

计量经济学

第八章

虚拟变量回归

问题的一般性描述

在前面各章的分析中，被解释变量主要是受可以直接度量的定量因素的影响，如收入、产出、商品需求量、价格、成本、资金、人数等。但现实经济生活中，影响被解释变量变动的因素，除了可以直接观测数据的定量变量外，可能还包括一些本质上为定性因素的影响，例如性别、种族、职业、季节、文化程度、战争、自然灾害、政府经济政策的变动等。

在实际的经济分析中，这些定性因素有时具有不可忽视的重要作用。例如，研究居民收入水平时，职业、性别、文化程度、就业的地域等因素，通常是值得考虑的影响因素。

因此，在计量经济学的建模中有必要将定量因素和定性因素同时纳入回归模型之内。

第八章 虚拟变量回归

本章主要讨论:

- 虚拟变量
- 虚拟解释变量的回归
- 虚拟被解释变量的回归 (选讲, 不包括)

第一节 虚拟变量

本节基本内容:

- 基本概念
- 虚拟变量设置规则

一、基本概念

定量因素：可直接测度、数值性的因素。

定性因素：属性因素，表征某种属性存在与否的非数值性的因素。

基本思想：

直接在回归模型中加入定性因素存在诸多的困难（那些困难？），是否可将这些定性因素进行量化，以达到定性因素能与定量因素有着相同作用之目的。

虚拟变量的定义

计量经济学中，将取值为0和1的人工变量称为虚拟变量。虚拟变量也称：哑元变量、定性变量等等。通常用字母D或DUM加以表示（英文中虚拟或者哑元Dummy的缩写）。

对定性变量的量化可采用虚拟变量的方式实现。

二、虚拟变量设置规则

虚拟变量的设置规则涉及三个方面：

- 1.“0”和“1”选取原则
- 2.属性（状态、水平）因素与设置虚拟变量数量的关系
- 3.虚拟变量在回归分析中的角色以及作用等方面的问题

“0” 和 “1” 选取原则

虚拟变量取“1”或“0”的原则，应从分析问题的目的出发予以界定。

从理论上讲，虚拟变量取“0”值通常代表比较的基础类型；而虚拟变量取“1”值通常代表被比较的类型。

- “0”代表基期（比较的基础，参照物）；
- “1”代表报告期（被比较的效应）。

例如，比较收入时考察性别的作用。当研究男性收入是否高于女性时，是将女性作为比较的基础（参照物），故有男性为“1”，女性为“0”。

例1

$$(1) \quad D = \begin{cases} 1 & \text{男} \\ 0 & \text{女} \end{cases}$$

$$(2) \quad D = \begin{cases} 1 & \text{改革开放以后} \\ 0 & \text{改革开放以前} \end{cases}$$

$$(3) \quad D_1 = \begin{cases} 1 & \text{天气阴} \\ 0 & \text{其他} \end{cases}$$

$$(4) \quad D_2 = \begin{cases} 1 & \text{天气雨} \\ 0 & \text{其他} \end{cases}$$

问题：

为何只选0、1，选2、3、4行吗？为什么？

属性的状态（水平）数与虚拟变量数量的关系

定性因素的属性既可能为两种状态，也可能为多种状态。例如，性别（男、女两种）、季节（4种状态），地理位置（东、中、西部），行业归属，所有制，收入的分组等。

$$\text{如: } (D_1, D_2) = \begin{cases} (1, 0) & \text{天气阴} \\ (0, 1) & \text{天气雨} \\ (0, 0) & \text{其 他} \end{cases}$$

虚拟变量数量的设置规则

1. 若定性因素具有 m ($m \geq 2$) 个相互排斥属性(或几个水平), 当回归模型有截距项时, 只能引入 $m-1$ 个虚拟变量;
2. 当回归模型无截距项时, 则可引入 m 个虚拟变量; 否则, 就会陷入“虚拟变量陷阱”。(为什么?)

一个例子(虚拟变量陷阱)

研究居民住房消费支出 Y_i 和居民可支配收入 X_i 之间的数量关系。回归模型的设定为： $Y_i = \alpha_0 + \beta_1 X_i + u_i$ (1)
现在要考虑城镇居民和农村居民之间的差异，如何办？
为了对“城镇居民”、“农村居民”进行区分，分析各自在住房消费支出 Y_i 上的差异，设 $D_{1i} = 1$ 为城镇；
 $D_{1i} = 0$ 为农村，则模型为

$$Y_i = \alpha_0 + \beta_1 X_i + \alpha_1 D_1 + u_i \quad (2)$$

(模型有截距，“居民属性”定性变量只有两个相互排斥的属性状态 ($m=2$)，故只设定一个虚拟变量。)

若对两个相互排斥的属性 “居民属性” ， 仍然引入 $m=2$ 个虚拟变量， 则有

$$D_{1i} = \begin{cases} 1 & \text{城镇居民} \\ 0 & \text{农村居民} \end{cases} \quad D_{2i} = \begin{cases} 1 & \text{农村居民} \\ 0 & \text{城镇居民} \end{cases}$$

则模型（1）为

$$Y_i = \alpha_0 + \beta_1 X_i + \alpha_1 D_1 + \alpha_2 D_2 + u_i \quad (3)$$

则对任一家庭都有： $D_1 + D_2 = 1$ $D_1 + D_2 - 1 = 0$ ，

即产生完全共线， 陷入了 “虚拟变量陷阱” 。

“虚拟变量陷阱” 的实质是： **完全多重共线性**。

综上所述：

1. 引入虚拟变量的个数与两个因素有关：一是定性变量的属性多少，一是有无截距项；
2. 对虚拟变量的运用要谨慎，虚拟变量的使用得当常能发挥积极的作用，但在模型中引入虚拟变量的数量要适当，引入的虚拟变量的数量过度，则可能带来负面的影响。

虚拟变量在回归模型中的角色

虚拟变量既可作为被解释变量，也可作为解释变量，分别称其为虚拟被解释变量和虚拟解释变量。虚拟被解释变量的研究是当前计量经济学研究的前沿领域，如MacFadden、Heckmen等人的微观计量经济学研究，大量涉及到虚拟被解释变量的分析。本课程只是讨论虚拟解释变量的问题

第二节 虚拟解释变量的回归

本节基本内容:

