Problem Set on Earnings Announcements (219B, Spring 2005)

Stefano DellaVigna

April 13, 2004

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 Earnings 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. Earnings surprise 1 is

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

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 $s^1_{t,k} = .01$, it means that the company earned unexpected profits equal to 1 percent of the value of the company.

Measure 2. Earnings surprise 2 is

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

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. Earnings surprise 3 is

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

The numerator is the difference between the earning announcement and the earning announcement 4 quarters before (the argument here is that there are seasonalities). The denominator $d_{t,k}$ is the standard deviation of the numerator 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? There are two broad methodologies to do this.

Methodology 1 – Quantiles. First, sort the earnings surprises into quantiles using one of the earnings surprise variables s^1, s^2, or s^3. Then compute the average stock return for each of the group of announcements. Figures 1a-1d in DellaVigna and Pollet (2004) do this (they also split announcements into Friday and non-Friday, but this is beyond the point here). This methodology allows to get a non-parametric plot of the relationship between returns and earnings surprises. On the other hand, it does not allow for control variables.

Methodology 2 – Regression framework. For simplicity, we consider an OLS regression:

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

The coefficient ϕ captures the response of stock returns to the surprise, assuming that the response is linear. Given that the response is not linear, this specfication works best if restricted in the interval $-.02 \leq s_{t,k} \leq .02$, or some comparable interval.

In either case, we can consider the response of stock returns at different horizons. To capture the immediate response, one could 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, one should 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 β 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 (VOLU*), aggregate volume information (VOLUA*). [VOLU31 is volume same day of earning announcement, VOLU32 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 ownsed 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 for each measure? (use SUM) How high is their correlation? (use PWCORR). Now 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?

2. **Short-run response, Methodology 1.** Use Methodology 1 and plot raw returns $r^{(0,1)}$ as a function of s^1. Sort the raw returns into 11 quantiles as follows. Define quantile 6 as the group of announcements with no surprise ($e_{t,k} = \hat{e}_{t,k}$). Divide the announcements

with negative surprises ($s^1 < 0$) in 5 equal-sized groups, with group 1 being the one with the most negative announcements and group 5 the one with least negative. Similarly, divide announcements with positive surprises ($s^1 > 0$) in 5 equal sized groups (groups 7 through 11). Group 11 will be the one with the most positive surprises. Plot raw returns $r^{(0,1)}$ for each of these quantiles.

3. **Short-run response, Methodology 1, Part 2.** Interpret the economic magnitudes in the plot in point 3. Does this plot in point 3 imply a linear relationship between s^1 and $r^{(0,1)}$? Provide at least one interpretation about the observed non-linearity. Can you think of ways to test your hypotheses on the non-linearity?

4. **Short-run response, Methodology 2.** Estimate specification (2) using raw returns $r^{(0,1)}$ 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 restrict the sample to $-.02 \leq s_{t,k} \leq .02$? What does this suggests about the underlying linearity assumption in (2)? (Refer to point 4) From now on, we will always use the sample restriction $-.02 \leq s_{t,k} \leq .02$ when running the OLS specification.

5. **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 is problematic, here as in most papers. 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. *Lesson: Think about the correlation structure of your errors, or you may vastly overestimate the precision of your estimates.*

6. **Post-earning announcement drift.** Use Methology 1 to plot raw returns $r^{(3,75)}$ as a function of the 11 quantiles in the earnings surprise variable s^1. [It is easiest to do the plot in Excel once you have generated the mean return for each quantile] What does the theory of efficient financial markets predict? What do you find? Measure the drift as the difference between the return for the highest quantile minus the return for the lowest quantile. Compute a standard error for this difference.

7. **Manipulation of earnings.** (DeGeorge, Patel, and Zeckhauser, 1999). Companies have some discretion in the accounting procedure, so they can manipulate the earnings release at the margin. Consider the numerator of the earnings surprise, $e_{t,k} - \hat{e}_{t,k}$. This is the earnings surprise per share. Plot the distribution of this variable for $-\$.1 \leq e_{t,k} - \hat{e}_{t,k} \leq \$.1$. (Excel histogram cent-by-cent would work, for example) Comment on whether the distribution has a discontinuous drop at $e_{t,k} - \hat{e}_{t,k} = 0$, and interpret it relating it to manipulation of earnings.

3 Assignment – part 2

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

1. **Drift II.** We now explore further the finding 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 earnings announcement. Consider the specification
$$r_{t,k}^{(0,1)} = \alpha + \phi s_{t-1,k} + \varepsilon_{t,k}, \tag{3}$$
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 ϕ should be zero. Now regress $s_{t,k}$ on $s_{t-1,k}$. What does this suggest about analysts? 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?

