Question 1

The unbalanced panel for `sic3 == 262` contains 48 observations; the balanced panel (firms that have data across all 4 years) contain 28 observations. There are only 7 firms that constitute the balanced panel.

The balanced panel has higher mean of all relevant variables:

Table 1: Mean Values of Outputs, Inputs, and Investments

	Unbalanced Panel (obs = 48)	Balanced Panel (obs = 28)
ldsal	7.37	7.82
lemp	2.56	3.03
ldnpt	7.13	7.69
ldrst	4.21	4.83
ldrnd	2.48	3.10
ldinv	4.82	5.40

Question 2: Balanced vs Unbalanced Panel

After dropping all observations where `sic3 == 357`, I create the relevant panel variable with `xtset index yr` where `index` is to identify individual firm effects and `yr` year of observation.

The results of the regression `xtreg ldsal ldnpt ldrst lemp i.yr, fe vce(robust)` are shown below, first for the unbalanced panel and then for the balanced panel:

```
                         Robust
    ldsal       Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]

    ldnpt     .1395981   .0210137     6.64    0.000     .0983734    .1808228
    ldrst     .0503836   .0217921     2.31    0.021     .0076319    .0931353
     lemp     .7345001    .026306    27.92    0.000     .6828928    .7861073

       yr
       78     .0383085   .0093939     4.08    0.000     .0198796    .0567374
       83     .0085651   .0161885     0.53    0.597    -.0231935    .0403237
       88     .2687079   .0197464    13.61    0.000     .2299693    .3074464

    _cons    3.941211    .0739566    53.29    0.000     3.796122    4.086299

  sigma_u    .4632147
  sigma_e   .17067234
      rho   .88047018   (fraction of variance due to u_i)
```

```
                        Robust
    ldsal |      Coef.    Std. Err.      t     P>|t|      [95% Conf. Interval]
----------+----------------------------------------------------------------
    ldnpt |   .1845772    .0400649     4.61   0.000      .1055805    .2635739
    ldrst |   .1024187    .0456585     2.24   0.026      .0123931    .1924444
     lemp |   .6775737    .0428838    15.80   0.000       .593019    .7621285
          |
       yr |
       78 |   .0478098    .0127993     3.74   0.000      .0225732    .0730464
       83 |  -.0134817    .0256455    -0.53   0.600     -.0640474    .0370839
       88 |   .2406897    .0294128     8.18   0.000       .182696    .2986835
          |
    _cons |   3.692007    .1691336    21.83   0.000      3.358523    4.025491
----------+----------------------------------------------------------------
  sigma_u |  .39110479
  sigma_e |  .16355168
      rho |  .85115554    (fraction of variance due to u_i)
```

The estimated coefficients of capital and R&D capital from the balanced panel regression are higher. This the balanced panel is by definition composed of firms that did not exit; they may be the most productive firms, who also benefit more from renting capital. This might explain why we observe these higher coefficients.

The coeffcient of labor is, however, lower in the balanced than the unbalanced panel. This may be because more capital-intensive firms are inherently more productive, and therefore require less labor. However, the difference is not extremely large in magnitude, but is very statistically significant.

This *might* be explained from productivity as well: if firms higher very productive workers, they may require fewer employees. There is not enough evidence to say that this is indeed the reason though.

Question 3: Survival

I create an exit dummy by sorting firms by index and, if a firm is not present in the time period $t + 1$, then exit $= 1$ in time period t for that firm. 1988, as the last year observed in the dataset, is left as NA, since we do not know if firms exit or stay after this of course. In particular note that this is *exit probability*, not *survival probability*: so it is by definition equal to one minus survival probability.

```
gen exit = 0 if yr != 88
bysort index: replace exit = 1 if yr[_n+1] != yr[_n] + 5 & yr[_n] != 88
```

I then run a probit regression as outlined in the problem set, from which I find the following coefficients:

Table 2: Output of Probit Regression of Exit ~ Labor, Capital, R&D Capital and Investment

Variable	Coefficient
lemp	-0.019 (n.s)
ldnpt	0.167
ldrst	-0.180
ldinv	-0.175
const	-0.108 (n.s.)

All coefficients are significant at the 5% level, except for those marked `(n.s)`.

From this simple regression, the results seem strange: more capital is associated with a higher probability that a firm exits in the next period. However, this is a biased estimate: precisely from this higher spending on capital or labor, a firm may have to exit if its sales are too low. Firms with very high sales can afford to spend more on these input costs. As a firm's sales decline, holding capital and labor fixed, the costs of these inputs may force the firm to exit.

The coefficients of investment and R&D capital are more intuitive. Those firms that find it a worthwhile idea to invest heavily or to spend on R&D probably have a higher probability of surviving. In addition, R&D may increase the demand for a firm's product through product improvements, or allow the firm to produce at a lower cost. Both of these mechanisms would reduce the probability of exit in the next period.

Question 4: Estimating the Production Function

I begin with the production function:

$$y_{it} = \beta^c ldnpt + \beta^{rc} ldrst + \beta^l lemp + \omega_{it} + e_{it}$$

First Stage

I estimate a first-stage equation using a third-order polynomial as outlined in the problem set, with help from the document `statahints.docsx`. This stage is done to allow the production function to be estimated non-parametrically, since we do not know the actual form. We don't know the form of the equation because there are some things which we cannot observe, that only firms observe. Secondly, we want to estimate productivity, and we will use the predicted values from this nonparametric regression in order to identify productivity.

Second Stage

I follow the assumption that productivity is given by:

$$\omega_{it} = \rho \omega_{it-1} + \xi_{it}$$

and use the moment condition:

$$E[\phi_{it+1} - \beta_0 - \beta_1^k k_{it+1} - \beta_2 k_{it+1}^{R\&D} - \beta_3^l l_{it+1} - \rho(\phi_{it} - \beta_0 - \beta_1^k k_{it} - \beta_2 k_{it}^{R\&D} - \beta_3^l l_{it})] \otimes \begin{bmatrix} k_{it+1} \\ k_{it+1}^{R\&D} \\ l_{it} \\ l_{it+1} \\ \phi \end{bmatrix}$$

Third Stage

I now use the moment condition:

$$E[\phi_{it+1} - \beta_0 - \beta_1^k k_{it+1} - \beta_2 k_{it+1}^{R\&D} - \beta_3^l l_{it+1} - \rho(\phi_{it} - \beta_0 - \beta_1^k k_{it} - \beta_2 k_{it}^{R\&D} - \beta_3^l l_{it} - (1-p))] \otimes \begin{bmatrix} k_{it+1} \\ k_{it+1}^{R\&D} \\ l_{it} \\ l_{it+1} \\ \phi \end{bmatrix}$$

Where $1 - p$ is the probability of survival, since I define p to be the probability of exit. Adjusting for probability of survival significantly changes the estimate of firm productivity. Before adjusting for this probability, the GMM estimation gives:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \rho \end{pmatrix} = \begin{pmatrix} 2.978 \\ 0.313 \\ 0.144 \\ 0.608 \\ 0.808 \end{pmatrix}$$

After, these estimated coefficients become:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \rho \end{pmatrix} = \begin{pmatrix} 3.88 \\ 0.290 \\ 0.263 \\ 0.649 \\ 0.463 \end{pmatrix}$$

Question 5: Paper Mill Productivity

I find an average productivity in the Paper Mill sector to be 2.54. Productivity remains relatively constant over time:

year	mean productivity of 12 paper mill firms
73	2.57
78	2.55
83	2.46
88	2.58

mean productivity of 12 firms in Paper Mill sector