DRAWING INFERENCES FROM MODELS and MODEL COMPARISON

Day 5

Stephan Lewandowsky and Chris Donkin

Overview for Today (Thursday)

Time	Instructor	Topic
9:00 – 10:30	Stephan Lewandowsky	Drawing conclusions from models
10:30 – 1:00	Chris Donkin	Information Criteria (AIC, BIC) How Bayes Factors work
1:00 – 2:00		lunch
2:00 – 3:00	Chris Donkin	Bayes Factors example (and maybe discussion of priors)
3:00 – dinner	All	meet mentors and discuss projects
7:00		Dinner

Outline

- Drawing inferences from models
- Comparing models on fit
- Model complexity
- Information criteria
- How Bayes Factors work
- Bayes Factors for cognitive models

DRAWING CONCLUSIONS FROM MODELING

DRAWING CONCLUSIONS FROM MODELING

DRAWING CONCLUSIONS FROM MODELING

Parameter interpretation

Model Comparison

Parameter Estimation

Parameter Estimation → Interpretation

- Parameters can be informative
  - does the accumulation rate in an accumulator model differ between conditions (or between individuals)?
  - do people pay more attention to some dimensions than others in categorization?
  - is the value function in prospect theory concave or convex (or linear)?

Interpretation (I.): Try and Break the Model

- Credit assignment
  - does a mechanism usefully contribute to a model?
  - does a parameter do what we think it is doing?
- Klaus exploring TBRS*
  - setting retrieval threshold to 0 did not change the predictions ($TBRS^0$ vs. TBRS*)
  - absolute activation does not matter
  - decay causes forgetting only in collusion with other processes

Interpretation (II.): Examine Individual Parameter Estimates

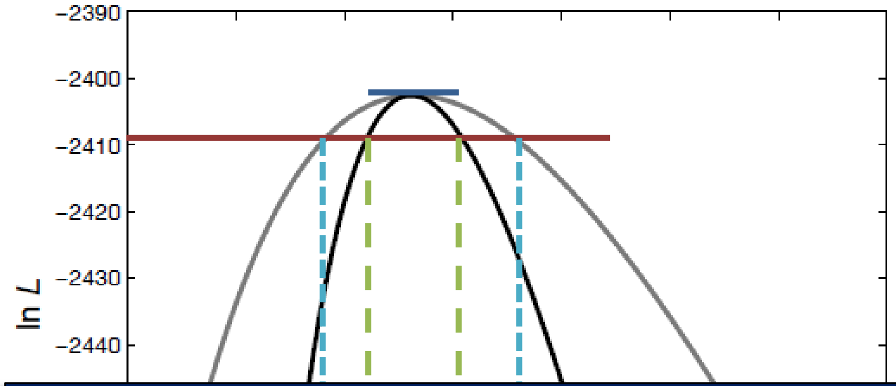
- Maximum likelihood estimation
- Fit the model to each person's data in each condition × session
- For each cell in condition × session factorial, calculate
  - mean paramete
  - standard error a

By extension: perform ANOVAs (or similar) on parameter estimates

Interpretation (III.): Standard Errors for Aggregate Data

- Bootstrapping: Resample from the data to simulate the sampling variability usually arising from random sampling from the population (Section 3.5.1 in F&L)
- Curvature of the likelihood surface

