* **自相关的概念，产生原因，如何检验，处理的方法，产生的影响。**

**概念：对于K元线性回归模型：,如果不同随机误差项之间存在相关关系，即cov(,,)!=0,则称模型存在自相关性**

**原因：**

1. **模型中遗漏了重要的解释变量(消费，收入；消费习惯，各期相关)**
2. **经济变量的惯性作用(前后期之间互相关联)**
3. **模型设定不当的影响**
4. **一些随机干扰因素的影响**

**检验：**

**主要相关性检验有四种：图示法、回归检验法、杜宾-瓦森检验法（D.W.）、拉格朗日检验（GB）。**

**最好的检验方法应该是GB检验，适用于高阶序列相关及模型中存在滞后变量的情形。D.W.检验中，存在一个不能确定的D.W.值区域，且仅能检测一阶自相关，对存在置后被解释变量的模型无法检验**

**处理：**

1. **使用ols+异方差自相关稳健的标准误，只改变标准误的估计值，不改变回归系数的估计值**
2. **使用ols+聚类稳健的标准误(面板数据)**
3. **可行广义最小二乘法**
4. **修改模型设定**

**影响：**

1. **自相关不影响OLS估计量的线性和无偏性，但使之失去有效性**
2. **自相关的系数估计量将有相当大的方差**
3. **自相关系数的T检验不显著**
4. **模型的预测功能失效**

* **异方差的概念，产生原因，如何检验，处理的方法，产生的影响**。

**概念：对于模型…如果出现 ，即对于不同样本点，随机误差项的方差不再是常数，而互不相同，则认为模型出现了异方差**

**原因:**

1. **模型中遗漏了某些解释变量**
2. **模型函数形式的设定误差**
3. **样本数据的测量误差**
4. **随机因素的影响**

**检验: 图示检验法、Goldfeld - Quandt 检验法、White检验法、Park检验法,Gleiser检验法。**

**处理:**

1. **使用ols+稳健标准误**
2. **广义最小二乘法**
3. **加权最小二乘法**
4. **可行广义最小二乘法**

**影响：如果线性回归模型的随机误差项存在异方差性，会对模型参数估计、模型检验及模型应用带来重大影响**

**1）不影响模型参数最小二乘估计值的无偏性**

**2）参数的最小二乘估计量不是一个有效的估计量**

**3）对模型参数估计值的显著性检验失效**

**4）模型估计式的代表性降低，预测精度精度降低。**

* **如何解决遗漏变量带来的偏差。多重共线性的概念，产生原因，如何检验，处理的方法，产生的影响。虚拟变量的概念以及它的作用，影响**。

**遗漏变量:遗漏的解释变量，与解释变量相关或者不相关，不相关不用考虑，不会影响。**

1. **增加控制变量**
2. **随机试验和自然实验(准实验法)**
3. **工具变量法**
4. **面板数据**

**概念：多重共线性是指线性回归模型中的解释变量之间由于存在精确**[相关关系](https://baike.baidu.com/item/%E7%9B%B8%E5%85%B3%E5%85%B3%E7%B3%BB/9227098)**或高度相关关系而使模型估计失真或难以估计准确。**

**原因：**

1. **经济变量相关的共同趋势**
2. **滞后变量的引入**
3. **样本资料的限制**

**检验:**

1. **相关系数检验**
2. **辅助回归模型检验**
3. **方差膨胀因子法**
4. **直观判断法**

**处理:**

1. **排除引起共线性的变量影响**
2. **差分法**
3. **减小参数估计量的方差岭回归法**
4. **简单相关系数检验法**

**影响:**

1. **参数估计量经济含义不合理**
2. **变量的显著性检验失去意义，可能将重要的解释变量排除在模型之外**
3. **模型的预测功能失效**

**虚拟概念：虚拟变量又称虚设变量、名义变量或**[哑变量](https://baike.baidu.com/item/%E5%93%91%E5%8F%98%E9%87%8F/9621990)**，用以反映质的属性的一个**[人工变量](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E5%8F%98%E9%87%8F/19127829)**，是量化了的自变量，通常取值为0或1。**

**作用：**

1. **分离异常因素的影响**
2. **检验不同属性类型对因变量的作用**
3. **提高模型的精度， 扩大了样本容量（增加了误差自由度，从而降低了误差方差）**

