经济计量学

1. 经济计量分析的八个步骤：
   （1） 建立一个理论假说
   （2） 收集数据
   （3） 设定数学模型
   （4） 设立统计或经济计量模型
   （5） 估计经济计量模型参数
   （6） 检验模型的适用性：模型设定检验
   （7） 检验源自模型的假设
   （8） 利用模型进行预测

2. 不同的数据类型：
   a） 时间序列数据：按时间跨度搜集得到的数据；
   b） 截面数据：一个或多个变量在某个时点上的数据集合；
   c） 合并数据：既包括时间序列数据又包括截面数据；
   d） 面板数据：同一个横截面单位的跨期调查数据。

参考课后习题：1.1、1.2

一元线性回归模型（双变量模型）

1. 总体回归和样本回归

总体回归函数： $Y_i = B_1 + B_2X_i + \mu_i$，$u_i$为随机误差项

样本回归函数：$Y_i = b_1 + b_2X_i + e_i$，$e_i = Y_i - \widehat{Y_i}$为残差项，是$u_i$的估计量

参考课后习题：2.1、2.4

2. 普通最小二乘法（OLS）

（1） 求解过程

$$Y_i = b_1 + \ b_2 X_i + e_i$$
$$e_i = Y_i - \widehat{Y}_i$$

$$minQ = min\sum e_i^2 = \sum (Y_i - \widehat{Y}_i)^2 = \sum (Y_i - b_1 - \ b_2 X_i)^2$$

$$\begin{cases} \dfrac{\partial Q}{\partial b_1} = 0 \\ \dfrac{\partial Q}{\partial b_2} = 0 \end{cases} \longrightarrow \begin{cases} nb_1 + b_2 \sum X_i = \sum Y_i \\ b_1 \sum X_i + b_2 \sum X_i^2 = \sum X_i Y_i \end{cases}$$

$$\begin{cases} b_2 = \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ b_1 = \bar{Y}_i - b_2 \bar{X} \end{cases}$$

（2） 性质

- 回归线经过样本均值点： $\bar{Y} = \hat{b}_1 + \hat{b}_2 \bar{X}$
- 残差的均值总为零； $E(e_i) = \frac{\sum e_i}{n} = 0$
- 残差与解释变量积求和为零:$\sum(X_i e_i) = 0$, 即两个变量不相关
- 残差与估计的 Y 值的积求和，值为零，$\sum(\hat{Y}_i e_i) = 0$

参考课后习题：2.23 至 2.28

3. 古典线性回归模型的基本假定

1） 回归模型是参数线性的，并且是正确设定的；

2） X_2和X_3与扰动项u不相关；

3） 误差项均值为 0，即$E(u_i) = 0$；

4） 同方差假定，即$var(u_i) = \sigma^2$；

5） 误差项u_i和u_j无自相关，即$cov(u_i, u_j) \neq 0$，$i \neq j$；

6） X_2和X_3不存在完全共线性；

7) 为了进行假设检验，假定随机误差项u服从均值为零，(同)方差为σ^2的正态分布。

4. OLS 估计量的性质 "BLUE": Best Linear Unbiased Estimator

- b_1和b_2是线性估计量，即为随机变量 Y 的线性函数
- b_1和b_2是无偏估计量，即$E(b_1) = B_1$，$E(b_2) = B_2$
- $E(\hat{\sigma}^2) = \sigma^2$, 即误差方差的 OLS 估计量是无偏的
- b_1和b_2是有效估计量，即$var(b1)/\ var(b2)$小于$B1/\ B2$的任意一个线性无偏估计量的方差

5. 假设检验

$$H_0: B_2 = B_2^*（例如：\mathrm{B}_2^* = 0）；\ H_1: B_2 \neq B_2^*$$

5.1 置信区间法（以双边检验为例）

置信区间内的区域为H_0的接受区域，以外的区域为H_0的拒绝区域

a) 显著性水平为α，即置信水平为$1-\alpha$

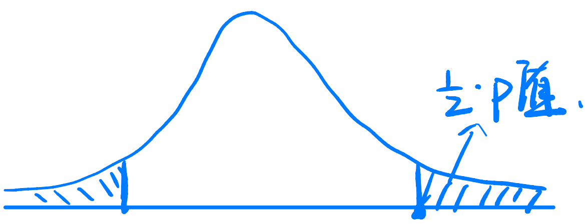
b) 查阅 t 分布表，确定 t 临界 $t\alpha_{/2}(n-2)$，使得$P\left[-t\alpha_{/2}(n-2) \leq \frac{\hat{b}_2 - B_2^*}{se(\hat{b}_2)} \leq t\alpha_{/2}(n-2)\right] = 1-\alpha$

