2-8

基于回归模型的预测

主讲人：范国斌

基于回归的预测

基本思想：

●经估计的计量经济模型可用于：经济结构分析　经济预测　政策评价　验证理论

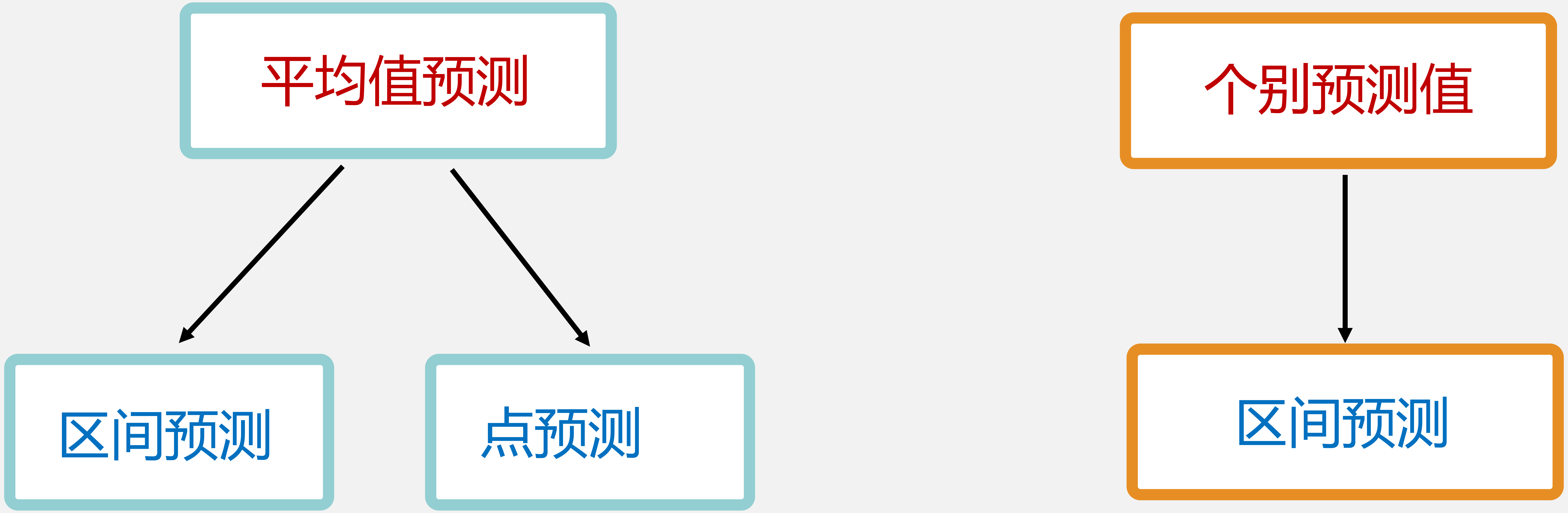
●运用计量经济模型作预测：指利用所估计的样本回归函数作预测工具，用解释变量的已知值或预测值，对预测期或样本以外的被解释变量的数值作出定量的估计。

●计量经济预测是一种条件预测：

条件：◆模型设定的关系式不变
◆所估计的参数不变
◆解释变量在预测期的取值已作出预测

预测的类型

- 对被解释变量Y的预测分为：平均值预测和个别值预测
- 对被解释变量Y的预测又分为：点预测和区间预测



Y 平均值的点预测

点预测:

用样本估计的总体参数值所计算的Y的估计值直接作为Y的预测值

方法： 将解释变量预测值直接代入估计的方程

$$\hat{Y}_F = \hat{\beta}_1 + \hat{\beta}_2 X_{F2} + \hat{\beta}_3 X_{F3} + \cdots + \hat{\beta}_K X_{Fk}$$

