算法选取最优调整参数时的 RMD 方法。

从表5.1、5.2中能明显得出如下得结论：在无污染情形中，ML 方法在参数估计中表现最好，但是 RML，以及具有较小调整参数的 RMD 方法和 ML 估计结果非常相近。这说明在这种情形下具有较小调整参数的 RMD 方法、RML 与 ML 方法是几乎一样有效的，其偏差和 MSE 相差都不多。而具有较大调整参数的 RMD 方法表现不好，表明调整参数的选取至关重要，而最优调整参数可根据本研究提供的算法得到。当随机效应  $v$  存在污染的情形下，ML 方法估计得到的方差  $\sigma$  具有大的偏差和 MSE，而 RML 方法的估计中变得较小。而本研究提出的

表 5.3 小域均值的估计偏差和 MSE

情形	ML	RML	RMD1	RMD2	RMD3	RMD.Opt.
		$K=2$	$\gamma=0.1$	$\gamma=0.2$	$\gamma=0.3$	$\bar{\gamma}=0.186$
(0, 0)	0.0061	0.0063	0.0059	0.0057	0.0054	0.0059
	0.3010	0.3052	0.3058	0.3076	0.3110	0.3061
(v, 0)	0.0023	0.0053	0.0043	0.0073	0.0085	0.0076
	0.4144	0.3216	0.3088	0.3546	0.3811	0.3610
(0, e)	0.0069	0.0078	0.0076	0.0076	0.0074	0.0075
	0.6990	0.5086	0.4767	0.4081	0.3899	0.4000
(v, e)	0.0034	0.0077	0.0035	0.0035	0.0041	0.0042
	1.3460	0.4820	0.4575	0.4871	0.4895	0.4618

RMD 方法明显优于 RML 方法，对所有参数的估计均有较小的偏差和 MSE，根据表中模拟数据，当调整参数  $\gamma = 0.026$  时，每个参数的估计效果都比较好。同样地，当模型误差中存在污染时，ML 方法估计得到的  $\sigma_e^2$  具有大的偏差和 MSE，RML 方法对离群值的影响控制较好，每个参数的估计偏差和 MSE 都较小。提出的 RMD 方法中，当  $\gamma = 0.1$  时，结果中模型方差的估计有所减少，但不及 RML 方法好，然而当  $\gamma \geq 0.2$  时，RMD 方法的表现优于 RML 的估计结果。在区域效应和模型误差均存在污染时，ML 对方差部分的估计受离群值观测影响严重以致于产生的很大的偏差和 MSE，RML 方法减小了离群值带来的影响，但表现并非最好，而提出的 RMD 方法显著好于 RML 方法，只需要选取合适的调整参数即可。

接下来，同样利用上述模拟中用到的数据集，我们考虑小区域均值的估计。我们取已知的辅助变量在第  $i$  个区域上的均值为  $\bar{X}_i = 1, i = 1, 2, \dots, m$ 。表5.3展示了几种方法得到的区域均值估计的偏差和 MSE 在区域上的平均值。在未污染数据的情况下，使用 ML 方法得到的 EBLUP 似乎是最有效的，正如预期的那样。利用 RML 方法得到的 REBLUP 估计量、以及本研究提到的 RMD 方法得到的估计量也被认为几乎和 EBLUP 一样有效。在其他三种有污染的情形下，通过表中数据可以得到和上述模拟中一致的结论，ML 方法得到的小区域均值具有大的偏差和 MSPE，RML 方法得到的小区域均值表现稍好，本研究中提到的 RMD 方法表现严格优于 RML 方法。

在未污染数据的情况下,使用 ML 方法获得的 EBLUP 似乎是最有效的,正如预期的那样。使用 RML 方法的 REBLUP 和提出的稳健 MDPDE(RMDPDE)也被认为几乎与 EBLUP 一样有效。在其他三种有异常值的情况下,可以通过表5.3中的数据得到与上述模拟一致的结论,用 ML 方法得到的小面积均值有较大的偏差和 MSPE,用 RML 方法得到的小面积均值性能略好,提出的 RMD 方法性能严格优于 RML 方法。

为了检验现有几种稳健方法对污染比例和污染分布方差的大小的表现,进一步模拟验证估计参数的 MSE 随污染比例、污染分布方差变化情况。同上述模拟步骤一样,考虑区域随机误差的分布被污染,单元随机误差的分布被污染以及单元随机误差和区域随机误差的分布均被污染三种情形下的估计效果。在模拟污染数据时,利用5.8.0.1中的模型生成数据,然后用上述提到的方法进行参数估计。重复模拟 500 次,考虑其 MSE 的平均值,绘制变化曲线图。

第一种情况,考虑随污染比例增加时估计的参数的 MSE 的表现。污染分布的方差被固定为 25,考虑污染比例在 0—0.5 之间以 0.02 为步长变化时,三种污染情形下估计参数的 MSE 的变化情况。另一种情形,考虑污染分布的方差增加时模型参数的 MSE 的表现。在污染分布中,污染比例被确定为 0.1,考虑污染分布的方差在 5—100 以 5 为步长增加时,三种污染情形下估计参数的 MSE 的变化情况。

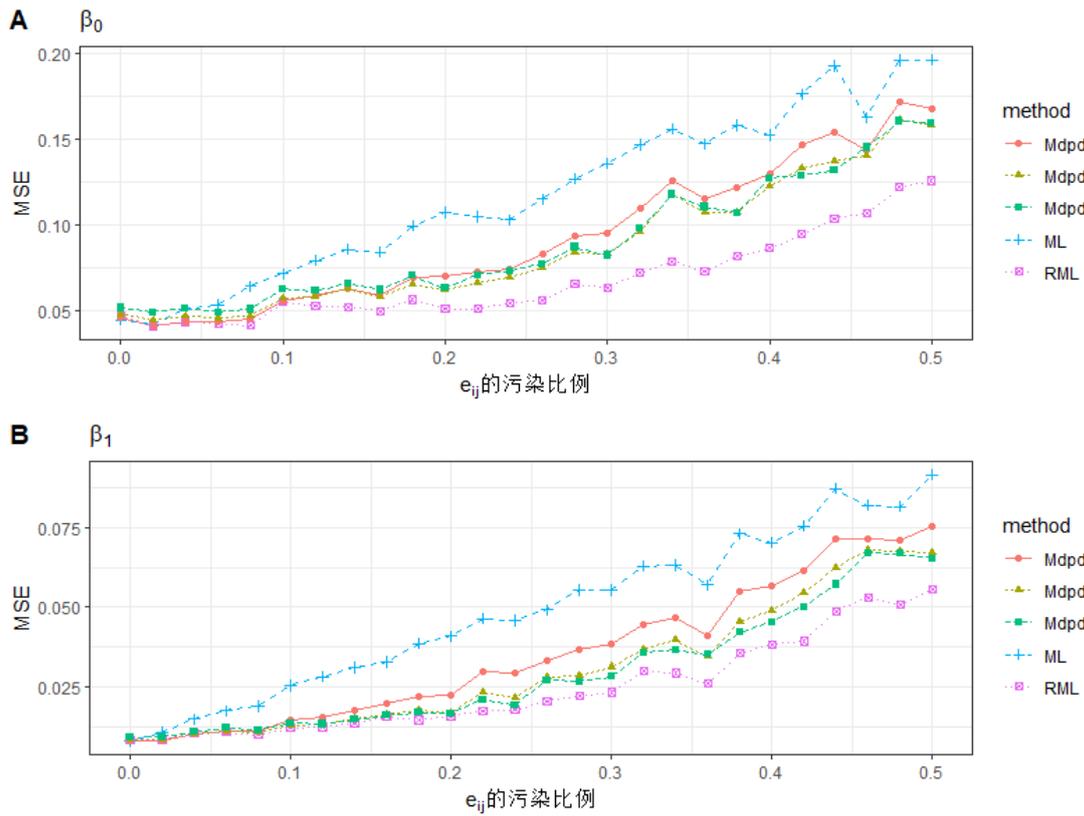


图 5.1 模型系数的 MSE 随  $e_{ij}$  的污染比例变化图

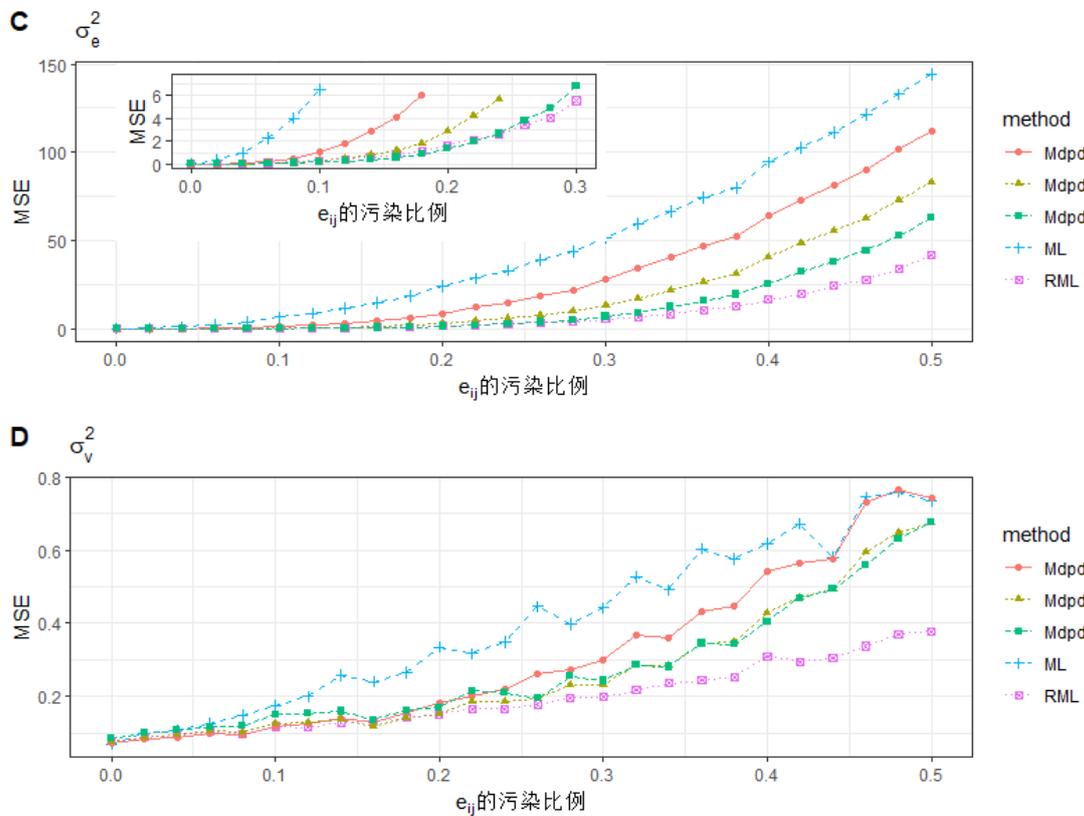


图 5.2 随机误差方差 MSE 随  $e_{ij}$  的污染比例变化图

从上述图形可以发现，当随机效应的污染比例从 0 至 0.5 增加时，小域模型中的四个参数的 MSE 均随污染比例的增加而增加。四个参数的 MSE 中，除了  $\sigma_e^2$  之外，其他的三个参数的 MSE 的变化均不太大，与我们的预期一样， $\sigma_e^2$  容易受  $e_{ij}$  的污染概率的影响，且估计的 MSE 比较大。在几种方法之间的相互比较中，ML 的表现最差，当调整参数  $\gamma$  的取值较小时 (0-0.2) 时，MDPD 方法和 RML 方法表现差不多，有些情形下 MDPD 方法表现更好，但当  $\gamma$  取值较大时，RML 方法要优于 MDPD 方法。

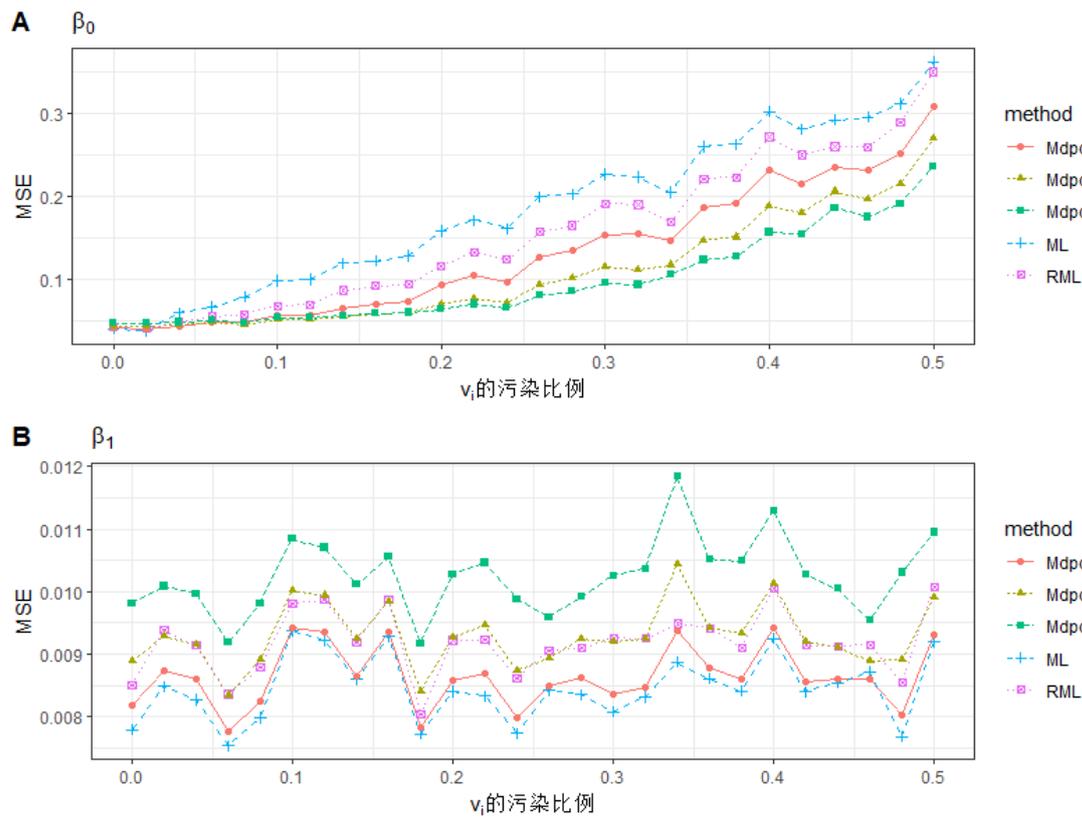


图 5.3 模型系数的 MSE 随  $v_i$  的污染比例变化图

结合图5.3,5.4不难发现：当区域随机效应  $v_i$  被污染时，参数  $\beta_0, \sigma_v^2$  的 MSE 的估计随污染比例的增加而增大，而参数  $\beta_1, \sigma_e^2$  的 MSE 的估计与污染比例的大小无关，且 MSE 较小。比较几种方法对四个参数 MSE 估计的表现， $\beta_1, \sigma_e^2$  的 MSE 均比较小，几种方法的差异不大，而对参数  $\beta_0, \sigma_v^2$  的 MSE 的估计中，ML 方法表现最为糟糕，RML 方法的表现稍好于 ML，几个 MDPD 方法的估计效果显著优于上述两个方法。进一步说明本研究给出的 MDPD 方法是有效的。

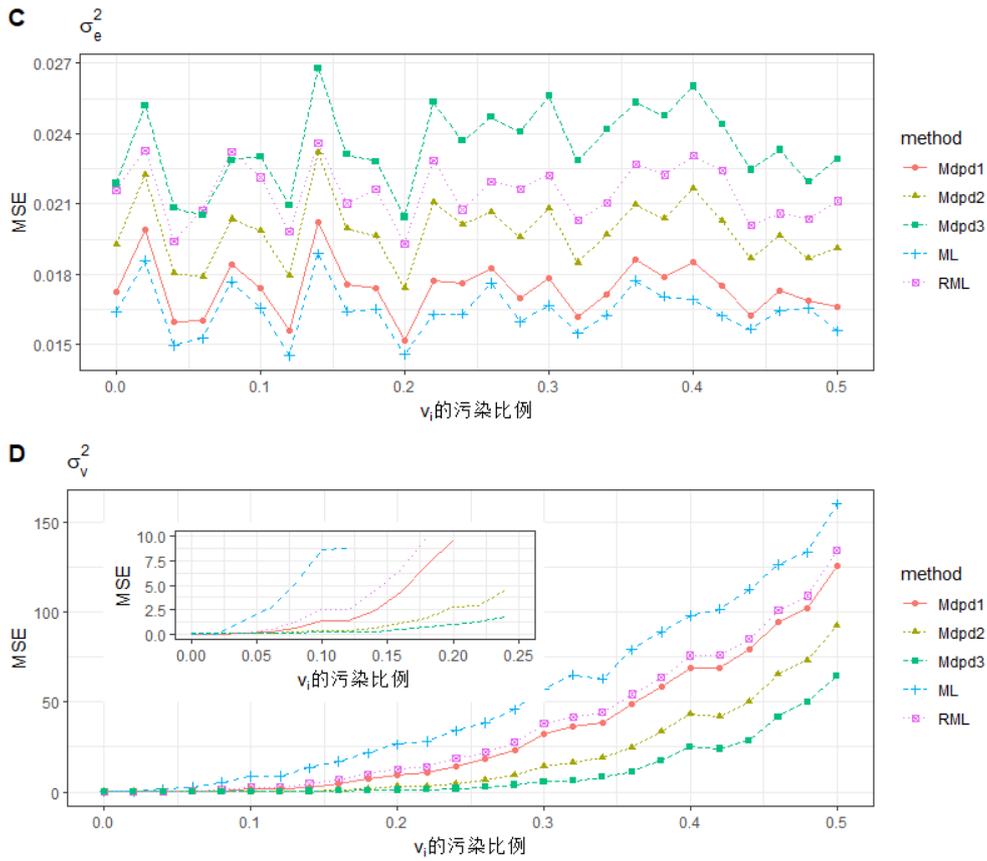


图 5.4 随机误差方差 MSE 随  $v_i$  的污染比例变化图

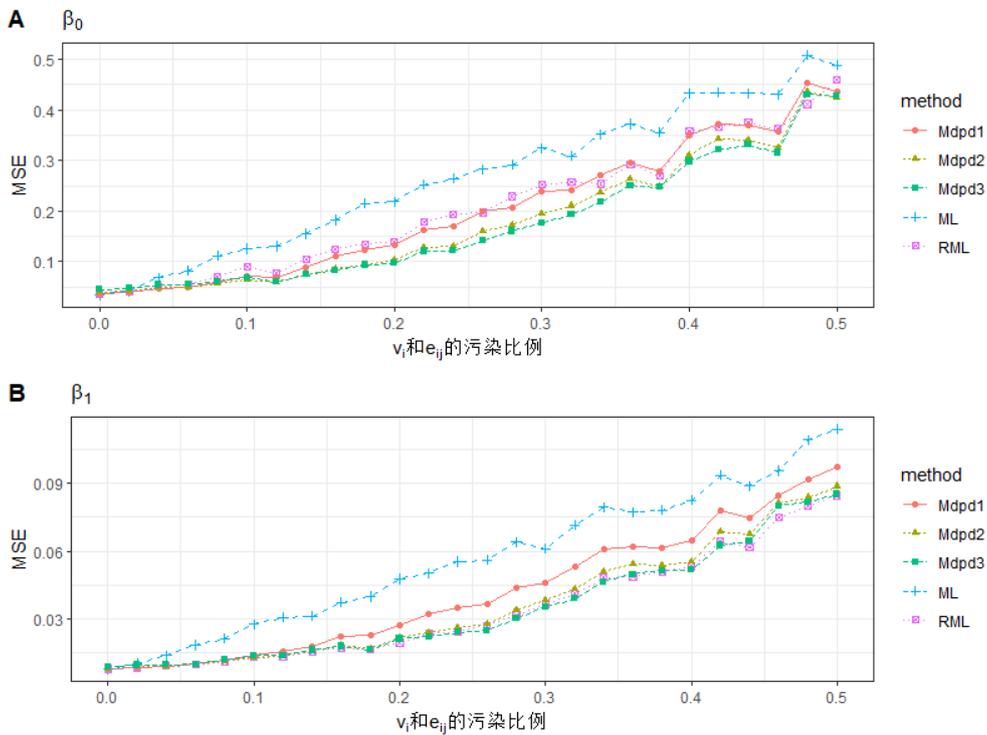


图 5.5 模型系数的 MSE 随  $v_i$  和  $e_{ij}$  的污染比例变化图

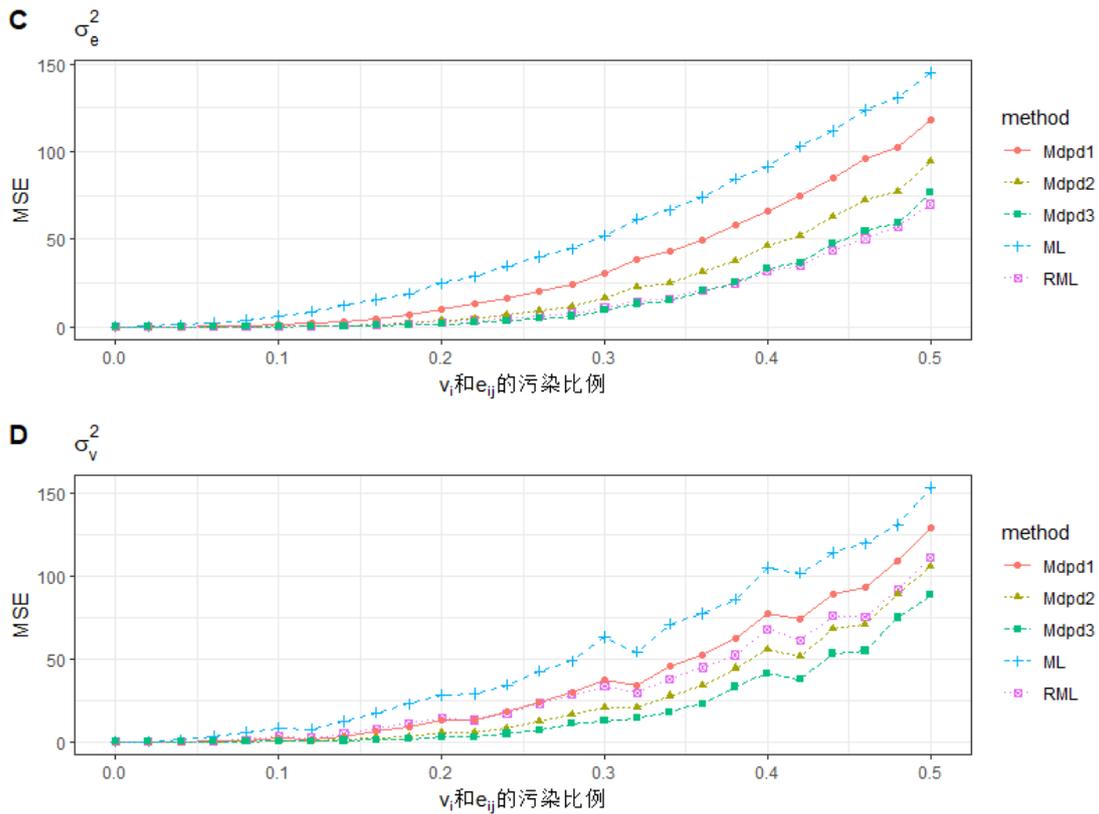


图 5.6 随机误差方差 MSE 随  $v_i$  和  $e_{ij}$  的污染比例变化图

当随机误差  $e_{ij}$  和区域随机误差  $v_i$  均存在污染时，估计参数的 MSE 随污染比例的变化被展示在了图5.5、5.6中。从图中可以看出，当污染概率增大时，参数的 MSE 也随之增大。几种估计方法比较而言，本研究提出的 MDPD 方法显然会优于 RML 和 ML 方法，当然 ML 方法表现最为糟糕。

当模型随机误差  $e_{ij}$  受污染，并且污染分布的方差增大时，四个参数的 MSE 的估计情况被呈现在图5.7,5.8中。从图中可以看到，当污染分布的方差增大时，只有 ML 方法得到的参数的 MSE 呈显著的增大趋势，而其他稳健估计方法得到的参数的 MSE 变化并不显著，说明在这种情形下稳健估计方法一致有效。几类稳健方法比较而言，提出的 MDPD 方法会优于 RML 方法，尤其在  $\sigma_e^2$  的 MSE 的估计中更加明显。

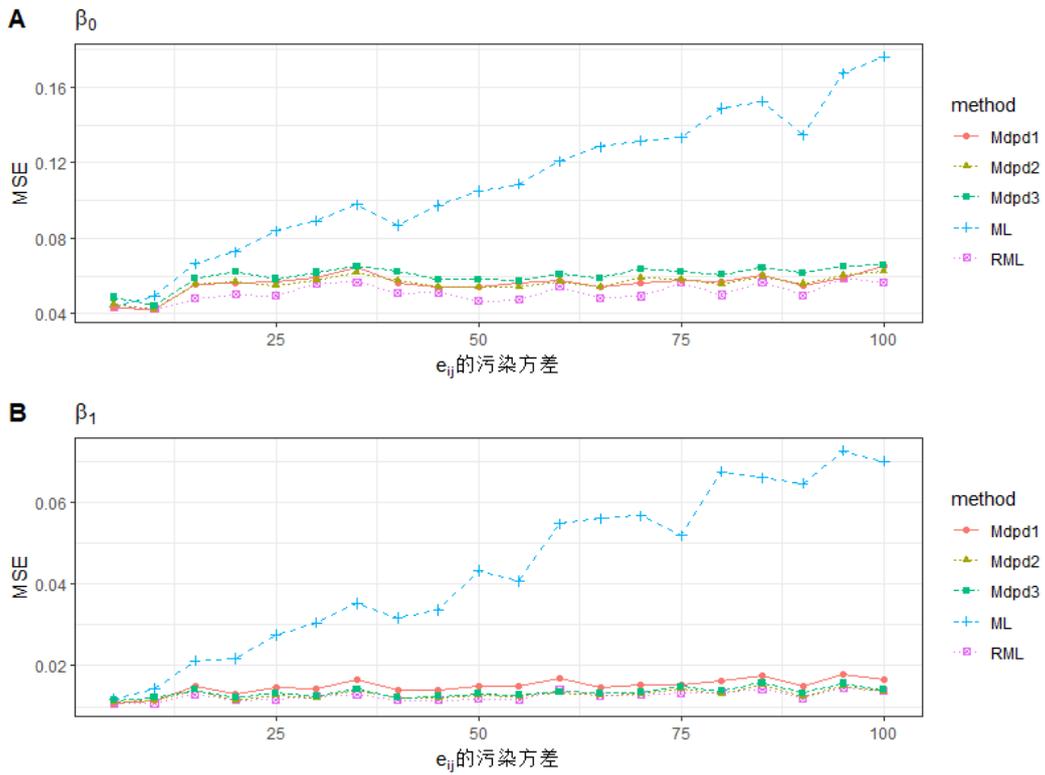


图 5.7 模型系数的 MSE 随  $e_{ij}$  的污染比例变化图

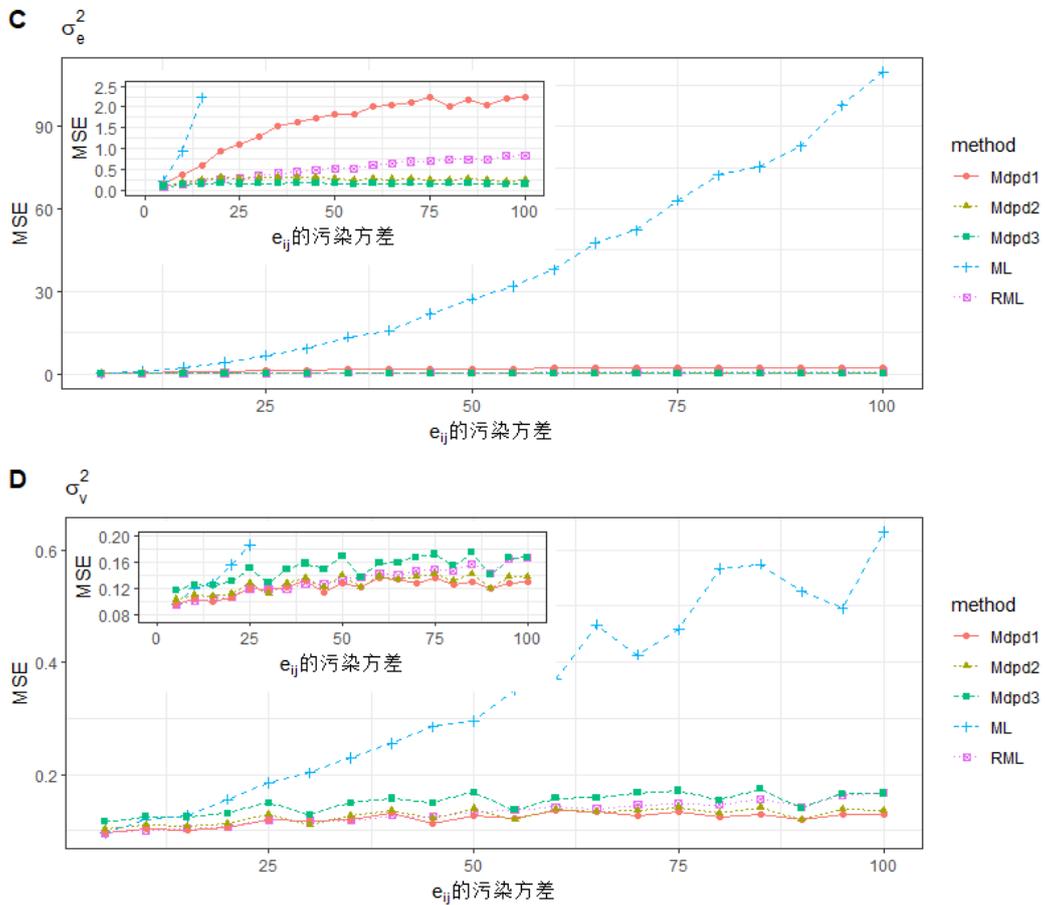


图 5.8 随机误差方差 MSE 随  $e_{ij}$  的污染比例变化图

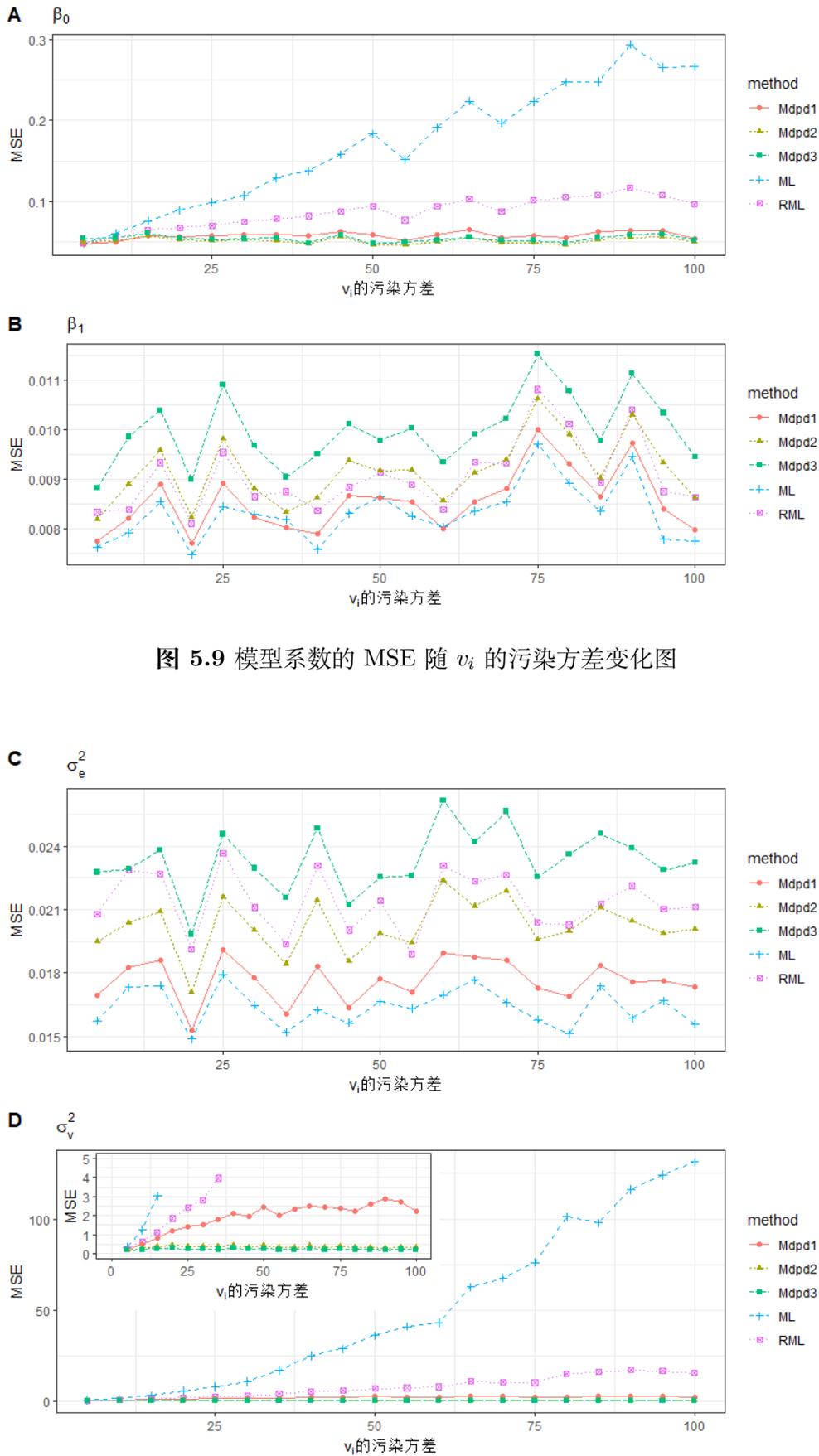


图 5.9 模型系数的 MSE 随  $v_i$  的污染方差变化图

图 5.10 随机误差方差 MSE 随  $v_i$  的污染方差变化图

当区域随机误差被污染，且污染分布的方差在 0—100 变化时，估计参数的 MSE 的变化情况被展示在图形 5.9 和 5.10 中。结合图形可知，参数  $\beta_1, \sigma_e^2$  的 MSE 与污染分布方差的大小无关。参数  $\beta_0, \sigma_v^2$  的 MSE 的估计中，ML 方法受污染分布方差影响较为严重，RML 方法的表现有所改进，但不如 MDPD 方法表现好。