c) 对括号中的区间重新整理可以得到b_2的预测区间为$\left[\hat{b}_2 - t\alpha_{/2}(n-2) \times se(\hat{b}_2),\ \hat{b}_2 + t\alpha_{/2}(n-2) \times se(\hat{b}_2)\right]$

d) 若b_2 落在置信区间，则接受零假设， b_2统计不显著；否则拒绝零假设，b_2统计上显著（不为 0)。

5.2 显著性检验法（以双边检验为例） P值: $p<\alpha$

- 给定显著性水平α
- 查 t 分布表得临界值$t\alpha_{/2}(n-2)$
- 计算样本统计量$t_{b_2} = \frac{\hat{b}_2 - B_2^*}{se(\hat{b}_2)} \sim t(n-2)$



- 比较t_{b_2}的绝对值大于$t\alpha_{/2}(n-2)$，则拒绝H_0；反之，接受零假设。

6. 拟合优度检验

TSS $= \sum (Y_i - \overline{Y})^2 = \sum y_i^2$，为总平方和，真实值围绕其均值的总变异；

ESS $= \sum (\hat{Y}_i - \overline{Y})^2 = \sum \hat{y}_i^2$，为解释平方和，估计值围绕其均值的变异；

RSS $= \sum (Y_i - \hat{Y})^2 = \sum e_i^2$，为残差平方和，Y 变异未被解释的部分。

$$\mathrm{TSS} = \mathrm{ESS} + \mathrm{RSS}$$

拟合优度指样本回归直线与样本观察值之间的拟合程度，用判定系数r^2的大小来表示。判定系数也称可决系数/测定系数，是指解释平方和占总平方和的比重，用来描述解释变量对被解释变量的解释程度。

判定系数：$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$

- $0 \le r^2 \le 1$
- r^2越大，拟合优度越好，解释变量对被解释变量的解释能力越强。
- 当 $r^2 = 1$ 时，Y_i的变化 100%由 X_i 作出解释，或者，回归直线的解释能力为 100%，拟合优度达到最高，所有的观察点都落在样回归直线上，是完全拟合。
- 当 $r^2 = 0$ 时，X_i 不能解释 Y_i 的任何变化，解释变量与被解释量之间没有线性关系。

相关系数：$r = \frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{(X_i-\overline{X})^2(Y_i-\overline{Y})^2}} = \frac{\sum x_i y_i}{\sqrt{x_i^2 y_i^2}} = \sqrt{r^2}$

参考课后习题：3.2、3.3、3.4、3.5、3.7、3.9、3.10、3.24

多元线性回归模型

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} \cdots + B_k X_{ki} + \mu_i$$

1. 多元线性回归模型的假定条件

（1） 回归模型是参数线性，且模型设定正确；

（2） X_j（j=1, 2⋯k) 与随机扰动项u_i不相关；

（3） 误差项均值为零 $E(u_i) = 0$；

（4） 同方差假定$Var(u_i) = \sigma^2$（常数）；

（5） 误差项u_i与u_j无自相关，$cov(u_i, u_j) = 0$

（6） 解释变量之间不存在高度线性相关关系；

（7） 为了假设检验，$u_i \sim N(0, \sigma^2)$

2. 多元回归的拟合优度

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{b_1\sum x_{1i}y_i + b_2\sum x_{2i}y_i + ... + b_k\sum x_{ki}y_i}{\sum y_i^2}$$

多元判定系数:R^2=ESS/TSS=1−RSS/TSS

存在的问题：模型中增加解释变量数量越多，R^2往往越大。但是，现实情况往往是，由增加解释变量个数引起的R^2的增大与拟合好坏无关，因此需校正。

在样本容量一定的情况下，增加解释变量将使得自由度减少，校正的思路为：将残差平方和与总平方和分别除以各自的自由度，以剔除变量个数对拟合优度的影响。

$$\overline{R^2} = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k}$$

其中，$\boldsymbol{n-k}$为残差平方和的自由度， $\boldsymbol{n-1}$为总体平方和的自由度。（注意：k为包括截距在内的解释变量个数）

