4.1 多重共线性

胡华平

## 1 教学目标

* 能够正确写出多重共线性数学表达式；
* 能够区分多重共线性的不同类型；
* 能够列举引起多重共线性的三个主要原因；
* 能够说明自变量相关性与模型多重共线性的联系；
* 能够正确说出多重共线性下OLS参数估计量公式的含义；
* 能够正确写出多重共线性下OLS估计量的方差和协方差公式；
* 能够写出方差膨胀因子（VIF）和容忍度（TOL）计算公式；
* 能够列举多重共线性理论后果和实际后果；
* 能够读懂主回归方程报告，正确得出诊断结论；
* 能够读懂矩阵相关系数和矩阵散点图，正确得出诊断结论；
* 能够正确计算方差膨胀因子（VIF）和容忍度（TOL），正确得出诊断结论；
* 能够读懂EViews软件给出的回归系数方差分解表，正确得出诊断结论；
* 能够综合应用多种诊断方法，对实际案例进行全面诊断；

## 2 课时安排

* 课时1（40分钟）：多重共线性的内涵与性质
* 课时2（40分钟）：多重共线性的问题与后果
* 课时3（40分钟）：多重共线性的诊断
* 课时4（40分钟）：多重共线性的矫正

## 3 教学内容

1. 多重共线性的内涵与性质：首先讲述多重共线性的基本概念、定义和数学表达式，然后重点区分多重共线性的不同类型，并梳理引起多重共线性的原因，最后概括多重共线性的特征。
2. 多重共线性的问题与后果：重点讲述多重共线性时的估计问题，然后梳理多重共线性理论后果和实际后果。
3. 多重共线性的诊断：将依次讲述多种诊断方法，分别是主回归方程报告观察法、矩阵相关系数和矩阵散点图观察法、辅助回归方程比较法、回归系数方差分解法。
4. 多重共线性的矫正：讲述多种矫正方法，分别是经济学和实践观察法、变量变换法、逐步最小二乘回归法、因子分析法等。

## 4 教学步骤

### 4.1 课时1：多重共线性的内涵与性质

#### 4.1.1 课堂引入

回顾经典假设CLRM，突出强调多元回归模型中的新增一条假设（学生容易忽视）；回顾经典假设下OLS方法的BLUE性质（简单板书），总结经典假设的作用和重要性。

#### 4.1.2 讲授内容

1. 多重共线性的含义：语义定义、区分共线性与多重共线性；
2. 多重共线性的数学表达：多重共线性数学表达式、完全共线性数学表达式；
3. 多重共线性的类型及多重共线性与变量相关的联系：无共线性、低度多重共线性、中度多重共线性、高度多重共线性；
4. 引发多重共线性的原因：样本数据的限制、模型设定不谨慎、相同时间趋势、过度决定

#### 4.1.3 案例分析

1. 朗读和分享蒙哥马利(Montgomery) 和佩克(Peck)等的观点，增强知识广度和观点认同。引入医药研究中从少数病人上收集大量变量信息的案例，说明过度决定问题，加深对形成原因的直观认知。
2. 案例情景类比：公司老板很不客气地解雇某一个员工，因为他缺少“核心竞争力”，很容易被同一部门其他人替代。通过情景类比指出多重共线性的核心问题是变量间的“共线性”——相互可替代。

#### 4.1.4 课程讨论

1. 为什么多重共线性问题是对CLRM假设的冲击？多重共线性如何表达？多重共线性有哪些种类？为什么会普遍存在多重共线性？
2. 表达式只有自变量X之间的关系，为什么还要考虑到因变量Y？多重共线性如果不能够避免，那有没有办法减弱吗？多重共线性就一定是坏事吗？

### 4.2 课时2：多重共线性的问题与后果

#### 4.2.1 课堂引入

回顾分析OLS估计的内容和估计过程、OLS估计的性质：BLUE。用“从样本推断总体”的故事线快速串讲OLS估计的步骤流程，回顾在CLRM经典假设下OLS方法“完美”地获得了BLUE估计性质。

