Assignment: Theory and Linear Models

How theory informs research design and analysis.

Stijn Masschelein

Table of contents

The models

A lot of theories in accounting and finance are about information that is being revealed for strategic reasons and how investors react to that information. The models about strategic disclosure I use for the assignment come from the book 'Games and Information' by Eric Rasmusen which is the best introduction to Game Theory that I am aware of.

In this assignment, I will explain two simple models of how companies can reveal information about their own performance to investors. These questions are relevant to voluntary disclosure problems in accounting and a lot of research questions on corporate social responsibility. They provide a toy world to think about the typical research designs in a lot of accounting and finance research.

The goal of introducing these models is to make them less scary. They are not that hard to understand at the conceptual level but they are often couched in advanced mathematics in research papers. Fortunately, Rasmusen does an excellent job explaining these models by using simple versions that reveal the key issues. These little games are really useful because they reveal unexpected implications of some assumptions that are not necessarily clear from verbal explanations. For instance, you might think that if a certain disclosure is 'cheap talk' that it is useless. That is not necessarily true! And I will show this in these assignments.

Second, the results of these models are relatively easy to simulate in `R`. This means that we can simulate data and run regressions on the data. This can help you to understand which variables you will need to collect, which tests are appropriate, what the underlying assumptions are of your test, and how you should interpret the data. The assignment questions are aimed at helping you through the simulations and the regression models.

Terminology

I need to explain some terminology before we dive into the models. Agency theory games deal with *players* in a game. In accounting and finance, we often deal with two players: firms and investors. *Nature* is a special, extra player that makes decisions that are beyond the control of any players that we are interested in. The actions of Nature can be thought of as random variation or uncertainty. These players can take some simple *actions* which we will represent by numbers. For instance, a firm can decide to disclose some information (disclosure = 1) or not (disclosure = 0). The actions taken by the players give rise to *payoffs* or outcomes and the players try to maximise their own payoff taking into account the other players' actions. A stable outcome is called the *equilibrium* when no player can improve their payoffs by changing

their own actions. There are often multiple equilibria for a game but not all of them are equally feasible or interesting.

For each model that I present, I will explain who the players are, which actions they take and the order of these actions, the payoffs, and finally the equilibrium that I am interested in.

Signalling

The first model that I want to present is a model where a firm can decide to hire a Big4 auditor to certify their financial statements. The Big4 auditor is more expensive and especially so when financial performance is bad. In such a situation hiring a Big4 auditor can be a signal that the firm has good underlying financial performance.

Players

One firm and two investors.

The order of play

0. Nature chooses the firm's performance $p \in \{1, 4\}$. Performance is observed by the firm, not by the investors.
1. The firm chooses a Big4 auditor or not $a \in \{0, 1\}$
2. The investors bid for the company $b(a)$ taking into account the audit choice.
3. The firm accepts the bid or rejects it.
4. The output equals p

Payoffs

$$\pi_{firm} = \begin{cases} b - 8a/p & \text{if the firm accepts a bid} \\ 0 & \text{if the firm rejects all bids} \end{cases}$$

$$\pi_{investor} = \begin{cases} p - b & \text{for the investor whose bid is accepted} \\ 0 & \text{for the other investor} \end{cases}$$

Separating Equilibrium

One equilibrium is where all high performing firms ($p = 4$) hire a Big4 auditor ($a = 1$) and the low performing firms ($p = 1$) do not hire a Big4 auditor ($a = 0$). In such a scenario, investors will bid 4 for the firms with a Big4 auditor ($b(1) = 4$) and 1 for firms without a Big4 auditor ($b(0) = 1$). I assume that investors will always bid their payoff to 0 because there is competition between investors, i.e. we have a competitive financial market.

A high performing firm that does not hire a Big4 auditor will receive a lower bid ($b(0) - b(1) = -3$) and the cost savings will not compensate for that loss ($0 - 8/4 = -2$). A low performing firm that hires a Big4 auditor will receive a higher bid ($b(1) - b(0) = 3$) but the extra cost outweigh ($8/1 - 0 = 8$) this benefit. As a result, neither the high performing firm nor the low performing firm have an incentive to deviate from the optimal decision.

