> 共 4 頁。標準常態分配表附於最後一頁。

Instructions:

- Write in Chinese or English only.

- Make sure all your answers are legible, unquestionably labeled, and clearly explained.

- A standard normal probability table is attached on the last page.

1. (10 points) A fast-food chain uses online and telephone questionnaires to estimate average customer satisfaction at one of its stores. Each customer reports an overall satisfaction rating between 1 and 10. The fast-food chain will solicit 66 online surveys and 42 telephone surveys. The results of each survey are independent draws from the same distribution. Let $\bar{X}$ and $\bar{Y}$ denote the sample mean responses to the online and telephone surveys, respectively. The fast-food chain would like to take a weighted average $\bar{C} = w\bar{X} + (1-w)\bar{Y}$ of these sample means to obtain a single estimator of average customer satisfaction. Determine the choice of weight w that leads to the most efficient unbiased estimator of average customer satisfaction.

解： 設母體平均滿意度為 μ，母體變異數為 σ^2。

確認不偏性：

$$\mathbb{E}[\bar{C}] = w\mathbb{E}[\bar{X}] + (1-w)\mathbb{E}[\bar{Y}] = w\mu + (1-w)\mu = \mu$$

對於任何 w，$\bar{C}$ 都是 μ 的不偏估計量。

計算 $\bar{C}$ 的變異數：

由於線上調查與電話調查獨立：

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{66}, \quad \text{Var}(\bar{Y}) = \frac{\sigma^2}{42}$$

$$\text{Var}(\bar{C}) = w^2\text{Var}(\bar{X}) + (1-w)^2\text{Var}(\bar{Y}) = w^2 \cdot \frac{\sigma^2}{66} + (1-w)^2 \cdot \frac{\sigma^2}{42}$$

最小化變異數：

對 w 微分並令其為零：

$$\frac{d\text{Var}(\bar{C})}{dw} = 2w \cdot \frac{\sigma^2}{66} - 2(1-w) \cdot \frac{\sigma^2}{42} = 0$$

$$\frac{w}{66} = \frac{1-w}{42}$$

$$42w = 66(1-w) = 66 - 66w$$

$$42w + 66w = 66 \implies 108w = 66 \implies w = \frac{66}{108} = \frac{11}{18}$$

驗證二階條件：

$$\frac{d^2\text{Var}(\bar{C})}{dw^2} = \frac{2\sigma^2}{66} + \frac{2\sigma^2}{42} > 0$$

確實為最小值。

答案：最有效率的不偏估計量權重為　$\boxed{w = \dfrac{11}{18} \approx 0.611}$

註：這個結果符合直覺——樣本數較大的調查（線上 66 份）應給予較大權重。

一般公式：當有兩個獨立的不偏估計量，最佳權重與各自變異數的倒數成比例：

$$w^* = \frac{1/\mathrm{Var}(\bar{X})}{1/\mathrm{Var}(\bar{X}) + 1/\mathrm{Var}(\bar{Y})} = \frac{n_X}{n_X + n_Y} = \frac{66}{66 + 42} = \frac{66}{108} = \frac{11}{18}$$

2. (20 points) A bank manager is analyzing the usage patterns of her bank's ATM. She observes that the amount of time (X, in minutes) a user spends at the ATM follows the probability distribution below:

x	$\mathbb{P}(X = x)$
2	0.5
3	0.2
4	0.3

The amounts of time spent by different users at the ATM are independent of one another. Calculate the probability that the total time spent by the first 120 users at the ATM is at least 330 minutes.

解：　步驟一：計算單一使用者時間的期望值與變異數

$$\mathbb{E}[X] = 2(0.5) + 3(0.2) + 4(0.3) = 1 + 0.6 + 1.2 = 2.8 \text{ 分鐘}$$

$$\mathbb{E}[X^2] = 4(0.5) + 9(0.2) + 16(0.3) = 2 + 1.8 + 4.8 = 8.6$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 8.6 - (2.8)^2 = 8.6 - 7.84 = 0.76$$

