

Initial results for correlation
between future tense and
earnings management

Aims

Assess the strength of pcftr on AAM, controlling for linguistic history.

Country-level AAM data, but language-level linguistic history data.

Solution: Multiple membership model.

Each country 'belongs' to multiple language family histories.

R package 'brms' implements a multiple membership model through stan.

Computation is expensive: Estimation is done through sampling an MCMC process along multiple chains. The chains eventually reach convergence.

First analysis

To get a first look at the data I analysed a random subset:

- 10,000 randomly selected datapoints (about 10% of data)

And used only short chains:

- 1000 runs of 'burnin' and 10,000 sample runs

Compare model with random intercepts for pcftr to model with random slopes for pcftr.

Model 1: random intercepts for each language family

AAM ~ 1 + pcftr +

invpro+pd+indiv+mas+ua+lto+

indul+ggr+SIZE+BTM+LEV+ROA+

ISSUE+MEET+LOSS+

(1 | gvkey) +

Random effect for company?

(1 | fyear) +

Random effect for year

(1 | indus) +

Random effect for industry

(1 | mm(G1, G2, G3,

Random effect for language family

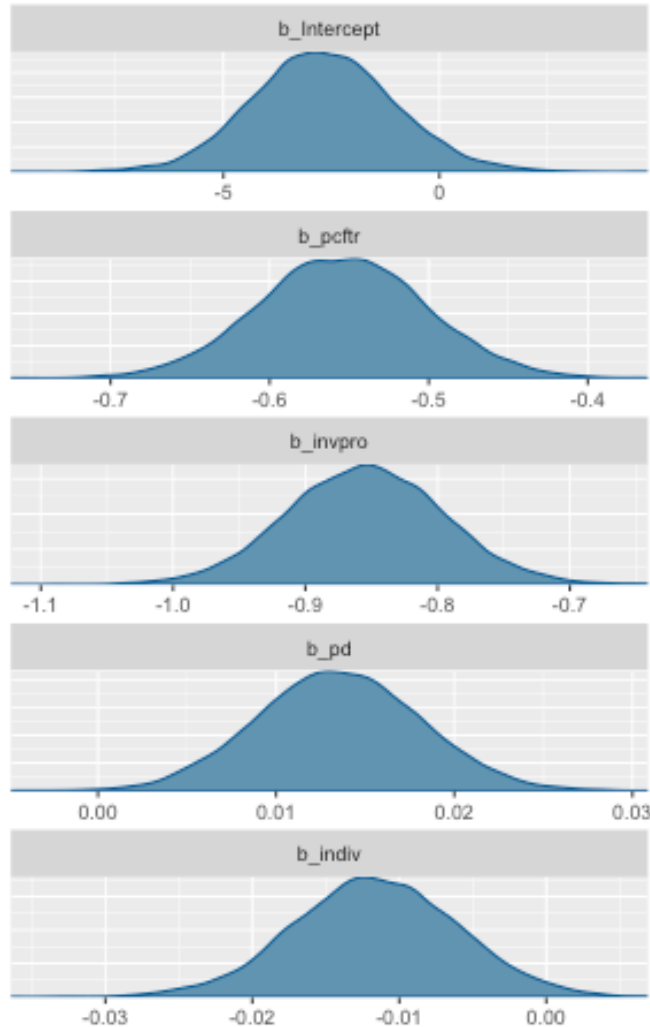
weights = cbind(G1.p, G2.p, G3.p))

Results

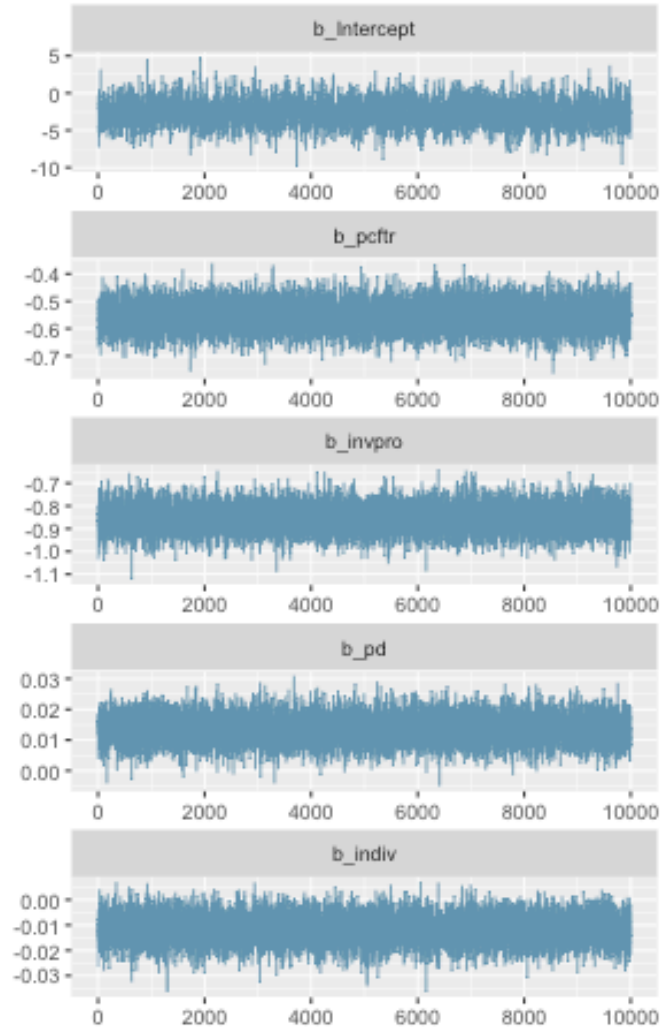
Distributions are close to normal
= good convergence

