7.1 上海证券交易所的上市公司有两千多家，每年从中随机抽取 500 家，连续抽取 5 年，记录公司有关的财务数据。这些财务数据是面板数据吗？为什么？

7.2 什么是不可观测的个体异质性？其重要性体现在哪里？

7.3 如果真实模型是随机效应模型，估计时采用固定效应方法，估计结果受到什么影响？反过来呢？

7.4 固定效应模型采用模型（7.6）进行估计。证明在假设 2 成立时，模型（7.6）的误差项 $\varepsilon_{it}=u_{it}-\bar{u}_i$ 满足

（1）$\mathrm{E}(\varepsilon_{it}\varepsilon_{js})=0, i\neq j,\ i,j=1,2,\cdots,N,t,s=1,2,\cdots,T$

（2）$\mathrm{E}(\varepsilon_{it}\varepsilon_{is})=-\sigma_u^2/T, t\neq s,\ i,j=1,2,\cdots,N,t=1,2,\cdots,T$

（3）$\mathrm{Var}(\varepsilon_{it})=[(T-1)/T]\sigma_u^2,\ i,j=1,2,\cdots,N,t=1,2,\cdots,T$

。这能说明什么问题？

7.5 随机效应模型采用模型（7.9）估计模型参数。与固定效应模型（7.6）相比，随机效应模型估计的效率是更高还是更低？为什么？

7.6* 如果假设 2 和假设 3 满足，证明模型（7.9）的误差项 $\varepsilon_{it}=(1-\lambda)\alpha_i+(u_{it}-\lambda\bar{u}_i)$ 满足同方差和不相关假设。

7.7 表 7.3 给出的是 45 家企业 1980 年到 1986 年的研发数据，其中 lpt 为企业申请的专利数量的对

表 7.3

lpt	lrd	fcod	year
3.806662	0.99361	1	1980
3.433987	0.95654	1	1981
3.433987	0.809	1	1982
3.332205	0.71507	1	1983
3.135494	0.6605	1	1984
3.555348	0.59358	1	1985
3.401197	1.06047	1	1986
3.044522	2.87721	2	1980
3.258097	3.04874	2	1981
3.850148	3.17782	2	1982
3.73767	3.24168	2	1983
4.110874	3.28292	2	1984
4.143135	3.31609	2	1985
4.043051	3.34218	2	1986
……	……	……	……

数，lrd 为研发费用的对数，fcod 表示企业编号，year 为年份。为研究研发投入对专利申请数据的影响，建立面板数据模型

$$lpt_{it}=\alpha_i+\beta_1 lrd_{it}+u_{it}$$
$$i=1,2,\cdots,45,t=1,2,\cdots,7$$

。α_i 表示个体异质性变量，满足假设 3，u_{it} 为模型误差项，满足假设 1 和假设 2。

（1）α_i 包含哪些因素？

（2）将数据结构调整为面板数据结果，其中 cross-section ID series 选变量 fcod，Date series 选变量 year。

（3）用固定效应方法估计模型。

（4）用随机效应方法估计模型，并根据估计结果计算变量变换中的权重参数 λ。

（5）采用 Hausman 检验确定适合数据的模型是固定效应还是随机效应。

7.8　表 7.4 给出的是航空公司 1198 条航线 1997-2000 年的有关数据[①]。

表 7.4

YEAR	ID	FARE	DIST	CONCEN	PASSEN	Y98	Y99	Y00
1997	1	106	528	0.8386	152	0	0	0
1998	1	106	528	0.8133	265	1	0	0
1999	1	113	528	0.8262	336	0	1	0
2000	1	123	528	0.8612	298	0	0	1
1997	2	104	861	0.5798	282	0	0	0
1998	2	105	861	0.5817	178	1	0	0
1999	2	115	861	0.7319	204	0	1	0
2000	2	129	861	0.5386	190	0	0	1
……	……	……	……	……	……	……	……	……

$fare$ 表示平均机票价格（美元），$dist$ 表示航程距离（英里），$passen$ 表示平均日乘客人数，$concen$ 表示航班集中度，y98、y99 和 y00 分别表示 98 年、99 年和 2000 年的年份虚拟变量。ID 表示航线编号。为研究机票价格与各个影响因素之间的关系，建立面板数据模型

$$lfare_{it} = \alpha_i + \beta_1 concen_{it} + \beta_2 ldist_i + \beta_3 lpassen_{it} + u_{it},$$
$$i = 1,2,\cdots,1198,\ t = 1,2,3,4$$

