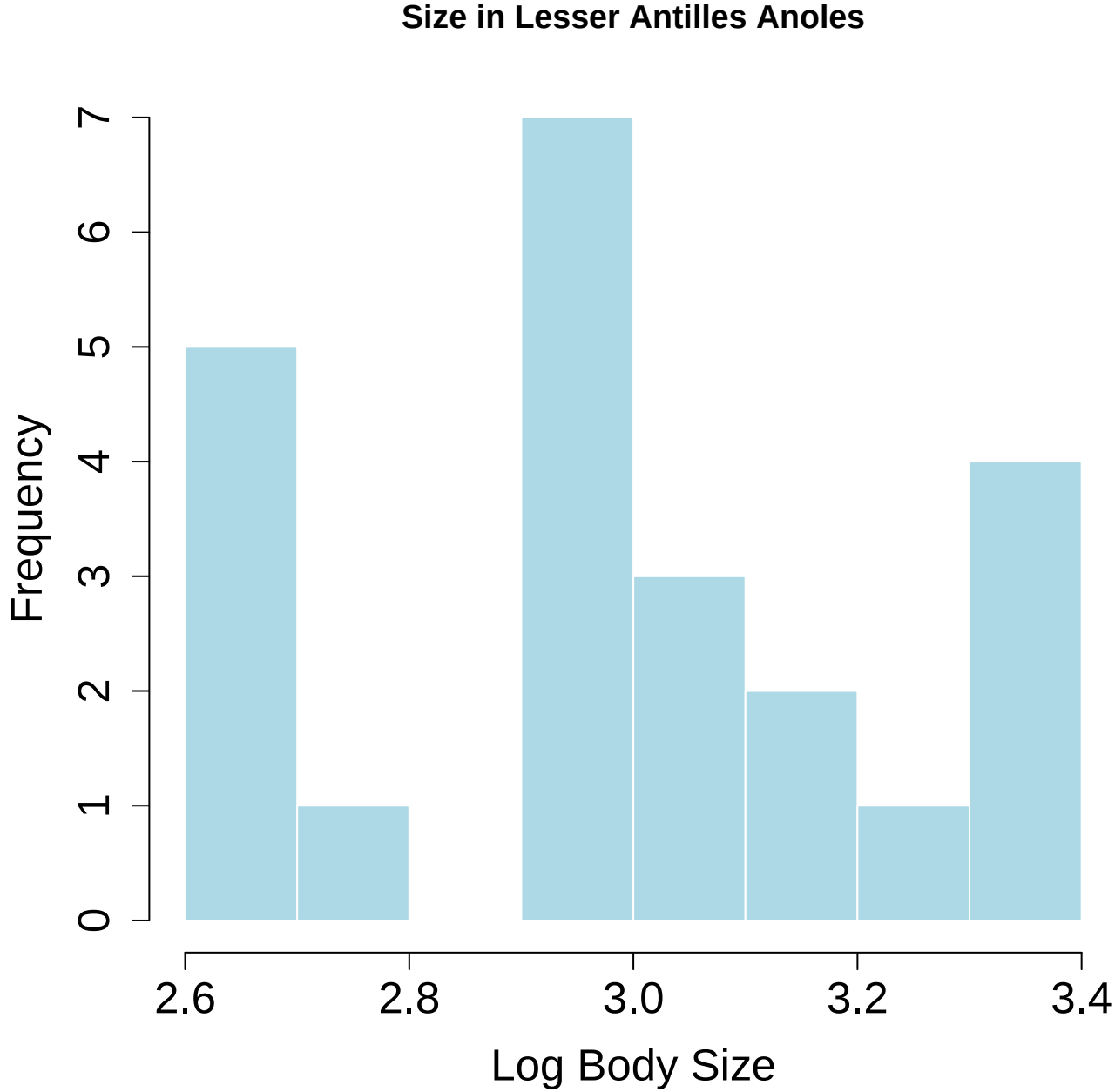
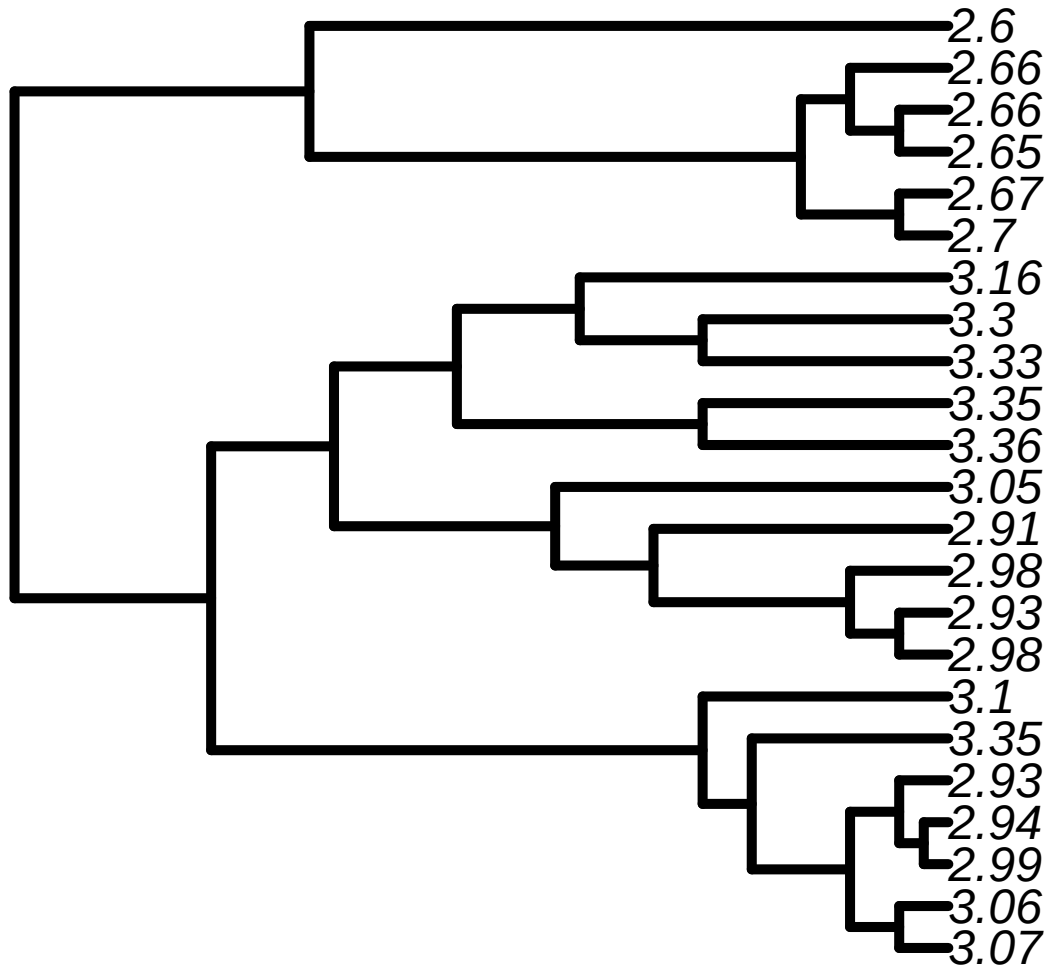
A new phylogenetic comparative method: detecting niches and transitions with continuous characters.

Carl Boettiger

UC Davis

June 26, 2010


Size in Lesser Antilles Anoles
Frequency
0
1
2
3
4
5
6
7
2.6
2.8
3.0
3.2
3.4
Log Body Size


2.6
2.66
2.66
2.65
2.67
2.7
3.16
3.3
3.33
3.35
3.36
3.05
2.91
2.98
2.93
2.98
3.1
3.35
2.93
2.94
2.99
3.06
3.07

Goals

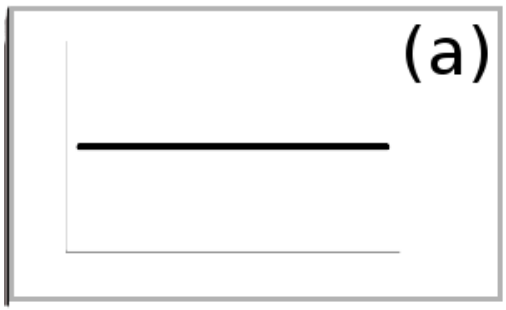
Goals

1. Demonstrate selecting models by information criteria is inadequate

Goals

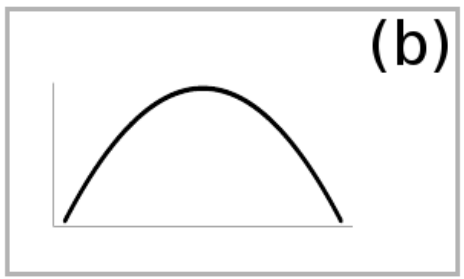
1. Demonstrate selecting models by information criteria is inadequate
2. I'll propose a more robust framework

