

# 1 Introduction

This problem set asks you to work with earning announcements and the reaction to the announcements. The first part of the problem set asks you to go through a series of fundamental steps to analyze the response of stock prices to earning surprises. The second part of the problem set offers a choice between a number of alternative topics.

## 1.1 Earning surprises

The main focus on the literature on earnings announcement has been on the response of investors to new information. Three main measures have been proposed in the literature to quantify the new information. The first two measures compare the earning announcement  $e_{t,k}$  for company  $k$  in quarter  $t$  with the corresponding analyst forecast  $\hat{e}_{t,k}$ . The last measure compares the earning announcement  $e_{t,k}$  with the earning announcement four quarters before,  $e_{t-4,k}$ . The analyst forecasts is defined as the median forecast among all the analysts that make a forecast in the last 45 (trading) days before the earning announcement. If an analyst amde multiple forecasts in this time horizon, we consider the most recent one.

**Measure 1.** Earning surprise 1 is

$$s_{t,k}^1 = \frac{e_{t,k} - \hat{e}_{t,k}}{p_{t,k}}. \quad (1)$$

The difference between the earning announcement and the forecast is divided by the lagged price of a share,  $p_{t,k}$ . The price of a share works as a renormalization factor: the earnings  $e$  are measured as earnings in dollar per share. The division by  $p$  implies that  $s^1$  is the earning surprise as fraction of the value of the company. To see this, multiply numerator and denominator of expression (1) by the number of share  $n_{t,k}$ :

$$s_{t,k}^1 = \frac{e_{t,k}n_{t,k} - \hat{e}_{t,k}n_{t,k}}{p_{t,k}n_{t,k}}.$$

In the numerator,  $e_{t,k}n_{t,k}$  is the total profit for quarter  $t$ , and  $\hat{e}_{t,k}n_{t,k}$  is the total forecasted profit. At the denominator is the market capitalization of a company,  $p_{t,k}n_{t,k}$ . The earning surprise measure, therefore, captures the unexpected profits as a share of total market value of the company. If  $s_{t,k}^1 = .01$ , it means that the company earned unexpected profits equal to 1 percent of the value of the company.

**Measure 2.** Earning surprise 2 is

$$s_{t,k}^2 = \frac{e_{t,k} - \hat{e}_{t,k}}{\hat{d}_{t,k}},$$

where  $\hat{d}_{t,k}$  is defined as the standard deviation between the earning forecasts of the analysts. This measure is therefore missing for companies with only one analyst, and in general for cases in which all the analysts agree in their assessment of the company's profits. This measure

captures the intuition that the surprise is larger for companies in which the analysts agreed in their forecasts.

**Measure 3.** Earning surprise 3 is

$$s_{t,k}^3 = \frac{e_{t,k} - e_{t-4,k} - \bar{e}_{t,t-4,k}}{d_{t,k}}.$$

The numerator is the difference between the earning surprise and the earning surprise 4 quarters before (the argument here is that there are seasonalities) minus an adjustment for average growth in earnings in a year. That is,  $\bar{e}_{t,t-4,k}$  is estimated as the average difference in earnings  $e_{s,k} - e_{s-4,k}$  for  $s = t-1, t-2, \dots, t-16$  (typically people stop at 4 years). The denominator  $d_{t,k}$  is the standard deviation of the earning announcements over the previous 16 quarters. (Note:  $d$  is very different from  $\hat{d}$  above)

## 1.2 Stock returns

Given a measure of earning surprise, the most important question is: how do investors react to the new information contained in the announcement? The specification chosen to look at this is typically a simple, OLS regression:

$$r_{t,k} = \alpha + \phi s_{t,k} + \varepsilon_{t,k}. \quad (2)$$

The idea is to look at how stock prices  $p$  move in response to the new information;  $r$ , the stock return, captures the change in stock prices. A specification allows us to look at stock returns at different horizons. To capture the immediate response, we typically look at  $r^{(0,0)}$ , that is, the stock return the same day as the announcement (measure as price at the close on day  $t$  minus the price at the close on day  $t-1$ ). However, since announcements are sometimes made after close, it is more common to look at  $r^{(0,1)}$ , that is the return for the same day and the next day. If one wants to look at the delayed response to the earning announcement, a typical measure is  $r^{(3,75)}$ , that is, the stock returns for the period  $(t+3, t+75)$ , where days are always meant as trading days. (this is finance!)

As for the measure of returns, three are typically used:

- RAW is just the unadjusted stock return:  $r_{t,k}$
- NET is the stock return minus the market stock return,  $r_{t,k} - r_{t,m}$
- CAR is the abnormal return defined as  $r_{t,k} - \hat{\beta} r_{t,m}$  where  $\beta$  is the correlation between stock  $k$  and the market. This beta are an unexplained passion of all financial economist, and they are unlikely to make a difference for a short-run event study.