- 加法类型
- 乘法类型
- 虚拟解释变量综合应用

在计量经济学中，通常引入虚拟变量的方式分为加法方式和乘法方式两种：即

$$Y_t = \alpha_0 + \beta X_t + u_t + \alpha_1 D$$

$$Y_t = \alpha + \beta_1 X_t + u_t + \beta_2 X_t D$$

对上式可
分解为：

原模型： $Y_i = \alpha + \beta X_i + u_i$

加法方式引入 $\alpha = \alpha_0 + \alpha_1 D$

乘法方式引入 $\beta = \beta_1 + \beta_2 D$

实质：加法方式引入虚拟变量改变的是截距；
乘法方式引入虚拟变量改变的是斜率。

一、加法类型

以加法方式引入虚拟变量时，主要考虑的问题是定性因素的属性和引入虚拟变量的个数。

分为四种情形讨论：

- (1) 解释变量只有一个定性变量而无定量变量，而且定性变量为两种相互排斥的属性；
- (2) 解释变量分别为一个定性变量（两种属性）和一个定量解释变量；

-
- (3) 解释变量分别为一个定性变量（两种以上属性）和一个定量解释变量；
 - (4) 解释变量分别为两个定性变量（各自分别是两种属性）和一个定量解释变量；

思考：

四种加法方式引入虚拟变量会产生什么效应？

(1) 一个两种属性定性解释变量而无定量变量的情形

模型形式: $Y_i = f(D_i) + \mu_i \Rightarrow \alpha = \alpha_0 + \alpha_1 D_i$

例如: $Y_i = \alpha_0 + \alpha_1 D_i + \mu_i$

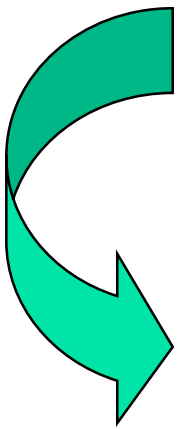
其中: $D_i = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$ (比较的基础: 农村)

那么: $E(Y_i | D_i = 1) = (\alpha_0 + \alpha_1)$

$E(Y_i | D_i = 0) = \alpha_0$

$Y_i = (\alpha_0 + \alpha_1) + \mu_i$ 城市

$Y_i = \alpha_0 + \mu_i$ 农村



(2) 一个定性解释变量（两种属性） 和一个定量解释变量的情形

模型形式 $Y_i = f(D_i, X_i) + \mu_i \Rightarrow \alpha = \alpha_0 + \alpha_1 D_i$

例如： $Y_i = \alpha_0 + \alpha_1 D_i + \beta X_i + \mu_i$

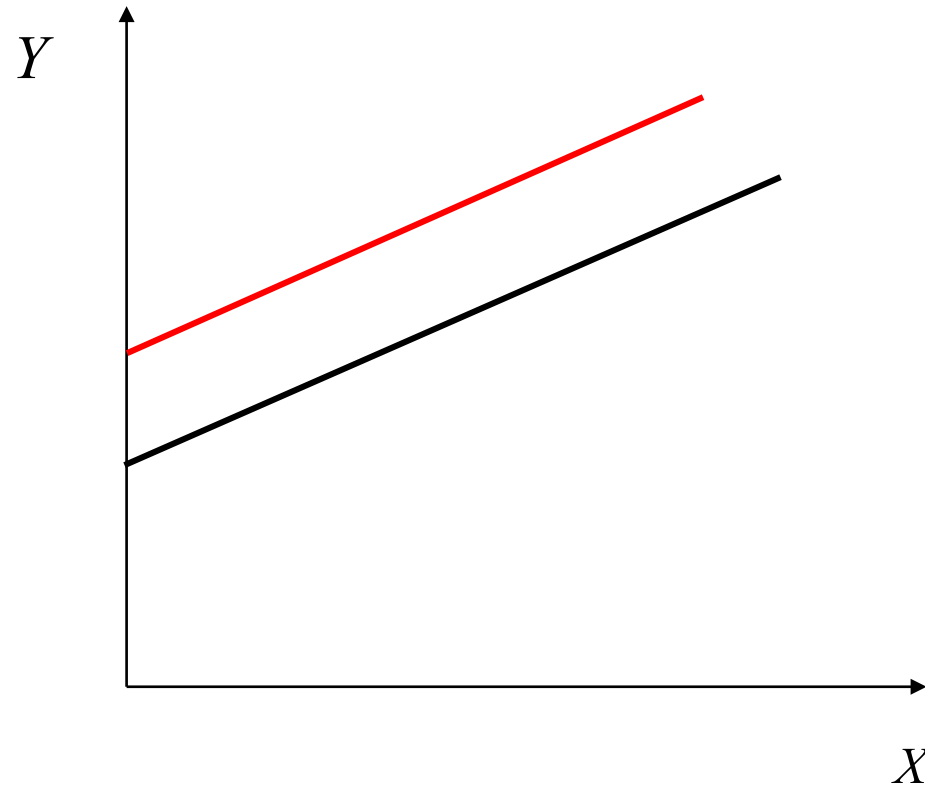
其中： Y —支出； X —收入； $D_i = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$

$$E(Y_i | X_i, D_i = 1) = (\alpha_0 + \alpha_1) + \beta X_i$$

$$E(Y_i | X_i, D_i = 0) = (\alpha_0) + \beta X_i$$

$$Y_i = (\alpha_0 + \alpha_1) + \beta X_i + \mu_i \quad \text{城市}$$

$$Y_i = \alpha_0 + \beta X_i + \mu_i \quad \text{农村}$$



共同的特征：截距发生改变（？）

(3) 一个定性解释变量（两种以上属性）和一个定量解释变量的情形

模型形式 $Y_i = f(X_i, D_1, D_2, \dots) + \mu_i$

（如：民族有56种特性；季度有4种特性）

例如：啤酒销量 Y 、人均收入 X 、季度 D ；

$$Y_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta X_i + \mu_i$$

$$\text{其中: } D_1 = \begin{cases} 1 & \text{一季度} \\ 0 & \text{其 它} \end{cases} \quad D_2 = \begin{cases} 1 & \text{二季度} \\ 0 & \text{其 它} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{三季度} \\ 0 & \text{其 它} \end{cases}$$

一季度: $E(Y_i | X_1, D_1 = 1, D_2 = D_3 = 0) = (\alpha_0 + \alpha_1) + \beta X_i$

二季度: $E(Y_i | X_1, D_2 = 1, D_1 = D_3 = 0) = (\alpha_0 + \alpha_2) + \beta X_i$

三季度: $E(Y_i | X_1, D_3 = 1, D_1 = D_2 = 0) = (\alpha_0 + \alpha_3) + \beta X_i$

四季度: $E(Y_i | X_1, D_1 = D_2 = D_3 = 0) = \alpha_0 + \beta X_i$

(基准: 四季度)