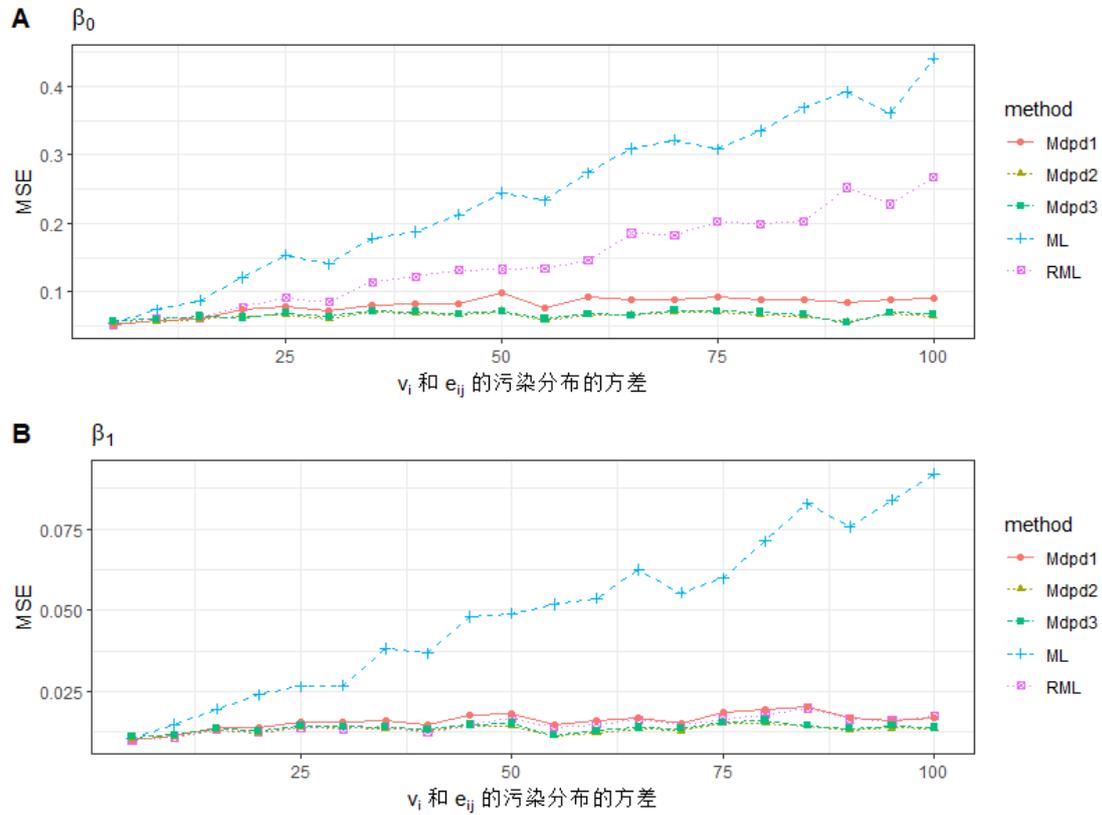


图 5.11 模型系数的 MSE 随  $v_i$  和  $e_{ij}$  的污染方差变化图

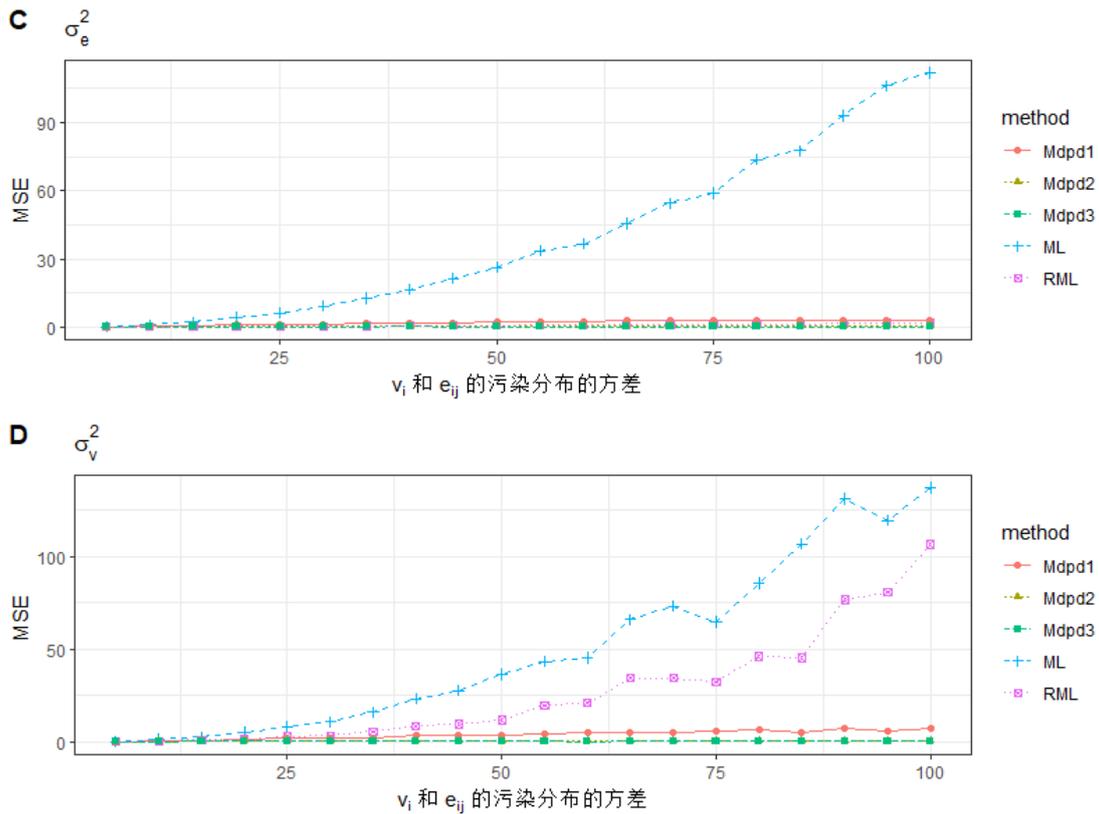


图 5.12 随机误差方差 MSE 随  $v_i$  和  $e_{ij}$  的污染方差变化图

当区域随机误差和模型随机误差项均存在污染，且污染比例为 0.1，污染分布的方差在 0—100 变化时，四个参数的 MSE 的估计被反映在图 5.11 和 5.12 中。由图可知，ML 方法表现最差，RML 方法的估计效果虽有所改善，但对参数  $\beta_0, \sigma_v^2$  的估计效果不好，综合来看，MDPD 方法对几个参数的估计均具有较好的稳健性。

结合上述图形和分组情形下的参数估计的 MSE 的变化趋势图可知。在基本的单元水平模型中，当随机误差是混合正态分布时，模型中会出现离群观测值。表 5.1 和表 5.3 中给出了不同的污染情形下，几种估计方法对模型系数、小域均值的估计 MSE。上述图形给出了当污染分布的污染比例增加，以及污染分布的方差变大时，模型系数的 MSE 的变化趋势。从图中可知，当污染比例增加时，模型中会出现更多的离群值，此时系数的估计均会随着污染比例的增加而增加，导致估计系数的 MSE 增大，估计效果变差。当污染分布的方差增加时，模型中离群值偏离其平均水平的程度将会增加，这类离群值的偏离程度会增加，从而导致系数估计的偏差更大，从图形结果上来看，大多数情形下，当污染分布的方差增加时，系数估计的 MSE 也会随之增加。接下来，聚焦于小域稳健估计量的 MSE

随污染比例和污染分布方差的变化情况。类似于上述图形，绘制几种估计方法在对目标变量估计的平均偏差和 MSE 变化图形，比较几种稳健估计方法的优劣。

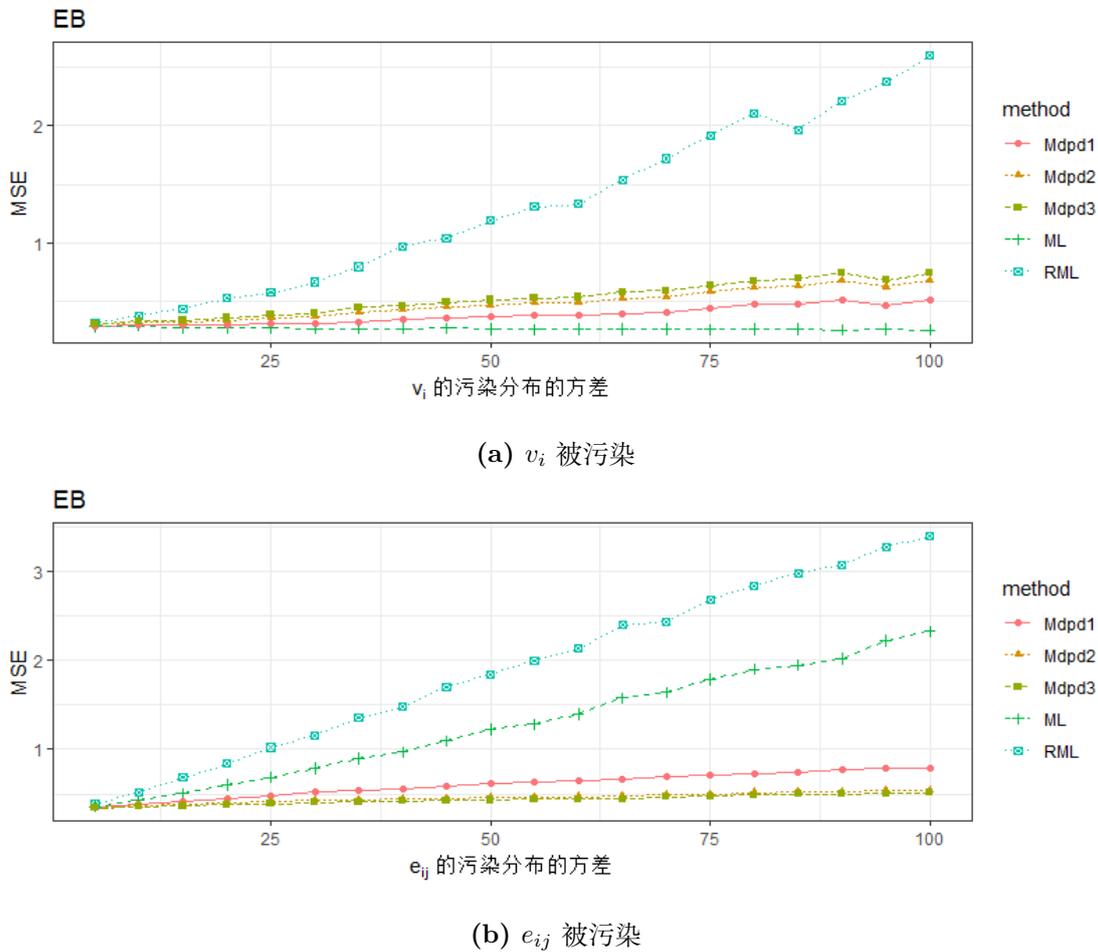


图 5.13 稳健估计的 MSE 随污染方差变化图

类似于上述系数估计的步骤，同样对于三种污染情形下的模拟数据，利用几类稳健估计方法对目标变量进行估计，并计算其平均 MSE。图形5.13绘制了在污染情形  $(v, 0), (0, e)$  下，利用几种稳健估计方法估计目标变量时的平均 MSE 随污染分布方差变化的图形。5.13a展示了当  $v_i$  被污染且污染分布的方差增加时，估计的目标变量的平均 MSE 的变化趋势，从图中可见，RML 方法对污染分布的方差最为敏感，随着污染分布方差的变化，其平均 MSE 呈显著上升趋势。反观其它几类稳健方法，随污染方差的增加，其 MSE 有上升的趋势，但是变化并不明显。其中 ML 方法表现最好。在图形5.13b中，当  $e_{ij}$  被污染且污染分布的方差增加时，几类估计的平均 MSE 均呈上升趋势，其中 RML 方法估计的 MSE 上升最为明显，ML 方法同样呈现出了明显的上升趋势，增加幅度较大，本研究提出的 MDPD 方法表现较为稳定，上升趋势不太显著，且增加的幅度并不大，足

以说明 MDPD 估计方法是有效的稳健估计方法。

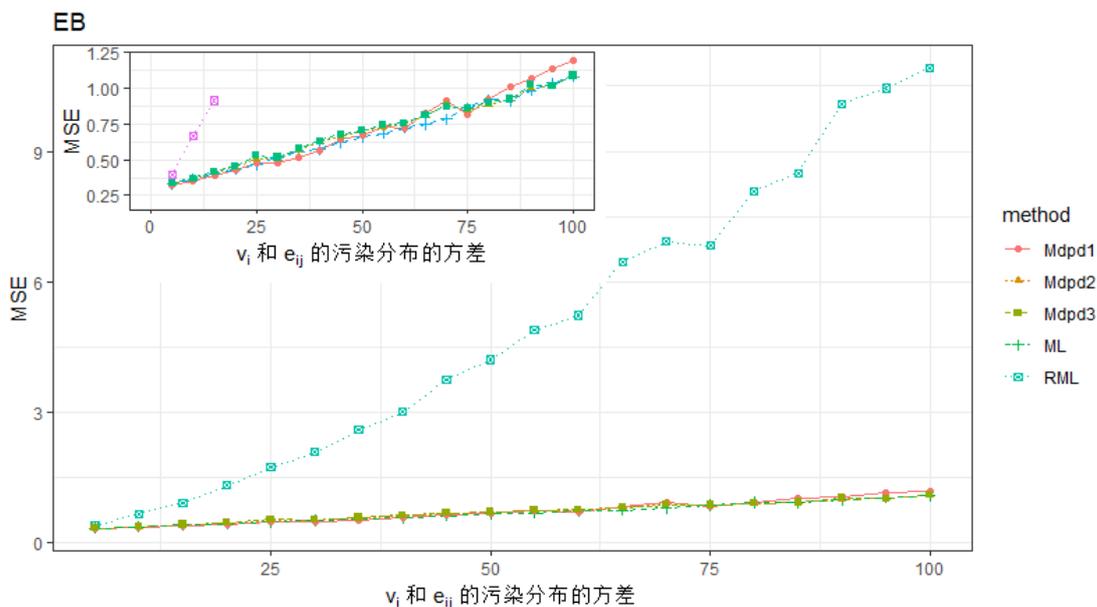
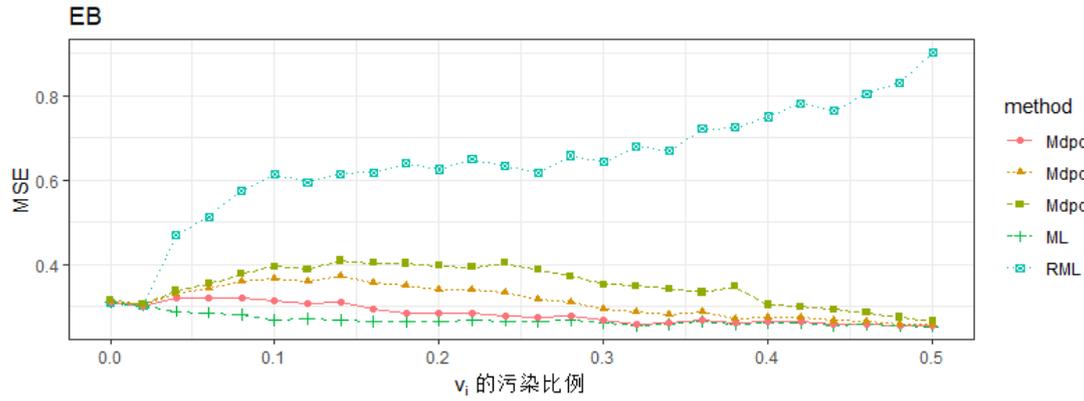


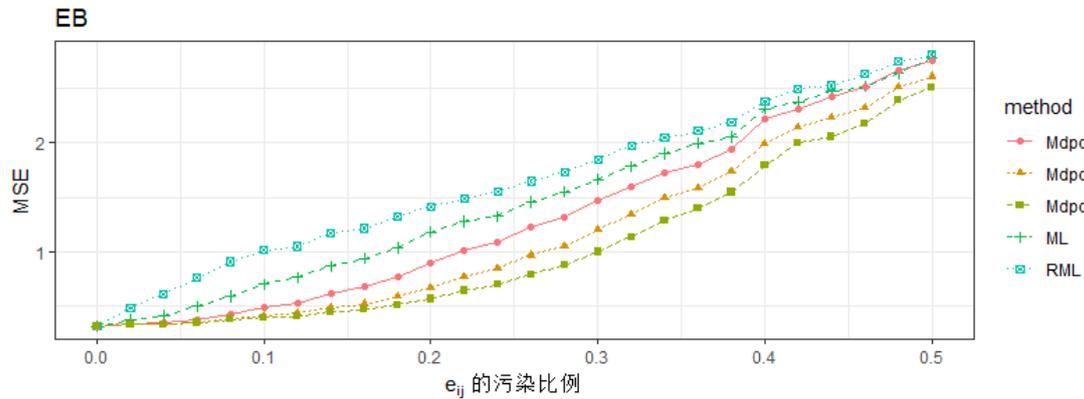
图 5.14 稳健估计量的 MSE 随  $v_i$  和  $e_{ij}$  的污染方差变化图

当模型误差和随机误差都存在离群观测时，图5.14给出了目标变量估计的平均 MSE 随误差方差的变化趋势，从图中可以看出，随着污染方差的增加，几种方法估计的 MSE 均具有上升趋势，其中 RML 的上升幅度最大，且非常明显，说明这一方法易受离群方差的影响。其它几个方法中，本研究提出的 MDPD 以及 ML 均表现较为稳定，增加幅度不大，足以说明本研究给出的估计方法是有效的。

上述结论是在污染分布的方差增加时，估计量的平均 MSE 的变化趋势，说明了当离群值与目标变量具有较大的偏差时，稳健估计方法仍然有效。下面来分析，当污染分布的污染比例增加，即观测数据中离群观测的数量增加时，稳健估计量的平均 MSE 变化情况，见图形5.15,5.16。



(a)  $v_i$  被污染



(b)  $e_{ij}$  被污染

图 5.15 稳健估计量的 MSE 随污染比例变化图

图5.15a中给出了当区域随机误差  $v_i$  的污染比例增加时，稳健估计均值的 MSE 的变化情形。从图中可见，RML 方法估计的 MSE 随污染比例的增加呈显著的上升趋势，但其变化幅度并不大。其它几种估计方法中，均呈现出先上升再下降的趋势，但其变化区间不大，变化的幅度有限，足以说明这几类稳健估计方法对区域随机误差的增加并不敏感，具有较好的稳健效果。在图5.15b中，当模型误差  $e_{ij}$  的污染比例增加时，估计的 MSE 随上升趋势，足以说明当模型误差的污染程度增加时，几种稳健估计方法都不太理想，尽管本研究提出的估计方法相比于 ML 和 RML 均有更小的 MSE。

图形5.16中给出了  $v_i$  和  $e_{ij}$  均被污染时，污染比例增加时，估计量的平均 MSE 的变化情况。从图中可见，RML 估计方法对污染比例的增加非常敏感，显著增加，且变化幅度较大。另外几种稳健估计方法的平均 MSE 也呈现出显著的上升趋势，其变化幅度相对较小，MDPD1 和 ML 的 MSE 最小。结合图5.15和5.16可知，几类稳健估计方法估计的 MSE 均会随着污染比例的增加而增加。尽管本研

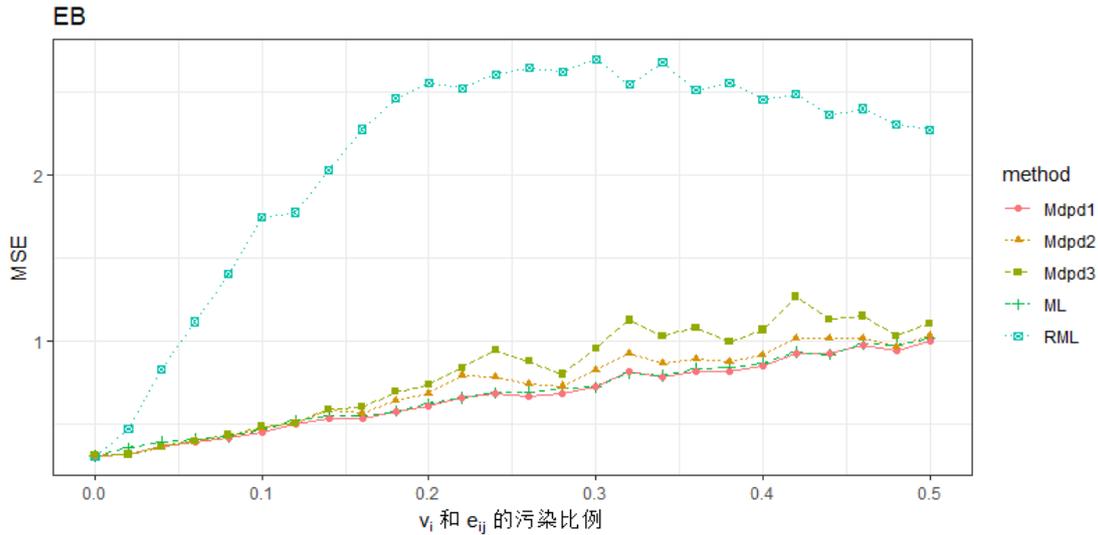


图 5.16 稳健估计量的 MSE 随  $v_i$  和  $e_{ij}$  的污染比例变化图

究提出的 MDPD 方法比其它估计方法更加有效, 其估计的 MSE 还是会增加, 因此建议在污染比例较小时使用 MDPD 估计方法会得到较为可靠的结果, 例如当污染比例小于 0.3 时, 提出的方法会显著有效。

接下来将重点考虑有限总体中单元水平模型的估计问题, 比较在有限总体中用上述稳健估计方法得到的区域均值的估计结果。在此比较上述模拟实验中的方法在  $m = 40$ ,  $N_i$  分别为 40, 80, 200 的情形下总体均值的表现。在总体数据集的模拟中, 第  $i$  个区域的  $N_i$  个单元中选取  $n_i = 4$  个单元作为随机样本。对每次模拟的数据集, 可以利用上述模拟中提到的稳健方法获取每个小区域上观测变量的均值, 最后比较 500 次模拟之后区域均值的平均估计效果。

表5.4中给出了有限总体中分布被污染的情形下, 区域均值的估计偏差和估计的 MSE。分别考虑了有限区域总体分别为 40, 80, 200 时, 在污染情形  $(0, 0)$ ,  $(v, 0)$ ,  $(0, e)$ ,  $(v, e)$  下, 区域均值的平均偏差和 MSE。

表 5.4 有限总体区域均值的估计偏差和 MSE

污染情形	ML	RML	RMD1	RMD2	RMD3
$N_i=40$		$K=2$	$\gamma=0.1$	$\gamma=0.2$	$\gamma=0.3$
$(0, 0)$	-0.0065	-0.0382	-0.0614	-0.0535	-0.0464
	0.2331	0.2620	0.2344	0.2339	0.2337
$(v, 0)$	-0.0057	-0.0162	0.1543	0.1583	0.1620
	0.6525	0.6885	0.6848	0.6855	0.6863
$(0, e)$	0.0731	0.0614	-0.0714	-0.0727	-0.0740
	0.2725	0.5016	0.2703	0.2711	0.2719
$(v, e)$	0.0738	0.0516	0.1345	0.1173	0.0985
	1.1644	1.9497	1.6813	1.6808	1.6821
$N_i=80$					
$(0, 0)$	-0.1310	-0.1500	0.0307	0.0280	0.0235
	0.1984	0.2737	0.1798	0.1784	0.1770
$(v, 0)$	-0.1413	-0.1548	0.2015	0.2217	0.2321
	0.7206	0.7453	0.7393	0.7375	0.7396
$(0, e)$	0.0014	-0.0078	0.0324	0.0334	0.0347
	0.2335	0.5899	0.2334	0.2329	0.2323
$(v, e)$	-0.0089	-0.0215	0.1973	0.1982	0.2015
	1.7002	1.9281	1.7466	1.7505	1.7559
$N_i=200$					
$(0, 0)$	0.0044	-0.0434	0.0227	0.0282	0.0335
	0.2415	0.3125	0.2430	0.2434	0.2435
$(v, 0)$	0.0686	0.0646	0.2812	0.2704	0.2579
	0.8123	0.8014	0.8868	0.8901	0.9041
$(0, e)$	0.4473	0.3949	0.0139	0.0124	0.0115
	0.4965	0.8335	0.2978	0.2977	0.2974
$(v, e)$	0.5115	0.5126	0.2812	0.2700	0.2598
	1.7897	1.9373	1.5979	1.5888	1.5884

表中每种情形下第一行表示的是估计量的平均偏差，第二行表示的是估计

量的平均 MSE。从模拟结果上看，总的来说我们提出的 RMDPD 方法在大部分情形下具有更小的 MSE，在某些情形下虽不及 RML 方法的估计效果，但相差不大。另外，当区域尺度变大时如  $N_i = 200$ ，RMDPD 方法得到的估计量具有更小的 MSE，估计效果好于传统的估计方法。

### 5.8.0.2 实际应用

在本节中，采用 Battese et al.<sup>[3]</sup> 中用于估计爱德华州 12 个县的玉米和大豆的种植面积的调查数据。该数据可以从 R 包 sae 中获得，其中包含了来自于 12 个县区的玉米和大豆的面积样本数据 37 个，以及每个区域上的玉米和大豆的像素值。用农场采访收集的数据最为因变量，卫星数据作为辅助变量，建立单元水平模型。

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_i + \tilde{e}_{ij}, \quad (5.21)$$

该模型可作为 (5.1) 的特殊情形，即取  $k_{ij} = 1, \mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2})^T$ ，和  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ 。这里  $y_{ij}$  表示玉米种植面积数量， $x_{ij1}$  和  $x_{ij2}$  分别表示第  $i$  个县种植玉米和大豆面积的像素值构成的辅助变量。

Battese et al.<sup>[3]</sup> 将 Hardin 县的一个观测值识别为离群值，并且他们在预测玉米和大豆的面积时只是将这一观测值简单的删除。Sinha et al.<sup>[28]</sup> 利用稳健估计方法对这一数据进行了分析，并给出了存在离群值时相应的预测值。在这里，运用提出的稳健估计方法 RMDPD 对该数据进行建模分析，分析离群观测对传统估计量的影响。

考虑到该数据中存在离群值，因此考虑用我们提出的稳健估计方法进行每个区域上玉米种植面积的估计和预测。另外，由于本数据中仅存在一个离群观测，因此在提出的稳健估计方法中，我们选取  $\gamma = 0.01 \ 0.05$  两种情形进行估计。

表 5.5 模型参数的估计及其标准误

Coefficients	ML	RML	Mdpd1	Mdpd2
Intercept( $\beta_0$ )	18.09(29.82)	28.68(27.30)	22.04 (14.44)	37.64(14.18)
Corn pixels( $\beta_1$ )	0.3657(0.0625)	0.3545(0.0574)	0.3614(0.0329)	0.3441(0.0334)
Soybeans pixels( $\beta_2$ )	-0.0302(0.0650)	-0.0676(0.0596)	-0.0427(0.0337)	-0.0942(0.0323)
$\sigma_e^2$	280.2(71.55)	213.7(63.94)	274.3(66.84)	243.11(68.38)
$\sigma_v^2$	47.80(56.51)	113.7(64.57)	52.64(50.32)	75.01(69.94)

在表5.5中, 利用 ML 方法、Sinha et al.<sup>[28]</sup> 提出的稳健方法、以及所提出的 MDPD 方法估计得到回归系数和随机误差的方差均被展示。同时表中括号内展示了根据第四节中的渐近分布得到的每个参数的标准误。结合表中数据比较而言, 在调整参数  $\gamma = 0.01$  时, 用 MDPD 方法估计得的系数介于 ML 方法和 RML 方法得到的估计值之间。当调整参数  $\gamma$  增加到 0.05 时, 模型估计的系数有了显著的变化。通过比较表中展示的参数的标准误, 可见用提到的方法估计得到的参数具有更小的标准误。

为了体现估计的效果, 我们比较了几种估计方法对每个区域上大豆种植面积平均亩数的预测。采用上述估计方法得到的 EBLUP 值被展示在表格5.6中, 其中括号内展示了用 500 个 Bootstrap 样本得到的估计量的 MSPE 的 Bootstrap 估计值。首先在估计结果上来看, 用本研究提到的估计方法对没有离群值区域的估计更加接近与 ML 估计的结果, 并且对区域 Hardin 的预测有了一定的改进。其次, 通过比较 Bootstrap MSPE, RMDPD 方法得到的 MSPE 值更小。可见, 提出的方法是有效的。

表 5.6 各县区玉米种植面积的预测值及其 Bootstrap MSPE

县区	样本量	估计值			
		ML	RML	MDPD1	MDPD2
Cerro Gordo	1	122.2(7.9)	124.2(9.7)	122.4(8.1)	122.5(7.9)
Hamilton	1	123.2(7.3)	125.6(9.6)	123.5(7.4)	124.2(7.4)
Worth	1	113.9(7.4)	106.1 (9.3)	113.6(7.1)	112.0 (7.6)
Humboldt	2	115.4(7.4)	112.7 (8.7)	115.0(6.8)	112.6 (6.6)
Franklin	3	136.1(7.0)	143.2 (7.3)	136.7(6.0)	139.1 (6.0)
Pocahontas	3	108.4(7.3)	113.0 (7.0)	108.5(6.0)	109.0 (5.9)
Winnebago	3	116.8(7.0)	113.0 (7.5)	116.8(6.0)	115.8 (6.1)
Wright	3	122.6(6.8)	123.1 (7.5)	122.5(6.0)	121.8 (5.8)
Webster	4	110.9(6.4)	116.1 (6.6)	111.1(5.4)	112.0 (5.4)
Hancock	5	124.4(5.9)	122.6 (6.1)	124.5(5.3)	124.5 (5.4)
Kossuth	5	113.4(5.9)	105.2 (6.2)	112.9(5.3)	110.7 (5.4)
Hardin	6	131.3(5.8)	142.2 (6.0)	131.5(5.4)	132.3 (5.7)

## 5.9 小结

在本章中，提出了一种针对存在离群观测值的单元水平模型的小区域稳健估计方法。通过引入 MDPD 方法，给出了解决具有离群观测和非正态分布误差的稳健估计方法。首先给出了单元水平模型参数的估计方程和渐近性质。其次，结合参数的渐近分布，给出了最优调整参数的选择程序。再次，给出了有限总体中单元和区域均值的 EBLUP 值及其估计量的 MSE。最后通过模拟数据和实际数据验证了所提出方法的优越表现。在模拟部分，模拟了分布被污染时的稳健估计，讨论了三种污染情形下几类稳健估计方法的效果。还特别讨论了污染比例变化、以及污染分布的方差变化时几类估计方法的 MSE 的变化情况。同时模拟结果表明，提出的方法能更好地解决这种离群情况。实际数据中，用了一个小区域估计中很经典的数据来说明提出方法的有效性，通过比较，提出的方法能够很好地处理离群观测这一特殊情况。进一步，还可以验证提出的方法针对随机效应服从其他有偏分布时也是有效的。在本研究中，当随机误差的分布被污染，且污染概率大于 0.3 时，提出的估计方法的表现一般，比 Sinha et al.<sup>[28]</sup> 中的稳健估计方法差，但是在这种情形下，几类方法得到的 MSE 均很大，稳健估计结果都不太有价值。下一步，本研究提出的估计方法还能用于指数分布情形的小区域估计问题中，当然，这些工作有待于进一步的研究和证明。

## 第 6 章 基于 $\gamma$ 散度的单元水平模型的稳健估计

本章主要研究单元水平模型的稳健小域估计, 将  $\gamma$  散度应用于单元水平模型, 研究单元水平中小域目标单元的估计问题, 给出了参数估计程序, 估计量的表达式, 以及估计量的 MSE。并通过模拟数据和实际数据对本研究中提出的估计方法进行了验证, 证明本研究中所提出的方法具有最小的 MSE 和良好的稳健性。最后本章在前述理论研究的基础上, 考虑稳健小域估计量的应用研究。将上述理论研究结果应用于中国家庭追踪调查数据, 探究家庭收入在小域上的变化情况。

类似于第4章的内容, 本章主要介绍利用最小化  $\gamma$  散度的方法对单元水平模型进行稳健估计。提出经验贝叶斯估计量, 推导估计量的 MSE, 并给出 MSE 的估计。并通过模拟实验和实际数据检验提出估计量的稳健性, 和常用的几类稳健估计方法进行比较。有关单元水平模型的基本理论框架和模型如同第5章中的相同, 在本章直接讨论基于  $\gamma$  散度的稳健估计。

### 6.1 基于 $\gamma$ 散度的 NER 模型的稳健估计方法

本章中, 重点讨论基于  $\gamma$  散度的单元水平模型的稳健估计问题, 类似于4.6节的内容, 将  $\gamma$  散度应用于单元水平模型的稳健估计中, 尝试解决单元水平模型中存在离群观测以及模型假设不满足的问题。

本研究通过引入  $\gamma$  散度和  $\gamma$  似然函数, 构建了基于单元水平模型的小域稳健估计方法, 得到了模型参数的稳健估计和小域目标变量的稳健估计。与现有的稳健估计方法相比, 能更好地处理区域随机误差和模型随机误差非正态的情形, 对于目标变量存在离群观测的情形, 具有更好的稳健性, 估计均方误差更小。在利用模拟数据进行验证中, 比较了不同误差分布情形下几类常用估计方法得到的估计量的均方误差, 并进一步探究了随着污染分布的方差和比率变化时, 所得估计量的均方误差变化情形。最后, 通过应用于经典的小域估计数据, 进一步验证了本研究提出方法的可行性。

### 6.2 $\gamma$ 散度在 NER 模型中的应用和模型参数估计

在第2.4节中, 给出了  $\gamma$  散度的定义、 $\gamma$  似然函数的定义, 并指出了其在稳健估计中的应用。第4章中应用  $\gamma$  散度研究了区域水平模型的稳健估计。并得到

了有效的稳健性。在本节中, 将其应用于单元水平模型, 探究其稳健效果。在本研究中, 利用文献 Nakagawa et al.<sup>[142]</sup> 中用到的  $\gamma$  似然函数来考虑单元水平模型的稳健性估计问题。  $\gamma$  似然函数的定义如下所示:

$$L_\gamma(y; \phi) = -\frac{1}{\gamma} \sum_{i=1}^m \frac{f_i(y_i; \phi)^\gamma}{(\int f_i(t; \phi)^{1+\gamma} dt)^{\frac{\gamma}{\gamma+1}}} - \frac{n}{\gamma}, \quad \gamma \in (0, 1) \quad (6.1)$$

其中  $f_i(y_i; \phi)$  是  $y_i$  的密度函数。  $\gamma$  是与稳健性相关的调整参数。 根据文献 [142] 的结论, 有

$$\lim_{\gamma \rightarrow 0} L_\gamma(y; \phi) = \sum_{i=1}^m \log f_i(y_i; \phi) = \log f(y; \phi)$$

因此,  $\gamma$  似然函数可以看作是对数似然函数的推广。

根据单元水平模型的基本假设, 有  $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \mathbf{V}_i)$ , 将 (5.3) 中的条件密度作为模型密度代入  $\gamma$  散度的定义中, 得到单元水平模型的  $\gamma$  似然函数,

$$L_\gamma(y; \boldsymbol{\phi}) = -\frac{1}{\gamma} \sum_{i=1}^m \left\{ \pi_\gamma |\mathbf{V}_i|^{-\frac{n_i \gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) \right\} - \frac{n}{\gamma}. \quad (6.2)$$

其中  $\pi_\gamma = \left(\frac{1+\gamma}{2\pi}\right)^{\frac{n_i \gamma}{2(1+\gamma)}}$ ,  $B_i = (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\beta})$ 。 根据公式 (5.4), 可得

$$B_i = \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta})^2 - \frac{\rho_i}{\sigma_e^2 a_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta}) \right\}^2.$$

基于  $\gamma$  散度的稳健估计方法中, 首先需要最小化目标函数 (6.2), 求未知参数  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  的估计量。 对式 (6.2) 的  $\gamma$  似然函数关于  $\boldsymbol{\theta}$  的每个分量求偏导数, 便可得参数估计方程:

$$\begin{aligned} & \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{\gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \mathbf{x}_{ij}^T = \\ & \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{\gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) \rho_i \bar{\mathbf{x}}_i^T \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (6.3) \\ & \frac{1}{1+\gamma} \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{3\gamma+2}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) a_i \cdot \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2 = \\ & \frac{1}{2} \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{\gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) \frac{1}{(a_i \cdot \sigma_v^2 + \sigma_e^2)^2} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right)^2, \quad (6.4) \end{aligned}$$

$$\frac{1}{1+\gamma} \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{3\gamma+2}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) (n_i \sigma_e^{2(n_i-1)} + (n_i - 1) a_i \cdot \sigma_v^2 \sigma_e^{2(n_i-2)}) \prod_{j=1}^{n_i} k_{ij}^2 =$$

$$\begin{aligned}
& - \sum_{i=1}^m \pi(\gamma) |\mathbf{V}_i|^{-\frac{\gamma}{2(1+\gamma)}} \exp\left(-\frac{\gamma}{2} B_i\right) \left\{ -\frac{1}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta})^2 + \frac{2\sigma_e^2 \sigma_v^2 + a_i \cdot \sigma_v^4}{(a_i \cdot \sigma_v^2 + \sigma_e^2)^2 \sigma_e^4} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right)^2 \right\}. \quad (6.5)
\end{aligned}$$

上述估计方程中,  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  是需要去估计的参数。我们可以利用牛顿-拉夫逊迭代算法对三个未知参数进行迭代求解, 具体的算法步骤将在附录中给出。为了方便, 通过上述迭代算法得到的估计值记作  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\sigma}_v^2, \hat{\sigma}_e^2)^T$ 。

如下的几个求导公式是求解方程以及后续渐近分布中会反复用到的, 在这里给出

$$\begin{aligned}
\frac{\partial |\mathbf{V}_i|}{\partial \boldsymbol{\beta}} &= 0, \quad \frac{\partial |\mathbf{V}_i|}{\partial \sigma_v^2} = a_i \cdot \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2, \\
\frac{\partial |\mathbf{V}_i|}{\partial \sigma_e^2} &= (n_i \sigma_e^{2(n_i-1)} + (n_i - 1) a_i \cdot \sigma_v^2 \sigma_e^{2(n_i-2)}) \prod_{j=1}^{n_i} k_{ij}^2, \\
\frac{\partial B_i}{\partial \boldsymbol{\beta}} &= -\frac{2}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \mathbf{x}_{ij}^T + \frac{2\sigma_v^2}{\sigma_e^2 (\sigma_v^2 + \frac{\sigma_e^2}{a_i})} \bar{\mathbf{x}}_i^T \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right), \\
\frac{\partial B_i}{\partial \sigma_v^2} &= -\frac{1}{a_i^2 (\sigma_v^2 + \frac{\sigma_e^2}{a_i})^2} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right)^2, \\
\frac{\partial B_i}{\partial \sigma_e^2} &= -\frac{1}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta})^2 + \frac{2\sigma_e^2 \sigma_v^2 + a_i \cdot \sigma_v^4}{(a_i \cdot \sigma_v^2 + \sigma_e^2)^2 \sigma_e^4} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right)^2,
\end{aligned}$$

### 6.3 模型参数的渐近分布

在推导模型参数的渐近分布中, 本节应用同第5.3中相同的记号和表达时。在本节中, 同样使用形如 (5.11) 和 (5.12) 的得分函数的定义,  $\mathbf{J}, \mathbf{K}, \mathbf{I}, \boldsymbol{\xi}$  等采用同样的定义。

接下来重点讨论估计参数的渐近性质。首先定义得分函数为  $\mathbf{u}_\theta(\mathbf{y}_i | \mathbf{x}_i) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_\theta(\mathbf{y}_i | \mathbf{x}_i)$ , 其中  $f_\theta(\mathbf{y}_i | \mathbf{x}_i)$  为式 (5.3) 中的条件密度函数,  $\mathbf{u}_\theta(\mathbf{y}_i | \mathbf{x}_i) =$

$(u_{\beta}^T(\mathbf{y}_i | \mathbf{x}_i), u_{\sigma_2^2}^T(\mathbf{y}_i | \mathbf{x}_i), u_{\sigma_e^2}^T(\mathbf{y}_i | \mathbf{x}_i))^T$  为了便于标记, 定义

$$\begin{aligned} \mathbf{J}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_{\theta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\theta}^T(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y} + \int_{\mathbf{y}} \{I_{\theta}(\mathbf{y}_i | \mathbf{x}_i) - \\ &\quad \gamma \mathbf{u}_{\theta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\theta}^T(\mathbf{y}_i | \mathbf{x}_i)\} \{g(\mathbf{y}_i | \mathbf{x}_i) - f_{\theta}(\mathbf{y}_i | \mathbf{x}_i)\} f_{\theta}^{\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}, \\ \mathbf{K}^{(i)} &= \int_{\mathbf{y}} \mathbf{u}_{\theta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\theta}^T(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{2\gamma}(\mathbf{y}_i | \mathbf{x}_i) g(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y} - \xi^{(i)} \xi^{(i)T}, \\ I_{\theta}(\mathbf{y}_i | \mathbf{x}_i) &= -\frac{\partial}{\partial \theta} \mathbf{u}_{\theta}(\mathbf{y}_i | \mathbf{x}_i), \quad \xi^{(i)} = \int_{\mathbf{y}} \mathbf{u}_{\theta}(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{\gamma}(\mathbf{y}_i | \mathbf{x}_i) g(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}^4. \end{aligned}$$

进一步定义  $\mathbf{J} = \lim \frac{1}{m} \sum_{i=1}^m \mathbf{J}^{(i)}$ ,  $\mathbf{K} = \lim \frac{1}{m} \sum_{i=1}^m \mathbf{K}^{(i)}$ , 同时不失一般性地给出下列正则假设条件:

- 存在参数空间  $\Theta$  中的开集  $\Lambda$  包含最佳拟合参数  $\theta$ , 使得  $\theta \in \Lambda$  时,  $\mathbf{J}$  是正定矩阵。
- 存在函数  $M_{jkl}(x, y)$  使得  $|\partial^3 \exp [(\mathbf{y} - \mathbf{x}\beta)^T \mathbf{V}_i^{-1}(\mathbf{y} - \mathbf{x}\beta)] / \partial \theta_j \partial \theta_k \partial \theta_l| \leq M_{jkl}(x, y)$ .