Types of models



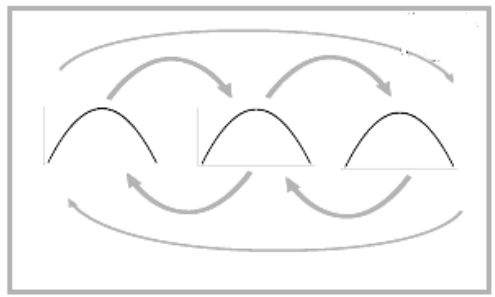
Brownian Motion

$$dX = \sigma dB$$



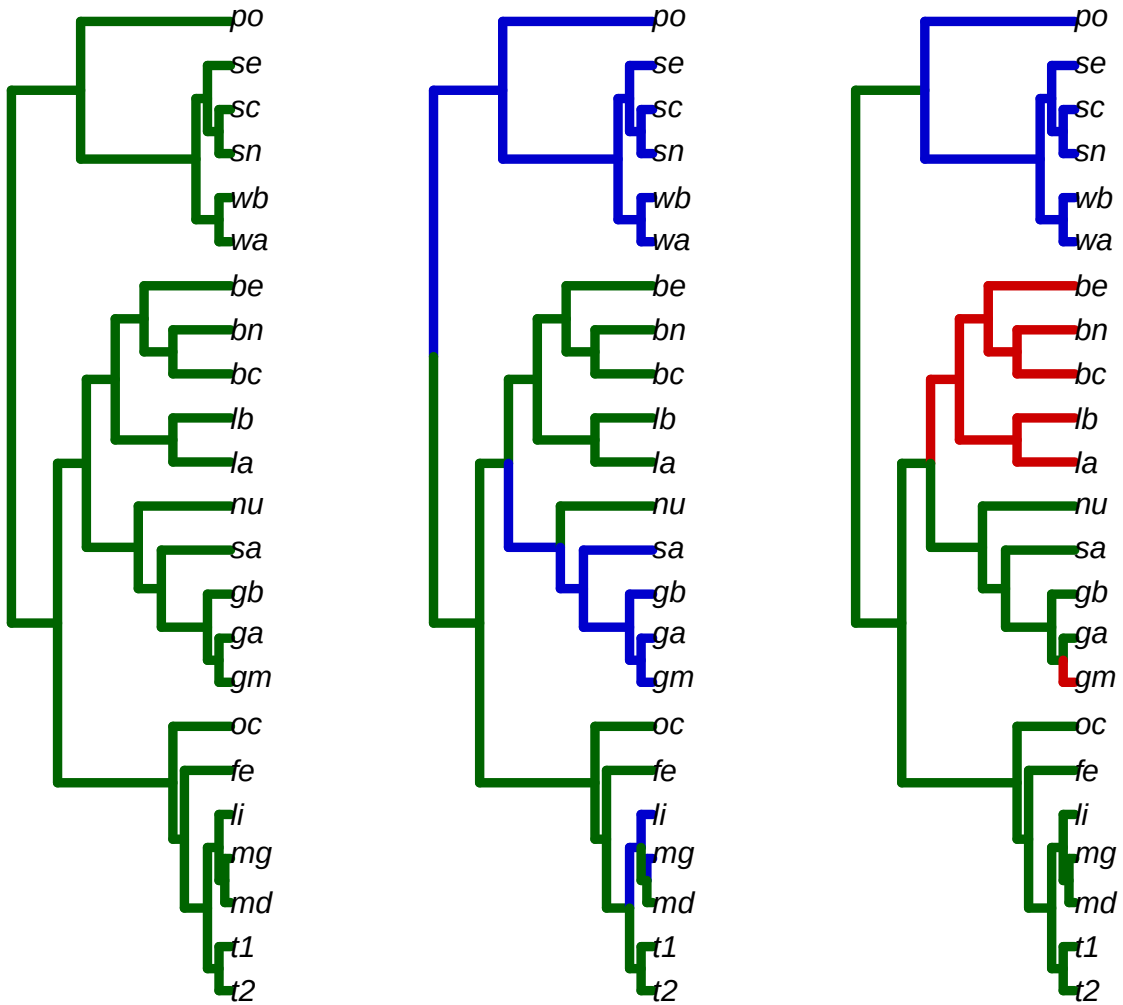
Ornstead-Uhlenbeck

$$dX = \alpha(\theta - X)dt + \sigma dB$$

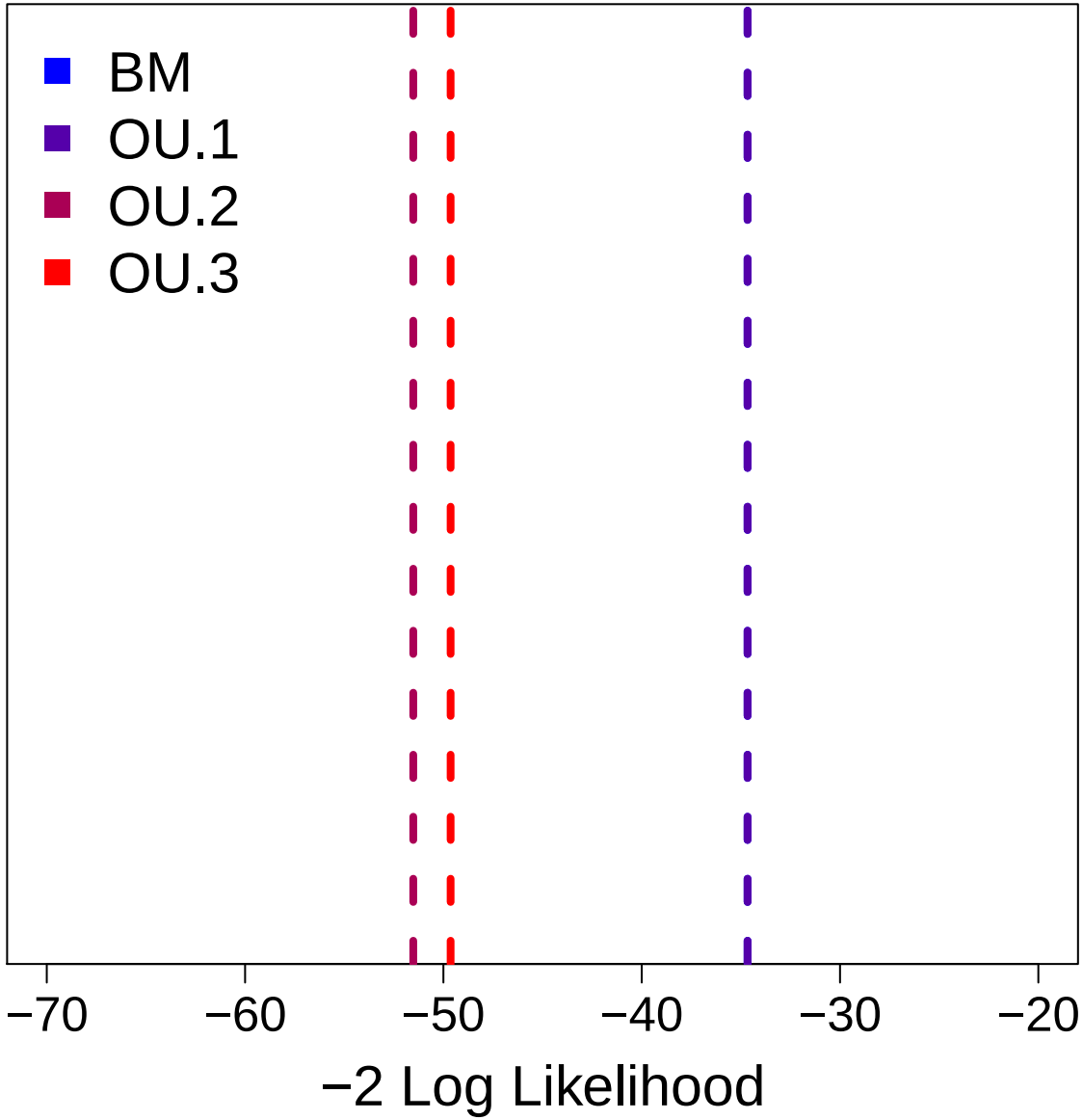


Regimes

Comparing Models

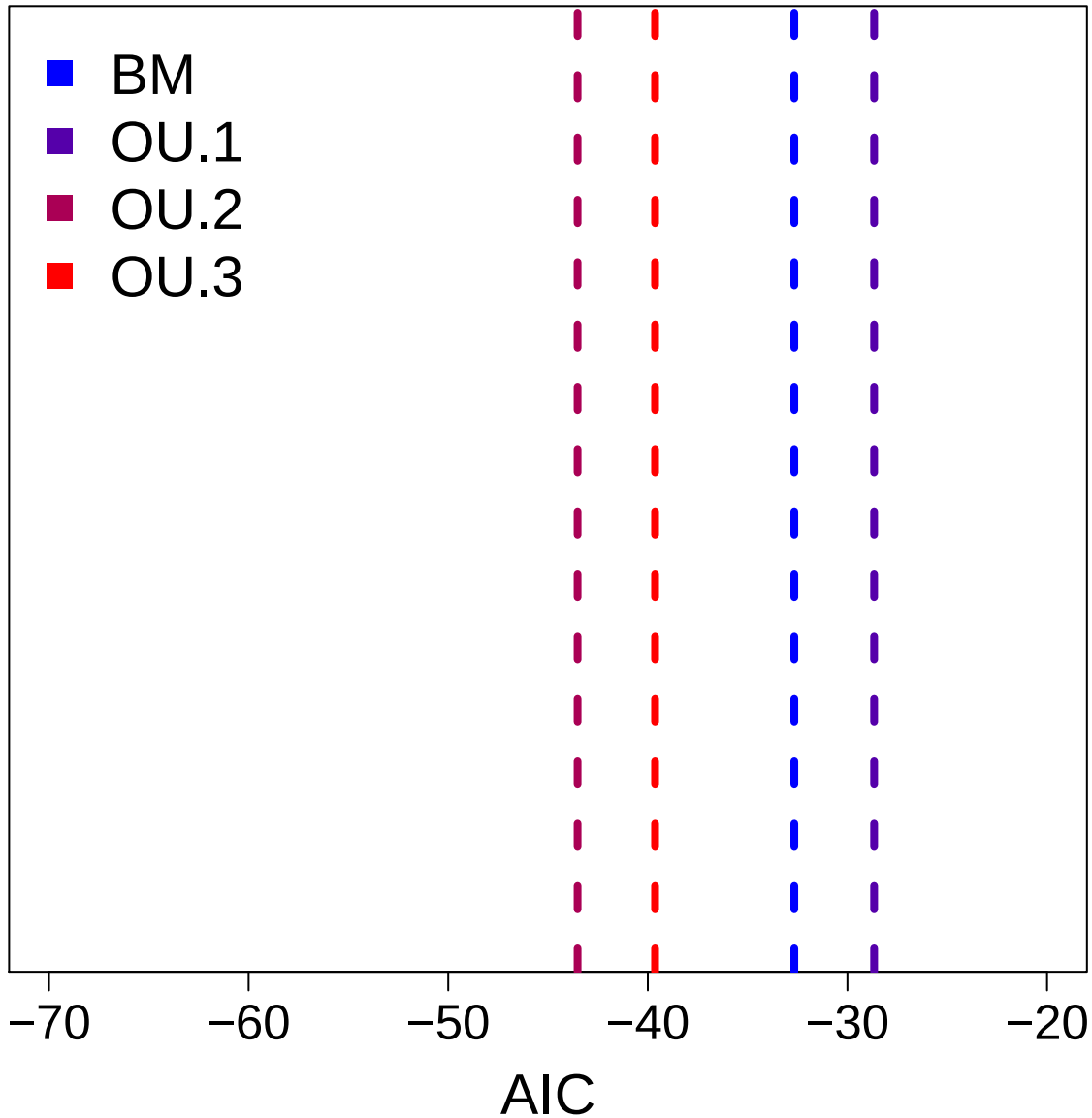


Comparing Models



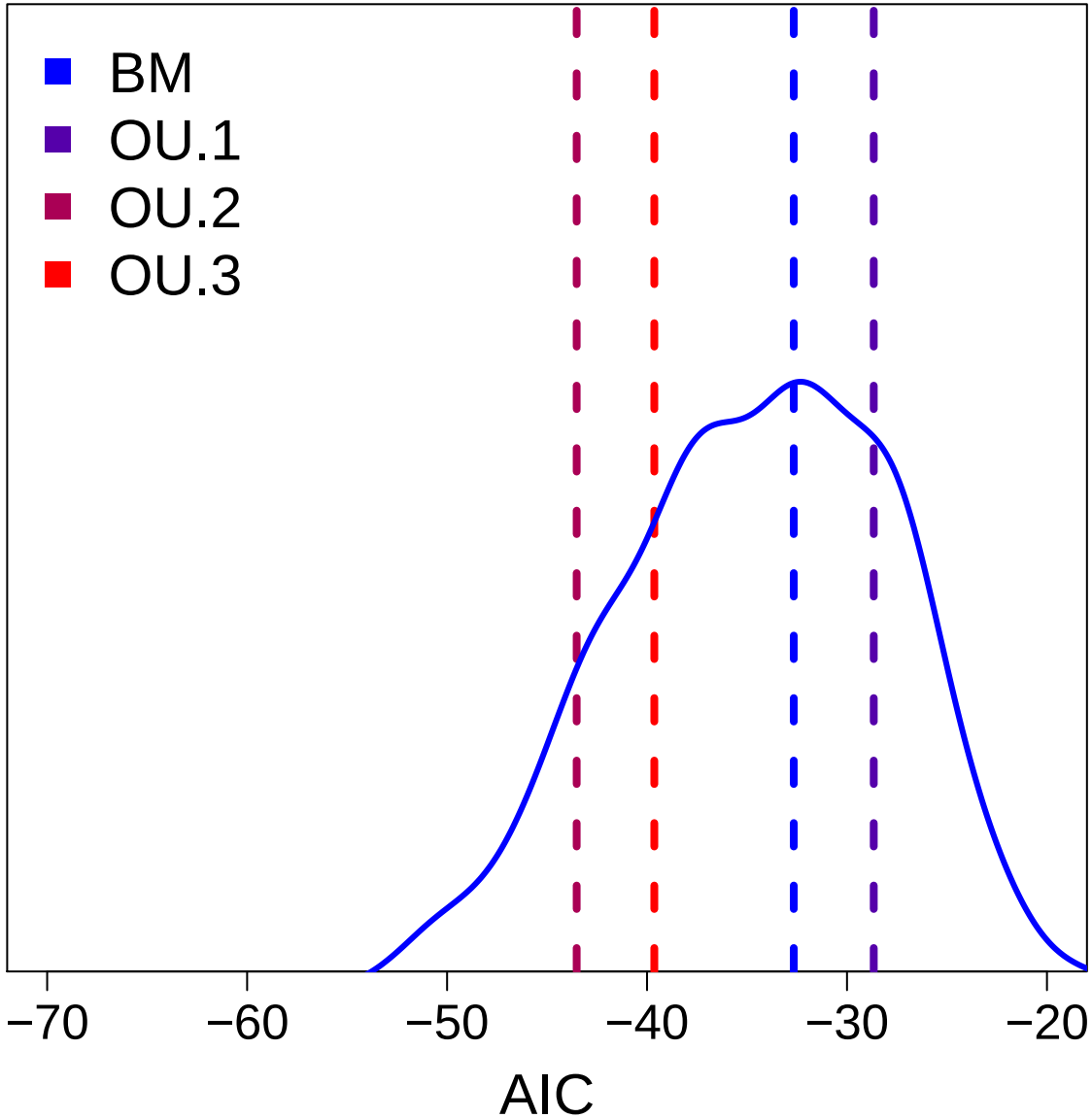
(Better Scores)