Key Assumptions

The key assumption in the model is that hiring a Big4 auditor is costly and more costly if the firm has bad financial performance. More specifically, the cost for the low performing firm must be larger than the additional benefit from getting a higher bid from the investors. Otherwise, the low performing firms might think it worthwhile to signal high performance by hiring the auditor. At that point, the investors could no longer distinguish between high and low performing firms and they would bid the average performance between the firms.

Another assumption is that the payoff of the investors is dependent on p. This means that the investors care about the performance of the firm. On the other hand, the investor's only care about the choice of auditor because it is a signal not because it affects the actual performance of the firm.

Applications

One interpretation of this model is that p is the unobserved economic value of the firm and b is the market value. The firm's can signal that they are highly valuable by paying for a costly Big4 auditor. This would mean that in equilibrium the companies with a Big4 auditor ($a = 1$) have a higher market value ($b = 4$).

These signalling models have a number of other applications. You can think of some forms of advertising as a game between a firm and a customer where advertising is a costly signal of the quality of the firm's products. You can also analyse dividends as a costly signal (i.e. less cash for the company) of the health of the company. In the assignment questions, we will interpret this model as a firm's choice to engage in corporate social responsibility (CSR) activities. That is, one model of CSR activities is that it is costly but only firms with good financial performance are able to perform these activities at a low enough cost.

Cheap Talk

The second example is a model of cheap talk. Cheap talk is often used as a term when firms can disclose information but there is no way to check whether the information is correct. The implicit implication is often that investors should not believe that information. However, the following model shows that if investors and firms both care about whether investors have the correct belief, that this implicit assumption is wrong.

The game is very simple, the firm discloses a report, r, about their performance, p, where both r and p can be any number between 0 and 10. The complication is in the payoffs which I will explain further on.

Players

A firm and an investor

Order of Play

0. Nature chooses firm performance as $p \sim \mathrm{U}[0, 10]$.
1. The firm chooses a report $r \in [0, 10]$.
2. The investor chooses to believe performance is b.

Payoffs

$$\pi_{firm} = c - (b - [p + 1])^2$$

$$\pi_{investor} = c - (b - p)^2$$

c is just a constant to make sure that firms and investors have a positive payoff. In Game Theory terms: c must be high enough so that firms are willing to play the game. It does not play any role in the actual game.

The interpretation of the payoffs is that the firm wants the investor to believe that they are doing a little better than they are actually doing but not by too much. The investor wants to be correct in their assessment of the performance of the firm. You can interpret this as follows. If the performance is going to be bad, the investor wants to short the company; if performance is going to be good, the investor wants to buy the company. Either way, they are making money. The firm's payoff can be interpreted as them preferring that the investors buy the company but if they misrepresent their financial situation too much, regulators are going to punish them in the long run. An additional interpretation is that the more firms misrepresent the more likely they are to be detected for misreporting.

Pooling equilibrium: Everybody Lies

All firms prepare a report that says $r = 10$, all investors believe that $b = 5$. In this equilibrium, the disclosure is not informative and investors will ignore it. Investors will just believe the expected value of the firm performance before they received the report. Firms always pretend that they are doing excellent. This is the typical equilibrium that people have in mind when they say that firm disclosures are just cheap talk. An alternative, but very similar equilibrium is that firms just report a random number without any connection to the performance. The investors will respond in the same by just ignoring the report.

Partial pooling equilibrium

If $p \in [0, 3]$, choose a report that says $r = 0$, if $p > 3$, choose a report that says $p = 10$.

If $r = 0$, the investors believe that $b = 1.5$. If $r = 10$, the investors believe that $b = 6.5$.

The details are .

Key Assumptions

The key assumption is that the payoffs of the investor and the firm partly overlap. The firm wants the investors to be roughly correct in their belief. That is, there is no complete conflict of interest between investors and firms.

