PS2 Solutions

Jingle Fu

Solution (a).

```
use GMdata.dta, clear
tabstat ldsal lemp ldnpt, statistics(mean median sd min max p5 p95) by(
    yr)

graph box ldsal, over(yr) title("Boxplot of ldsal by Year")
graph export "boxplot_ldsal.png", replace

graph box lemp, over(yr) title("Boxplot of lemp by Year")
graph export "boxplot_lemp.png", replace

graph box ldnpt, over(yr) title("Boxplot of ldnpt by Year")
graph export "boxplot_ldnpt.png", replace
```

- **Summary Statistics by Year:** For each year, detailed summaries (mean, median, standard deviation, minimum, maximum, 5th percentile, and 95th percentile) are computed for `ldsal`, `lemp`, and `ldnpt`.

Table 1: Summary Statistics by Year

Year	Variable	Mean	Median	SD	Min	Max	5th Pctl	95th Pctl
73	ldnpt	4.863000	4.687852	1.924382	-0.208976	10.66034	1.962434	8.160343
	ldsal	6.001581	5.911699	1.797126	-0.857350	11.36805	3.226227	8.849676
	lemp	1.714698	1.656417	1.559433	-2.577022	6.698169	-0.693147	4.227228
78	ldnpt	4.332366	4.100965	2.187433	-1.389284	10.80406	1.071379	8.066000
	ldsal	5.552179	5.394975	1.965634	-0.317301	11.53196	2.383898	8.813699
	lemp	1.262622	1.131402	1.725453	-3.352407	6.732211	-1.382302	4.143135
83	ldnpt	4.515011	4.266926	2.308885	-1.150816	11.06617	0.954807	8.344097
	ldsal	5.520881	5.340935	1.994305	0.587181	11.31305	2.441314	8.778238
	lemp	1.134527	0.955511	1.845955	-3.772261	6.538140	-1.795768	4.244200
88	ldnpt	4.226000	3.836088	2.355197	-1.228260	11.11161	0.648894	8.197926
	ldsal	5.688195	5.348608	2.030207	-0.243908	11.69840	2.710890	9.217271
	lemp	0.963538	0.755403	1.867374	-3.194183	6.64079	-1.795768	4.259859

- **Findings:**

  - **ldnpt:** In 1973 and 1983, the median is around 4.5, while in 1978 it is about 4.0, and in 1988 it is slightly lower at around 3.8–4.0.The mean and median values exhibit a slight downward trend from 1973 to 1988. Although there is a minor rebound in 1983, the overall progression shows a decrease.
  - **ldsal:** This variable shows a relatively stable pattern. Although the mean drops from 6.002 in 1973 to 5.552 in 1978, it then hovers around 5.52-5.69 by 1988. The median values follow a similar pattern. The median values are quite stable across all years, hovering around 5.5, except for a slight decline in 1977 relative to 1973.
  - **lemp:** lemp shows the most pronounced decline. The mean drops steadily from 1.715 in 1973 to 0.964 in 1988, and the median follows suit. The median, 5th percentile, and 95th percentile all show a downward trend over time, with the highest values in 1973 and the lowest in 1988. Moreover, the increasing standard deviation (from 1.559 to 1.867) points to rising dispersion in employment outcomes.

