Residual Information for Absolute Model Fit

Feb 2020

Charles Driver

Max Planck Institute for Human Development

- Put bluntly...

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,
  - without a head.

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,
  - without a head.
- Put differently:

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,
  - without a head.
- Put differently:
  - Precision and specific forms of misspecification are sometimes overemphasised,

- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,
  - without a head.
- Put differently:
  - Precision and specific forms of misspecification are sometimes overemphasised,
  - other, potentially more substantial, forms of misspecification go undetected because they are orthogonal to the modelling approach.

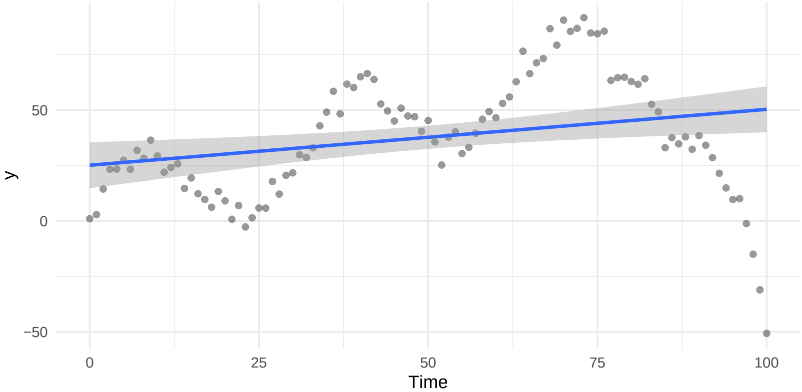
- Put bluntly...
  - Modelling in psychology is too often concerned about the Emperors' new shoelaces,
  - while he's running around naked,
  - in the cold,
  - without a head.
- Put differently:
  - Precision and specific forms of misspecification are sometimes overemphasised,
  - other, potentially more substantial, forms of misspecification go undetected because they are orthogonal to the modelling approach.
- I hope we can improve on this by approaching model fit differently.

Model fit – simple, deterministic

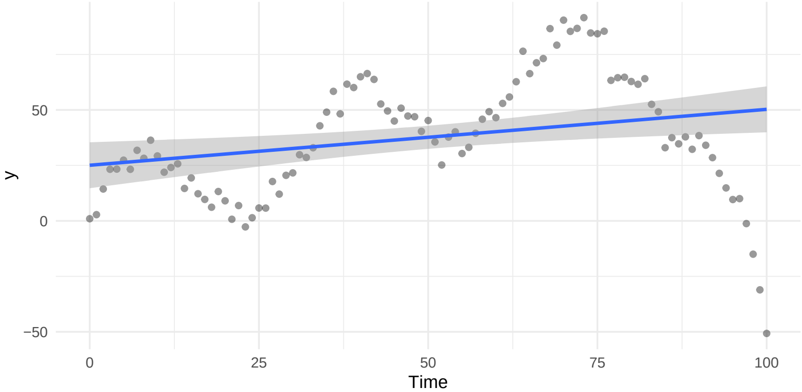
Model fit – simple, deterministic

- Straightforward with simple, deterministic systems and yes / no questions – prediction errors that can't be explained by instrumentation imply the model is inadequate.

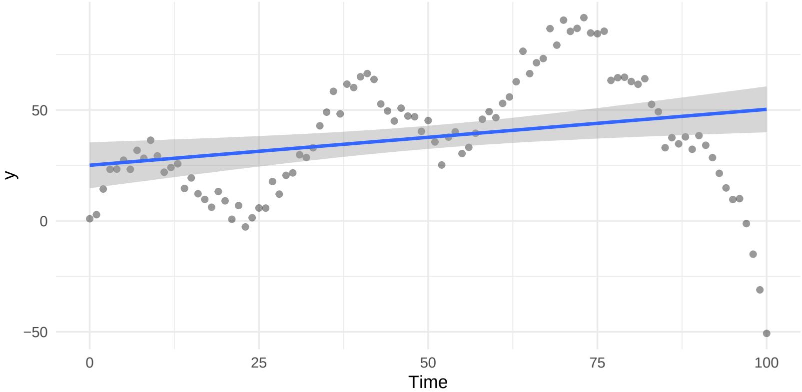
- Not so straightforward when we know in advance the model is imperfect.



