题型

单选(10\*2)

判断(10\*2)（记得把判断题都看一遍）

名词解释（5\*4）

简答题（4\*5）

计算和分析（6,6,8）

#以下出现的k均为解释变量个数，不包括截距项（书上的是包括的）

**外生变量**i.e.输入变量：完全由经济系统外部确定并输入系统的变量，只对系统产生影响而不受系统影响。在经济计量模型中，是与模型的随机扰动项不相关的变量。

# 一、

**经济计量学**：经济的测度。利用经济理论、数学、统计推断等工具对经济现象进行分析的一门社会科学。

**时间序列/截面数据**：按时间跨度收集到的/一个或多个变量在某一时点上的数据集合。（**合并数据**-既包括时又包括截；**面板/纵向/微观面板数据**-特殊类型的合并数据#同一个横截面单位的跨期调查数据）

经济计量学建立步骤：经济理论-理论的数学模型-理论的计量经济模型-数据（书上这一步在第二步）-计量经济模型的估计-（核查适用性#模型设定检验）-假设检验-预报或预测-利用模型进行控制或制定政策

# 二、双变量模型

**回归分析**：研究一个变量（被解释）关于另一个/些变量（解释）的具体依赖关系的计算方法和理论。（目的：根基自变量取值估计因变量均值+检验假设+预测）

总体回归函数PRF：在给定解释变量Xi条件下被解释变量Y的期望轨迹称为总体回归线，相应的函数称为PRF，说明被解释变量Y的平均状态i.e.总体条件期望随解释变量X变化的规律。

样本回归函数与总体回归函数之间的区别

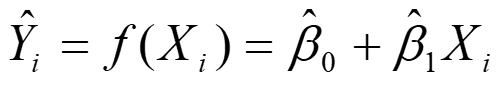
　　1.总体回归直线是未知的，且只有一条。而样本回归直线是已知的，且可以有许多条（从总体中可以取许多样本，每抽取一组样本，便可以拟合一条直线）。

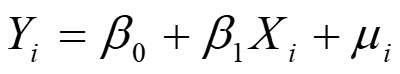
　　2.总体回归函数中β1、β2是未知参数，表现为常数；而样本回归函数中的 和

是随机变量，其具体数值随所抽取的样本观测值不同而不同。

　　3.总体回归函数中的ut是Yt与未知总体回归直线之间的纵向距离，它是不可直接观测的；而样本回归函数中的et是Yt与样本回归直线之间的纵向距离，当根据样本观测值拟合出样本回归直线后，可以计算出et的具体数值。ut和et在各自直线上的位置分别见下图a和b所示。

随机误差项/扰动项：记为观察值Yi围绕在它期望值的离差，为不可观测的随机变量（μi’包含的因素影响-解释变量中被忽略因素/变量观测值观测误差/模型设定误差/其它因素的影响#设定μi的原因-理论含糊/数据欠缺/节省原则）

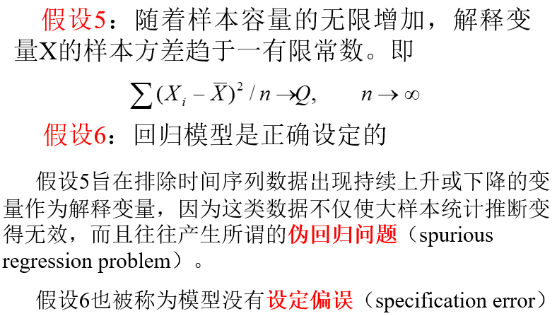
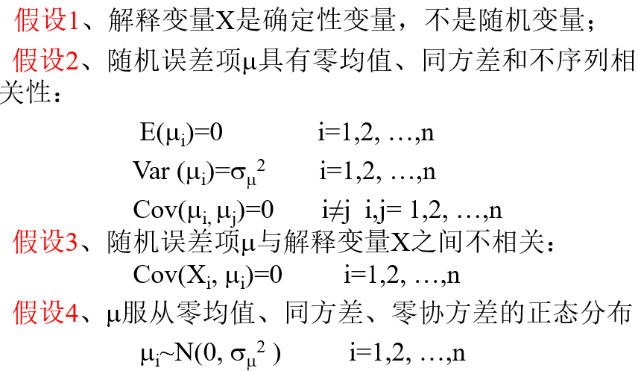
样本回归函数SRF：用取自总体的样本散点图拟合出总体回归线的近似替代，即样本回归线，其函数形式为，称为SRF。（残差项，代表其他影响Yi的随机因素集合，即μi的估计量）(样本回归模型：样本回归函数的随机形式，成为计量经济模型)