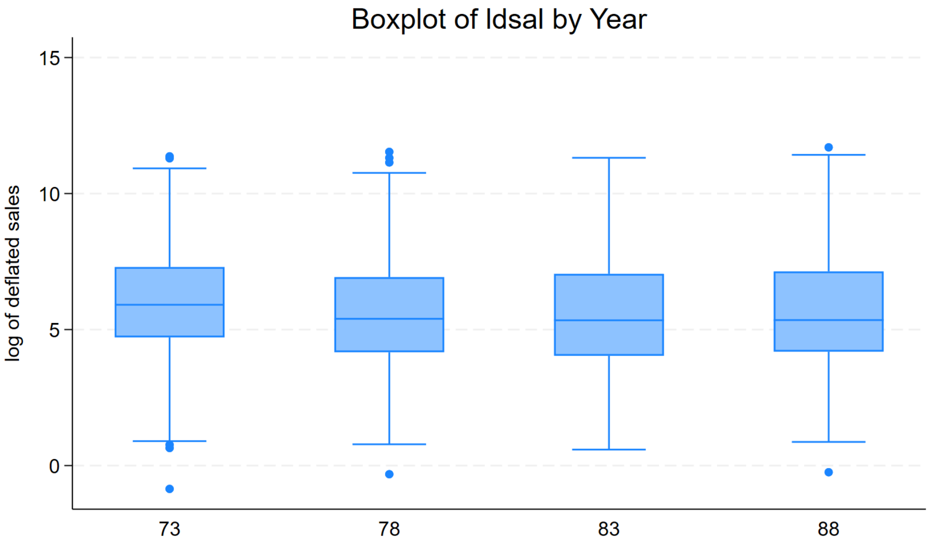
- **Box Plots:** Side-by-side box plots for each variable across the years reveal:
  - For `ldsal`: Median sales dip in the late 1970s and early 1980s then rise by 1988, with an increasing number of high-value outliers.
  - For `lemp`: A downward trend in median employment, indicating that the typical firm employs fewer workers in later years.
  - For `ldnpt`: A similar downward drift in capital, with consistent high-capital outliers.

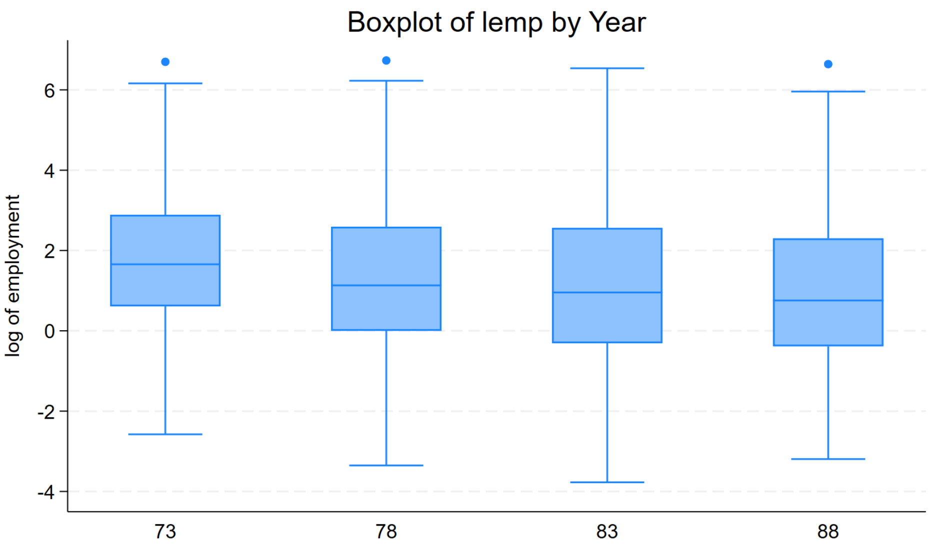
- **Conclusion:** There is no clear monotonic trend in output; however, the distributions of labor and capital suggest a shift towards smaller firms in later years, while a few firms experienced very high sales by 1988.