对所有的  $\theta \in \Lambda$  成立, 且  $\int_x \int_y |M_{jkl}(x, y)| g(y | x) h(x) dy dx < \infty$ 。

**定理 6.1** 在上述正则性条件下, 存在  $\hat{\theta}$  使得下列结论依概率 1 成立,

1.  $\hat{\theta}$  一致收敛于  $\theta$ ;
2.  $\hat{\theta}$  的渐近分布为  $\sqrt{m}(\hat{\theta} - \theta) \sim N_{p+2}(0, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1})$ .

该定理说明了估计参数的一致性和渐近分布, 详细的证明步骤可参见附录。在上述渐近分布的计算中,  $\mathbf{J}^{(i)}$ ,  $\mathbf{K}^{(i)}$  中均含有未知的真实密度函数  $g(\mathbf{y}_i | \mathbf{x}_i)$ , 因此直接进行计算是难以做到的。在基本假设条件中, 我们假定密度族  $f_{\theta}(\mathbf{y}_i | \mathbf{x}_i)$  中包含真实的密度函数, 因此在  $\mathbf{J}^{(i)}$ ,  $\mathbf{K}^{(i)}$ ,  $\xi^{(i)}$  的定义中, 用  $f_{\theta}(\mathbf{y}_i | \mathbf{x}_i)$  替换  $g(\mathbf{y}_i | \mathbf{x}_i)$  进行估计。有关渐近方差的具体计算可参见补充材料。从  $\mathbf{J}^{(i)}$ ,  $\mathbf{K}^{(i)}$ ,  $\xi^{(i)}$  的定义中不难发现, 当我们把矩阵  $\mathbf{J}^{(i)}$  认为是  $\gamma$  的函数时, 矩阵  $\mathbf{K}^{(i)}$  可以表示为:  $\mathbf{K}^{(i)} = \mathbf{J}^{(i)}(2\gamma) - \xi^{(i)} \xi^{(i)T}$ , 因此, 矩阵  $\mathbf{K}$  可以表示为:

$$\mathbf{K} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{J}^{(i)}(2\gamma) - \xi^{(i)} \xi^{(i)T}.$$

通过上述协方差矩阵的计算, 可知,  $\gamma$  增大时参数方差增大, 这说明 MDPD 估计量的效率随  $\gamma$  增大而减小。这一点进一步验证了优化参数  $\gamma$  用于控制 MDPD 估计量的效率和鲁棒性之间的权衡, 当  $\gamma$  增加时稳健性增强效率降低。然而, 后面的模拟实验表明这种效率的损失并不大。

#### 6.4 调整参数的选取

从上述结果中不难发现, 利用 DPD 方法迭代得到的参数  $\theta$  的估计表达式中, 含有未知的调整参数  $\gamma$ 。而  $\gamma$  的选取决定着稳健性和有效性的平衡, 当  $\gamma$  的取值越接近于 1, 则估计得到的参数具有更强的稳健性, 反之, 稳健性越弱, 有效性越强。因此选取适当的调整参数  $\gamma$  是稳健估计中很关键的因素。目前关于调整参数选取的方法主要有两个方面, 一个是根据有效性和稳健性的占比关系, 研究者根据自己的需求确定两者之间的占比, 在此基础上, 选取最优的调整参数。例如 Sugasawa<sup>[77]</sup> 在利用 DPD 方法进行区域水平的估计时, 利用该方法选取调整参数。另一种方法是基于数据驱动的参数选择方法, Warwick et al.<sup>[92]</sup> 通过最小化估计参数的 MSE 得到最优的调整参数, 但是该方法依赖于初始值的选取, 不同的参数初始值可能会产生不同的调整参数。Basak et al.<sup>[93]</sup> 在 Warwick et al.<sup>[92]</sup> 的基础上, 提出了不依赖于初始值的调整参数选择方法, 在本研究中, 我们采用 Basak et al.<sup>[93]</sup> 中提到的参数选取方法, 选择最优的调整参数, 用于构建稳健小区区域估计量。对于未知参数  $\theta$  的真实值  $\theta^*$ , 最优的调整参数  $\gamma$  通过最小化 MDPD 估计量的 MSE 来获得, 即

对于未知参数  $\theta$  的真实值  $\theta^*$ , 最优的调整参数  $\gamma$  通过最小化 MDPD 估计量的 MSE 来获得, 即

$$E \left\{ \left( \hat{\theta}_\gamma - \theta^* \right)^T \left( \hat{\theta}_\gamma - \theta^* \right) \right\} = N^{-1} \text{tr} \left\{ \mathbf{J}_\gamma^{-1} \left( \theta_\gamma \right) \mathbf{K}_\gamma \left( \theta_\gamma \right) \mathbf{J}_\gamma^{-1} \left( \theta_\gamma \right) \right\} + \left( \theta_\gamma - \theta^* \right)^T \left( \theta_\gamma - \theta^* \right). \quad (6.6)$$

由公式 (6.6) 易知, 公式中含有未知参数  $\theta^*$ , 此时关于最优调整参数  $\gamma$  的选取通常有两种方法。一种方法是认为估计  $\hat{\theta}_\gamma$  就是参数  $\theta$  的真实值, 即只要求上式第一项的最小值即可得到最优的调整参数。这种方法易于操作, 但是由于  $\hat{\theta}_\gamma$  是渐近趋向于  $\theta$  的, 因此直接替换会产生较大的误差。另一种方法是给定一个  $\theta^*$  的初始值  $\theta_p$ , 再最小化 (6.6) 求解最优的参数, Warwick et al.<sup>[92]</sup> 就是利用这种方法, 并给出了 Wick-Jones (WJ) 算法。然而, WJ 算法严重依赖于初始  $\theta_p$  的选取, 初始参数的选择直接决定了最优参数的选取。为了克服上述两个选取方法的不足, Basak et al.<sup>[93]</sup> 提出了迭代 WJ 算法 (IWJ), 用来计算最优的调整参数, 本研究中, 同样利用这种方法选取最优调整参数。具体算法如下所示:

**算法 4 IWJ 算法**

**输入:** 设置参数的初始值  $\gamma, \boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \sigma_e^{2(0)}, \sigma_v^{2(0)})$ .

**重复:**

1: 利用 WJ 算法最小化正文中的 (9) 式, 并在  $\gamma$  的取值区间  $I_\gamma$  内更新  $\gamma$ :

$$\gamma^{(i+1)} = \min_{\gamma \in I_\gamma} E \left\{ \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^{*(i)} \right)^T \left( \hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}^{*(i)} \right) \right\}.$$

2: 固定  $\gamma^{(i+1)}$ , 将其代入最小化  $\gamma$  散度的估计方程并更新  $\hat{\boldsymbol{\theta}}: \boldsymbol{\theta}^{*(i+1)} \leftarrow \text{RMG}(\boldsymbol{\theta}^{*(i+1)})$ .

3: 重复 1,2, 直至  $|\gamma^{(i+1)} - \gamma^{(i)}| < \epsilon$  或  $|\boldsymbol{\theta}^{*(i+1)} - \boldsymbol{\theta}^{*(i)}| < \epsilon^*$ , 其中  $\epsilon$  和  $\epsilon^*$  估计参数的精度。

**输出:**  $\boldsymbol{\theta}^{(m+1)}$

**6.5 稳健 EB 估计量**

本节考虑满足小域模型 (5.1) 的目标变量  $y_{ij}$  的稳健预测量。同样假定  $y_i$  在辅助变量  $x$  的条件下服从正态分布, 即  $y_i \sim N(x_i^T \boldsymbol{\beta}, V_i)$ , 为了推导便于推导, 将目标变量  $\mathbf{y}_i$  划分成抽样部分和未抽样部分:

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{is} \\ \mathbf{y}_{ir} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{is}^T \\ \mathbf{x}_{ir}^T \end{bmatrix} \boldsymbol{\beta} + v_i \begin{bmatrix} \mathbf{1}_{is} \\ \mathbf{1}_{ir} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{is} \\ \mathbf{e}_{ir} \end{bmatrix},$$

其中下标  $r$  表示未抽样单元, 这里假设前  $n_i$  个单元为抽样的单元, 否则, 总可以通过调整顺序写成上述形式。同样地, 协方差矩阵可划分为:  $\mathbf{V}_i = \begin{bmatrix} \mathbf{V}_{is} & \mathbf{V}_{isr} \\ \mathbf{V}_{irs} & \mathbf{V}_{ir} \end{bmatrix}$ , 其

中  $\mathbf{V}_{is} = \mathbf{R}_{is} + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T = \text{diag}_{1 \leq j \leq n_i} + (k_{ij}^2) \sigma_e^2 + \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T$ ,  $\mathbf{V}_{isr} = \sigma_v^2 \mathbf{1}_{n_i} \mathbf{1}_{N_i - n_i}^T$ ,  $\mathbf{V}_{irs} = \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{n_i}^T$ ,  $\mathbf{V}_{ir} = \mathbf{R}_{ir} + \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}^T = \text{diag}_{1 \leq j \leq N_i} (k_{ij}^2) \sigma_e^2 + \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}^T$ 。结合上述单元水平模型的假设, 在抽样数据  $\mathbf{y}_{is}$  的条件下,  $\mathbf{y}_{ir}$  的条件分布为

$$\mathbf{y}_{ir} | \mathbf{y}_{is} \sim N(\boldsymbol{\mu}_{ir|s}, \mathbf{V}_{ir|s}), \tag{6.7}$$

其条件均值和条件协方差矩阵分别为:

$$\begin{aligned} \boldsymbol{\mu}_{ir|s} &= \boldsymbol{\mu}_{ir} + \mathbf{V}_{irs} \mathbf{V}_{is}^{-1} (\mathbf{y}_{is} - \boldsymbol{\mu}_{is}) = \mathbf{X}_{ir} \boldsymbol{\beta} + \sigma_v^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{n_i}^T \mathbf{V}_{is}^{-1} (\mathbf{y}_{is} - \mathbf{X}_{is} \boldsymbol{\beta}) \\ \mathbf{V}_{ir|s} &= \mathbf{V}_{ir} - \mathbf{V}_{irs} \mathbf{V}_{is}^{-1} \mathbf{V}_{isr} = \sigma_v^2 (1 - \rho_{ir}) \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}^T + \sigma_e^2 \mathbf{R}_{ir}, \end{aligned}$$

其中  $\mathbf{a}_{ir} = (a_{in_i+1}, \dots, a_{iN_i})^T$ ,  $a_{ir} = \sum_{j=n_i+1}^{N_i} a_{ij}$ ,  $\rho_{ir} = \frac{a_{ir} \cdot \sigma_v^2}{\sigma_e^2 + a_{ir} \cdot \sigma_v^2}$ 。

一般地, 小域估计的目的就是利用抽样数据  $y_{is}$  和辅助信息  $X$  来估计小域上与目标变量  $\mathbf{y}_i$  相关的函数  $\tau(\mathbf{y}_i)$  的值, 例如区域上的均值, 分位数等。由于

$\mathbf{y}_i$  中未抽样部分的值并不知道, 因此有必要考虑  $\mathbf{y}_{ir}$  的条件分布。假设模型参数  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  是已知的, 则在单元水平模型条件下, 根据 Rao et al.<sup>[1]</sup> 的结论,  $\tau(\mathbf{y}_i)$  的最佳无偏估计 (BP) 为  $\hat{\tau}^{BP} = E_{y_r}[\tau | \mathbf{y}_{is}]$ 。

在实际应用中, 模型参数  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_v^2, \sigma_e^2)^T$  被其一致估计量  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\sigma}_v^2, \hat{\sigma}_e^2)^T$  替换, 且变量  $\hat{y}_{ir}$  由式 (6.7) 的分布所生成。则  $\tau_i$  的经验贝叶斯估计 (EBP) 为:

$$\hat{\tau}_i^{ebp} = \frac{1}{N_i} \left\{ \sum_{j \in s_i} h(y_{ij}) + \sum_{j \in r_i} E_{y_r} [h(y_{ij}) | \mathbf{y}_{is}; \hat{\boldsymbol{\theta}}] \right\}.$$

一般来说, 函数  $h$  可能是不可解析表示的, 或者当  $n_i$  较小时, 上式的期望的计算可能会涉及多重积分, 因此上式直接求期望得到表达式通常是不现实的。通常使用蒙特卡洛 (Monte Carlo) 方法来计算  $\tau_i$  的 EBP 值。对于未抽样的样本  $y_{ij}, j \in r_i$ , 生成  $A$  个  $\hat{y}_{ij}^{(a)} = \hat{\mu}_{ij|s} + \hat{v}_{ij|s}, a \in 1, \dots, A$ , 其中  $\hat{\mu}_{ij|s}$  和  $\hat{v}_{ij|s}$  由式 (10) 中的分布生成。则  $\tau_i$  的 EBP 的蒙特卡洛近似估计为:

$$\hat{\tau}_i^{ebp} \approx \frac{1}{A} \sum_{a=1}^A \tau_i^{(a)}, \quad \tau_i^{(a)} = \frac{1}{N_i} \left\{ \sum_{j \in s_i} h(y_{ij}) + \sum_{j \in r_i} h(y_{ij}^{(a)}) \right\}. \quad (6.8)$$

## 6.6 区域均值的 EB 估计量

特别地, 当上述估计中的函数  $h$  为平均值函数时, 对应的估计值就为区域均值  $\bar{Y}_i = \sum_{i=1}^{N_i} y_{ij}/N_i$ 。结合 (10) 中的条件分布得  $\hat{y}_{is}^{ebp} = \mathbf{y}_{is}, \hat{y}_{ir}^{ebp} = \hat{\boldsymbol{\mu}}_{ir|s} = \mathbf{X}_{ir}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\sigma}}_v^2 \mathbf{1}_{N_i-n_i} \mathbf{1}_{n_i}^T \hat{\mathbf{V}}_{is}^{-1} (\mathbf{y}_{is} - \mathbf{X}_{is}\hat{\boldsymbol{\beta}})$ 。因此区域均值的 EBP 为:

$$\begin{aligned} \hat{Y}_i^{ebp} &= \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{y}_{ij}^{ebp} = \frac{1}{N_i} \sum_{j \in s_i} y_{ij} + \frac{1}{N_i} \sum_{j \in r_i} (\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{v}_i) \\ &= f_i \hat{Y}_i + \frac{1}{N_i} \sum_{j \in U_i} (\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{v}_i) - f_i \frac{1}{n_i} \sum_{j \in s_i} (\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{v}_i) \\ &= (1 - f_i) \left[ \bar{\mathbf{x}}_i^T \hat{\boldsymbol{\beta}} + \hat{v}_i \right] + f_i \left[ \hat{Y}_i + (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_{is})^T \hat{\boldsymbol{\beta}} \right], \end{aligned}$$

其中  $f_i = n_i/N_i$  是区域抽样比率,  $\hat{Y}_i = \sum_{j \in s_i} y_{ij}/n_i, \bar{\mathbf{x}}_{is} = \sum_{j \in s_i} \mathbf{x}_{ij}/n_i, \bar{\mathbf{x}}_i = \sum_{j \in U_i} \mathbf{x}_{ij}/N_i$ 。

## 6.7 稳健估计量的 MSE

估计单元水平模型的均方预测误差 (MSPE) 通常是比较困难的, 原因有二: 其一, 单元水平模型中误差项和未抽中样本单元所服从的真实分布并不知道, 从

而无法获取密度函数进行 MSPE 的计算。其二，有时即便知道未抽样单元的分布，由于单元水平模型涉及的是单元层级的数据，在计算其期望的过程中会出现多重积分，这对 MSPE 的计算造成了极大的挑战。

在本研究中，利用 Sinha et al.<sup>[28]</sup>, Hall et al.<sup>[102]</sup> 中提到的参数 Bootstrap 方法来估计 MSPE。具体估计步骤如下：

1. 用 MDPD 方法求参数  $\theta = (\beta^T, \sigma_v^2, \sigma_e^2)^T$  的估计值  $\hat{\theta} = (\hat{\beta}^T, \hat{\sigma}_v^2, \hat{\sigma}_e^2)^T$ ;
2. 生成独立同分布的区域随机误差  $v_i^* \sim N(0, \hat{\sigma}_i^2)$ ,  $i = 1, \dots, m$ , 生成与  $v_i^*$  相互独立的随机误差  $e_{ij}^* \sim N(0, \hat{\sigma}_e^2 k_{ij}^2)$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, m$ ;
3. 根据模型 (5.1) 生成目标变量的 Bootstrap 总体,  $y_{ij}^* = x_{ij}^T \hat{\beta} + v_i^* + e_{ij}^*$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, m$ ;
4. 记  $y_i^{P*} = (y_{i1}^*, \dots, y_{iN_i}^*)$  为区域  $i$  上的目标变量 Bootstrap 总体, 计算  $\tau_i^* = h(y_i^{P*})$ ,  $i = 1, \dots, m$ ;
5. 从 Bootstrap 总体  $y_i^{P*}$  中选取样本  $y_s^*$ , 并根据样本数据  $y_s^*$  利用 MDPD 方法估计模型参数, 得到  $\hat{\theta}^* = (\hat{\beta}^{*T}, \hat{\sigma}_v^{2*}, \hat{\sigma}_e^{2*})^T$ ;
6. 结合步骤 5 中的估计结果, 利用式 (6.8) 的方法估计  $\tau_i^{*ebp}$ ;
7. 重复步骤 2–6 共  $B$  次, 对应第  $b \in 1, \dots, B$  次的目标函数值分别为  $\tau_i^*(b)$  和  $\tau_i^{*ebp}(b)$ , 则估计量  $\tau_i^{*ebp}$  的 Bootstrap MSE 为:

$$\text{MSE}(\tau_i^{*ebp}) = B^{-1} \sum_{b=1}^B \left[ \tau_i^{*ebp}(b) - \tau_i^*(b) \right]^2.$$

## 6.8 应用举例

本节中同时使用模拟数据和实际数据来说明本研究提出的稳健估计方法的有效性，并和现有的估计方法进行对比。

### 6.8.1 数值模拟

一般地，对于正态分布而言，称形如  $(1 - \eta)N(\mu_1, \sigma^2) + \eta N(\mu_2, \sigma_1^2)$  的分布为混合正态分布，该分布用于刻画具有厚尾分布特征的数据。在本研究中，用混合正态分布  $(1 - \eta)N(0, \sigma^2) + \eta N(0, \sigma_1^2)$  来模拟正态性假定不满足的情况，其中  $N(0, \sigma^2)$  为真实分布， $N(0, \sigma_1^2)$  称为污染分布， $\eta$  称为污染比例。一般情况下

$\sigma_1^2 > \sigma^2$ , 当污染比例  $\eta = 0$  时, 表示真实分布未被污染。在本节中, 讨论模型误差和随机误差均是混合正态分布时的小域估计问题。

本研究中, 我们做和 Sinha et al.<sup>[28]</sup> 相同的模拟实验, 即具有单个辅助变量的单元水平模型:  $y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}$ ,  $i = 1, \dots, k, j = 1, \dots, n$ , 取区域数  $k = 40$ , 每个区域上样本量  $n = 4$ , 辅助变量  $x_{ij} \sim N(1, 1)$ 。在本模拟中, 选取区域随机误差  $v_i$  和随机误差  $e_{ij}$  均来源于污染比例  $\eta = 0.1$ , 污染分布的方差  $\sigma_1^2 = 25$  的混合正态分布。考虑四种情形下的估计效果, 分别记作  $\{(0, 0), (0, v), (0, e), (v, e)\}$ , 其中  $(0, 0)$  表示两个误差均没有被污染,  $(0, v)$  表示只有区域随机误差被污染等。在模拟中设置回归系数  $(\beta_0, \beta_1) = (1, 1)$ , 在上述每种情形中, 进行 500 次模拟, 考虑估计参数以及小域均值的 MSE 的变化。在模拟实验中, 重点比较传统的极大似然方法 (ML)、Sinha et al.<sup>[28]</sup> 中提到的稳健估计方法 (RML)、以及提出的取不同调整参数 ( $\gamma = 0.1, 0.2, 0.3$ ) 时的稳健最小化  $\gamma$  散度的方法 (RMG)。

表 6.1 四种污染情形下模型参数估计的偏差和 MSE

情形	参数	ML	RML	RMG 1	RMG 2	RMG 3	RMG.opt
(0,0)	$\beta_0 = 1$	-0.0010	-0.0029	-0.0008	-0.0014	-0.0025	-0.0035
		0.0346	0.0343	0.0348	0.0369	0.0405	0.0375
	$\beta_1 = 1$	0.0017	-0.0017	0.0010	0.0005	0.0001	0.0003
		0.0076	0.0077	0.0078	0.0083	0.0091	0.0085
	$\sigma_v^2 = 1$	-0.0406	-0.0390	-0.0384	-0.0406	-0.0456	-0.0312
		0.0769	0.0803	0.0790	0.0860	0.0969	0.0861
(v,0)	$\sigma_e^2 = 1$	-0.0052	-0.0048	-0.0025	-0.0022	-0.0037	-0.0045
		0.0164	0.0174	0.0166	0.0183	0.0211	0.0190
	$\beta_0 = 1$	0.0029	-0.0037	-0.0059	-0.0108	-0.0135	-0.0136
		0.0962	0.0780	0.0544	0.0496	0.0528	0.0517
	$\beta_1 = 1$	0.0022	-0.0007	0.0026	0.0034	0.0036	0.0040
		0.0084	0.0083	0.0085	0.0092	0.0099	0.0097
	$\sigma_v^2 = 1$	2.3795	1.6838	0.9269	0.3647	0.1984	0.2427
		8.5421	4.3143	1.4393	0.3588	0.2204	0.2015
	$\sigma_e^2 = 1$	-0.0144	-0.0232	-0.0140	-0.0137	-0.0145	-0.0153
	0.0154	0.0174	0.0159	0.0181	0.0212	0.0199	

(0,e)	$\beta_0 = 1$	0.0037	-0.0143	-0.0042	-0.0095	-0.0133	-0.0117
		0.0737	0.0487	0.0548	0.0544	0.0597	0.0552
	$\beta_1 = 1$	-0.0056	-0.0063	-0.0005	0.0017	0.0028	0.0016
		0.0244	0.0129	0.0145	0.0126	0.0127	0.0128
	$\sigma_v^2 = 1$	-0.0573	-0.1769	-0.0558	-0.0511	-0.0534	-0.0444
		0.1846	0.1208	0.1191	0.1225	0.1355	0.1183
(v,e)	$\sigma_e^2 = 1$	2.2893	0.6826	0.9554	0.4484	0.2797	0.3965
		6.0234	0.5354	1.0862	0.2778	0.1362	0.2294
	$\beta_0 = 1$	0.0079	-0.0122	-0.0089	-0.0041	0.0001	-0.0053
		0.1407	0.1110	0.0808	0.0723	0.0748	0.0736
	$\beta_1 = 1$	0.0053	-0.0032	0.0063	0.0063	0.0054	0.0049
		0.0269	0.0144	0.0146	0.0127	0.0132	0.0138
(v,e)	$\sigma_v^2 = 1$	2.3661	1.5949	1.0528	0.4330	0.2396	0.5363
		8.8796	4.3080	1.8641	0.5182	0.3265	0.5882
	$\sigma_e^2 = 1$	2.3477	0.7721	0.9802	0.4270	0.2480	0.6143
		6.2115	0.6866	1.1589	0.2660	0.1283	0.7582

<sup>1</sup> 上述 RMG1, RMG2, RMG3, RMG.Opt 表示用最小化  $\gamma$  散度的方法, 且其调整参数分别为 0.1, 0.2, 0.3 以及通过算法选取得到的最优调整参数。上述结果是在重复 500 次模拟之后得到的平均结果。

首先比较四种污染情形下的模型参数的估计。表 1 中给出了几种稳健估计方法在不同污染情形下的参数估计的偏差和均方误差, 其中每种方法对应的第一列表示估计的偏差, 第二列表示相应的 MSE。

从表6.1中很明显可以得出结论: 当随机误差未被污染时, ML 方法在参数估计中表现最好, 但是 RML 以及具有较小调整参数的 RMG 方法和 ML 估计结果非常相近。说明在这种情形下, 具有较小调整参数的 RMD 方法、RML 与 ML 方法是几乎一样有效的, 其偏差和 MSE 相差都不多。具有较大调整参数的 RMD 方法表现不好, 恰好表明调整参数的选取至关重要, 而最优调整参数可根据我们提供的算法得到。

当随机效应  $v$  存在污染的情形下, ML 方法估计得到的方差  $\sigma_v^2$  具有大的偏差和 MSE, RML 方法的估计结果较 ML 方法虽有所改善, 但与 RMG 方法相比, 其 MSE 仍然较大。综合几类方法的模拟结果不难发现, RMG 方法明显优

于 RML 方法, 对所有参数的估计均有较小的 MSE, 且依照本文算法选取调整参数时, 每个参数的估计效果都比较好。同样地, 当模型误差中存在污染时, ML 方法估计得到的  $\sigma_e^2$  具有大的偏差和 MSE, RML 方法对离群值的影响控制较好, 每个参数的估计偏差和 MSE 都较小。RMG 方法中, 当  $\gamma = 0.1$  时, 结果中模型方差的估计有所减少, 但不及 RML 方法好, 然当  $\gamma \geq 0.2$  时, RMG 方法的表现优于 RML 的估计结果。

在区域效应和模型误差均被污染时, ML 对方差部分的估计受离群值观测影响严重, 以致于产生很大的偏差和 MSE, RML 方法减小了离群值带来的影响, 但表现并非最好, RMG 方法显著好于 RML 方法, 只需要选取合适的调整参数即可。

接下来, 同样利用上述数据集, 考虑小域均值的估计。取已知的辅助变量在第  $i$  个区域上的均值为  $\bar{x}_i = 1, i = 1, 2, \dots, m$ 。表 6.2 给出了几种稳健方法得到的区域均值估计的偏差和 MSE 在区域上的平均值。在未污染数据的情况下, 使用 ML 方法得到的 EBLUP 似乎是最有效的, 这一点和我们的预期结果相同。利用 RML 方法得到的 REBLUP 估计量、以及本文提到的 RMG 方法得到的估计量和 EBLUP 一样有效。在其他三种有污染的情形下, 通过表中数据可以得到和上述模拟中一致的结论, ML 方法得到的小区域均值具有大的偏差和 MSPE, RML 方法得到的小区域均值表现稍好, RMG 方法表现严格优于 RML 方法。

表 6.2 四种污染情形下均值估计的偏差及其 MSE

情形	ML	RML	RMG	RMG2	RMG3	RMG.opt
(0,0)	-0.0009	-0.0010	-0.0009	-0.0008	-0.0007	-0.0008
	0.3043	0.3043	0.3107	0.3121	0.3146	0.3124
(v,0)	-0.0021	-0.0024	-0.0027	-0.0032	-0.0038	-0.0036
	0.3067	0.2996	0.3070	0.3084	0.3108	0.3034
(0,e)	0.0033	0.0013	0.0018	0.0016	0.0015	0.0015
	0.6837	0.4326	0.4802	0.4055	0.3869	0.3926
(v,e)	-0.0025	-0.0021	-0.0038	-0.0032	-0.0026	-0.0025
	0.7084	0.5608	0.5548	0.4856	0.4076	0.4235

为了检验现有几种稳健方法对污染比例和污染分布方差的大小的表现, 进一步模拟验证估计参数的 MSE 随污染比例、污染分布方差  $\sigma_1^2$  变化情况。同上述模拟步骤一样, 考虑区域随机误差的分布被污染, 单元随机误差的分布被污染

以及单元随机误差和区域随机误差的分布均被污染三种情形下的估计效果。在模拟污染数据时，利用 (5.1) 中的模型生成数据，然后用上述提到的方法进行参数估计。重复模拟 500 次，考虑其 MSE 的平均值，绘制变化曲线图。第一种情况，考虑随污染比例增加时估计的参数的 MSE 的表现。污染分布的方差被固定为 25，考虑污染比例在 0—0.5 之间以 0.02 为步长变化时，三种污染情形下估计参数的 MSE 随污染比例的变化情况。另一种情形，考虑污染分布的方差增加时模型参数的 MSE 的表现。在污染分布中，污染比例被确定为 0.1，考虑污染分布的方差在 5—100 以 5 为步长增加时，三种污染情形下估计参数的 MSE 随污染方差的变化情况。

首先，考虑在污染情形  $(0, e), (v, 0), (v, e)$  下，将污染方差固定为 25，污染比例在 0—0.5 变化时估计的模型参数 MSE 的变化情况。在图6.1和图6.2给出了在污染情形  $(0, v)$  下的估计结果。

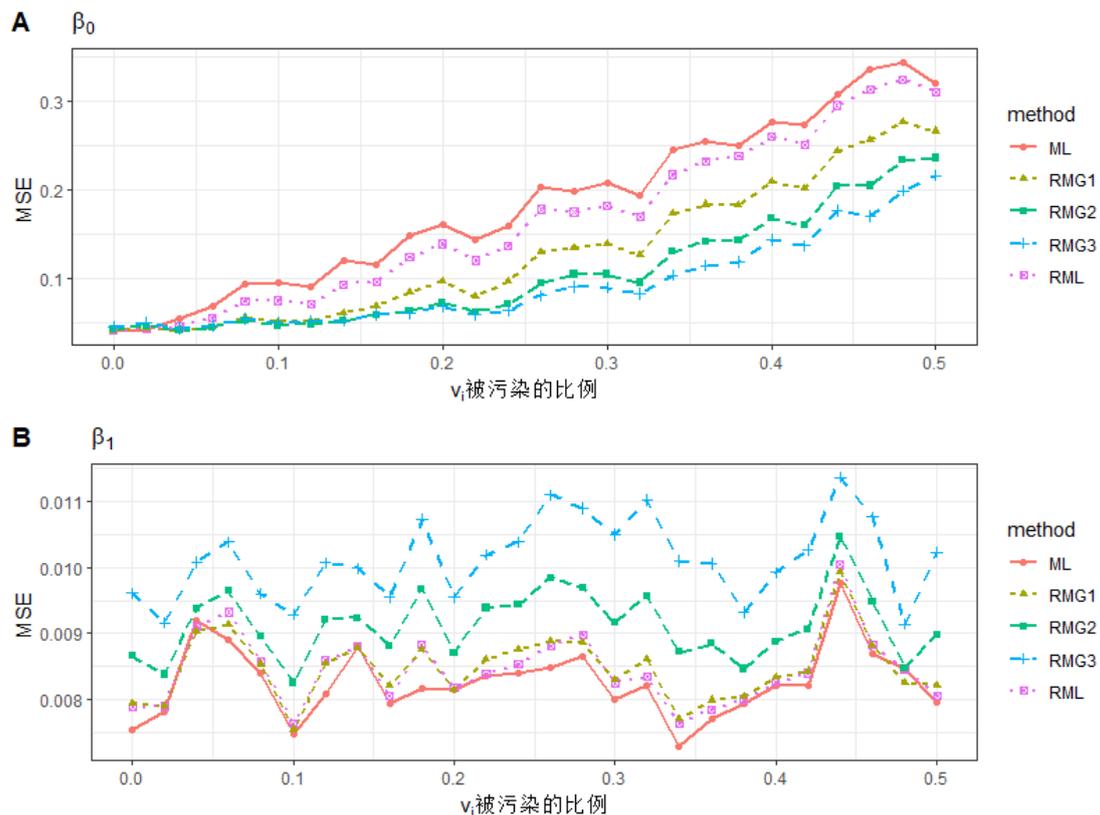


图 6.1 模型系数的 MSE 随  $v_i$  的污染比例变化图

结合图6.1,6.2我们不难发现：当区域随机效应  $v_i$  被污染时，参数  $\beta_0, \sigma_v^2$  的 MSE 的估计随污染比例的增加而增大，而参数  $\beta_1, \sigma_e^2$  的 MSE 的估计与污染比例的大小无关，且 MSE 较小。比较几种方法对四个参数 MSE 估计的表现， $\beta_1, \sigma_e^2$