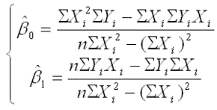
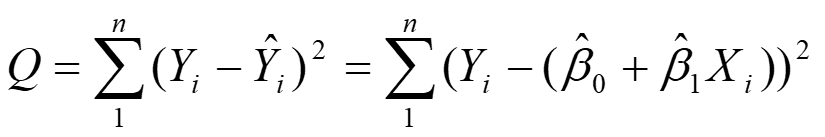
**一元线性回归模型**：只有一个解释变量(i=1...n)，Y为被解，X为解， β0和β1为待估参数，μ为随。

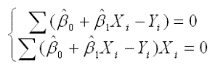
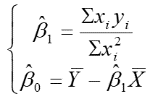
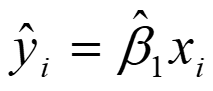
线性回归模型基本假设(4+2)：

4-解释变量X是确定性变量（非随机）；随机误差项μ具有零均值、同方差和不序列相关性；随机误差项μ与解释变量X间不相关；μ服从零均值、同方差和零协方差的正态分布.（经典线性回归模型-满足前4个假设/经典假设的线性回归模型）

2-随着样本容量无限增加，解释变量X样本方差趋于一有限常数（防止伪回归）；回归模型正确设定（无设定偏误）。

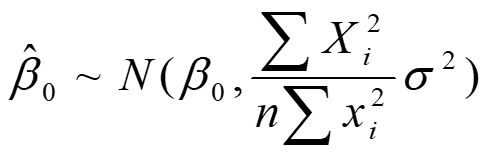


**普通最小二乘法OLS**：在给定样本观察值下，以残差平方和最小为原则来估计回归模型中系数的方法。

正规方程组→普通最小二乘估计量→离差形式; #样本回归函数离差形式

**最佳线性无偏估计BLUE**：具有小样本性质（线性性-估计量是否为另一随机变量的线性函数；无偏性-它的均值或期望值是否等于总体真实值；有效性-它是否在所有线性无偏估计量中具有最小方差）的估计量)#当不满足小样本性质时，需进一步考察估计量的大样本或渐近性质（渐近无偏性/一致性/渐近有效性-样本容量趋于无穷大时，是否它的均值序列趋于总体真值/它是否依概率收敛于总体真值/是否它在所有一致估计量中有最小渐近方差）#注意无偏性除了回归系数还包括误差方差

**高斯马尔科夫定理**：在给定经典线性回归假定下，最小二乘估计量是具有最小方差的线性无偏估计量i.e.BLUE。

  ；μ的方差i.e.总体方差的OLS估计量

**假设检验**：事先对总体参数或总体分布形式作出一个假设，然后判断样本信息与原假设是否有显著差异，从而决定是否接受或否定原假设。

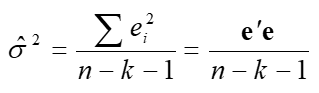
**参数检验的置信区间估计**：考察一个以样本参数估计值为中心的区间以多大可能性包含着参数真实值，来判断估计值可近似替代真值的程度。（为估计离β有多近，预先选择一个概率α i.e.显著性水平，并求一个正数δ，使得随机区间（）i.e.置信区间包含真值概率为1-α i.e.置信系数/度#临界值-置信区间端点）（缩小置信区间方法-提高样本容量/模型拟合优度/样本观测值分散度#针对多元）

**拟合优度检验**：对样本回归直线与样本观测值之间拟合程度的检验（**判定系数/可决系数R2**-度量拟合优度指标，R2∈[0,1]，越接近1拟合优度越高#拟合优度=ESS/TSS，余相关系数1-R2表示未被X解释的Y变异比例）（总体平方和i.e.Y的总离差；解释/回归平方和估计Y值围绕其均值的变异；残差平方和Y变异未被解释的部分）

