

We would like you to use the following datasets to produce your best answer to the following question: are analysts' annual forecast updates a linear function of their quarterly forecast errors?

You have two files, *fundq* and *det_epsus*, provided in both csv and xlsx formats. The file *fundq* (based on Compustat) contains data on S&P 500 firms' earnings announcements for the fiscal quarter ended on Mar 31, 2020. All included firms have calendar fiscal years (i.e., fiscal year ended on Dec 31, 2020). The file contains the following variables:

Variable	Definition
<i>gvkey</i>	Compustat company identifier
<i>datadate</i>	Quarter-end date (Mar 31, 2020 for all observations)
<i>conm</i>	Company name
<i>rdq</i>	Quarterly earnings announcement date
<i>cik</i>	U.S. SEC company identifier
<i>ate</i>	Total assets (as of Mar 31, 2020)
<i>cshoq</i>	Common shares outstanding (as of Mar 31, 2020)
<i>ticker</i>	I/B/E/S company identifier

The file *det_epsus* (based on I/B/E/S) contains data on analyst EPS forecasts announced from Jan 1, 2020 to Jun 30, 2020. The file only contains forecasts for the immediate quarter or year (i.e., *fpi* is 1 or 6).

The file contains the following variables:

Variable	Definition
<i>ticker</i>	I/B/E/S company identifier (of the firm being forecast)
<i>analys</i>	Analyst identifier
<i>fpi</i>	Forecast period indicator (1 for year, 6 for quarter)
<i>value</i>	Forecast EPS value
<i>fpedats</i>	Forecast period end date
<i>anndats</i>	Forecast announcement date
<i>actual</i>	Actual EPS value for the fiscal period per I/B/E/S

The data could be messy, so use your judgment to transform variables or to restructure the data (e.g., winsorization or truncation for treatment of outliers) by going further than the minimum requirement of the following instructions.

Task 1

Merge the file *det_epsus* into the file *fundq* by creating at least the following variables:

- Analysts who provided forecasts for the firm.
- Actual EPS value for the fiscal quarter ended Mar 31, 2020 per I/B/E/S.
- Last quarterly EPS forecast for the fiscal quarter ended Mar 31, 2020 each analyst announced for the firm between Jan 1, 2020 and Mar 31, 2020.
- Last annual EPS forecast for the fiscal year ended Dec 31, 2020 each analyst announced for the firm between Jan 1, 2020 and Mar 31, 2020.
- First annual EPS forecast for the fiscal year ended Dec 31, 2020 each analyst announced for the firm between the earnings announcement date (*rdq*) and Jun 30, 2020.

Task 2

Using the data created in Task 1, define the two main variables: (1) quarterly EPS forecast error (forecasted EPS minus actual EPS) and (2) annual EPS forecast revision around the Q1 earnings announcement (i.e., new forecast minus old forecast).

- Both variables should be scaled by total assets per share as of Mar 31, 2020 ($atq \div cshoq$).
- You are encouraged to define more variables for control purposes (e.g., number of analysts following the firm, number of firms the analyst is following, analyst forecast dispersion).

Task 3

Provide descriptive statistics, both in tables and figures (e.g., scatter plots, histograms, correlograms), you deem useful in answering the question: is analysts' annual forecast updating a linear function of their quarterly forecast errors? However, at a minimum, you should provide:

- Table with basic descriptive statistics for the two main variables.
- Scatter plot of the two main variables.

Task 4

Design regression tests you deem useful in answering the question: is analysts' annual forecast updating a linear function of their quarterly forecast errors? You are free in designing your regression tests, but you

should demonstrate and explain the purpose of the inclusion of multiple independent variables, clustering of standard errors, and inclusion of fixed effects in some of the specifications.