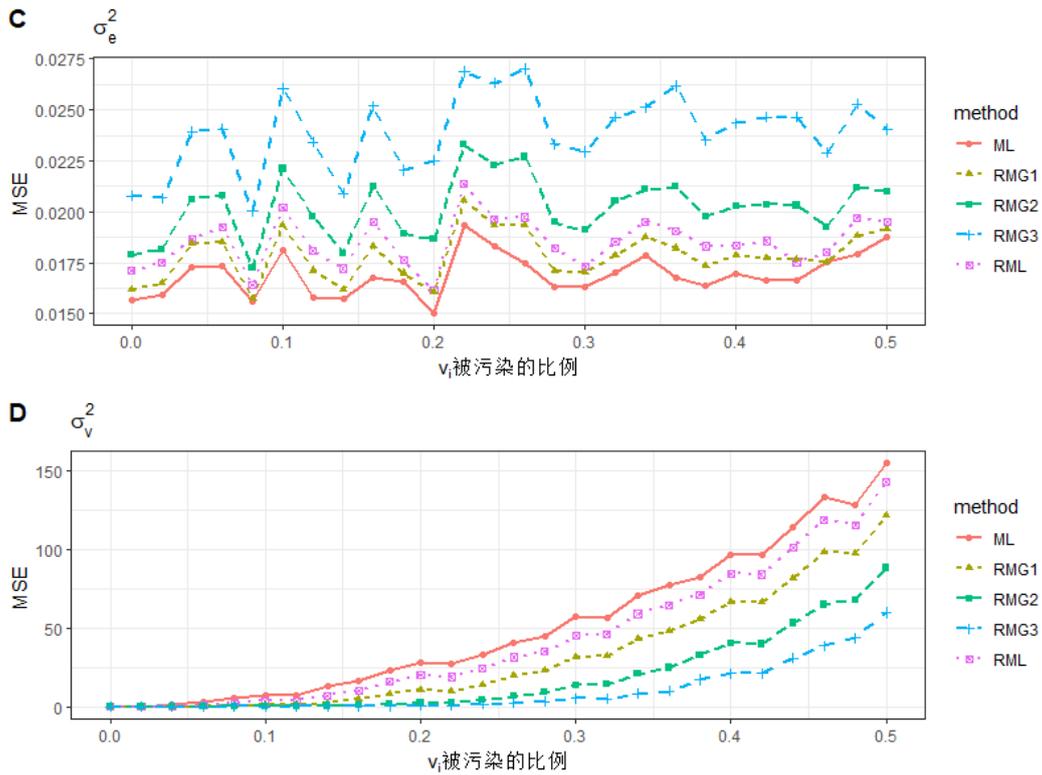


图 6.2 随机误差方差 MSE 随  $v_i$  的污染比例变化图

的 MSE 均比较小，几种方法的差异不大，而对参数  $\beta_0, \sigma_v^2$  的 MSE 的估计中，ML 方法表现最为糟糕，RML 方法的表现稍好于 ML，几个 RMG 方法的估计效果显著优于上述两个方法。进一步说明我们给出的 RMG 方法是有效的。

在情形 (0, e) 下，估计参数 MSE 的变化如图6.3,6.4所示。

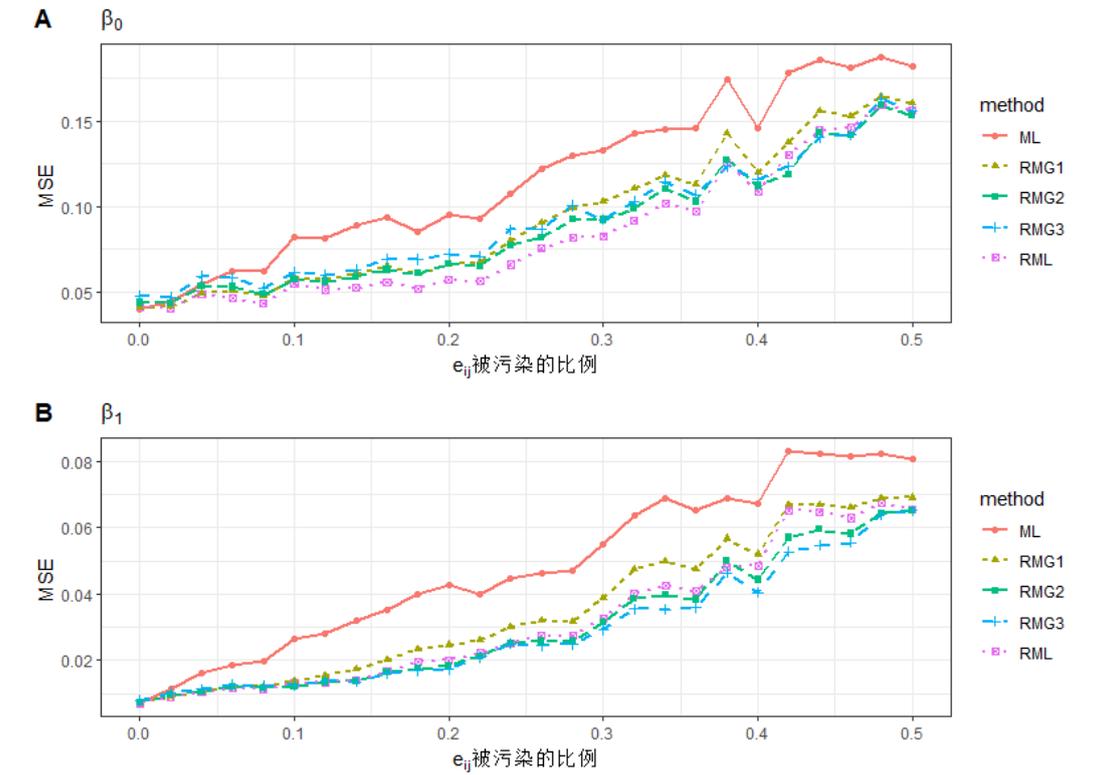


图 6.3 模型系数的 MSE 随  $e_{ij}$  污染比例变化图

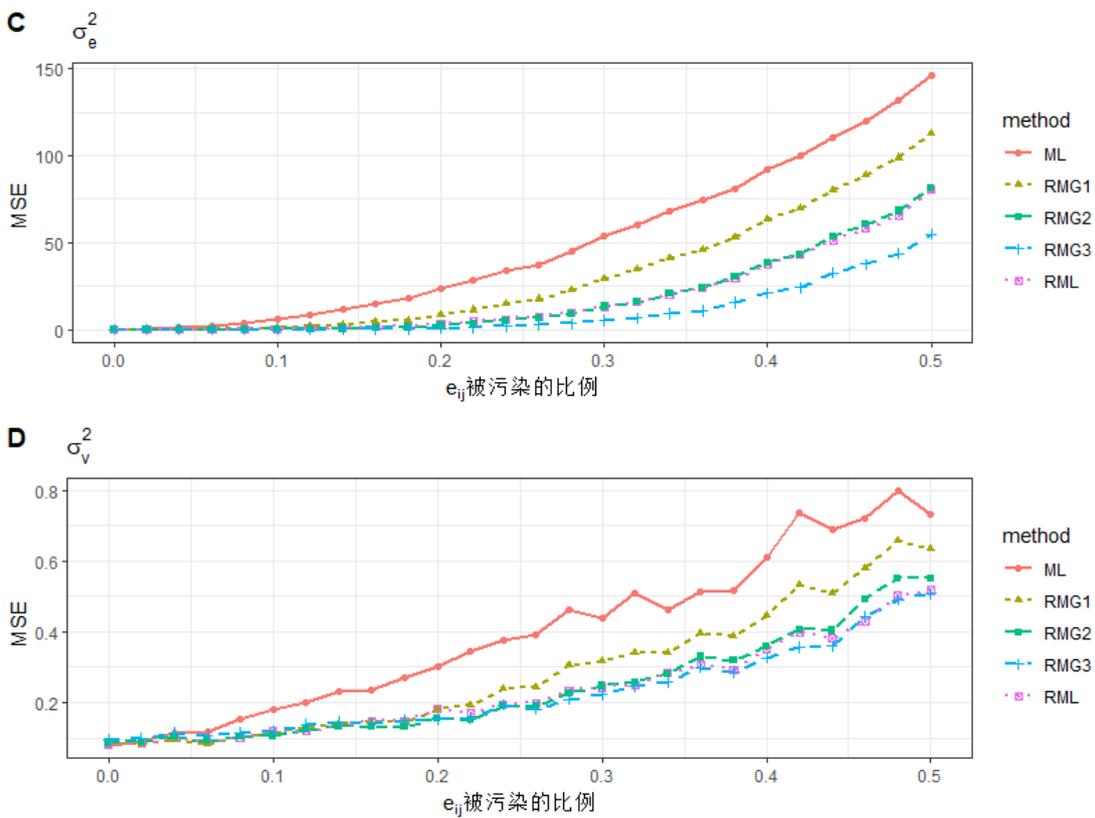


图 6.4 估计方差的 MSE 随  $e_{ij}$  污染比例变化图

从图6.3和6.4可以发现，当随机效应的污染比例从 0 至 0.5 增加时，小域模型中的四个参数的 MSE 均随污染比例的增加而增加。四个参数的 MSE 中，除了  $\sigma_e^2$  之外，其他的三个参数的 MSE 的变化均不太大，与我们的预期一样， $\sigma_e^2$  容易受  $e_{ij}$  的污染概率的影响，且估计的 MSE 比较大。在几种方法之间的相互比较中，ML 的表现最差，当调整参数  $\gamma$  的取值较小时 (0 - 0.2) 时，RMG 方法和 RML 方法表现差不多，有些情形下 RMG 方法表现更好，但当  $\gamma$  取值较大时，RML 方法要优于 RMG 方法。

在污染情形为  $(v, e)$  时，随着污染比例的变化，模型系数估计的 MSE 随其变化如图6.5,6.6所示。

图6.5和图6.6是估计参数的 MSE 随污染比例  $\eta$  变化的图形。从图中可以看出，当污染概率增大时，参数的 MSE 均呈现出显著的上升趋势。几种估计方法比较而言，RMG 方法显然会优于 RML 和 ML 方法，当然 ML 方法表现最为糟糕。

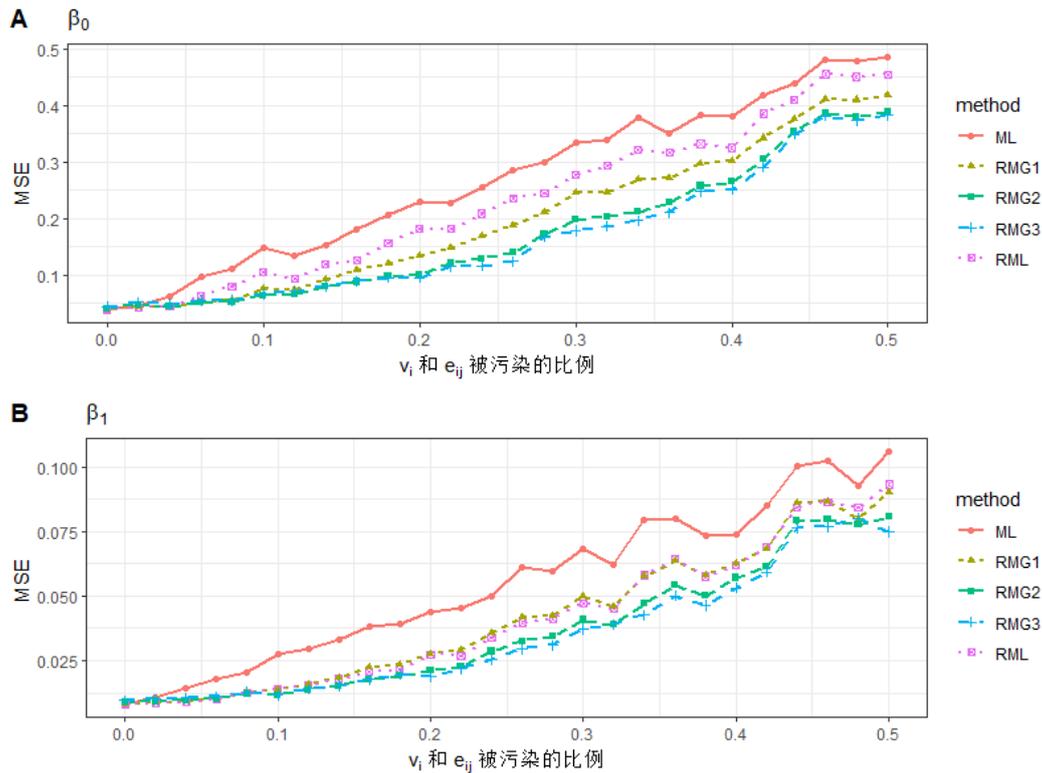


图 6.5 模型系数的 MSE 随污染比例变化图

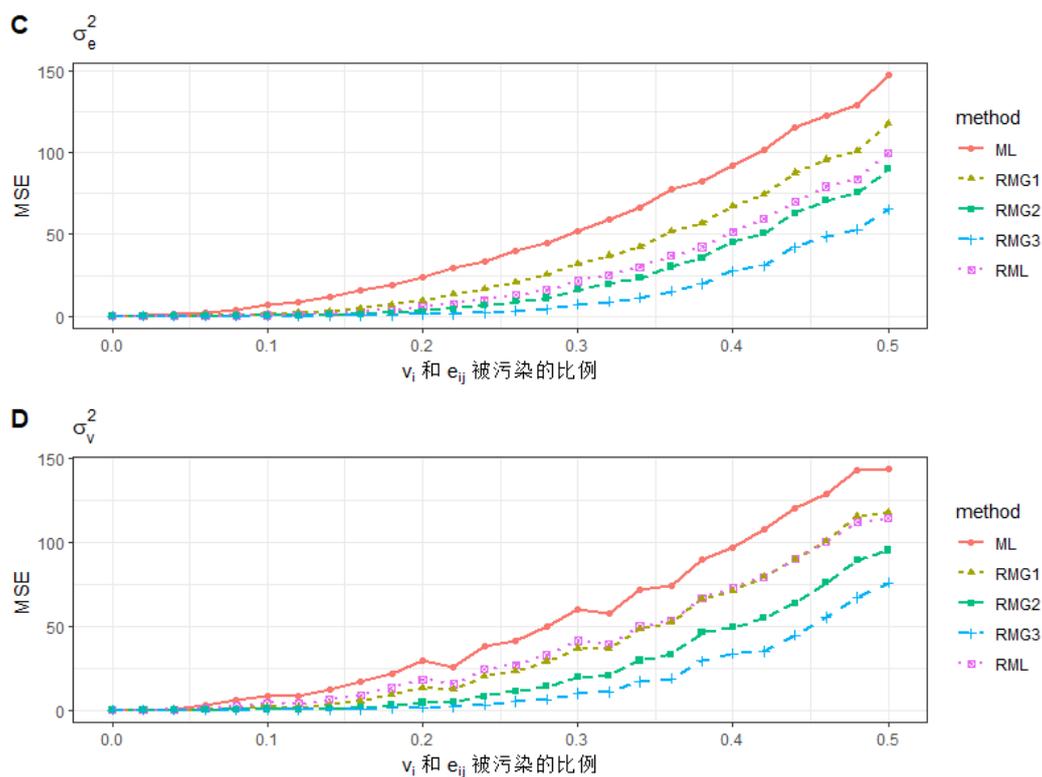


图 6.6 估计方差的 MSE 随污染比例变化图

其次，我们考虑在污染情形  $(v, 0)$ ,  $(0, e)$  和  $(v, e)$  下，将污染比例固定为 0.1，污染方差在 1 – 100 变化时估计的模型参数 MSE 的变化情况。在图6.7和图6.8给出了在污染情形  $(0, v)$  下的估计结果。

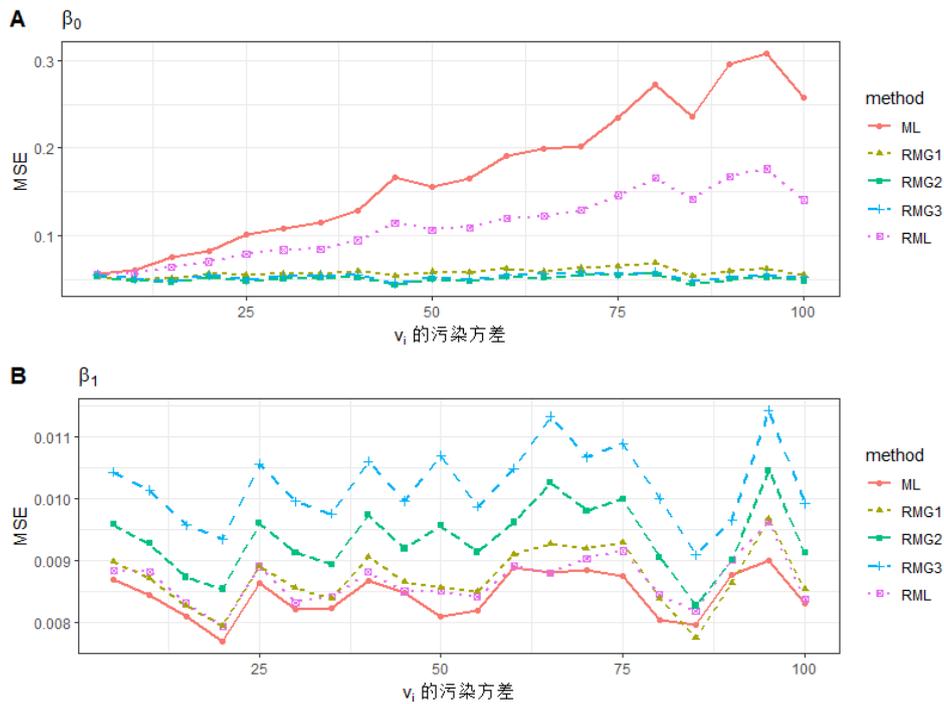


图 6.7 模型系数的 MSE 随污染分布方差变化图

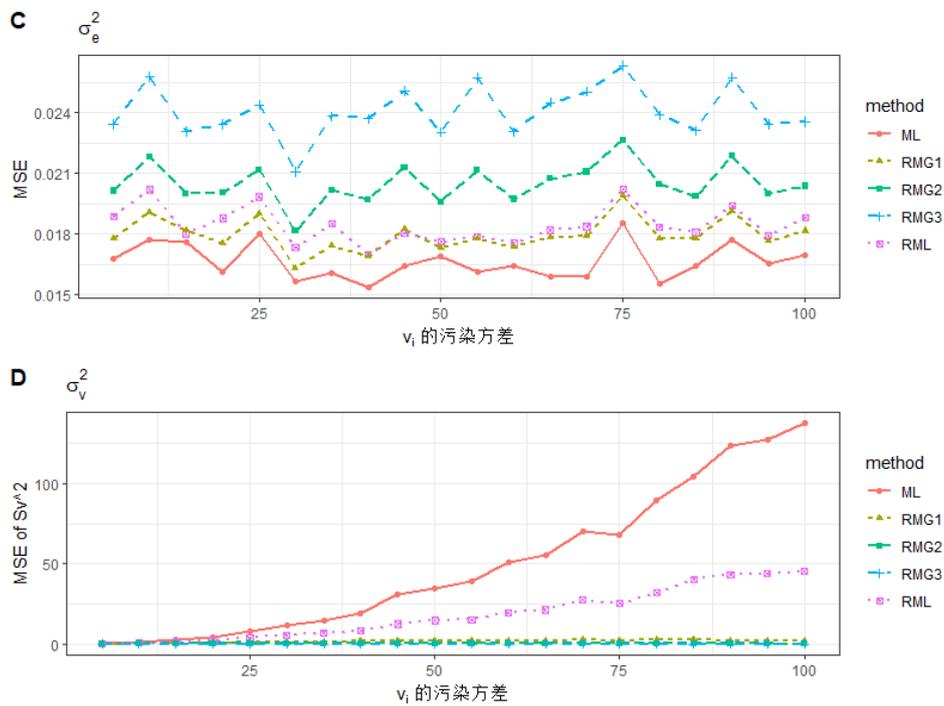


图 6.8 估计方差的 MSE 随污染分布方差变化图

结合图6.7,6.8形可知，参数  $\beta_1, \sigma_e^2$  的 MSE 与污染分布方差的大小无关。参数  $\beta_0, \sigma_v^2$  的 MSE 的估计中，ML 方法受污染分布方差影响较为严重，RML 方法的表现有所改进，但不如 RMG 方法表现好。

在污染情形  $(0, e)$  的下, 估计的模型参数的 MSE 的变化如6.9,6.10所示。

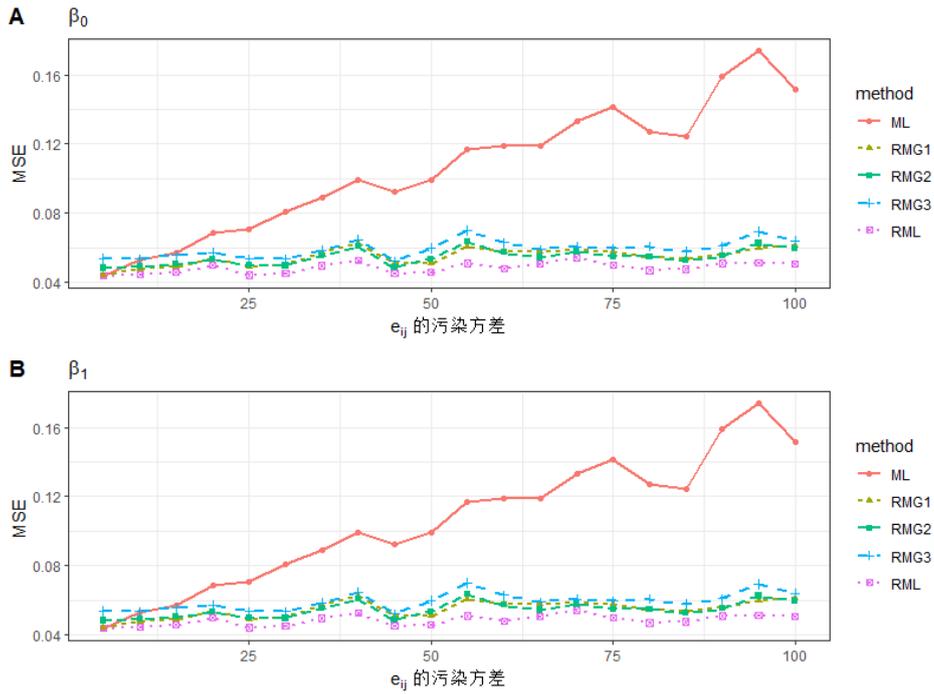


图 6.9 模型系数的 MSE 随污染分布方差变化图

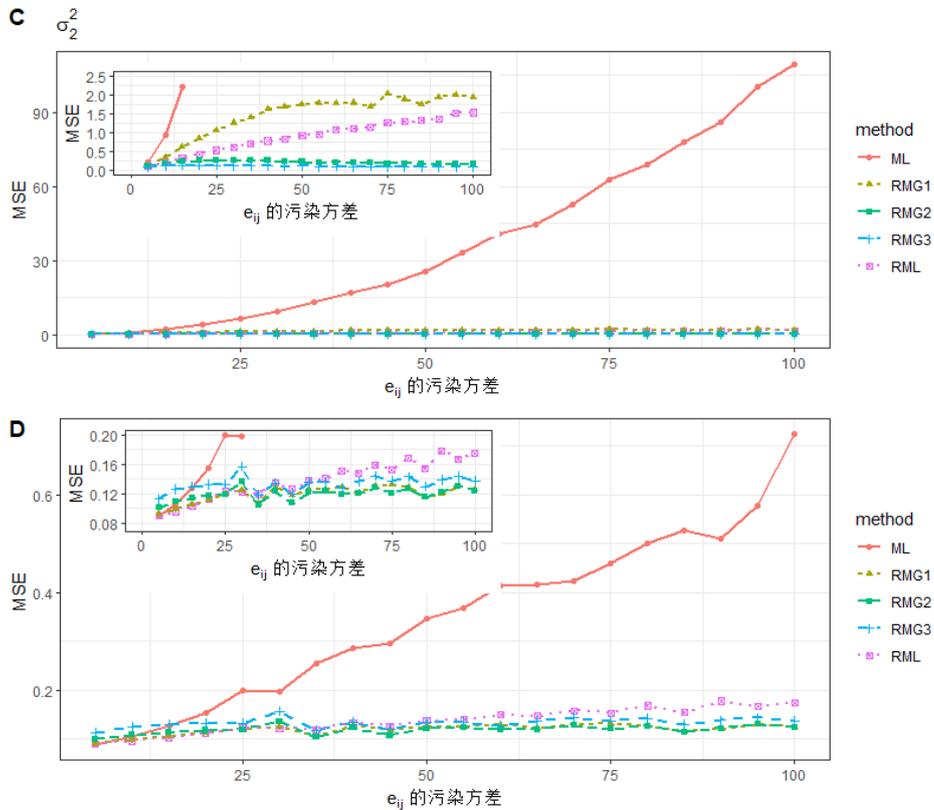


图 6.10 估计方差的 MSE 随污染分布方差变化图

当模型随机误差  $e_{ij}$  受污染, 并且污染分布的方差增大时, 四个参数的 MSE 的估计情况被呈现在图形 6.9, 6.10 中。从图中可以看到, 当污染分布的方差增大时, 只有 ML 方法得到的参数的 MSE 呈显著的增大趋势, 而其他稳健估计方法得到的参数的 MSE 变化并不显著, 说明在这种情形下稳健估计方法一致有效。几类稳健方法比较而言, 提出的 RMG 方法会优于 RML 方法, 尤其在  $\sigma_e^2$  的 MSE 的估计中更加明显。

在污染情形  $(v, e)$  中, 若固定其污染比例为 0.1。其模型参数的 MSE 随随机效应的污染分布的方差的变化如图 6.11, 6.12。

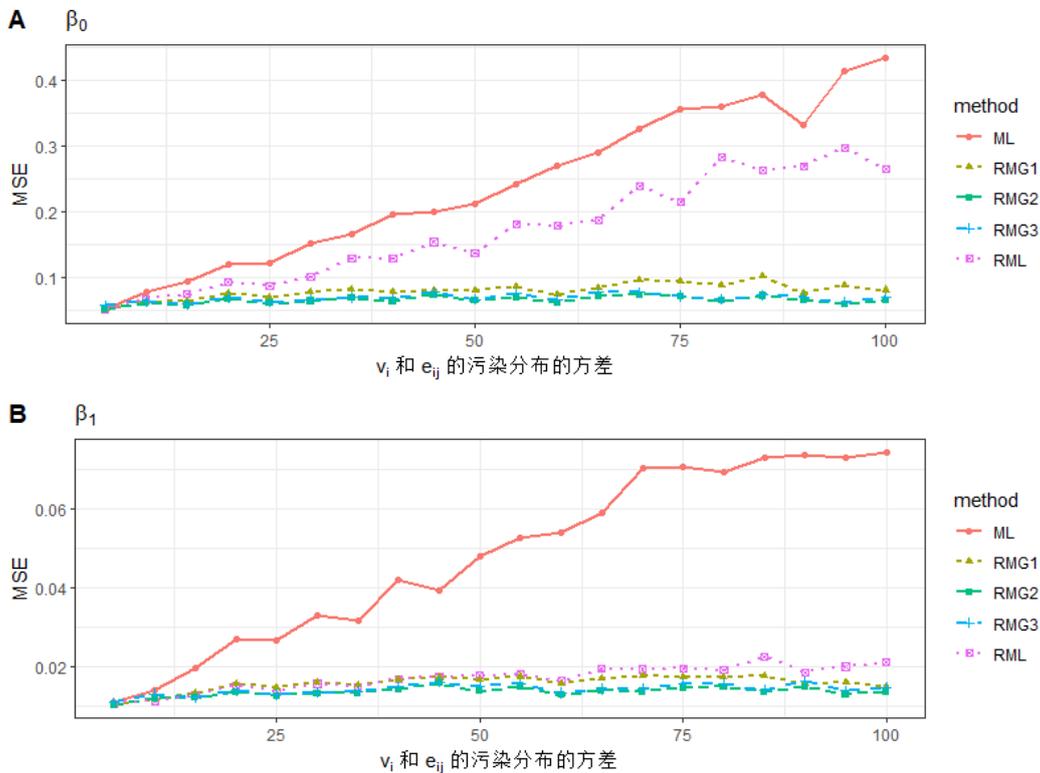


图 6.11 模型系数的 MSE 随污染分布方差变化图

从图中变化趋势可以得出, ML 估计方法对异常值比较敏感, 当污染方差增大时, 估计参数的 MSE 也随之增大。RML 方法较 ML 而言有所改善, 但对  $\beta_0, \sigma_v^2$  的估计中, MSE 仍会受污染方差变化的影响。而 RMG 方法对四个参数的估计均具有较好的稳健性。结合图形比较而言, RMG 方法要优于 ML 和 RML 方法。

综合上述几类估计方法, 可见无论是污染比例的增加, 还是污染分布方差的增加, 对模型系数的估计都会产生较大的影响, 某些情形下产生的影响非常显著。但综合几类稳健估计方法来看, 提出的 RMG 估计方法是有效的, 不仅能够

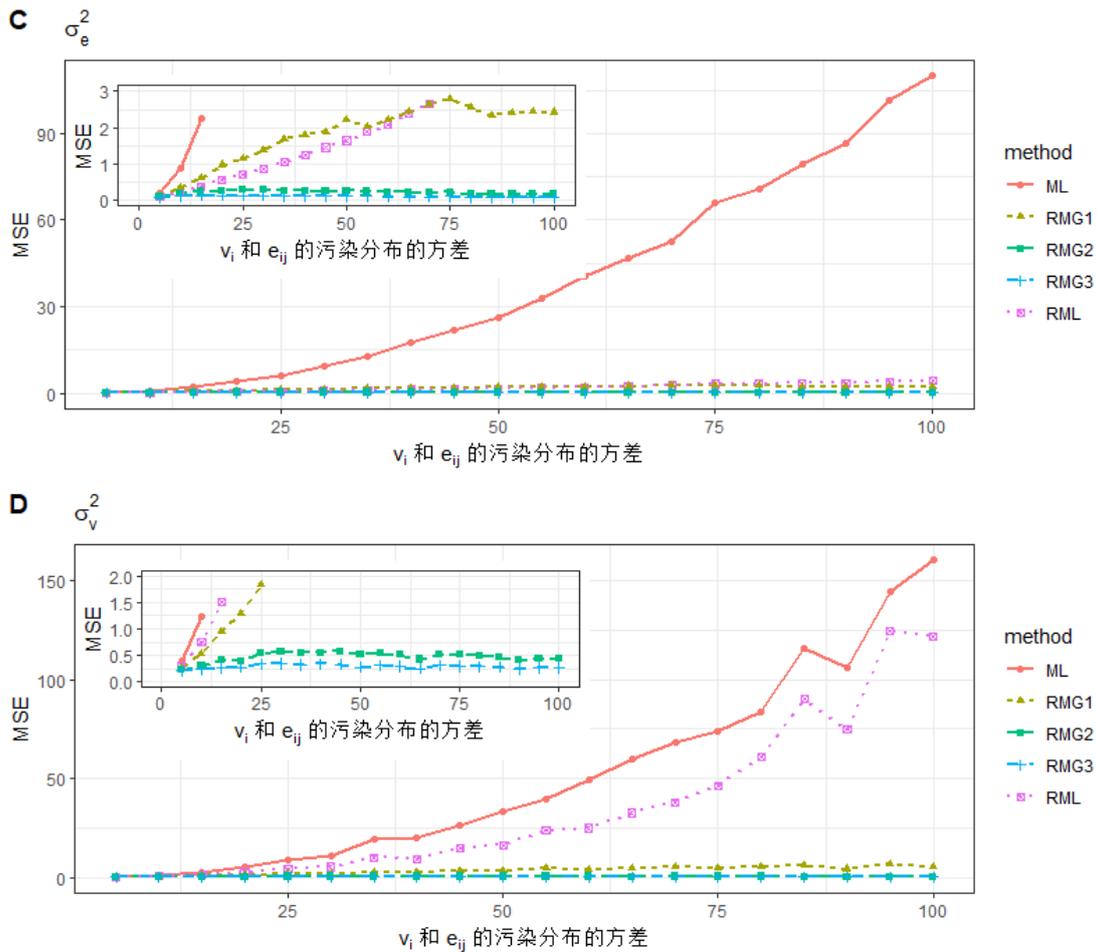
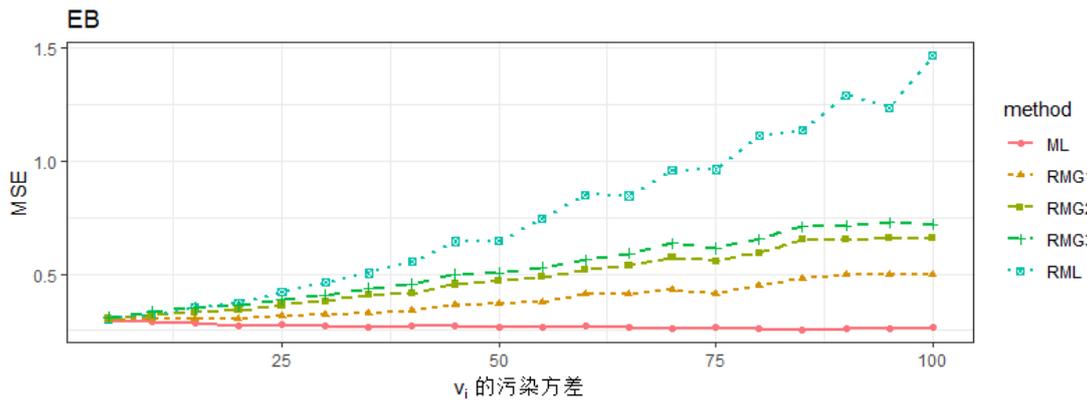


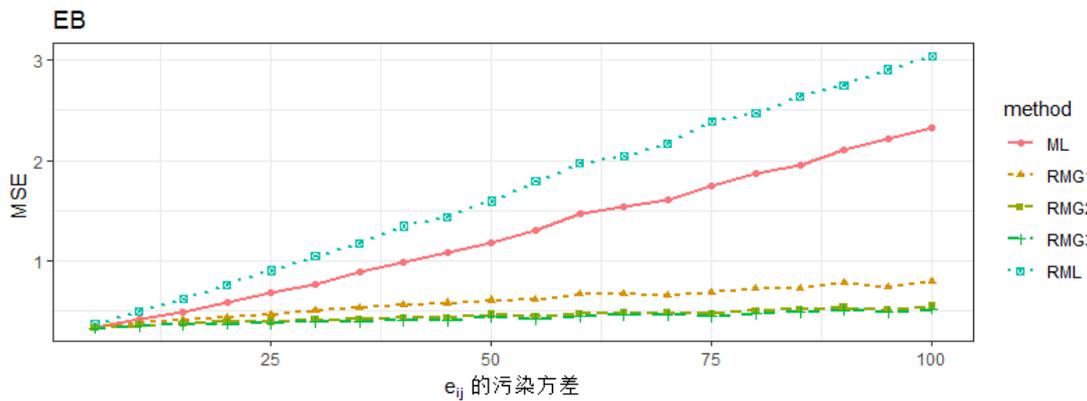
图 6.12 估计方差的 MSE 随污染分布方差变化图

使得系数估计的 MSE 比较小，同时对污染比例的增加和污染分布方差的增加表现并不敏感，在大多数情形下能达到稳健估计的目的。然而，除了估计系数具有一定的稳健性之外，还需要对目标变量的估计也具有一定的稳健性也是至关重要的，因此接下来考虑目标变量的估计的 MSE 的平均值随污染比例和污染分布方差的变化情形。

图6.13展示了当污染分布的方差增加时，污染情形分别为  $(v, 0)$ ,  $(0, e)$  时，目标变量 EB 估计值的平均 MSE 变化图。从图6.13a中可知，随着污染方差的增加，其 EB 估计的平均 MSE 呈增加趋势。其中 RML 估计的平均 MSE 增长较为明显，显著大于其它几类估计方法。ML 方法的估计 MSE 几乎没有增加趋势，说明该方法对区域随机效应的污染分布方差并不敏感，提出的 RMG 估计虽有上升趋势，但随着污染分布方差的增加其 MSE 的增加幅度较小，变化不大。图6.13b中可以看出，在  $e_{ij}$  被污染时，RML、ML 方法估计的 MSE 呈显著的增加趋势，且上升幅度较大，而提出的 RMG 方法估计的 MSE 增加较小，该种情形下，本研