Interpretation

The interpretation of the partial pooling equilibrium is the most interesting and I am going to focus on this equilibrium. First of all, notice that the separation between low performing firms and the other firms works because the low performers would misreport by too much, i.e. the investor would believe $b = 6.5$ while the true performance is $p \leq 3$. At best the firm's payoff would be $\pi_{firm} = c - (2.5)^2$.

One interpretation is that b represents the stock price of the firm. If investor's belief, b, is their best estimate of the value of the firm that is what they are willing to pay for it. The amount of money they make is not necessarily given by whether they pay a lot for the stock or not but by how correct they are in their belief. For instance, if they short the stock and it turns out the company was more valuable than b, they lose money. If they buy the stock and it turns out that the value of the firm was lower, they also lose money.

It might sound weird that the firm itself might want the stock price to be not too different from the true value. However, in a strong regulatory environment, firm's might risk legal action if the stock price is too high and they did not warn investors enough about the potential

overvaluation. Even the previously (and now again) richest man in the world once said that his largest asset was probably overvalued.

> Tesla stock price is too high imo.

Another reason why firms might care about roughly conveying the right information to investors is that disclosure is a repeated game. If the firm exaggerates their performance by too much, the investors will eventually find out and might not believe future reports by the firm.

As explained before, the key assumption is that the firm and investor care about reporting the true performance. If the firm's payoff function would include $[p + 10]$ instead of $[p + 1]$, the firm would have a much stronger incentive to exaggerate the report and a partial pooling equilibrium would not be possible. If the payoff function would have $[p + .1]$ and thus a smaller incentive to exaggerate. The firms would now choose between more than 2 reporting regimes. The intuition for this is clearest when you think about a situation where the payoff for the firm and the investor are the same. In this case the firm is going to always report truthfully and the investor is going to believe them.

This means that the cheap talk partial pooling equilibrium is going to be a better explanation in an institutional environment where the firm knows that they will be punished for misrepresenting their financial situation and investors know that firms are not trying to outright defraud them. So, you can think of the payoffs as an assumption that investors expect a no fraud environment.

Conclusion

In general, I do not want you to take these models too seriously. Simple model like these will never capture all the complexities of the real world. However, while they might seem ridiculously simplified, they can help you to think through the implications of assumptions that are necessary to predict a certain outcome in the data that you collect. They also tell you what the key assumptions are. In your thesis, you can defend these key assumptions based on prior research or by testing them with your data. The role of your literature review and robustness checks is to convince the reader that the assumptions in your theory are defensible in your setting.

These type of models can help you in general to better identify what you can and cannot test with your data. For instance, both the signalling separation equilibrium and the cheap talk partial pooling equilibrium predict that low performers will behave differently from high performers. That means that you should at least see some difference in the actions in the data. As another example, the cheap talk partial pooling equilibrium predicts that firms will exaggerate their performance when it is not terrible. This is something we observe in the real world.

Finally, these models can provide an alternative explanation to your findings. You might have a complex theory with many variables and moving parts but should you believe your theory if a simple model can explain your findings as well?

Assignment 1: Theory and Regressions

In the first assignment, you are going to simulate data from these two models and run a number of regression models in the setting of CSR. This will give you some practice in running simulations and regression models in `R`.

Set-up

You need to make one change in the `setup` chunk. You have to set your student number equal to the `student_number` object in the code.

The remainder of the code sets up the `R` environment. The first three lines, check whether you have the `cowplot` package installed, and if not it will be installed. Then, I load the `tidyverse` and `cowplot` packages and set the theme for the plots to the default from the `cowplot` package. I do not like the `ggplot` default and prefer the more sparse theme from `cowplot`. I use your student number to set the random seed which makes sure that while all the random simulations are random, they do not change every time you rerun the code. This can make it easier to figure out problems with simulations.