Comparing Models: AIC



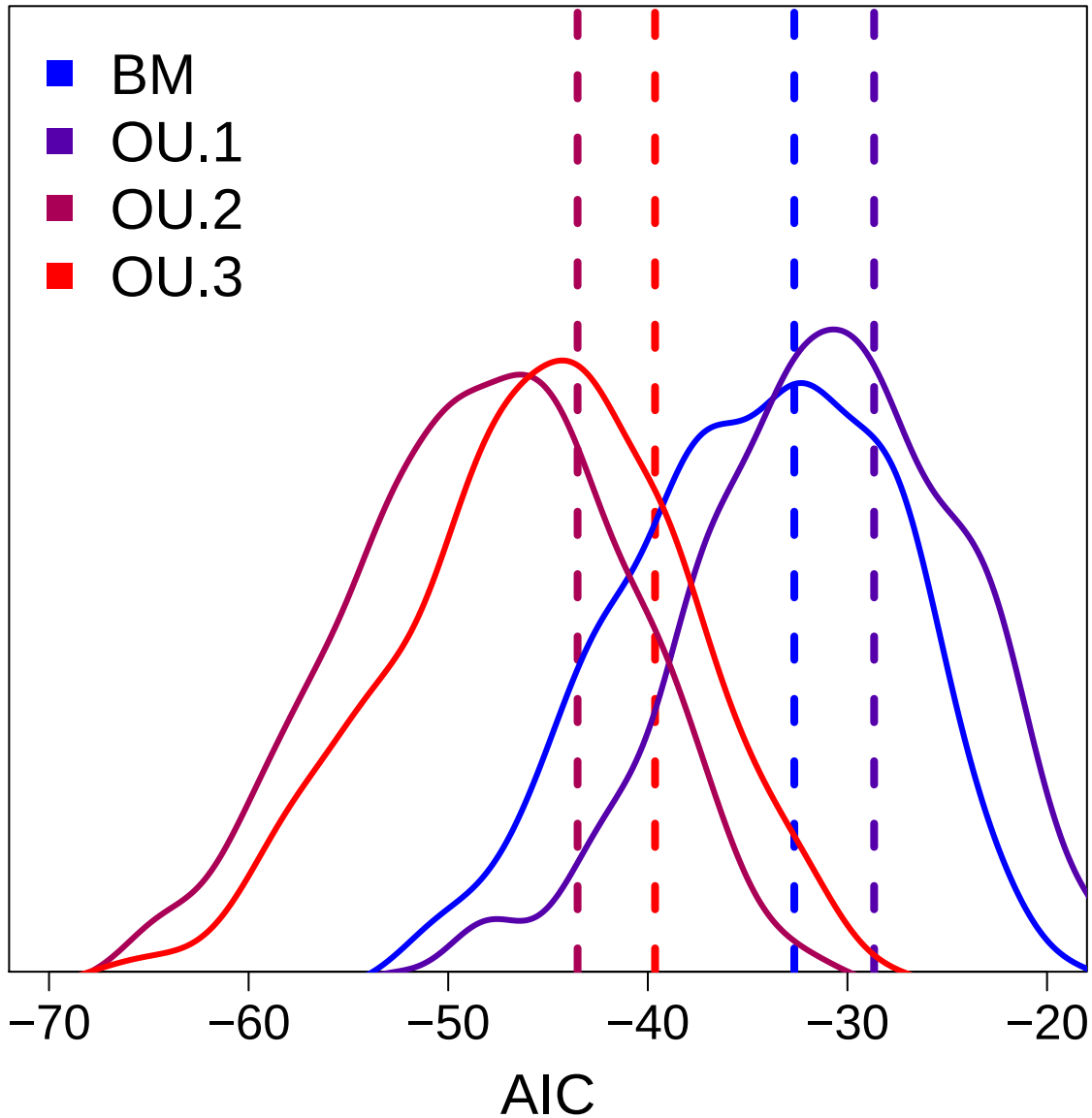
(Better Scores)

Comparing Models: Estimating Uncertainty



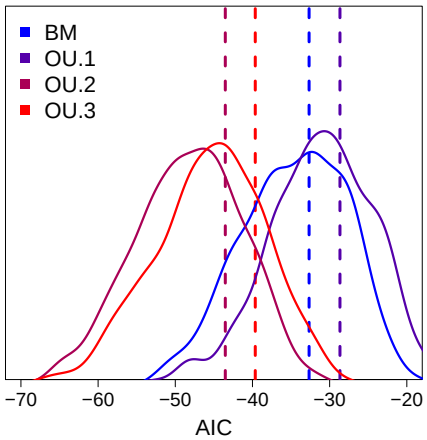
(Better Scores)

Comparing Models: Uncertainty dominates

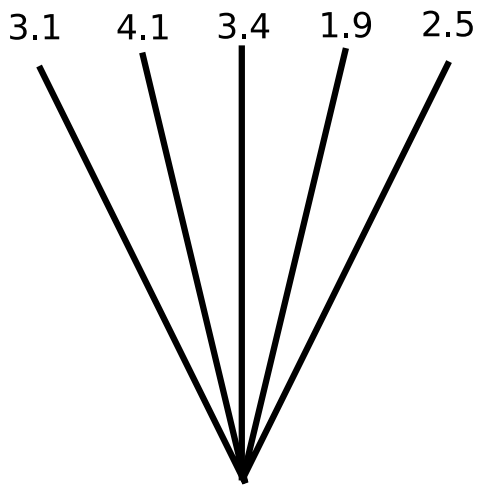


(Better Scores)

Sources of this uncertainty

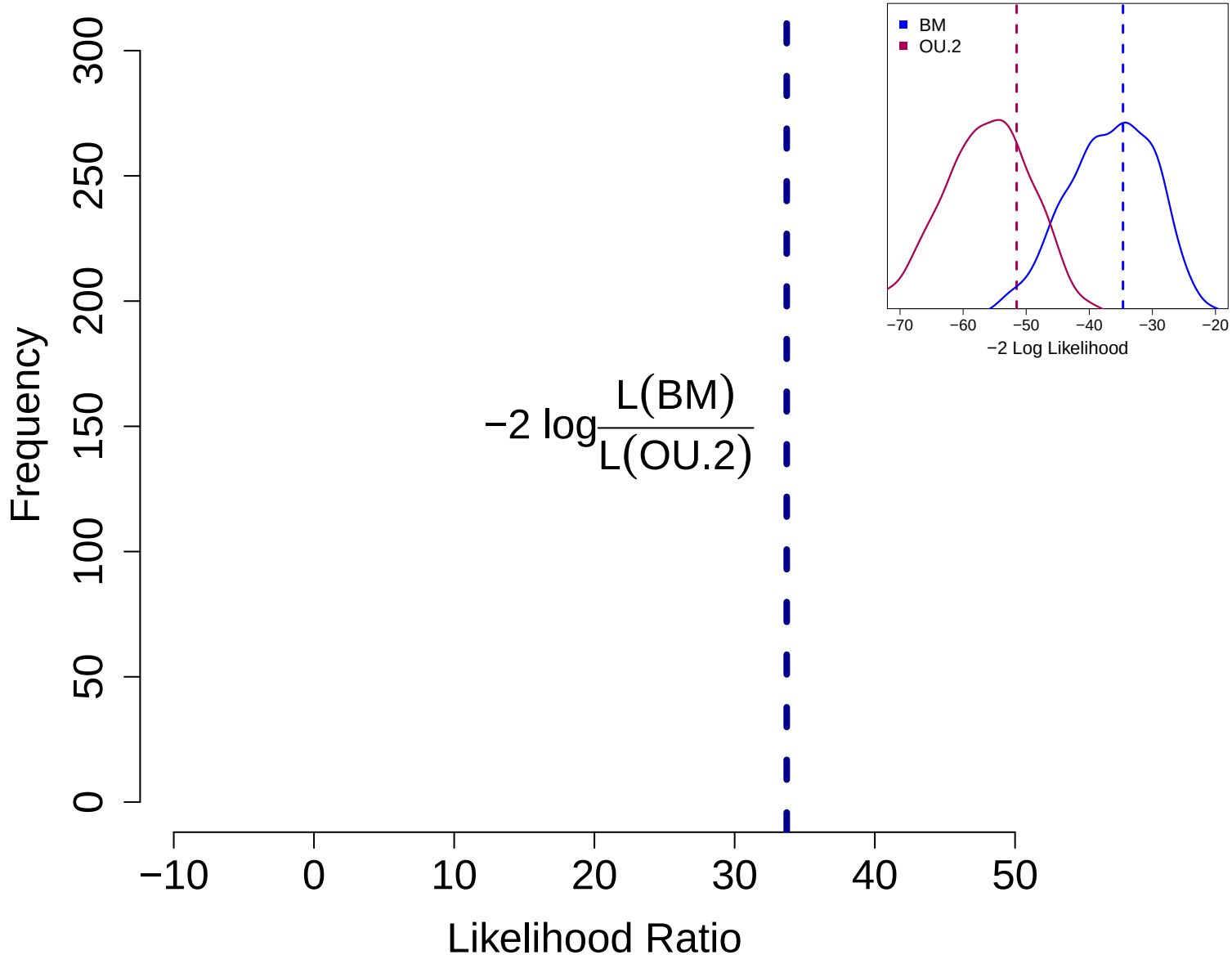


- Small datasets
- Uninformative topology
- Model details (*i.e.* high rates)