(4) 两个定性解释变量（均为两种属性）和一个定量解释变量的情形

例：分析啤酒销售量 Y 受到人均收入 X 、季节 D_1 以及居民属性 D_2 的影响。

$$Y_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \beta X_i + \mu_i$$

$$\text{其中： } D_1 = \begin{cases} 1 & \text{夏季} \\ 0 & \text{冬季} \end{cases} \quad D_2 = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$$

(比较的基础—冬季、农村)

夏季、城市居民

$$E(Y_i | X_i, D_1 = 1, D_2 = 1) = (\alpha_0 + \alpha_1 + \alpha_2) + \beta X_i$$

夏季、农村居民

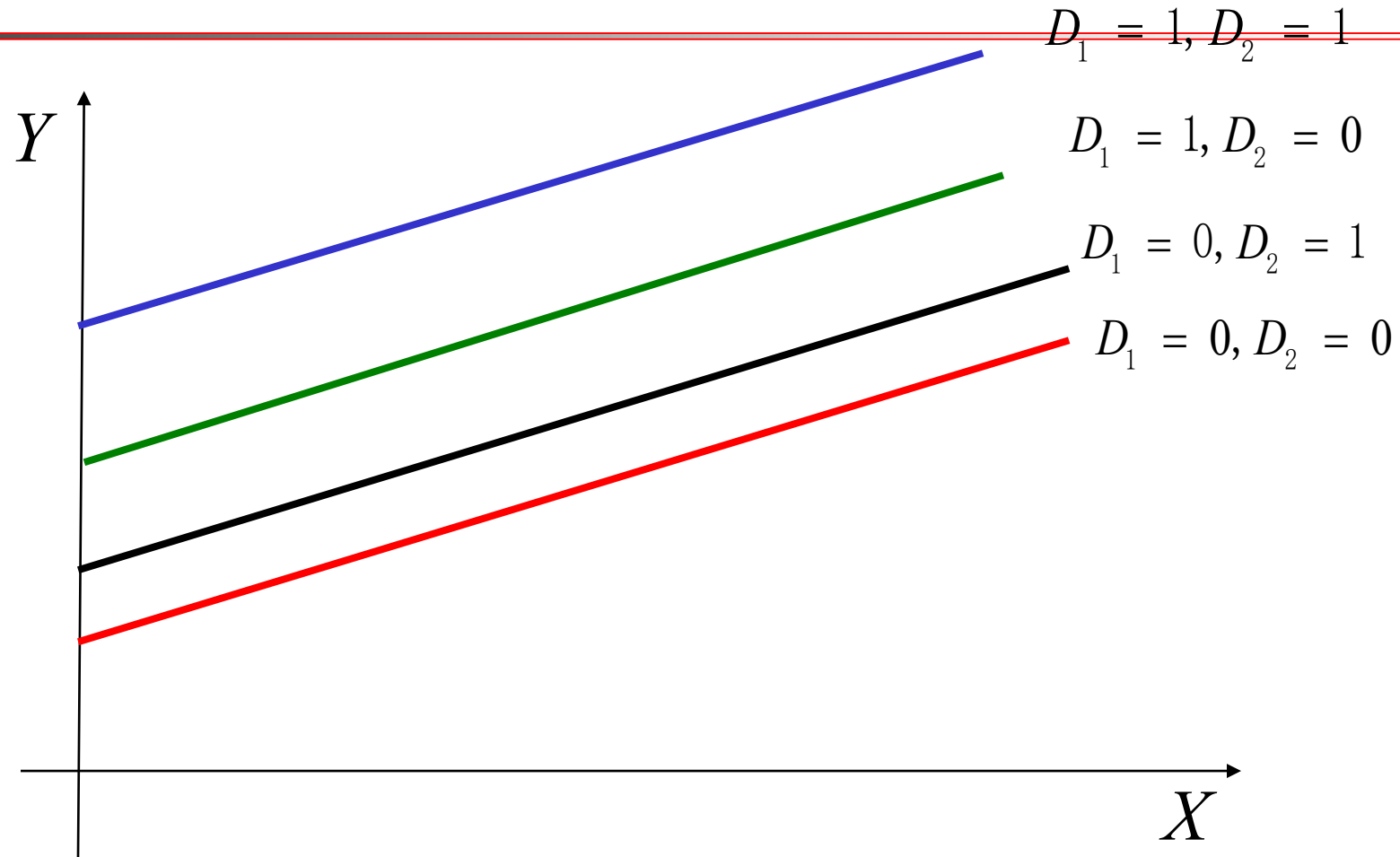
$$E(Y_i | X_i, D_1 = 1, D_2 = 0) = (\alpha_0 + \alpha_1) + \beta X_i$$

冬季、城市居民

$$E(Y_i | X_i, D_1 = 0, D_2 = 1) = (\alpha_0 + \alpha_2) + \beta X_i$$

冬季、农村居民

$$E(Y_i | X_i, D_1 = 0, D_2 = 0) = \alpha_0 + \beta X_i$$



上述图形的前提条件是什么？

运用**OLS**得到回归结果，再用**t**检验讨论因素是否对模型有影响。

加法方式引入虚拟变量的一般表达式：

$$Y_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \dots + \alpha_k D_{kt} + \beta X_t + u_t$$

基本分析方法：条件期望。

$$E(Y_t / D_{1t}, D_{2t}, \dots, D_{kt}) = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \dots + \alpha_k D_{kt} + \beta X_t$$

加法方式引入虚拟变量的主要作用为：

1. 在有定量解释变量的情形下，主要改变方程截距；
2. 在没有定量解释变量的情形下，主要用于方差分析。

二、乘法类型

基本思想

以乘法方式引入虚拟变量时，是在所设立的模型中，将虚拟解释变量与其它解释变量的乘积，作为新的解释变量出现在模型中，以达到其调整设定模型斜率系数的目的。或者将模型斜率系数表示为虚拟变量的函数，以达到相同的目的。

乘法引入方式：

- (1) 截距不变；
- (2) 截距和斜率均发生变化；

分析手段：仍然是条件期望。

(1) 截距不变的情形

模型形式:

$$Y_t = f(X_t, D_t X_t) + u_t \Rightarrow \alpha = \alpha, \beta = \beta_1 + \beta_2 D$$