性质：1）如果 k>1，则$\overline{R^2} \le R^2$。随着模型中解释变量个数的增加，校正判定系数越来越小于R^2。2）R^2总为正，$\overline{R^2}$可能为负。

	自由度
ESS	k−1

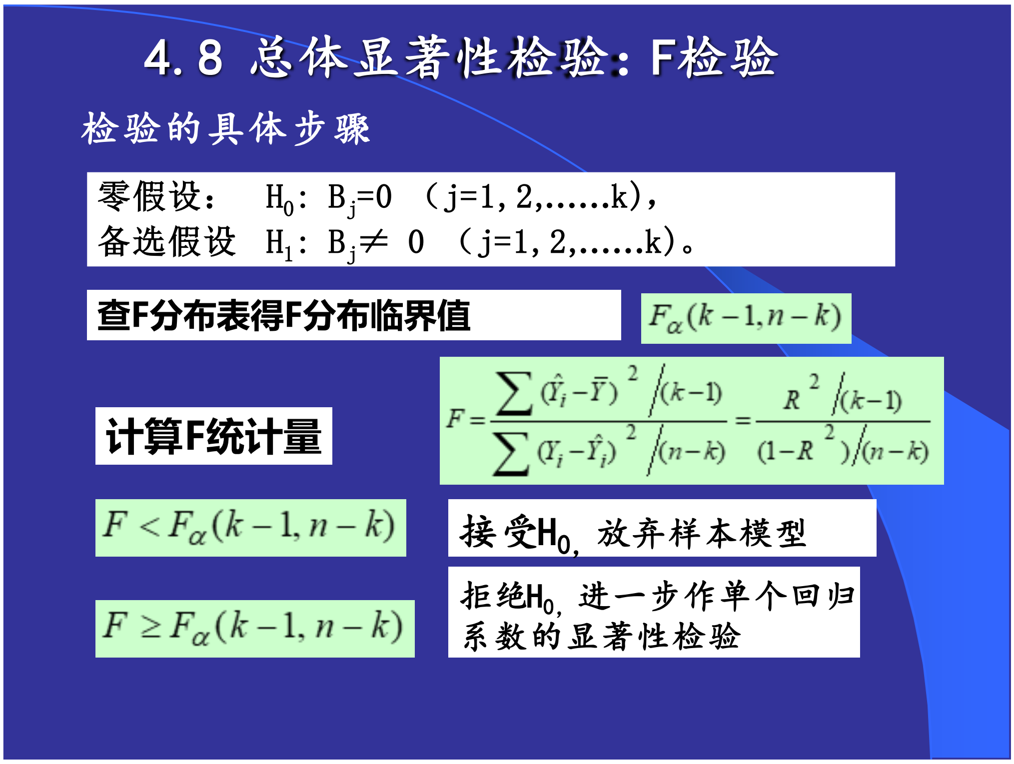
RSS	n-k
TSS	n-1

3. 假设检验

3.1 偏回归系数的假设检验

参考一元线性回归

3.2 检验联合假设：$B_2 = B_3 = 0$或者$R^2 = 0$



$$F=\frac{ESS/(k-1)}{RSS/(n-k)} \sim F(k-1,n-k)$$

F 和R^2的关系：$F=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$

两个统计量同向变动：R^2=0，F=0；$R^2 \to 1, F \to \infty$。

参考课后习题：4.3、4.8、4.9、4.10

回归模型的函数形式

- 模型选择的重点在于
    - 解释变量之间的相关性（理论基础）
    - 预期的解释变量系数的符号
    - 统计显著性
    - 考虑类似与弹性系数等衍生度量工具
    - 最后再选择恰好有较高的判决系数的模型
- 不同函数类型

函数形式	回归模型	斜率系数的含义	Y 对 X 的弹性
线性	$Y_i = B_0 + B_1X_i + \mu_i$	X 每变动一个单位，平均 Y 的绝对变动量	$B_1\frac{X}{Y}$
线性-对数	$Y_i = B_0 + B_1 lnX_i + \mu_i$	X 每增加 1%，Y 的绝对变动量	$B_1\frac{1}{Y}$
双对数	$lnY_i = B_0 + B_1 lnX_i + \mu_i$	弹性	B_1
对数-线性	$lnY_i = B_0 + B_1X_i + \mu_i$	Y 的增长率	B_1X