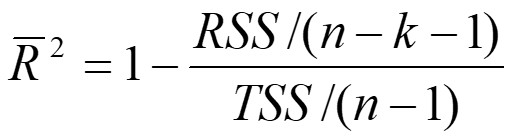
三、

**多元线性回归模型**：在线性回归模型中解释变量有多个。

**偏回归系数βj**：在其他解释变量保持不变的情况下，Xj每变化1个单位时E(Y)的变化。（多元线性回归模型基本假定：假定1加上各X间互不相关i.e.无多重共线性问题）

 or；

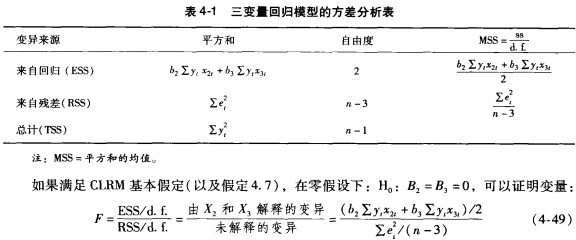
样本容量问题：样本最小容量必须不少于模型中解释变量个数#含常数项，即n≥k+1；模型的良好性质只有在大样本下才能得到理论上证明#n≥30或n≥3(k+1)

**调整的可决系数**：将残差平方和与总离差平方和分别除以各自自由度，以剔除变量个数对拟合优度影响。

方程显著性检验i.e. **F检验**：对模型中被解释变量与解释变量间的线性关系在总体上是否显著成立作出推断，即检验模型Yi=β0+β1X1i+β2X2i+...+βkXki+μi i=1,2, …,n中的参数βj是否显著不为0。（先设H0和H1- 则拒绝H0）#所以检验H0:β0=β1=...=βk=0等价于检验R2=0。

变量显著性检验i.e. t检验|t|> tα/2(n-k-1)则拒绝βi=0 #置信区间

方差分析：对TSS各个组成部分进行研究（ESS由回归模型解释，RSS不能由模型解释）

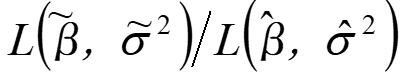
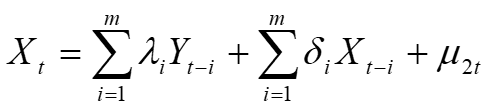


四、

**受约束回归**：模型施加约束条件后进行回归。# #m恰为约束条件个数

**邹氏参数稳定性检验**：分别以两连续时间序列作为两个样本回归，得到相应残差平方RSS1和RSS2，将两序列并为一个大样本回归，得残差平方和RSSR，若F值大于临界值则拒绝零假设，认为发生了结构变化，参数非稳定。

**邹氏预测检验**：若n2<k，则先用前一时间段n1个样本估计原模型，再用估计参数对后一时间段n2个样本预测，若误差较大则说明参数发生了变化。

非线性约束检验：最大似然比检验（两个似然函数值差异是否够大，受约束与无约束的比值#很小/接近1-差距大/小）；沃尔德检验（估计无约束模型，测量无约束回归不满足约束条件的程度）；拉格朗日乘数检验（估计受约束模型，若某一约束为真，则该约束条件对最大似然函数值影响很小，若为各约束条件为真，LM统计量服从自由度为约束条件个数的渐近卡方分布）；**格兰杰因果检验**（要求估计，X（或Y）各滞后项前参数整体为零-X（或Y）不是Y（或X）的格兰杰原因。如针对，分别做包含与不包含X滞后项的回归，记RSSU和RSSR，若大于临界值则拒绝H0，认为X是Y的格兰杰原因。#对滞后期长度很敏感，以无序列相关来选滞后期）

五、

**双对数模型/不变弹性模型**：被解释变量和解释变量都以对数形式出现，形如（双对数线性-变量间是线性的；多元对数模型即多个解释变量）

**弹性系数**：一变量变动1%所引起的另一变量变动的百分比，在双对数模型中，解释变量前的回归系数即被解释变量相对于解释变量弹性系数的估计值。（**偏弹性系数**-多元双对数模型中的偏斜率系数，度量了在其他变量不变时被解释变量对某一解释变量的偏弹性）（斜率系数-X每变动一单位，Y均值的变动量）