(a)  $v_i$  被污染



(b)  $e_{ij}$  被污染

图 6.13 稳健 EB 估计的 MSE 随污染分布方差变化图

究提出的估计方法显著好于其它两类方法。

当污染比例增加时，图6.14给出了目标变量的 EB 估计量的平均 MSE 随污染比例的增加的变化情形。从6.14a中可以看出，随着污染比例的增加，几类方法估计的 MSE 均出现了不规则变化，比较而言，RML 估计的 MSE 较大，ML 估计方法的 MSE 较小，RMG1 的 MSE 和 ML 估计的结果较为接近，且其 MSE 比较小。当  $e_{ij}$  的污染比例增加时，EB 估计的平均 MSE 呈显著的增加趋势，且在该情形下，本研究提出的 RMG 的三种估计结果均好于 ML 和 RML 的估计，再次验证提出的估计方法的稳健性。

图6.15给出了在污染情形  $(v, e)$  下目标变量 EB 估计量的 MSE 随污染分布方差和污染分布比例的变化图。当区域随机效应和模型随机误差的分布均被污染时，从6.15a中可见，随着污染分布的方差的增加，平均 MSE 均有上升趋势。但其中只有 RML 估计的 MSE 增加较为显著，其它几类方法的估计结果均表现较为稳定，随着污染分布方差的增加，其 EB 估计的平均 MSE 变化并不大，体

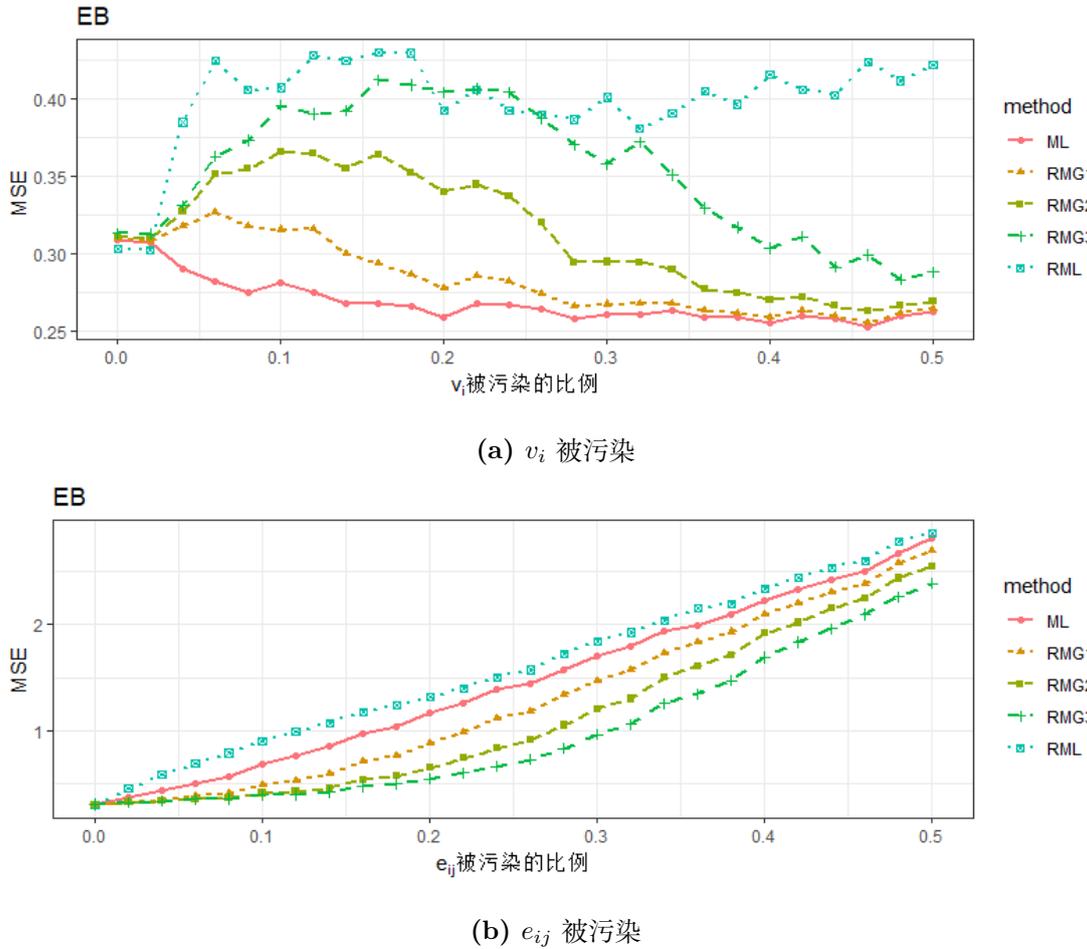
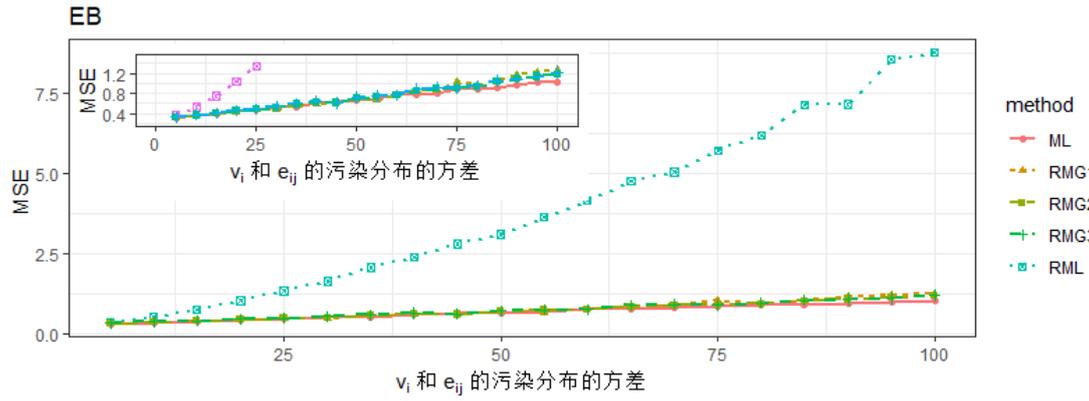


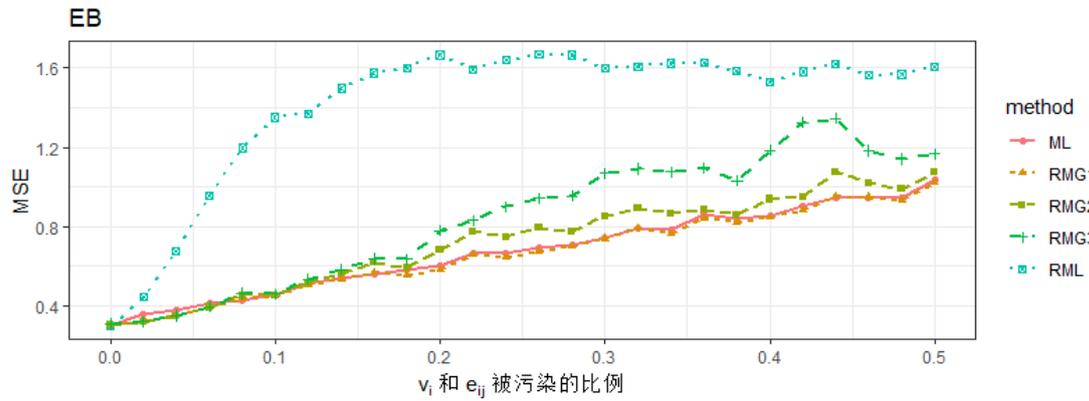
图 6.14 稳健 EB 估计的 MSE 随污染分布比例变化图

现出一定的稳健性。在图6.15b中，随着污染比例的增加，平均 MSE 有显著的上升趋势，且在几种方法中，RML 的估计 MSE 最大，提出的 RMG 估计方法和 ML 估计结果较为接近，增加幅度较小。

接下来考虑有限总体中小域均值的估计。比较在有限总体中用上述稳健估计方法得到的区域平均值的估计结果。设区域个数  $m=40$ ，考虑区域总体数量  $N_i$  分别为 40,80,200 的情形下总体均值估计的表现。从第  $i$  个区域的  $N_i$  个单元中选取  $n_i = 4$  个单元作为随机观测样本。对每次模拟的数据集，可以利用上述模拟中提到的稳健方法获取每个区域上观测变量的均值，最后比较 500 次模拟之后区域均值的平均估计效果。



(a)  $v_i$  被污染



(b)  $e_{ij}$  被污染

图 6.15 稳健 EB 估计的 MSE 随污染分布变化图

表 6.3 不同污染情形下有限总体均值估计的偏差及其 MSE

Ni	情形	ML	RML	RMG 1	RMG 2	RMG 3					
40	(0,0)	-0.0040	0.1908	-0.0032	0.1993	-0.0042	0.1898	-0.0043	0.1903	-0.0043	0.1911
	(v,0)	0.0004	0.3743	0.0017	0.2099	-0.0019	0.2687	-0.0012	0.2544	0.0001	0.2269
	(0,e)	0.0014	0.4769	0.0057	0.3335	0.0010	0.3064	0.0011	0.2937	0.0009	0.2703
	(v,e)	-0.0033	0.6536	0.0006	0.6140	-0.0024	0.4650	-0.0021	0.4417	-0.0034	0.4237
80	(0,0)	-0.0008	0.1984	0.0014	0.2077	-0.0003	0.1978	0.0002	0.1981	0.0008	0.1989
	(v,0)	-0.0020	0.3903	-0.0042	0.2242	-0.0026	0.2908	-0.0024	0.2735	-0.0022	0.2432
	(0,e)	0.0072	0.5008	0.0065	0.3289	0.0066	0.3454	0.0062	0.3296	0.0059	0.2984
	(v,e)	0.0032	0.6445	0.0042	0.6282	0.0010	0.4776	0.0003	0.4717	0.0007	0.4553
200	(0,0)	-0.0009	0.2032	-0.0004	0.2132	-0.0013	0.2032	-0.0018	0.2039	-0.0023	0.2052
	(v,0)	-0.0069	0.3934	-0.0048	0.2260	-0.0057	0.2971	-0.0042	0.2786	-0.0030	0.2472
	(0,e)	0.0068	0.4866	0.0028	0.3057	0.0046	0.3455	0.0031	0.2703	0.0023	0.2447
	(v,e)	0.0086	0.6592	0.0063	0.6077	0.0069	0.4316	0.0062	0.4979	0.0058	0.4606

从模拟结果上看,总的来说 RMG 方法在大部分情形下具有更小的 MSE,在某些情形下虽不及 RML 方法的估计效果,但相差不大。另外,当区域尺度变大时如  $N_i = 200$ , RMG 方法得到的估计量具有更小的 MSE,估计效果好于传统的估计方法。

### 6.8.2 实际数据

下面我们采用被 Battese et al.<sup>[3]</sup> 给出的数据来验证本研究提出的稳健估计方法。该数据可以从 R 包 sae 中获得,其中包含了来自于 12 个县区的玉米和大豆的面积样本数据 37 个,以及每个区域上的玉米和大豆的像素值。用农场采访收集的数据最为因变量,卫星数据作为辅助变量,建立单元水平模型。

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + v_i + \tilde{e}_{ij},$$

该模型可以作为模型 (1) 中  $k_{ij} = 1, x_{ij} = (1, x_{ij1}, x_{ij2})^T$  的特殊情形。其中  $y_{ij}$  表示第  $i$  个县第  $j$  个区域的玉米(大豆)的亩数,  $x_{ij1}$  和  $x_{ij2}$  分别表示第  $i$  个县第  $j$  个区域的玉米、大豆的像素值。

Battese et al.<sup>[3]</sup> 将 Hardin 县的值识别为离群值,并且在预测玉米和大豆的面积时只是将这一观测值简单的删除。<sup>[28]</sup> 中利用稳健估计方法对这一数据进行了分析,并给出了存在离群值时相应的预测值。在这里,我们运用本研究提出的方法对该数据进行建模分析,对每个区域上玉米种植面积的进行估计和预测。由于本数据中仅存在一个离群观测,因此在提出的稳健估计方法中,选取  $\gamma = 0.01, 0.05$  两种情形进行估计。在表 6.4 中给出了利用 ML 方法、RML 方法、以及本研究的 RMG 方法估计得到回归系数和随机误差的方差。同时表中括号内给出了根据渐近分布得到的每个参数的标准误。结合表中数据比较而言,在调整参数  $\gamma = 0.01$  时,用 RMG 方法估计得的系数介于 ML 方法和 RML 方法得到的估计值之间。当调整参数  $\gamma$  增加到 0.05 时,模型估计的系数有了显著的变化。通过比较表中展示的参数的标准误,可见用本研究的方法估计得到的参数具有更小的标准误。

表 6.4 模型参数的估计及其标准误

参数	ML	RML	RMG1	RMG2
截距 $\beta_0$	18.09(29.82)	28.68(27.30)	22.02(15.31)	36.88(14.41)
玉米像素 $\beta_1$	0.3657(0.0625)	0.3545(0.0574)	0.3614(0.0347)	0.3445(0.0335)
大豆像素 $\beta_2$	-0.0302(0.0650)	-0.0676(0.0596)	-0.0426(0.0350)	-0.0908(0.0329)
$\sigma_e^2$	280.2(71.55)	213.7(63.94)	274.5(70.35)	248.5(67.58)
$\sigma_v^2$	47.80(56.51)	113.7(64.57)	51.4(49.33)	65.2(55.09)

为了体现估计的效果，我们比较了几种估计方法对每个区域上大豆种植面积平均亩数的预测。采用上述估计方法得到的 EBLUP 值在表6.5中，其中括号内展示了用 500 个 Bootstrap 样本得到的估计量的 MSPE 的估计值。首先在估计结果上来看，RMG 对没有离群值区域的估计更接近的 ML 结果，并且对区域 Hardin 的预测有了一定的改进。其次，通过比较 Bootstrap MSPE，RMG 得到的 MSPE 值更小。可见，RMG 方法是有效的。

表 6.5 区域玉米种植面积的预测及其 Bootstrap MSE

县区	样本量	预测值及其 MSE			
		ML	RML	RMG1	RMG2
Cerro Gordo	1	122.2(7.9)	124.2(9.7)	122.4(6.6)	122.4(7.4)
Hamilton	1	123.2(7.3)	125.6(9.6)	123.5(6.3)	124.0(6.7)
Worth	1	113.9(7.4)	106.1(9.3)	113.7(5.4)	112.8(6.1)
Humboldt	2	115.4(7.4)	112.7(8.7)	115.1(5.5)	113.2(6.0)
Franklin	3	136.1(7.0)	143.2(7.3)	136.6(4.6)	138.3(4.9)
Pocahontas	3	108.4(7.3)	113.0(7.0)	108.5(4.7)	108.6(5.2)
Winnebago	3	116.8(7.0)	113.0(7.5)	116.8(4.9)	116.2(5.3)
Wright	3	122.6(6.8)	123.1(7.5)	122.5(4.9)	121.8(5.0)
Webster	4	110.9(6.4)	116.1(6.6)	111.1(4.3)	111.5(5.0)
Hancock	5	124.4(5.9)	122.6(6.1)	124.5(4.1)	124.7(4.5)
Kossuth	5	113.4(5.9)	105.2(6.2)	113.0(3.9)	111.4(4.6)
Hardin	6	131.3(5.8)	142.2(6.0)	131.5(4.1)	132.3(4.6)

## 6.9 小结

在本研究中,我们提出了一种针对存在离群观测值的单元水平模型小域稳健估计方法。通过引入最小化  $\gamma$  散度方法,给出了解决具有离群观测和非正态分布误差的稳健估计方法。首先给出了单元水平模型参数的估计方程和渐近性质。其次,结合参数的渐近分布,给出了最优调整参数的选择程序。再次,给出了有限总体中单元和区域均值的 EBLUP 值和估计量的 MSE。最后通过模拟数据和实际数据验证了提出方法的优越表现。在模拟部分,模拟了分布被污染时的稳健估计,讨论了三种污染情形下几类稳健估计方法的效果。还特别讨论了污染比例变化、以及污染分布的方差变化时几类估计方法的 MSE 的变化情况。同时模拟结果表明,本研究的方法能更好地解决这种离群情况。实际数据中,用一个小区域估计中很经典的数据说明提出方法不仅十分有效,还能够很好地处理离群观测这一特殊情况。进一步的验证表明,本研究的方法针对随机效应服从其他有偏分布时也是有效的。当随机误差的分布被污染,且污染概率大于 0.3 时,本研究的方法表现一般,比 Sinha et al.<sup>[28]</sup> 中的稳健估计方法差,但是在这种情形下,几类方法得到的 MSE 均很大,稳健估计结果都不太有价值。下一步,我们尝试将本方法进一步推广应用于指数分布情形的小区域估计问题。

## 6.10 应用研究

### 6.10.1 数据来源

本部分的应用数据来源于北京大学中国社会社科调查中心实施的中国家庭追踪调查 (China Family Panel Studies, CFPS) 项目,详细数据可见中国家庭追踪调查网站 <https://opendata.pku.edu.cn/dataverse/CFPS>。该项目旨在通过跟踪收集个体、家庭、社区三个层次的数据,反映中国社会、经济、人口、教育和健康的变迁,为学术研究和公共政策分析提供数据基础<sup>[149]</sup>。本研究中采用 2018 年的追踪调查数据,将其看成是一个实际的有限总体,并从中抽样形成小域特征的数据,检测本文提出的理论方法在该有限总体中的估计效果。

2018 年中国家庭追踪调查数据中的家庭经济调查部分的样本覆盖中国 31 个省/市/自治区,包含 14218 户家庭的样本数据。本研究采用其中的变量家庭纯收入 (fnicome) 作为小域目标变量,以家庭劳动人口数量 (fmlcount)、劳动力人口的教育水平 (edu)、消费性支出 (pce) 为辅助变量,检验前面章节中提到的稳健小域估计方法。在这里选取了和家庭纯收入相关的三个变量作为辅助变量,以省为区域划分变量。在统计到的数据中有些变量是无回答的,因此在样本数据中删

表 6.6 各省份的样本量

北京	天津	河北	山西	内蒙古	辽宁	吉林	黑龙江	上海	江苏	浙江
128	95	715	539	9	1226	246	394	707	251	287
安徽	福建	江西	山东	河南	湖北	湖南	广东	广西	海南	重庆
254	142	243	647	1328	195	375	1157	251	7	151
四川	贵州	云南	西藏	陕西	甘肃	青海	宁夏	新疆		
627	374	379	5	245	1504	2	6	33		

除这些无回答样本，总共得到 12522 个有效样本。样本在各个区域上的分布如下表 6.6 所示。

从表 6.6 中的数据可见，有部分省份的样本量较少，因此在抽取样本时，本文中对小样本省份采取全部抽样，而其它省份抽取 20 个样本作为小域上的样本量，并依此为小域样本数据。

### 6.10.2 小域模型及其估计

结合小域样本数据，建立单元水平模型

$$fincome_{ij} = \beta_0 + \beta_1 fmlcount_{ij} + \beta_2 edu_{ij} + \beta_3 pce_{ij} + v_i + e_{ij} \quad (6.9)$$

其中  $i = 1, 2, \dots, 31, j = 1, 2, \dots, n_i$ , 且  $n_i$  表示第  $i$  个区域上的样本数量。

对原始数据进行分析之前，对数据进行预处理，这里主要包括无回答样本的删除，数据的标准化等。为了表现变量间的关系，绘制出变量纯收入与变量消费性支出之间的散点图 (6.16)，从散点图不难发现，观测数据中存在大量的异常观测，虽然二者呈现出一定的关系，但是在离群值存在情形下显得这种关系并不突出。进一步绘制变量家庭纯收入的 QQ 图 (6.17)，从 QQ 图可以看出，目标变量并不服从正态分布，因此该数据不满足传统的 NER 模型中的正态性假设条件。因此需要进一步探求稳健估计方法，下面我们探究几类稳健方法的估计效果。

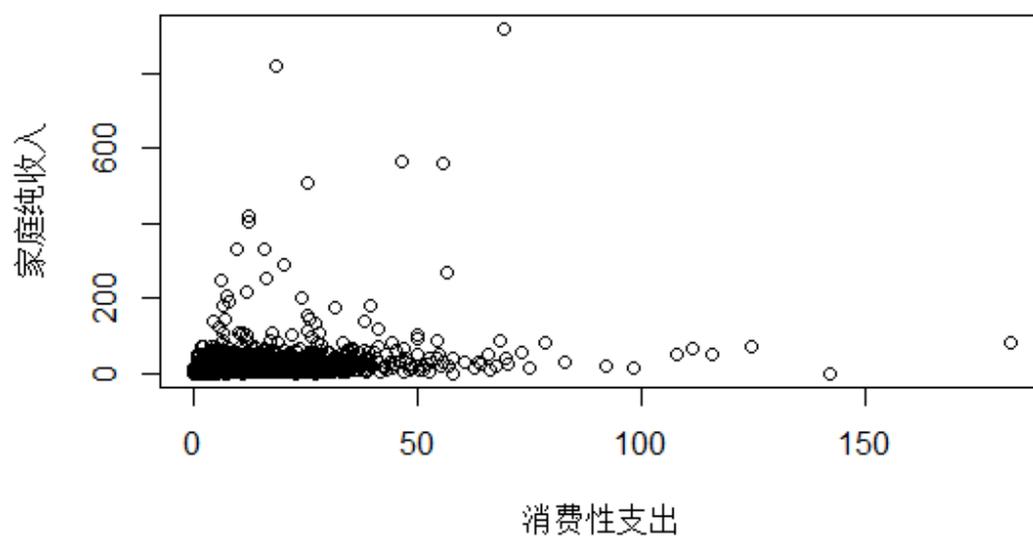


图 6.16 家庭纯收入与消费性支出之间的散点图

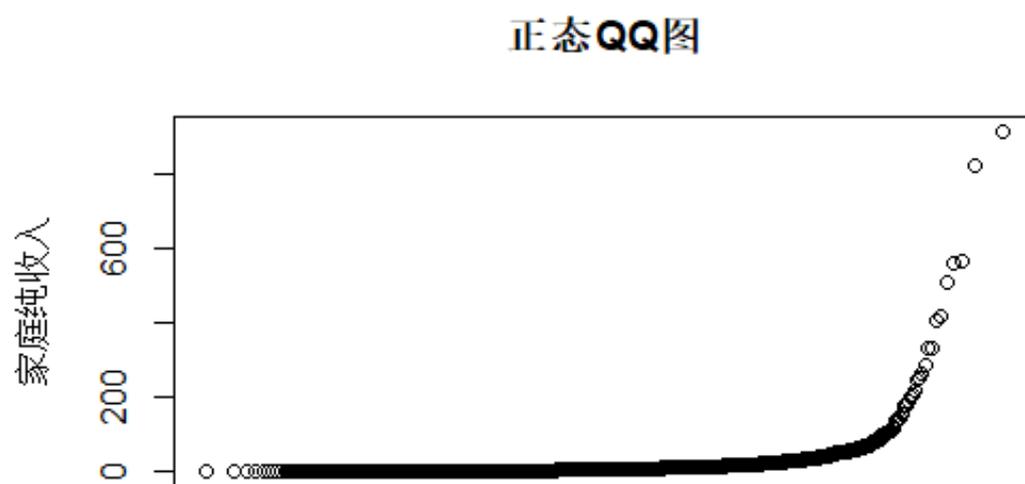


图 6.17 家庭纯收入的 QQ 分布图

分别用 ML 方法, RML 方法、DPD 方法以及  $\gamma$  散度方法 (GRML) 来构建小域估计模型, 得到小域模型的系数估计, 见表6.7。

表 6.7 模型系数的估计

变量	ML	RML	MDPD	GRML
Intercept	3.452	3.113	3.220	2.458
edu	0.020	0.150	0.190	0.267
pce	1.524	0.257	0.306	0.373
fmlcount	-0.927	0.093	0.080	0.192

表6.7中得到的模型系数的估计均是显著的，即通过假设检验可以确定表中的估计值均对因变量有显著的影响。观察表中系数的估计值，三种稳健估计方法具有较为接近的系数估计，对于传统的 ML 估计方法，其估计系数与其它稳健估计方法有较大的差异。且对于变量家庭劳动人口数量 (fmlcount) 而言，其和家庭纯收入之间应该是正向的关系，而 ML 方法给出了负的系数估计，这与现实逻辑是不符合的，即系数的估计是不准确的。

在上述估计的基础上，进行残差分析，分别绘制 ML 方法和 RML 方法估计得到的残差的 QQ 图，见图 6.18 所示。从图中可见，ML 估计的残差中具有异常值，且残差不满足正态分布的假设。而在 RML 方法中，虽然这种估计有所改善，但模型的正态性假定仍然不满足，因此这两种模型下的小域估计并不稳健，因此本文中提出的估计方法的研究是非常有必要的。下面将几种方法的估计结果进行比较分析。

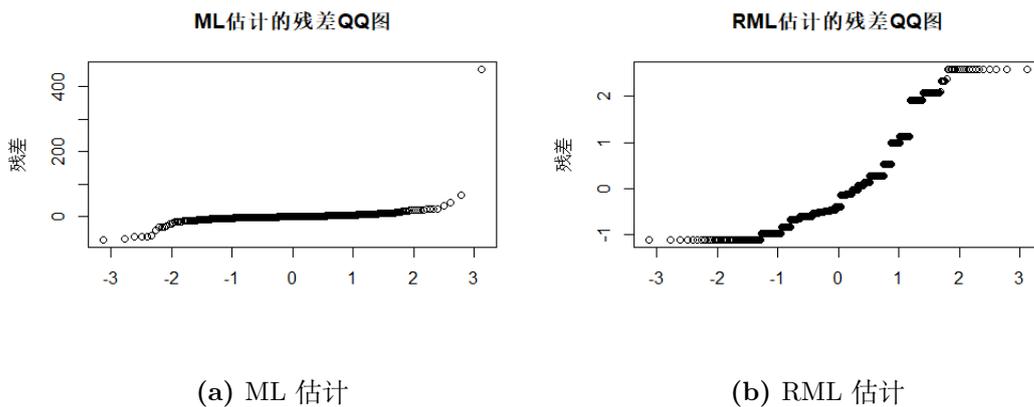


图 6.18 估计残差 QQ 图

根据上述拟合模型，来预测每个省份的家庭收入的平均值，并结合实际值来比较，比较几种方法在小域估计中的表现。

表 6.8 各省份家庭平均收入的小域估计

省份	平均家庭收入	ML	RML	MDPD	GRML
北京	13.17	17.13	6.96	8.119875	9.64
天津	10.32	12.03	6.30	8.49	9.59
河北	6.32	7.03	5.79	6.73	7.18
山西	5.77	7.79	5.76	5.40	5.74
内蒙古	4.90	7.85	5.35	5.32	5.76
辽宁	7.40	8.76	5.70	6.67	7.25
吉林	6.24	8.15	5.54	6.64	7.16
黑龙江	7.27	8.99	5.82	5.94	6.46
上海	21.77	7.38	12.59	14.66	17.19
江苏	14.83	15.13	6.89	9.50	10.99
浙江	16.02	7.57	20.46	23.63	19.06
安徽	9.56	10.47	6.24	9.98	11.12
福建	10.65	6.05	6.75	7.49	10.75
江西	7.11	8.07	5.79	7.35	7.93
山东	7.98	5.43	5.73	6.07	7.53
河南	7.22	7.71	5.93	6.39	6.84
湖北	12.17	13.25	6.65	8.91	10.16
湖南	9.74	12.20	6.51	7.26	8.22
广东	9.90	10.79	6.20	6.52	7.26
广西	6.04	6.58	5.60	6.20	6.51
海南	9.99	5.56	6.45	7.34	9.41
重庆	6.99	8.95	5.71	5.41	5.85
四川	6.47	8.09	6.08	6.63	7.15
贵州	7.69	8.97	6.00	5.91	6.39
云南	7.38	9.72	6.12	7.15	7.86
西藏	116.8	22.20	6.89	23.56	29.54
陕西	8.74	9.06	5.94	7.27	7.95
甘肃	6.29	7.55	6.02	6.06	6.48
青海	6.60	24.30	8.16	6.41	9.89
宁夏	3.87	7.53	4.55	4.33	4.56
新疆	7.99	9.74	5.29	5.85	6.37

表 6.9 各省份家庭平均收入估计的 MSE

省份	ML	RML	MDPD	GRML	省份	ML	RML	MDPD	GRML
北京	13.93	42.48	3.365	1.321	湖北	32.52	10.41	2.954	0.816
天津	4.483	18.08	3.882	0.2256	湖南	3.112	11.65	0.131	0.062
河北	8.385	0.5334	0.048	2.573	广东	9.631	14.62	3.307	0.061
山西	6.559	0.037	0.673	1.431	广西	0.081	0.301	0.601	1.129
内蒙古	3.846	1.172	0.057	1.374	海南	21.74	8.271	5.722	0.005
辽宁	0.002	3.723	0.239	0.246	重庆	1.730	2.112	0.858	0.0056
吉林	24.41	4.701	0.867	0.091	四川	0.358	0.303	4.838	0.043
黑龙江	2.311	2.907	0.595	0.398	贵州	0.789	3.077	1.364	0.203
上海	214.9	111.5	89.23	0.056	云南	15.06	1.915	3.931	0.152
江苏	66.13	12.34	26.60	1.225	西藏	1212	12139	7761	189.3
浙江	4.915	75.53	4.976	3.043	陕西	0.272	8.882	0.127	2.822
安徽	14.06	12.40	2.705	0.0038	甘肃	4.166	0.185	0.728	4.548
福建	6.319	21.85	3.546	4.435	青海	272.6	94.59	14.03	2.169
江西	0.569	1.989	0.714	0.0017	宁夏	2.359	0.195	2.485	1.262
山东	0.283	3.067	0.0112	0.775	新疆	3.090	7.999	0.288	1.365
河南	0.205	2.038	0.204	0.094					

根据辅助信息的取值, 利用几类小域估计方法得到的稳健估计结果见表6.8。从几类方法的比较上来看, 其 ML 方法估计的结果较真实值有较大的偏差, RML 方法有所改善, 从总体的估计偏差上来看, MDPD 方法和 GRML 方法的估计效果有所改善, 这一结论和前面理论分析的结果相一致。当然由于观测数据中存在较多的离群值, 部分省份的估计值仍然不够准确, 因此本方法仍然存在不足之处, 仍有改进的空间。

为评价几种估计方法的估计效果, 分别利用上述几种方法计算各个小域估计量的 MSE。在利用上述几类方法进行稳健估计时, 为了更加准确体现每种方法估计量的 MSE, 重复 100 次上述抽样过程, 每次重复中选择不同的样本作为小域估计的样本, 最后根据估计量计算每次估计的 MSE。其平均 MSE 的变化情况如表6.9所示。

结合表6.9中估计量 MSE 的变化趋势不难看出, 在样本量较小的区域, 其

MSE 相对较大。再比较几类估计方法的估计效果,从每个区域上的 MSE 值大小可以得出,ML 方法估计得到的 MSE 相对较大,RML 方法对 ML 估计的效果有所改善,本文中提出的 RDPD 方法和 GRML 的 MSE 均比现有的估计方法的 MSE 小,这一点和表6.8中小域目标估计量的结论是一致的。通过本例的应用说明本研究中提出的估计方法是有效的,在稳健小域估计中具有一定的效果。

## 第 7 章 结论与展望

本章将对本研究的内容进行简单的总结,给出主要研究结论。最后,本章给出了本研究后续的研究展望。

### 7.1 研究总结

本研究聚焦于稳健小域估计问题,主要的包含小域估计方法的比较、稳健小域估计方法的比较、密度幂散度族在小域估计模型中的应用等几个方面的研究。提出了一种针对存在离群观测值的单元水平模型稳健小域估计方法。本文通过引入 DPD 散度、 $\gamma$  散度,给出了解决具有离群观测和非正态分布误差的稳健估计方法。对主要结论进行归纳,包括如下几个方面:

第一,对基本的小域估计方法、小域稳健估计方法进行了比较研究。通过对现有小域估计方法的分析发现,小域估计量对区域随机效应的方差较为敏感。当观测数据存在离群值或者随机效应是有偏分布时,现有估计量具有较大的偏差,且容易受异常值影响。而现有的稳健估计方法中,一部分方法是针对于特殊情形下的小域模型进行估计的方法,不具有普适性。另一部分稳健估计方法容易出现过度收缩的情形。因此本文中引入了密度幂散度族的估计方法,给出了 DPD 散度和  $\gamma$  散度的定义、相关的性质以及如何进行估计等。

第二,本文研究了区域水平模型的稳健小域估计。首先将 DPD 散度引入了区域水平模型,通过推广似然函数的表达,给出了稳健估计量的具体公式。结合估计量的表达式,给出了估计量的 MSE 及其估计。其次,将  $\gamma$  散度应用于区域水平模型,给出了估计量的表达式和 MSE。最后,给出了模型调整参数的选择算法和估计量 MSE 估计的程序。通过模拟数据和实际数据,验证了本研究中提出方法的优越性。结果一致表明在适当的调整参数中,本文提出的方法对于模型的系数的估计和小域目标的估计均具有最小的 MSE,表现出了稳健性。

第三,本文研究了单元水平模型的稳健估计,提出了稳健估计方法。给出了模型参数的估计方法,给出了有限总体中单元和区域均值的 EBLUP 值和估计量的 MSE。通过模拟数据和实际数据验证了提出方法的优越表现。在模拟部分,模拟了分布被污染时的稳健估计,讨论了三种污染情形下几类稳健估计方法的效果。还特别讨论了污染比例变化、以及污染分布的方差变化时几类估计方法的 MSE 的变化情况。同时模拟结果表明,本研究中的方法能更好地解决这种离群

情况。实际数据中,用一个小区域估计中很经典的数据说明本文提出方法不仅十分有效,还能够很好地处理离群观测这一特殊情况。进一步的验证表明,本文的方法针对随机效应服从其他有偏分布时也是有效的。当随机误差的分布被污染,且污染概率大于 0.3 时,本文的方法表现一般,比 Sinha 和 Rao (2009) 中的稳健估计方法差,但是在这种情形下,几类方法得到的 MSE 均很大,稳健估计结果都不太有价值。最后,将本文中提出的理论方法应用于中国家庭追踪调查数据,进一步验证了本研究的理论结果。

在本研究的模拟实验部分,仅展示了模型误差项分布来自于混合正态分布时的估计结果。实际上,当模型误差项分布属于其它情形的有偏估计时,提出的估计方法都是适用的。例如当模型误差来自于  $t$  分布、对数正态分布、伽马分布等情形时,提出的稳健估计方法具有更小的偏差和均方误差,是一致有效的。这一点进一步说明该方法的普适性,能够在更一般的模型假设条件或存在离群观测时使用,并得到较为理想的估计结果。

## 7.2 应用场景

小域估计作为抽样调查研究的一个分支,在精细化的区域估计以及小域单元的估计中发挥着重要的作用,成为抽样调查中不可或缺的估计技术之一。本研究重点考虑小域估计模型随机效应和误差项有偏时的小域估计问题,提出了稳健估计方法。下面简单说明本研究中提出的方法的实际应用背景。

随着大数据时代的来临以及人工智能的出现,有人会担心是不是小域估计就没有存在的意义,这类方法会失去其存在的价值呢?答案当然是否定的。首先,区域目标变量的数据获取本身就不易,这类数据的收集通常需要花费较大的成本,这也是采用小域估计估计最本质的原因。其次,大数据本身的特点也为区域估计带来了挑战,如大数据的大容量使得处理数据变得异常困难、大数据的低密度性就难以保证获取高质量的数据,同时大数据通常是片段性的,获取的数据也通常不完整,还是需要借助于小域估计方法来获得更加精确的估计。再次,大数据的真实性以及如何获取目标变量的大数据样本都有待进一步的验证和研究。因此,小域估计在当下仍然具有一定的研究价值和实际意义。