✓ 当被解释变量相同时，不同模型的R^2可以比较。

参考课后习题 5.1、5.5、5.17

虚拟变量回归模型

△ 一致回归、平行回归、并行回归和相异回归（课本P144）

- 虚拟变量设定的原则
    - 包含m种特征的因素，引入（$m-1$）个虚拟变量预以描述。

- 差别截距和差别斜率

$$Y_i = \beta_0 + B_0D_i + \beta_1X_i + B_1(D_iX_i) + \mu_i$$

参考课后习题：6.4、6.5、6.6、6.7、6.10、6.12、6.16

模型选择

1. 好模型具备的性质

（1） 简约性：模型应该尽可能的简单

（2） 可识别性：给定一组数据，估计的参数值必须是唯一的

（3） 拟合优度：模型中包含的解释变量应尽可能地解释应变量地变化

（4） 理论一致性：模型中的系数符号正确

（5） 预测能力：理论预测和实践相吻合

2. 设定误差的类型

（1） 遗漏相关变量

（2） 包括不必要的变量

（3） 采用错误的函数形式

（4） 度量误差

3. 遗漏相关变量的后果

对于$Y_i = B_1 + B_2X_{2i} + B_3X_{3i} + \mu_i$，遗漏相关变量$X_3$，估计$Y_i = A_1 + A_2X_i + v_i$

（1） 对有偏性和一致性的影响

a） 如果X_3和X_2相关，则估计系数a_1和a_2是有偏的，同时也是不一致的

b） 如果X_3和X_2不相关，则a_1是有偏的，但a_2是无偏且一致的

（2） 误差方差是真实值σ^2的有偏估计量

（3） 通常估计的a_2的方差是真实估计量b_2方差的有偏估计量

（4） 通常的置信区间和假设检验过程不可靠

4. 包括不相关变量的后果

OLS 估计量是无偏且一致的，是无效的（因为方差会变大），不是 BLUE

σ^2的估计量是正确的

t检验和 F 检验仍然有效

△MWD 检验 & RESET

(课本P175 – P178)

参考课后习题：7.3、7.4、7.5、7.7、7.10、7.11、7.18

多重共线性

1. 完全多重共线性

   会导致无法估计单个回归系数及其标准误，也不能根据样本进行统计推断（假设检验）

2. （不完全）多重共线性的理论后果

   - 近似共线性的情况下，并没有违背古典线性回归模型的假设，OLS 估计量仍然是无偏的。
   - 近似共线性并未破坏 OLS 估计量的最小方差性。
   - 即使在总体回归方程中变量X之间不是线性相关的，但在某个样本中，X变量之间可能线性相关（多重共线性是一个样本特征）。

3. （不完全）多重共线性的实际后果

   - OLS 估计量的方差和标准误较大，精确度下降。
   - 由于标准误变大，置信区间变宽。
   - 由于标准误变大，t值变小，t值不显著。
   - R^2值较高，但t值并不都是统计显著的。
   - OLS 估计量及其标准误对数据的微小变化非常敏感，很不稳定。
   - 回归系数符号有误。
   - 难以评估各个解释变量对回归平方和（ESS）或者 R^2的贡献。当两个变量高度共线性时，一个变量变化，另一个变量也随之自动变化，无法明确区分每一个变量的贡献。

4. （不完全）多重共线性的诊断

   - R^2较高但 t 值显著的不多。（典型特征）具体而言，通过观察回归结果，如果存在：1）估计值的大小和符号与经济理论相违背；2）若R^2,F 值均很大,而各 t 值均偏小,则认为存在多重共线性。

- 解释变量两两高度相关（局限性：变量两两相关系数较低时，仍可能存在共线性）。
- 偏~~相~~关系数 r 很高。(不要求)
- 做从属回归或者辅助回归。诊断的思想：利用某个解释变量与其余的解释变量进行回归，如果判定系数 R^2 很大，F 检验显著，则 X 可用其他解释变量的线性组合表示，即 X 与其他解释变量存在多重共线，应将 X 从解释变量中排除。

多重共线性诊断的经验法则

- 4. 做辅助回归。
  - 考虑Y对X_2- X_6的解释变量回归。如果回归结果表明存在多重共线性，但就需要找出那些变量可能是其他变量的线性或者近似线性组合。具体步骤如下：
    - （1）作X_2对其他剩余解释变量的回归，求样本判定系数
    - （2）作X_3对其他剩余解释变量的回归，求样本判定系数
    - （3）对模型中其余解释变量继续以上步骤。
  - 如果解释变量不是其他变量的线性组合，回归的判定系数不会显著不为零。
  - 原假设为：某个样本判定系数显著为零。拒绝原假设表示存在共线性。