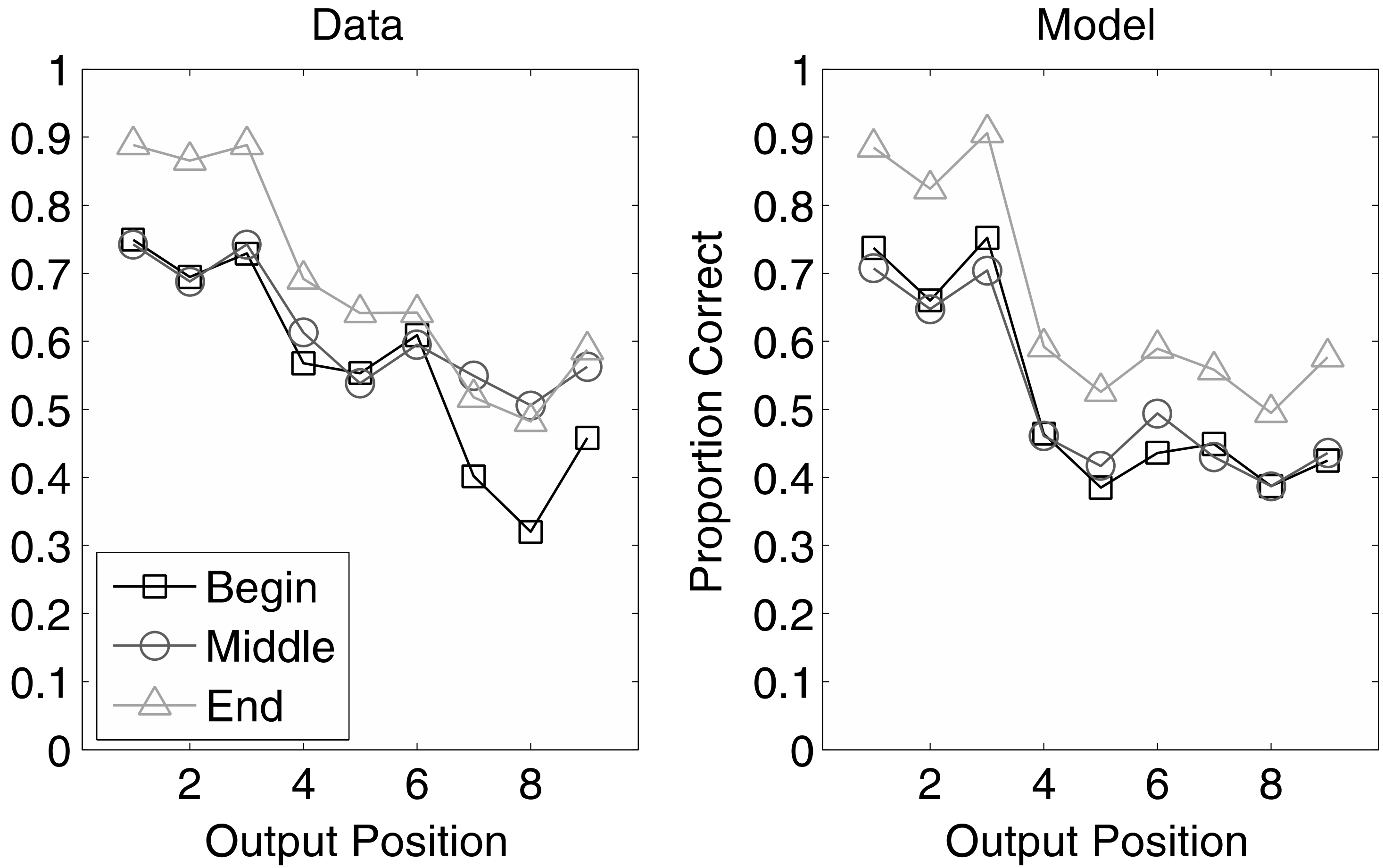
Curvature of the Likelihood Surface



- Measure how quickly *lnL* changes as each parameter changes
- More steeply peaked surface gives us higher confidence in estimate

Model Comparison

Model Fits Data!