The `eval_assignment1` variable is set to `FALSE` for the moment. If you are happy with all the `R` code for assignment 1, you can set it to `TRUE` to render the whole document. It should not take a lot of time.

```
if (!require(cowplot)){
  install.packages("cowplot", repos = "https://cloud.r-project.org/")
}
```

```
Loading required package: cowplot
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
```

```
-- Conflicts -------------------------------------- tidyverse_conflicts() --
x dplyr::filter()   masks stats::filter()
x dplyr::lag()      masks stats::lag()
x lubridate::stamp() masks cowplot::stamp()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(cowplot)
theme_set(theme_cowplot())
################################################################
###### Change the following line to your student number ######
################################################################
student_number <- 00076637
set.seed(student_number)
eval_assignment1 <- FALSE
```

CSR as signalling.

First, we can treat CSR activities as a signal of the underlying performance of the firm. Given the assumptions of the signalling model, we want a CSR activity that (a) is costly for the firm, (b) less costly for firms that are performing well and (c) investors care about the underlying performance. Donating money to local charities can be seen as an activity that fulfils these assumptions because donating money is more costly for firms that are in a bad financial situation. In the separation equilibrium firms will donate money to local charities when they are high performers but not when they are low performers. As a result, the stock price return of the high performing companies is higher than for the low performing companies.

Questions

1. *Simulation*: Complete the code in the code chunk `crs-signalling` in the answers section. I added indicators, `.x`, `.y`, `.z` or `...` when you need to add changes. When I use `.x` and friends, you only need to fill in one variable or function. When I use `...`, you will need to use more code than just one variable.

   The code works as follows.

   - We are going to create a dataset with 1000 observations.

   - `high_performance` is 1 with probability .5 for each of the observations.

   - `donation` is 1 when the firm is a high performing firm and 0 otherwise. The signalling model represents the donation by a.

   - `return` or b in the signalling model is 4 when the firm donates and 1 otherwise.

- In reality, we will never observe the underlying values perfectly; maybe there are other factors at work, maybe our data collection is imperfect. So, we will add some randomness to the variables. Donations are wrongly observed 10% of the time. We code this as a 90% chance that we have a correct observations: `rbinom(N, 1, .9)`. If we have a correct observation, `observed_donation` is `donation`, otherwise `observed_donation` is the opposite of `donation`.

- `observed_return` is normally distributed with a mean of `return` and a standard deviation of 3.

- We put the observed variables in a data frame as `donation` and `return` that we call `sig` (short for signalling).

- We also use `mutate` to make a new variable `donated` which equals 'yes' when we observe that the firm has donated and 'no' otherwise. This will make it easier to plot the data.

2. *Plotting*: Let's plot the basic effect with `donated` on the x-axis and `observed_return` on the y-axis. We call the figure `sig_plot`.

3. *Regression*: Run a linear regression with `observed_return` as the dependent variable and `donated` as the independent variable. Use `summary` to show the results of the regression.

4. *Interpretation*: explain your answer with 1 or 2 sentences for each question below.

   a. Based on the analysis and the underlying theory, a firm that is not donating money in the sample, can earn a higher stock return by donating money. True or False?

   b. Based on the analysis and the underlying theory, if you know that a firm is high performing, they should donate money if they want to increase their stock return. True or False? Explain your answer.

Answers

1. Data simulation

```
N <- 1000
sig <- tibble(
  high_performance = rbinom(N, .x, .y)) %>%
  mutate(
    donation = ifelse(high_performance == 1, .x, .y),
    return = ifelse(...),
    observed_donation = ifelse(rbinom(N, 1, .9) == 1,
                               donation, 1 - donation),
    observed_return = rnorm(N, return, 3),
```

```
    donated = ifelse(observed_donation == .x, .y, .z)
  )
glimpse(sig)
```

2. Plotting

```
sig_plot <- ggplot(sig, aes(x = .x, y = .y)) +
    geom_jitter(width = .3)
plot(sig_plot)
```

3. Regression

```
sig_reg <- lm(..., data = sig)
summary(sig_reg)
```

4. Interpretation

    a.
    b.

CSR as cheap talk.