（4）分别对样本判决系数R_i^2进行检验，检验方法如下：H_0: $R_i^2=0$，即X_i与其余解释变量并**不存在**共线性

（A）给定显著性水平α；

（B）查分布表得理论分布临界点 $F_\alpha(k-1,n-k)$

（C）计算样本统计量：
$$F=\frac{R_i^2/(k-1)}{(1-R_i^2)/(n-k)}$$

（D）比较判断：

如果 $F<F_\alpha(k-1,n-k)$ ，接受H_0

如果 $F\ge F_\alpha(k-1,n-k)$ ，拒绝H_0

- 方差膨胀因子（VIF）检测法：将存在多重共线时回归系数估计量的方差与无多重共线时回归系数估计量的方差对比而得出的比值系数。计

算公式为：$VIF\ (\beta_i)=\frac{1}{1-R_i^2}$。$R_i^2$是$X_i$对其余解释变量进行辅助回归的样本判定系数。如果$X_i$与其余解释变量都不相关,则$R_i=0$，那么其方差膨胀因子等于1；如果$X_i$与其余解释变量存在一定程度的相关关系，那么其方差膨胀因子大于1。经验认为方差膨胀因子大于5，多重共线性的程度就很严重了。

5. （不完全）多重共线性的补救措施
   - 从模型中删掉一个变量
   - 获取额外的数据或新的样本
   - 重新考虑模型：差分法
   - 参数的先验信息
   - 变量变换
   - 其他补救措施

参考课后习题：8.8、8.9、8.10、8.20、8.27

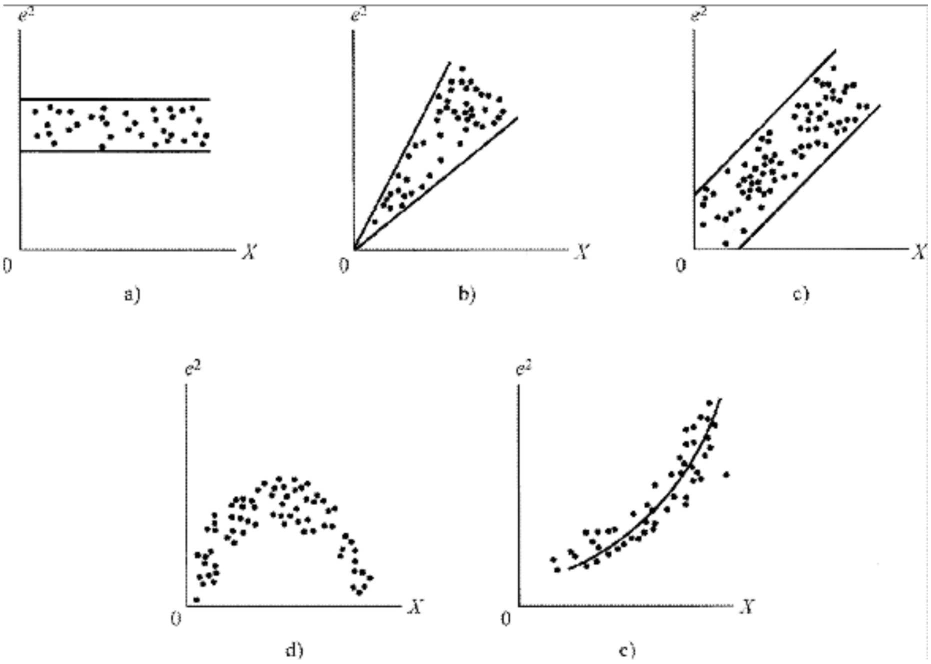
异方差

当回归模型中随机误差项的方差不是常数,即$Var(\mu_i)=\sigma_i^2\neq$ 常数时,称随机误差项的方差为非齐次性或为异方差性。

1. 异方差的后果
   - OLS 估计量仍然是线性的；
   - OLS 估计量也是无偏的，即 $E(b_1)=B_1$，$E(b_2)=B_2$；
   - OLS 估计量不再有最小方差性，即不再是有效的；
   - 估计量的方差通常是有偏的。
   - 偏差的产生是因为 $\hat{\sigma}^2=\Sigma e_i^2/(n-2)$（以双变量模型为例）不再是真实$\sigma^2$的无偏估计量。
   - 建立在 t 分布和 F 分布之上的置信区间和假设检验是不可靠的。