步驟二：設 S_{120} 為 120 位使用者的總時間

$$S_{120} = X_1 + X_2 + \cdots + X_{120}$$

由於各 X_i 獨立同分布：

$$\mathbb{E}[S_{120}] = 120 \times 2.8 = 336 \text{ 分鐘}$$

$$\mathrm{Var}(S_{120}) = 120 \times 0.76 = 91.2$$

$$\sigma_{S_{120}} = \sqrt{91.2} \approx 9.55$$

步驟三：應用中央極限定理

當 $n = 120$ 足夠大時，S_{120} 近似常態分配：

$$S_{120} \overset{\text{approx}}{\sim} N(336, 91.2)$$

步驟四：計算 $\mathbb{P}(S_{120} \geq 330)$

標準化：

$$Z = \frac{S_{120} - 336}{\sqrt{91.2}} = \frac{S_{120} - 336}{9.55}$$

$$\mathbb{P}(S_{120} \geq 330) = \mathbb{P}\left(Z \geq \frac{330 - 336}{9.55}\right) = \mathbb{P}\left(Z \geq \frac{-6}{9.55}\right) = \mathbb{P}(Z \geq -0.628)$$

由對稱性：

$$\mathbb{P}(Z \geq -0.628) = \mathbb{P}(Z \leq 0.628)$$

查表： $Z = 0.63$ 時，$\Phi(0.63) \approx 0.7357$（從標準常態表，$\Phi(-0.63) \approx 0.2643$，故 $\Phi(0.63) = 1 - 0.2643 = 0.7357$）

答案： $\mathbb{P}(S_{120} \geq 330) \approx \boxed{0.7357}$（或約 73.6%）

註： 若使用連續性校正，則計算 $\mathbb{P}(S_{120} \geq 329.5)$：

$$Z = \frac{329.5 - 336}{9.55} = -0.68 \implies \mathbb{P}(Z \geq -0.68) = \Phi(0.68) \approx 0.7517$$

3. (20 points) A power company claims that it has successfully reduced the average carbon dioxide emissions from one of its facilities to 24,000 metric tons per day to meet its obligations in an emissions trading program. However, an emissions regulator suspects that the actual average is lower, at 23,000 metric tons per day. Assume the regulator's estimate is correct, and that daily emissions are independent and identically distributed with a standard deviation of 2,200 metric tons. Determine the number of days required to perform a one-sided hypothesis test with a Type I error probability of 0.01 and a Type II error probability of 0.02.

解： 設定假設檢定：

監管機構懷疑排放量比公司聲稱的更低，因此：

$$H_0 : \mu = 24{,}000 \text{(公司聲稱的排放量)}$$
$$H_1 : \mu < 24{,}000 \text{(實際排放量較低)}$$

這是左尾檢定。

給定參數：

- $\mu_0 = 24{,}000$（虛無假設下的平均值）

- $\mu_1 = 23{,}000$（對立假設下的真實平均值）

- $\sigma = 2{,}200$

- $\alpha = 0.01$（型一錯誤機率）

- $\beta = 0.02$（型二錯誤機率）

臨界值的設定：

設檢定的臨界值為 c，當 $\bar{X} < c$ 時拒絕 H_0。

型一錯誤條件：（在 H_0 為真時錯誤拒絕）

$$\alpha = \mathbb{P}(\bar{X} < c \mid \mu = 24{,}000) = 0.01$$

標準化：$\dfrac{c - 24{,}000}{\sigma/\sqrt{n}} = z_{0.01} = -2.33$（查表：$\Phi(-2.33) \approx 0.01$）

$$c = 24{,}000 - 2.33 \cdot \frac{2{,}200}{\sqrt{n}} \tag{1}$$

型二錯誤條件：（在 H_1 為真時未能拒絕）

$$\beta = \mathbb{P}(\bar{X} \geq c \mid \mu = 23{,}000) = 0.02$$

即 $\mathbb{P}(\bar{X} < c \mid \mu = 23{,}000) = 1 - 0.02 = 0.98$

標準化：$\dfrac{c - 23{,}000}{\sigma/\sqrt{n}} = z_{0.98} = 2.05$（查表：$\Phi(2.05) \approx 0.98$，或 $\Phi(-2.05) \approx 0.02$）

