Group Assignment #3

Team Name

Team Member 1,
Team Member 2,
Team Member 3,
Team Member 4,
Team Member 5

Section I. Description of model and variables

Our prediction model is as follows:

$$Earn_{t+1} = \alpha_0 + \alpha_1 EARN_t + \alpha_2 Research + \alpha_3 ATO + \alpha_4 SpecialItems \quad (1)$$

$$AFE_{t+1} = \beta_0 + \beta_1 AFE_t + \beta_2 Research + \beta_3 ATO + \beta_4 SpecialItems \quad (2)$$

where *Earn* is earnings measured as a fraction of assets. *AFE* is analyst forecast error as a fraction of assets. *Research* is research and development expense as a fraction of assets. *ATO* is the measure of asset turnover from Soliman (2016). *SpecialItems* is the amount of special items as a fraction of assets.

Formal definitions of the variables are as follows:

Earn	Compustat IB / Compustat AT
AFE	((Ibes ACTUAL – Ibes CONSENSUS) * Ibes IBESSHROUT) / Compustat AT.
Research	Compustat XRD / Compustat AT. Missing values of XRD are set to zero.
ATO	Compustat SALE / NOA. Definition is from Soliman (2016). Following Soliman (2016), observations with values below (above) the 1st (99th) percentile are set equal to the percentile value.
NOA	(Compustat AT – Compustat CHE – Compustat IVAO) – (Compustat AT – (Compustat DLTT+ Compustat DLC) – (Compustat CEQ+ Compustat PSTK) – Compustat MIB). Definition is from Soliman (2016). Missing values are replaced with zeros.
SpecialItems	Compustat SPI / Compustat AT. Missing values of SPI are set to zero.

All regressions are estimated on 83,382 observations with non-missing data from 1993 to 2016.

Section II. Predictions and Intuition

We include current period versions of earnings ($Earn_t$) and analyst forecast error (AFE_{t+1}) because… If earnings are persistent, and analysts do not perfectly learn from their prior mistakes, we predict both variables will have positive coefficients, α_1 and $\beta_1 > 0$.

We include research and development expect because… If research and develop is associated with higher future earnings we expect $\alpha_2 > 0$; and analysts do not fully incorporate this into their forecasts, we expect $\beta_2 > 0$.

Following Soliman (2016) we include asset turnover because…. If…. we expect…

We include special items because… If special items represent one time revenues or expenses that are not recurring we expect $\alpha_4 = 0$. If analysts do not understand that special items are one-time items that do not persist, then their forecasts for firms with large special items will be too optimistic and these firms will subsequently miss such forecasts. Consequently, we expect $\beta_4 < 0$.

Section III. Discussion of Results

Table 1 reports the number of observations with non-missing values of all variables from 1993 to 2016, and Table 2 shows mean, standard deviation, and 25th, 50th, and 75th percentiles of the variables used in our analysis. Tables 3 show results from estimating equations (1) and (2).

…DISCUSS...

Section IV. Out-of-sample testing

We begin out-of-sample testing by calculating the absolute analyst forecast error, $|AFE_{2017}|$, and averaging across our sample of 2,214 firms in 2017 to arrive at the average absolute analyst forecast error, *Analyst* $ABFE_{2017}$. Similarly, we calculate the squared analyst forecast error, $(AFE_{2017})^2$, and average across our sample of 2,214 firms in 2017 to arrive at the mean-squared analysts forecast error, *Analyst* $MSFE_{2017}$.

Analyst $ABFE_{2017}$ = 0.018
Analyst $MSFE_{2017}$ = 0.029

We then repeat this procedure using the model described in Table 3 to forecast earnings in 2017 using 2016 data. *Model* $ABFE_{2017}$ is the average of $|Earn_{2017} - Predicted\_Earn_{2017}|$ across our sample of 2,214 firms in 2017. *Model* $MSFE_{2017}$ is the average of $(Earn_{2017} - Predicted\_Earn_{2017})^2$ across our sample of 2,214 firms in 2017.

Model $ABFE_{2017}$ = 0.072
Model $MSFE_{2017}$ = 0.035

…DISCUSS...

Table 1. Number of Observations

This table presents number of firms, by year, with non-missing values of all variables used to estimate equations 1 and 2.

Year	*Number of Firms*
1991	2,214
1992	3,196
1993	3,461
1994	3,674
1995	4,116
1996	4,085
1997	3,823
1998	3,696
1999	3,321
2000	3,063
2001	3,083
2002	3,138
2003	3,257
2004	3,370
2005	3,401
2006	3,298
2007	3,208
2008	3,147
2009	3,134
2010	3,191
2011	3,195
2012	3,295
2013	3,396
2014	3,357
2015	3,263
2016	2,214
TOTAL	83,382

Table 2. Descriptive Statistics

This table presents mean, standard deviation, and 25th, 50th, and 75th percentiles of variables used in our analysis.

Variable	*Mean*	*Std*	*25th*	*Median*	*75th*
EARN	0.000	0.254	0.003	0.031	0.072
AFE	-0.001	0.195	-0.001	0.000	0.003
RESEARCH	0.047	0.650	0.000	0.000	0.041
ATO	1.760	4.605	0.526	1.422	2.639

SPECIALITEMS	-0.016	0.309	-0.007	0.000	0.000

Table 3. Regression Results.