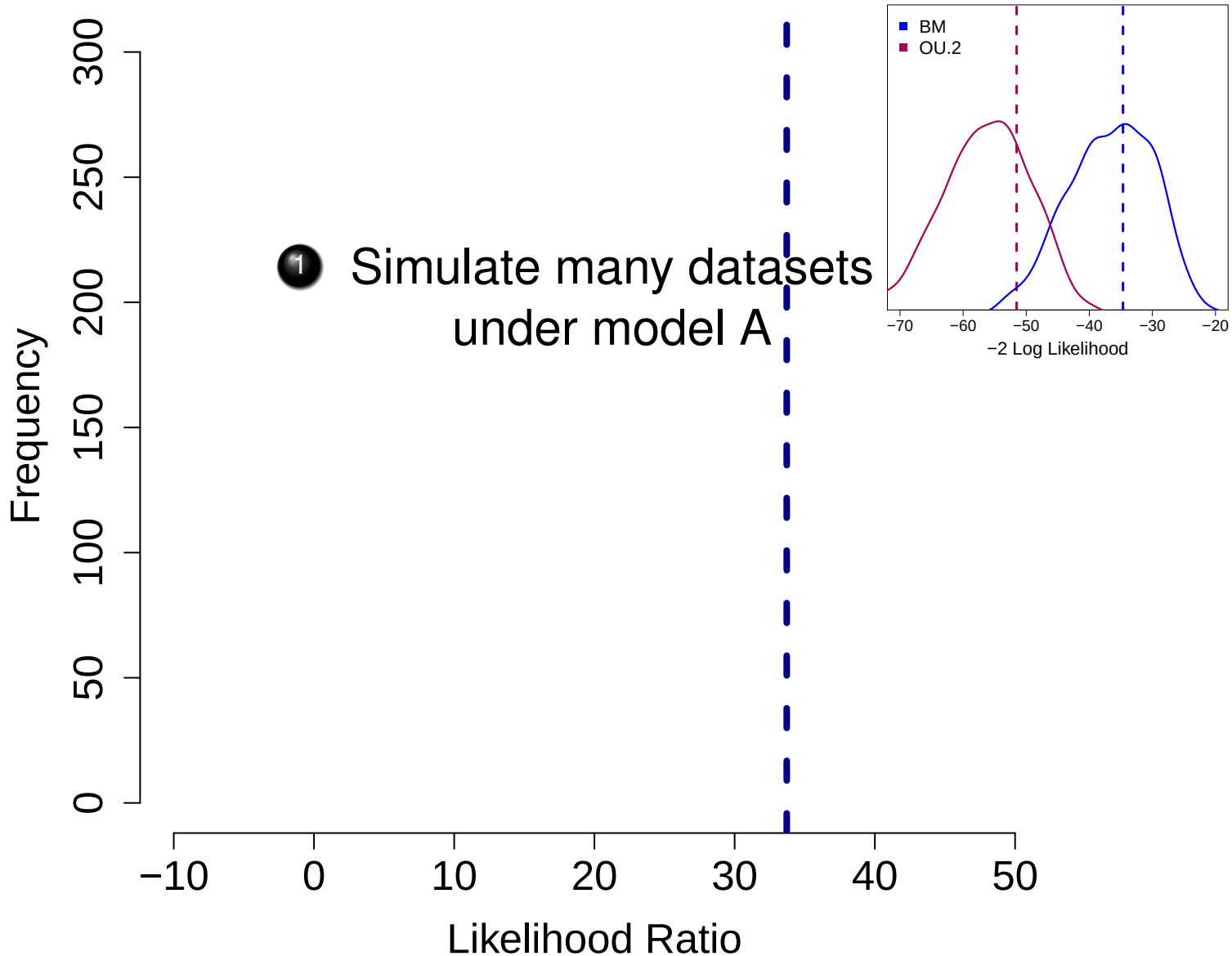


Information criteria alone may be misleading

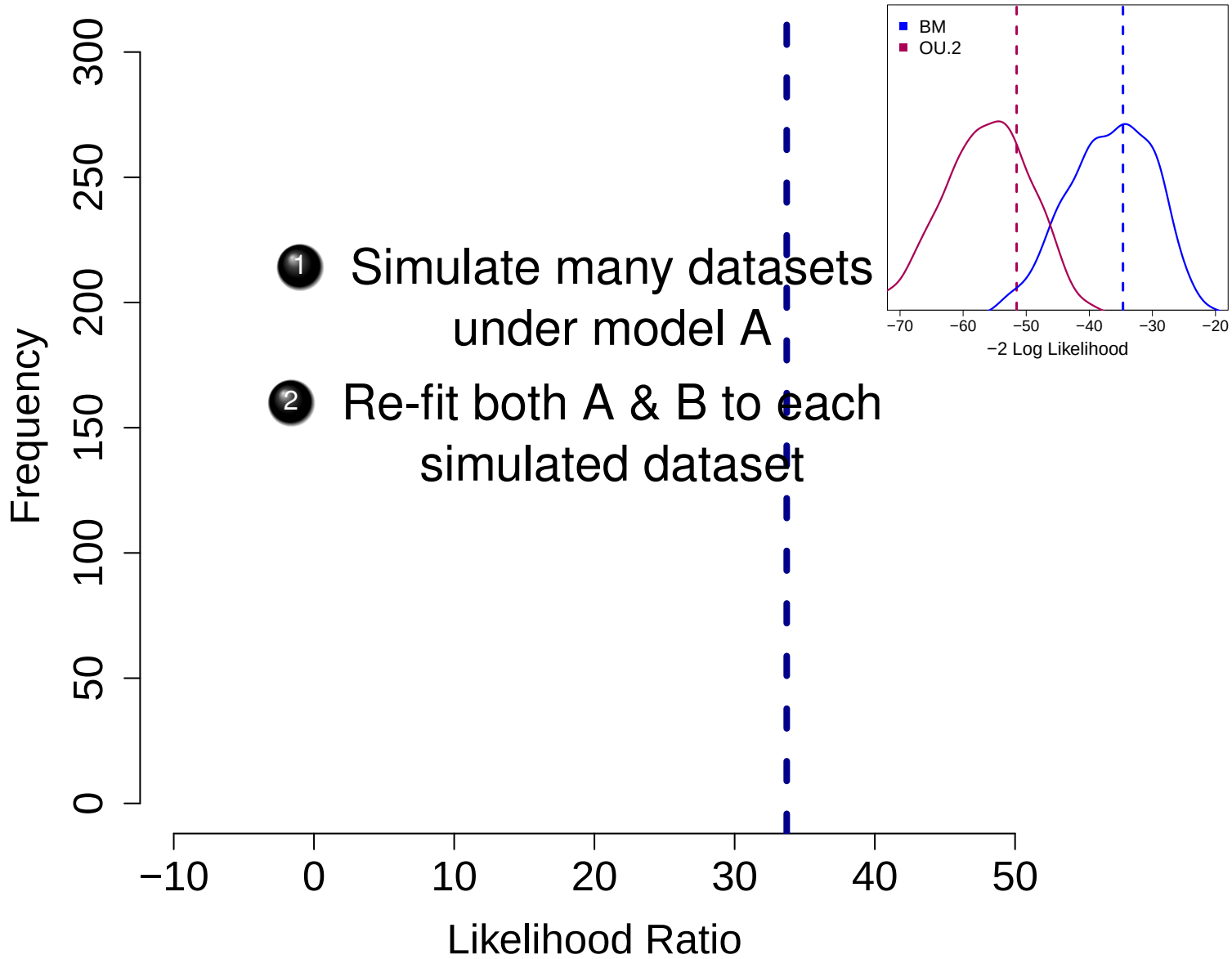
A Better Way: Comparing Models Directly



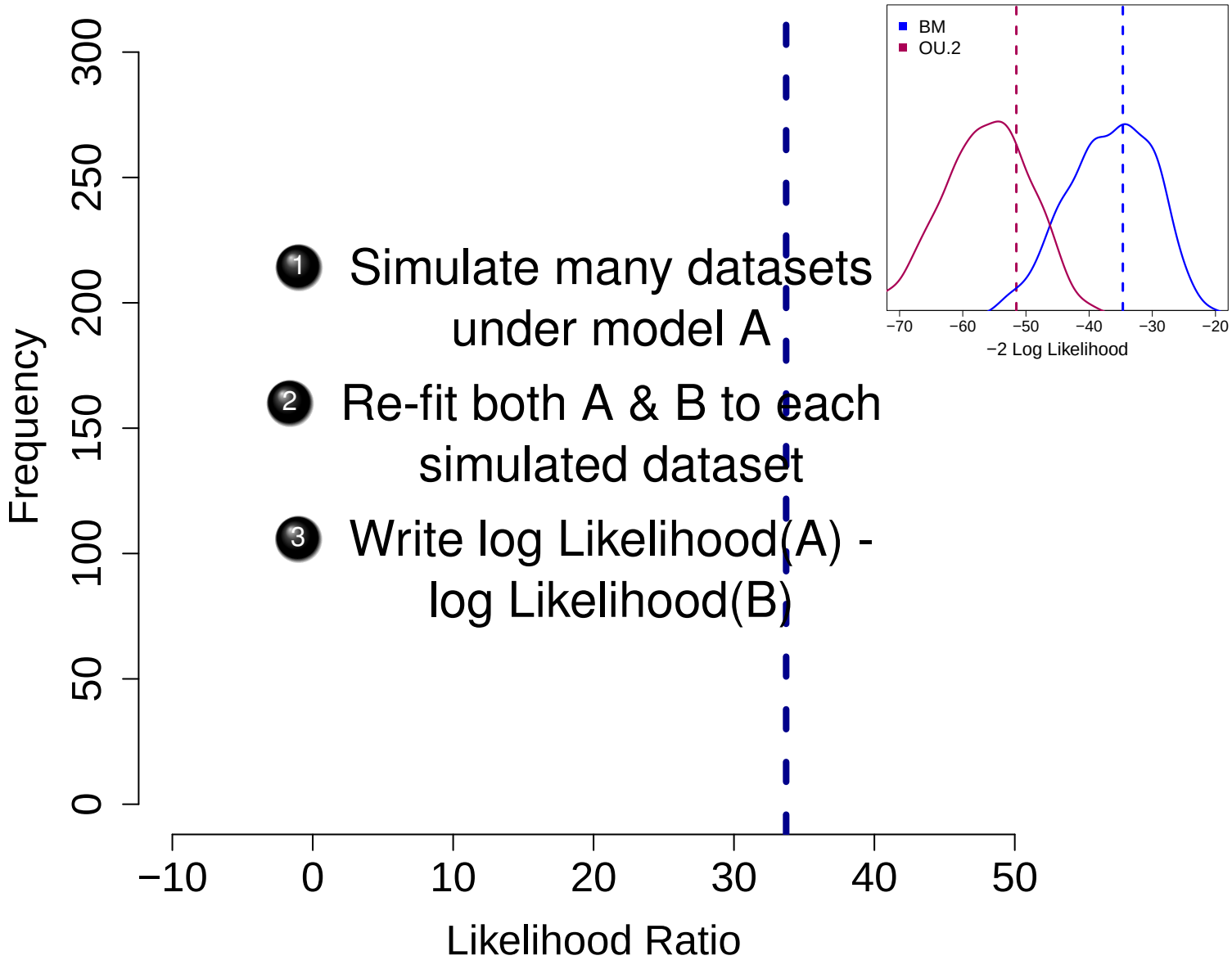
Method


1
Simulate many datasets
under model A
Frequency
0
50
100
150
200
250
300
−10
0
10
20
30
40
50
Likelihood Ratio
BM
OU.2
−70
−60
−50
−40
−30
−20
−2 Log Likelihood