$$c = 23{,}000 + 2.05 \cdot \frac{2{,}200}{\sqrt{n}} \tag{2}$$

聯立求解：

由 $(1) = (2)$：

$$24{,}000 - 2.33 \cdot \frac{2{,}200}{\sqrt{n}} = 23{,}000 + 2.05 \cdot \frac{2{,}200}{\sqrt{n}}$$

$$24{,}000 - 23{,}000 = 2.05 \cdot \frac{2{,}200}{\sqrt{n}} + 2.33 \cdot \frac{2{,}200}{\sqrt{n}}$$

$$1{,}000 = (2.05 + 2.33) \cdot \frac{2{,}200}{\sqrt{n}} = 4.38 \cdot \frac{2{,}200}{\sqrt{n}}$$

$$\sqrt{n} = \frac{4.38 \times 2{,}200}{1{,}000} = \frac{9{,}636}{1{,}000} = 9.636$$

$$n = (9.636)^2 = 92.85$$

答案：需要 $\boxed{n = 93}$ 天。（無條件進位，因為需要至少達到指定的檢定力）

一般公式：對於單尾檢定，所需樣本量為：

$$n = \left(\frac{(z_\alpha + z_{1-\beta})\sigma}{\mu_0 - \mu_1} \right)^2 = \left(\frac{(2.33 + 2.05) \times 2{,}200}{1{,}000} \right)^2 \approx 93$$

4. A researcher is analyzing the factors influencing whether individuals participate in a job training program.

The binary variable y_i equals 1 if individual i participates and 0 otherwise. The researcher uses the following probit model:

$$\mathbb{P}(y_i = 1|x_i) = \Phi(x_i^\top \beta), \tag{1}$$

where

- $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution,
- $x_i = [x_{i1}, x_{i2}, \ldots, x_{ik}]^\top$ is a $k \times 1$ vector of explanatory variables (e.g., age, education, and income), and
- $\beta = [\beta_1, \beta_2, \ldots, \beta_k]^\top$ is a vector of coefficients to be estimated.

(a) (8 points) Derive the log-likelihood function and the corresponding first-order conditions for the probit model given a sample of n observations.

解：個別觀測值的概似函數：

對於觀測值 i：

$$\mathbb{P}(y_i = 1|x_i) = \Phi(x_i^\top \beta), \quad \mathbb{P}(y_i = 0|x_i) = 1 - \Phi(x_i^\top \beta)$$

合併寫成：

$$L_i(\beta) = [\Phi(x_i^\top \beta)]^{y_i}[1 - \Phi(x_i^\top \beta)]^{1-y_i}$$

對數概似函數：

假設觀測值獨立，總對數概似函數為：

$$\boxed{\ell(\beta) = \sum_{i=1}^{n} \left\{ y_i \ln \Phi(x_i^\top \beta) + (1 - y_i) \ln[1 - \Phi(x_i^\top \beta)] \right\}}$$

一階條件（Score Function）：

令 $\phi(\cdot)$ 為標準常態密度函數。利用 $\frac{d\Phi(z)}{dz} = \phi(z)$：

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \left\{ y_i \frac{\phi(x_i^\top \beta)}{\Phi(x_i^\top \beta)} x_i - (1 - y_i) \frac{\phi(x_i^\top \beta)}{1 - \Phi(x_i^\top \beta)} x_i \right\}$$

整理：

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \left\{ \frac{y_i - \Phi(x_i^\top \beta)}{\Phi(x_i^\top \beta)[1 - \Phi(x_i^\top \beta)]} \phi(x_i^\top \beta) \right\} x_i$$

一階條件：令 $\frac{\partial \ell}{\partial \beta} = 0$：

$$\boxed{\sum_{i=1}^{n} \frac{[y_i - \Phi(x_i^\top \beta)]\phi(x_i^\top \beta)}{\Phi(x_i^\top \beta)[1 - \Phi(x_i^\top \beta)]} x_i = 0}$$

或等價地寫成：

$$\sum_{i=1}^{n} \lambda_i x_i = 0, \quad 其中 \lambda_i = \frac{[y_i - \Phi(x_i^\top \beta)]\phi(x_i^\top \beta)}{\Phi(x_i^\top \beta)[1 - \Phi(x_i^\top \beta)]}$$

此為非線性方程組，需以數值方法（如 Newton-Raphson 或 BFGS）求解。

(b) (4 points) Derive the expression for the marginal effect of a continuous explanatory variable x_{ik} on the probability $\mathbb{P}(y_i = 1|x_i)$.

