Taewon Hwang(2020711771)
Industrial Organization 1
Final Exam

Question 1

The data used in this practice contain information on sales and input factors of U.S. R&D performing firms. Observations are taken at a 5-year interval, starting from 1973 and ending in 1988. A total of 1,400 firms are investigated, and amongst these firms, 121 different items, distinguished by a 3 digit SIC code, are produced. Variables included in the dataset regard sales and conventional input factors, namely employment and capital, and also include R&D, R&D capital, and investment, allowing the investigation of R&D investment behavior of firms of different products over a 20-year time period. Note that all values (excluding employment) are deflated to a base year. The unit of raw values of employment is thousands, the others are in million USD; the dataset is composed of log transformed variables.

A total of 2,971 observations across 1,400 firms are provided in the full panel. Corresponding summary statistics are given in **Table 1**. However, if we drop firms that having 'missing' observations, that is, firms with less than 4 observations, the number of observations is greatly reduced to 856, with only 214 firms left. Summary statistics of this balanced panel are given in **Table 2**. Note that 'balanced' corresponds to those firms that have existed during all 4 points of observation. Finally, summary statistics for firms that have existed during at least two of the points of observation are given in **Table 3**.

Table 1
Summary Statistics (Full Panel)

	Mean	S.D.	Min.	Median	Max.
log(Sales)	5.67	1.96	-0.86	5.53	11.70
log(Employment)	1.26	1.78	-3.77	1.11	6.73
log(Capital)	4.47	2.22	-1.39	4.21	11.11
log(R&D capital)	3.40	2.03	-4.29	3.18	9.97
log(R&D)	1.79	2.05	-5.31	1.63	8.43
log(Investment)	2.67	2.17	-3.84	2.51	8.99
Observations	2971				

Table 2
Summary Statistics (Balanced Panel)

	Mean	S.D.	Min.	Median	Max.
log(Sales)	6.91	1.84	1.66	7.06	11.70
log(Employment)	2.41	1.62	-2.07	2.65	6.73
log(Capital)	5.92	2.06	0.81	6.03	11.11
log(R&D capital)	4.89	1.93	0.06	5.11	9.97
log(R&D)	3.22	1.99	-2.71	3.40	8.43
log(Investment)	4.07	2.06	-2.08	4.23	8.89
Observations	856				

Table 3
Summary Statistics (2-Year Panel)

	Mean	S.D.	Min.	Median	Max.
log(Sales)	5.92	1.94	-0.86	5.80	11.70
log(Employment)	1.51	1.73	-3.77	1.35	6.73
log(Capital)	4.77	2.18	-1.39	4.57	11.11
log(R&D capital)	3.67	2.01	-2.61	3.46	9.97
log(R&D)	2.02	2.05	-4.31	1.85	8.43
log(Investment)	2.93	2.17	-3.84	2.77	8.89
Observations	2440				

Changes in summary statistics due to restrictions on number of observations are generally consistent with intuition when one assumes that dropping out of the panel signifies exiting the market - perhaps due to incompetence.

As the panel is modified from the full panel to a panel with at least 2 observations per firm to a panel with observations for all periods per firm, it can be seen that the mean values of variables are changing in a consistent direction. Indeed, increased volumes in sales, employment, capital, R&D, investment etc. may likely indicate less probability of failure in the market. A point to address is the significantly higher mean and median values of R&D, R&D capital, and investment of firms that have stayed in the market over the full period of interest (balanced panel) compared to those of the full dataset.

Question 2

In this section, the balanced panel is used in analysis. This implies that firms that have exited the market due to poor performance etc., are not represented, potentially inducing bias in the estimators.