This table presents results from estimating equations (1) and (2). All variables are as defined in Appendix A. *t*-statistics appear in parentheses. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

Variable	Base Model	Equation (1)
EARN	0.89***	0.95***
	(254.33)	(272.53)
Research	.	–0.28***
	.	(–64.85)
ATO	.	0.0002
	.	(1.15)
SpecialItems	.	–0.70***
	.	(–77.15)
N	83,382	83,382
Adj R^2 (%)	43.69	47.86

Variable	Base Model	Equation (2)
AFE	0.41	0.28***
	(125.58)	(125.65)
Research	.	–0.003
	.	(–0.80)
ATO	.	–0.0002
	.	(–1.12)
SpecialItems	.	0.003
	.	(0.49)
N	83,382	83,382
Adj R^2 (%)	15.90	15.92

Appendix A

```
*These tell SAS the location of the data;

libname sav "~/fin3380/";


**********************;
* ADD FUTURE EARNINGS *;
**********************;

proc sql;
create table addfutureyear as
select a.*, b.ib as ib_p1, b.at as at_p1
from sav.Compustat1980to2021 a
inner join sav.Compustat1980to2021 b
on a.gvkey=b.gvkey and b.fyear=a.fyear+1;
quit;

************;
* ADD IBES *;
************;

proc sql;
create table addibes as
select a.*, b.actual, b.consensus, b.ibesshrout
from addfutureyear a
inner join sav.ibes1993to2021 b
on a.permno=b.permno
and a.datadate=b.datadate;
quit;

proc sql;
create table addibesfut as
select a.*, b.actual as actual_p1, b.consensus as consensus_p1, b.ibesshrout
as ibesshrout_p1
from addibes a
inner join sav.ibes1993to2021 b
on a.permno=b.permno
and year(a.datadate)+1=year(b.datadate);
quit;

/*You can use SQL and its coalesce function for the same purpose*/
data varlist; set addibesfut; where at>0;

earn_p1=ib_p1/at_p1;
earn=ib/at;
AFE_p1=((actual_p1-consensus_p1)*ibesshrout_p1)/at_p1;
AFE=((actual-consensus)*ibesshrout)/at;

if XRD=. then XRD=0;
if CHE=. then CHE=0;
if IVAO=. then IVAO=0;
if DLTT=. then DLTT=0;
if DLC=. then DLC=0;
if CEQ=. then CEQ=0;
```

```
if PSTK=. then PSTK=0;
if MIB=. then MIB=0;
if SPI=. then SPI=0;

Research=XRD/AT;
SpecialItems=SPI/AT;
NOA = (AT - CHE - IVAO) - (AT - (DLTT+DLC) - (CEQ+PSTK) - MIB);
ATO = SALE / NOA;
run;

proc means data=varlist mean std p1 p25 p50 p75 p99;
var earn afe research ato specialitems;
quit;

*****************;
* Winsorization *;
*****************;

data varlist; set varlist;
if ATO>21.3838515 then ATO=21.3838515; if .<ATO<-17.9972798 then ATO=-
17.9972798;
run;

data sav.groupassign3;
set varlist;
if nmiss(earn_p1, earn, afe_p1, afe, research, specialitems, ato)=0;
if fyear>1991; *only 4 firms made sample in 1991;
run;

***********;
* TABLE 1 *;
***********;

proc freq data=sav.groupassign3;
tables fyear;
quit;

***********;
* TABLE 2 *;
***********;

proc means data=sav.groupassign3 mean std p1 p25 p50 p75 p99;
var earn afe research ato specialitems;
quit;

***********;
* TABLE 3 *;
***********;

proc reg data=sav.groupassign3;
model earn_p1 = earn ;
quit;

proc reg data=sav.groupassign3;
model earn_p1 = earn research ato specialitems;
quit;
```

```
proc reg data=sav.groupassign3;
model afe_p1 = afe;
quit;

proc reg data=sav.groupassign3;
model afe_p1 = afe research ato specialitems;
quit;

***************************;
* OUT OF SAMPLE PREDICTION *;
***************************;
data groupassign3; set sav.groupassign3;
keep earn_p1 earn AFE_p1 AFE research ato specialitems noa fyear permno gvkey
actual consensus
eps:;
run;

proc corr data=groupassign3 spearman;
var actual consensus;
with eps:;
run;

proc reg data=groupassign3 outest=params noprint;
model earn_p1 = earn research ato specialitems;
where fyear<=2017;
quit;

proc sql;
create table outofsample as
select distinct a.permno, a.gvkey, a.fyear, a.earn_p1, afe_p1,
a.earn*b.earn+a.research*b.research+a.ato*b.ato+
a.specialitems*b.specialitems+Intercept as predicted_earn
from groupassign3 as a inner join params as b
on a.fyear=2018;
quit;

data outofsample; set outofsample;
predicted_earn2019=predicted_earn;
modelABFE=abs(earn_p1-predicted_earn2019);
modelMSFE=(earn_p1-predicted_earn2019)**2;
analystABFE=abs(afe_p1);
analystMSFE=afe_p1**2;
run;

proc means data=outofsample mean;
var analystABFE analystMSFE modelABFE modelMSFE;
quit;
```