Coefficient for pcftr is robustly below 0

Distribution of coefficient estimates

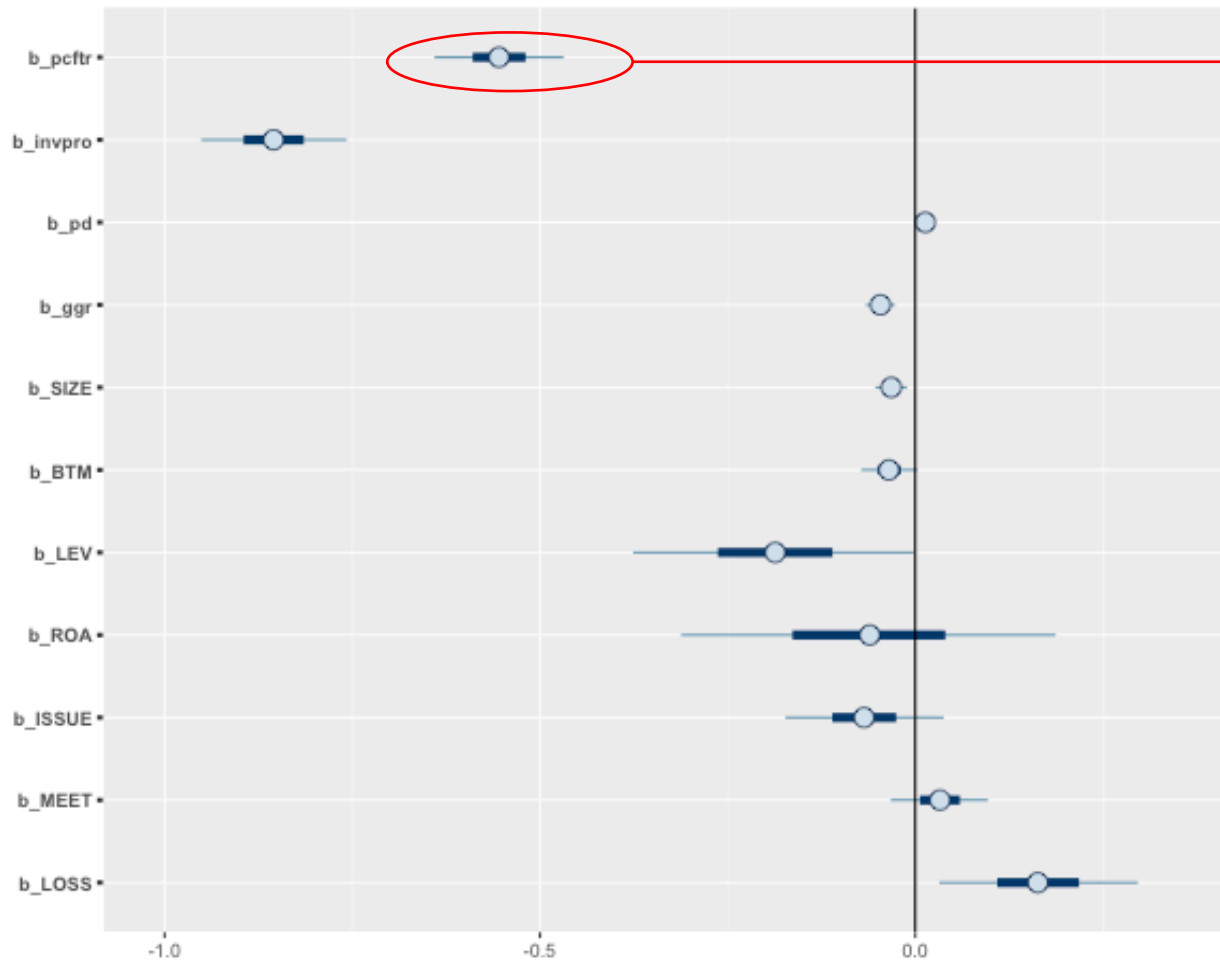


Trace of estimates in chain



Parameter estimates don't show big "jumps"
= good convergence

Results



Distribution of coefficient estimates for pcftr:
Confidence interval for main effect of pcftr is above zero