(i-a) Production Function Estimation

The model used in estimation of the production function is as follows:

$$y_{it} = \beta_1 l_{it} + \beta_2 k_{it} + \beta_3 g_{it} + d_t + d_{t,357} + \alpha_i + \omega_{it} \tag{1}$$

where d_t denotes time fixed effects and $d_{t,357}$'s are time dummy variables for item 357 (computer). Results of pooled OLS regression (column (1)), fixed effects regression (column (2)), and random effects regression (column (3)) are given in **Table 4**. Note that robust standard errors are implemented via the argument `vce(cluster index)` in order to control for correlation between observations within a single firm. Further, as all variables - both RHS and LHS - are provided in log transformed values, coefficients should be interpreted accordingly as a percentage change in sales being correlated with a unit percentage change in the independent variable.

(i-b) Hausman Test

The Hausman test for model specification is implemented to investigate whether there is a statistically significant difference between the fixed and random effects estimators, where the null hypothesis is that there is no systematic difference between the two. That is, both fixed effects estimators and random effects are consistent. Specifically, in this practice, a modified version of the Hausman test is implemented using the function `rhausman`. This executes a cluster-robust version of the Hausman test which is efficient under less demanding assumptions regarding efficiency of the tested models.

The resulting cluster-robust Hausman test statistics are 35.02. Following a χ^2 distribution with d.f. 9, the calculated p-values are 0.001, indicating that the null hypothesis can be rejected at a 5% confidence level, thus confirming the inconsistency of the random effects estimators.

Table 4
Panel Model Regressions of log Sales (Balanced Panel)

	(1) Pooled OLS	(2) FE	(3) RE
log(Employment)	0.496***	0.685***	0.598***
	(0.04)	(0.04)	(0.04)
log(Capital)	0.460***	0.180***	0.335***
	(0.03)	(0.04)	(0.03)
log(R&D capital)	0.0335	0.0989*	0.0645**
	(0.02)	(0.04)	(0.02)
Year 1973	-0.163***	-0.244***	-0.196***
	(0.03)	(0.03)	(0.03)
Year 1978	-0.150***	-0.196***	-0.169***
	(0.02)	(0.02)	(0.02)
Year 1983	-0.288***	-0.255***	-0.272***
	(0.02)	(0.02)	(0.02)
Year 1973 & Computer	-3.235***	-3.421***	-3.295***
	(0.08)	(0.06)	(0.07)
Year 1978 & Computer	-2.064***	-2.286***	-2.143***
	(0.07)	(0.04)	(0.06)
Year 1983 & Computer	-0.689***	-0.944***	-0.791***
	(0.07)	(0.05)	(0.07)
Year 1988 & Computer	0.240***		0.144**
	(0.06)		(0.05)
Constant	3.050***	3.961***	3.404***
	(0.11)	(0.17)	(0.11)
Observations	856	856	856
R^2	0.97	0.93	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(ii) Productivity

In this section, autocorrelation between OLS residuals are investigated. While a multitude of tests for serial correlation can be suggested, I take an intuitive approach by regressing the OLS residuals on their $(t-1)$ and $(t-2)$ lagged values. Note that $(t-1)$ in this dataset refers to values generated from the preceding observation (if available), which would be 5 years earlier. The equation used in estimation is given below, and results of linear regression are given in **Table 5**.

$$\hat{\omega}_{i,t} = \rho\hat{\omega}_{i,(t-j)} + \epsilon_{i,t} \tag{2}$$

where $j \in [1, 2]$ denotes 5-year lagged values and 10-year lagged values, respectively.

Results in **Table 5** show that there clearly exist significant autocorrelation in the residuals, suggesting inconsistency and bias in the OLS estimators. Note that the magnitude of correlation decreases for $(t-2)$ lagged values, which agrees with conventional ideas on serial correlation.

Table 5
Linear Regressions of OLS Residuals (Balanced Panel)

	(1)	(2)
5 Year Lag	0.772***	
	(0.02)	
10 Year Lag		0.615***
		(0.03)
Observations	642	428

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(iii) Firm Heterogeneity

Results from sections **(i)** and **(ii)** suggest that both the pooled OLS estimators and random effects estimators are inconsistent. **(ii)** suggests that this inconsistency is caused by heterogeneity between firms, which must be controlled for in order to execute robust estimation of coefficients and further inference.