* **工具变量的概念，能够解决什么问题。有效工具变量应该满足的条件以及有关工具变量的各种检验方式。**

**概念: 如能将内生变量分成两部分，一部分与扰动项相关，另一部分与扰动项不相关，可用与扰动项不相关的那部分得到一致估计。这种分离常借助另一“工具变量”来实现。**

**能够解决解释变量和扰动项相关的回归估计问题从而得到一致估计。**

**需要满足：**

1. **相关性：工具变量与内生解释变量相关，即。**
2. **外生性：工具变量与扰动项不相关，即。**

**检验：**

1. **弱工具变量检验(包含很少与有关的信息，利用这部分信息进行的工具变量法估计就不准确，即样本容量很小也很难收敛到真实的参数值**
2. **过度识别检验**
3. **究竟用ols还是工具变量法:对解释变量内生性的检验(豪斯曼)**
4. **不可识别检验(检验秩条件是否成立)**

* **二值选择模型和多值选择模型的概念。了解logit模型和probit模型，以及这两个模型在生物统计中的一些应用。**

**如果被解释变量为离散变量或者虚拟变量时，使用离散选择模型(二值或者多值)，也就是面板二值选择模型。**

**以二值选择（被解释变量取值为0或1）为例，当被解释变量取1的概率为标准正态分布时，使用probit模型；当被解释变量取1的概率为logistic分布时，使用logit模型。**

**当模型的被解释变量为二值变量时，线性回归方法一般不再适用，需要采用其他方法。二值选择模型关注的是自变量的变动对因变量取值的概率的影响。**

**实际应用中，常用Probit模型和Logit模型对二值选择模型进行估计，两者的区别在于对连接函数具体形式的设定不同。多值:多项p和多项l模型，条件和混合l模型。应用:比如是否吸烟，喝酒对肠癌等病的影响，以及某些二分类因素对某疾病的影响等**

* **断尾回归的概念，断尾回归可以使用OLS， MLE估计吗？如果可以使用，需要注意什么问题。归并回归的概念，了解Tobit模型。**

**概念:对于线性模型，假设由于某种原因，只有满足的数据才能观测到，因此，当时，没有任何有关的数据，例如企业销售收入，统计局只统计>=1000的情况，这样被假释变量则在左边断尾。**

**OLs回归中，扰动项和解释变量相关，导致不一致的估计，收敛不到真实的系数值，同时可能出现预测值<=(断点处)的情形；mle可以得到一致估计，由于随机变量断尾后，其概率分布也会随着发生变化，OLS回归会导致不一致的结果，所以一般用MLE进行估计。极大似然估计需要建立似然函数，然后取对数求导，解似然方程。被解释变量受限的另一种情况是对于线性模型，当时，所有都被归于，这种数据为归并数据，虽然有所有数据，但是时一个离散点和连续分布，Ols是不一致的估计，tobit就是mle估计，写出整个样本的似然函数，然后mle估计。**

* **面板数据的概念，分类，主要的优点以及会带来的问题。了解面板数据中的固定效应估计量和一阶差分估计量，并尝试证明这两个估计量随时点变化会发生什么变化**。**在处理面板数据时，应该使用固定效用还是随机效用模型。**

**概念：面板数据，也叫“平行数据”，是指在时间序列上取多个截面，在这些截面上同时选取样本观测值所构成的样本数据。或者说他是一个的数据矩阵，记载的是个时间节点上，个对象的某一数据指标。**

**分类：**

1. **混合估计模型(在时间和截面上都不存在显著性差异，就可以把面板数据混合在一起用最小二乘法)**
2. **固定效应模型(对不同截面或者时间序列，模型的截距是不同的。个体，时点，个体时点)**
3. **随机效应模型(是与不相关的随机变量，不同个体有不同)**