基于模型的小域估计方法通过“借力”的方式达到小域变量的估计这一目标。因此获取和目标变量相关的辅助信息成为了小域估计的关键环节。通常这些辅助信息来源于一些更大型的调查或者行政记录,然而这些数据都有一定的时效性,高质量的辅助信息的获取就显得尤为重要。在大数据时代,可以很好地利

用大数据信息来获取小域辅助信息,虽然小域目标的获取可能非常困难,但是可以利用大数据获取与之相关的信息作为辅助信息,用于建立小域估计模型得到目标变量的小域估计。例如可以利用移动手机的使用信息作为辅助变量来估计全国各区域上人口流动率的研究;可以利用交通监控信息作为辅助信息估计各区域中交通事故发生率的估计;用卫星图像数据作为辅助信息估计全国各县区的粮食种植面积、产量等。足以说明,大数据时代下,小域估计的辅助信息获取会变得更加容易,小域估计方法也会具有更大的应用空间。

实际应用中,很多场景下都会遇到离群观测以及目标观测是有偏的情形,本研究中提到的稳健小域估计方法是解决这类小域估计的有力武器。同时,这类小域估计问题在社会、经济、教育、卫生等领域均有广泛的应用,为我国的经济统计发挥着重要的作用。

稳健小域估计方法可用于住户调查数据分析。在住户调查中,通常选取一部分住户作为样本用于调查分析,但这一部分住户通常比较少,属于典型的小域估计问题。例如在甘肃省榆中县住户调查数据中,仅仅包含了 130 户样本数据,这相比较于全县 46 万户居民数据就非常少,小样本乃至无样本问题就会非常突出。另外,住户调查中由于调查户分布均比较分散,因此很容易出现有偏乃至离群样本,这便是本研究中所设计方法的应用背景。结合住户调查数据的区域家庭人均收入、人均消费、恩格尔系数等关键指标的小域估计问题均可以用本研究中提到的方法。

稳健估计方法可用于社会调查的区域估计问题。在政府统计工作中,通常会关注某些区域上的失业率、贫困率等问题,因此在成本有限条件下的小域估计是非常有效的估计手段。由于失业、贫困这一类问题本身具有一定的特殊性,受众多因素的影响,从而使得观测数据中容易出现异常观测。在该情形下本研究提出的稳健估计方法能够得到非常有效的估计结果。

稳健估计方法可用于卫生健康调查的区域估计中。如某种疾病的发病率、传染率以及医疗保险覆盖率等。为了得到小域上这些指标的准确估计,可以借助于小域估计方法,利用本文中提出的稳健估计方法进行估计。

## 7.3 研究不足与展望

### 7.3.1 不足

本研究中通过将密度幂散度族应用于小域估计模型,改进了 BLUP 的精度和稳健性。但是,本研究仍然存在不足之处。主要有如下几方面:

(1) 本研究中将密度幂散度模型应用于基本小域模型,对于更加复杂的线性混合模型、广义线性混合模型的稳健估计没有深入讨论。在小域估计中,混合线性模型和广义线性混合模型也是非常重要的模型之一,有非常广泛的应用,如果将密度幂散度族方法应用于该类模型,也是一项非常有意义的研究。

(2) 针对污染非常严重的观测数据,本文中提出的方法对其估计效果一般。在本研究的模拟实验中发现,对于来自于混合正态总体的观测数据,当污染比例和污染方差变大时,现有的几种稳健估计量的 MSE 均呈现出发散趋势。虽然本文提出的估计量具有相对较小的 MSE,但是该情形下估计量 MSE 仍然较大(尽管这种趋势难以避免)。

(3) 在本文的模拟研究部分,仅展示了当随机效应、模型误差项是混合正态分布情形下的稳健估计结果。实际上,通过作者的模拟发现,当随机效应和模型误差服从其它的有偏分布( $t$ 分布、 $\Gamma$ 分布)时,本研究中提出的估计方法仍然具有较高的稳健性。这些模拟结果在本文中并没有一一列举。

### 7.3.2 展望

本文通过将稳健估计中的密度幂散度应用于小域估计问题,得到了基本小域估计模型的稳健估计。在后续的研究中,可以聚焦于如下几个方面的研究。

可以尝试将本研究方法进一步推广应用于指数分布情形的小区域估计问题。在小域估计中,混合线性模型和广义线性混合模型也是非常重要的模型之一,有非常广泛的应用背景。将本文中提出的基于密度幂散度族的小域估计方法应用于混合线性模型和广义线性模型,进一步完善稳健小域估计的理论。也为更加复杂的应用场景提供可选的估计方法。

在本研究的基础上,还可以做如下工作:有关小域稳健估计的应用方面,Tang et al.<sup>[150]</sup>提出了一种全局——局部收缩限样的方法来刻画小域估计中的随机效应,在贝叶斯分析的框架下给出了参数和目标变量的后验估计。与本文提出的稳健估计不同,该方法主要适用于区域数量较大的区域水平模型。Kurisu et al.<sup>[151]</sup>利用了和本文同样的估计方法,给出了区域水平模型的置信区间的估计。类似地,可以在本文的基础上给出单元水平模型的置信区间估计。在本文的稳健估计方法中,其中调整参数的选取采用了 IWJ 算法,而 Sugasawa et al.<sup>[94]</sup>给出了最小化密度幂类方法的调整参数选择的另外一种方法。因此,本文的研究框架下,同样可以采用 Sugasawa 提出的方法来选择调整参数,从而达到稳健小域估计的目的。

在小域估计的稳健估计中, 可以尝试应用模型平均的方法。Bates et al.<sup>[152]</sup> 提出模型平均的方法比起单一的估计模型具有更高的估计效率。此后, 模型平均方法在统计预测中得到了广泛应用, 其中较为典型的应用方法有 Hansen<sup>[153]</sup> 给出的最小二乘模型平均方法、预测模型平均等, Zhang<sup>[154]</sup> 完善了最小二乘模型平均的理论基础。为提高小域估计的准确度, 模型平均的方法在小域估计中的应用研究将是一个非常有意义工作。然而目前鲜有学者致力于这种方法的研究, 因而基于模型平均方法的小域估计研究充满了挑战和机遇。另外, 基于机器学习和深度学习方法的小域估计研究也是今年来的研究热点, 如 Kontokosta et al.<sup>[155]</sup>, Singleton et al.<sup>[156]</sup> 等人将机器学习的方法应用于小域估计中探究小域上的应用问题。



## 参考文献

- [1] RAO J N K, MOLINA I. Small area estimation[M]. New York: John Wiley & Sons, 2015.
- [2] FAY III R E, HERRIOT R A. Estimates of income for small places: an application of james-stein procedures to census data[J]. Journal of the American Statistical Association, 1979, 74(366a): 269–277.
- [3] BATTESE G E, HARTER R M, FULLER W A. An error-components model for prediction of county crop areas using survey and satellite data[J]. Journal of the American Statistical Association, 1988, 83(401): 28–36.
- [4] RONCHETTI E M, HUBER P J. Robust statistics[M]. Hoboken, NJ, USA: John Wiley & Sons, 2009.
- [5] ROJAS-PERILLA N, PANNIER S, SCHMID T, et al. Data-driven transformations in small area estimation[J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 2020, 183(1): 121–148.
- [6] GUADARRAMA M, MOLINA I, RAO J. A comparison of small area estimation methods for poverty mapping[J]. Statistics in Transition New Series, 2016, 17(1): 41–66.
- [7] ESTEBAN M D, LOMBARDÍA M J, LÓPEZ-VIZCAÍNO E, et al. Small area estimation of average compositions under multivariate nested error regression models [J]. Test, 2023: 1–26.
- [8] ARIAS-SALAZAR A, GUTIÉRREZ A, MANCERO X, et al. Small area estimation for composite indicators: the case of multidimensional poverty incidence[J]. arXiv preprint arXiv:2304.03901, 2023.
- [9] HOSSEINI S E, SHAHSAVANI D, RABIEI M R, et al. Small area estimation using a semiparametric spatial model with application in insurance[J]. Symmetry, 2022, 14(10): 2194. DOI: [10.3390/sym14102194](https://doi.org/10.3390/sym14102194).
- [10] MÜNNICH R, BURGARD P, VOGT M. Small area estimation for population counts in the german census 2011[J]. Section on Survey Research Methods JSM, 2009: 181–190.
- [11] KREUTZMANN A K, MAREK P, RUNGE M, et al. The fay–herriot model for multiply imputed data with an application to regional wealth estimation in germany[J]. Journal of Applied Statistics, 2022, 49(13): 3278–3299.

- [12] YOUNG L J, CHEN L. Using small area estimation to produce official statistics [J]. *Stats*, 2022, 5(3): 881–897.
- [13] GREEN P C, HOGG D W, WATSON B, et al. Small area estimation in diverse timber types using multiple sources of auxiliary data[J]. *Journal of Forestry*, 2022, 120(6): 646–659.
- [14] ANJOY P, CHANDRA H. Spatial nonstationary hierarchical bayes estimation of small area proportions[J]. *Communications in Statistics-Theory and Methods*, 2023, 52(7): 2161–2181.
- [15] WÜRZ N, SCHMID T, TZAVIDIS N. Estimating regional income indicators under transformations and access to limited population auxiliary information[J]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2022, 185: 1679–1706.
- [16] ZHOU Y, LIU Y. The geography of poverty: Review and research prospects[J]. *Journal of Rural Studies*, 2022, 93: 408–416.
- [17] GHOSH T, GHOSH M, MAPLES J J, et al. Multivariate global-local priors for small area estimation[J]. *Stats*, 2022, 5(3): 673–688.
- [18] MOLINA I, CORRAL P, NGUYEN M. Estimation of poverty and inequality in small areas: review and discussion[J]. *TEST*, 2022: 1–24.
- [19] 周巍, 朱荣, 张锦水. 遥感辅助的农作物种植面积小域估计方法研究[J]. *统计研究*, 2015(7): 81–86.
- [20] 巩红禹, 王若男, 王春枝. 小域估计及基准校正贫困测度中的应用[J]. *统计与决策*, 2021(10): 42–46.
- [21] 王小宁. 小域估计在问卷分割中的应用研究[J]. *统计与信息论坛*, 2021, 36(9): 3–10.
- [22] 赵建华. 小域估计方法在小品种农作物播种面积统计中的应用[J]. *统计研究*.
- [23] 张圆, 周兰兰. 分层贝叶斯模型下我国疾病死亡率的小域估计方法[J]. *统计与决策*, 2019, 35(19): 15–19.
- [24] 赵喜仓, 康俊良. 小区域估计方法在政府统计中的应用研究[J]. *中国集体经济*, 2014 (15): 63–64.
- [25] 于力超. 小区域贫困指标估计方法的比较与模拟[J]. *统计与决策*, 2020, 36(11): 32–36.
- [26] 刘乐平, 潘松权, 任晓美. 基于分层贝叶斯分析的残疾率小域估计方法[J]. *统计研究*, 2010, 27(3): 83–88.
- [27] GHOSH M, MAITI T, ROY A. Influence functions and robust bayes and empirical bayes small area estimation[J]. *Biometrika*, 2008, 95(3): 573–585.

- [28] SINHA S K, RAO J N K. Robust small area estimation[J]. Canadian Journal of Statistics, 2009, 37(3): 381–399.
- [29] GHOSH M, RAO J. Small area estimation: an appraisal[J]. Statistical Science, 1994, 9(1): 55–76.
- [30] GHOSH M, NANGIA N, KIM D H. Estimation of median income of four-person families: a bayesian time series approach[J]. Journal of the American Statistical Association, 1996, 91(436): 1423–1431.
- [31] RAO J N, YU M. Small-area estimation by combining time-series and cross-sectional data[J]. Canadian Journal of Statistics, 1994, 22(4): 511–528.
- [32] COCHRAN W G. Sampling techniques[M]. New York: John Wiley & Sons, 1977.
- [33] RAO J. On deriving mean square errors and their non-negative unbiased estimators in finite population sampling[J]. Journal of the Indian Statistical Association, 1979, 17: 125–136.
- [34] DEVILLE J C, SÄRNDAL C E. Calibration estimators in survey sampling[J]. Journal of the American Statistical Association, 1992, 87(418): 376–382.
- [35] HANSEN M H, HURWITZ W N, MADOW W G. Sample survey methods and theory[J]. Methods and Applications, 1953, 1: 15–32.
- [36] LEVY P S. The use of mortality data in evaluating synthetic estimates[J]. Proceedings of the American Statistical Association, Social Statistics Section, 1971, 328: 331.
- [37] GONZALEZ M E, WAKSBERG J. Estimation of the error of synthetic estimates [M]. Suitland: US Census Bureau [custodian], 1973.
- [38] SCHAIBLE W L. Choosing weights for composite estimators for small area statistics[J]. Proceedings of the section on survey research methods, American Statistical Association, 1978: 741–746.
- [39] PURCELL N P, KISH L. Estimation for small domains[J]. Biometrics, 1979, 35(2). DOI: [10.2307/2530340](https://doi.org/10.2307/2530340).
- [40] JIANG J, LAHIRI P, WAN S M. A unified jackknife theory for empirical best prediction with m-estimation[J]. The Annals of Statistics, 2002, 30(6): 1782–1810.
- [41] DATTA G S, GHOSH M. Bayesian prediction in linear models: Applications to small area estimation[J]. The Annals of Statistics, 1991: 1748–1770.
- [42] GHOSH A, BASU A. Robust estimation in generalized linear models: the density power divergence approach[J]. Test, 2016, 25(2): 269–290.
- [43] DATTA G S, LAHIRI P. Robust hierarchical bayes estimation of small area

- characteristics in the presence of covariates and outliers[J]. *Journal of Multivariate Analysis*, 1995, 54(2): 310–328.
- [44] DATTA G S. Model-based approach to small area estimation[J]. *Handbook of statistics*, 2009, 29: 251–288.
- [45] RODRIGUEZ M, LEIVA V, HUERTA M, et al. An asymmetric area model-based approach for small area estimation applied to survey data[J]. *REVSTAT-Statistical Journal*, 2021, 19(3): 399–420.
- [46] FAY R E. Application of multivariate regression to small domain estimation[J]. *Small Area Statistics*, 1987: 91–102.
- [47] BILTON P A. Tree-based models for poverty estimation: a thesis presented in partial fulfilment of the requirements for the degree of doctor of philosophy in statistics at massey university, manawatu[D]. [S.l.]: Massey University, 2016.
- [48] CHAKRABORTY A, DATTA G S, MANDAL A. Robust hierarchical bayes small area estimation for the nested error linear regression model[J]. *International Statistical Review*, 2018, 87(S1): S158–S176. DOI: [10.1111/insr.12283](https://doi.org/10.1111/insr.12283).
- [49] 于力超. 基于模型的小域估计方法及其拓展——由单变量推广到多变量情形[J]. *数理统计与管理*, 2019, 38(03): 473–482.
- [50] 胡玉琴. 域与小域估计理论的综述[J]. *统计与决策*, 2009(14): 157–160.
- [51] ERICKSEN E P, KADANE J B. Estimating the population in a census year 1980 and beyond[J]. *Journal of the American Statistical Association*, 1985, 80(389): 98–109.
- [52] DICK P. Modelling net undercoverage in the 1991 canadian census[J]. *Survey Methodology*, 1995, 21(1): 45–54.
- [53] GHOSH M. Hierarchical and empirical multivariate bayes analysis in small[C]// *Proceedings: Annual Research Conference*. Suitland: US Department of Commerce, Bureau of the Census, 1991: 63.
- [54] 胡丹丹, 金勇进, 艾小青. 解决小域估计的一种新思路: 样本外推[J]. *调研世界*, 2014(9): 44–46.
- [55] 李腾魏传华苏宇楠. 基于空间误差分量 fay-herriot 模型的小域估计[J]. *统计与信息论坛*, 2018, 033(011): 11–16.
- [56] 张会敏. 小域估计理论的国际动态与应用[J]. *统计与决策*, 2015(5): 32–36.
- [57] BURGARD J P, ESTEBAN M D, MORALES D, et al. Small area estimation under a measurement error bivariate fay-herriot model[J]. *Statistical Methods & Applications*, 2021, 30: 79–108.

- [58] STUKEL D M, RAO J. On small area estimation under two-fold nested error regression models[J]. *Journal of Statistical Planning and Inference*, 1999, 78(1-2): 131–147.
- [59] GHOSH M, NATARAJAN K, STROUD T, et al. Generalized linear models for small-area estimation[J]. *Journal of the American Statistical Association*, 1998, 93(441): 273–282.
- [60] MORI L, FERRANTE M R. Small area estimation under unit-level generalized additive models for location, scale and shape[J]. *arXiv preprint arXiv:2302.00108*, 2023.
- [61] FABRIZI E, FERRANTE M R, TRIVISANO C. Bayesian small area estimation for skewed business survey variables[J]. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 2018, 67(4): 861–879.
- [62] MANANDHAR B, NANDRAM B. Hierarchical bayesian models for continuous and positively skewed data from small areas[J]. *Communications in Statistics-Theory and Methods*, 2021, 50(4): 944–962.
- [63] JIANG J, RAO J S. Robust small area estimation: An overview[J]. *Annual Review of Statistics and Its Application*, 2020, 7: 337–360.
- [64] CHAMBERS R, CHANDRA H, SALVATI N, et al. Outlier robust small area estimation[J]. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 2014: 47–69.
- [65] CHAMBERS R L. Outlier robust finite population estimation[J]. *Journal of the American Statistical Association*, 1986, 81(396): 1063–1069.
- [66] GHOSH M, LAHIRI P. Robust empirical bayes estimation of means from stratified samples[J]. *Journal of the American Statistical Association*, 1987, 82(400): 1153–1162.
- [67] BELL W R, HUANG E T. Using the t-distribution to deal with outliers in small area estimation[C]//*Proceedings of Statistics Canada Symposium*. Ottawa, Canada: Statistics Canada, 2006: 45–52.
- [68] GHOSH M, MAITI T, ROY A. Influence functions and robust bayes and empirical bayes small area estimation[J]. *Biometrika*, 2008, 95(3): 573–585.
- [69] MOLINA I, MARHUENDA Y. sae: an r package for small area estimation.[J]. *R J.*, 2015, 7(1): 81.
- [70] CHAMBERS R, TZAVIDIS N. M-quantile models for small area estimation[J]. *Biometrika*, 2006, 93(2): 255–268.

- [71] BIANCHI A, FABRIZI E, SALVATI N, et al. Estimation and testing in m-quantile regression with applications to small area estimation[J]. *International Statistical Review*, 2018, 86(3): 541–570.
- [72] FABRIZI E, SALVATI N, TRIVISANO C. Robust bayesian small area estimation based on quantile regression[J]. *Computational Statistics & Data Analysis*, 2020, 145: 106900. DOI: <https://doi.org/10.1016/j.csda.2019.106900>.
- [73] OPSOMER J D, CLAESKENS G, RANALLI M G, et al. Non-parametric small area estimation using penalized spline regression[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2008, 70(1): 265–286.
- [74] GERSHUNSKAYA J, LAHIRI P. Robust small area estimation using a mixture model[C]//*Proceedings of the Section on Survey Research Methods, American Statistical Association*. Washington, USA: Section on Survey Research Methods – JSM, 2010: 2783–2796.
- [75] GERSHUNSKAYA J, LAHIRI P. Robust empirical best small area finite population mean estimation using a mixture model[J]. *Calcutta Statistical Association Bulletin*, 2017, 69(2): 183–204.
- [76] 刘展, 潘莹丽, 蔡雯;. 混合样本的小域模型推断研究[J]. *统计与决策*, 2021, 37(20): 5–9.
- [77] SUGASAWA S. Robust empirical bayes small area estimation with density power divergence[J]. *Biometrika*, 2020, 107(2): 467–480.
- [78] 庞智强, 王朝旭, 牛玺娟. 基于  $\gamma$  散度的单元水平模型小域稳健估计[J]. *统计与信息论坛*, 2023, 38(03): 3–15.
- [79] 金勇进, 赵雪慧. 对抽样调查解决多层次估计问题的探讨[J]. *统计研究*, 2003(12): 55–58.
- [80] 冯士雍. 抽样调查应用与理论中的若干前沿问题[J]. *统计与信息论坛*, 2007(1): 5–13.
- [81] 吕萍. 分位数回归模型在小域估计中的应用[J]. *统计教育*, 2009(1): 56–59.
- [82] 贺建风, 付永超, 熊健. 基于分层贝叶斯广义线性模型的小域估计方法研究[J]. *数理统计与管理*, 2019(2): 247–260.
- [83] 李腾, 魏传华, 于力超. 基于混合地理加权 fay-herriot 模型的小域估计[J]. *应用数学*, 2019(2): 339–348.
- [84] 刘小惠, 马海强, 蒋继明. 存在潜在模型偏误时小域估计的不确定性评价: 一带一路法[J]. *中国科学: 数学*, 2022, 52(07): 779–808.
- [85] BASU A, HARRIS I R, HJORT N L, et al. Robust and efficient estimation by minimising a density power divergence[J]. *Biometrika*, 1998, 85(3): 549–559.

- [86] GHOSH A, BASU A. Bootstrap payes estimation using the density power divergence [J]. *Annals of the Institute of Statistical Mathematics*, 2016, 68(2): 413–437.
- [87] RIANI M, ATKINSON A C, CORBELLINI A, et al. Robust regression with density power divergence: Theory, comparisons, and data analysis[J]. *Entropy*, 2020, 22(4): 399. DOI: [10.3390/e22040399](https://doi.org/10.3390/e22040399).
- [88] JONES M, HJORT N L, HARRIS I R, et al. A comparison of related density-based minimum divergence estimators[J]. *Biometrika*, 2001, 88(3): 865–873.
- [89] FUJISAWA H, EGUCHI S. Robust parameter estimation with a small bias against heavy contamination[J]. *Journal of Multivariate Analysis*, 2008, 99(9): 2053–2081.
- [90] NAKAGAWA T, HASHIMOTO S. Robust bayesian inference via  $\gamma$ -divergence[J]. *Communications in Statistics Theory and Methods*, 2020, 49(2): 343–360.
- [91] KAWASHIMA T, FUJISAWA H. Robust and sparse regression via  $\gamma$ -divergence [J]. *Entropy*, 2017, 19(11): 608.
- [92] WARWICK J, JONES M C. Choosing a robustness tuning parameter[J]. *Journal of Statistical Computation and Simulation*, 2005, 75(7): 581–588.
- [93] BASAK S, BASU A, JONES M. On the ‘optimal’ density power divergence tuning parameter[J]. *Journal of Applied Statistics*, 2021, 48(3): 536–556.
- [94] SUGASAWA S, YONEKURA S. On selection criteria for the tuning parameter in robust divergence[J]. *Entropy*, 2021, 23(9): 1147. DOI: [10.3390/e23091147](https://doi.org/10.3390/e23091147).
- [95] HENDERSON C R, KEMPTHORNE O, SEARLE S R, et al. The estimation of environmental and genetic trends from records subject to culling[J]. *Biometrics*, 1959, 15(2): 192–218.
- [96] PRASAD N N, RAO J N. The estimation of the mean squared error of small-area estimators[J]. *Journal of the American Statistical Association*, 1990, 85(409): 163–171.
- [97] KACKAR R N, HARVILLE D A. Unbiasedness of two-stage estimation and prediction procedures for mixed linear models[J]. *Communications in Statistics Theory and Methods*, 1981, 10(13): 1249–1261.
- [98] KACKAR R N, HARVILLE D A. Approximations for standard errors of estimators of fixed and random effects in mixed linear models[J]. *Journal of the American Statistical Association*, 1984, 79(388): 853–862.
- [99] DATTA G S, LAHIRI P. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems[J]. *Statistica Sinica*, 2000, 10(2): 613–627.

- [100] DAS K, JIANG J, RAO J. Mean squared error of empirical predictor[J]. *The Annals of Statistics*, 2004, 32(2): 818–840.
- [101] PFEFFERMANN D, TILLER R. Bootstrap approximation to prediction mse for state–space models with estimated parameters[J]. *Journal of Time Series Analysis*, 2005, 26(6): 893–916.
- [102] HALL P, MAITI T. Nonparametric estimation of mean-squared prediction error in nested-error regression models[J]. *The Annals of Statistics*, 2006, 34(4): 1733–1750.
- [103] HALL P, MAITI T. On parametric bootstrap methods for small area prediction [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 68(2): 221–238.
- [104] GONZÁLEZ-MANTEIGA W, LOMBARDÍA M J, MOLINA I, et al. Bootstrap mean squared error of a small-area eblup[J]. *Journal of Statistical Computation and Simulation*, 2008, 78(5): 443–462.
- [105] RAO J. Eb and eblup in small area estimation[M]//*Empirical Bayes and Likelihood Inference*. New York,USA: Springer, 2001: 33–43.
- [106] MORALES D, ESTEBAN M D, PÉREZ A, et al. A course on small area estimation and mixed models: Methods, theory and applications in r[M]. New York,USA: Springer Nature, 2021.
- [107] TUKEY J W. A survey of sampling from contaminated normal distributions in i. olkin et al.(eds.) *contributions to probability and statistics*. hstanford[M]. California: Stanford University Press, 1960.
- [108] HUBER P J. Robust estimation of a location parameter[J]. *The Annals of Mathematical Statistics*, 1964, 35(1): 73–101.
- [109] HAMPEL F R. *Contributions to the theory of robust estimation*[M]. California: University of California, Berkeley, 1968.
- [110] CRESSIE N, HAWKINS D M. Robust estimation of the variogram: I[J]. *Journal of the international Association for Mathematical Geology*, 1980, 12(2): 115–125.
- [111] SCOTT D W. Parametric statistical modeling by minimum integrated square error[J]. *Technometrics*, 2001, 43(3): 274–285.
- [112] HAMPEL F R. Robust estimation: A condensed partial survey[J]. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1973, 27(2): 87–104.
- [113] HAMPEL F R. The influence curve and its role in robust estimation[J]. *Journal of the American Statistical Association*, 1974, 69(346): 383–393.

- [114] MARONNA R A, MARTIN R D, YOHAI V J, et al. Robust statistics: theory and methods (with r)[M]. New York: John Wiley & Sons, 2019.
- [115] HUBER P J. Robust statistics: volume 523[M]. New York: John Wiley & Sons, 2004.
- [116] FELLNER W H. Robust estimation of variance components[J]. *Technometrics*, 1986, 28(1): 51–60.
- [117] 金蛟, 李瞳辉, 徐帅帅, 等. 测量误差模型稳健方法及应用研究[J]. *统计研究*, 2021, 038(011): 150–160.
- [118] MATLOFF N S. Use of regression functions for improved estimation of means[J]. *Biometrika*, 1981, 68(3): 685–689.
- [119] JIANG J, NGUYEN T. Small area estimation via heteroscedastic nested-error regression[J]. *Canadian Journal of Statistics*, 2012, 40(3): 588–603.
- [120] RAO J N, SINHA S K, DUMITRESCU L. Robust small area estimation under semi-parametric mixed models[J]. *Canadian Journal of Statistics*, 2014, 42(1): 126–141.
- [121] LOMBARDÍA M J, SPERLICH S. Semiparametric inference in generalized mixed effects models[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2008, 70(5): 913–930.
- [122] DATTA G, DELAIGLE A, HALL P, et al. Semi-parametric prediction intervals in small areas when auxiliary data are measured with error[J]. *Statistica Sinica*, 2018, 28(4): 2309–2335.
- [123] JIANG J, LAHIRI P. Mixed model prediction and small area estimation[J]. *Test*, 2006, 15(1): 1–96.
- [124] HAWKINS D M. Identification of outliers: volume 11[M]. London: Chapman and Hall, 1980.
- [125] AGGARWAL C C. An introduction to outlier analysis[M]//*Outlier Analysis*. New York: Springer, 2017: 1–34.
- [126] HUBER P J. Robust statistics[M]//*International encyclopedia of statistical science*. New York: Springer, 2011: 1248–1251.
- [127] CHATRCHI G. Robust estimation of variance components in small area estimation[D]. Ottawa, Canada: Carleton University, 2012.
- [128] SCHOCH T. Robust unit-level small area estimation: A fast algorithm for large datasets[J]. *Austrian Journal of Statistics*, 2012, 41(4): 243–265.
- [129] ROUSSEEUW P J, HAMPEL F R, RONCHETTI E M, et al. Robust statistics:

- the approach based on influence functions[M]. New York: John Wiley & Sons, 2011.
- [130] PARDO L. Statistical inference based on divergence measures[M]. Florida: Chapman and Hall/CRC, 2018.
- [131] BREGMAN L M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming[J]. USSR Computational Mathematics and Mathematical Physics, 1967, 7(3): 200–217.
- [132] STUMMER W, VAJDA I. On bregman distances and divergences of probability measures[J]. IEEE Transactions on Information Theory, 2012, 58(3): 1277–1288.
- [133] BASU A, SHIOYA H, PARK C. Statistical inference: the minimum distance approach[M]. Florida: CRC press, 2011.
- [134] LINDSAY B G. Efficiency versus robustness: the case for minimum hellinger distance and related methods[J]. The Annals of Statistics, 1994, 22(2): 1081–1114.
- [135] GHOSH A, BASU A. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression[J]. Electronic Journal of statistics, 2013, 7: 2420–2456.
- [136] PATRA S, MAJI A, BASU A, et al. The power divergence and the density power divergence families: the mathematical connection[J]. Sankhya B, 2013, 75(1): 16–28.
- [137] BASU A, MANDAL A, MARTIN N, et al. Testing composite hypothesis based on the density power divergence[J]. Sankhya B, 2018, 80(2): 222–262.
- [138] ROY S, CHAKRABORTY K, BHADRA S, et al. Density power downweighting and robust inference: some new strategies[J]. arXiv preprint arXiv:1910.12372, 2019.
- [139] SARACENO G, GHOSH A, BASU A, et al. Robust estimation under linear mixed models: The minimum density power divergence approach[J]. arXiv preprint arXiv:2010.05593, 2020.
- [140] FUJISAWA H, EGUCHI S. Robust parameter estimation with a small bias against heavy contamination[J]. Journal of Multivariate Analysis, 2008, 99(9): 2053–2081.
- [141] JONES M, HJORT N L, HARRIS I R, et al. A comparison of related density-based minimum divergence estimators[J]. Biometrika, 2001, 88(3): 865–873.
- [142] NAKAGAWA T, HASHIMOTO S. Robust bayesian inference via  $\gamma$ -divergence[J]. Communications in Statistics Theory and Methods, 2019, 49(2): 343–360.

- [143] EFRON B. Tweedie' s formula and selection bias[J]. *Journal of the American Statistical Association*, 2011, 106(496): 1602–1614.
- [144] SUGASAWA S, KUBOKAWA T. Small area estimation with mixed models: a review[J]. *Japanese Journal of Statistics and Data Science*, 2020, 3(2): 693–720.
- [145] SINHARAY S, STERN H S. Posterior predictive model checking in hierarchical models[J]. *Journal of Statistical Planning and Inference*, 2003, 111(1-2): 209–221.
- [146] DIAO L, SMITH D D, DATTA G S, et al. Accurate confidence interval estimation of small area parameters under the fay-herriot model[J]. *Scandinavian Journal of Statistics*, 2014, 41(2): 497–515.
- [147] CHATTERJEE S, LAHIRI P, LI H. Parametric bootstrap approximation to the distribution of eblup and related prediction intervals in linear mixed models[J]. *Annals of Statistics*, 2008, 36: 1221–1245.
- [148] ARORA V, LAHIRI P. On the superiority of the bayesian method over the blup in small area estimation problems[J]. *Statistica Sinica*, 1997: 1053–1063.
- [149] 谢宇, 胡婧炜, 张春泥. 中国家庭追踪调查: 理念与实践[J]. *社会*, 2014, 34(02): 1–32.
- [150] TANG X, GHOSH M, HA N S, et al. Modeling random effects using global–local shrinkage priors in small area estimation[J]. *Journal of the American Statistical Association*, 2018, 113: 1476–1489.
- [151] KURISU D, ISHIHARA T, SUGASAWA S. Adaptively robust small area estimation: Balancing robustness and efficiency of empirical bayes confidence intervals [J]. *arXiv preprint arXiv:2108.11551*, 2021.
- [152] BATES J M, GRANGER C. The combination of forecasts[J]. *Journal of the Operational Research Society*, 1969, 20(4): 451–468.
- [153] HANSEN B E. Least squares model averaging[J]. *Econometrica*, 2007, 75: 1175–1189.
- [154] ZHANG X. A new study on asymptotic optimality of least squares model averaging [J]. *Econometric Theory*, 2020: 1–20.
- [155] KONTOKOSTA C E, HONG B, JOHNSON N E, et al. Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities[J]. *Computers, Environment and Urban Systems*, 2018, 70: 151–162.
- [156] SINGLETON A, ALEXIOU A, SAVANI R. Mapping the geodemographics of digital inequality in great britain: An integration of machine learning into small area estimation[J]. *Computers, Environment and Urban Systems*, 2020, 82: 101486.

- [157] LOHR S L, RAO J. Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models[J]. *Biometrika*, 2009, 96(2): 457–468.
- [158] LEHMANN E. Springer texts in statistics: Elements of large-sample theory[M]. New York: Springer, 2004.