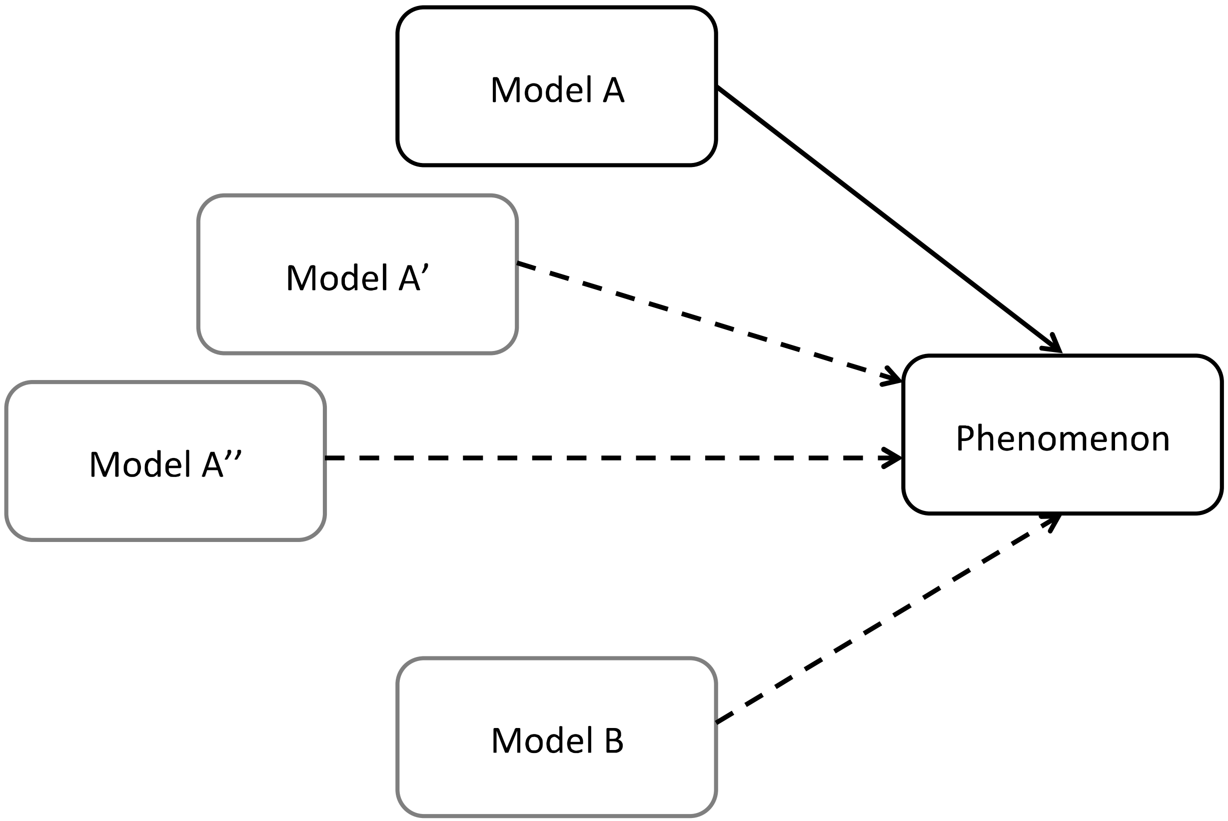


Farrell (2012), *Psychological Review*

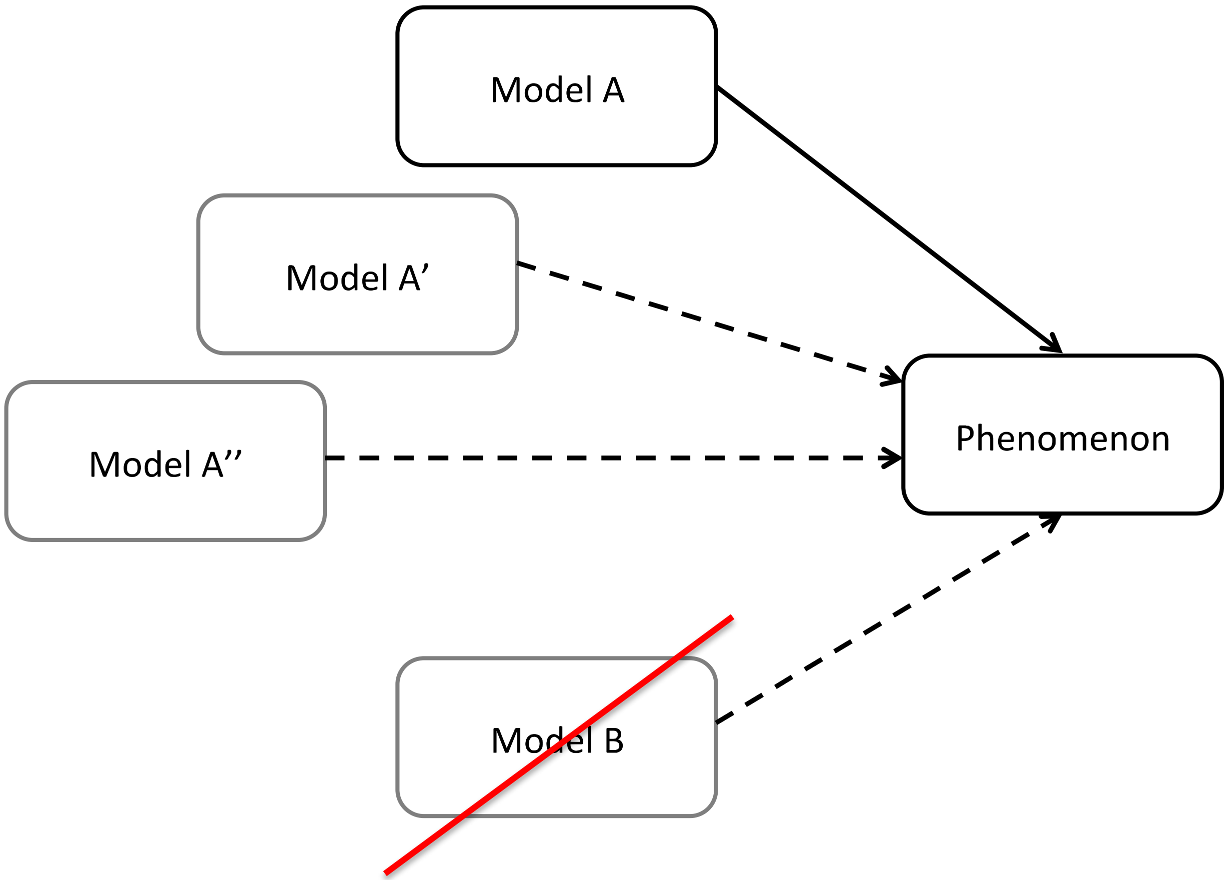
Model Fits Data!
What Does This Tell Us?