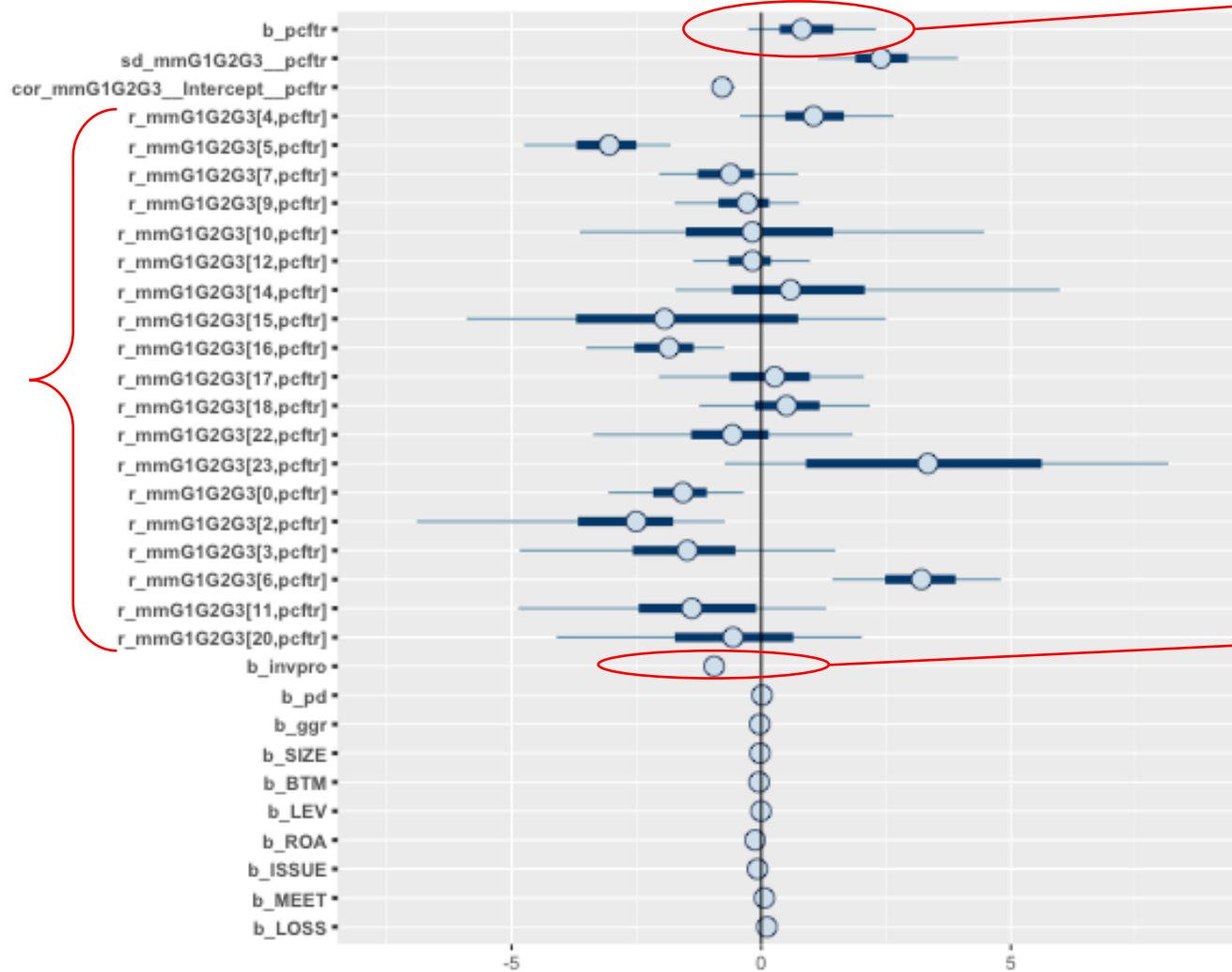
Coefficient estimates

Model 2: allow effect of pcftr to vary by language family (random slopes)