## 论文中相关定理的证明

### A.1 第四章中的定理证明

**证明 1 (引理4.1的证明)** 由于

$$\begin{aligned} E \left\{ (y_i - x_i^T \beta)^c s_i^k \right\} &= \frac{V_i^{k\alpha}}{\{2\pi(A + D_i)\}^{1/2}} \int_{-\infty}^{\infty} (t - x_i^T \beta)^c \exp \left\{ -\frac{(k\alpha + 1)(t - x_i^T \beta)^2}{2(A + D_i)} \right\} dt \\ &= \frac{V_i^{k\alpha}}{(k\alpha + 1)^{1/2}} E(Z^c), \end{aligned}$$

其中  $Z \sim N\{0, (A + D_i)/(k\alpha + 1)\}$ 。因此, 当  $c$  是奇数时, 上述期望值为 0。另外, 当  $c = 2j, j = 0, 1, 2, \dots$  时, 有  $E(Z^{2j}) = (2j - 1)!!(A + D_i)^j (k\alpha + 1)^{-j}$ , 从而引理成立。

**证明 2 (定理4.1的证明)** 因  $\tilde{\theta}_i = E(\theta_i | y_i)$ , 有

$$\begin{aligned} E \left\{ (\tilde{\theta}_i^R - \theta_i)^2 \right\} &= E \left\{ (\tilde{\theta}_i - \theta_i)^2 \right\} + E \left\{ (\tilde{\theta}_i^R - \tilde{\theta}_i)^2 \right\} \\ &= \frac{AD_i}{A + D_i} + E \left\{ (\tilde{\theta}_i^R - \tilde{\theta}_i)^2 \right\} \equiv g_{1i}(A) + g_{2i}(A). \end{aligned}$$

由于  $\tilde{\theta}_i^R - \tilde{\theta}_i = (A + D_i)^{-1} D_i (y_i - x_i^T \beta) (1 - \pi_\gamma V_{i\gamma} s_i)$ , 故

$$g_{2i}(A) = \frac{D_i^2}{(A + D_i)^2} E \left\{ (y_i - x_i^T \beta)^2 (1 - \pi_\gamma V_{i\gamma} s_i)^2 \right\},$$

结合引理4.1的内容, 便可得  $g_{2i}(A)$  的表达式。

**证明 3 (定理4.2的证明)** 由于参数  $\hat{\beta}_\gamma$  和  $\hat{A}_\gamma$  是由估计方程 (4.6) 解得。为了方便, 记式 (4.6) 中得第一个公式为  $F_\beta$ , 第二个公式为  $F_A$ , 那么根据 (4.6) 有如下的结果:

$$\begin{aligned} \frac{\partial F_\beta}{\partial \beta^T} &= \pi(\gamma) \sum_{i=1}^m \left\{ V_{i\gamma} \frac{x_i x_i^T s_i}{(A + D_i)^2} (\gamma(y_i - x_i^T \beta)^2 - (A + D_i)) \right\}, \\ \frac{\partial F_A}{\partial \beta} &= \pi(\gamma) \sum_{i=1}^m \left\{ V_{i\gamma} \frac{x_i s_i (y_i - x_i^T \beta)}{(A + D_i)^3} \left( \gamma(y_i - x_i^T \beta)^2 - \frac{2 + 3\gamma}{1 + \gamma} (A + D_i) \right) \right\}, \\ \frac{\partial F_A}{\partial A} &= \pi(\gamma) \sum_{i=1}^m \left\{ \frac{\gamma V_{i\gamma} s_i (y_i - x_i^T \beta)^4}{2(A + D_i)^4} - \frac{2 + 3\gamma}{1 + \gamma} \frac{V_{i\gamma} s_i (y_i - x_i^T \beta)^2}{(A + D_i)^3} + \frac{2 + 3\gamma}{2(1 + \gamma)^2} \frac{V_{i\gamma} s_i}{(A + D_i)^2} \right\}. \end{aligned}$$

利用引理4.1可知,  $E(\partial F_\beta / \partial \beta^T) = mJ_\beta$ ,  $E(\partial F_A / \partial \beta) = 0$  和  $E(\partial F_A / \partial A) = -mJ_A$ 。并且结合引理4.1, 有  $E(F_\beta F_\beta^t) = K_\beta$ ,  $E(F_\beta F_A) = 0$  和  $E(F_A^2) = K_A$ 。因此,  $\hat{\beta}_\alpha$  和  $\hat{A}_\alpha$  是渐近独立的, 并且其渐近方差分别为  $J_\beta^{-1} K_\beta J_\beta^{-1}$  和  $J_A^{-1} K_A J_A^{-1}$ 。

**证明 4 (定理4.3的证明)** 根据定义,  $MSE M_i = E \left\{ \left( \hat{\theta}_i^R - \theta_i \right)^2 \right\}$  可以分解为

$$M_i = E \left\{ \left( \tilde{\theta}_i^R - \theta_i \right)^2 \right\} + 2E \left\{ \left( \tilde{\theta}_i^R - \theta_i \right) \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right) \right\} + E \left\{ \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right)^2 \right\},$$

根据定理4.2可知, 上述分解中的第一项为  $g_{1i}(A) + g_{2i}(A)$ 。上式第二项的期望为 0, 因此只需估计上式的第三项, 这里采用 Taylor 展式对其进行展开可得:

$$\hat{\theta}_i^R - \tilde{\theta}_i^R = \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T} \left( \hat{\phi}_\gamma - \phi \right) + \frac{1}{2} \left( \hat{\phi}_\gamma - \phi \right)^T \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi_* \partial \phi_*^T} \left( \hat{\phi}_\gamma - \phi \right),$$

其中  $\phi_*$  与  $\phi$  和  $\hat{\phi}_\gamma$  相关。那么

$$E \left\{ \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right)^2 \right\} = E \left[ \left\{ \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T} \left( \hat{\phi}_\gamma - \phi \right) \right\}^2 \right] + R_1 + R_2,$$

其中  $R_1 = E \left\{ \left( \hat{\phi}_\gamma - \phi \right)^T \left( \partial \tilde{\theta}_i^R / \partial \phi \right) \left( \hat{\phi}_\gamma - \phi \right)^T \left( \partial^2 \tilde{\theta}_i^R / \partial \phi_* \partial \phi_*^T \right) \left( \hat{\phi}_\gamma - \phi \right)^T \right\}$  和  $R_2 = E \left[ \left\{ \left( \hat{\phi}_\gamma - \phi \right)^T \left( \partial^2 \tilde{\theta}_i^R / \partial \phi_* \partial \phi_*^T \right) \left( \hat{\phi}_\gamma - \phi \right)^T \right\}^2 \right] / 4$ 。为了推导上式, 这里引入如下的引理。

**引理 7.1** 在正则性条件下,  $E \left( \left| \hat{\phi}_{\gamma(k)} - \phi_k \right|^r \right) = O \left( m^{-r/2} \right)$  对任意  $r > 0$  和  $k = 1, \dots, p+1$  成立, 其中  $\hat{\phi}_{\gamma(k)}$  是  $\hat{\phi}_\gamma$  的第  $k$  个元素。

根据一致可积性和定理 4.2可知,  $E \left( m^r \left| \hat{\phi}_{\gamma(k)} - \phi_k \right|^r \right) = O(1)$  在正则性条件下成立, 因而上述引理成立。

通过直接计算可得

$$\begin{aligned} \frac{\partial \tilde{\theta}_i^R}{\partial \beta} &= -\frac{\pi_\gamma V_{i\gamma} D_i s_i x_i}{(A + D_i)^2} \left( \gamma (y_i - x_i^T \beta)^2 - (A + D_i) \right), \\ \frac{\partial \tilde{\theta}_i^R}{\partial A} &= -\frac{\pi_\gamma V_{i\gamma} D_i s_i (y_i - x_i^T \beta)}{2(A + D_i)^3} \left\{ \gamma (y_i - x_i^T \beta)^2 - \frac{3\gamma + 2}{1 + \gamma} (A + D_i) \right\}. \end{aligned}$$

另外有

$$\begin{aligned} \frac{\partial^2 \tilde{\theta}_i^R}{\partial \beta \partial \beta^T} &= -\frac{\gamma \pi_\gamma V_{i\gamma} D_i s_i x_i x_i^T}{(A + D_i)^3} \left( \gamma (y_i - x_i^T \beta)^3 - 3(A + D_i)(y_i - x_i^T \beta) \right), \\ \frac{\partial^2 \tilde{\theta}_i^R}{\partial A^2} &= -\frac{\pi_\gamma V_{i\gamma} D_i s_i (y_i - x_i^T \beta)}{2(A + D_i)^5} \left\{ \frac{\gamma^2}{2} (y_i - x_i^T \beta)^4 - \frac{5\gamma^2 + 4\gamma}{1 + \gamma} (y_i - x_i^T \beta)^2 (A + D_i) \right. \\ &\quad \left. + \frac{15\gamma^2 + 22\gamma + 8}{2(1 + \gamma)^2} (A + D_i)^2 \right\}. \end{aligned}$$

其中

$$\begin{aligned} R_1 &= \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} \sum_{\ell=1}^{p+1} E \left\{ \left( \frac{\partial \tilde{\theta}_i^R}{\partial \phi_j} \right) \left( \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi_k \partial \phi_\ell} \right) \left( \hat{\phi}_{\gamma(j)} - \phi_j \right) \left( \hat{\phi}_{\gamma(k)} - \phi_k \right) \left( \hat{\phi}_{\gamma(\ell)} - \phi_\ell \right) \right\} \\ &\equiv \sum_{j=1}^{p+1} \sum_{k=1}^{p+1} \sum_{\ell=1}^{p+1} U_{1jkl}. \end{aligned}$$

结合 Hölder 不等式

$$\begin{aligned} |U_{1jkl}| &\leq E \left\{ \left| \left( \frac{\partial \tilde{\theta}_i^R}{\partial \phi_j} \right) \left( \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi_k^* \partial \phi_\ell^*} \right) \right|^4 \right\}^{1/4} E \left\{ \left| \left( \hat{\phi}_{\gamma(j)} - \phi_j \right) \left( \hat{\phi}_{\gamma(k)} - \phi_k \right) \left( \hat{\phi}_{\gamma(\ell)} - \phi_\ell \right) \right|^{4/3} \right\}^{3/4} \\ &\leq E \left( \left| \frac{\partial \tilde{\theta}_i^R}{\partial \phi_j} \right|^8 \right)^{1/8} E \left( \left| \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi_k^* \partial \phi_\ell^*} \right|^8 \right)^{1/8} \prod_{a \in \{j, k, \ell\}} E \left( \left| \hat{\phi}_{\gamma(a)} - \phi_a \right|^4 \right)^{1/4}. \end{aligned}$$

因  $E \left( \left| \partial \tilde{\theta}_i^R / \partial \phi_j \right|^8 \right) < \infty$ , 根据引理 7.1 可知  $E \left( \left| \partial^2 \tilde{\theta}_i^R / \partial \phi_k^* \partial \phi_\ell^* \right|^8 \right) < \infty$ ,  $R_1 = o(m^{-1})$ . 同理可知  $R_2 = o(m^{-1})$ .

$$\begin{aligned} E \left[ \left\{ \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T} \left( \hat{\phi}_\gamma - \phi \right) \right\}^2 \right] &= \text{tr} \left[ E \left( \frac{\partial \tilde{\theta}_i^R}{\partial \phi} \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T} \right) E \left\{ \left( \hat{\phi}_\gamma - \phi \right) \left( \hat{\phi}_\gamma - \phi \right)^T \right\} \right] + o(m^{-1}) \\ &= \frac{1}{m} \text{tr} \left\{ E \left( \frac{\partial \tilde{\theta}_i^R}{\partial \beta} \frac{\partial \tilde{\theta}_i^R}{\partial \beta^T} \right) J_\beta^{-1} K_\beta J_\beta^{-1} \right\} + \frac{1}{m} E \left\{ \left( \frac{\partial \tilde{\theta}_i^R}{\partial A} \right)^2 \right\} J_A^{-1} K_A J_A^{-1} + o(m^{-1}). \end{aligned}$$

结合定理 4.2 有

$$\begin{aligned} E \left( \frac{\partial \tilde{\theta}_i^R}{\partial \beta} \frac{\partial \tilde{\theta}_i^R}{\partial \beta^T} \right) &= \frac{\pi_\gamma^2 V_i^{2\gamma} D_i^2 x_i x_i^T}{(A + D_i)^2} \frac{3\gamma^2 + 2\gamma + 1}{(2\gamma + 1)^{5/2}}, \\ E \left\{ \left( \frac{\partial \tilde{\theta}_i^R}{\partial A} \right)^2 \right\} &= \frac{\pi_\gamma^2 V_i^{2\gamma} D_i^2}{4(A + D_i)^3} \frac{15\gamma^4 + 36\gamma^3 + 34\gamma^2 + 16\gamma + 4}{(2\gamma + 1)^5 (1 + \gamma)^2}. \end{aligned}$$

得到  $E \left\{ \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right)^2 \right\} = m^{-1} g_{3i}(A) + m^{-1} g_{4i}(A) + o(m^{-1})$ . 因  $E \left\{ \left( \tilde{\theta}_i^R - \theta_i \right) \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right) \right\}$ ,

$$\begin{aligned} E \left\{ \left( \tilde{\theta}_i^R - \theta_i \right) \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right) \right\} &= E \left\{ \left( \tilde{\theta}_i^R - \tilde{\theta}_i \right) \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right) \right\} \\ &= \frac{D_i}{A + D_i} E \left\{ \left( 1 - \pi_\gamma V_{i\gamma} s_i \right) \left( y_i - x_i^T \beta \right) \left( \hat{\theta}_i^R - \tilde{\theta}_i^R \right) \right\}. \end{aligned}$$

通过 Taylor 展式

$$\hat{\theta}_i^R - \tilde{\theta}_i^R = \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T} \left( \hat{\phi}_\gamma - \phi \right) + \frac{1}{2} \left( \hat{\phi}_\alpha - \phi \right)^T \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi \partial \phi^T} \left( \hat{\phi}_\alpha - \phi \right) + R_3,$$

其中

$$R_3 = \frac{1}{6} \sum_{k=1}^{p+1} \sum_{j=1}^{p+1} \sum_{\ell=1}^{p+1} \frac{\partial^3 \tilde{\theta}_i^R}{\partial \phi_k^* \partial \phi_j^* \partial \phi_\ell^*} \left( \hat{\phi}_{\gamma(k)} - \phi_k \right) \left( \hat{\phi}_{\gamma(j)} - \phi_j \right) \left( \hat{\phi}_{\gamma(\ell)} - \phi_\ell \right),$$

类似与  $R_1$  的证明, 有  $E\{(1 - \pi_\gamma V_{i\gamma} s_i) u_i R_3\} = o(m^{-1})$ 。那么,

$$\begin{aligned} & E\left\{\left(\tilde{\theta}_i^R - \theta_i\right)\left(\hat{\theta}_i^R - \tilde{\theta}_i^R\right)\right\} \\ &= \frac{D_i}{A+D_i} E\left\{\left(1 - \pi_\gamma V_{i\gamma} s_i\right)\left(y_i - x_i^T \beta\right) \frac{\partial \tilde{\theta}_i^R}{\partial \phi^T}\left(\hat{\phi}_\gamma - \phi\right)\right\} + \\ & \quad \frac{D_i}{2(A+D_i)} E\left\{\left(1 - \pi_\gamma V_{i\gamma} s_i\right)\left(y_i - x_i^T \beta\right)\left(\hat{\phi}_\gamma - \phi\right)^T \frac{\partial^2 \tilde{\theta}_i^R}{\partial \phi \partial \phi^T}\left(\hat{\phi}_\gamma - \phi\right)\right\} + o\left(m^{-1}\right) \\ & \equiv T_1 + T_2 + o\left(m^{-1}\right), \end{aligned}$$

其中

$$\begin{aligned} T_2 &= \frac{D_i}{2m(A+D_i)} \text{tr}\left[E\left\{\left(1 - \pi_\gamma V_{i\gamma} s_i\right)\left(y_i - x_i^T \beta\right) \frac{\partial^2 \tilde{\theta}_i^R}{\partial \beta \partial \beta^T}\right\} J_\beta^{-1} K_\beta J_\beta^{-1}\right] \\ & \quad + \frac{D_i K_A}{2m(A+D_i) J_A^2} E\left\{\left(1 - \pi_\gamma V_{i\gamma} s_i\right)\left(y_i - x_i^T \beta\right) \frac{\partial^2 \tilde{\theta}_i^R}{\partial A^2}\right\} + o\left(m^{-1}\right). \end{aligned}$$

根据文献 *Lohr et al.*<sup>[157]</sup> 的结果, 有

$$\begin{aligned} E\left(\hat{\beta}_\gamma - \beta \mid y_i\right) &= b_\beta - m^{-1} \pi_\gamma V_{i\gamma} (A+D_i)^{-1} J_\beta^{-1} x_i s_i (y_i - x_i^T \beta) + o_p\left(m^{-1}\right), \\ E\left(\hat{A}_\gamma - A \mid y_i\right) &= b_A - m^{-1} J_A^{-1} \left\{\frac{\pi_\gamma V_{i\gamma} (y_i - x_i^T \beta)^2 s_i}{(A+D_i)^2} - \frac{V_{i\gamma} s_i}{1+\gamma} \frac{1}{A+D_i}\right\} + o_p\left(m^{-1}\right), \end{aligned}$$

其中  $b_\beta = \lim_{m \rightarrow \infty} m E\left(\hat{\beta}_\gamma - \beta\right)$  和  $b_A = \lim_{m \rightarrow \infty} m E\left(\hat{A}_\gamma - A\right)$ , 所以

$$\begin{aligned} T_1 &= -\frac{D_i}{m(A+D_i)^2} E\left\{\pi_\gamma V_{i\gamma} s_i (1 - \pi_\gamma V_{i\gamma} s_i) (y_i - x_i^T \beta)^2 \frac{\partial \tilde{\theta}_i^R}{\partial \beta^T} J_\beta^{-1} x_i\right\} \\ & \quad - \frac{D_i}{m(A+D_i)^3 J_A} E\left\{V_{i\gamma} s_i (1 - \pi_\gamma V_{i\gamma} s_i) (y_i - x_i^T \beta) \frac{\partial \tilde{\theta}_i^R}{\partial A} \left(\pi_\gamma (y_i - x_i^T \beta)^2 - \frac{(A+D_i)}{1+\gamma}\right)\right\} \\ & \quad + o\left(m^{-1}\right). \end{aligned}$$

结合引理 4.1 的内容,  $E\left\{\left(\tilde{\theta}_i^R - \theta_i\right)\left(\hat{\theta}_i^R - \tilde{\theta}_i^R\right)\right\} = m^{-1} g_{5i}(A) + o\left(m^{-1}\right)$ , 至此证明完成。

## B.2 5.1中的证明和推导

为了后续推导计算的方便, 首先给出5.1相关变量的导数的推导, 如下偏导数用于求解 MDPDE 估计方程

$$\begin{aligned} \frac{\partial |\mathbf{V}_i|}{\partial \beta} &= 0, \quad \frac{\partial |\mathbf{V}_i|}{\partial \sigma_v^2} = a_i \sigma_e^{2(n_i-1)} \prod_{j=1}^{n_i} k_{ij}^2, \\ \frac{\partial |\mathbf{V}_i|}{\partial \sigma_e^2} &= (n_i \sigma_e^{2(n_i-1)} + (n_i - 1) a_i \sigma_v^2 \sigma_e^{2(n_i-2)}) \prod_{j=1}^{n_i} k_{ij}^2, \end{aligned}$$

$$\begin{aligned}\frac{\partial B_i}{\partial \boldsymbol{\beta}} &= -\frac{2}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta}) \mathbf{x}_{ij}^T + \frac{2\sigma_v^2}{\sigma_e^2 (\sigma_v^2 + \frac{\sigma_e^2}{a_i})} \bar{\mathbf{x}}_i^T \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta}) \right), \\ \frac{\partial B_i}{\partial \sigma_v^2} &= -\frac{1}{a_i^2 (\sigma_v^2 + \frac{\sigma_e^2}{a_i})^2} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta}) \right)^2, \\ \frac{\partial B_i}{\partial \sigma_e^2} &= -\frac{1}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta})^2 + \frac{2\sigma_e^2 \sigma_v^2 + a_i \sigma_v^4}{(a_i \sigma_v^2 + \sigma_e^2)^2 \sigma_e^4} \left( \sum_{j=1}^{n_i} a_{ij} (y_{ij} - \mathbf{x}_{ij} \boldsymbol{\beta}) \right)^2,\end{aligned}$$

下面给出定理5.1的证明

**证明 5 (定理 5.1的证明)** 借助于 Ghosh et al.<sup>[135]</sup> 中定理 3.1 中关于极大似然估计一致性的证明, 令  $\hat{\boldsymbol{\theta}}$  是  $\boldsymbol{\theta}$  的 RMDPD 估计量, 那么

$$\frac{\partial}{\partial \boldsymbol{\theta}} \hat{d}_\gamma(f_\theta, g) = \frac{\partial}{\partial \boldsymbol{\theta}} \left[ \frac{1}{m} \sum_{i=1}^m \int_y f_\theta^{1+\gamma}(y | \mathbf{x}_i) dy - \frac{1+\gamma}{m\gamma} \sum_{i=1}^m f_\theta^\gamma(y_i | \mathbf{x}_i) \right] = 0.$$

进一步, 将其写成  $M$  估计量的估计方程形式, 即

$$\sum_{i=1}^m \Psi_{\hat{\boldsymbol{\theta}}}(y_i | \mathbf{x}_i) = 0. \quad (\text{B.2.1})$$

其中

$$\Psi_\theta(y_i | \mathbf{x}_i) = u_\theta(y_i | \mathbf{x}_i) f_\theta^\gamma(y_i | \mathbf{x}_i) - \int_y u_\theta(y | \mathbf{x}_i) f_\theta^{1+\gamma}(y | \mathbf{x}_i) dy,$$

令  $\boldsymbol{\theta}_g$  是  $\boldsymbol{\theta}$  的真实值, 根据  $E(\sum_{i=1}^m \Psi_{\hat{\boldsymbol{\theta}}}(y_i | \mathbf{x}_i)) = 0$  可知

$$\sum_{i=1}^m \left[ \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^\gamma(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{1+\gamma}(y | \mathbf{x}_i) dy \right] = 0. \quad (\text{B.2.2})$$

对式 (B.2.1) 进行泰勒展开可得

$$\begin{aligned}\frac{1}{m} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) + \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_\theta(y_i | \mathbf{x}_i) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_g} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_g) + R_m &= 0 \\ \text{或者 } \sqrt{m} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_g) &= - \left[ \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_\theta(y_i | \mathbf{x}_i) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_g} \right]^{-1} \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) + \sqrt{m} R_m \right].\end{aligned} \quad (\text{B.2.3})$$

其中  $R_N$  是余项. 利用若大数定律有

$$\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_\theta(y_i | \mathbf{x}_i)$$

$$\begin{aligned}
& \xrightarrow{p} \lim_{m \rightarrow \infty} \mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_{\boldsymbol{\theta}}(y_i | \mathbf{x}_i) \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{E} \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \left( u_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} - \int u_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} \right) \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{E} \left[ -I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} + \gamma u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{\gamma} - \int \{ -I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} + (1+\gamma) u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} \} \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \left[ - \int I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} g + \gamma \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{\gamma} g + \int I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} - (1+\gamma) \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} \right] \\
& \xrightarrow{p} - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \left[ \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} + \int (I_{\boldsymbol{\theta}} - \gamma u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T) (g - f_{\boldsymbol{\theta}}) f_{\boldsymbol{\theta}}^{\gamma} \right].
\end{aligned}$$

因此

$$\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_{\boldsymbol{\theta}}(y_i | \mathbf{x}_i) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_g} \xrightarrow{p} - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m J^{(i)} = -J. \quad (\text{B.2.4})$$

根据 (B.2.2), 有

$$\begin{aligned}
& \mathbb{E} \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \right] \\
& = \frac{1}{\sqrt{m}} \sum_{i=1}^m \left[ \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{\gamma}(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{1+\gamma}(y | \mathbf{x}_i) dy \right] \\
& = 0.
\end{aligned} \quad (\text{B.2.5})$$

那么,

$$\begin{aligned}
& V \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \right] \\
& = \frac{1}{m} \sum_{i=1}^m V [\Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i)] \\
& = \frac{1}{m} \sum_{i=1}^m \left[ \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) u_{\boldsymbol{\theta}_g}^T(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{2\gamma}(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \xi^{(i)} \xi^{(i)T} \right] \\
& = \frac{1}{m} \sum_{i=1}^m K^{(i)}.
\end{aligned} \quad (\text{B.2.6})$$

根据 Lehmann<sup>[158]</sup> 以及式 (B.2.5) 和 (B.2.6), 利用中心极限定理可得

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \overset{a}{\sim} N(0, K). \quad (\text{B.2.7})$$

在正则性条件下, 易知公式 (B.2.3) 中的余项满足  $\sqrt{m}R_m = o_p(1)$ . 故而, 结合公式 (B.2.3) 和 (B.2.7), 可得式 (B.2.3)

$$\sqrt{N} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_g) \overset{a}{\sim} N(0, J^{-1} K J^{-1}).$$

至此，完成证明。

### B.2.1 渐近分布协方差均值的计算

根据  $\xi^{(i)}$ ,  $\mathbf{u}_\theta$  的定义:

$$\xi^{(i)} = \int_{\mathbf{y}} \mathbf{u}_\theta(\mathbf{y} | \mathbf{x}_i) f_\theta^{\gamma+1}(\mathbf{y} | \mathbf{x}_i) d\mathbf{y},$$

有

$$\begin{aligned} \xi_\beta^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_\beta(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij}\beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= 0, \end{aligned}$$

结合式 (5.13), 有

$$\begin{aligned} \xi_{\sigma_v^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} \times M(1 + \gamma), \\ &\quad + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \times M(a_i \sigma_e^2 + a_i^2 \sigma_v^2), \\ &= -\frac{M a_i \gamma}{2(\sigma_e^2 + a_i \sigma_v^2)}, \\ \xi_{\sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 \right. \\ &\quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \times M(1 + \gamma), \\ &\quad + \frac{1}{2\sigma_e^4} \times M(n_i \sigma_e^2 + a_i \sigma_v^2), \\ &\quad - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M(a_i \sigma_e^2 + a_i^2 \sigma_v^2), \\ &= -\frac{M\gamma [n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2]}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)}. \end{aligned}$$

B.2.2 模型中矩阵  $\mathbf{J}^{(i)}$  的计算

根据  $\mathbf{J}^{(i)}$  的定义, 有

$$\begin{aligned}
\mathbf{J}_{\sigma_v^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_v^2}^2(\mathbf{y} | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ -\frac{a_{i\cdot}}{2(\sigma_e^2 + a_{i\cdot}\sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 \right]^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \frac{a_{i\cdot}^2}{4(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^2} \int_{\mathbf{y}_i} f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i - \frac{a_{i\cdot}}{2(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^3} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&\quad + \frac{1}{4(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^4} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^4 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \frac{a_{i\cdot}^2 M(1+\gamma)}{4(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^2}, \\
&\quad - \frac{a_{i\cdot}}{2(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^3} \times M(a_{i\cdot}\sigma_e^2 + a_{i\cdot}^2\sigma_v^2) \\
&\quad + \frac{1}{4(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^4} \times \frac{3M(a_{i\cdot}\sigma_e^2 + a_{i\cdot}^2\sigma_v^2)^2}{1+\gamma} \\
&= \frac{Ma_{i\cdot}^2(\gamma^2 + 2)}{4(\sigma_e^2 + a_{i\cdot}\sigma_v^2)^2(1+\gamma)}.
\end{aligned}$$

结合5.13,

$$\begin{aligned}
\mathbf{J}_{\beta}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\beta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\beta}^T(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij} \beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta) \right] \\
&\quad \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij} \beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta) \right]^T f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \int_{\mathbf{y}_i} a_{ij}^2 (y_{ij} - x_{ij} \beta)^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&\quad + \sum_{j \neq j'} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \int_{\mathbf{y}_i} a_{ij} a_{ij'} (y_{ij} - x_{ij} \beta) \\
&\quad (y_{ij'} - x_{ij'} \beta) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= M \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} (\sigma_e^2 + a_{ij} \sigma_v^2) \\
&\quad + M \sum_{j \neq j'} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} a_{ij'} \sigma_v^2, \\
&= M \sigma_e^2 \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} \\
&\quad + M \sigma_v^2 \sum_{j, j'=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} a_{ij'} \\
&= M \sigma_e^2 \sum_{j=1}^{n_i} \frac{1}{\sigma_e^4} a_{ij} x_{ij} x_{ij}^T + M \sigma_e^2 \left[ -\frac{2a_i^2 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \\
&\quad + M \sigma_v^2 \left[ \frac{a_i^2}{\sigma_e^4} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T - \frac{2a_i^3 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^4 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \\
&= M \left[ \frac{\sigma_e^2}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \frac{\sigma_v^2 a_i^2}{\sigma_e^4} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T + (\sigma_e^2 + a_i \sigma_v^2) \left\{ -\frac{2a_i^2 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right\} \right] \\
&= M \sigma_e^{-4} \left[ \sigma_e^2 \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \sigma_v^2 a_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T - 2\sigma_v^2 a_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{(\sigma_e^2 + a_i \sigma_v^2)} \right] \\
&= M \sigma_e^{-4} \left[ \sigma_e^2 \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \sigma_v^2 a_i^2 \left( \frac{a_i \sigma_v^2}{(\sigma_e^2 + a_i \sigma_v^2)} - 1 \right) \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \right].
\end{aligned}$$

$$\begin{aligned}
\mathbf{J}_{\sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} u_{\sigma_e^2}^2(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \beta)^2 \right]
\end{aligned}$$

$$\begin{aligned}
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\boldsymbol{\beta}) \right\}^2 \Big] f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& = \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \int_{y_i} f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& + \frac{1}{4\sigma_e^8} \int_{y_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& + \frac{\sigma_v^4 (2\sigma_e^2 + a_i \sigma_v^2)^2}{4\sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^4} \int_{y_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\boldsymbol{\beta}) \right\}^4 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} \int_{y_i} a_{ij} (y_{it} - x_{it}\beta)^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& + \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2) \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)^3} \int_{y_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\boldsymbol{\beta}) \right\}^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \sum_{j=1}^{n_i} \int_{y_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\boldsymbol{\beta}) \right\}^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& = \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M(1 + \gamma) \\
& + \frac{1}{4\sigma_e^8} \frac{M \{ (n_i^2 + 2n_i) \sigma_e^4 + 2(n_i + 2) a_i \sigma_e^2 \sigma_v^2 + 3a_i^2 \sigma_v^4 \}}{1 + \gamma} \\
& + \frac{\sigma_v^4 (2\sigma_e^2 + a_i \sigma_v^2)^2}{4\sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^4} \times \frac{3M (a_i \sigma_e^2 + a_i^2 \sigma_v^2)^2}{1 + \gamma} \\
& - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)} \times \sum_{j=1}^{n_i} M (\sigma_e^2 + a_{ij} \sigma_v^2) \\
& + \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2) \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)^3} \times M (a_i \sigma_e^2 + a_i^2 \sigma_v^2) \\
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M \left\{ \frac{(n_i a_i + 2a_i) \sigma_e^4 + (n_i a_i^2 + 2a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (a_i^2 - a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
& = \frac{M(\gamma - 1) (n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \\
& + \frac{M \{ (n_i^2 + 2n_i) \sigma_e^4 + 2(n_i + 2) a_i \sigma_e^2 \sigma_v^2 + 3a_i^2 \sigma_v^4 \}}{4\sigma_e^8 (1 + \gamma)} \\
& + \frac{3M a_i^2 \sigma_v^4 (2\sigma_e^2 + a_i \sigma_v^2)^2}{4(1 + \gamma) \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2}
\end{aligned}$$

$$-\frac{\sigma_v^2(2\sigma_e^2 + a_i\sigma_v^2)}{2\sigma_e^8(\sigma_e^2 + a_i\sigma_v^2)^2} \times M \left\{ \frac{(n_i a_{i\cdot} + 2a_{i\cdot})\sigma_e^4 + (n_i a_{i\cdot}^2 + 2a_{i\cdot}^2 + 3\sum_{j=1}^{n_i} a_{ij}^2)\sigma_e^2\sigma_v^2}{1 + \gamma} + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij}(a_{i\cdot}^2 - a_{ij}a_{i\cdot} + a_{ij}^2)}{1 + \gamma} \right\}.$$

根据5.13

$$\begin{aligned} \mathbf{J}_{\beta, \sigma_v^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_\beta(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij}\beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right] \\ &\quad \times \left[ -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= 0. \end{aligned}$$

同样地,

$$\begin{aligned} \mathbf{J}_{\beta, \sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_\beta(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij}\beta) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right] \\ &\quad \times \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 \right. \\ &\quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\beta) \right\}^2 \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= 0. \end{aligned}$$

$$\begin{aligned} \mathbf{J}_{\sigma_v^2, \sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 \right] \\ &\quad \times \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 \right. \\ &\quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it}\beta) \right\}^2 \right] f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \int_{\mathbf{y}_i} f_\theta^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \sum_{j=1}^{n_i} \int_{\mathbf{y}_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^4} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^4 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& - \frac{a_i}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} \int_{\mathbf{y}_i} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& + \left[ \frac{a_i \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^3} - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^3} \right] \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& = \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M(1 + \gamma) \\
& + \frac{1}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M \left\{ \frac{(n_i a_i + 2a_i) \sigma_e^4 + (n_i a_i^2 + 2a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (a_i^2 - a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^4} \times \frac{3M (a_i \sigma_e^2 + a_i^2 \sigma_v^2)^2}{1 + \gamma} \\
& - \frac{a_i}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} \times M (n_i \sigma_e^2 + a_i \sigma_v^2) \\
& + \left[ \frac{a_i \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^3} - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^3} \right] \times M (a_i \sigma_e^2 + a_i^2 \sigma_v^2) \\
& = M(1 + \gamma) \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \\
& + \frac{M}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \frac{(n_i a_i + 2a_i) \sigma_e^4 + (n_i a_i^2 - 4a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (-a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
& - \frac{M a_i (n_i \sigma_e^2 - (n_i - 1) a_i \sigma_v^2)}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2}.
\end{aligned}$$

### B.2.3 矩阵 $\mathbf{J}^{(i)}$ 的计算

$$\begin{aligned}
& \int_{\mathbf{y}_i} a_{ij} a_{ij'} (y_{ij} - x_{ij}\beta) (y_{ij'} - x_{ij'}\beta) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& = \int_{z_i} a_{ij} a_{ij'} z_{ij} z_{ij'} f_{\theta}^{1+\gamma}(z_i | 0) dz_i, \text{ where } f_{\theta}(z_i | 0) \text{ is } N_{n_i}(0, \mathbf{V}_i) \\
& = (2\pi)^{-\frac{n_i \gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \int_{z_i} a_{ij} a_{ij'} z_{ij} z_{ij'} (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1 + \gamma}{2} z_i' \mathbf{V}_i^{-1} z_i \right\} dz_i
\end{aligned}$$

$$\begin{aligned}
&= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i}{2}} \int_{z_i} a_{ij} a_{ij'} z_{ij} z_{ij'} (2\pi)^{-\frac{n_i}{2}} \left| \frac{\mathbf{V}_i}{1+\gamma} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} z_i' \left( \frac{\mathbf{V}_i}{1+\gamma} \right)^{-1} z_i \right\} dz_i \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i+2}{2}} a_{ij} a_{ij'} \mathbf{V}_{i-jj'} \\
&= M a_{ij} a_{ij'} \mathbf{V}_{i-jj'}. \tag{B.2.8}
\end{aligned}$$

其中

$$\begin{aligned}
M &= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i+2}{2}} \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} (1+\gamma)^{-\frac{n_i+2}{2}} \left( \sigma_e^{2(n_i-1)} (\sigma_e^2 + a_i \sigma_v^2) \prod_{j=1}^{n_i} k_{ij}^2 \right)^{-\frac{\gamma}{2}} \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} (1+\gamma)^{-\frac{n_i+2}{2}} \sigma_e^{2(-\frac{\gamma(n_i-1)}{2})} (\sigma_e^2 + a_i \sigma_v^2)^{-\frac{\gamma}{2}} \prod_{j=1}^{n_i} k_{ij}^{-\gamma}. \tag{B.2.9}
\end{aligned}$$

当  $j = j'$ , 式 (B.2.8) 可以表示为

$$\int_{\mathbf{y}_i} a_{ij}^2 (y_{ij} - x_{ij}\boldsymbol{\beta})^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i = M a_{ij} (\sigma_e^2 + a_{ij}\sigma_v^2). \tag{B.2.10}$$

对于  $j \neq j'$ , 结合 (B.2.8), 有

$$\int_{\mathbf{y}_i} a_{ij} a_{ij'} (y_{ij} - x_{ij}\boldsymbol{\beta}) (y_{ij'} - x_{ij'}\boldsymbol{\beta}) f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i = M a_{ij} a_{ij'} \sigma_v^2. \tag{B.2.11}$$

对任意的两个整数  $r$  和  $s$ , 其中  $(r+s)$  是奇数时有

$$\begin{aligned}
&\int_{\mathbf{y}_i} a_{ij}^r a_{ij'}^s (y_{ij} - x_{ij}\boldsymbol{\beta})^r (y_{ij'} - x_{ij'}\boldsymbol{\beta})^s f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{z_i} a_{ij}^r a_{ij'}^s z_{ij}^r z_{ij'}^s f_{\boldsymbol{\theta}}^{1+\gamma}(z_i | 0) dz_i, \text{ where } f_{\boldsymbol{\theta}}(z_i | 0) \text{ is } N_T(0, \mathbf{V}_i) \\
&= 0 \tag{B.2.12}
\end{aligned}$$

另

$$\begin{aligned}
\int_{\mathbf{y}_i} f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i &= \int_{z_i} f_{\boldsymbol{\theta}}^{1+\gamma}(z_i | 0) dz_i, \text{ where } f_{\boldsymbol{\theta}}(z_i | 0) \text{ is } N_T(0, \mathbf{V}_i) \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \int_{z_i} (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1+\gamma}{2} z_i' \mathbf{V}_i^{-1} z_i \right\} dz_i \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i}{2}} \int_{z_i} (2\pi)^{-\frac{n_i}{2}} \left| \frac{\mathbf{V}_i}{1+\gamma} \right|^{-\frac{1}{2}} \\
&\quad \exp \left\{ -\frac{1}{2} z_i' \left( \frac{\mathbf{V}_i}{1+\gamma} \right)^{-1} z_i \right\} dz_i \\
&= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i}{2}}
\end{aligned}$$

$$= M(1 + \gamma), \quad (\text{B.2.13})$$

所以

$$\begin{aligned} & \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{z_i} \left\{ \sum_{j=1}^{n_i} a_{ij} z_{ij} \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(z_i | 0) dz_i, \text{ where } f_{\boldsymbol{\theta}}(z_i | 0) \text{ is } N_{n_i}(0, \mathbf{V}_i) \\ &= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} \int_{z_i} \left\{ \sum_{j=1}^{n_i} a_{ij} z_{ij} \right\}^2 (2\pi)^{-\frac{n_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1+\gamma}{2} z_i' \mathbf{V}_i^{-1} z_i \right\} dz_i \\ &= (2\pi)^{-\frac{n_i\gamma}{2}} |\mathbf{V}_i|^{-\frac{\gamma}{2}} (1+\gamma)^{-\frac{n_i}{2}} \int_{z_i} \left\{ \sum_{j=1}^{n_i} a_{ij} z_{ij} \right\}^2 (2\pi)^{-\frac{n_i}{2}} \left| \frac{\mathbf{V}_i}{1+\gamma} \right|^{-\frac{1}{2}} \\ & \quad \exp \left\{ -\frac{1}{2} z_i' \left( \frac{\mathbf{V}_i}{1+\gamma} \right)^{-1} z_i \right\} dz_i \\ &= M(1+\gamma) \text{E} \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^2 \right) \\ &= M(a_i \sigma_e^2 + a_i^2 \sigma_v^2). \end{aligned} \quad (\text{B.2.14})$$