We can also think of some CSR reporting as cheap talk. One example would be a CSR report or an integrated report that has not been certified by an external auditor. The main assumptions that we make in the cheap talk model is that both the firm and the investor care about getting investor to believe the true state of the firm but the firm prefers the investors to believe that the firm is doing slightly better than the actual CSR performance.

In the separation equilibrium of our model, the firm is choosing between two options: not issuing a report ($r = 0$) or issuing a glowing CSR report ($r = 10$) depending on their underlying CSR performance. In this world, all firms with performance $10 > p > 3$ are doing some form of greenwashing[1].

---

[1]You could also argue that all the bad performing firms are taking a big bath

Questions

1. *Simulation*

   - We still have `N = 1000` from our previous simulation.

   - `csr_performance` has `N` observations and is randomly uniformly distributed between 0 and 10.

   - `csr_report` is 1 if `csr_performance` is larger than 3 and 0 otherwise.

   - However, we have to hand-collect the data. We have to look up the reports and there is only a 60% chance that we will find it if it exits. Let's create a variable `observed_report` which is 1, .6 of the time. Our data will contain a 1 for `csr_report_observed` if there is a `csr_report` and if the report is observed.

   - Now, we construct a variable `scandals` which indicates how many CSR scandals the firms had after the publication of the report. Scandals or incidents are often used as a post-hoc measure of CSR performance. For no particular reason, we assume that there are 20 periods and in each period a firm has a `scandal_probability` that they will have a scandal in that period. `scandal_probability` is calculated as a function of the CSR performance and we assume that a higher CSR performance is associated with lower `scandal_probability`. I use the `plogis` function which is the R function for the logit transformation. This function takes the value and transforms it to a number between 0 and 1. The higher the value, the closer to 1; the lower the value, the closer to 0. `plogis(csr_performance)` can be interpreted as the probability of no scandal, so `1- plogis(csr_performance)` is `scandal_probability`.

   - The `return` on the stock is the belief that investors have about the stock in the cheap talk model. So, if `csr_report` is equal to 1, the return on the stock of the firm will be 6.5 and 1.5 otherwise. The `observed_return` in our dataset is normally distributed with a mean equal to `return` and a standard deviation of 5.

   - We can put the observed csr report and the observed returns in a dataset, together with the number of scandals. We use `mutate` to create a variable, `reported`, that equals 'yes' if we observe a csr report and 'no' otherwise.

2. *Plotting*: Make a similar plot as for the signalling model and call it `cheap_plot`. Whether a company has an observed csr report should be on the x-axis and the observed returns should be on the y-axis.

3. *Regressions*

   - Create a regression, `cheap_lm1`, with `observed return` as the dependent variable and `csr_report_observed` as the independent variable.

- Create a regression, `cheap_lm2`, which is similar to `cheap_lm1` but adds scandal as a control variable.

- Create a regression, `cheap_lm3`, with `observed return` as the dependent variable and `scandals` as the independent variable.

- Create a new dataset, `only_reports`, with only the firms that have a csr report.

- Create a regresion, `cheap_lm4`, with the same model as `cheap_lm3` but only use the data in `only_reports`.

4. In the `cheap-msummary` chunk I use (and install) the `modelsummary` package to get a nice html output in the rendered document. If you run the code in the `R` console it will give you a preview in your default browser. You need to wrap all the models in a list to get them to use next to each other. I added the traditional p-value starts with `stars = TRUE`. `msummary` includes a lot more *goodness-of-f*it indices, you can omit the least important ones with `gof_omit = "IC|Log|Adj"`. We are typically not too interested in these. If you want to knit the document with the code, you will first have to set `eval_assignment1 <- TRUE` in the setup chunk above. Once you have the results, you can interpret them.

   a. Say that a researcher is interested whether CSR reports are informative for investors, should they report regression (1) or regression (2)? Explain in couple of sentences.

   b. Regression (3) and (4) have a different result for the relation between scandals and return. Explain in a couple of sentences why that is the case.