## 2 Assignment – part 1

In the dataset `earn219bshort.dta`, which you find zipped on the webpage of the class, I have already merged for you the information from Compustat, CRSP, and IBES. I have also generated the forecast of earnings  $\hat{e}$ . The data set that you see includes therefore information on earnings (MEDACT and GAAP, use MEDEST), earning forecast (MEDEST), stock returns (RAWWIN\*—raw returns, NETWIN\*—returns net of market reutns, CARWIN\*—returns adjusted for correlation with market), volume information (VOLUME\*), aggregate volume information (VOLUME\*). [VOLUME31 is volume same day of earning announcement, VOLUME32 is volume next trading day, etc.] It also contains number of analysts (NEST), standard deviation of earning forecast (STDEST), SIC code of industry (SICCODE), company name (CONAME), price of shares (LAGPRICE), number of shares outstanding (LAGSHR), fraction of share owned by institutional investors (FRINST).

In order to make the data set small enough, it contains a random sample of half of the initial data set. Contact me if you would like to work with the original data set.

1. **Measure of surprise.** Construct measures  $s^1$ ,  $s^2$ , and  $s^3$  in the dataset. What is the average? (use SUM) How high is their correlation? (use PWCORR)
2. **Outliers.** Consider the distribution of these measures. (Use SUM VARNAME,D) Does it seem that the variables have extreme outliers? Construct variables obtained from  $s^1$ ,  $s^2$  and  $s^3$  by trimming (dropping) 2 percent on either tail of the distribution. What is the correlation between the trimmed measures?
3. **Returns I.** Estimate specification (2) using raw returns  $r^{(0,0)}$  as the dependent variable and using the  $s^1$  as a measure of surprise. What is the size (and significance) of the coefficient? How do things change if you use the trimmed surprise variable? What does this suggests about the trimming procedure? From now on, we will always use the trimmed variables.
4. **Clustering.** In running a simple OLS, so far you have made the assumption that all the observations are i.i.d. draws from a Normal distribution. This may be a little problematic, here as in most places. In particular, you may be concerned about the correlation of errors across companies making an announcement on the same day. A way to relax this assumption is to cluster observations by day of announcement  $t$ . In Stata, you add to your regression specification `”, ROBUST CLUSTER(T)”`. How do the point estimates change? How about the standard errors? Argue that the increase in the standard errors due to clustering means that we were neglecting a positive correlation and ‘overcounting’ observations. From now on, maintain the clustering by  $t$  in your specifications. *Important lesson: Think about the correlation structure of your errors, or you may vastly overestimate the precision of your estimates.*
5. **Returns II.** Reestimate specification (2) using the (trimmed) measures 2 and 3 of earning surprises. In which specification is the  $R^2$  higher? Compare the third measure

with the other two. Notice that the third measure does not use at all the forecasts of analysts. Do the analyst forecasts help in increasing the explanatory power? What happens if you run a specification with all three surprise measures in it? Do they all remain significant predictors?

6. **Drift.** Estimate specification (2) using raw (trimmed) returns  $r^{(3,75)}$  as the dependent variable and using the  $s^1$  as a measure of surprise. What does the theory of efficient financial markets predict? What do you find?

### 3 Assignment – part 2

In this second part we use the data set on earning announcements to explore a dozen of different questions. Pick four of the questions below and address them.

1. **Timing of stock response to earning announcement.** Consider specification (2) with (trimmed) surprise measure 1 and net returns  $r_{t,k} - r_{t,m}$  as the dependent variable. Now we focus on when stock prices react to the news contained in the earning announcement. Repeat the regression with returns at (0,0), (1,1), (2,2), (3,75). Is the coefficient  $\phi$  positive for the (2,2) horizon? How about for the (3,75) horizon? How do you interpret the results? Now do the regression with returns at (-1,-1), (-2,-2), and (-30,-3). Do you find any positive coefficients? What does this suggest about the possibility that the part of the information contained in the earning surprise was leaked to the market in the days before the announcement?
2. **Post-earning announcement drift.** We now explore a finding in the previous task, that is, that earning surprises forecast stock returns over the horizon (3,75). This is called the post-earnings announcement drift. We now analyze how much of the drift occurs at the next earning announcement. Consider the specification

$$r_{t,k}^{(0,1)} = \alpha + \phi s_{t-1,k} + \varepsilon_{t,k}, \quad (3)$$

that is, you regress the stock response at time of an announcement on the earning surprise at the *previous* announcement. What is the result for  $\hat{\phi}$ ? Why is it surprising that  $\hat{\phi}$  is positive? Argue that in efficient financial market  $\phi$  should be zero. Now regress  $s_{t,k}$  on  $s_{t-1,k}$ . Argue that this result suggests that analysts do not adjust in a timely fashion their earnings forecasts. Give two possible reasons for this. Can this analyst bias help explain the result in specification (3)? Replicate regression (3) using the earning surprise 2 announcements ago, 3 announcements ago, and 4 announcements ago. How are the patterns?