例：研究消费支出 Y 受收入 X 、年份状况 D 的影响

$$Y_t = \alpha + \beta_1 X_t + \beta_2 (D_t X_t) + \mu_t$$

其中： Y - 消费支出； X - 收入； $D_t = \begin{cases} 1 & \text{反常年份} \\ 0 & \text{正常年份} \end{cases}$

反常年份 $E(Y_t | X_t, D_t = 1) = \alpha + (\beta_1 + \beta_2) X_t$

正常年份 $E(Y_t | X_t, D_t = 0) = \alpha + \beta_1 X_t$

在正常年份的基础上进行比较，（只有斜率系数发生改变）。

(2) 截距和斜率均发生变化

模型形式:

$$Y_i = f(X_t, D_t, D_t X_t) \Rightarrow \alpha = \alpha_0 + \alpha_1 D, \beta = \beta_1 + \beta_2 D$$

例，同样研究消费支出 Y 、收入 X 、年份状况 D 间的影响关系。

$$Y_t = \alpha_0 + \beta_1 X_t + \alpha_1 D_t + \beta_2 (D_t X_t) + \mu_t$$

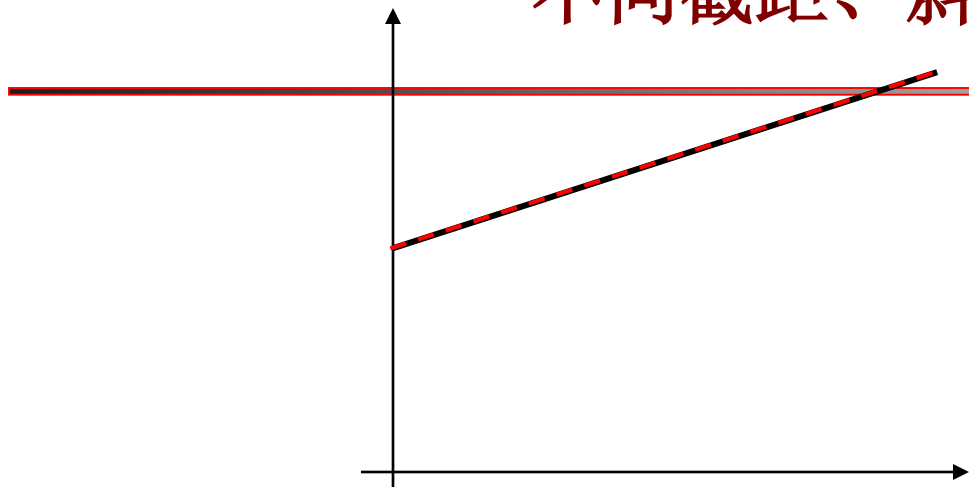
其中： Y —消费支出； X —收入； $D_t = \begin{cases} 1 & \text{反常年份} \\ 0 & \text{正常年份} \end{cases}$

$$\text{反常年份} \quad E(Y_t | X_t, D_t = 1) = (\alpha_0 + \alpha_1) + (\beta_1 + \beta_2) X_t$$

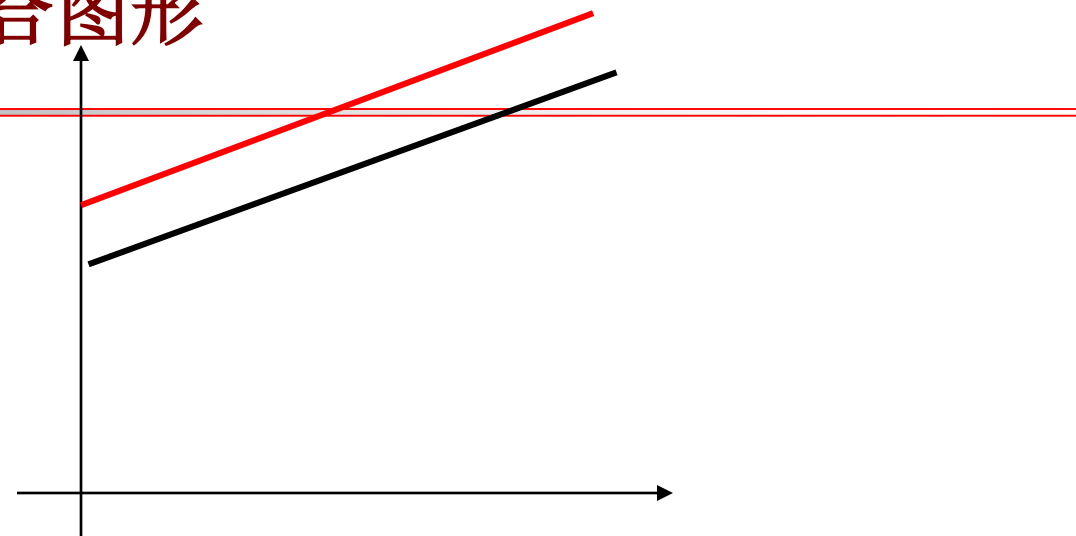
$$\text{正常年份} \quad E(Y_t | X_t, D_t = 0) = \alpha + \beta_1 X_t$$