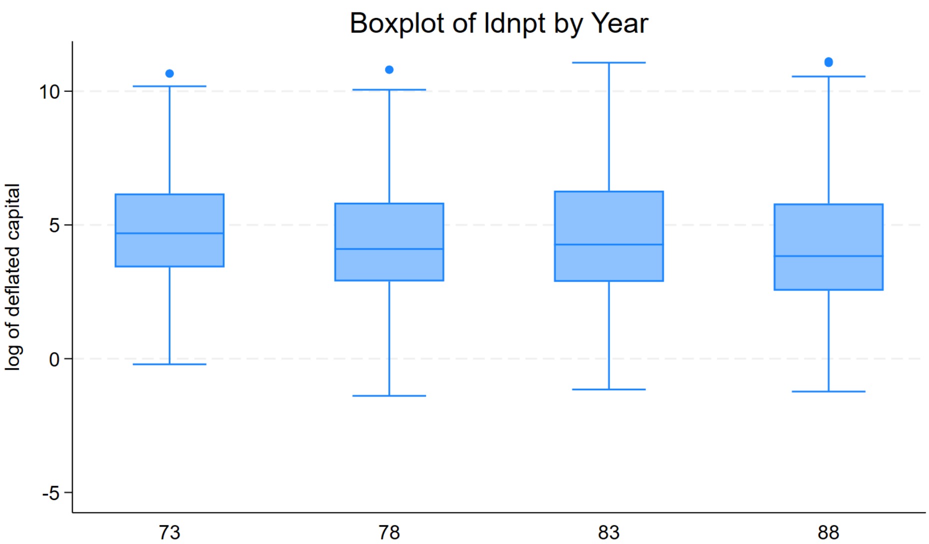
Solution (b).

Number of firms dropped when creating a balanced panel: 1186

```
egen tag = tag(index)
count if tag == 1
local total_firms = r(N)
display "Total number of firms in the unbalanced panel: " `total_firms'

bysort index: egen n_obs = count(yr)

list index n_obs if n_obs < 4, sepby(index)

keep if n_obs == 4

```


Boxplot of ldsal by Year
log of deflated sales
15
10
5
0
73
78
83
88


Boxplot of lemp by Year
log of employment
6
4
2
0
-2
-4
73
78
83
88


Boxplot of ldnpt by Year
log of deflated capital
10
5
0
-5
73
78
83
88

```
egen tag_bal = tag(index)
count if tag_bal == 1
local balanced_firms = r(N)
display "Number of firms in the balanced panel: " `balanced_firms'

local lost_firms = `total_firms' - `balanced_firms'
display "Number of firms dropped when creating a balanced panel: " `lost_firms'
```

Solution (c).

Consider the panel data model

$$\text{ldsal}_{it} = \alpha_i + \beta_1 \,\text{lemp}_{it} + \beta_2 \,\text{ldnpt}_{it} + u_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T,$$

where α_i is an unobserved firm-specific effect and u_{it} is an idiosyncratic error term. In the random effects (RE) framework, α_i is treated as a random variable that is uncorrelated with the regressors lemp_{it} and ldnpt_{it}.

1. Model Reformulation

Define the composite error term as

$$\varepsilon_{it} = \alpha_i + u_{it} - \beta_0.$$

Thus, the model can be rewritten as

$$\text{ldsal}_{it} = \beta_0 + \beta_1 \,\text{lemp}_{it} + \beta_2 \,\text{ldnpt}_{it} + \varepsilon_{it},$$

with β_0 being an overall intercept (if included).

2. Error Structure

Under RE, the variance of the composite error is:

$$\text{Var}(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_u^2,$$

and for two different time periods $t \neq s$ for the same firm, the covariance is:

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_\alpha^2.$$

This intra-firm correlation suggests the use of Generalized Least Squares (GLS).

3. Quasi-Demeaning Transformation

To account for the correlation in the error structure, a quasi-demeaning transformation is applied. For each firm i, define the time averages:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} \text{ldsal}_{it}, \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it},$$

where x_{it} represents each regressor (i.e., lemp_{it} and ldnpt_{it}).
The transformed variables are:

$$y^*_{it} = \text{ldsal}_{it} - \theta\,\bar{y}_i, \quad x^*_{it} = x_{it} - \theta\,\bar{x}_i,$$

with the transformation parameter θ defined as:

$$\theta = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}}.$$

After this transformation, the model becomes:

$$y^*_{it} = \beta_0(1-\theta) + \beta_1\,\text{lemp}^*_{it} + \beta_2\,\text{ldnpt}^*_{it} + \varepsilon^*_{it}.$$

Applying ordinary least squares (OLS) to the transformed model yields the RE estimator:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x^*_{it}x^{*\prime}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} x^*_{it}y^*_{it}\right).$$

4. Assumptions for Consistency

For the RE estimator to be consistent, the following assumptions must hold:

1. **Exogeneity:**

$$E\left[u_{it} \mid \text{lemp}_{i1}, \text{ldnpt}_{i1}, \ldots, \text{lemp}_{iT}, \text{ldnpt}_{iT}, \alpha_i\right] = 0.$$

   Equivalently, the regressors are uncorrelated with both the idiosyncratic error u_{it} and the individual effect α_i.