2. **Drift III.** Again on drift. We now go back and plot again raw returns $r^{(3,75)}$ as a function of quantiles in the earnings surprise variable. However, instead of sorting into quantile based on the earnings surprise $s_{t,k}^{1}$, we sort into ten deciles based on the raw return at announcement, $r_{t,k}^{(0,1)}$. The immediate stock response is of course an alternative measure of good/bad news at announcement. Comment on the difference between this graph and the graph in point 7 above. Which specification gives the largest earnings drift, as measured as in point 7?

3. **Manipulation of earnings II.** (DeGeorge, Patel, and Zeckhauser, 1999). The earning surprise per share that we consider at the previous point is not the only obvious target of attention for investors. Two other obvious variables are the earning per share itself, $e_{t,k}$, and the difference from the previous year, same quarter, $e_{t,k} - e_{t-4,k}$. Again, do a plot for each of these two variables. Is there a discontinuity at zero? Where does the discontinuity appear to be larger? What does this suggest about what investors pay most attention to?

4. **Trading volume I.** I have provided you with data on a trading volume measure, that is, the value of the shares exchanged in a day for a company. 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:
$$\log\left(v_{t,k}^{(s,s)}\right) - \log\left(v_{t,k}^{(-10,-5)}\right) = \alpha + \varepsilon_{t,k}$$
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. **Trading volume II.** We now look at the increase in abnormal volume as a function of the earning surprise. Run a specification controlling for the 11 quantiles of the earnings surprise (omit quantile 6, it's easiest to interpret the coefficients):

$$\log\left(v_{t,k}^{(s,s)}\right) - \log\left(v_{t,k}^{(-10,-5)}\right) = \alpha + \sum_{d=1}^{11} \phi_d s_{t,k}^d + \varepsilon_{t,k}$$

What are the results for $s = 0$ (same day increase in volume)? How does the volume response vary depending on the earnings surprise, that is, what are you finding on the ϕ_ds? What are the interpretations of this result in terms of attention and information content? Do the results on the ϕ_ds vary for $s = 2$ or $s = 5$ (two or five days later)?

6. **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 standards errors? What does this suggest about the clustering that one should adopt in order to be conservative?

7. **Response over time to earnings announcement.** Consider specification (2) with the usual sample restriction and 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 ϕ significantly 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?

8. **Non-linearities I.** Consider again specification (2) with the usual sample restriction and surprise measure 1 and net returns $r_{t,k} - r_{t,m}$ as the dependent variable. So far we have assumed that the relationship between stock returns and earning surprises is linear. We now allow for non-linear specifications. Consider 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

$$r_{t,k}^{(0,1)} = \alpha + \phi_0 s_{t,k}^1 + \phi_1 D_{t,k} + \phi_2 s_{t,k}^1 * D_{t,k} + \varepsilon_{t,k}. \tag{4}$$

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 (4). 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 $\hat{\phi}_2$?

9. **Non-linearities II.** Following up on the previous point, consider other non-linear specifications. Propose one that specification that makes sense given the results with the quantile method. Run the regression and comment on increases in R^2. Alternatively, do a kernel regression or 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 (4)? Was this latter specification a good approximation?

10. **Different surprise measures II.** Reestimate specification (2) using the measures 2 and 3 of earning surprises (usual sample restrictions). 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?

11. **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 specification (2). Notice that the coefficient ϕ 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. *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.*

12. **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?

4 Names of Variables

Brief explanation of variables. In square parentheses are the ones that you will not need for the problem set

T - Date of earning announcement

[TC and TI - Date of earning announcement according to Compustat and IBES respectively]

NEST-number of analysts following stock

STDEST-standard deviation of analist forecasts about earning announcement

MEDEST-Median earning forecast (IBES)

MEDACT-Earning announcement (IBES)

CONAME-Company name

[GAAP-Earning announcement (Compustat)]

SICCODE-SIC code of company making announcement

PERMNO-Identifier number of company making announcement (CRSP)

RAWWIN*-Raw return of stock k on Window * around earning announcement

NETWIN*-Return of stock k on Window * around earning announcement minus aggregate stock

CARWIN*-Return of stock k on Window * around earning announcement minus β * aggregate stock

Window Explanation: Type SUM CARWIN*,D. (0,1) for example means return between the announcement day and the next day.

LAGPRICE-Price of a share of company k right before announcement

LAGSHR-Number of shares outstanding of company k right before announcement

VOLU*-Volume of shares of company k traded (in $) on Window * around announcement day. VOLU31 is volume traded on announcement day, VOLU32 is volume traded on the trading day follwoing the announcement day, etc.

VOLUA*-Total volume of shares traded (in $) on Window * around announcement day

Time indicators

FRINST - Fraction of market value of company k held by institutional investor