在正常年份基础上比较，截距和斜率系数都改变，为什么？

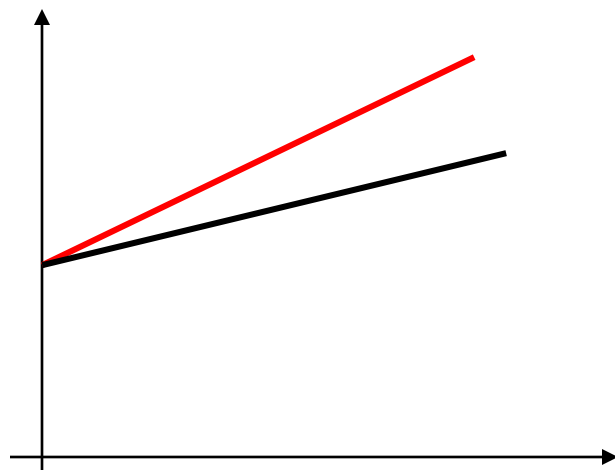
不同截距、斜率的组合图形



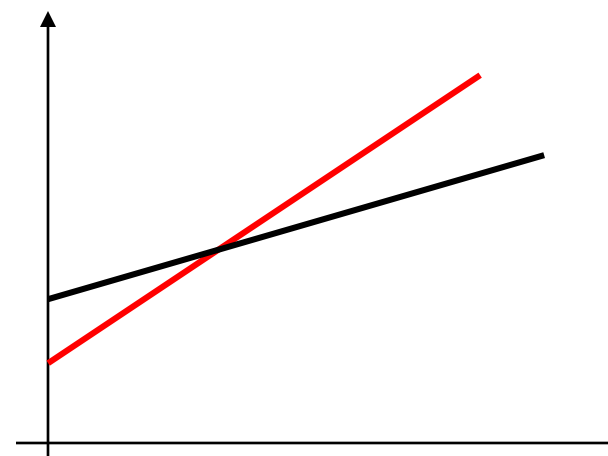
重合回归：截距斜率均相同



平行回归：截距不同斜率相同



共点回归：截距相同斜率不同



交叉（不同）回归：截距斜率均不同

三、虚拟解释变量综合应用

所谓综合应用是指将引入虚拟解释变量的加法方式、乘法方式进行综合使用。

基本分析方式仍然是条件期望分析。

本节主要讨论

- (1) 结构变化分析;
- (2) 交互效应分析;
- (3) 分段回归分析

(1) 结构变化分析

结构变化的实质是检验所设定的模型在样本期内是否为同一模型。显然，平行回归、共点回归、不同的回归三个模型均不是同一模型。

平行回归模型的假定是斜率保持不变（加法类型，包括方差分析）；

共点回归模型的假定是截距保持不变（乘法类型，又被称为协方差分析）；

不同的回归的模型的假定是截距、斜率均为变动的（加法、乘法类型的组合）。

例：比较改革开放前、后我国居民（平均）“储蓄—收入”总量关系是否发生了变化？

模型的设定形式为：

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \quad (1)$$

其中： Y_t 为储蓄总额， X_t 为收入总额。

$$D = \begin{cases} 1 & \text{改革开放后} \\ 0 & \text{改革开放前} \end{cases}$$

回归方程:

$$\text{改革开放后} \quad E(Y_t | X_t, D=1) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \quad (2)$$

$$\text{改革开放前} \quad E(Y_t | X_t, D=0) = \alpha_1 + \beta_1 X_t \quad (3)$$

显然，只要 α_2 、 β_2 不同时为零，上述模型就能刻画改革开放前后我国居民储蓄收入模型结构是否发生变化。

问题：

1. 本例中，平行、共点回归、不同的回归三模型
的经济学背景解释是什么？
2. 如何进行结构变化判断？
3. 是否可对**(2)**、**(3)**分别进行 **OLS** 估计？为什么？
4. 若分别对**(2)**、**(3)**进行 **OLS** 估计应注意什么？

(2) 交互效应分析

交互作用：

一个解释变量的边际效应有时可能要依赖于另一个解释变量。为此，Klein和Morgen(1951)提出了有关收入和财产在决定消费模式上相互作用的假设。他们认为消费的边际倾向不仅依赖于收入，而且也依赖于财产的多少——较富有的人可能会有不同的消费倾向。

为了捕获该影响，设 $C = \alpha + \beta Y + u$ 。假设边际消费倾向 β 依赖于财产 Z 。一个简单的表示方法就是 $\beta = \beta_1 + \beta_2 Z$ 。代入消费函数，有：

$$C = \alpha + \beta_1 Y + \beta_2 YZ + u$$

由于 YZ 捕获了收入和财产之间的相互作用而被称为交互作用项。

显然，刻画交互作用的方法，在变量为数量(定量)变量时，是以乘法方式引入虚拟变量的。

例：是否发展油菜籽生产与是否发展养蜂生产的差异对农副产品总收益的影响研究。

模型设定为：

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (1)$$

其中： Y_i (农副产品收益)； X_i (农副产品投入)

$$D_2 = \begin{cases} 1 & \text{发展油菜籽生产} \\ 0 & \text{其他} \end{cases}; \quad D_3 = \begin{cases} 1 & \text{发展养蜂生产} \\ 0 & \text{其他} \end{cases}$$

(1) 式中，以加法形式引入虚拟变量暗含何假设？

(1) 式以加法形式引入，暗含的假设为：
菜籽生产和养蜂生产是分别独立地影响农副产品生产总收益。但是，在发展油菜籽生产时，同时也发展养蜂生产，所取得的农副产品生产总收益，可能会高于不发展养蜂生产的情况。即在是否发展油菜籽生产与养蜂生产的虚拟变量 D_{2i} 和 D_{3i} 间，很可能存在着一定的交互作用，且这种交互影响对被解释变量农副产品生产收益会有影响。

问题：如何刻画同时发展油菜籽生产和养蜂生产的交互作用？

基本思想：在模型中引入相关的两个变量的乘积。

区别之处在于，上页定义中的交互效应是针对数量变量，而现在是定性变量，又应当如何处理？

为了反映交互效应，将（1）变为：

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{2i} D_{3i} + \beta X_i + u_i$$

同时发展油菜籽和
养蜂生产：

$$Y_i = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i + u_i$$

发展油菜籽生产：

$$Y_i = (\alpha_1 + \alpha_2) + \beta X_i + u_i$$

发展养蜂生产：

$$Y_i = (\alpha_1 + \alpha_3) + \beta X_i + u_i$$

基础类型：

$$Y_i = \alpha_1 + \beta X_i + u_i$$

如何检验交互效应是否存在？

看 $(D_{2i}D_{3i})$ 系数 α_4 对应的 t 值：

$$\text{即检验：} \begin{cases} H_0: \alpha_4 = 0 \\ H_1: \alpha_4 \neq 0 \end{cases}$$

若拒绝原假设，即交互效应对 Y 产生了影响（应该引入模型）。

(3) 分段回归分析

作用：提高模型的描述精度。

虚拟变量也可以用来代表数量因素的不同阶段。

分段线性回归就是类似情形中常见的一种。

一个例子： 研究不同时段我国居民的消费行为。

实际数据表明，1979年以前，我国居民的消费支出呈缓慢上升的趋势；从1979年开始，居民消费支出为快速上升趋势。

如何刻画我国居民在不同时段的消费行为？

基本思路： 采用乘法方式引入虚拟变量的手段。
显然，1979年是一个转折点，可考虑在这个转折点作为虚拟变量设定的依据。若设 $X^* = 1979$ ，
当 $t < X^*$ 时可引入虚拟变量。（为什么选择1979作为转折点？）