- Model “works”
- But not much else
- At best, *sufficiency*
- Many publications only fit single model

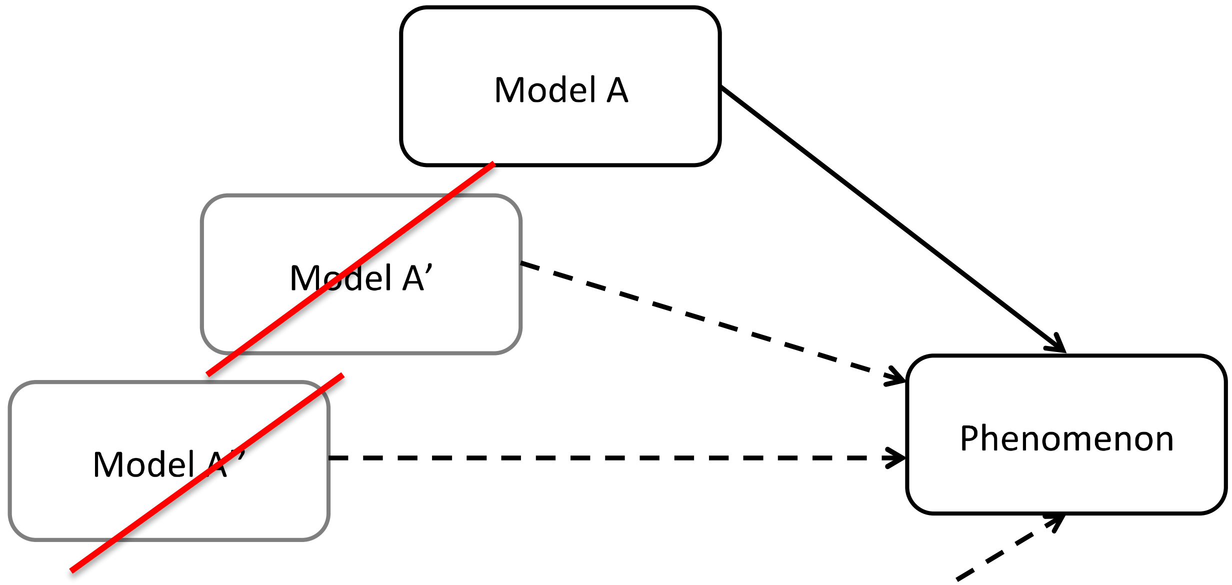
Model comparison



Model comparison



Model comparison



How do we quantitatively determine whether a model predicts the data (better than another one)?