Question 3

In this section, the full panel and 2-year panel are used in analysis. Using these panels, the market exiting behavior of firms can also be witnessed and modeled into estimation, allowing the removal of potential bias generated by thriving firms. Also, as requested in the following sections, these panels allow the investigation of conditions in which firms choose to exit the market.

(i) Selection

Equation (1) is estimated using the full panel and 2-year panel, and results are given in **Table 6**. Pooled OLS estimates are provided for both panels (columns (1)) and (4)), and first differences estimates are provided for the full panel (column (2)).

Here, I investigate the effects of using full, 2-year, and balanced panels on estimators. Note that it can be perceived that balanced panels are highly selective, as only data on the survivors over a 20-year period are used. The 2-year panel is less so, and the full panel can be seen as not selective.

Since evidence of serial correlation can be found in the data, the OLS estimators are likely to be inconsistent, and should be disregarded. For a rigorous investigation, a comparison between first differenced model estimators (**Table 6** column (2)) and fixed effects estimators (**Table 4** column (2)) may be most suitable. Nevertheless, for a more intuitively tasteful blueprint, let us first compare the three pooled OLS estimators.

It is clear that a move to a more selective panel corresponds to a smaller impact of labor inputs, and increasing impact of capital inputs on sales. From these results, one may deduce that firms that fail to survive have insufficient capital stock and could be dependent on labor inputs. These results can also agree with the change in behavior of these variables between correctly specified models.

The correlation of R/D capital on sales or survival seem *ad hoc* in the pooled OLS models, but ultimately is statistically significant in the correctly specified model using the balanced panel, which may indicate that a lack of R&D may be a contributing factor to exiting the market.

(ii) Survival

In columns (3) and (5) of **Table 6**, the predicted probability of survival is included to the model as an independent variable. The predicted probability is calculated by first constructing a variable which signifies if the firm shall exist in the next period, that is, if an observation is recorded for $(t+1)$. Note that jumps between time periods were treated as still being extant in the market, and observations from 1988 were omitted, as existence

in the next period is unknown.

Next, a Probit model is implemented so that existence, as generated above, is modeled as a function of k, g, and i. Using the obtained parameters, one may predict the fitted values $\hat{p}_{it}$, which give the predicted probability of survival of firm i in time $(t+1)$. Note that predicted probabilities of survival were calculated separately for respective panels.

As can be observed in columns (3) and (5), in both panels, probability of survival has a statistically significant positive correlation with sales. This, too, agrees well with our general line of thought; a firm with a higher probability of continuing business may exhibit a larger sales volume.

Table 6
Regressions of log (differenced) Sales (Full Panel, 2-Year Panel)

	Full Panel			2-Year Panel	
	(1) Pooled OLS	(2) First Diff.	(3) First Diff. Survival	(4) Pooled OLS	(5) Pooled OLS Survival
log(Employment)	0.578***			0.541***	0.512***
	(0.02)			(0.02)	(0.02)
log(Capital)	0.372***			0.392***	0.424***
	(0.02)			(0.02)	(0.02)
log(R&D capital)	0.0380***			0.0534***	0.0158
	(0.01)			(0.01)	(0.01)
D.log(Employment)		0.740***	0.737***		
		(0.03)	(0.03)		
D.log(Capital)		0.116***	0.117***		
		(0.02)	(0.02)		
D.log(R&D capital)		0.0414	0.0391		
		(0.02)	(0.02)		
Pr(Survival)			0.153***		1.275***
			(0.04)		(0.18)
Year 1973	-0.192***			-0.181***	-0.170***
	(0.02)			(0.02)	(0.02)
Year 1978	-0.169***	-0.212***	-0.209***	-0.149***	-0.134***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Year 1983	-0.267***	-0.289***	-0.281***	-0.267***	-0.219***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Year 1973 & Computer	-3.211***			-3.221***	-3.246***
	(0.05)			(0.05)	(0.06)
Year 1978 & Computer	-1.973***	1.134***	1.126***	-2.025***	-2.079***
	(0.05)	(0.03)	(0.03)	(0.05)	(0.05)
Year 1983 & Computer	-0.689***	1.318***	1.308***	-0.683***	-0.750***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
Year 1988 & Computer	0.466***	1.064***	1.056***	0.343***	0.321***
	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)
Constant	3.365***	0.258***	0.145***	3.264***	2.237***
	(0.06)	(0.01)	(0.04)	(0.07)	(0.16)
Observations	2971	1502	1502	2440	2440
R^2	0.96	0.85	0.85	0.96	0.97