依据上述思路，有如下描述我国居民在不同时段消费行为模型：

$$Y_t = \beta_0 + \beta_1 t + \beta_2 (t - X^*) D + u_t$$

其中：

$$D = \begin{cases} 1 & t \geq X_t^* \\ 0 & t < X_t^* \end{cases} \quad (t=1955, 1956, \dots, 2004)$$

居民消费趋势方程：

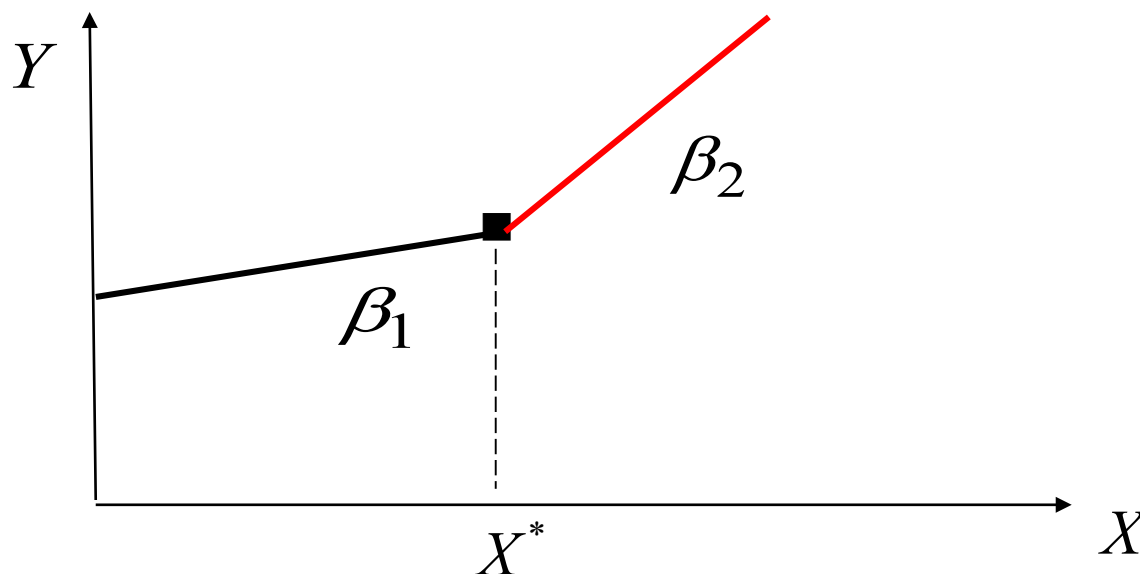
1979年以前： $Y_t = \beta_0 + \beta_1 t + u_t$

1979年以后： $Y_t = \beta_0 - \beta_2 X^* + (\beta_1 + \beta_2)t + u_t$

分析

1979年之前，回归模型的斜率为 β_1 ；

1979年之后，回归模型的斜率为 $\beta_1 + \beta_2$ ；



若统计检验表明， β_2 显著不为零，则我国居民的消费行为在1979年前后发生了明显改变。

第三节 案例分析

为了考察改革开放以来中国居民的储蓄存款与收入的关系是否已发生变化，以城乡居民人民币储蓄存款年底余额代表居民储蓄（ Y ），以国民总收入GNI代表城乡居民收入，分析居民收入对储蓄存款影响的数量关系，并建立相应的计量经济学模型。