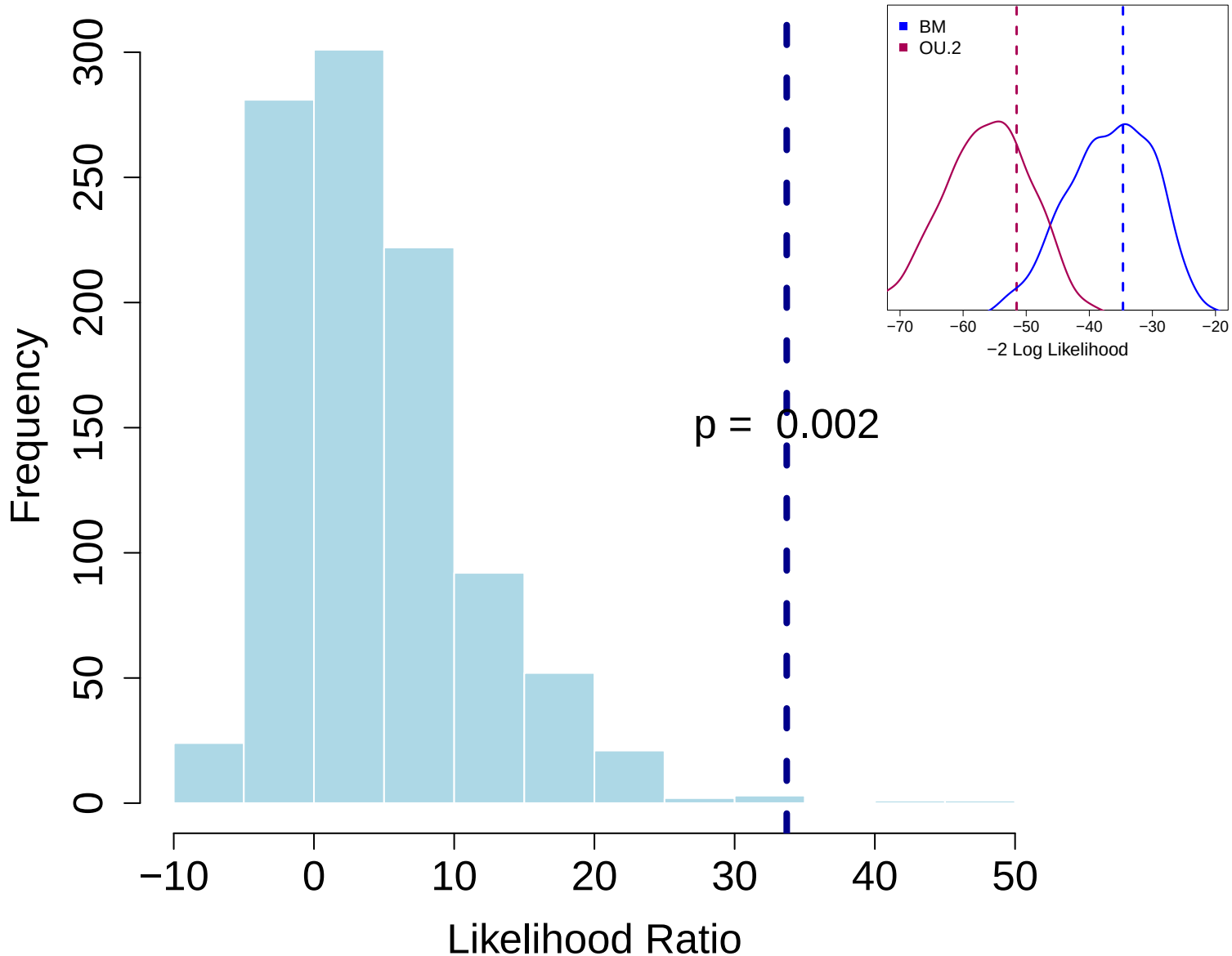
Method


Frequency
0
50
100
150
200
250
300
−10
0
10
20
30
40
50
Likelihood Ratio
1 Simulate many datasets under model A
2 Re-fit both A & B to each simulated dataset
BM
OU.2
−70
−60
−50
−40
−30
−20
−2 Log Likelihood

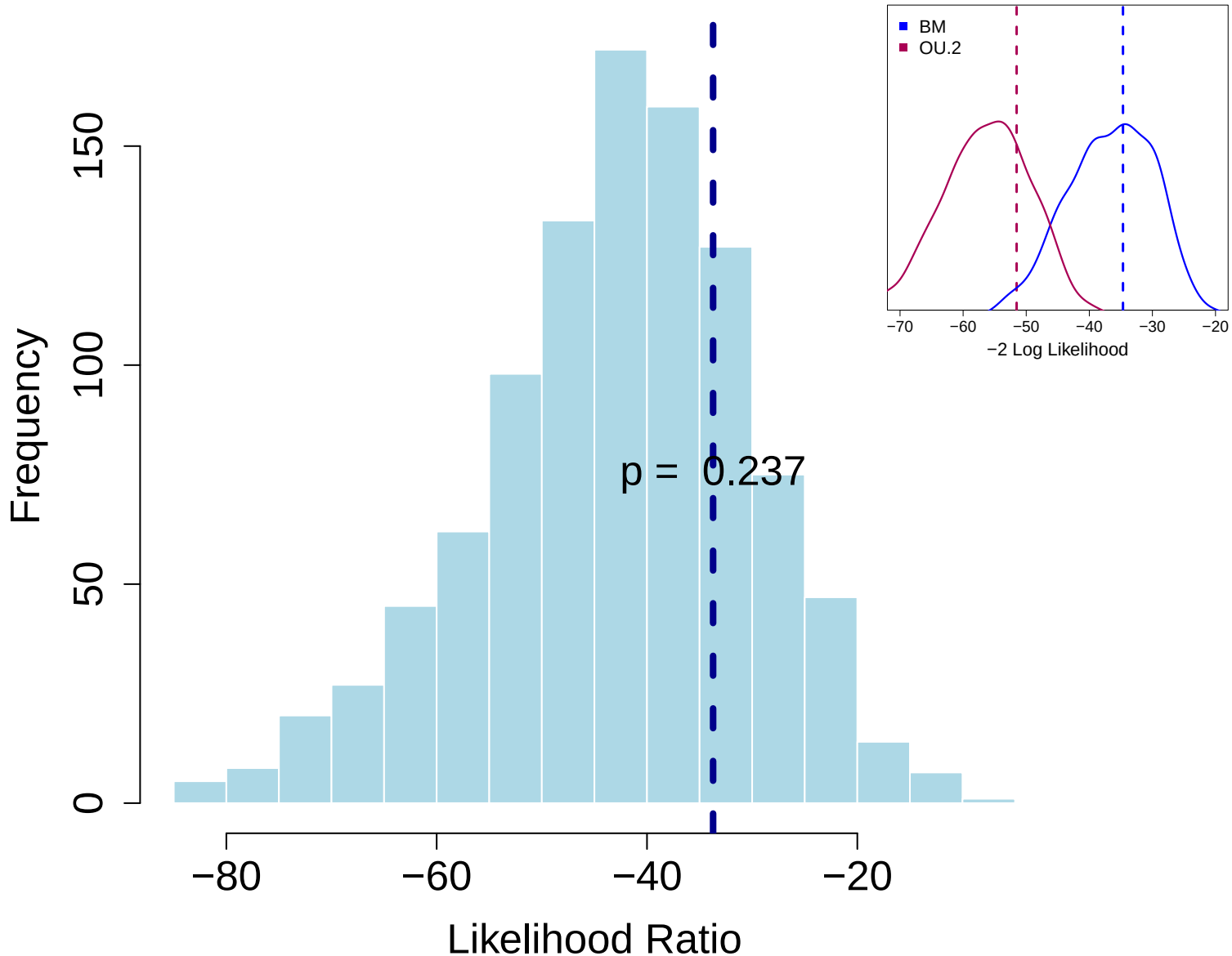
Method


Frequency
0
50
100
150
200
250
300
−10
0
10
20
30
40
50
Likelihood Ratio
1 Simulate many datasets under model A
2 Re-fit both A & B to each simulated dataset
3 Write log Likelihood(A) - log Likelihood(B)
BM
OU.2
−70
−60
−50
−40
−30
−20
−2 Log Likelihood

BM vs OU.2

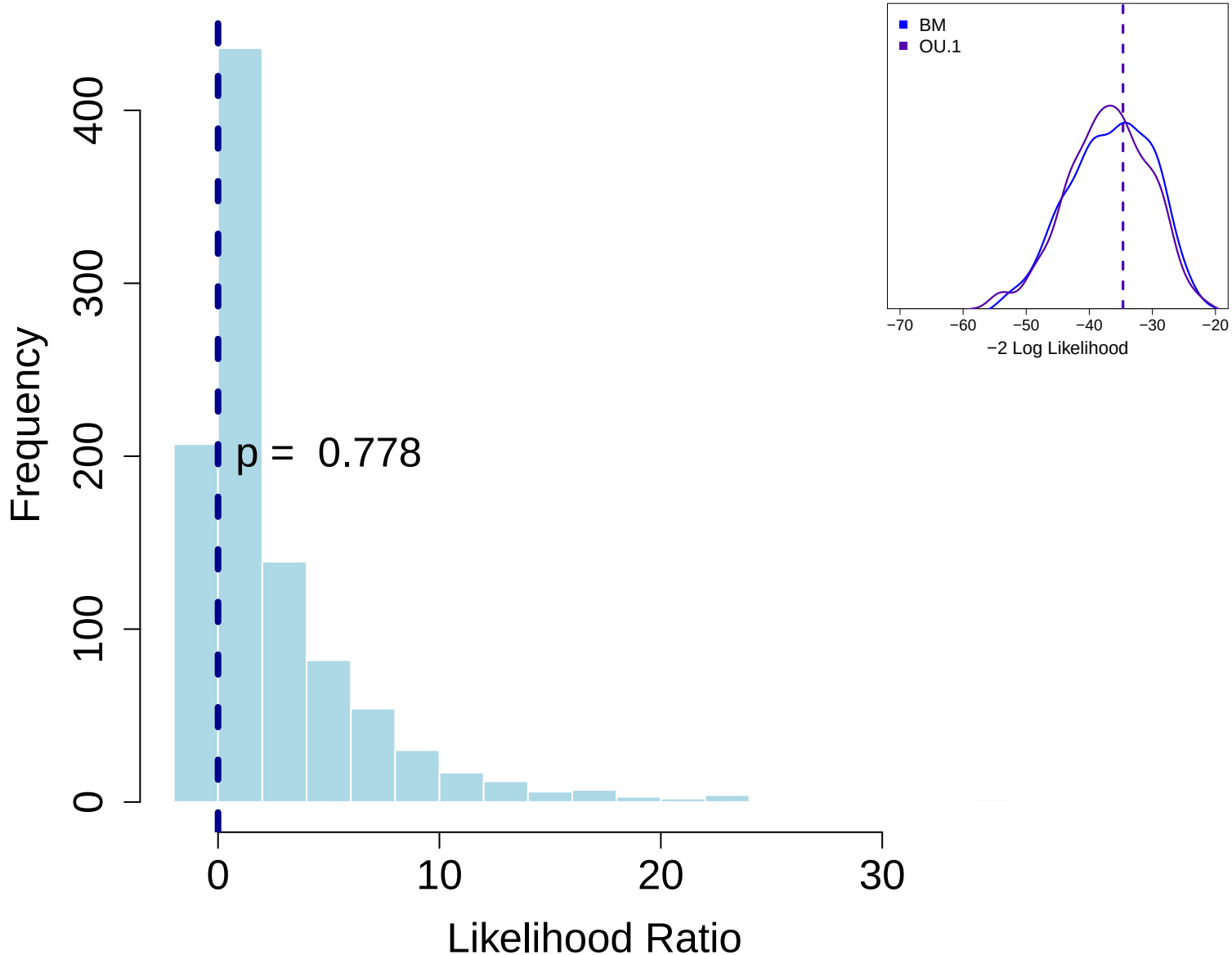

Frequency
0
50
100
150
200
250
300
−10
0
10
20
30
40
50
Likelihood Ratio
p = 0.002
BM
OU.2
−70
−60
−50
−40
−30
−20
−2 Log Likelihood