或

$$\hat{Y}_F = \boldsymbol{X}_F \hat{\boldsymbol{\beta}}$$

这样计算的 $\hat{Y}_F$ 是一个点估计值

Y 平均值的区间预测

基本思想：

●预测的目标值是真实平均值，由于存在抽样波动，预测的平均值 $\hat{Y}_F$ 是随机变量，不一定等于真实平均值 $E(Y_F|X_F)$，还需要对 $E(Y_F|X_F)$ 作区间估计

●为对Y的平均值作区间预测，必须确定平均值点预测值 $\hat{Y}_F$ 的抽样分布

●必须找出点预测值 $\hat{Y}_F$ 与预测目标值 $E(Y_F|X_F)$ 的关系，即找出与二者都有关的统计量

具体作法（从 $\hat{Y}_F$ 的分布分析）

由 $\hat{Y}_F = \boldsymbol{X}_F\hat{\boldsymbol{\beta}}$，$\hat{Y}_F$ 服从正态分布(为什么?)

已知

$$E(\hat{Y}_F) = E(Y_F | X_F) = X_F\beta$$

可以证明

$$Var(\hat{Y}_F) = \sigma^2 \boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'$$

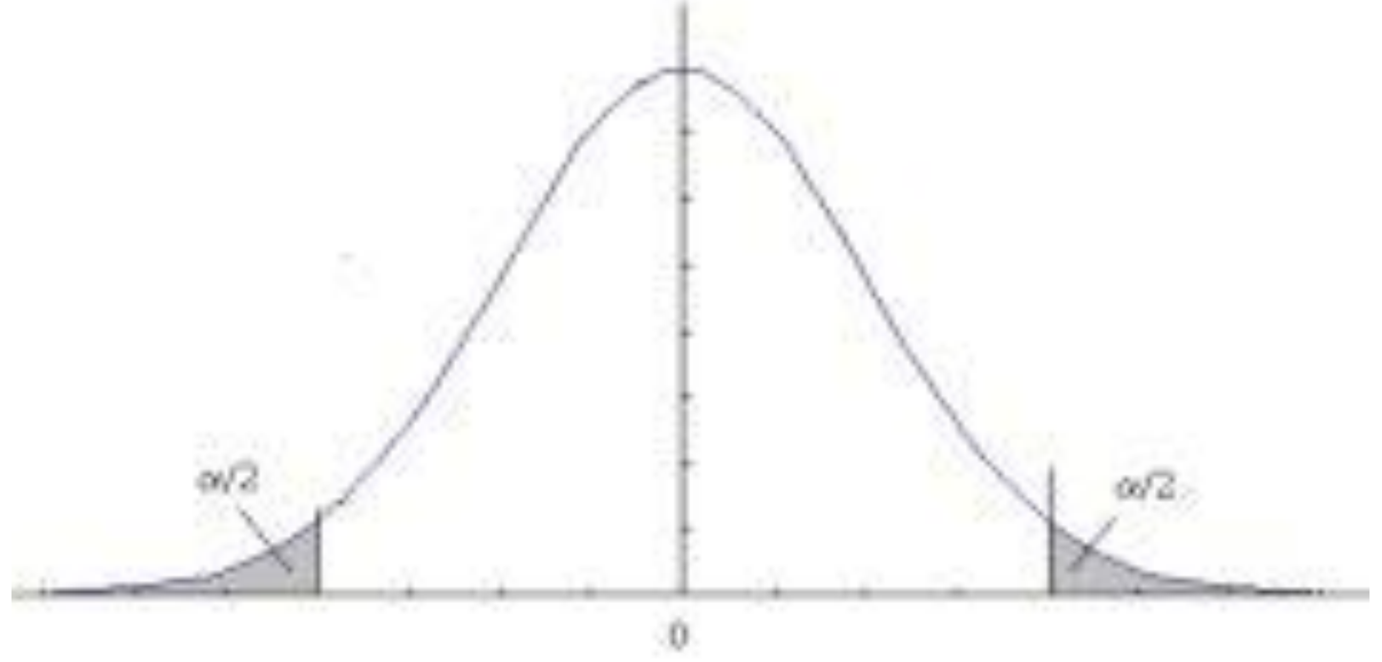
当 σ^2 未知时，只得用 $\hat{\sigma}^2 = \sum e_i^2 / (n-k)$ 代替，这时将 $\hat{Y}_F$ 标准化有

$$t = \frac{\hat{Y}_F - E(Y_F | \mathbf{X}_F)}{\hat{\sigma}\sqrt{\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}} \sim t(n-k)$$