Polynomial Law of Sensation

ABSTRACT: *A new theory proposes that sensation grows as a polynomial function of physical intensity. The theory reproduces all of the published data perfectly without error. The degree of the polynomial is independent of whether category ratings or magnitude estimations are used as the dependent variable; it is independent of stimulus range, number of categories, value of the standard, first stimulus, modulus, stimulus spacing, and all other contextual features of the experiment except the number of stimuli. Because the polynomial law always provides a superior fit to the data, it should supersede the logarithmic and power laws of sensation.*

Ri Law University

Judgments of the sensory magnitude of stimuli varying in one dimension of physical intensity have often been fit by functions of the form:

$$R = a\,(\phi - \phi_0)^k + b, \qquad (1)$$

where R is the subjective estimate of magnitude, ϕ is the physical intensity, and a, ϕ_0, k, and b are constants. One attractive feature of Equation 1 is that it can be used to fit a large variety of curves and will give a reasonable approximation to almost any set of data that might reasonably be obtained in psychophysical experiments.

Consequently, during my stay in the United States, I have fit a new theory to all of the psychophysical data published to date. The new theory fits all of these data perfectly without error. Furthermore, the new theory provides an index of the data that is independent of the experimental details.

I have therefore concluded that sensation grows not as the log, not as a power, but instead as a polynomial function of stimulus intensity:

$$R = a_0 + a_1\phi + a_2\,\phi^2 + a_3\phi^3 + \ldots + a_k\phi^k, \quad (2)$$

where R is the sensation, ϕ is the physical value, $a_0, a_1, a_2, a_3, \ldots a_k$ are constants, and k is the degree of the polynomial. Although there are many constants to estimate, it is proposed that the degree is the only one of scientific importance, hence the only one that need be reported.

Nihm Has Had Major Impact

American Psychologist
Volume 32, Issue 9, September 1977, Page 782

The Polynomial Law

Douglas G. Detterman[a] and Stephen K. Reed[b]
[a]Case Western Reserve University, Cleveland, OH, US
[b]Case Western Reserve University, Cleveland, OH, US

We were very impressed with Professor Sue Doe Nihm's (November 1976) polynomial law of sensation, which states that the degree of the polynomial is always one less than the number of stimuli. However, a distinguished visitor to our university, Professor Hoff Witt of the Frohliche Hochschule, has found that the law applies not only to psychophysical data but to psychologi
Nihm's law, we h
as the Nihm-Witt
course, that psyc
of all psychologic
work, has convin

- Is Nihm over-hyped?
- If so, why?
- Don't we want models that handle data?

