

Pillar 1 - Conformal Prediction

DARS

2019-05-09

1 Set up

2 Course selection

I select courses with large and small n, and large and small CV error. I do not select courses whose lasso model contains less than 3 non-zero coefficients (df).

- COR1002: large n
- COR1004: large n and smaller CV error (mean absolute error) than COR1002
- SCI3003: small n and large CV error
- SCI2040: small n and small CV error
- SSC3044: small CV error
- SSC3038: small CV error and n twice as large as SSC3044
- SCI2018: large CV error
- SCI2010: large CV error and n twice as large as SCI2018

```
## # A tibble: 8 x 4
##   target      n cv_error    df
##   <chr>   <dbl>   <dbl> <dbl>
## 1 SSC3044   136     0.382    13
## 2 SSC3038   272     0.398     7
## 3 SCI2040    29     0.546     9
## 4 COR1004  1998     0.669    22
## 5 COR1002  2067     0.998    20
## 6 SCI2010   417     1.41    18
## 7 SCI2018   198     1.62    14
## 8 SCI3003    30     1.83    14
```

3 Conformal Prediction setp by step

3.1 Cross-validation

3.2 Training, calibration and test sets

3.3 Fit lasso on training set

3.4 Compute non-conformity scores on callibration test

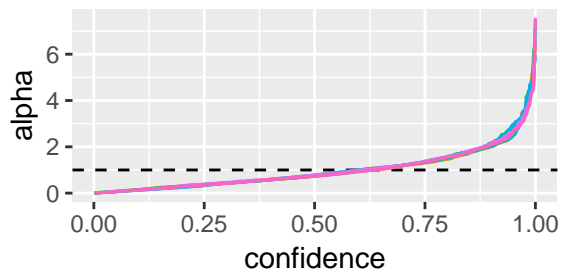
3.5 Evaluate coverage on test set

Grade prediction on test set

4 Loop through all courses

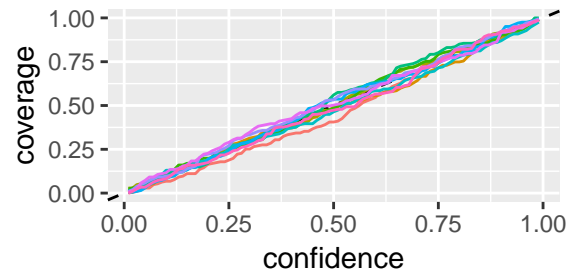
COR1002 (n = 2067)

A line per CV calibration sets



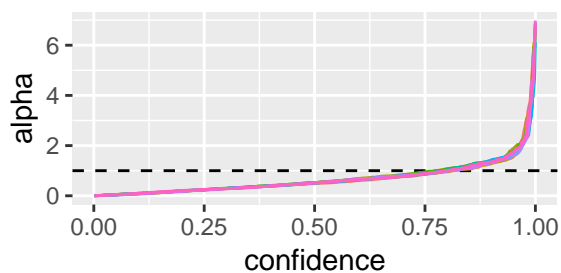
COR1002 (n = 2067)

A line per CV test sets



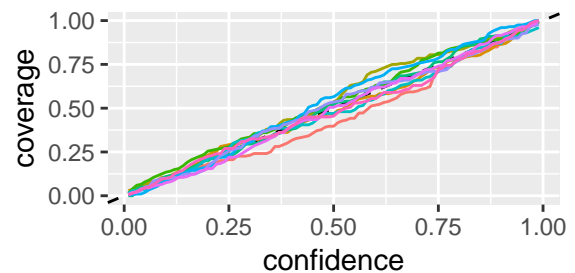
COR1004 (n = 1998)