2. **Independence of Individual Effects and Regressors:**

$$E\left[\alpha_i \mid \text{lemp}_{it}, \text{ldnpt}_{it}\right] = 0 \quad \text{for all } t.$$

3. **Homoskedasticity and No Serial Correlation:** The idiosyncratic errors u_{it} are assumed to be homoskedastic and serially uncorrelated (conditional on α_i). $\mathbb{V}[u_{it}|x_{it}] = \sigma_u^2$, and $\text{Cov}[u_{it}, u_{is}|lemp_{it}, lemp_{is}] = 0$ for $t \neq s$.

4. **Correct Specification:** The model is correctly specified with no omitted variables that are correlated with the regressors.

- The RE estimator is obtained by running a pooled OLS regression where the firm-specific effect α_i is absorbed into the error term.

- **Results (balanced panel):**

Table 2: Question (c): Random Effects Estimator

	(1) RE		(2) RE Robust	
	b	se	b	se
log of employment	0.509	0.037	0.509	0.041
log of deflated capital	0.484	0.029	0.484	0.035
Observations	856		856	

 – β_1 (Labor elasticity): about 0.509.
 – β_2 (Capital elasticity): about 0.484.

- **Interpretation:** A 1% increase in labor is associated with a 0.5087% increase in sales, and a 1% increase in capital with a 0.4839% increase in sales. The sum (approximately 0.989) is close to constant returns.
- **Assumptions:** Consistency requires that the unobserved firm effect α_i is uncorrelated with both `lemp` and `ldnpt`. This is likely violated since more productive firms may choose different input levels.

```
xtset index yr

xtreg ldsal lemp ldnpt, re
xtreg ldsal lemp ldnpt, re vce(robust)
```

Solution (d).

To eliminate the unobserved firm-specific effect α_i, we apply the within (time-demeaning) transformation. Define the time averages for each firm i as:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} \mathrm{ldsal}_{it}, \quad \bar{x}_{i1} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{lemp}_{it}, \quad \bar{x}_{i2} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{ldnpt}_{it},$$

and

$$\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}.$$

Subtracting these averages from the original model gives:

$$\mathrm{ldsal}_{it} - \bar{y}_i = \beta_1\Big(\mathrm{lemp}_{it} - \bar{x}_{i1}\Big) + \beta_2\Big(\mathrm{ldnpt}_{it} - \bar{x}_{i2}\Big) + \Big(u_{it} - \bar{u}_i\Big).$$

Let the demeaned variables be:

$$\tilde{y}_{it} = \mathrm{ldsal}_{it} - \bar{y}_i, \quad \tilde{\mathrm{lemp}}_{it} = \mathrm{lemp}_{it} - \bar{x}_{i1}, \quad \tilde{\mathrm{ldnpt}}_{it} = \mathrm{ldnpt}_{it} - \bar{x}_{i2},$$

and

$$\tilde{u}_{it} = u_{it} - \bar{u}_i.$$

Thus, the transformed model is:

$$\tilde{y}_{it} = \beta_1 \, \tilde{\text{lemp}}_{it} + \beta_2 \, \tilde{\text{ldnpt}}_{it} + \tilde{u}_{it}.$$

The FE within estimator is obtained by applying Ordinary Least Squares (OLS) to the demeaned model:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it} \right),$$

where $\tilde{x}_{it} = \begin{pmatrix} \tilde{\text{lemp}}_{it} \\ \tilde{\text{ldnpt}}_{it} \end{pmatrix}$.