注意:

构建平均值的预测区间

显然这样的 t 统计量与 $\hat{Y}_F$ 和 $E(Y_F|X_F)$ 都有关。 给定显著性水平α，查 t 分布表，得自由度 n - k 的临值 $t_{\alpha/2}(n-k)$ ，则有



$$P(-t_{\alpha/2} \le t = \frac{\hat{Y}_F - E(Y_F|X_F)}{\hat{SE}(\hat{Y}_F)} \le t_{\alpha/2}) = 1-\alpha$$

$$p\{[\hat{Y}_F - t_{\alpha/2}\hat{SE}(\hat{Y}_F)] \le E(Y_F|X_F) \le [\hat{Y}_F + t_{\alpha/2}\hat{SE}(\hat{Y}_F)]\} = 1-\alpha$$

Y平均值的置信度为$1-\alpha$ 的预测区间为

$$(\hat{Y}_F - t_{a/2}\hat{s}\sqrt{X_F(X'X)^{-1}X'_F}, \hat{Y}_F + t_{a/2}\hat{s}\sqrt{X_F(X'X)^{-1}X'_F})$$

被解释变量个别值预测

基本思想：

- $\hat{Y}_F$ 是对Y平均值的点预测。

- 由于存在随机扰动 u_i 的影响，Y的平均值并不等于Y的个别值

- 为了对Y的个别值 Y_F 作区间预测，需要寻找与点预测值 $\hat{Y}_F$ 和预测目标个别值 Y_F 有关的统计量，并要明确其概率分布

具体作法：

已知剩余项 $e_F = Y_F - \hat{Y}_F$ 是与预测值 $\hat{Y}_F$ 及个别值 Y_F 都有关的变量，并且已知 e_F 服从正态分布，且可证明 $E(e_F)=0$