表8.1 国民总收入与居民储蓄存款

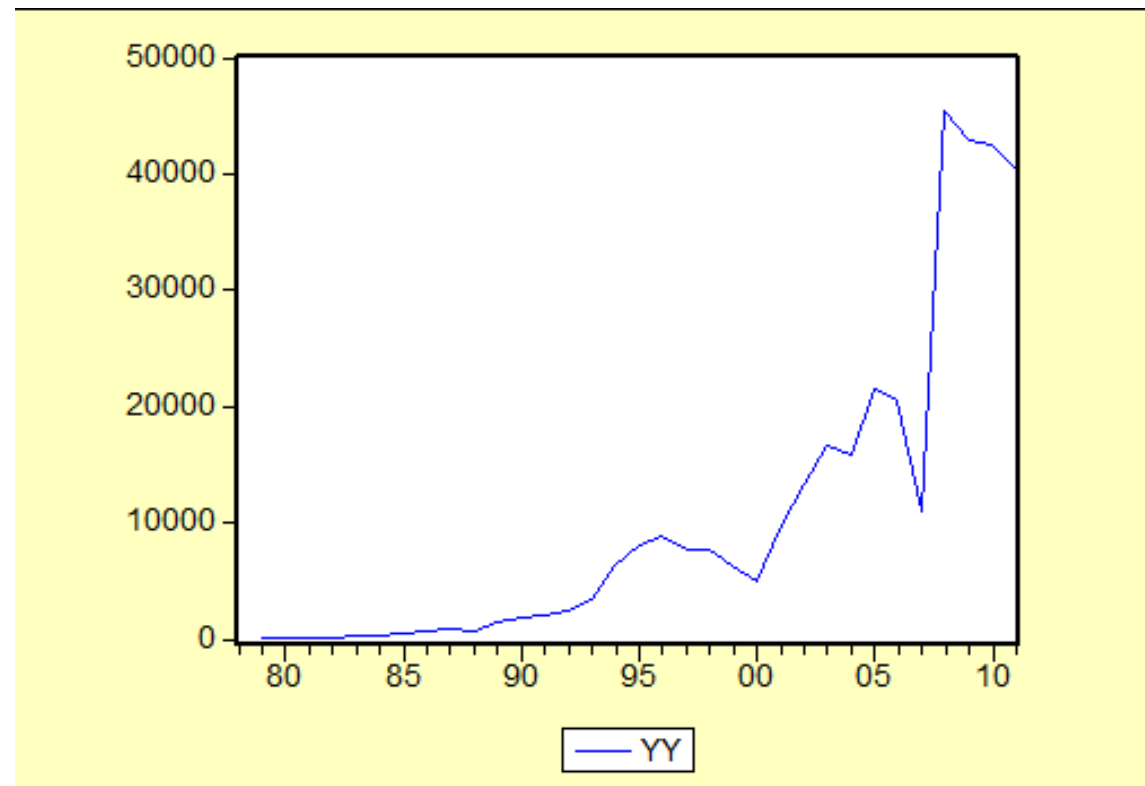
单位：亿元

年 份	国民总收入 (GNI)	城乡居民人民币储蓄 存款年底余额 (Y)	城乡居民人民币 储蓄存款增加额 (YY)	年 份	国民总收入 (GNI)	城乡居民人民 币储蓄存款年 底余额 (Y)	城乡居民人民币 储蓄存款增加额 (YY)
1978	3645.2	210.6	NA	1995	59810.5	29662.3	8143.5
1979	4062.6	281.0	70.4	1996	70142.5	38520.8	8858.5
1980	4545.6	399.5	118.5	1997	78060.8	46279.8	7759.0
1981	4889.5	532.7	124.2	1998	83024.3	53407.5	7615.4
1982	5330.5	675.4	151.7	1999	88479.2	59621.8	6253.0
1983	5985.6	892.5	217.1	2000	98000.5	64332.4	4976.7
1984	7243.8	1214.7	322.2	2001	108068.2	73762.4	9457.6
1985	9040.7	1622.6	407.9	2002	119095.7	86910.6	13233.2
1986	10274.4	2237.6	615.0	2003	135174.0	103617.3	16631.6
1987	12050.6	3073.3	835.7	2004	159586.7	119555.4	15929.4
1988	15036.8	3801.5	728.2	2005	184088.6	141051.0	21496.8
1989	17000.9	5196.4	1374.2	2006	213131.7	161587.3	20544.0
1990	18718.3	7119.8	1923.4	2007	251483.2	172534.2	10967.1
1991	21826.2	9241.6	2121.8	2008	302853.4	217885.4	45353
1992	26937.3	11759.4	2517.8	2009	340320	260771.3	42885.9
1993	35260.0	15203.5	3444.1	2010	399759.5	303302.5	42417.1
1994	48108.5	21518.8	6315.3	2011	472115	343635.9	40333.4

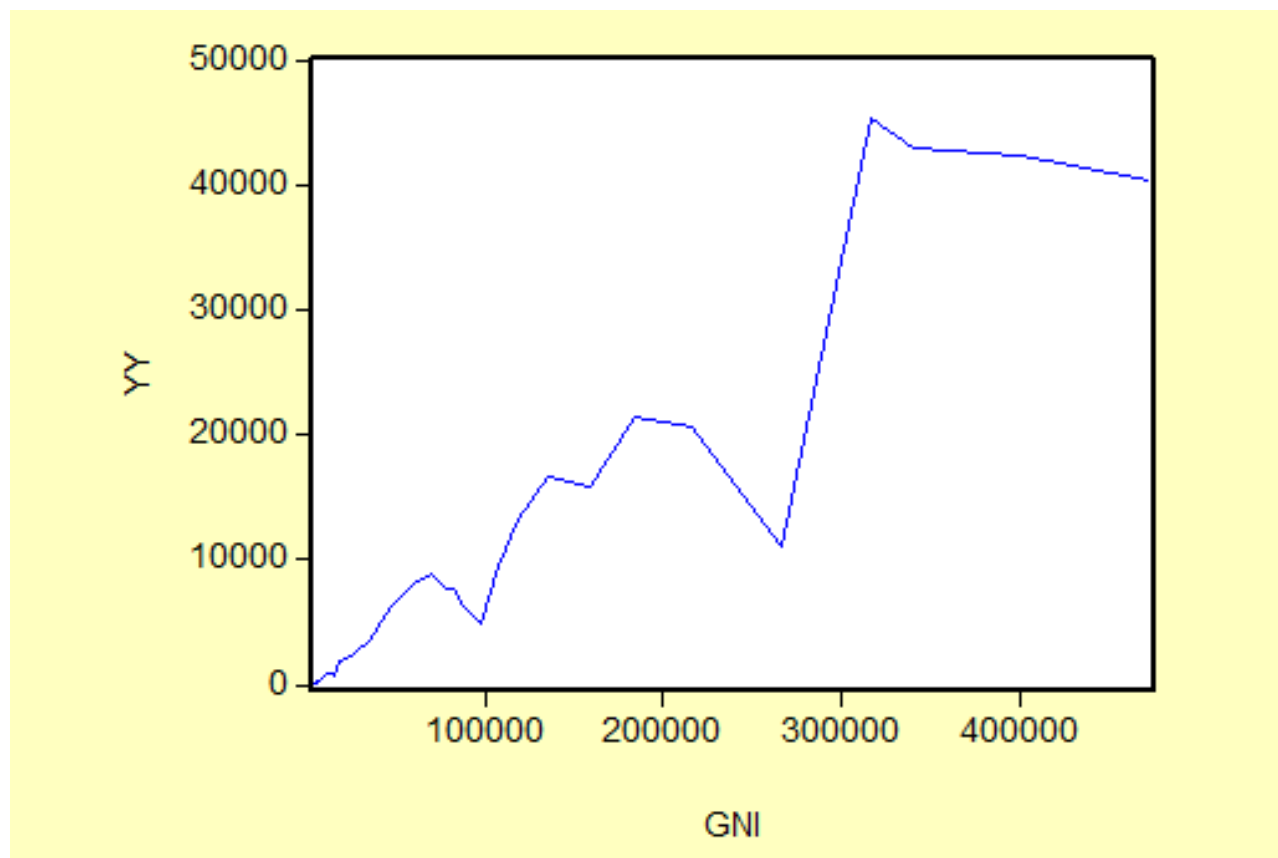
数据来源：根据《中国统计年鉴》和中经网统计数据库数据整理。

为了研究1978—2011年期间城乡居民储蓄存款随收入的变化规律是否有变化, 考证城乡居民储蓄存款、国民总收入随时间的变化情况, 如图所示:

从上图中，尚无法得到居民的储蓄行为发生明显改变的详尽信息。若取居民储蓄的增量（**YY**），并作时序图如下。从图中可以看出，城乡居民的储蓄行为表现出了明显的阶段特征：在**1996年**、**2000年**、**2005年**、**2007年**和**2009年**有五个明显的转折点。