Assumptions for Consistency

For the FE estimator to be consistent, the following assumptions must hold:

1. **Strict Exogeneity:**

$$E\Big(u_{it} \mid \text{lemp}_{i1}, \text{ldnpt}_{i1}, \ldots, \text{lemp}_{iT}, \text{ldnpt}_{iT}, \alpha_i\Big) = 0 \quad \text{for all } t.$$

2. **No Perfect Collinearity:** The demeaned regressors $\tilde{\text{lemp}}_{it}$ and $\tilde{\text{ldnpt}}_{it}$ are not perfectly collinear.

3. **Large Number of Cross-Sections:** Consistency is achieved as $N \to \infty$ while T remains fixed.

Under these conditions, the FE estimator is consistent even if the individual effects α_i are correlated with the regressors in the original model.

- The FE-W estimator is computed by demeaning the data to remove the firm-specific effect α_i.

- **Results (balanced panel):**

Table 3: Question (d): Fixed Effects (Within) Estimator

	(1) FE		(2) FE Robust	
	b	se	b	se
log of employment	0.765	0.062	0.765	0.079
log of deflated capital	0.408	0.047	0.408	0.060
Observations	856		856	

  - β_1: approximately 0.765.

  - β_2: approximately 0.408.

- **Interpretation:** Within a firm, a 1% increase in labor results in about a 0.765% increase in sales, and a 1% increase in capital results in about a 0.408% increase in sales. The sum (approximately 1.17) suggests increasing returns to scale in the within-firm variation.

- **Assumptions:** Requires strict exogeneity of the error term (i.e., no feedback from shocks to future inputs) after controlling for α_i. Although more realistic than RE, this condition may still be violated in practice.

```
xtreg ldsal lemp ldnpt , fe
xtreg ldsal lemp ldnpt , fe vce(robust)
```

Solution (e).
To eliminate the unobserved firm-specific effect α_i, we take first differences. For $t \geq 2$, define the first-differenced variables as:

$$\Delta \text{ldsal}_{it} = \text{ldsal}_{it} - \text{ldsal}_{i,t-1},$$

$$\Delta \text{lemp}_{it} = \text{lemp}_{it} - \text{lemp}_{i,t-1}, \quad \Delta \text{ldnpt}_{it} = \text{ldnpt}_{it} - \text{ldnpt}_{i,t-1},$$

$$\Delta u_{it} = u_{it} - u_{i,t-1}.$$

Then, the differenced model is given by:

$$\Delta \text{ldsal}_{it} = \beta_1 \, \Delta \text{lemp}_{it} + \beta_2 \, \Delta \text{ldnpt}_{it} + \Delta u_{it}.$$

Denote $\Delta y_{it} = \Delta \text{ldsal}_{it}$ and

$$\Delta x_{it} = \begin{pmatrix} \Delta \text{lemp}_{it} \\ \Delta \text{ldnpt}_{it} \end{pmatrix}.$$

The FE First Difference estimator is then:

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it} \, \Delta x_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it} \, \Delta y_{it} \right).$$

For $\hat{\beta}_{FD}$ to be consistent, the following assumptions are required:

1. **Strict Exogeneity in Differences:**

$$E\left(\Delta u_{it} \mid \Delta x_{i1}, \Delta x_{i2}, \ldots, \Delta x_{iT}\right) = 0 \quad \text{for all } t.$$

2. **No Perfect Collinearity:** The differenced regressors must not be perfectly collinear.

3. **Large Number of Cross-Sections:** The estimator is consistent as $N \to \infty$ while T remains fixed.

- The FE-FD estimator is obtained by differencing consecutive observations, eliminating α_i.
- **Results (balanced panel):**

Table 4: Question (e): Fixed Effects (FD) Estimator

	(1)		(2)	
	D_ldsal		D_ldsal	
	b	se	b	se
D_lemp	0.928	0.056	0.928	0.066
D_ldnpt	0.003	0.040	0.003	0.046
Observations	642		642	

  - β_1: approximately 0.928.
  - β_2: approximately 0.003.