解：邊際效應的定義：

連續變數 x_{ik} 對 $\mathbb{P}(y_i = 1 | x_i)$ 的邊際效應為：

$$\frac{\partial \mathbb{P}(y_i = 1 | x_i)}{\partial x_{ik}} = \frac{\partial \Phi(x_i^\top \beta)}{\partial x_{ik}}$$

應用連鎖法則：

$$\frac{\partial \Phi(x_i^\top \beta)}{\partial x_{ik}} = \phi(x_i^\top \beta) \cdot \frac{\partial (x_i^\top \beta)}{\partial x_{ik}} = \phi(x_i^\top \beta) \cdot \beta_k$$

邊際效應公式：

$$\boxed{\mathrm{ME}_{ik} = \phi(x_i^\top \beta) \cdot \beta_k}$$

重要特性：

- 邊際效應依賴於 x_i 的值（透過 $\phi(x_i^\top \beta)$），因此對不同個體有不同的邊際效應。

- $\phi(\cdot) > 0$，所以邊際效應的正負號與 β_k 一致。

- 邊際效應在 $x_i^\top \beta = 0$ 時最大（此時 $\phi(0) \approx 0.399$）。

- 常見的報告方式：
  - 在平均值處的邊際效應 (MEM)：$\phi(\bar{x}^\top \beta) \cdot \beta_k$
  - 平均邊際效應 (AME)：$\frac{1}{n} \sum_{i=1}^{n} \phi(x_i^\top \beta) \cdot \beta_k$

5. A researcher is studying the relationship between years of education (Edu_i) and earnings (Earn_i) using the following linear regression model:

$$\mathrm{Earn}_i = \beta_0 + \beta_1 \mathrm{Edu}_i + \beta_2 \mathrm{Ability}_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

in which $\mathrm{Ability}_i$ is positively correlated with both Edu_i and Earn_i, and the error term $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ is uncorrelated with both Edu_i and $\mathrm{Ability}_i$. However, the researcher does not observe $\mathrm{Ability}_i$.

(a) (6 points) Suppose that the researcher considers a simplified model:

$$\mathrm{Earn}_i = b_0 + b_1 \mathrm{Edu}_i + \epsilon_i, \quad i = 1, 2, \ldots, n. \tag{3}$$

Will the ordinary least squares (OLS) estimate of b_1 in the regression above be a consistent estimator for β_1? If not, in which direction does the bias occur? Explain your answer with relevant equations.

解：OLS 估計量 $\hat{b}_1$ 對 β_1 不一致。

推導：

真實模型 (2) 代入簡化模型 (3)：

$$\mathrm{Earn}_i = \beta_0 + \beta_1 \mathrm{Edu}_i + \beta_2 \mathrm{Ability}_i + \epsilon_i$$

在模型 (3) 中，誤差項實際上是：

$$u_i = \beta_2 \mathrm{Ability}_i + \epsilon_i + (\beta_0 - b_0)$$

OLS 估計量的機率極限：

$$\text{plim}\hat{b}_1 = \beta_1 + \beta_2 \cdot \frac{\text{Cov}(\text{Edu}_i, \text{Ability}_i)}{\text{Var}(\text{Edu}_i)}$$

遺漏變數偏誤公式：

$$\text{plim}\hat{b}_1 = \beta_1 + \beta_2 \cdot \delta_1$$

其中 δ_1 是將 Ability_i 對 Edu_i 迴歸的係數。

偏誤方向分析：

- $\beta_2 > 0$：能力對收入有正向影響

- $\text{Cov}(\text{Edu}_i, \text{Ability}_i) > 0$：能力與教育正相關（題目已給定）

- 因此 $\delta_1 > 0$

結論：

$$\text{plim}\hat{b}_1 = \beta_1 + \underbrace{\beta_2 \cdot \delta_1}_{>0} > \beta_1$$