AAM ~ 1 + pcftr +
 invpro+pd+indiv+mas+ua+lto+
 indul+ggr+SIZE+BTM+LEV+ROA+
 ISSUE+MEET+LOSS+
 (1 | gvkey) +
 (1 | fyear) +
 (1 | indus) +
 (1 + **pcftr** | mm(G1, G2, G3, # genres
 weights = cbind(G1.p, G2.p, G3.p))

Model 2: allow effect of pcftr to vary by language family (random slopes)

Effect of pcftr
varies widely
between
language
families

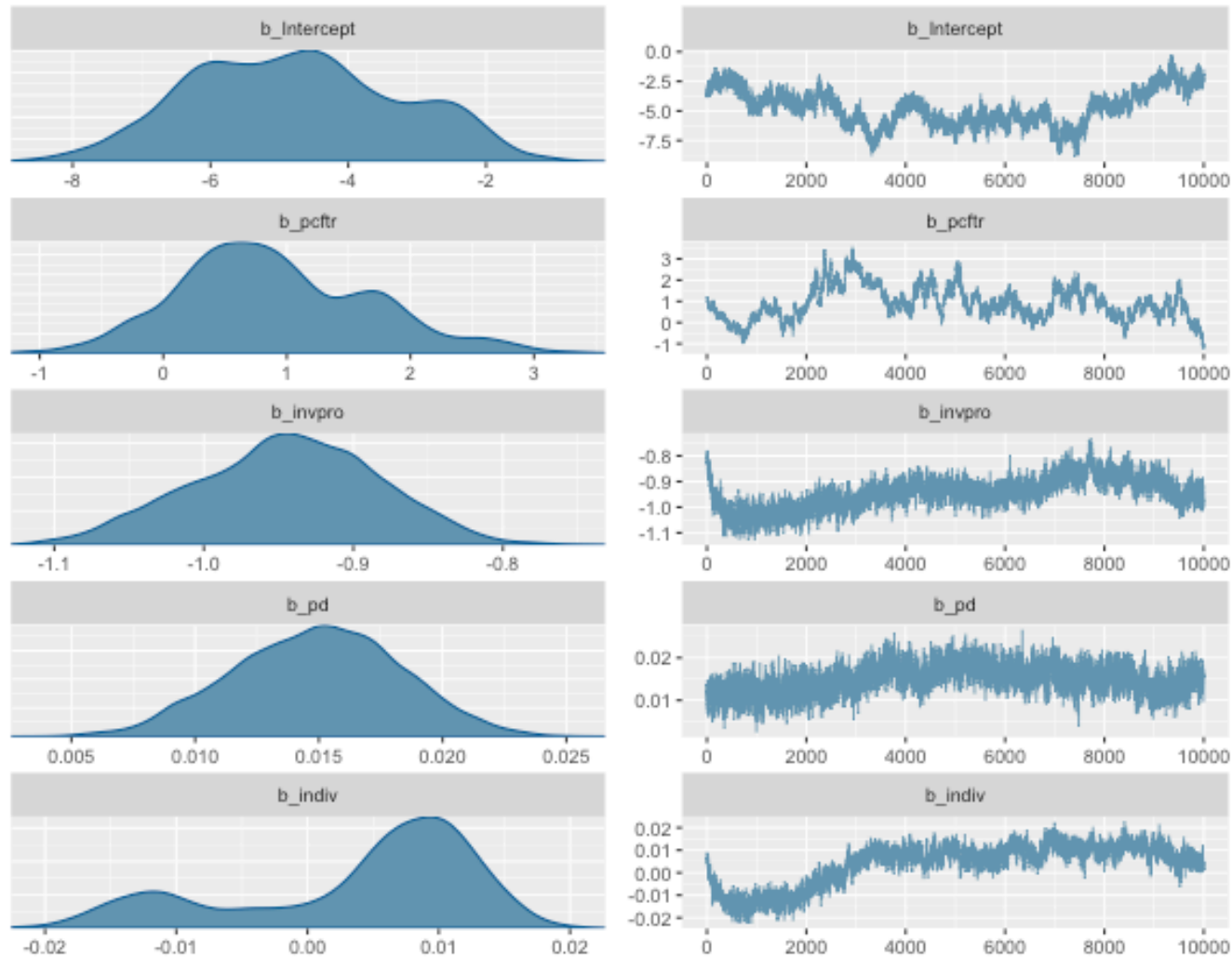


Confidence interval for main
effect of pcftr now overlaps
zero
= the effect of pcftr is not
robust to controls for
language family

Effect of invpro robustly
different from zero

However ... Model 2 does not converge well yet

Distributions have multiple solutions



Estimates are still converging

Solution:

- Run chain for longer
- Run more chains

This needs to be done on our cluster computer

Some technical problems at the moment ...

Issues:

- What other variables to include in the model?
- What should the source of the language family data be?
- Do random slopes make sense?