同样地,

$$\begin{aligned} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^4 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i &= M(1+\gamma) \text{E} \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^4 \right) \\ &= \frac{3M(a_i \sigma_e^2 + a_i^2 \sigma_v^2)^2}{1+\gamma}. \end{aligned} \quad (\text{B.2.15})$$

$$\begin{aligned} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\boldsymbol{\beta})^2 \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i &= \int_{z_i} \left\{ \sum_{j=1}^{n_i} a_{ij} z_{ij}^2 \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(z_i | 0) dz_i \\ &= M(1+\gamma) \text{E} \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij}^2 \right]^2 \right) \\ &= \frac{M \{ (n_i^2 + 2n_i) \sigma_e^4 + 2(n_i + 2) a_i \sigma_e^2 \sigma_v^2 + 3a_i^2 \sigma_v^4 \}}{1+\gamma}, \end{aligned} \quad (\text{B.2.16})$$

$$\int_{\mathbf{y}_i} a_{ij}' (y_{ij}' - x_{ij}'\boldsymbol{\beta})^2 \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i$$

$$\begin{aligned}
&= M(1 + \gamma) E \left( a_{ij'} s_{ij'}^2 \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^2 \right) \\
&= \frac{M \{ (a_i + 2a_{ij'}) \sigma_e^4 + (a_i^2 + 2a_{ij'} a_i + 4a_{ij}^2 - a_{ij'}^2) \sigma_e^2 \sigma_v^2 + 3a_{ij'} (a_i^2 - a_{ij'} a_i + a_{ij'}^2) \sigma_v^4 \}}{1 + \gamma}.
\end{aligned} \tag{B.2.17}$$

### B.2.4 相关积分的期望

假设  $s_i \sim N_{n_i} \left( 0, \frac{\mathbf{V}_i}{1+\gamma} \right)$ , 那么

$$\begin{aligned}
V \left( \sum_{j=1}^{n_i} a_{ij} s_{ij} \right) &= \sum_{j=1}^{n_i} V(a_{ij} s_{ij}) + \sum_{j \neq j'} \text{COV}(a_{ij} a_{ij'} s_{ij} s_{ij'}) \\
&= \frac{1}{1 + \gamma} \sum_{j=1}^{n_i} a_{ij}^2 \mathbf{V}_{i-jj} + \frac{1}{1 + \gamma} \sum_{j \neq j'} a_{ij} a_{ij'} \mathbf{V}_{i-jj'} \\
&= \frac{1}{1 + \gamma} \sum_{j=1}^{n_i} a_{ij}^2 (k_{ij}^2 \sigma_e^2 + \sigma_v^2) + \frac{1}{1 + \gamma} \sum_{j \neq j'} a_{ij} a_{ij'} \sigma_v^2 \\
&= \frac{1}{1 + \gamma} \left( \sigma_e^2 \sum_{j=1}^{n_i} a_{ij} + \sigma_v^2 \sum_{j=1}^{n_i} a_{ij}^2 \right) + \frac{1}{1 + \gamma} \sigma_v^2 \sum_{j \neq j'} a_{ij} a_{ij'} \\
&= \frac{1}{1 + \gamma} (a_i \cdot \sigma_e^2 + a_i^2 \cdot \sigma_v^2),
\end{aligned}$$

其中  $\mathbf{V}_{i-jj'}$  是  $\mathbf{V}_i$  的第  $j$  行  $j'$  列的元素。

因此,  $\sum_{j=1}^{n_i} a_{ij} s_{ij} \sim N \left( 0, \frac{1}{1+\gamma} (a_i \cdot \sigma_e^2 + a_i^T a_i \sigma_v^2) \right)$ . 且

$$E \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^2 \right) = \frac{1}{1 + \gamma} (a_i \cdot \sigma_e^2 + a_i^2 \cdot \sigma_v^2), \tag{B.2.18}$$

$$E \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^4 \right) = \frac{3}{(1 + \gamma)^2} (a_i \cdot \sigma_e^2 + a_i^2 \cdot \sigma_v^2)^2. \tag{B.2.19}$$

和

$$\begin{aligned}
E \left( \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij}^2 \right]^2 \right) &= E \left( \sum_{j=1}^{n_i} a_{ij}^2 s_{ij}^4 + \sum_{j \neq j'} a_{ij} a_{ij'} s_{ij}^2 s_{ij'}^2 \right) \\
&= \sum_{j=1}^{n_i} \frac{3a_{ij}^2 \mathbf{V}_{i-jj}^2}{(1 + \gamma)^2} + \sum_{j \neq j'} \frac{a_{ij} a_{ij'} (\mathbf{V}_{i-jj} \mathbf{V}_{i-j'j'} + 2\mathbf{V}_{i-jj'})}{(1 + \gamma)^2} \\
&= \frac{3 \sum_{j=1}^{n_i} a_{ij}^2 (k_{ij}^2 \sigma_e^2 + \sigma_v^2)^2}{(1 + \gamma)^2} + \frac{\sum_{j \neq j'} a_{ij} a_{ij'} (k_{ij}^2 \sigma_e^2 + \sigma_v^2) (k_{ij'}^2 \sigma_e^2 + \sigma_v^2) + 2 \sum_{j \neq j'} a_{ij} a_{ij'} \sigma_v^4}{(1 + \gamma)^2} \\
&= \frac{2 \sum_{j=1}^{n_i} (\sigma_e^2 + a_{ij} \sigma_v^2)^2 + \left( \sum_{j=1}^{n_i} (\sigma_e^2 + a_{ij} \sigma_v^2) \right)^2 + 2 \sum_{j \neq j'} a_{ij} a_{ij'} \sigma_v^4}{(1 + \gamma)^2}
\end{aligned}$$

$$\begin{aligned}
&= \frac{2n_i\sigma_e^4 + 4\sigma_e^2\sigma_v^2 \sum_{j=1}^{n_i} a_{ij} + 2\sigma_v^4 \sum_{j=1}^{n_i} a_{ij}^2 + n_i^2\sigma_e^4 + 2n_i a_i \sigma_e^2\sigma_v^2 + a_i^2\sigma_v^4 + 2\sigma_v^4 \sum_{j \neq j'} a_{ij} a_{ij'}}{(1+\gamma)^2} \\
&= \frac{(n_i^2 + 2n_i)\sigma_e^4 + 2(n_i + 2)a_i \sigma_e^2\sigma_v^2 + 3a_i^2\sigma_v^4}{(1+\gamma)^2}. \tag{B.2.20}
\end{aligned}$$

对  $j' = 1, 2, \dots, n_i$  有

$$\begin{aligned}
&E \left( a_{ij'} s_{ij'}^2 \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^2 \right) \\
&= E \left( a_{ij'}^3 s_{ij'}^4 + a_{ij'} s_{ij'}^2 \sum_{j \neq j'} a_{ij}^2 s_{ij}^2 + a_{ij'}^2 s_{ij'}^3 \sum_{j \neq j'} a_{ij} s_{ij} + a_{ij'} s_{ij'}^2 \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} s_{ij} s_{ij''} \right) \\
&= \left\{ 3a_{ij'}^3 \mathbf{V}_{i-j'}^2 + \left( a_{ij'} \mathbf{V}_{i-j'} \sum_{j \neq j'} a_{ij}^2 \mathbf{V}_{i-jj} + 2a_{ij'} \sum_{j \neq j'} a_{ij}^2 \mathbf{V}_{i-jj'} \right) + 3a_{ij'}^2 \mathbf{V}_{i-j'} \sum_{j \neq j'} a_{ij} \mathbf{V}_{i-jj} + \right. \\
&\quad \left. \left( a_{ij'} \mathbf{V}_{i-j'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \mathbf{V}_{i-jj''} + 2a_{ij'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \mathbf{V}_{i-jj'} \mathbf{V}_{i-j'j''} \right) \right\} \frac{1}{(1+\gamma)^2} \\
&= \left\{ 2a_{ij'}^3 (k_{ij'}^2 \sigma_e^2 + \sigma_v^2)^2 + a_{ij'} (k_{ij'}^2 \sigma_e^2 + \sigma_v^2) \sum_{j=1}^{n_i} a_{ij}^2 (k_{ij}^2 \sigma_e^2 + \sigma_v^2) + 2a_{ij'} \sum_{j \neq j'} a_{ij}^2 \sigma_v^4 \right. \\
&\quad + 3a_{ij'}^2 (k_{ij'}^2 \sigma_e^2 + \sigma_v^2) \sum_{j \neq j'} a_{ij} \sigma_v^2 + a_{ij'} (k_{ij'}^2 \sigma_e^2 + \sigma_v^2) \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \sigma_v^2 \\
&\quad \left. + 2a_{ij'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \sigma_v^4 \right\} \frac{1}{(1+\gamma)^2} \\
&= \left\{ 2a_{ij'} (\sigma_e^2 + a_{ij'} \sigma_v^2)^2 + (\sigma_e^2 + a_{ij'} \sigma_v^2) \left( a_i \sigma_e^2 + \sigma_v^2 \sum_{j=1}^{n_i} a_{ij}^2 \right) + 2a_{ij'} \sum_{j \neq j'} a_{ij}^2 \sigma_v^4 \right. \\
&\quad + 3a_{ij'} (\sigma_e^2 + a_{ij'} \sigma_v^2) \sum_{j \neq j'} a_{ij} \sigma_v^2 + (\sigma_e^2 + a_{ij'} \sigma_v^2) \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \sigma_v^2 \\
&\quad \left. + 2a_{ij'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \sigma_v^4 \right\} \frac{1}{(1+\gamma)^2} \\
&= \left\{ (a_i + 2a_{ij'}) \sigma_e^4 + \left( 4a_{ij'}^2 + \sum_{j=1}^{n_i} a_{ij}^2 + a_{ij'} a_i + 3a_{ij'} \sum_{j \neq j'} a_{ij} + \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \right) \sigma_e^2 \sigma_v^2 \right. \\
&\quad \left. + \left( 2a_{ij'}^3 + 2a_{ij'} \sum_{j \neq j'} a_{ij}^2 + a_{ij'} \sum_{j=1}^{n_i} a_{ij}^2 + 3a_{ij'}^2 \sum_{j \neq j'} a_{ij} + 3a_{ij'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} \right) \sigma_v^4 \right\} \frac{1}{(1+\gamma)^2} \\
&= \left\{ (a_i + 2a_{ij'}) \sigma_e^4 + (a_i^2 + 2a_{ij'} a_i + 4a_{ij'}^2 - a_{ij'}^2) \sigma_e^2 \sigma_v^2 + 3a_{ij'} (a_i^2 - a_{ij'} a_i + a_{ij'}^2) \sigma_v^4 \right\} \frac{1}{(1+\gamma)^2}. \tag{B.2.21}
\end{aligned}$$

在上述计算中，用到了公式

$$\sum_{j \neq j' \neq j''} a_{ij} a_{ij''} = \sum_{j, j''} a_{ij} a_{ij''} - \sum_{j=j'' \neq j'} a_{ij}^2 - \sum_{j=j' \neq j''} a_{ij'} a_{ij''} - \sum_{j''=j' \neq j} a_{ij'} a_{ij} - a_{ij'}^2$$

$$\begin{aligned}
&= a_i^2 - \sum_{j \neq j'} a_{ij}^2 - 2 \sum_{j \neq j'} a_{ij'} a_{ij} - a_{ij'}^2 \\
&= a_i^2 - \sum_{j=1}^{n_i} a_{ij}^2 - 2a_{ij'} \sum_{j \neq j'} a_{ij}.
\end{aligned}$$

同样地,

$$\mathbb{E} \left( a_{ij'} s_{ij'} \left[ \sum_{j=1}^{n_i} a_{ij} s_{ij} \right]^2 \right) = \mathbb{E} \left( a_{ij'}^3 s_{ij'}^3 + a_{ij'}^2 s_{ij'}^2 \sum_{j \neq j'} a_{ij} s_{ij} + a_{ij'} s_{ij'} \sum_{j \neq j' \neq j''} a_{ij} a_{ij''} s_{ij} s_{ij''} \right) = 0. \quad (\text{B.2.22})$$

### C.3 6.1中的证明和推导

下面我们给出定理6.1的证明

**证明 6** 这里的部分证明可参见<sup>[135]</sup>中定理 3.1 的关于极大似然估计的一致性的证明。令  $\hat{\theta}$  是  $\theta$  的 RMG 估计量, 那么

$$\frac{\partial}{\partial \theta} \hat{d}_\gamma(f_\theta, g) = \frac{\partial}{\partial \theta} \left[ \frac{1}{m} \sum_{i=1}^m \int_y f_\theta^{1+\gamma}(y | \mathbf{x}_i) dy - \frac{1+\gamma}{m\gamma} \sum_{i=1}^m f_\theta^\gamma(y_i | \mathbf{x}_i) \right] = 0,$$

因此, 可将其写成  $M$  估计量的估计方程, 即

$$\sum_{i=1}^m \Psi_{\hat{\theta}}(y_i | \mathbf{x}_i) = 0. \quad (\text{C.3.23})$$

其中

$$\Psi_\theta(y_i | \mathbf{x}_i) = u_\theta(y_i | \mathbf{x}_i) f_\theta^\gamma(y_i | \mathbf{x}_i) - \int_y u_\theta(y | \mathbf{x}_i) f_\theta^{1+\gamma}(y | \mathbf{x}_i) dy,$$

令  $\theta_g$  是  $\theta$  的真实值, 由  $\mathbb{E}(\sum_{i=1}^m \Psi_{\hat{\theta}}(y_i | \mathbf{x}_i)) = 0$  可知

$$\sum_{i=1}^m \left[ \int_y u_{\theta_g}(y | \mathbf{x}_i) f_{\theta_g}^\gamma(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \int_y u_{\theta_g}(y | \mathbf{x}_i) f_{\theta_g}^{1+\gamma}(y | \mathbf{x}_i) dy \right] = 0. \quad (\text{C.3.24})$$

对公式 (C.3.23) 进行泰勒展开, 可得

$$\begin{aligned}
&\frac{1}{m} \sum_{i=1}^m \Psi_{\theta_g}(y_i | \mathbf{x}_i) + \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \theta} \Psi_\theta(y_i | \mathbf{x}_i) \Big|_{\theta=\theta_g} (\hat{\theta} - \theta_g) + R_m = 0 \\
&\text{or } \sqrt{m} (\hat{\theta} - \theta_g) = - \left[ \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \theta} \Psi_\theta(y_i | \mathbf{x}_i) \Big|_{\theta=\theta_g} \right]^{-1} \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\theta_g}(y_i | \mathbf{x}_i) + \sqrt{m} R_m \right].
\end{aligned} \quad (\text{C.3.25})$$

其中  $R_N$  是余项。利用弱大数定律有

$$\begin{aligned}
& \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_{\boldsymbol{\theta}}(y_i | \mathbf{x}_i) \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_{\boldsymbol{\theta}}(y_i | \mathbf{x}_i) \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{E} \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \left( u_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} - \int u_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} \right) \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{E} \left[ -I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} + \gamma u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{\gamma} - \int \{ -I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} + (1+\gamma) u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} \} \right] \\
& \xrightarrow{p} \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \left[ - \int I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{\gamma} g + \gamma \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{\gamma} g + \int I_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\gamma} - (1+\gamma) \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} \right] \\
& \xrightarrow{p} - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \left[ \int u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T f_{\boldsymbol{\theta}}^{1+\gamma} + \int (I_{\boldsymbol{\theta}} - \gamma u_{\boldsymbol{\theta}} u_{\boldsymbol{\theta}}^T) (g - f_{\boldsymbol{\theta}}) f_{\boldsymbol{\theta}}^{\gamma} \right].
\end{aligned}$$

因此

$$\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial \boldsymbol{\theta}} \Psi_{\boldsymbol{\theta}}(y_i | \mathbf{x}_i) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_g} \xrightarrow{p} - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m J^{(i)} = -J. \quad (\text{C.3.26})$$

结合公式 (C.3.24) 可得

$$\begin{aligned}
& \mathbb{E} \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \right] \\
& = \frac{1}{\sqrt{m}} \sum_{i=1}^m \left[ \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{\gamma}(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{1+\gamma}(y | \mathbf{x}_i) dy \right] \\
& = 0. \quad (\text{C.3.27})
\end{aligned}$$

那么

$$\begin{aligned}
& V \left[ \frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \right] \\
& = \frac{1}{m} \sum_{i=1}^m V [\Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i)] \\
& = \frac{1}{m} \sum_{i=1}^m \left[ \int_y u_{\boldsymbol{\theta}_g}(y | \mathbf{x}_i) u_{\boldsymbol{\theta}_g}^T(y | \mathbf{x}_i) f_{\boldsymbol{\theta}_g}^{2\gamma}(y | \mathbf{x}_i) g(y | \mathbf{x}_i) dy - \xi^{(i)} \xi^{(i)T} \right] \\
& = \frac{1}{m} \sum_{i=1}^m K^{(i)}. \quad (\text{C.3.28})
\end{aligned}$$

根据 Lehmann<sup>[158]</sup> 的结论, 结合式 (C.3.27) 和 (C.3.28), 利用中心极限定理, 我们可得

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m \Psi_{\boldsymbol{\theta}_g}(y_i | \mathbf{x}_i) \stackrel{a}{\sim} N(0, K). \quad (\text{C.3.29})$$

在正则性条件下, 易知公式 (C.3.25) 中的余项  $\sqrt{m}R_m = o_p(1)$ 。因此结合公式 (C.3.25) 和 (C.3.29), 可得

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_g) \stackrel{a}{\sim} N(0, J^{-1}KJ^{-1}).$$

至此, 定理证毕。

### C.3.1 矩阵 $\mathbf{J}^{(i)}$ 的计算

根据  $\mathbf{J}^{(i)}$  的定义,

$$\begin{aligned} \mathbf{J}_{\sigma_v^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_v^2}^2(\mathbf{y} | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ -\frac{a_i}{2(\sigma_e^2 + a_i\sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i\sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 \right]^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \frac{a_i^2}{4(\sigma_e^2 + a_i\sigma_v^2)^2} \int_{\mathbf{y}_i} f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i - \frac{a_i}{2(\sigma_e^2 + a_i\sigma_v^2)^3} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^2 \\ &\quad f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i + \frac{1}{4(\sigma_e^2 + a_i\sigma_v^2)^4} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right\}^4 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \frac{a_i^2 M(1+\gamma)}{4(\sigma_e^2 + a_i\sigma_v^2)^2} \\ &\quad - \frac{a_i}{2(\sigma_e^2 + a_i\sigma_v^2)^3} \times M(a_i\sigma_e^2 + a_i^2\sigma_v^2) \\ &\quad + \frac{1}{4(\sigma_e^2 + a_i\sigma_v^2)^4} \times \frac{3M(a_i\sigma_e^2 + a_i^2\sigma_v^2)^2}{1+\gamma} \\ &= \frac{Ma_i^2(\gamma^2 + 2)}{4(\sigma_e^2 + a_i\sigma_v^2)^2(1+\gamma)}. \end{aligned}$$

结合文中的定义有

$$\begin{aligned} \mathbf{J}_{\boldsymbol{\beta}}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\boldsymbol{\beta}}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\boldsymbol{\beta}}^T(\mathbf{y}_i | \mathbf{x}_i) f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij}x_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) - \frac{a_i\sigma_v^2\bar{\mathbf{x}}_i}{\sigma_e^2(\sigma_e^2 + a_i\sigma_v^2)} \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right] \\ &\quad \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij}x_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) - \frac{a_i\sigma_v^2\bar{\mathbf{x}}_i}{\sigma_e^2(\sigma_e^2 + a_i\sigma_v^2)} \sum_{j=1}^{n_i} a_{ij}(y_{ij} - x_{ij}\boldsymbol{\beta}) \right]^T f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\ &= \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij}x_{ij}^T - \frac{2a_i\sigma_v^2x_{ij}\bar{\mathbf{x}}_i^T}{\sigma_e^4(\sigma_e^2 + a_i\sigma_v^2)} + \frac{a_i^2\sigma_v^4\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T}{\sigma_e^4(\sigma_e^2 + a_i\sigma_v^2)^2} \right] \int_{\mathbf{y}_i} a_{ij}^2(y_{ij} - x_{ij}\boldsymbol{\beta})^2 f_{\boldsymbol{\theta}}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j \neq j'} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \int_{\mathbf{y}_i} a_{ij} a_{ij'} (y_{ij} - x_{ij} \boldsymbol{\beta}) \\
 & (y_{ij'} - x_{ij'} \boldsymbol{\beta}) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
 = & M \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} (\sigma_e^2 + a_i \sigma_v^2), \\
 & + M \sum_{j \neq j'} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} a_{ij'} \sigma_v^2, \\
 = & M \sigma_e^2 \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} \\
 & + M \sigma_v^2 \sum_{j, j'=1}^{n_i} \left[ \frac{1}{\sigma_e^4} x_{ij} x_{ij'}^T - \frac{2a_i \sigma_v^2 x_{ij} \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^2 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] a_{ij} a_{ij'} \\
 = & M \sigma_e^2 \sum_{j=1}^{n_i} \frac{1}{\sigma_e^4} a_{ij} x_{ij} x_{ij}^T + M \sigma_e^2 \left[ -\frac{2a_i^2 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \\
 & + M \sigma_v^2 \left[ \frac{a_i^2}{\sigma_e^4} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T - \frac{2a_i^3 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^4 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right] \\
 = & M \left[ \frac{\sigma_e^2}{\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \frac{\sigma_v^2 a_i^2}{\sigma_e^4} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T + (\sigma_e^2 + a_i \sigma_v^2) \left\{ -\frac{2a_i^2 \sigma_v^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \right\} \right] \\
 = & M \sigma_e^{-4} \left[ \sigma_e^2 \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \sigma_v^2 a_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T - 2\sigma_v^2 a_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T + \frac{a_i^3 \sigma_v^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T}{(\sigma_e^2 + a_i \sigma_v^2)} \right] \\
 = & M \sigma_e^{-4} \left[ \sigma_e^2 \sum_{j=1}^{n_i} a_{ij} x_{ij} x_{ij}^T + \sigma_v^2 a_i^2 \left( \frac{a_i \sigma_v^2}{(\sigma_e^2 + a_i \sigma_v^2)} - 1 \right) \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \right].
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{J}_{\sigma_e^2}^{(i)} & = \int_{\mathbf{y}_i} u_{\sigma_e^2}^2(y | \mathbf{x}_i) f_{\theta}^{1+\gamma}(y_i | \mathbf{x}_i) dy_i \\
 & = \int_{\mathbf{y}_i} \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta})^2 \right. \\
 & \quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \boldsymbol{\beta}) \right\}^2 \right]^2 f_{\theta}^{1+\gamma}(y_i | \mathbf{x}_i) dy_i \\
 & = \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \int_{\mathbf{y}_i} f_{\theta}^{1+\gamma}(y_i | \mathbf{x}_i) dy_i \\
 & \quad + \frac{1}{4\sigma_e^8} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta})^2 \right\}^2 f_{\theta}^{1+\gamma}(y_i | \mathbf{x}_i) dy_i \\
 & \quad + \frac{\sigma_v^4 (2\sigma_e^2 + a_i \sigma_v^2)^2}{4\sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^4} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \boldsymbol{\beta}) \right\}^4 f_{\theta}^{1+\gamma}(y_i | \mathbf{x}_i) dy_i
 \end{aligned}$$

$$\begin{aligned}
& - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2 \sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} \int_{y_i} a_{ij} (y_{it} - x_{it} \beta)^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& + \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2) \sigma_v^2 (2 \sigma_e^2 + a_i \sigma_v^2)}{2 \sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)^3} \int_{y_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \beta) \right\}^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
& - \frac{\sigma_v^2 (2 \sigma_e^2 + a_i \sigma_v^2)}{2 \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \sum_{j=1}^{n_i} \int_{y_i} a_{ij} (y_{ij} - x_{ij} \beta)^2 \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \beta) \right\}^2 f_\theta^{1+\gamma} (y_i | \mathbf{x}_i) dy_i \\
= & \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4 \sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M(1 + \gamma) \\
& + \frac{1}{4 \sigma_e^8} \frac{M \{ (n_i^2 + 2 n_i) \sigma_e^4 + 2(n_i + 2) a_i \sigma_e^2 \sigma_v^2 + 3 a_i^2 \sigma_v^4 \}}{1 + \gamma} \\
& + \frac{\sigma_v^4 (2 \sigma_e^2 + a_i \sigma_v^2)^2}{4 \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^4} \times \frac{3 M (a_i \sigma_e^2 + a_i^2 \sigma_v^2)^2}{1 + \gamma} \\
& - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2 \sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)} \times \sum_{j=1}^{n_i} M (\sigma_e^2 + a_{ij} \sigma_v^2) \\
& + \frac{(n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2) \sigma_v^2 (2 \sigma_e^2 + a_i \sigma_v^2)}{2 \sigma_e^6 (\sigma_e^2 + a_i \sigma_v^2)^3} \times M (a_i \sigma_e^2 + a_i^2 \sigma_v^2) \\
& - \frac{\sigma_v^2 (2 \sigma_e^2 + a_i \sigma_v^2)}{2 \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M \left\{ \frac{(n_i a_i + 2 a_i) \sigma_e^4 + (n_i a_i^2 + 2 a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3 \sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (a_i^2 - a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
= & \frac{M(\gamma - 1) (n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2)^2}{4 \sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \\
& + \frac{M \{ (n_i^2 + 2 n_i) \sigma_e^4 + 2(n_i + 2) a_i \sigma_e^2 \sigma_v^2 + 3 a_i^2 \sigma_v^4 \}}{4 \sigma_e^8 (1 + \gamma)} \\
& + \frac{3 M a_i^2 \sigma_v^4 (2 \sigma_e^2 + a_i \sigma_v^2)^2}{4 (1 + \gamma) \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \\
& - \frac{\sigma_v^2 (2 \sigma_e^2 + a_i \sigma_v^2)}{2 \sigma_e^8 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M \left\{ \frac{(n_i a_i + 2 a_i) \sigma_e^4 + (n_i a_i^2 + 2 a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3 \sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (a_i^2 - a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{J}_{\beta, \sigma_v^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\beta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right] \\
&\quad \times \left[ -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right\}^2 \right] f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= 0.
\end{aligned}$$

同样地,

$$\begin{aligned}
\mathbf{J}_{\beta, \sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\beta}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} a_{ij} x_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) - \frac{a_i \sigma_v^2 \bar{\mathbf{x}}_i}{\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right] \\
&\quad \times \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta})^2 \right. \\
&\quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \boldsymbol{\beta}) \right\}^2 \right] f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\mathbf{J}_{\sigma_v^2, \sigma_e^2}^{(i)} &= \int_{\mathbf{y}_i} \mathbf{u}_{\sigma_v^2}(\mathbf{y}_i | \mathbf{x}_i) \mathbf{u}_{\sigma_e^2}(\mathbf{y}_i | \mathbf{x}_i) f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \int_{\mathbf{y}_i} \left[ -\frac{a_i}{2(\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2(\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right\}^2 \right] \\
&\quad \times \left[ -\frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)} + \frac{1}{2\sigma_e^4} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta})^2 \right. \\
&\quad \left. - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{2\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{it} - x_{it} \boldsymbol{\beta}) \right\}^2 \right] f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&= \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \int_{\mathbf{y}_i} f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&\quad + \frac{1}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \sum_{j=1}^{n_i} \int_{\mathbf{y}_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta})^2 \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right\}^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
&\quad - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^4} \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij} \boldsymbol{\beta}) \right\}^4 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i
\end{aligned}$$

$$\begin{aligned}
& - \frac{a_i}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} \int_{\mathbf{y}_i} \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta)^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
& + \left[ \frac{a_i \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^3} - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^3} \right] \int_{\mathbf{y}_i} \left\{ \sum_{j=1}^{n_i} a_{ij} (y_{ij} - x_{ij}\beta) \right\}^2 f_{\theta}^{1+\gamma}(\mathbf{y}_i | \mathbf{x}_i) d\mathbf{y}_i \\
= & \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M(1 + \gamma) \\
& + \frac{1}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \times M \left\{ \frac{(n_i a_i + 2a_i) \sigma_e^4 + (n_i a_i^2 + 2a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (a_i^2 - a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
& - \frac{\sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^4} \times \frac{3M (a_i \sigma_e^2 + a_i^2 \sigma_v^2)^2}{1 + \gamma} \\
& - \frac{a_i}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)} \times M (n_i \sigma_e^2 + a_i \sigma_v^2) \\
& + \left[ \frac{a_i \sigma_v^2 (2\sigma_e^2 + a_i \sigma_v^2)}{4\sigma_e^4 (\sigma_e^2 + a_i \sigma_v^2)^3} - \frac{n_i \sigma_e^2 + (n_i - 1) a_i \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^3} \right] \times M (a_i \sigma_e^2 + a_i^2 \sigma_v^2) \\
= & M(1 + \gamma) \frac{n_i a_i \sigma_e^2 + (n_i - 1) a_i^2 \sigma_v^2}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \\
& + \frac{M}{4\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2} \left\{ \frac{(n_i a_i + 2a_i) \sigma_e^4 + (n_i a_i^2 - 4a_i^2 + 3 \sum_{j=1}^{n_i} a_{ij}^2) \sigma_e^2 \sigma_v^2}{1 + \gamma} \right. \\
& \quad \left. + \frac{3\sigma_v^4 \sum_{j=1}^{n_i} a_{ij} (-a_{ij} a_i + a_{ij}^2)}{1 + \gamma} \right\} \\
& - \frac{M a_i (n_i \sigma_e^2 - (n_i - 1) a_i \sigma_v^2)}{2\sigma_e^2 (\sigma_e^2 + a_i \sigma_v^2)^2}.
\end{aligned}$$



## 攻读学位期间发表的学术论文与研究成果

### 已发表 (或正式接受) 的学术论文:

1. Temperature trend analysis and extreme high temperature prediction based on weighted Markov Model in Lanzhou[J]. Natural Hazards, 2021, 108(1).
2. 基于  $\gamma$  散度的单元水平模型小域稳健估计 [J]. 统计与信息论坛, 2023, 38(3): 3-15.
3. Robust Small Area Estimation for Unit Level Model with Density Power Divergence[J], PLOS ONE.

### 参加的研究项目及获奖情况:

- 国家社科基金重点项目“乡村治理绩效测度与评价研究”(20ATJ006) (参与);
- 甘肃省优秀研究生“创新之星”项目“基于小域估计的住户调查分析”(2021CXZX-698) (主持);



# 致谢

经过几轮修改，我的博士论文终于完成，这意味着我的求学生涯暂时也将告一段落。常常听到有人说，读书是一件最幸福的事，其实，并不是这样的，至少读博不是这样的。四年的读博生活实属不易，只有亲自经历了方能知道其中的滋味，工作、学习、生活上的努力对我今后的人生道路一定有莫大的帮助。这四年中，更像是进行一场“修行”，我收获的不仅是学术上的提升，更多的是思想和心态上的转变。这一路走来，更多的是收获和感激，想说的话有太多太多，想感谢的人也太多太多。

首先我要特别感谢我的导师庞智强教授。庞教授学识渊博，阅历丰富，治学严谨，是一位特别负责的老师。在我四年的博士生涯中倾注了大量的心血，在我工作、学习和生活等各个方面给予了许多关怀和教诲。我的毕业论文能够顺利完成，与老师的精心指导是密不可分的。从论文的选题、框架的搭建和最终的定稿，每一个过程，每一个环节，老师都悉心指导，提出了许多宝贵的建议。在此附老师给我修改论文的冰山一角（图7.1），从图中就可以看出老师严谨的治学态度和高深的学术造诣。从老师身上，我深切体会到学术研究的意义不应以发表几篇论文为目的，最重要的是要解决现实生活中遇到的问题。这四年中，老师带领我多次参加实地调研活动，资助我参加各类学术交流，让我开阔了眼界，有更多机会接触到一些学术圈中的大牛。师恩深重，终身难忘，在此谨向我的导师致以诚挚的感谢和崇高的敬意！

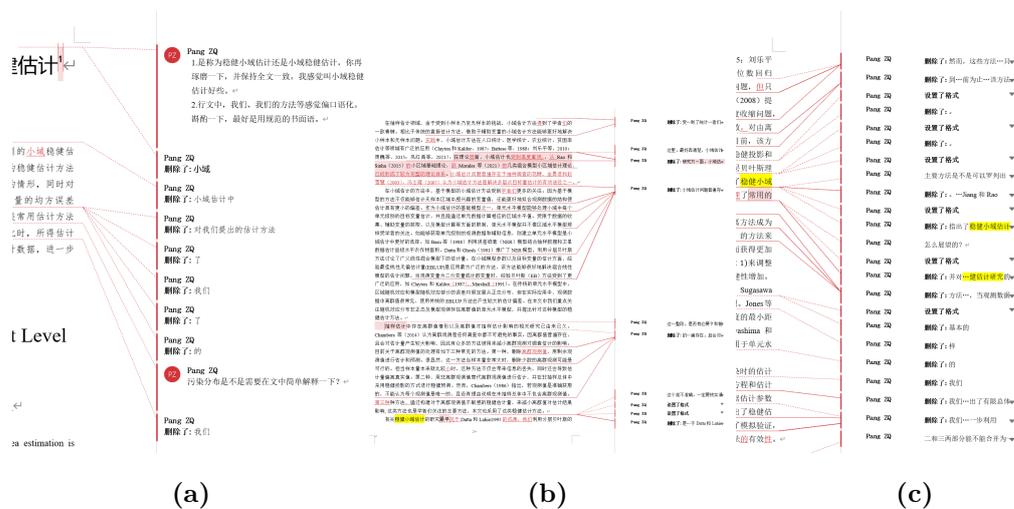


图 7.1 老师修改的部分论文截图

其次我要感谢我的爱人牛玺娟。正是她默默而无私的付出才使我有信心和毅力完成论文写作，她的体谅和包容是我顺利完成学业的前提和基础。我们相识于硕士阶段，能够在相识中相知、相恋、结婚，再到一起读博，我们共同度过了整整十年时间，这确实是修来的缘分。感谢我的爱人这些年为我生儿育女，给了我一个温暖的家。曾经有位名人说过，“人生的道路虽然漫长，但紧要处常常只有几步，特别是在年轻的时候”。感谢我们在正确的时间里遇到了彼此，在这段艰难奋战的日子里互相勉励并共同克服学习上的难关，这些将是我记忆里最美的风景。

在博士毕业之际，我要感谢我的父母和岳父母，感谢他们一路走来的帮助和陪伴。世间没有所谓的岁月静好，只不过是有人替你负重前行。感谢他们在身后默默地支持，在我们夫妻双双读博的日子里主动伸出援助之手，他们无微不至地照顾我们的两个孩子，帮助我们共同度过这段不平凡的岁月。同时，他们的存在也让我们切身体会到了上有老下有小的普通老百姓的幸福生活。倘若没有他们的支持与鼓励，我无法顺利完成学业，这份感激之情，永远铭记于我心中。

感谢 2019 级博士班的同学们，感谢他们陪我一起走过这段人生难忘的历程！我们都是同路人，相同的经历和相似的环境，让我们学会了相濡以沫，同舟共济的这四年是我们无法忘怀的美好记忆。感谢共同学习、讨论和生活的同学和师弟师妹，在学习讨论中我们通力合作，在日常生活中其乐融融，让我充满活力，不断向前，在此感谢他们在长期学习中给予的支持、理解和帮助。

感谢青海师范大学的领导对我攻读学位的大力支持，感谢他们为我读博提供的方便，使我边工作边学习完成了学位的攻读。同时，感谢兰州财经大学统计学院各位领导、老师在我博士学位攻读期间对我的关心、支持和帮助！

最后，再次感谢所有帮助过我和关心过我的人，我将永远带着感恩的心，好好努力，好好生活！

2023 年 5 月

王朝旭