$\hat{b}_1$ 有**向上偏誤（upward bias）**。

直觀解釋：由於能力較高的人傾向於接受更多教育，當我們不控制能力時，教育的係數會「吸收」部分能力對收入的效果，導致高估教育的真實回報。

(b) (6 points) Suppose that the researcher has access to two measured variables, $\widetilde{\text{Ability}}_i$ and $\widehat{\text{Ability}}_i$:

$$\widetilde{\text{Ability}}_i = \text{Ability}_i + u_i, \quad \text{and} \quad \widehat{\text{Ability}}_i = \text{Ability}_i + v_i, \tag{4}$$

where error terms $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ and $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ are uncorrelated with Edu_i, Ability_i, ϵ_i, and each other. To minimize the effects of these errors, the researcher calculates the average of these two measures:

$$\overline{\text{Ability}_i} = \frac{\widetilde{\text{Ability}}_i + \widehat{\text{Ability}}_i}{2}, \tag{5}$$

and considers the following model:

$$\text{Earn}_i = \gamma_0 + \gamma_1 \text{Edu}_i + \gamma_2 \overline{\text{Ability}_i} + e_i, \quad i = 1, 2, \ldots, n. \tag{6}$$

Will the ordinary least squares (OLS) estimate of γ_1 in the regression above be a consistent estimator for β_1? Will the OLS estimate of γ_2 be a consistent estimator for β_2? Explain your answer with relevant equations.

解：分析 $\overline{\text{Ability}}_i$ 的結構：

$$\overline{\text{Ability}_i} = \frac{\widetilde{\text{Ability}}_i + \widehat{\text{Ability}}_i}{2} = \text{Ability}_i + \frac{u_i + v_i}{2}$$

令 $w_i = \frac{u_i + v_i}{2}$，則：

$$\overline{\text{Ability}_i} = \text{Ability}_i + w_i$$

其中 $\mathbb{E}[w_i] = 0$，$\text{Var}(w_i) = \frac{\sigma_u^2 + \sigma_v^2}{4}$。

將真實模型代入：

$$\text{Earn}_i = \beta_0 + \beta_1\text{Edu}_i + \beta_2(\overline{\text{Ability}_i} - w_i) + \epsilon_i$$

$$= \beta_0 + \beta_1\text{Edu}_i + \beta_2\overline{\text{Ability}_i} + (\epsilon_i - \beta_2 w_i)$$

模型 (6) 的誤差項為 $e_i = \epsilon_i - \beta_2 w_i$。

$\hat{\gamma}_1$ 的一致性：

關鍵是檢查 $\text{Cov}(\text{Edu}_i, e_i)$：

$$\text{Cov}(\text{Edu}_i, e_i) = \text{Cov}(\text{Edu}_i, \epsilon_i) - \beta_2\text{Cov}(\text{Edu}_i, w_i) = 0 - 0 = 0$$

由於 Edu_i 與 e_i 不相關，$\hat{\gamma}_1$ 是 β_1 的一致估計量。

$\hat{\gamma}_2$ 的一致性：

關鍵是檢查 $\text{Cov}(\overline{\text{Ability}_i}, e_i)$：

$$\text{Cov}(\overline{\text{Ability}_i}, e_i) = \text{Cov}(\text{Ability}_i + w_i, \epsilon_i - \beta_2 w_i)$$

$$= \text{Cov}(\text{Ability}_i, \epsilon_i) - \beta_2\text{Cov}(\text{Ability}_i, w_i) + \text{Cov}(w_i, \epsilon_i) - \beta_2\text{Var}(w_i)$$

$$= 0 - 0 + 0 - \beta_2\text{Var}(w_i) = -\beta_2 \cdot \frac{\sigma_u^2 + \sigma_v^2}{4} \neq 0$$

由於 $\overline{\text{Ability}_i}$ 與誤差項 e_i 相關（經典的測量誤差問題），$\hat{\gamma}_2$ 對 β_2 不一致。

具體而言，$\hat{\gamma}_2$ 會有**衰減偏誤（attenuation bias）**，向零收縮：

$$\text{plim}\hat{\gamma}_2 = \beta_2 \cdot \frac{\text{Var}(\text{Ability}_i)}{\text{Var}(\text{Ability}_i) + \frac{\sigma_u^2 + \sigma_v^2}{4}} < \beta_2$$

(c) (6 points) Besides the estimators above, how would you consistently estimate β_1 and β_2? Provide a detailed explanation supported by relevant equations and formulas.