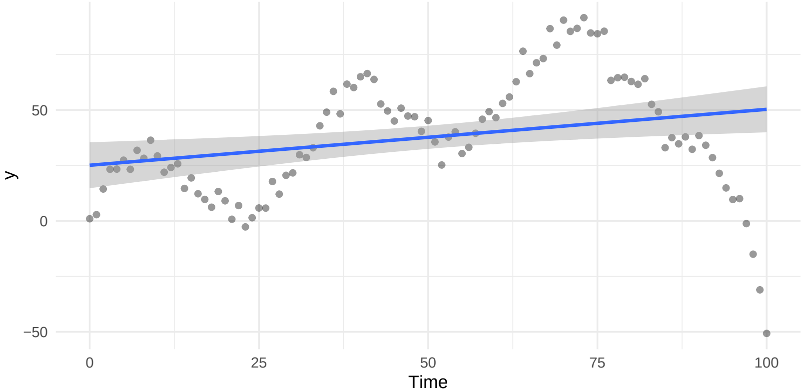
- Not so straightforward when we know in advance the model is imperfect.
- What is ‘wrong’ with this model?



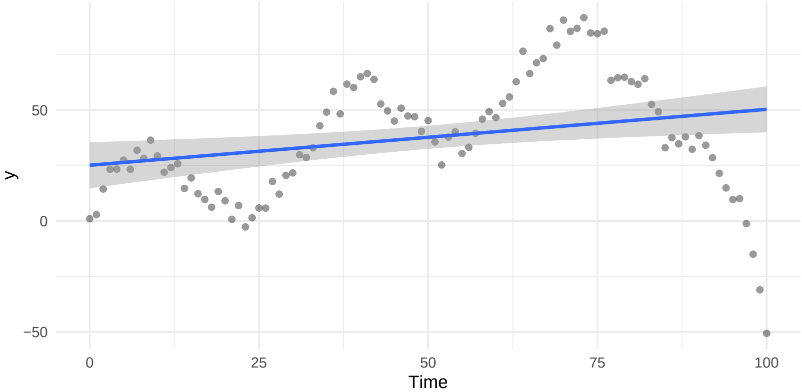
- Not so straightforward when we know in advance the model is imperfect.
- What is ‘wrong’ with this model?
  - Bad predictions for new data in some conditions.



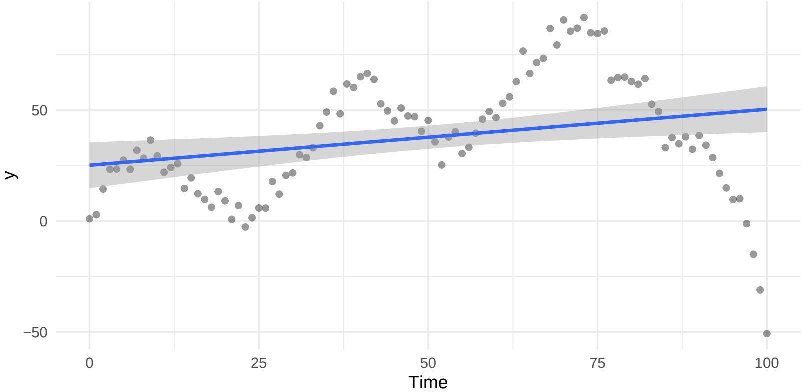
- Not so straightforward when we know in advance the model is imperfect.
- What is ‘wrong’ with this model?
  - Bad predictions for new data in some conditions.
  - Interpreted carefully, it offers extremely limited insight into the phenomenon and makes poor use of data.



- Not so straightforward when we know in advance the model is imperfect.
- What is 'wrong' with this model?
  - Bad predictions for new data in some conditions.
  - Interpreted carefully, it offers extremely limited insight into the phenomenon and makes poor use of data.
  - Interpreted loosely, it leads to crazy inferences.