2. 判断是否存在异方差

➢ 残差的图形检验



➢ 帕克检验（Park test）

a）拟合回归方程，计算残差e_i。

b）计算残差平方e_i^2。

c）用原始模型中的一个解释变量分别做$lne_i^2 = B_1 + B_2 \ lnX_i + \nu_i$的回归，或者做$e_i^2$对$\widehat{Y}_i$的回归。

d）作统计检验，判断异方差是否存在。零假设：不存在异方差即：$B_2 = 0$。

e）若接受零假设，则B_1可以理解为同方差σ^2的一个给定值

➢ 格莱泽检验（Glejser test）

a） 用普通最小乘法估计模型，计算出相应的残差e_i

b） 建立残差绝对值$|e_i|$，对每个解释变量的各种形式的回归方程

$$|e_i| = B_1 + B_2 X_i + v_i$$
$$|e_i| = B_1 + B_2 \sqrt{X_i} + v_i$$
$$|e_i| = B_1 + B_2 \cdot \frac{1}{X_i} + v_i$$

c) 检验每个回归方程参数的显著性，如果参数显著地不为 0，则存在异方差，相反表示误差项满足同方差假定

➢ 怀特的一般异方差检验

a) OLS求残差e_i

b) 辅助回归：$e_i^2 = A_1 + A_2X_2 + A_3\ X_3 + A_4X_2^2 +\ A_5X_3^2 +\ A_6X_2\ X_3 + v_i$

c) 得出辅助回归方程的R^2值，计算统计量nR^2，零假设下（不存在异方差）$nR^2 \sim \chi_{k-1}$，自由度(k-1)等于辅助回归方程中解释变量的个数（k 包括截距项）

d) 做出判断：如果nR^2 很大，则拒绝零假设。否则，则不能拒绝零假设，即接受不存在异方差。

➢ ~~异方差的其他检验方法~~ （不要求）

3. 补救措施

（1） 当σ_i^2已知时：加权最小二乘法（WLS）

$$Y_i = B_1 +\ B_2X_i + \mu_i$$
$$\frac{Y_i}{\sigma_i} = B_1(\frac{1}{\sigma_i}) +\ B_2(\frac{X_i}{\sigma_i}) + \frac{\mu_i}{\sigma_i}$$

（2） 当σ_i^2未知时：

a）情形 1：误差方差与σ_i^2成比例，$E(u_i^2) = \sigma^2 X_i$

平方根变换：$\frac{Y_i}{\sqrt{X_i}} = B_1(\frac{1}{\sqrt{X_i}}) +\ B_2(\frac{X_i}{\sqrt{X_i}}) + \frac{\mu_i}{\sqrt{X_i}}$

b）情形 2：误差方差与X_i^2成比例，$E(u_i^2) = \sigma^2 {X_i}^2$

$$\frac{Y_i}{X_i} = B_1(\frac{1}{X_i}) +\ B_2(\frac{X_i}{X_i}) + \frac{\mu_i}{X_i}$$

（3） 重新设定模型

例如，将线性模型转换为对数模型$lnY_i = B_1 + B_2 lnX_i + \mu_i$

4. 怀特异方差校正后的标准误和 t 统计量

   没有改变回归系数的值，改变了标准误

参考课后习题：9.1、9.2、9.4、9.5、9.6、9.7、9.8、9.10、9.14、9.16、9.19

自相关

当回归模型中随机误差项u_i与其滞后项存在着相关关系，即$Cov(\mu_i, \mu_j) \neq 0\ (i \neq j)$时，称随机误差项存在自相关或者序列相关。也可以表示为：$E(\mu_i \mu_j) \neq 0\ (i \neq j)$。

1. 原因
   - 惯性
   - 模型设定误差
   - 蛛网现象
   - 数据处理

2. 后果
   - 最小二乘估计量仍然是线性的和无偏的。
   - 最小二乘估计量不是有效的，不是 BLUE。
   - OLS 估计量的方差是有偏的。
   - 通常所用的 t 检验和 F 检验是不可靠的。
   - 计算得到的误差方差，$\hat{\sigma}^2 = RSS/d.f.$（残差平方和/自由度），是真实$\sigma^2$的有偏估计量，很可能低估了真实的$\sigma^2$。
   - 通常计算的R^2不能测度真实的R^2。
   - 通常计算的预测方差和标准误也是无效的。