解：方法：工具變數法（Instrumental Variables, IV）

可以使用其中一個測量變數作為另一個的工具變數。

策略：使用 $\widehat{\text{Ability}_i}$ 作為 $\widetilde{\text{Ability}_i}$ 的工具變數（或反過來）。

設定迴歸模型：

$$\text{Earn}_i = \beta_0 + \beta_1\text{Edu}_i + \beta_2\widetilde{\text{Ability}_i} + \tilde{e}_i$$

其中 $\tilde{e}_i = \epsilon_i - \beta_2 u_i$。

驗證 $\widehat{\text{Ability}_i}$ 作為 IV 的有效性：

1. 相關性條件（Relevance）：

$$\text{Cov}(\widehat{\text{Ability}_i}, \widetilde{\text{Ability}_i}) = \text{Cov}(\text{Ability}_i + v_i, \text{Ability}_i + u_i) = \text{Var}(\text{Ability}_i) \neq 0 \quad \checkmark$$

2. 外生性條件（Exogeneity）：

$$\text{Cov}(\widehat{\text{Ability}_i}, \tilde{e}_i) = \text{Cov}(\text{Ability}_i + v_i, \epsilon_i - \beta_2 u_i) = 0 \quad \checkmark$$

（因為 v_i 與 ϵ_i、u_i 都不相關）

兩階段最小平方法（2SLS）：

第一階段：將 $\widetilde{\text{Ability}}_i$ 對工具變數迴歸

$$\widetilde{\text{Ability}}_i = \pi_0 + \pi_1 \text{Edu}_i + \pi_2 \widehat{\text{Ability}}_i + \eta_i$$

得到預測值 $\widetilde{\text{Ability}}_i^*$。

第二階段：用預測值進行迴歸

$$\text{Earn}_i = \beta_0 + \beta_1 \text{Edu}_i + \beta_2 \widetilde{\text{Ability}}_i^* + e_i$$

結論：使用 2SLS，$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 都是一致估計量。

替代方法：也可以同時使用 $\widetilde{\text{Ability}}_i$ 和 $\widehat{\text{Ability}}_i$ 作為 Ability_i 的工具變數，用廣義矩估計（GMM）進行估計，可以提高效率。

---

6. Consider the following dynamic panel data model:

$$y_{it} = \mu_i + \alpha y_{i,t-1} + \epsilon_{it}, \quad i = 1, 2, \ldots, n, \quad \text{and} \quad t = 2, 3, \ldots, T, \tag{7}$$

where $T \geq 4$,

- y_{it} is the dependent variable for individual i at time t,
- $y_{i,t-1}$ is the lagged dependent variable,
- μ_i represents the individual-specific effect, and
- ϵ_{it} is the idiosyncratic error term, assumed to be i.i.d. $N(0, \sigma^2)$.

Taking the first difference of equation (7) eliminates the individual-specific effect, resulting in:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \epsilon_{it}, \quad i = 1, 2, \ldots, n, \quad \text{and} \quad t = 3, 4, \ldots, T, \tag{8}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ represents the change in y over time. The first-difference (FD) estimator of α, denoted as $\hat{\alpha}_{FD}$, is obtained by applying ordinary least squares (OLS) to the transformed variables Δy_{it} and $\Delta y_{i,t-1}$:

$$\hat{\alpha}_{FD} = \frac{\sum_{i=1}^n \sum_{t=3}^T \Delta y_{i,t-1} \Delta y_{it}}{\sum_{i=1}^n \sum_{t=3}^T \Delta y_{i,t-1}^2}. \tag{9}$$

(a) (8 points) (Nickell, 1981, *Econometrica*) Consider the scenario where $n \to \infty$ while T remains fixed. Explain whether $\hat{\alpha}_{FD}$ is consistent. Justify your answer.