#### 4.2.2 讲授内容

1. 多重共线性的估计问题：完全多重共线性时、不完全多重共线性时、OLS估计量的方差和协方差、方差膨胀因子（VIF）、容忍度（TOL）。
2. 多重共线性的理论后果：微数缺测性、多维样本；多重共线性下OLS估计量性质；BLUE性质和LUE性质的关系。
3. 多重共线性的实际后果：方差和协方差，估计精度；方差膨胀因子(VIF)和容忍度(TOL)；置信区间；系数的t检验；判定系数$R^{2}$；0LS估计对数据变化的敏感度；

#### 4.2.3 案例分析

1. 案例1：“人为模拟”不同多重共线性情形，演示OLS点估计的置信区间不断放大；模拟微小样本变动后数值的不稳定性。
2. 案例2：再次通过图形化消费（Y）、收入(X2)和财富(X3)的案例数据，分析和展现多重共线性问题的各种“看得到”的影响。
3. 案例3：朗读和分享克里斯托弗.阿肯 (Christopher Achen)等的观点，介绍更加专业的“微数缺测性”和“多维样本”概念及研究结论，进一步给出为什么“多重共线性”下仍能保持BLUE性质的理由。

#### 4.2.4 课程讨论

提问：既然多重共线性下OLS估计量仍然保持BLUE性质，那么大谈特谈多重共线性究竟是为了什么？多重共线性是庸人自扰么？

讨论：只要不是“完全共线性”，就不会破坏OLS估计的BLUE性质；但是“多重共线性”仍会带来不少“麻烦”。

### 4.3 课时3：多重共线性的诊断

#### 4.3.1 课堂引入

复习回顾：EViews报告结果解读；多重共线性的理论和实际后果

导入新课：怎样识别多重共线性带来了“大麻烦”还是“小麻烦”？

#### 4.3.2 讲授内容

1. 观察主回归报告：诊断标准：$R^{2}$值高，F显著，但显著的t比率少（经典法则）
2. 观察矩阵相关系数和矩阵散点图：矩阵相关系数；矩阵散点图 诊断标准：相关系数$>0.9$%，越高则越严重；散点图呈直线模式，越多则越严重。
3. 比较辅助回归方程：辅助方程、方差膨胀因子(VIF)、容许度(TOL)。诊断标准：比较辅助回归方程的$R\_{i}^{2}$与主回归方程$R^{2}$；比较方差膨胀因子(VIF)与容许度(TOL)。
4. 回归系数方差分解法：特征根；病态数K；方差分解比率VDP 诊断标准：至少一个病态数 $K\leq 0.001$，多个斜率系数的 $VDP\geq 0.5$

#### 4.3.3 案例分析

1. 案例1：介绍工人就业情况的朗利数据及变量，强调该案例将用于验证后续所有诊断方法。展示EViews的主回归报告结果，请同学回答如何解读$R^{2}$值、F显著和t检验结论，巩固旧知识，综合回答新问题。
2. 案例2：进行图表展示与互动，展示EViews软件计算得到的矩阵相关系数和矩阵散点图，提出明确的诊断标准。
3. 案例3：用“盲人摸象”，来类比多重共线性诊断可以多么“不靠谱”；以及用“全面体检时刻检查身体状态”来说明系统性诊断的重要性。

#### 4.3.4 课程讨论

什么情况下，多重共线性诊断会变得极为困难？

通过更深刻的提问，激发学习更复杂多重共线性问题诊断方法的兴趣，扩展更多前沿分析思想和工具。

### 4.4 课时4：多重共线性的矫正

#### 4.4.1 课堂引入

复习回顾：多重共线性的本质、多重共线性的诊断方法。

导入新课：如果多重共线性严重，怎么办？引发思考，导出本节课内容。

#### 4.4.2 讲授内容

1. 利用先验信息补救：先验信息、专家经验
2. 调整数据使用策略：横截面与时间序列数据并用、补充新数据
3. 剔除问题变量：简单删除（先验知识）、逐步删除（逐步回归分析法）
4. 变量替换：一阶差分变换、比率变换
5. 多项式回归模型的补救：离差变换、正交多项式
6. 其他方法：岭回归（Ridge Regression）、因子分析（Factor Analysis）