**其它:**

**动态，系变数，面板数据的向量自回归，非均衡面板数据，离散**

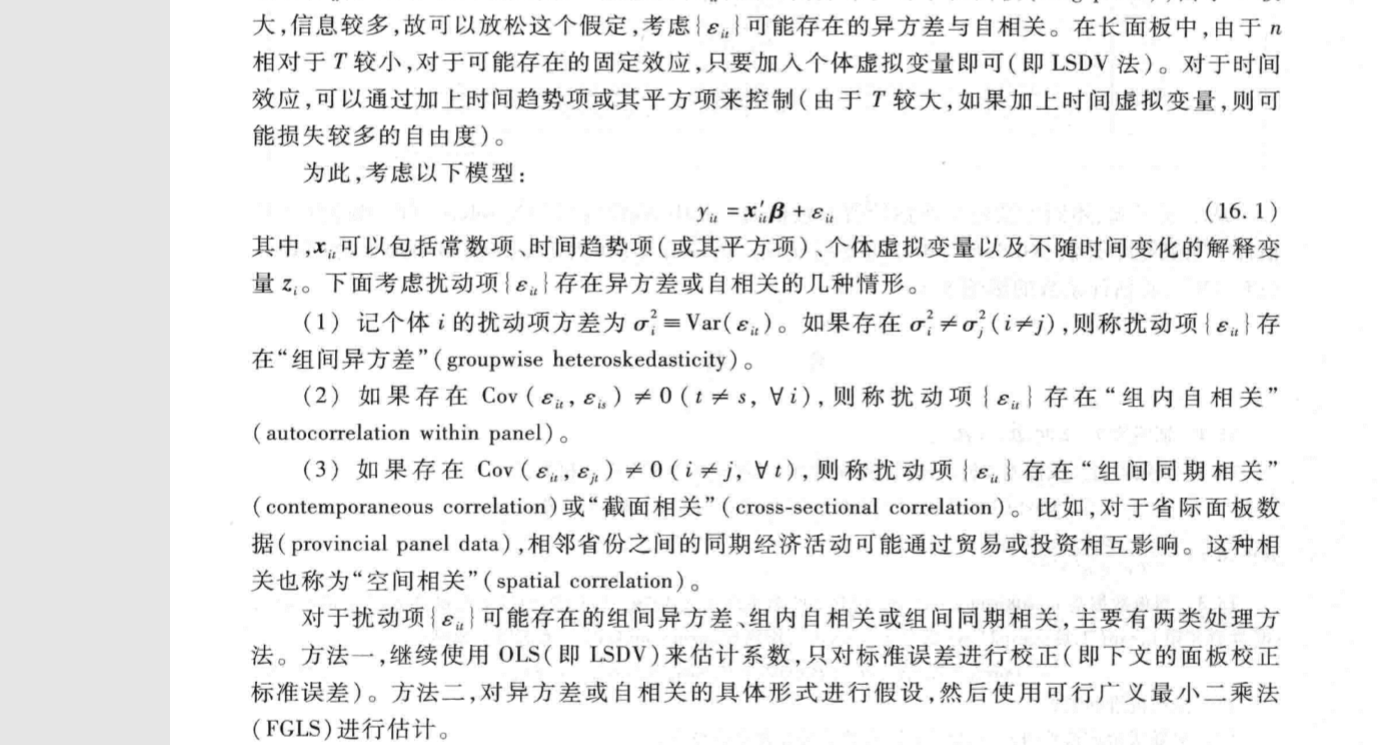
**优缺点:**

1. **面板数据可以解决遗漏变量问题。遗漏变量通常由于不可观察的个体差异或“异质性”导致，如果这一异质性不随时间变化，那么面板数据便提供了解决遗漏变量的利器，而这是截面数据不能解决的。**
2. **面板数据提供了个体的动态行为信息。面板数据兼具横截面和时间两个维度，可以解决横截面数据和面板数据单独不能解决的问题。**
3. **两个维度的数据使得面板样本容量大幅增加，与横截面相比，明显提高估计的精确度,因为很多估计量和检验都是在大样本下得到的渐进分布。**

**样本数据通常不满足独立同分布的假定，因为同一个体在不同期的扰动项一般存在自相关，收集成本较高，不易获得。原假设****与****、****不相关(即****为正确模型)，成立****，****都成立，但****更有效，不成立则****成立（豪斯曼,不适用异方差）**

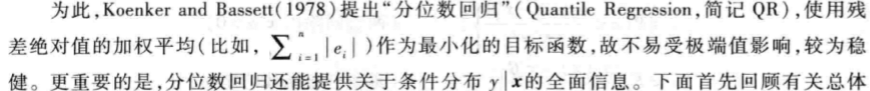
* **长面板数据的特点，和短面板数据的差异。对长面板数据，样本容量较大时，可建立什么模型，简要叙述该模型。**

**特点：长面板数据时间维度T较大，信息较多，可以放宽扰动项独立同分布的假定。短面板时间维度较小，每个个体的信息较少，无法探讨扰动项是否存在自相关，故假定其独立同分布。**



* **为什么需要分位数回归。分位数回归的估计方法是什么**。

**之前的回归模型中，考察对的条件期望的影响，是均值回归，关心的是对整个条件分布的影响，而只是刻画集中趋势的一个指标。如不是对称分布，则很难反映整个条件分布全貌。**



**方法：**

**1)点估计:单纯形算法(参数有较好稳定性，大型数据速度慢)，内点算法(对大量观察值和少量变量的数据集运算效率高)，平滑算法(适合处理有大量观察值和很多变量的数据集)**

**2)区间估计：直接估计(根据估计出来的回归分位系数的渐进正态性来计算置信区间)，秩得分法(简单，对大数据处理慢)，重复抽样(使用MCMB，能进行高效率运算，克服直接发和秩得分法的缺陷，小样本计算的参数估计值不稳定)**

* **非参与半参估计的意义。常用的核函数有哪些，并了解核密度估计和最优带宽的各种性质及两者之间的关联。**

**意义：参数估计法，即假设总体服从带未知参数的某个具体分布(比如正态分布),然后估计这些参数。但其对模型设定所作的假设较强，可能导致较大的‘设定误差’，比如，真实总体并非正太，甚至偏离正态较远，则在正态分布前提下所作的统计推断可能有较大偏差，而非参一般不对模型的具体分布做任何假定，故更为稳健，但其要求样本容量较大，而估计值收敛到真实值速度也较慢，半参则是一种折中，他降低了对样本容量的要求，又具有一定的稳健性，非参和半参与传统参数估计方法是互补关系，当后者不适用时，则考虑前者。**