3. **Time-varying effects and measurement error.** We now explore a different aspect of the findings in point 1. Break down the sample in three time periods, 1984-1989, 1990-94, and 1995-2002 and re-run the regression in point 1. Notice that the coefficient

$\phi$  of returns (0,0) on earning surprises is quite a bit higher in the later than in the earlier period. How about returns at (-1,-1)? How would you explain this? Part of the explanation is measurement error in the date of announcement. A team of Berkeley undergrads used newswires to locate the exact time of the announcement for about 1,500 announcements. This information is recorded by the variable  $tn$ . Compare the variable  $tn$  to the (reported) date of announcement in IBES, as recorded by the variable  $t$ . How close are the two dates for the pre-1990 and the post-1990 period? Argue that measurement error in the date can explain part of the differences in the results of the return regressions in the three different periods. *Very important lesson: Do not trust the quality of the data. Go out of your way to check it. You should know your data like your pockets. [Caroline Hoxby]*

4. **Volume data I.** In the data set, I have provided you with data on a trading volume measure, that is, the value of the shares exchanged in a day. I would like you to examine what happens to volume of trading in response to earning surprises. What do you expect to find? Denote by  $v_{t,k}^{(s,s)}$  the value of the shares of company  $k$  traded  $s$  days after the day of announcement in quarter  $t$ . You will run a specification like:

$$\log(v_{t,k}^{(s,s)}) - \log(v_{t,k}^{(-10,-5)}) = \alpha + \varepsilon_{t,k}$$

Notice that the dependent variable is the difference between log volume around the announcement date and log volume the week before the announcement. Why is it important to control for baseline volume? Run the regression for  $s = 0$ . How do you interpret the estimated  $\hat{\alpha}$ ? Now run the same regression for  $s = -2, -1, 1, 2, 3, 4, 5$ . How are trading patterns around announcement date? What does this suggest about the diffusion of information after the announcement? Why is this pattern different from the pattern for returns?

5. **Volume data II.** We now look at the increase in abnormal volume as a function of the earning surprise. To start with, you may try to run a specification like:

$$\log(v_{t,k}^{(s,s)}) - \log(v_{t,k}^{(-10,-5)}) = \alpha + \phi s_{t,k} + \varepsilon_{t,k} \quad (4)$$

What are the results for  $s = 0$  (same day increase in volume)? Why may there be problems with this specification? Would you expect the increase in volume to be a linear function of the earning surprise? Divide the earning surprises in quintiles. Run a specification with four dummies for the quintiles (omit the lowest). What do the result suggest? What are the interpretations of this result in terms of attention and information content? Is the linearity assumption in (4) a good assumption?

6. **Non-linearities in stock response.** So far we have assumed that the relationship between stock returns and earning surprises is linear, as in (2). We now allow for a piece-wise linear specification. Define  $D_{t,k}$  a dummy equal to one if the earning surprise  $s_{t,k}$  is positive. Run the following specification

$$r_{t,k}^{(0,1)} = \alpha + \phi_0 s_{t,k}^1 + \phi_1 D_{t,k} + \phi_2 s_{t,k}^1 * D_{t,k} + \varepsilon_{t,k}. \quad (5)$$

How do you interpret the coefficients  $\hat{\phi}_0$ ,  $\hat{\phi}_1$ ,  $\hat{\phi}_2$ ? Draw the (approximate) plot of stock returns as a function of earning surprise implied by the specification (5). Can you reject the linearity assumption implicit in (2)? How much higher is the  $R^2$  in this more general specification? Can you give a behavioral interpretation to the coefficient  $\hat{\phi}_1$ ? How about the coefficient  $\hat{\phi}_2$ ? (you may want to relate this to the Degeorges, Patel, Zeckhauser paper on the reading list under Firm Response)

7. **Non-linearities in stock response II.** Experience with other non-linear specifications. In particular, do a kernel regression of returns  $r_{t,k}^{(0,1)}$  on the surprise  $s_{t,k}^1$ . (to do this, you will need to download the package kernreg. Type "Search kernel" in Stata for explanations). How does this fully non-parametric specification match up with specification (5)? Was this latter specification a good approximation?
8. **Size of firms.** Run the regressions in point 1 separately for firms with few analysts and firms with many analysts? Do you see any differences? How about the regressions on drift in point 2? Try also to split the data by fraction of institutional investors (this measure is highly positively correlated with firm size). Are the results different in the two samples?
9. **Clustering II.** Above I have suggested that you allow for correlation across announcements in one day by clustering by time  $t$ . You may also be concerned about the correlation of errors over time for the same company. You can check this by running specification (2) with ", ROBUST CLUSTER(PERMNO)", that is, you cluster by company identifier. (you cannot cluster on both contemporaneously) What happens to standard errors? What does this suggest about the clustering that one should adopt in order to be conservative?
10. **Open-ended.** Have you noticed any other interesting phenomenon in the data? Write about it. Is this related to a feature of the trading environment, to an informational story, to a behavioral story? Any general lessons?