- **Interpretation:** Year-to-year changes indicate that a 1% increase in labor is associated with about a 0.928% increase in sales, while a 1% increase in capital is associated with only a 0.003% increase. The large difference in capital elasticity compared to the FE-W estimator may suggest that capital adjustments have a slower or more complex dynamic in the short-run.
- **Assumptions:** Consistency depends on the absence of serial correlation in the errors and that changes in inputs are not influenced by past shocks.

```
use "GMdata_balanced.dta", clear
sort index yr

by index: gen D_ldsal = ldsal - ldsal[_n-1]
by index: gen D_lemp = lemp - lemp[_n-1]
by index: gen D_ldnpt = ldnpt - ldnpt[_n-1]

by index: drop if _n==1

reg D_ldsal D_lemp D_ldnpt, robust
eststo fd_robust

reg D_ldsal D_lemp D_ldnpt, vce(cluster index)
eststo fd_robust_cluster

esttab fd_robust fd_robust_cluster using question_e.tex, replace label
    booktabs title("Question (e): Fixed Effects (FD) Estimator") ///
     cells("b(fmt(3)) se(fmt(3))") keep(D_lemp D_ldnpt)
```

Solution (f).
Define the time-demeaned variables as:

$$\tilde{y}_{it} = y_{it} - \bar{y}_i, \quad \tilde{x}_{it} = x_{it} - \bar{x}_i,$$

with

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}.$$

Then the transformed model becomes:

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{u}_{it}, \quad \text{where} \quad \tilde{u}_{it} = u_{it} - \bar{u}_i.$$

The Fixed Effects (FE) estimator is given by:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{y}_{it}\right).$$

Since $\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{u}_{it}$, we have:

$$\hat{\beta}_{FE} - \beta = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right).$$

Under standard regularity conditions, and with T fixed and $N \to \infty$, it holds that:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}' \xrightarrow{p} Q,$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it} \xrightarrow{d} N(0, \Sigma),$$

where

$$Q = E\left[\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right],$$

and

$$\Sigma = \text{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right).$$

Hence, by Slutsky's theorem,

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right)$$

converges in distribution to:

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) \xrightarrow{d} N\Big(0,\, Q^{-1}\Sigma Q^{-1}\Big).$$

Solution (g).
Recall from the asymptotic distribution of the FE estimator that

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) \xrightarrow{d} N\Big(0,\, Q^{-1}\Sigma Q^{-1}\Big),$$

where

$$Q = \lim_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}', \quad \Sigma = \lim_{N\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{u}_{it}\right).$$

An estimator for the asymptotic variance of $\hat{\beta}_{FE}$ is given by:

$$\widehat{\mathrm{Var}}(\hat{\beta}_{FE}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\hat{u}_{it}^2\right) \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'\right)^{-1},$$

where $\hat{u}_{it}$ are the residuals from the FE regression. Consequently, the standard error for the jth coefficient is given by:

$$\mathrm{se}(\hat{\beta}_j) = \sqrt{\left[\widehat{\mathrm{Var}}(\hat{\beta}_{FE})\right]_{jj}}.$$

This robust variance estimator is used in empirical software to report standard errors that are consistent even under heteroscedasticity and autocorrelation of the error term.

- The robust (clustered) standard errors are computed using the formula

  $$\hat{V} = \widehat{Q}^{-1}\widehat{\Sigma}\,\widehat{Q}^{-1},$$

  where $\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} X_i^{*\prime}\hat{u}_i\hat{u}_i' X_i^{*}$.

- With the balanced panel, the estimated standard errors are approximately:
  - $\mathrm{SE}(\hat{\beta}_1) \approx 0.079$,
  - $\mathrm{SE}(\hat{\beta}_2) \approx 0.060$.

- These values indicate high statistical significance of the estimated coefficients.

```
xtreg ldsal lemp ldnpt , fe robust
```

Solution (h).