**线性趋势模型**：Y对t的回归，t按时间顺序度量为趋势变量，形如

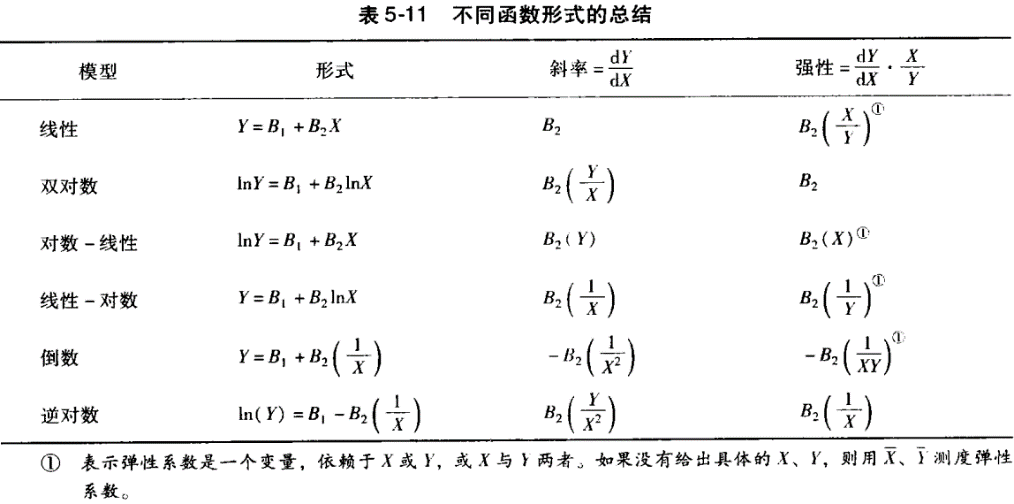
半对数模型：仅有一个变量以对数形式出现（**对数-线性模型i.e.增长模型**-被解释变量对数形式，解释变量线性形式，形如#线性-对数模型反过来）

**倒数模型**：解释变量以倒数形式出现，形如...

**多项式模型**：等式右边只有一个解释变量，但却以不同次幂出现，可看作多元回归模型（也可纳入其他解释变量及其高次幂），形如

过原点的回归：零截距模型（需要充分理论保证）

**标准化变量**：均值为0，方差为1。（**Beta系数**-标准化解释变量的回归系数，用B\*表示，标准化回归元每增加一个标准差，被解释标准化变量将增加B\*个标准差）

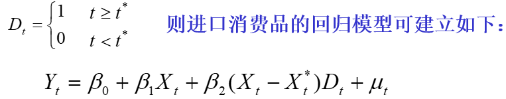


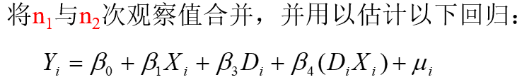
#模型的建立需要正确的理论、合适可用的数据、对各种模型统计性质的完整理解以及经验判断。不能仅强调R2（而且两个模型的因变量形式也不一定相同）