。其中 $lfare$、$ldist$ 和 $lpassen$ 分别表示机票价格、航程距离和日均乘客数的对数。α_i 表示个体异质性变量，满足假设 3，u_{it} 为模型误差项，满足假设 1 和假设 2。

（1）α_i 包含哪些因素？

（2）将数据结构调整为面板数据结果，其中 cross-section ID series 选变量 ID，Date series 选变量 year。

（2）用固定效应方法估计模型。

（3）由于变量 $ldist_i$ 不随年份变化，不能直接估计 β_2。用年份虚拟变量和 $ldist_i$ 的乘积作为自变量，对模型实施固定效应估计，并分析航程距离对票价影响随年份变化的情况。

（4）用随机效应方法估计模型，并根据估计结果计算变量变换中的权重参数 λ。

（5）采用 Hausman 检验确定适合数据的模型是固定效应还是随机效应。

◆ 参考答案

1. 不是面板数据。面板数据要求连续观测 5 年的 500 家上市公司保持不变，而不是每年从两千多家公司随即抽样。从面板数据模型（7.1）看出，个体异质性 α_i 不随时间变化，是同一公司的一致性，这就要求不同年份上的个体（抽样上市公司）保持不变。
2. 个体异质性是指横截面个体的不随时间变化的特征，这些特征没有包含在模型自变量之中。个体异质性包括不可观测而无法放入自变量中的个体异质性，也包括可以观测但没有自变量放入模型的异质性。个体异质性的重要性体现在对模型参数估计的影响，如果与模型自变量相关，不加处理会引起内生性，导致 OLS 估计的不一致；如果与自变量不相关，尽管不会引起内生性，不加利用会降低 OLS 估计的效率。。
3. 采用固定效应方法估计随机效应的模型，估计结果是一致估计，但估计效率降低，同时，不随时间变量的自变量对应的回归系数无法估计。采用随机效应方法估计固定效应模型，估计结果不是一致估计，估计结果不可信。。

[①]数据来源：伍德里奇 著/费建平 译，《计量经济学导论》（第四版），人民大学出版社，2010.

4. 由于假设 2 满足，当$i \neq j$时，u_{it}与u_{js}不相关，因此u_{it}与$\bar{u}_j$不相关，$\bar{u}_i$与$\bar{u}_j$不相关。u_{it}只与平均式$\bar{u}_i$中的u_{it}相关。据此得出

（1）$\mathrm{E}(\varepsilon_{it}\varepsilon_{js}) = \mathrm{E}(u_{it} - \bar{u}_i)(u_{js} - \bar{u}_j) = \mathrm{E}(u_{it}u_{js}) - \mathrm{E}(u_{it}\bar{u}_j) - \mathrm{E}(\bar{u}_i u_{js}) + \mathrm{E}(\bar{u}_i\bar{u}_j) = 0$

（2）

$$\begin{aligned}\mathrm{E}(\varepsilon_{it}\varepsilon_{is}) = \mathrm{E}(u_{it} - \bar{u}_i)(u_{is} - \bar{u}_i) &= \mathrm{E}(u_{it}u_{is}) - \mathrm{E}(u_{it}\bar{u}_i) - \mathrm{E}(\bar{u}_i u_{is}) + \mathrm{E}(\bar{u}_i^2) \\ &= 0 - \frac{1}{T}\mathrm{E}(u_{it}^2) - \frac{1}{T}\mathrm{E}(u_{is}^2) + \frac{1}{T^2}\mathrm{E}(\sum\nolimits_{t=1}^{T} u_{it})^2 \\ &= 0 - \frac{2}{T}\sigma_u^2 + \frac{1}{T^2}\sum\nolimits_{t=1}^{T}\mathrm{E}(u_{it}^2) \\ &= -\frac{1}{T}\sigma_u^2\end{aligned}$$