Table 5: Question (g): FE Regression with Robust SE

	(1)	
	log of deflated sales	
	b	se
log of employment	0.765	0.079
log of deflated capital	0.408	0.060
Observations	856	

Consider a balanced panel with N firms and T time periods per firm. The Fixed Effects (FE) estimator is obtained by first demeaning the data:

$$\tilde{y}_{it} = y_{it} - \bar{y}_i, \quad \tilde{x}_{it} = x_{it} - \bar{x}_i,$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}.$$

The FE estimator is then given by

$$\hat{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{y}_{it}\right).$$

To account for within-firm correlation and obtain robust standard error estimates, we employ a clustered bootstrap procedure as follows:

1. **Resampling:** For $b = 1, \ldots, B$ (with $B = 1000$), draw a bootstrap sample by resampling N firms with replacement from the original sample. For each selected firm, include all T observations.

2. **Estimation:** Compute the FE estimator on the bootstrap sample to obtain $\hat{\beta}^{*(b)}$.

3. **Variance Estimation:** The bootstrap estimator for the variance of $\hat{\beta}$ is

$$\widehat{\mathrm{Var}}_{\mathrm{boot}}(\hat{\beta}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}^{*(b)} - \bar{\hat{\beta}}^{*}\right)\left(\hat{\beta}^{*(b)} - \bar{\hat{\beta}}^{*}\right)',$$

   where

$$\bar{\hat{\beta}}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^{*(b)}.$$

4. **Standard Errors:** The standard error for the jth coefficient is

$$\mathrm{se}(\hat{\beta}_j) = \sqrt{\left[\widehat{\mathrm{Var}}_{\mathrm{boot}}(\hat{\beta})\right]_{jj}}.$$

This clustered bootstrap procedure yields standard error estimates that are robust to within-cluster (firm) correlation.

Table 6: Question (h): FE Bootstrap Estimates (Manual)

	(1)	
	b	se
_bs_1	0.765	0.080
_bs_2	0.408	0.060
Observations	856	

- The clustered bootstrap resamples firms (clusters) with replacement, preserving the panel structure.
- With $B = 1000$ bootstrap replications, the bootstrap standard errors are found to be:
  - $\text{SE}(\hat{\beta}_1) \approx 0.080$,
  - $\text{SE}(\hat{\beta}_2) \approx 0.060$.
- The bootstrap results confirm the accuracy of the analytical cluster-robust standard errors.

```
use "GMdata_balanced.dta", clear

program define fe_est, rclass
    preserve
    bysort index: egen ymean = mean(ldsal)
    bysort index: egen x1mean = mean(lemp)
    bysort index: egen x2mean = mean(ldnpt)

    gen yd = ldsal - ymean
    gen x1d = lemp - x1mean
    gen x2d = ldnpt - x2mean

    quietly regress yd x1d x2d, noconstant
    return scalar b1 = _b[x1d]
    return scalar b2 = _b[x2d]
    restore
end

bootstrap r(b1) r(b2), reps(1000) cluster(index) nodots: fe_est
```

Solution (i).
Consider the panel data model:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}, \quad i = 1, \ldots, N, \ t \in \mathcal{T}_i,$$

where $\mathcal{T}_i$ denotes the set of time periods for which firm i is observed and $T_i = |\mathcal{T}_i|$.
For each firm i, define the time averages based on the available observations:

$$\bar{y}_i = \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} y_{it}, \quad \bar{x}_i = \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} x_{it}.$$

The within (demeaning) transformation yields:

$$\tilde{y}_{it} = y_{it} - \bar{y}_i, \quad \tilde{x}_{it} = x_{it} - \bar{x}_i, \quad t \in \mathcal{T}_i.$$

The transformed model is:

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{u}_{it}, \quad \text{with} \quad \tilde{u}_{it} = u_{it} - \bar{u}_i,$$

and

$$\bar{u}_i = \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} u_{it}.$$

The Fixed Effects (FE) estimator is then given by:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \tilde{x}_{it} \tilde{y}_{it} \right).$$

Under standard assumptions (strict exogeneity, no perfect collinearity, etc.), $\hat{\beta}_{FE}$ is consistent for β even when the panel is unbalanced.