# 六、 虚拟变量

**虚拟变量**：根据定性因素属性类型构造只取0或1的人工变量。（虚拟变量模型i.e**.**方差分析模型-仅包含定性变量或虚拟变量的模型；

协方差分析模型-既包括定量解释变量又包括定性解释变量）（交互作用虚拟变量-两个虚拟变量的乘积，表示两个定性变量的联合影响）

虚拟变量引入模型方式：加/乘法方式-截距/斜率的变化；临界指标

重合/相异回归-两个回归相同/完全不同；平行/汇合回归-差异仅在截距/斜率（如β4显著则说明两时期斜率不同）

**虚拟变量设定原则**：每一定性变量所需的虚拟变量个数要比该定性变量的类别数少1，即如果有m个定性变量，只在模型中引入m-1个虚拟变量或者不要截距项（不符合这条原则则陷入虚拟变量陷阱）

**七、**

**模型设定误差**：模型中被解释变量和解释变量之间的关系设定有误

遗漏相关变量/包含无关变量/错误函数形式偏误：采用遗漏相关变量/包含无关变量/错误函数形式的模型进行估计而带来的偏误

**非相关变量**：没有具体理论表明应该把这些变量包括到模型中

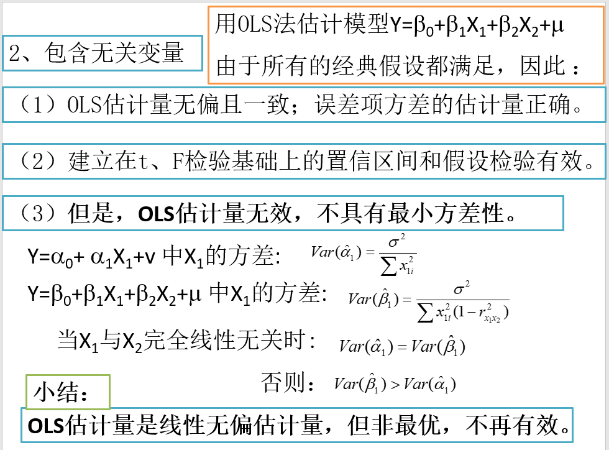
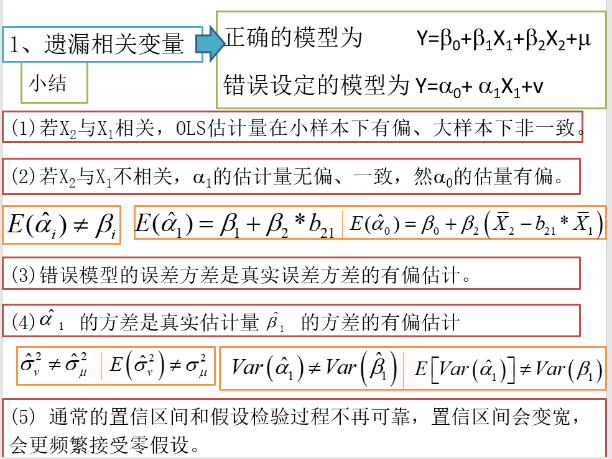
数据的度量误差：臆断的（伪造的）、外推的、内插的（使用模型推测）、围绕某个系统样式（产生于某种规则）

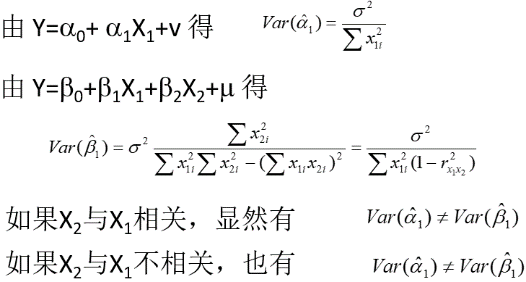
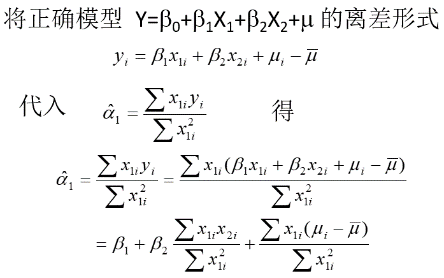
残差图示法/RESET检验：见下

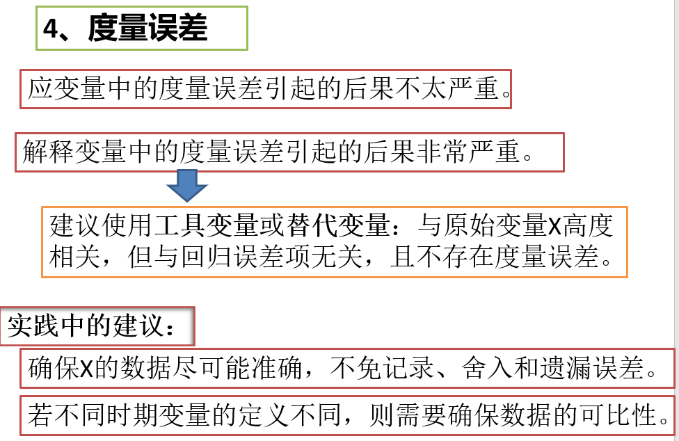
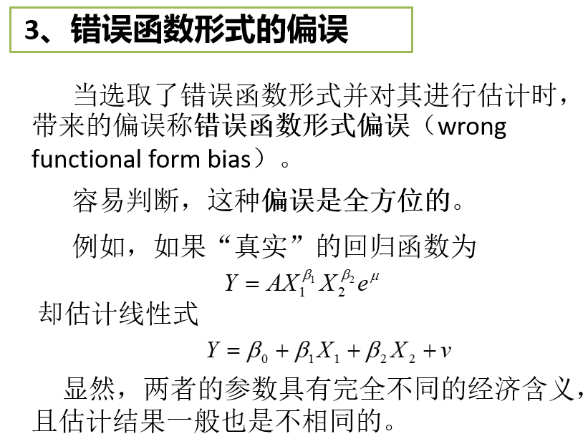
“好”的模型特性：简约性（模型是对现实的抽象，尽可能简洁）、可识别性（每个参数估计值唯一）、拟合优度（对样本数据拟合程度较好）、理论一致性（参数估计值符号与理论相符）、预测能力（具有良好的预测能力）