再从城乡居民储蓄存款增量与国民总收入之间关系的散布图看（下图），也呈现出了相同的阶段性特征。



为了分析居民储蓄行为在1996年—2011年不同时期的数量关系，以1996、2000、2005、2007、2009年度的五个转折点作为依据，分别引入虚拟变量 D_1 、 D_2 、 D_3 、 D_4 、 D_5 ，这五个年度所对应的GNI分别为70142.5，98000.5，184088.6，251483.2和340320亿元。据此，我们设定了如下以加法和乘法两种方式同时引入虚拟变量的模型：

$$YY_t = \beta_1 + \beta_2 GNI_t + \beta_3 (GNI_t - 70142.5) D_{1t} + \beta_4 (GNI_t - 98000.5) D_{2t} \\ + \beta_5 (GNI_t - 184088.6) D_{3t} + \beta_6 (GNI_t - 251483.2) D_{4t} + \beta_7 (GNI_t - 340320.0) D_{5t} + u_t$$

$$D_{1t} = \begin{cases} 1 & t = 1996 \text{年以后} \\ 0 & t = 1996 \text{年及以前} \end{cases} \quad D_{2t} = \begin{cases} 1 & t = 2000 \text{年以后} \\ 0 & t = 2000 \text{年及以前} \end{cases} \quad D_{3t} = \begin{cases} 1 & t = 2005 \text{年以后} \\ 0 & t = 2005 \text{年及以前} \end{cases}$$

$$D_{4t} = \begin{cases} 1 & t = 2007 \text{年以后} \\ 0 & t = 2007 \text{年及以前} \end{cases} \quad D_{5t} = \begin{cases} 1 & t = 2009 \text{年以后} \\ 0 & t = 2009 \text{年及以前} \end{cases}$$

对上式进行回归后，有：

Dependent Variable: YY
 Method: Least Squares
 Date: 08/11/13 Time: 21:35
 Sample(adjusted): 1979 2011
 Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-698.8397	792.4822	-0.881836	0.3859
GNI	0.132747	0.025281	5.250827	0.0000
(GNI-70142.5)*D1	-0.187502	0.093237	-2.011019	0.0548
(GNI-98000.5)*D2	0.235090	0.093022	2.527254	0.0179
(GNI-183617.4)*D3	-0.279974	0.056512	-4.954266	0.0000
(GNI-266422)*D4	0.552660	0.069078	8.000503	0.0000
(GNI-340320)*D5	-0.514980	0.056164	-9.169241	0.0000
R-squared	0.975981	Mean dependent var	10428.57	
Adjusted R-squared	0.970438	S.D. dependent var	13612.43	
S.E. of regression	2340.462	Akaike info criterion	18.53992	
Sum squared resid	1.42E+08	Schwarz criterion	18.85736	
Log likelihood	-298.9086	F-statistic	176.0794	
Durbin-Watson stat	2.865889	Prob(F-statistic)	0.000000	

$$\hat{Y}_t = -698.8397 + 0.132747GNI_t - 0.187502(GNI_t - 70142.5)D_{1t} + 0.23509(GNI_t - 98000.5)D_{2t}$$

$se = (792.4822)$	(0.025281)	(0.093237)	(0.093022)
$t = (-0.8818)$	(5.2508)	(-2.0110)	(2.5273)

$$- 0.279974(GNI_t - 183617.4)D_{3t} + 0.55266(GNI_t - 266422)D_{4t} - 0.51498(GNI_t - 340320)D_{5t}$$

$se = (0.056512)$	(0.069078)	(0.056164)
$t = (-4.9543)$	(8.0005)	(-9.1692)

$$R^2=0.975981 \quad \bar{R}^2 = 0.970438 \quad F=176.0794 \quad DW=2.8659$$

由于各虚拟变量前的参数估计量的t值均大于临界值，表明各解释变量的斜率系数在显著性水平或下显著地不等于0。于是，居民人民币储蓄存款年增加额的回归模型分别为：

$$\hat{Y}_t = \begin{cases} -698.8397 + 0.132747GNI_t + \varepsilon_t & t \leq 1996 \\ 12453.01934 - 0.054755GNI_t + \varepsilon_t & 1996 < t \leq 2000 \\ -10585.91821 + 0.180335GNI_t + \varepsilon_t & 2000 < t \leq 2005 \\ 40822.17974 - 0.099639GNI_t + \varepsilon_t & 2005 < t \leq 2007 \\ -106418.6028 + 0.453021GNI_t + \varepsilon_t & 2007 < t \leq 2009 \\ 68839.3908 - 0.361959GNI_t + \varepsilon_t & 2009 < t \end{cases}$$

- 表明六个时期居民储蓄增加额的回归方程在统计意义上确实有很大不同。1996年以前国民总收入每增加1亿元，平均说来居民储蓄存款的增加额为0.1327亿元，居民储蓄增加额随国民总收入增长的变动相对稳定。1996年以后随着国民总收入的增长居民储蓄存款的增加额在不同年份有增有减，呈现出明显的周期性变化。上述模型同城乡居民储蓄存款与国民总收入之间的散布图吻合，与这一时段中国的实际经济运行状况也是相符的。

需要指出的是，在上述建模过程中，主要是从教学的目的出发，说明运用虚拟变量的规则和方法，没有考虑通货膨胀等因素，也没有考虑时间序列数据的特殊问题。而在实证分析中，储蓄函数还应当考虑如通货膨胀、消费行为变动等其他因素以及平稳性问题。

第八章 小结

1. 虚拟变量是人工构造的取值为**0**和**1**的作为属性变量代表的变量。
2. 虚拟变量个数的设置有一定规则：在有截距项的模型中，若定性因素有 m 个相互排斥的类型，只能引入 m 个虚拟变量，否则会陷入所谓“虚拟变量陷阱”，产生完全的多重共线性。

-
3. 在计量经济模型中，加入虚拟解释变量的途径有两种基本类型：一是加法类型；二是乘法类型。以加法方式引入虚拟变量改变的是模型的截距；以乘法方式引入虚拟变量改变的是模型的斜率。
 4. 解释变量只有一个分为两种相互排斥类型的定性变量而无定量变量的回归，称为方差分析模型。

-
5. 解释变量包含一个分为两种类型定性变量的回归时，只使用了一个虚拟变量；解释变量包含一个两种以上类型的定性变量的回归时，定性变量有 m 种类型，依据虚拟变量设置规则引入了 $m-1$ 个虚拟变量。
 6. 解释变量包含两个（或 k 个）定性变量的回归中，可选用了两个（或 k 个）虚拟变量去表示，这并不会出现“虚拟变量陷阱”。

7. 以乘法形式引入虚拟解释变量的主要作用在于：
对回归模型结构变化的检验；定性因素间交互
作用的影响分析；分段线性回归等。

第八章 结束了!



THANKS