#### 4.4.3 案例分析

1. 案例1：类比说明：“中药好，西药好，还是中西医结合？”用“医生”角色类比模型矫正工作。引用谚语：引用谚语“按下葫芦起了瓢”，强调矫正的后果，让观点更鲜活、更明确。
2. 案例2：通过工人就业情况的朗利数据案例，展示简单删除和逐步删除的过程和结果。实验展示：展现EViews软件分析结果，比较剔除变量前后模型结果的差异。
3. 案例3：文献扩展：提供关于岭回归方法的若干重要文献，提示用到矩阵代数知识。实验展示：展现EViews软件进行因子分析的实验结果，总结分析过程和结论。

#### 4.4.4 课程讨论

提问：矫正多重共线性问题，可能会引起哪些“副作用”？

讨论：经验矫正方法与定量矫正方法可以互补。矫正方法可能带来副作用，不要对多重共线性问题过度反应。

## 5 评估方式

形成性评估：讨论、案例分析、小组合作等

总结性评估：课堂测试、小组互评等

## 6 课程总结

采用课堂问答的形式提高学生的课程反馈和掌握程度。

1. 课时1（多重共线性的内涵与性质）是在掌握“一元线性回归模型”和“多元线性回归模型”的全部内容后的扩展，与这些已学知识的有效衔接极为关键。相关概念和定义容易产生混淆，需要进行多维度阐述。引起多重共线性的原因有很多，分享引导阅读相关文献资料将很有帮助。
2. 课时2（多重共线性的问题与后果）是在“线性回归模型OLS估计方法”和“OLS估计的BLUE性质”基础上的扩展，与这些已学知识的有效衔接极为关键。OLS估计的BLUE性质涉及到复杂而严谨的数学推导，涉及到多重共线性问题后会进一步增大学习难度。
3. 课时3（多重共线性的诊断）中所有诊断方法都紧密围绕自变量(X)之间关系这一点展开。主回归方程报告观察法、矩阵相关系数和矩阵散点图观察法，相对比较直观，基本上是对已学知识的综合应用。辅助回归方程比较法和回归系数方差分解法属于全新内容，其诊断步骤和过程相对复杂。
4. 课时4（多重共线性的矫正）中各种矫正手段并不总是有效，需要基于已学的“多重共线性诊断”结论。经济学和实践观察法、变量变换法，操作相对比较主观，需要良好的经济学理论基础和经济社会经验。逐步最小二乘回归法和因子分析法，主要依赖计算机软件程序来控制，背后运作原理和流程相对复杂。

## 7 教学资源

### 7.1 网络资源

* 在线课程：《计量经济学》在线开放课程(胡华平主讲，全球公开访问) <https://www.huhuaping.com/course/course-em/>
* 多媒体教学资源：西北农林科技大学在线网络教学平台《计量经济学》(需要学校权限访问) <http://eol.nwsuaf.edu.cn/meol/jpk/course/layout/newpage/index.jsp?courseId=35257>

### 7.2 图书文献

* 李子奈, 潘文卿. 计量经济学[M]. 第5版. 北京: 高等教育出版社, 2020.
* 古扎拉蒂. 计量经济学基础[M]. 第5版.北京:中国人民大学出版社, 2011.
* 伍德里奇. 计量经济学导论: 现代观点[M].第6版. 北京: 中国人民大学出版社, 2018.

### 7.3 公开课平台

* 爱课程网：http://www.icourses.cn/home/
* 北京大学MOOCs课程：http://mooc.pku.edu.cn
* 新浪公开课：http://open.sina.com.cn/
* 网易公开课：http://open.163.com/
* 西北农林科技大学尔雅通识课程网址：http://nwsuaf.fanya.chaoxing.com/portal
* 西北农林科技大学网络教学综合平台http://eol.nwsuaf.edu.cn/eol/homepage/common/