设定误差类型：遗漏相关变量（过低拟合）、包含不必要变量（过拟合）、错误设定函数形式、度量误差。（产生原因：对所研究问题相关理论了解不深、未关注本领域前期研究成果、研究中缺乏相关数据、测量误差）

设定偏误后果（模型估计结果与实际有偏差，其性质与程度和模型设定偏误类型密切相关）



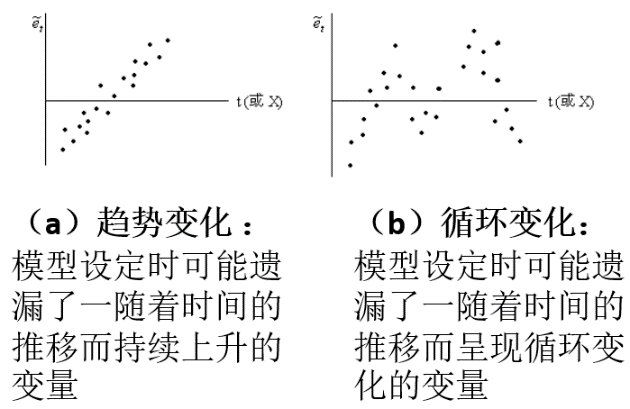
#1（1）；1（4）推导



**模型设定偏误检验**

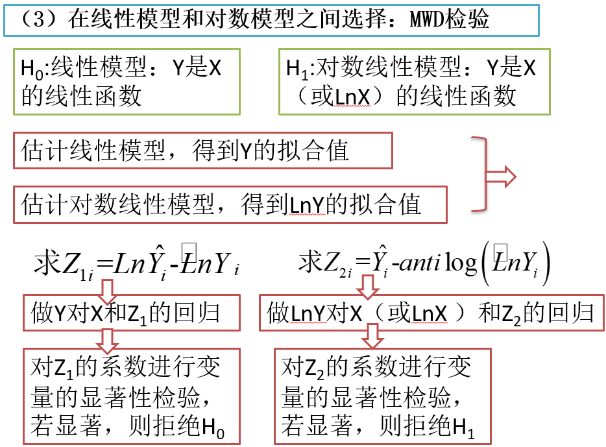
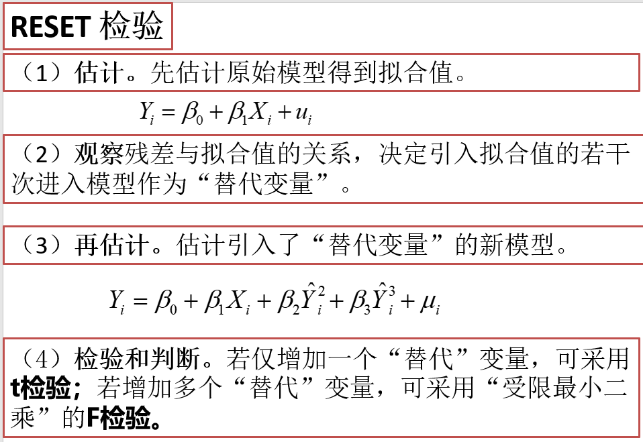
1. 检验是否含有无关变量：检验无关变量系数显著性（t检验+F检验；t/F检验：检验某1/若干个变量是否应/应同时包括在模型中）
2. 检验是否有相关变量的遗漏或函数形式设定偏误