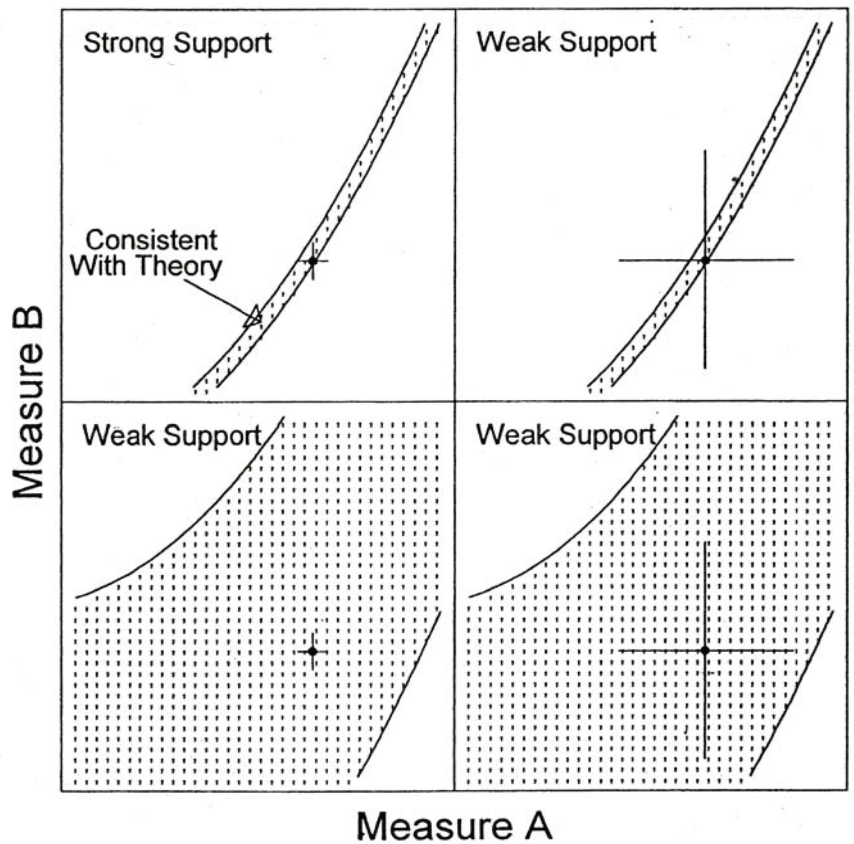
Nihm in R

MODEL COMPLEXITY

Nihm Revisited

- A powerful theory is not necessarily a good thing
- Powerful = flexible:
  - can fit any data
  - unfalsifiable
- Some tensions
  - parameter estimation: striving for a good fit
  - we want to fit signal
  - we don’t want to fit noise
  - we don’t want too good a fit
  - must account for flexibility in models

A Reminder



Roberts & Pashler (2000)

Where are we going with this?

- Knowing a model accounts for some data tells us something
- Knowing a model accounts for data better (or worse) than other models tells us more
- But we need to take into account the flexibility/complexity of models when comparing them—pure fit isn't everything

MODEL COMPLEXITY (MODEL FLEXIBILITY)

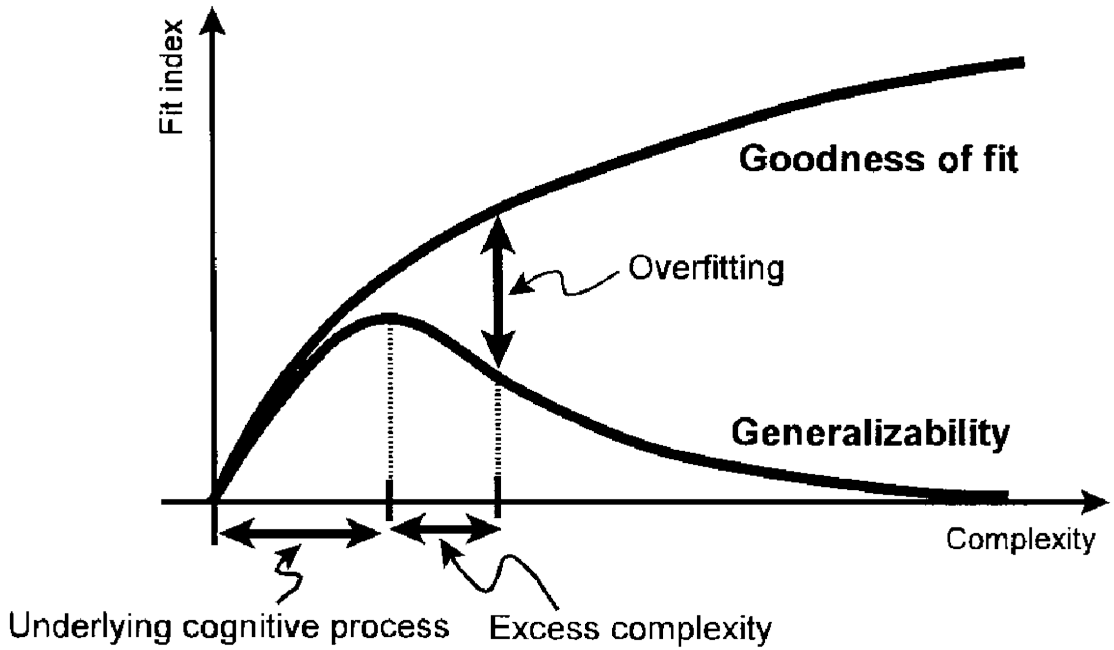
What Factors Determine Complexity?
(Myung & Pitt, 1997; Li & Lewandowsky, 1996)

1. Number of parameters
2. Expansion of parameter space
   - bounding parameters gives less complex model
3. Functional form of the model
   - how “wriggly” is the model?
   - do the predictions change much as the parameters change?
   - (average) curvature of the likelihood surface

Why is Complexity Potentially Bad?