- The full dataset (unbalanced panel) is used to re-estimate the FE model.
- **Results (unbalanced panel):**

Table 7: Question (i): FE Estimator on Unbalanced Panel

	(1) FE		(2) FE Robust	
	b	se	b	se
log of employment	0.753	0.035	0.753	0.061
log of deflated capital	0.311	0.026	0.311	0.037
Observations	2971		2971	

  - $\beta_1 \approx 0.753$ with a clustered standard error of about 0.061.

  - $\beta_2 \approx 0.311$ with a clustered standard error of about 0.037.

- **Interpretation:** Compared to the balanced panel, the labor coefficient is similar, but the capital coefficient is lower. This may reflect sample selection effects; firms with incomplete data (often newer or smaller firms) exhibit lower capital productivity.

```
use "GMdata.dta", clear
xtset index yr

xtreg ldsal lemp ldnpt , fe

xtreg ldsal lemp ldnpt , fe vce(robust)
```

Solution (j).
Let $\hat{\beta}_{FE}$ be the Fixed Effects (FE) estimator obtained from the original panel data model:

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}, \quad i = 1, \ldots, N, \ t \in \mathcal{T}_i.$$

For each firm i, define the time-demeaned variables as:

$$\tilde{y}_{it} = y_{it} - \bar{y}_i, \quad \tilde{x}_{it} = x_{it} - \bar{x}_i,$$

with

$$\bar{y}_i = \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} y_{it}, \quad \bar{x}_i = \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} x_{it}.$$

The FE estimator is then given by:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \tilde{x}_{it} \tilde{y}_{it} \right).$$

To account for within-firm correlation, we use a clustered bootstrap procedure:

1. **Resampling:** For $b = 1, \ldots, B$ (with $B = 1000$), draw a bootstrap sample by sampling N firms with replacement from the original data. For each selected firm, include all T_i observations.

2. **Estimation:** Compute the FE estimator on the bth bootstrap sample to obtain $\hat{\beta}_{FE}^{*(b)}$.

3. **Bootstrap Variance:** The variance of $\hat{\beta}_{FE}$ is estimated by

$$\widehat{\mathrm{Var}}_{\mathrm{boot}}(\hat{\beta}_{FE}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\beta}_{FE}^{*(b)} - \bar{\hat{\beta}}^* \right) \left(\hat{\beta}_{FE}^{*(b)} - \bar{\hat{\beta}}^* \right)',$$

where

$$\bar{\hat{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{FE}^{*(b)}.$$

4. **Standard Errors:** The standard error for the jth coefficient is given by

$$\text{se}(\hat{\beta}_j) = \sqrt{\left[\widehat{\text{Var}}_{\text{boot}}(\hat{\beta}_{FE})\right]_{jj}}.$$

This clustered bootstrap method provides standard error estimates that are robust to intra-cluster (firm-level) correlation.

Table 8: Question (j): FE Bootstrap on Unbalanced Panel (Manual)

	(1)	
	b	se
_bs_1	0.753	0.058
_bs_2	0.311	0.037
Observations	2971	

- The clustered bootstrap is also applied to the unbalanced panel.
- With $B = 1000$ replications, the bootstrapped standard errors are approximately:
  - $\text{SE}(\hat{\beta}_1) \approx 0.058$,
  - $\text{SE}(\hat{\beta}_2) \approx 0.037$.
- These values closely match the analytical robust standard errors, reinforcing the robustness of our estimates.

```
use "GMdata.dta", clear
xtset index yr

program define fe_est2, rclass
    preserve
    bysort index: egen ymean2 = mean(ldsal)
    bysort index: egen x1mean2 = mean(lemp)
    bysort index: egen x2mean2 = mean(ldnpt)

    gen yd2 = ldsal - ymean2
    gen x1d2 = lemp - x1mean2
    gen x2d2 = ldnpt - x2mean2

    quietly regress yd2 x1d2 x2d2, noconstant
    return scalar b1 = _b[x1d2]
```

```
    return scalar b2 = _b[x2d2]
    restore
end

bootstrap r(b1) r(b2), reps(1000) cluster(index) nodots: fe_est2
```