残差图示法：做出残差序列与t或某解释变量X的散点图，考察是否有规律变动，以判断是否遗漏重要解释变量或选取错误函数形式



**RESET检验**：由于不知遗漏了哪个变量（否则将其引入检验显著性即可），用所设定模型中的被解释变量Y的估计值的若干次幂来充当其替代变量。（优：简单易行，可用于判断模型设定是否错误-诊断工具；缺：不能帮助我们选择正确模型）

1. 在线性模型和对数模型之间选择：**MWD检验**



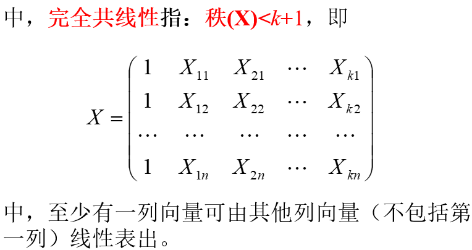
# 八、 多重共线性

**多重共线性**：某两个或多个解释变量之间出现了相关性

**完全共线性**：存在*c*1*X*1i+*c*2*X*2i+…+*c*k*X*ki=0，*i*=1,2,…,*n*，其中*c*i不全为0，则称解释变量间存在…；（不如近似共线性常见）

**近似共线性/交互相关**：存在*c*1*X*1i+*c*2*X*2i+…+*c*k*X*ki+*v*i =0， *i*=1,2,…,*n*，其中*c*i不全为0，*v*i为随机误差项，则称…

简单相关系数法/综合统计检验法/判定系数检验法/逐步回归法：见下



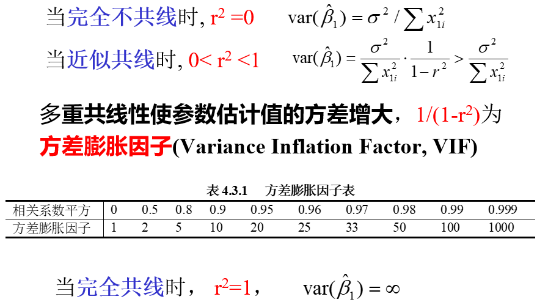
多共产生原因（实际经济问题）：经济变量相关的**共同趋势**（时间序列样本-繁荣增长衰退下降；横截面-生产函数中资本投入和劳力投入高度相关，大企业二者都大，小企业小）；**滞后**变量引入（用于反映真实经济关系）；样本资料**限制**（完全符合理论模型要求的数据难收集，特定样本可能存在多共；时间序列数据样本-简单线性模型往往存在多共，横截面-好点）

多共后果：OLS估计量方差和标准误较大；置信区间变宽；R2较高但t值不显著；OLS估计量及其标准误不稳定；回归系数符号有误（经济含义不合理），难以评估各个解释变量对RSS贡献。

#不完全共线性不违背任何基本假设，OLS估计量保持BLUE性质；问题在于OLS是最好的但非最完美的，统计推断上无法给出真正有用信息

1. 检验多共是否存在

2个解释变量-简单相关系数法（X1与X2相关系数绝对值|r|接近1说明多共强）；

多个解释变量-综合统计检验法（OLS法下R2和F值较大但t值小，说明存在共线性s.t.解释变量对Y独立作用不能分辨）；方差膨胀因子

1. 判明存在多共范围

**判定系数检验法**（使模型中每个解释变量以其余解释变量为解释变量进行回归，若某一回归的R^2或F值较大，说明其与其他X间存在共线性or排除某一解释变量Xj，若新模型拟合优度与包含Xj时十分接近则说明Xj与其他有共线性）；

**逐步回归法**（以Y为被解释变量，逐个引入解释变量来回归，若R^2变化显著则说明新引入变量为独立解释变量，否则其与其他有多共）

多共补救方法：

排除引起共线性的变量；

差分法（增量线性关系弱于总量）；

减小参数估计量方差（增加样本量；岭回归-以引入偏误为代价减小方差）；

重新考虑模型/参数先验信息/变量变换

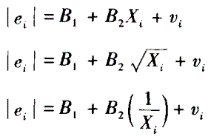
# 九、异方差

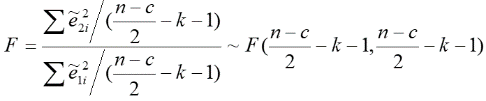
**同/异方差**：总体回归函数中，μi具有相同的方差/方差随观察值不同而变化。（异方差(Xi)类型-单调递增/减/复杂型：随X增大而增大/减小/变化呈复杂形式）

异方差后果：OLS估计量仍线性无偏，但非有效i.e.非BLUE（大小样本亦如此）；OLS估计量的方差有偏（方向无法先验确定），因为不再是真实的无偏估计量，因此建立在t分布和F分布之上的置信区间和假设检验不可靠，模型预测失效