OU.2 vs BM: Simulating under OU.2

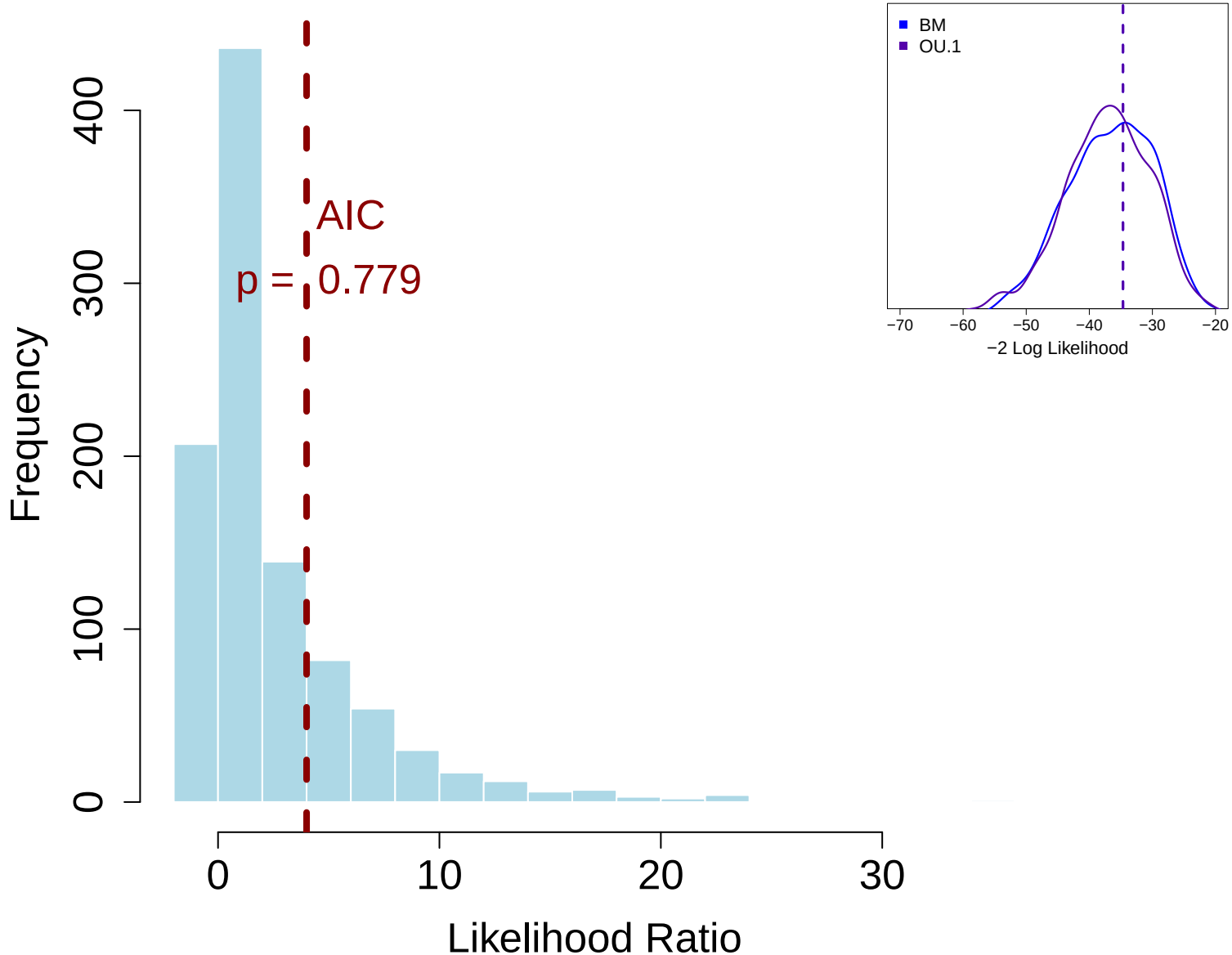


Model A rejects Model B.
Model B doesn't reject Model A.

How about two very similar models? BM vs OU.1

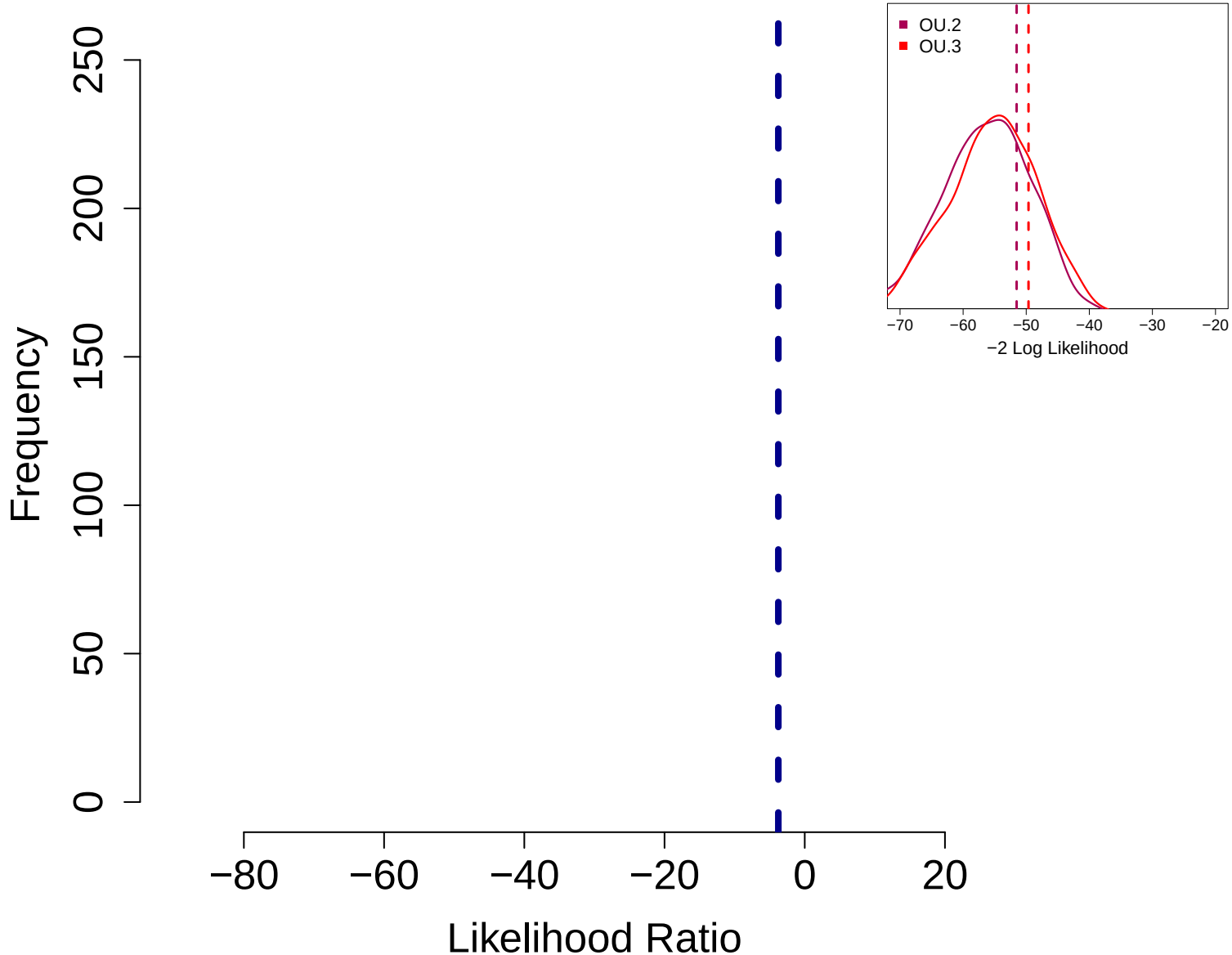


How would AIC rule compare?