3. 自相关的诊断

- 图形法
- 德宾-沃森 d 统计量
  - ➢ 德宾-沃森 d 统计量的构造：残差递差（一阶差分）的平方和与残差平方和的比值（注意：由于残差递差时失去一个观察值，样本容量为 n-1）

$$d=\frac{\sum_{i=2}^{n}(e_i-e_{i-1})^2}{\sum_{i=1}^{n}e_i^2}$$

  - ➢ 德宾-沃森 d 检验的基本假设
    - ✓ 回归模型包括截距项。
    - ✓ 解释变量 X 为非随机变量。
    - ✓ 随机扰动项u_i产生的机制为：$u_t=\rho u_{t-1}+v_t$ $(-1\leq\rho\leq 1)$，（马尔科夫一阶自回归过程 AR（1）过程），其中ρ为自相关系数。
    - ✓ 回归模型的解释变量不包含被解释变量的滞后值。
  - ➢ 自相关系数ρ与德宾-沃森 d 检验

（3）自相关系数ρ 的估计值：

$$\hat{\rho}=\frac{\sum_{i=2}^{n}e_ie_{i-1}}{\sum_{i=1}^{n}e_i^2}$$

当样本容量n充分大时，有 $d\approx 2(1-\hat{\rho})$

（4）德宾-沃森d检验的基本原理

如果计算的d值接近于零，则表明存在着正的自相关；如果接近于4，则表明存在负的自相关；d值越接近于2，则说明越倾向于无自相关。

➢ 德宾-沃森 d 检验的步骤

a） 进行 OLS 回归获得e_i

b） 计算 d 值

c） 根据样本容量和解释变量的个数，从 D-W 表中查找d_L和d_U

d） 根据规则进行判断

表 10-3 德宾－沃森 d 检验：判定规则

零假设	判断	如果
无正自相关	拒绝	$0 < d < d_L$
无正自相关	无法判断	$d_L \leq d \leq d_U$
无负自相关	拒绝	$4 - d_L < d < 4$
无负自相关	无法判断	$4 - d_U \leq d \leq 4 - d_L$
无正或负自相关	接受	$d_U < d < 4 - d_U$

- ~~游程检验~~ （不要求）

4. 自相关的补救

4.1 广义最小二乘法

对回归模型$Y_i = B_1 + B_2 X_i + \mu_i$，假设误差项$\mu_i$自相关产生的机制为$u_t = \rho u_{t-1} + v_t$（$-1 \leq \rho \leq 1$），$v_t$是符合经典线性回归假设的随机误差项。

$$\boxed{Y_i = B_1 + B_2 X_i + \mu_i}$$

$$\Downarrow$$

$$\boxed{\rho Y_{i-1} = \rho B_1 + B_2 \rho X_{i-1} + \mu_{i-1}}$$

$$\Downarrow$$

$$\boxed{\begin{aligned}&(Y_i - \rho Y_{i-1}) = B_1(1-\rho) + B_2(X_i - \rho X_{i-1})\\&+(\mu_i - \rho\mu_{i-1})\end{aligned}}$$

$$\Downarrow$$

$$Y_i^* = B_1^* + B_2^* X_i^* + v_i$$

4.2 普瑞斯-文斯顿变换

为了避免丢失第一个样本观察值，可以对Y和X的第一个观察值做如下变换。

$$Y_1^* = \sqrt{1-\rho^2} \times Y_1$$

$$X_1^* = \sqrt{1-\rho^2} \times X_1$$

5. 如何估计ρ

5.1 $\rho = 1$：一阶差分法

误差项之间完全正相关，在该假设下，广义差分方程变为一阶差分方程

$$Y_i = B_1 + B_2 X_i + \mu_i$$

$$\Downarrow$$

$$Y_{i-1} = B_1 + B_2 X_{i-1} + \mu_{i-1}$$

$$\Downarrow$$

$$\Delta Y_i = B_2 \ \Delta X_i + v_i$$

5.2 从德宾-沃森 d 统计量中估计ρ

$$d \approx 2(1-\hat{\rho}) \qquad \hat{\rho} \approx 1 - \frac{d}{2}$$

5.3 从 OLS 残差e_i中估计ρ

$$u_t = \rho u_{t-1} + v_t$$

$$e_t = \hat{\rho} e_{t-1} + v_t$$

参考课后习题：10.5、10.6、10.8、10.13、10.14、10.15