解：$\hat{\alpha}_{FD}$ **在** $n \to \infty$、T **固定時不一致。**

原因：$\Delta y_{i,t-1}$ 與 $\Delta \epsilon_{it}$ 相關

展開：

$$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$$
$$\Delta \epsilon_{it} = \epsilon_{it} - \epsilon_{i,t-1}$$

計算共變異數：

$$\text{Cov}(\Delta y_{i,t-1}, \Delta \epsilon_{it}) = \text{Cov}(y_{i,t-1} - y_{i,t-2}, \epsilon_{it} - \epsilon_{i,t-1})$$

由於 $y_{i,t-1}$ 取決於 $\epsilon_{i,t-1}$（透過動態結構），而 ϵ_{it} 與 $y_{i,t-2}$ 及更早的項無關：

$$= \text{Cov}(y_{i,t-1}, -\epsilon_{i,t-1}) + \text{Cov}(-y_{i,t-2}, -\epsilon_{i,t-1})$$

$$= -\text{Cov}(y_{i,t-1}, \epsilon_{i,t-1}) + \text{Cov}(y_{i,t-2}, \epsilon_{i,t-1})$$

由於 $y_{i,t-1} = \mu_i + \alpha y_{i,t-2} + \epsilon_{i,t-1}$：

$$\text{Cov}(y_{i,t-1}, \epsilon_{i,t-1}) = \text{Var}(\epsilon_{i,t-1}) = \sigma^2$$

而 $y_{i,t-2}$ 不包含 $\epsilon_{i,t-1}$（它只取決於 $\epsilon_{i,t-2}$ 及更早的項）：

$$\text{Cov}(y_{i,t-2}, \epsilon_{i,t-1}) = 0$$

因此：

$$\text{Cov}(\Delta y_{i,t-1}, \Delta \epsilon_{it}) = -\sigma^2 \neq 0$$

結論：由於解釋變數 $\Delta y_{i,t-1}$ 與誤差項 $\Delta \epsilon_{it}$ 相關，OLS 估計量 $\hat{\alpha}_{FD}$ 有偏且不一致。

當 $n \to \infty$ 而 T 固定時，此相關性不會消失，這就是著名的 **Nickell 偏誤**。偏誤方向通常為負（$\text{plim}\hat{\alpha}_{FD} < \alpha$）。

(b) (4 points) Now consider the scenario where both $n \to \infty$ and $T \to \infty$. Determine whether $\hat{\alpha}_{FD}$ is a consistent estimator, and explain your reasoning.

解：當 $n \to \infty$ 且 $T \to \infty$ 時，$\hat{\alpha}_{FD}$ 是一致的。

理由：

Nickell 偏誤的大小與 T 有關。可以證明偏誤項為 $O(1/T)$ 量級。

具體而言，當 T 固定時：

$$\text{plim}_{n \to \infty} \hat{\alpha}_{FD} = \alpha + \frac{偏誤項}{T-1} + O(1/T^2)$$

當 $T \to \infty$ 時：

- 分子中 $\text{Cov}(\Delta y_{i,t-1}, \Delta \epsilon_{it}) = -\sigma^2$ 對每個 t 都成立

- 但分母 $\text{Var}(\Delta y_{i,t-1})$ 在對 t 求和時的增長速度使得偏誤項相對於總變異趨近於零

更正式地說：

$$\text{plim}_{n,T \to \infty} \hat{\alpha}_{FD} = \alpha$$

當 $T \to \infty$ 時，每個個體有足夠多的時間序列觀測值，使得相關性的「有限樣本」問題被稀釋。

結論：若 n 和 T 同時趨向無窮大（例如 $n/T \to c$ 或其他適當的速率），$\hat{\alpha}_{FD}$ 是一致估計量。

(c) (8 points) (Anderson and Hsiao, 1982, *Journal of Econometrics*) The Anderson and Hsiao (AH) estimator of α, denoted as $\hat{\alpha}_{AH}$, is obtained by applying two stage least squares (2SLS) to the transformed variables Δy_{it} and $\Delta y_{i,t-1}$, using $\Delta y_{i,t-2}$ as the instrumental variable (IV):

$$\hat{\alpha}_{AH} = \frac{\sum_{i=1}^{n} \sum_{t=4}^{T} \Delta y_{i,t-2} \Delta y_{it}}{\sum_{i=1}^{n} \sum_{t=4}^{T} \Delta y_{i,t-2} \Delta y_{i,t-1}}. \tag{10}$$

Again, consider the scenario where $n \to \infty$ while T remains fixed. Discuss whether $\hat{\alpha}_{AH}$ is a consistent estimator for α, and provide justification for your conclusion.