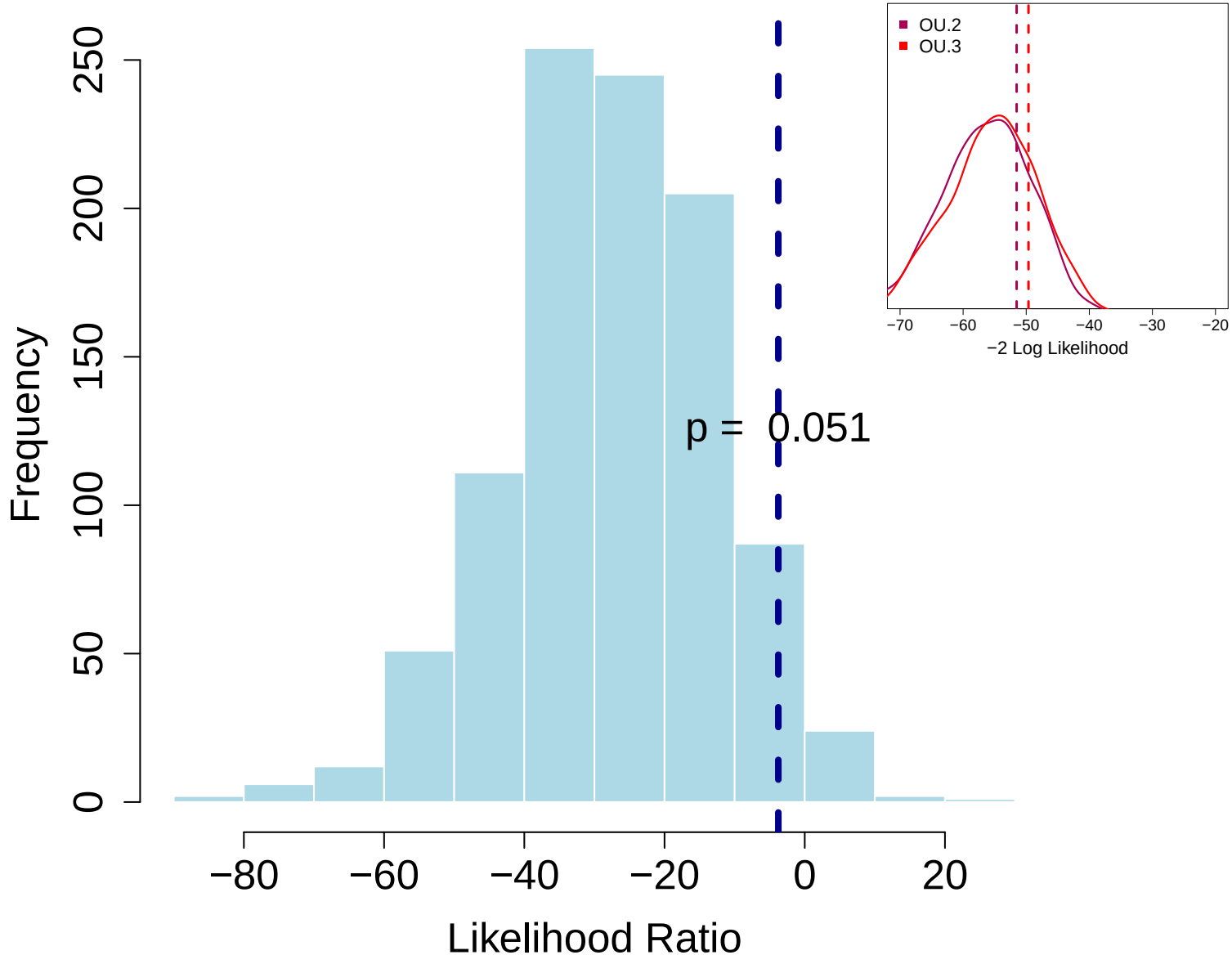


When data is insufficient to distinguish, method can say *“I don’t know”*

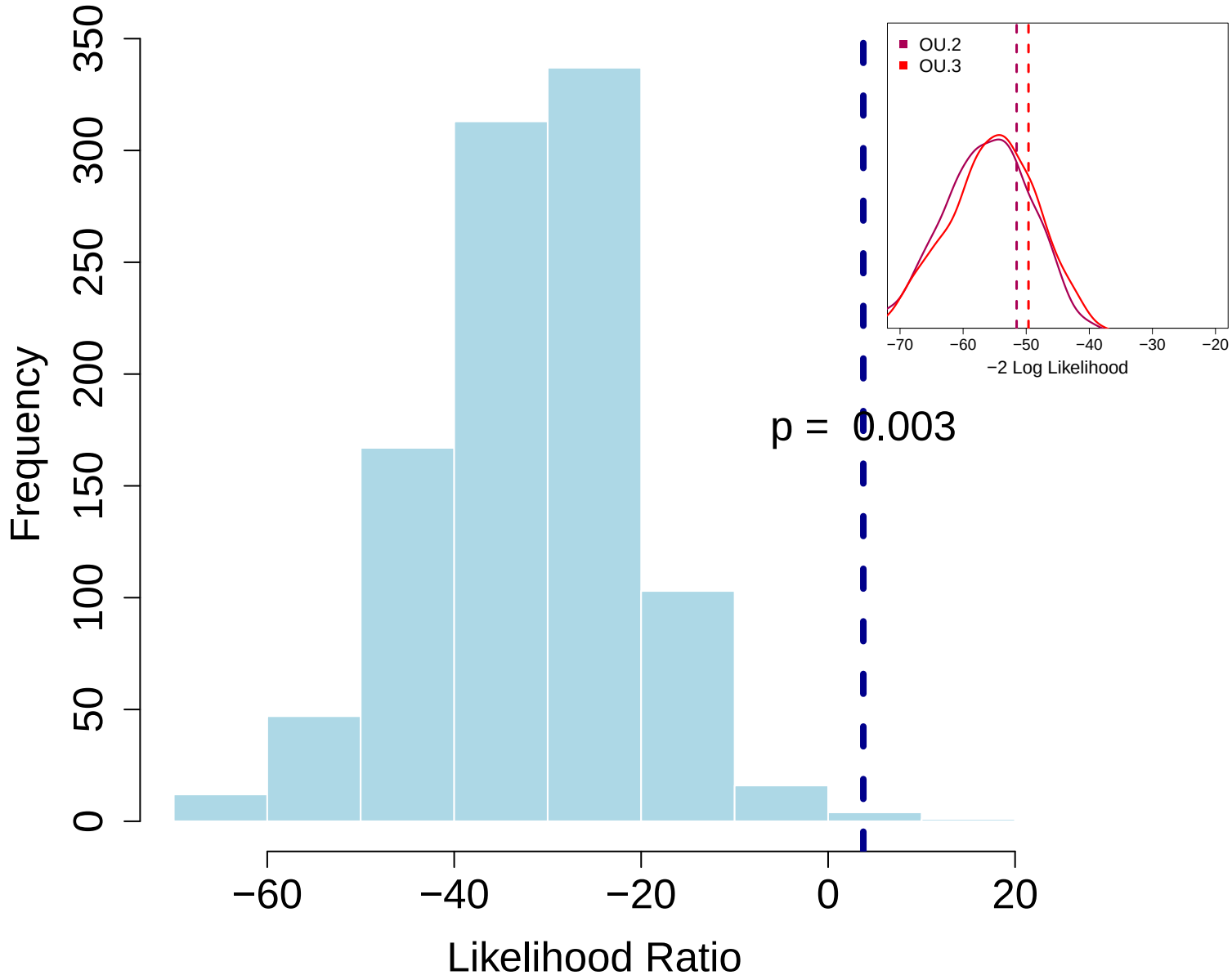
Can we distinguish between OU.2 and OU.3?



Simulate under OU.2 and compare to OU.3 . . .

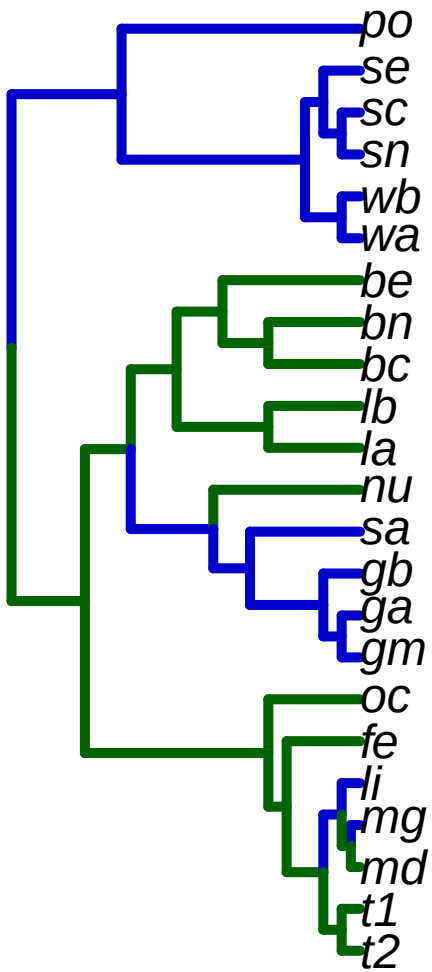


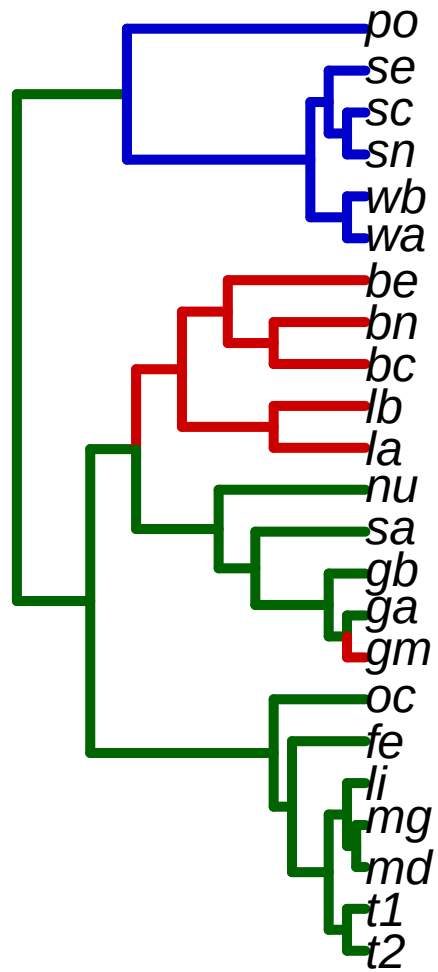
OU.3 vs OU.2: Now we're preferring OU.2!



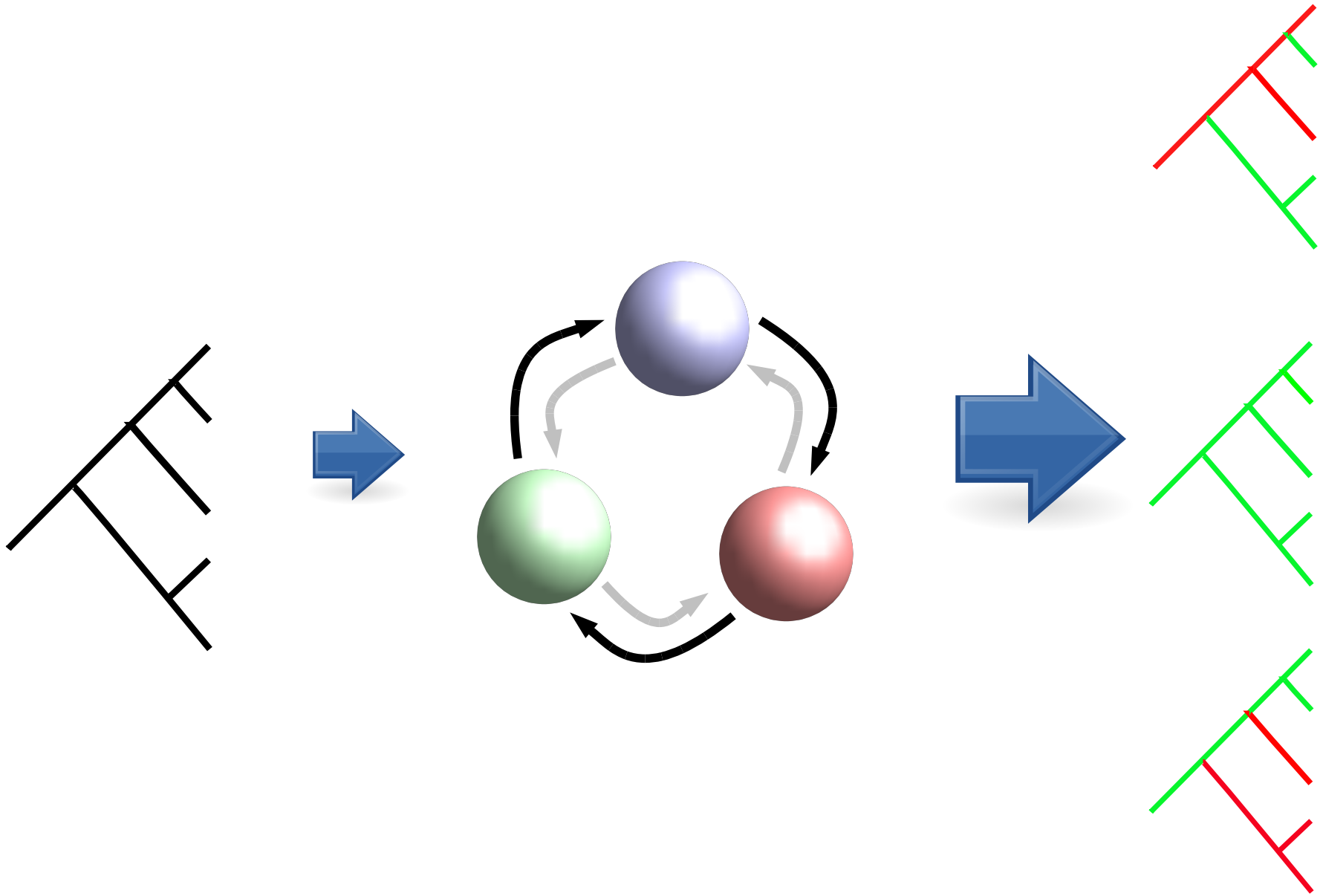
Model A rejects Model B.
Model B rejects Model A.

Non-nested models



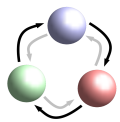


Replacing paintings with a transition model



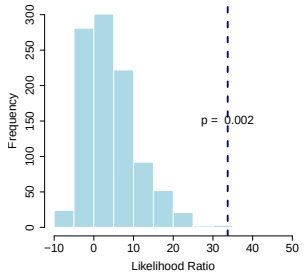
- All models are nested
- Estimate *number* of niches
- Also estimate *rates* of transitions

A hard problem in two easy pieces

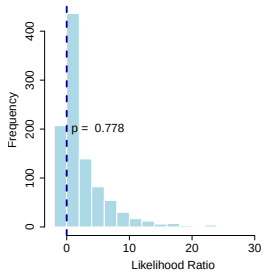
$$P(\text{  } \mid \text{  }) =$$
$$P(\text{  } \mid \text{  })P(\text{} \mid \text{})$$