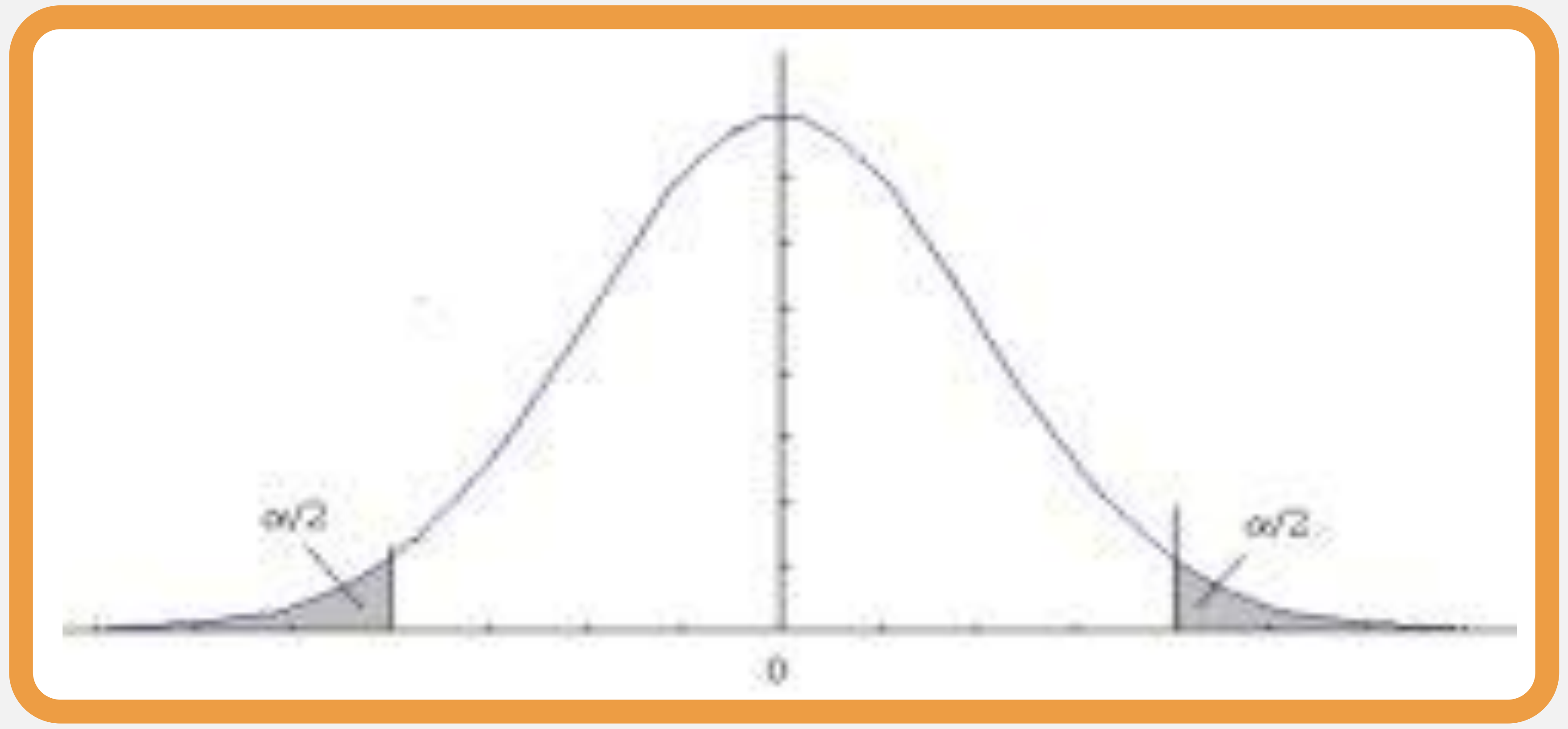
$$Var(e_F)=\sigma^2[1+\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F']$$

当用 $\hat{\sigma}^2=\sum e_i^2/(n-k)$ 代替 σ^2 时，对 e_F 标准化的变量 t 为

$$t=\frac{e_F-E(e_F)}{\hat{SE}(e_F)}=\frac{Y_F-\hat{Y}_F}{\hat{\sigma}\sqrt{1+\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}}\sim t(n-k)$$

构建个别值的预测区间

给定显著性水平 α，查 t 分布表得自由度为n-k的临界值 $t_{a/2}(n-k)$，则有



$$P\{[\hat{Y}_F - t_{\alpha/2}\hat{SE}(e_F)] \le Y_F \le [\hat{Y}_F + t_{\alpha/2}\hat{SE}(e_F)]\} = 1-\alpha$$

因此，Y的个别值的置信度为$1-\alpha$的预测区间上下限为

$$\hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}$$

被解释变量Y区间预测的特点

（1）Y平均值的预测值与真实平均值有误差，主要是受抽样波动影响

预测区间

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}$$

Y个别值的预测值与真实个别值的差异,不仅受抽样波动影响，而且还受随机扰动项的影响

预测区间

$$Y_F = \hat{Y}_F \mp t_{\alpha/2}\hat{\sigma}\sqrt{1+\boldsymbol{X}_F(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_F'}$$

被解释变量Y区间预测的特点(续)

在一元回归中：

$$X_F(X'X)^{-1}X'_F=\frac{1}{n}+\frac{(X_F-\bar{X})^2}{\sum x_i^2}$$

（2）平均值和个别值预测区间都不是常数，是随 X_F 的变化而变化的，当 $X_F=\bar{X}$ 时, 预测区间最小。

（3）预测区间上下限与样本容量有关，当样本容量 n → ∞ 时, 个别值的预测区间只决定于随机扰动的方差。

预测区间

$$Y_F=\hat{Y}_F\mp t_{\alpha/2}\,\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_F-\bar{X})^2}{\sum x_i^2}}$$