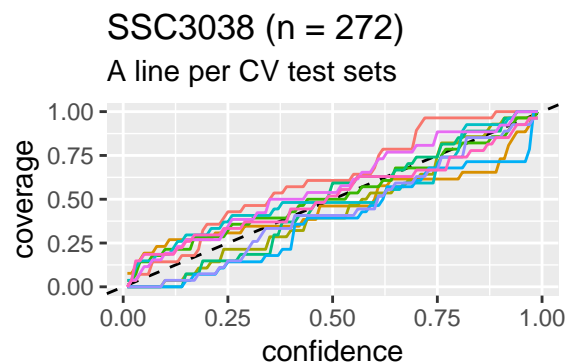
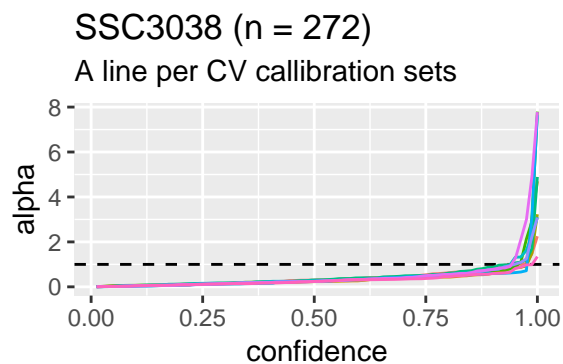
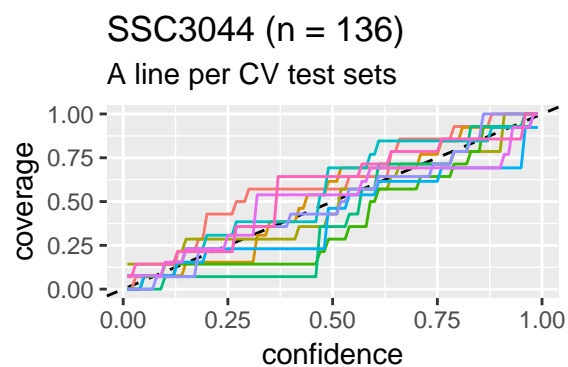
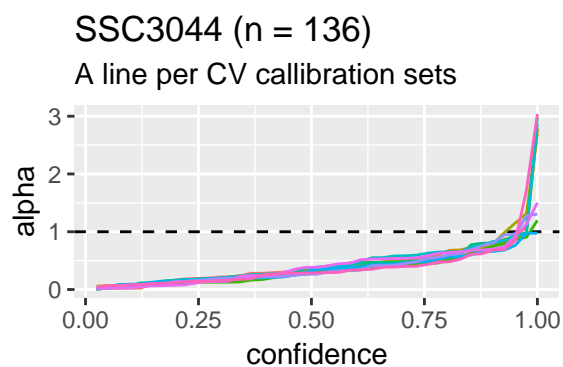
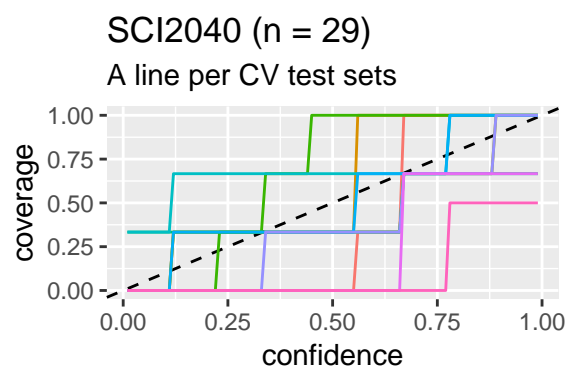
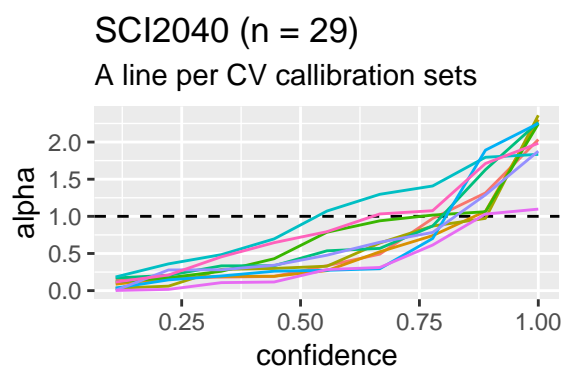
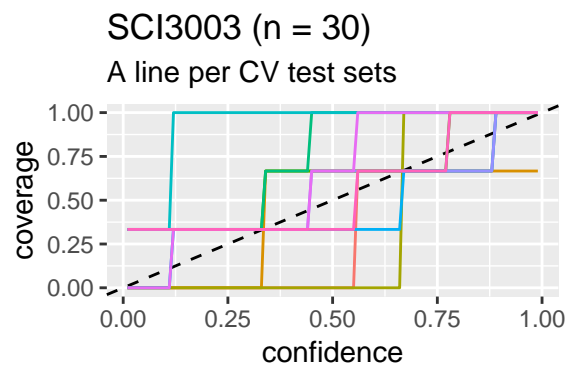
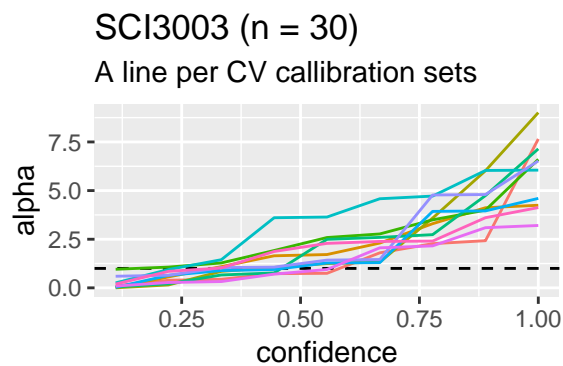
A line per CV calibration sets