异方差检验（即检验μi方差同解释变量观测值间的相关性及其形式）：

图示法；

**帕克/Glejzer检验**（，选择关于X的不同的函数形式，对方程进行估计并进行显著性检验，若存在一种形式使B2显著则有异方差；帕克-；Glejzer-）；

**G-Q检验**（将n对样本按Xi大小排队，去掉序列中间c个观察值，对子样做回归得，若F大于临界值则拒绝同方差假设）；

**White检验**（先对原模型做OLS回归得残差序列，再做辅助回归，同方差假设下有，h为上式解释变量数#解释变量太多使自由度减少时可去掉交叉项）

**加权最小二乘法**: 对原模型加权，使之变成一个新的不存在异方差性的模型，然后采用OLS估计其参数。

异方差修正：σ2已知-加权最小二乘法,作变换, 同方差；σ2未知-与成比例，则等式左右两边同时除以，同方差。

方差稳定变换：方差齐性的一种变换，用于消除异方差。

# 十、 自相关

**序列相关性**：对于不同样本点，随机误差项间存在某种相关性，即E()≠0,i≠j。按时间或空间排列的观察值之间的相关关系。自相关通常与时间序列数据有关（截面-空间相关）

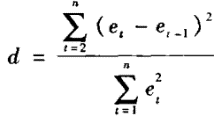
序列相关性产生原因：经济变量固有惯性；模型设定偏误；蛛网现象；数据处理(e.g.从月度数据平滑出季度)

序列相关性后果：同异方差+低估真实+计算的R2不能测度真实R2

序列相关性检验：

图示法（时序图）；

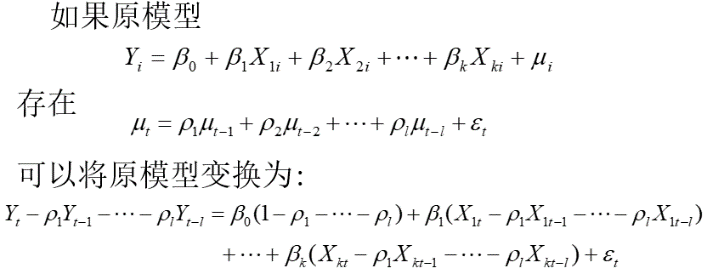
回归检验法（等等，若存在一种函数形式使方程显著成立，则有序列相关性）；

**DW检验**（残差递差平方和与残差平方和之比#注意其样本容量为n-1，，跟由n和k决定的临界值dL和dU比较）（假定条件：解释变量X非随机；解释变量中不含滞后因变量；回归模型有截距项；μi为一阶自回归形式μi=ρμi-1+εI，-1≤ρ≤1#ρ为自相关系数，该式称为马尔可夫一阶自回归过程，记为AR(1)过程）；拉格朗日乘数检验/GB检验（对方程做辅助回归，零假设为。该方程的满足,给定，查找临界值，若大于临界值则拒绝零假设）

序列相关补救：

**广义最小二乘法GLS**：对于模型Y=Xβ+ μ有Cov(μ，μ’)=σ2，对称正定阵有=DD’，变换原模型D-1Y=D-1X β +D-1μ，有同方差性和随机误差项相互独立性。#GLS估计量

**广义差分法**：将原模型变换为满足OLS法的差分模型再进行OLS估计。

（即广义差分模型）

随机误差项相关系数估计：**科克伦-奥科特迭代法**（以一元线性模型为例，先用OLS法得到μ的近似估计值，再用OLS法估计*μ*i=*ρ*1μi-1+*ρ*2*μ*i-2+…*ρ*L*μ*i-L+*ε*i，得...,，将其代入广义差分模型作OLS估计得，，再将它代入原模型得到μi新的近似估计值，再次估计*μ*i=*ρ*1μi-1+*ρ*2*μ*i-2+…*ρ*L*μ*i-L+*ε*i得...,）；**durbin两步法**（先变换差分模型为作OLS估计得...,，将其代入差分模型得，，于是）

普瑞斯-文斯顿变换：对有自相关的第一个观察值所作的变换，。

**虚假序列相关**：随机项的序列相关来自遗漏重要解释变量或模型设定偏误。

建议混个眼熟的附录