Answers

1. *Simulation*

```
cheap <- tibble(
  csr_performance = runif(.x, .y, .z)) %>%
  mutate(
    csr_report = ifelse(..., 1, 0),
    observed_report = rbinom(N, 1, 0.6),
    csr_report_observed = csr_report * observed_report,
    scandal_probability = 1 - plogis(csr_performance),
    scandals = rbinom(N, 20, scandal_probability),
    return = ifelse(...),
    observed_return = rnorm(...),
    reported = ifelse(csr_report_observed == 1, "yes", "no"))
```

2. *Plotting*

```
cheap_plot <- ggplot(cheap, aes(..)) +
geom_jitter(width = .3)
plot(.x)
```

3. *Regressions*

```
cheap_lm1 <- lm(..., data = cheap)
cheap_lm2 <- lm(...,
                data = cheap)
cheap_lm3 <- lm(...)
only_reports <- filter(...)
cheap_lm4 <- lm(..., data = only_reports)
```

4. *Interpretations*

```
if (!require(modelsummary)){
  install.packages("modelsummary", repos = "https://cloud.r-project.org/")
}
library(modelsummary)
msummary(list(cheap_lm1, cheap_lm2, cheap_lm3, cheap_lm4),
         gof_omit = "IC|Log|Adj", stars = TRUE)
```

```
a.
b.
```

Appendix

The partial pooling equilibrium in the Cheap Talk game

In the equilibrium, firms only make two choices. They either report that performance is perfect, $r = 10$, or that performance is terrible, $r = 0$. You can also interpret the latter report as not disclosing any performance at all.

Given that is the case, we want to find from which performance, x, the firm will report $r = 10$. We know that investors will belief the expected value of the interval that they think the firm is in. That is if, $r = 0$, investor belief that $b = x/2$ and if $r = 10$, investors belief that $b = (10 + x)/2$.

That means that the firm's payoff when they report $r = 0$, can be written as

$$\pi_{firm,r=0} = c - (x/2 - [p + 1])^2$$

The payoff when the firm reports $r = 10$, is equal to

$$\pi_{firm,r=10} = c - ([10 + x]/2 - [p + 1])^2$$

At the point, where $p = x$, the firm does not care about the difference between $r = 0$ or $r = 10$. This means that we need to set $p = x$ and $\pi_{firm,r=0}$ to $\pi_{firm,r=10}$.

$$\pi_{firm,r=0} = \pi_{firm,r=10}$$
$$([p + 1] - x/2)^2 = ([10 + x]/2 - [p + 1])^2$$
$$(x/2 + 1)^2 = ([10 - x]/2 - 1)^2$$
$$x/2 + 1 = 5 - x/2 - 1$$
$$x = 3$$

The three report version of the partial pooling equilibrium in the Cheap Talk game

Let's assume that the function for the firms is given by

$$\pi_{firm} = c - (b - (p + t))^2$$

where t is a measure of the level of the conflict of interest between the firm and the investors.

Let's assume that there is level of performance x and y with $x > y$ and where for performance $p < x$, the firm does not report performance, for $x < p < y$ the firm reports medium good performance, and for $p > y$ the firm reports excellent performance.

The condition for x is

$$x + t - \frac{x}{2} = \frac{x + y}{2} - x + t$$
$$2(x + t) - x = \frac{y}{2}$$
$$2t + x = \frac{y}{2}$$

The condition for y is

$$(y + t) - \frac{y + x}{2} = \frac{10 + y}{2} - (y + t)$$
$$2(y + t) - \frac{x}{2} - y = 5$$
$$y = 5 - 2t + \frac{x}{2}$$

15

Combining both conditions:

$$4t + 2x = 5 - 2t + \frac{x}{2}$$

$$6t - 5 = -\frac{3}{4}x$$

$$5 - 6t = \frac{3}{4}x$$

$$x = 2(\frac{5}{3} - 2t)$$

$$y = 5 - 2t + 2(\frac{5}{3} - 2t)$$

$$= \frac{25}{3} - 6t$$