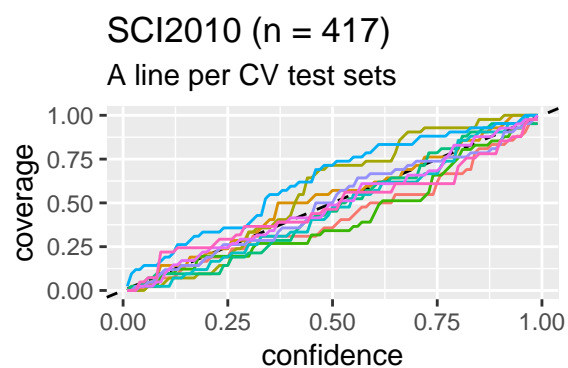
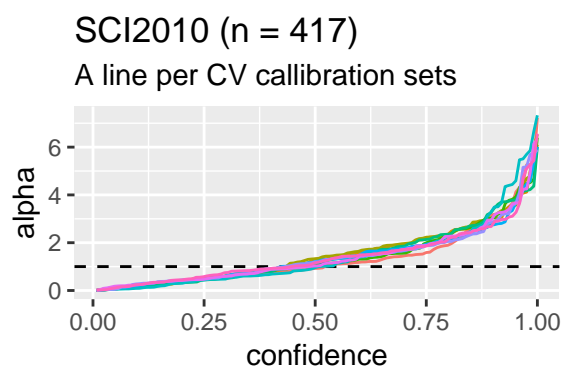
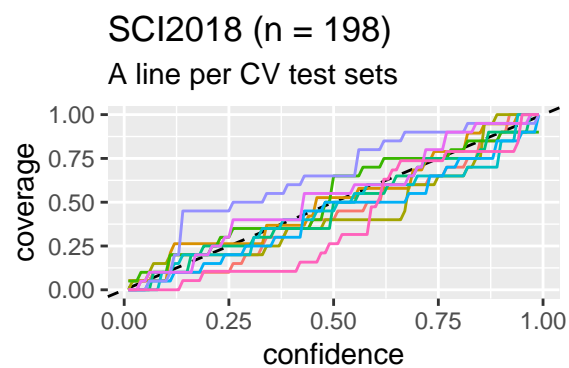
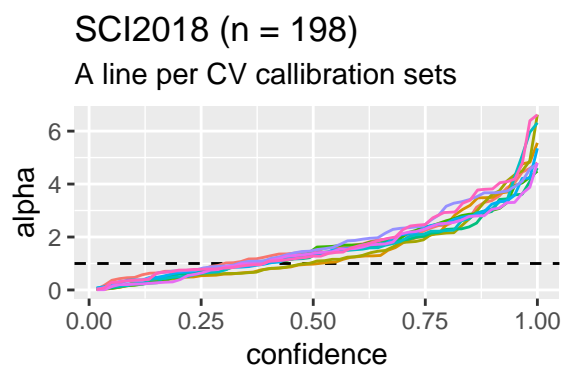


COR1004 (n = 1998)

A line per CV test sets







5 Figures