（3）

$$\mathrm{Var}(\varepsilon_{it}) = \mathrm{E}(u_{it} - \bar{u}_i)^2 = \mathrm{E}(u_{it}^2 - 2\mathrm{E}(u_{it}\bar{u}_i) + \mathrm{E}(\bar{u}_i^2) = \sigma_u^2 - \frac{2}{T}\sigma_u^2 + \frac{1}{T}\sigma_u^2 = \frac{T-1}{T}\sigma_u^2$$

。这说明误差项ε_{it}是同方差，但不是不相关，模型（7.6）的 OLS 估计不满足马尔科夫性，不是最优估计。

5. 由于随机效应估计采用的是模型（7.9），模型的误差项满足同方差和不相关假设，OLS 估计满足马尔科夫性，是最优估计。由习题 4 的结论可知，（7.6）的 OLS 估计不是最优估计。因此，与固定效应模型相比，随机效应模型估计的效率更高。

6. 根据假设 3，对任意的i, j, t α_i与u_{jt}和$\bar{u}_j$都不相关，当$i \neq j$时，α_i与α_j不相关。再根据假设 2，当$i \neq j$时，u_{it}与u_{js}不相关，因此u_{it}与$\bar{u}_j$不相关，$\bar{u}_i$与$\bar{u}_j$不相关。u_{it}只与平均式$\bar{u}_i$中的u_{it}相关。利用习题 4 证明中的有关推导得出，当$i \neq j$时，

$$\begin{aligned}\mathrm{E}(\varepsilon_{it}\varepsilon_{js}) &= \mathrm{E}[(1\text{-}\lambda)\alpha_i + (u_{it} - \lambda\bar{u}_i)][(1\text{-}\lambda)\alpha_j + (u_{js} - \lambda\bar{u}_j)] \\ &= \mathrm{E}[(1\text{-}\lambda)^2\alpha_i\alpha_j] + \mathrm{E}[(1\text{-}\lambda)\alpha_i(u_{js} - \lambda\bar{u}_j)] \\ &\quad + \mathrm{E}[(u_{it} - \lambda\bar{u}_i)(1\text{-}\lambda)\alpha_j] + \mathrm{E}[(u_{it} - \lambda\bar{u}_i)(u_{js} - \lambda\bar{u}_j)] \\ &= 0\end{aligned}$$

。$i = j, t \neq s$时，注意到λ的定义（7.10），得出

$$\begin{aligned}\mathrm{E}(\varepsilon_{it}\varepsilon_{is}) &= \mathrm{E}[(1\text{-}\lambda)\alpha_i + (u_{it} - \lambda\bar{u}_i)][(1\text{-}\lambda)\alpha_i + (u_{is} - \lambda\bar{u}_i)] \\ &= \mathrm{E}[(1\text{-}\lambda)^2\alpha_i^2] + \mathrm{E}[(1\text{-}\lambda)\alpha_i(u_{is} - \lambda\bar{u}_i)] \\ &\quad + \mathrm{E}[(u_{it} - \lambda\bar{u}_i)(1\text{-}\lambda)\alpha_i] + \mathrm{E}[(u_{it} - \lambda\bar{u}_i)(u_{is} - \lambda\bar{u}_i)] \\ &= (1-\lambda)^2\sigma_\alpha^2 - 2(\lambda/T)\sigma_u^2 + (\lambda^2/T)\sigma_u^2 \\ &= 0\end{aligned}$$

这表明误差项ε_{it}不存在序列相关。进一步

$$\begin{aligned}\mathrm{Var}(\varepsilon_{it}) &= \mathrm{E}[(1\text{-}\lambda)\alpha_i + (u_{it} - \lambda\bar{u}_i)]^2 \\ &= \mathrm{E}[(1\text{-}\lambda)^2\alpha_i^2] + 2\mathrm{E}[(1\text{-}\lambda)\alpha_i(u_{it} - \lambda\bar{u}_i)] + \mathrm{E}[(u_{it} - \lambda\bar{u}_i)^2] \\ &= (1\text{-}\lambda)^2\sigma_\alpha^2 + \sigma_u^2 - 2(\lambda/T)\sigma_u^2 + (\lambda^2/T)\sigma_u^2]\end{aligned}$$

，表明ε_{it}为同方差。

7.7、7.8 略。