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Question 4

In this section, the full panel is used for analysis, as the OP(1996) estimator places a strong emphasis on the inclusion of exiting firms in estimation.

(i)

Table 7 reports coefficients of l and dummy variables of a regression of y on l, dummy variables, and a second order polynomial in k, g, and i. The equation used in estimation is as follows:

$$y_{it} = \beta_1 l_{it} + d_t + d_{t,357} + \phi(k_{it}, g_{it}, i_{it}) + \alpha_i + \omega_{it} \tag{3}$$

where $\phi(k_{it}, g_{it}, i_{it})$ denotes a second order polynomial in k, g, and i.

Table 7
Regression of log Sales on second order polynomial of k, g, i (Full Panel)

	coef.	s.e.
log(Employment)	0.584***	(0.02)
Year 1973	-0.169***	(0.02)
Year 1978	-0.153***	(0.02)
Year 1983	-0.220***	(0.02)
Year 1973 & Computer	-3.245***	(0.05)
Year 1978 & Computer	-2.037***	(0.05)
Year 1983 & Computer	-0.757***	(0.05)
Year 1988 & Computer	0.408***	(0.05)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(ii)

Table 8 reports coefficients of a regression of ϕ on k and g, where ϕ denotes the value of the polynomial computed in **(i)**.

Table 8
Regression of fitted log Sales on k, g (Full Panel)

	coef.	s.e.
log(Capital)	0.718***	(0.01)
log(R&D capital)	0.154***	(0.01)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(iii)

The construction of predicted probability of survival follows the steps suggested in **Question 3 (ii)**.

(iv)

Table 9 reports coefficients of a nonlinear squares model for the second stage. That is, $y_{it} - \hat{\beta}_1 l_{it} - \hat{d}_t - \hat{d}_{t,357}$ is regressed on k, g, h, h^2. Column (2) of **Table 9** includes a second order polynomial in both h and predicted probability of survival. The equations used in estimation for respective columns are as follows:

$$y_{it} - \hat{\beta}_1 l_{it} - \hat{d}_t - \hat{d}_{t,357} = f_1(k_{it}, g_{it}, h_{it}, h_{it}^2) \tag{4}$$

$$y_{it} - \hat{\beta}_1 l_{it} - \hat{d}_t - \hat{d}_{t,357} = f_2(k_{it}, g_{it}, h_{it}, h_{it}^2, \hat{P}_{it}, \hat{P}_{it}^2, h\hat{P}_{it}) \tag{5}$$

Table 9
Nonlinear squares estimation of fitted log Sales (Full Panel)

	(1)		(2)	
	coef.	s.e.	coef.	s.e.
Constant	-0.305***	(0.03)	-0.440*	(0.21)
log(Capital)	0.365***	(0.01)	0.378***	(0.01)
log(R&D capital)	0.0403***	(0.01)	-0.0610***	(0.02)
H	0.0574*	(0.03)	0.0792	(0.09)
H^2	0.0265*	(0.01)	-0.00300	(0.01)
Pr(Survival)			-0.418	(0.62)
$Pr(Survival)^2$			1.435**	(0.49)
H×Pr(Survival)			-0.115	(0.12)
N	2971		2971	
R^2	0.84		0.84	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Olley, G. S., Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. Econometrica, 64 (6), 1263-1298.