- Complex models fit noise as well as “true” underlying data-generating process
- Complex models will always fit data well …
- ... so well, in fact, that they may be be *overfitting*

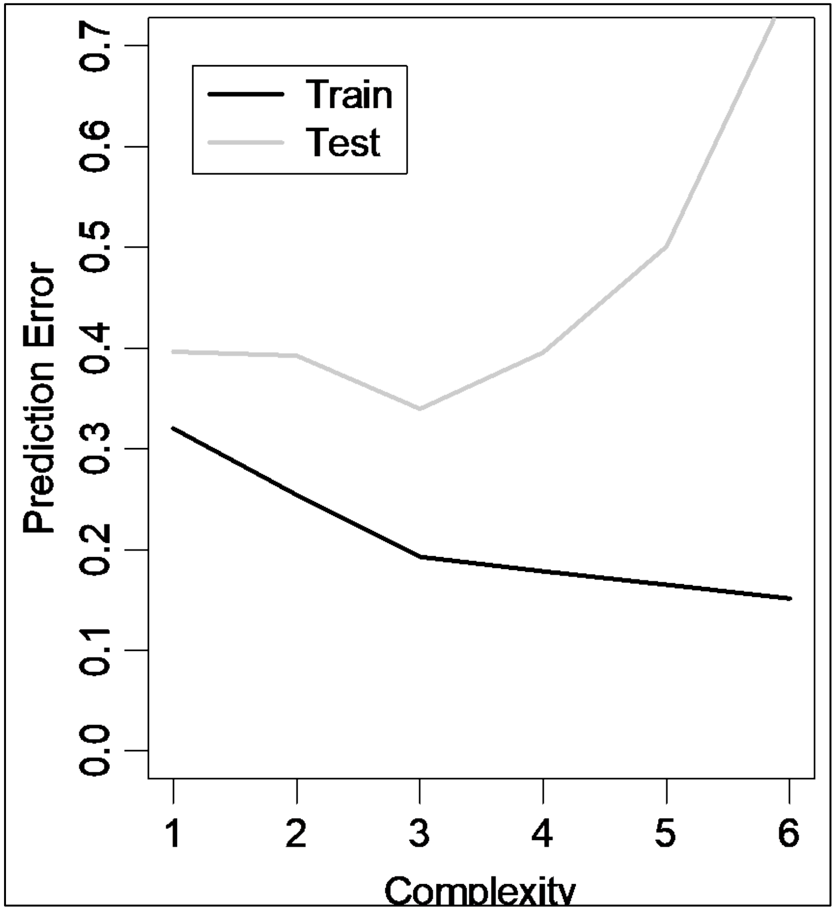
Fitting vs. Overfitting



g & Zhang (2002)

- Generalization is key
- "Cross-validation": apply model to data it has not seen before

Cross-Validation: Predicting Data That Have not Been Fit



Cross-Validation in R

k-fold Cross-Validation

- Split data into *k* equally sized segments (*s*)
- For each segment *s*
  - fit the model to all other segments
  - examine the (mis)fit for the target segment *s*
- One reasonable *k* is 10
- *k* = *N* is called leave-one-out cross validation

Asymptotically equivalent to AIC (Stone, 1977)

Over to Chris …

Model comparison needs to take complexity into account

- Likelihood ratio test
- **Information Criteria**
  - Akaike’s Information Criterion
  - Bayesian Information Criterion
- Cross-validation
- Minimum description length
  - Normalised maximum likelihood
- Landscaping
- **Bayes Factors**

- Complexity means that we can’t just use log-likelihood (deviance) to compare models
- We need to account for differences between the models in complexity
  - Number of parameters
  - Functional form

Exercise

- Data coming from a polynomial
- Your job: find the best model
- Polynomial, or any other model
- Test data set (from known generating process)

Exercise

- Run the cross validation procedure on Steven's and Polynomial law using leave-one-out cross validation
  - Minimize mean squared error
  - You'll fit each model 10 times, each time leaving out a different data point
  - For each fit you'll get a GOF value for the training set (the 9 points you fit) and the validation set (the one point you didn't fit)
  - Take the average of the GOF for the 10 validation sets
  - If you have time, plot the validation predictions against the data

Isn’t all this a problem for mathematical modelling?

- All this flexibility is hidden in verbal theorizing and reasoning from mental models
  - Submission “Theory X predicts a particular pattern of data. In fact, we observed the reverse pattern”
  - Reviewer 3 (author of Theory X): “Well, actually, theory would be consistent with reverse pattern if we additionally assume...”
- Verbal theories predict qualitative patterns
- We can *quantify* flexibility in models, just like we quantify the fit, using mathematical/computational models