Summary

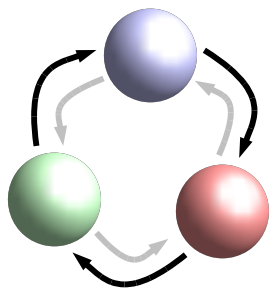
1. Quantifiable, robust model choice



2. Identify when data is insufficient

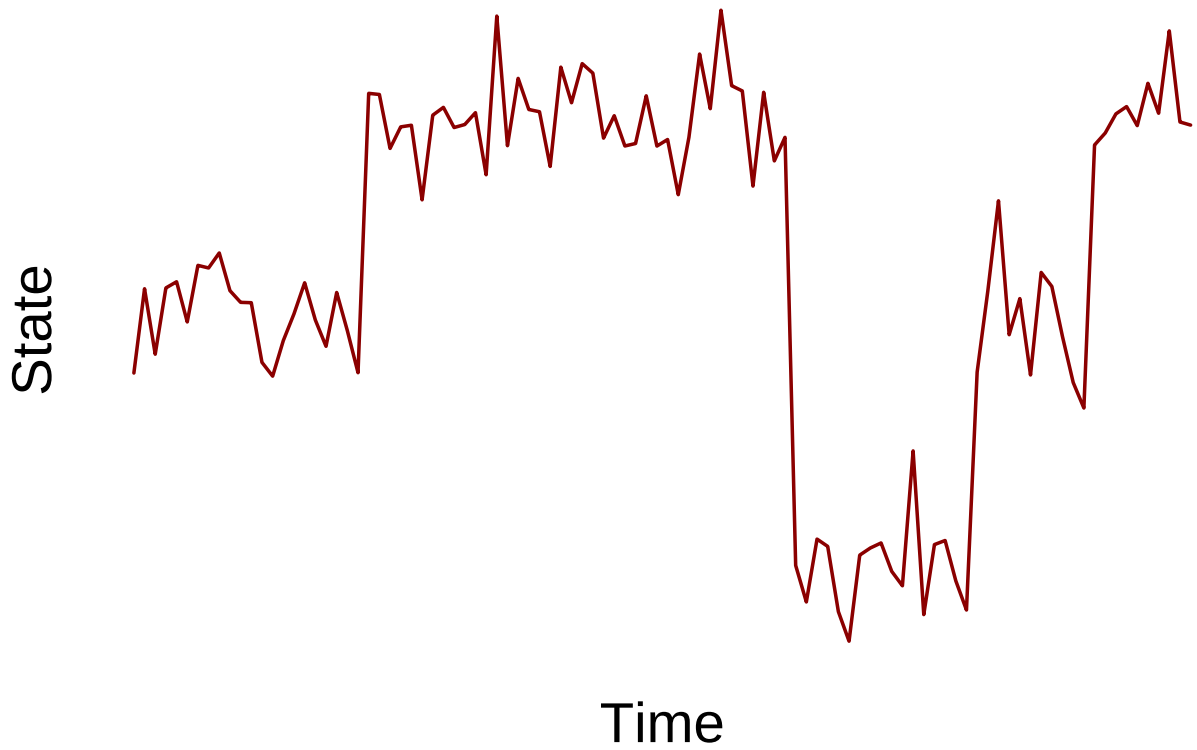


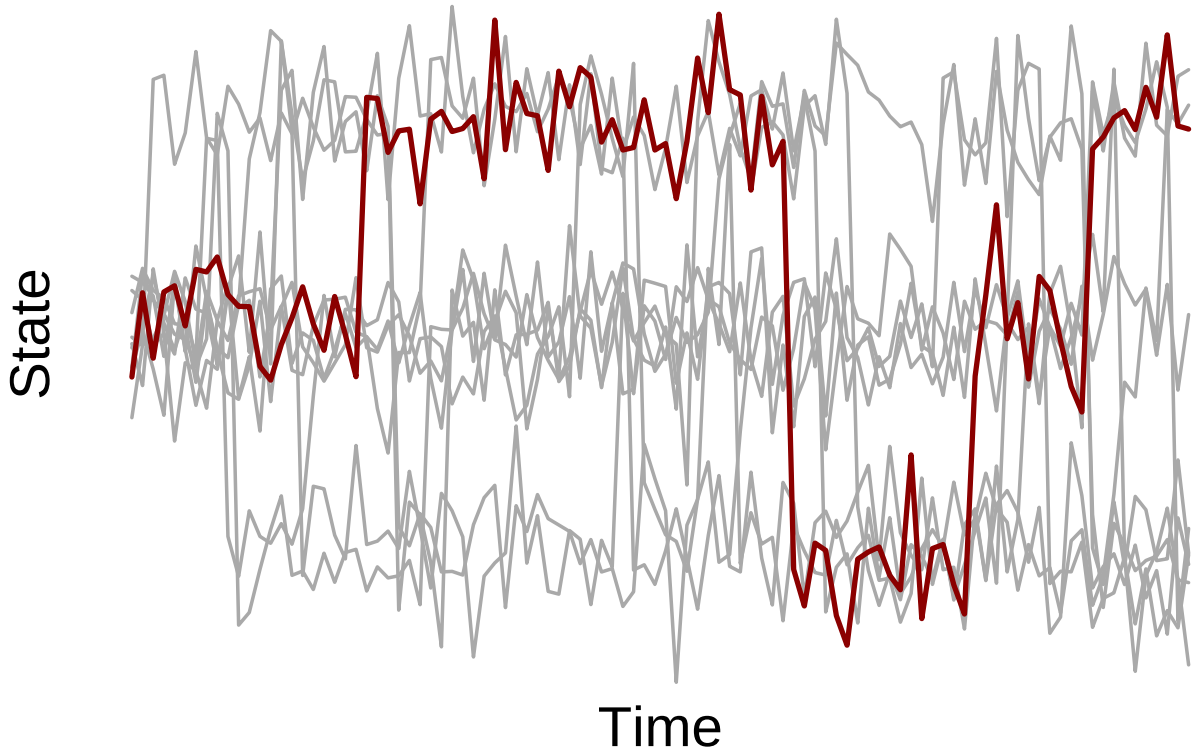
3. New framework avoids painting & non-nested comparison

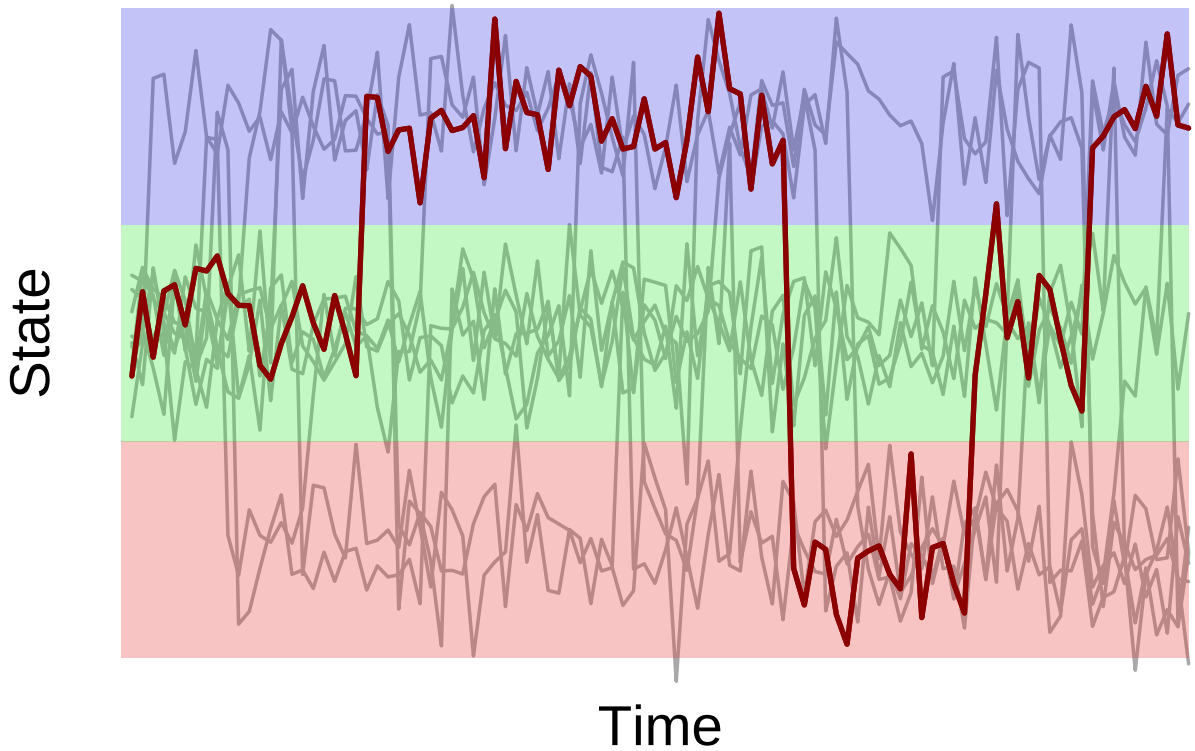


Thanks!

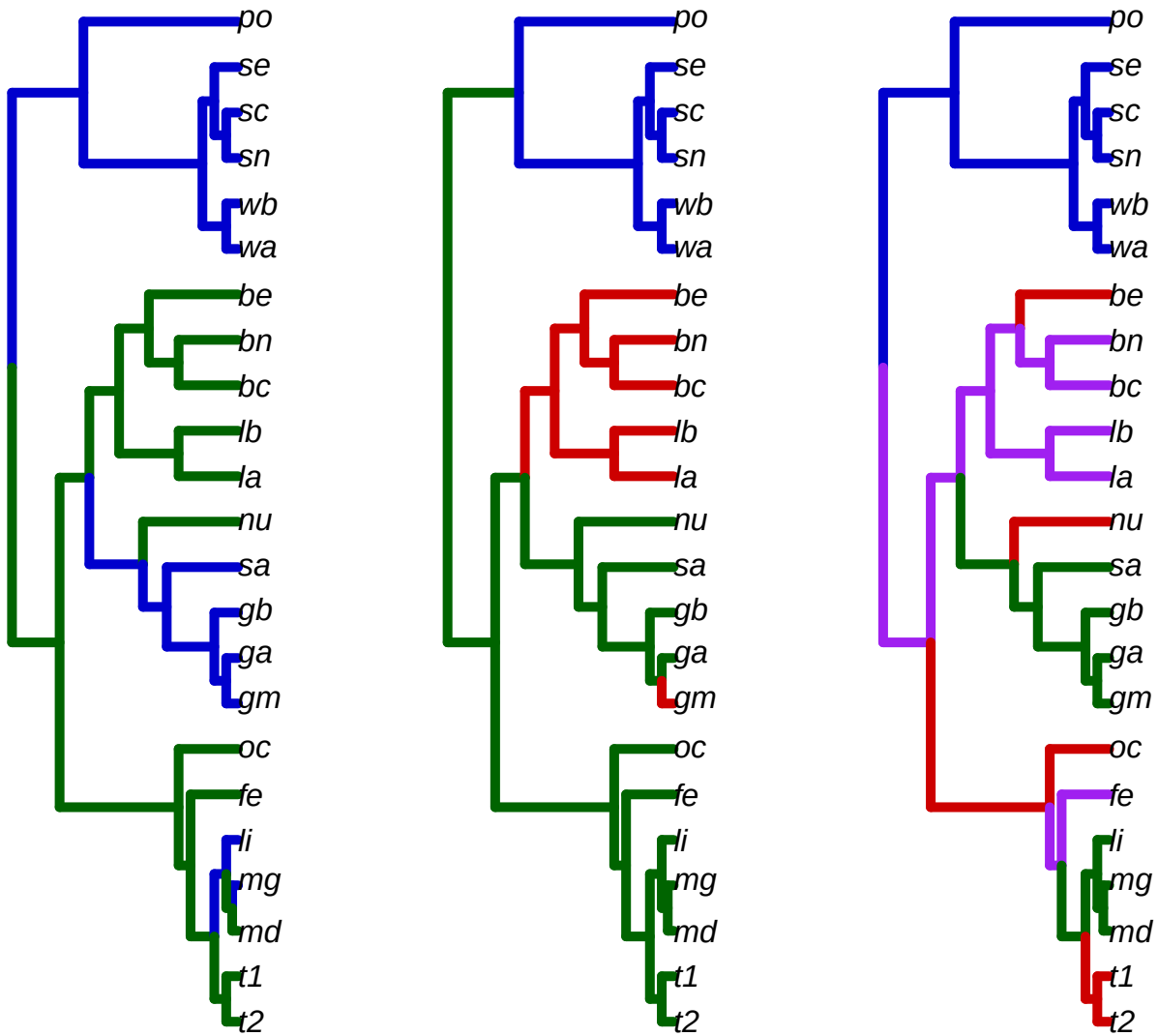
- Chris Martin
- Peter Wainwright
- Samantha Price
- Roi Holzman
- Graham Coop
- Peter Ralph
- Alan Hastings
- DoE CSGF


State
Time


State
Time


State
Time

Comparing Models



OU.3 vs OU.4: Why you should mistrust painting trees