**核函数: 均匀核，三角核，二次核，四次核，三权核，三三核，高斯核。核密度估计性质：一致性，渐近正态性**

**联系: 带宽的大小决定了核密度估计函数（KDE）的平滑（smooth）程度，带宽越小越undersmooth，带宽越大越oversmooth果带宽不是固定的，而是根据样本的位置而变化(其变化取决于估计的位置（balloon estimator）或样本点（逐点估计pointwise estimator）)，则会产生一种特别有力的方法，称为自适应或可变带宽的核密度估计。自适应带宽的核密度估计方法是在固定带宽核密度函数的基础上，通过修正带宽参数为而得到的.**

* **了解空间计量模型，以及各个模型的优缺点、适用范围。**

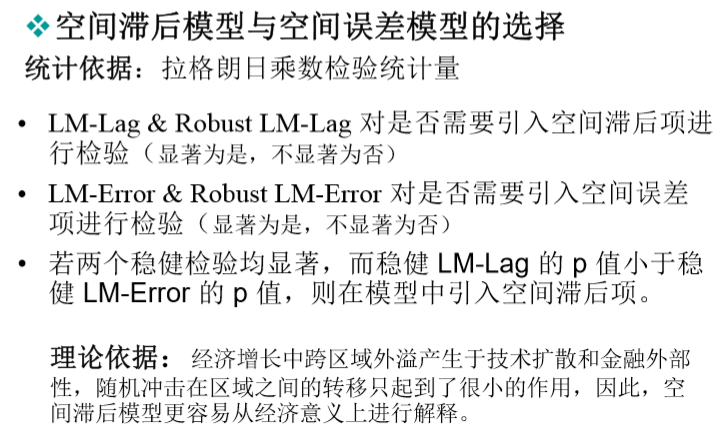
**空间自回归模型(滞后，最大似然估计) 探测由于各种空间溢出而产生的空间自相关，这种空间溢出来自于区域间存在的实质性空间相互作用，如技术扩散、 要素转移、知识交流等产生的扩散和极化效应。在经典线 性回归模型中引入空间滞后因变量：**

****

**空间误差模型：空间自相关在传统模型中往往被认为是噪音，它实际上 度量了邻近单元因变量的误差冲击对本单元观测值的影响 程度。空间误差模型中，空间自相关反映在误差项中：**



**一般的空间计量模型(自回归和误差结合起来)：**



**顾名思义，空间误差模型说明空间影响是在误差中，也就是说空间权重矩阵放在无法检测到的误差项中；空间之后模型则表明空间关联是在时间中，空间权重矩阵出现在以前的相关变量中。**

**空间杜宾模型：除了邻近地区的空间溢出效应外，如果在空间上邻近区域的解释变量对区域行为也有影响，需要采用空间杜宾 模型（SDM）：**

**SDM模型反映了区域行为的空间依赖性除直接受到邻 近其他地区行为的影响外，还来源于可能相互依赖的解 释变量等外生变量的间接影响。**

**地理加权回归：试图利用空间的非稳态性，使得变量间的关系随着空间的变化而变化；模型通过对回归方程残差项加权体现其空间变化。**

* **如何使用多项Logit回归，混合Logit模型，嵌套Logit模型对数据进行处理，并对结果进行分析。**

****

**解释变量只随个体而变，不随方案而变，然后用MLE进行估计（多项）。但有些解释变量可能既随个体而变，也随方案而变，这种解释变量称为’随方案而变’，既包括同时随方案与个体而变的变量，也包括随方案而变但不随个体而变的变量（条件），将解释变量不随方案而变的多项L和随方案而变的条件L同时发生就是混合的L。如果将多值选择模型中的任何两个方案单独挑出来，都是二值 logit模型，次假定称为“无关方案的独立性(IIA)’在实践中，如果不同方案之间很类似，则IIA假定不一定满足，这是多项，条件，混合共同缺点，此时考虑使用嵌套L 步骤：首先判断数据的基本特征和数据格式，通过描述性统计粗略判断要研究的关系 进行回归 IIA检验 相对风险比率 预测结果**

* **实例分析题来自第十二章**