解： 當 $n \to \infty$ 而 T **固定時，**$\hat{\alpha}_{AH}$ **是一致估計量。**

驗證工具變數 $\Delta y_{i,t-2}$ **的有效性：**

1. 相關性條件（Relevance）：$\text{Cov}(\Delta y_{i,t-2}, \Delta y_{i,t-1}) \neq 0$

$$\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$$
$$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$$

由於動態結構，$y_{i,t-1}$ 與 $y_{i,t-2}$ 相關，$y_{i,t-2}$ 與 $y_{i,t-3}$ 相關：

$$\text{Cov}(\Delta y_{i,t-2}, \Delta y_{i,t-1}) \neq 0 \quad \checkmark$$

2. 外生性條件（Exogeneity）：$\text{Cov}(\Delta y_{i,t-2}, \Delta \epsilon_{it}) = 0$

$$\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$$
$$\Delta \epsilon_{it} = \epsilon_{it} - \epsilon_{i,t-1}$$

由於 ϵ_{it} 是 i.i.d.，且 $y_{i,t-2}$ 和 $y_{i,t-3}$ 只取決於 $\epsilon_{i,t-2}$ 及更早的誤差項：

$$\text{Cov}(y_{i,t-2}, \epsilon_{it}) = 0, \quad \text{Cov}(y_{i,t-2}, \epsilon_{i,t-1}) = 0$$

$$\text{Cov}(y_{i,t-3}, \epsilon_{it}) = 0, \quad \text{Cov}(y_{i,t-3}, \epsilon_{i,t-1}) = 0$$

因此：

$$\text{Cov}(\Delta y_{i,t-2}, \Delta \epsilon_{it}) = 0 \quad \checkmark$$

IV 估計量的一致性：

$$\hat{\alpha}_{AH} = \frac{\sum_{i,t} \Delta y_{i,t-2} \Delta y_{it}}{\sum_{i,t} \Delta y_{i,t-2} \Delta y_{i,t-1}}$$

將 $\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \epsilon_{it}$ 代入分子：

$$\sum_{i,t} \Delta y_{i,t-2} \Delta y_{it} = \alpha \sum_{i,t} \Delta y_{i,t-2} \Delta y_{i,t-1} + \sum_{i,t} \Delta y_{i,t-2} \Delta \epsilon_{it}$$

取機率極限（當 $n \to \infty$）：

$$\text{plim}\,\hat{\alpha}_{AH} = \alpha + \frac{\text{plim}\frac{1}{n(T-3)}\sum_{i,t} \Delta y_{i,t-2} \Delta \epsilon_{it}}{\text{plim}\frac{1}{n(T-3)}\sum_{i,t} \Delta y_{i,t-2} \Delta y_{i,t-1}}$$

由外生性條件，分子的極限為 0，分母的極限不為 0：

$$\text{plim}\,\hat{\alpha}_{AH} = \alpha + \frac{0}{\text{非零常數}} = \alpha$$

結論：Anderson-Hsiao 估計量 $\hat{\alpha}_{AH}$ 在 $n \to \infty$、T 固定時是 α 的一致估計量。這是因為使用了適當的工具變數 $\Delta y_{i,t-2}$，該工具變數與解釋變數相關但與誤差項不相關。

註：也可以使用 $y_{i,t-2}$（水準值而非差分）作為工具變數，同樣可以得到一致估計量。

Standard normal probability table

（負 Z 值的左尾機率 $\Phi(Z) = P(Z \le z)$）

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

*For $Z \le -3.50$, the probability is less than or equal to 0.0002.