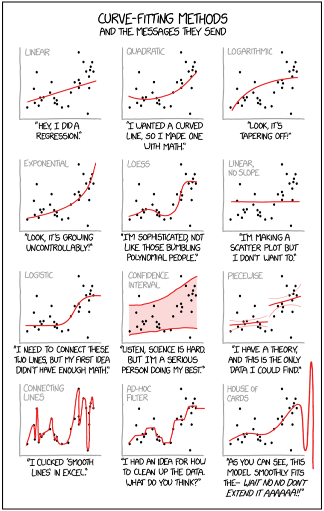


- Not so straightforward when we know in advance the model is imperfect.
- What is ‘wrong’ with this model?
    - Bad predictions for new data in some conditions.
    - Interpreted carefully, it offers extremely limited insight into the phenomenon and makes poor use of data.
    - Interpreted loosely, it leads to crazy inferences.
    - Uncertainty about the parameter tells us nothing about the ‘trueness’ of the parameter.



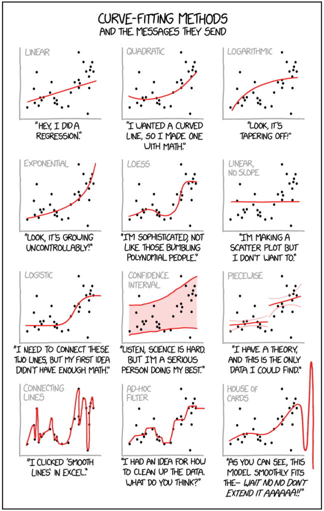
How to quantify model fit then?

- Lots of technical concerns, but at base:

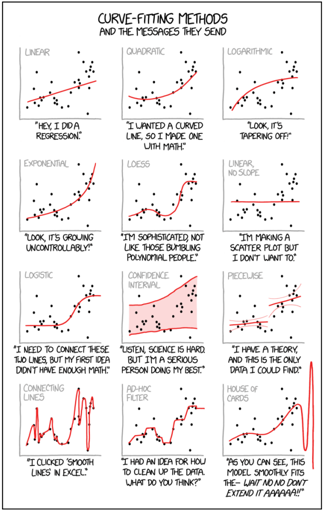


How to quantify model fit then?

- Lots of technical concerns, but at base:
  - How well does the fitted model predict new data?



- Lots of technical concerns, but at base:
  - How well does the fitted model predict new data?
  - How well do alternative models predict new data?



What should the alternative models be?

- Theory based:

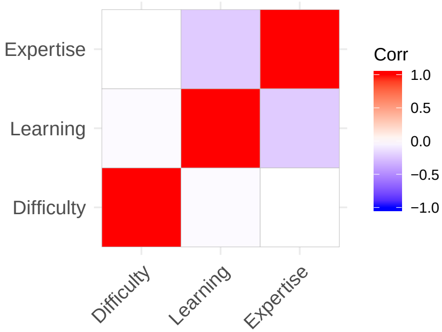
- Theory based:
    - Contrast with competing models – likelihood / evidence, information criteria, cross validation.

- Theory based:
  - Contrast with competing models – likelihood / evidence, information criteria, cross validation.
  - Tells us how much better our model is, but not whether it's actually any good.

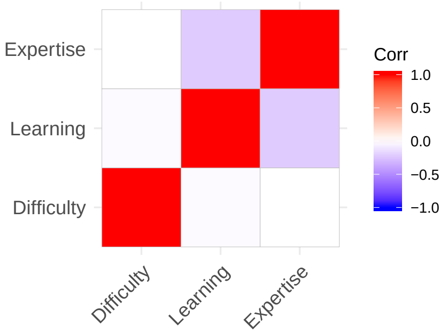
- Theory based:
  - Contrast with competing models – likelihood / evidence, information criteria, cross validation.
  - Tells us how much better our model is, but not whether it's actually any good.
- Data based:

- Theory based:
  - Contrast with competing models – likelihood / evidence, information criteria, cross validation.
  - Tells us how much better our model is, but not whether it's actually any good.
- Data based:
  - Arbitrary 'standard' model – e.g. saturated covariance matrix.

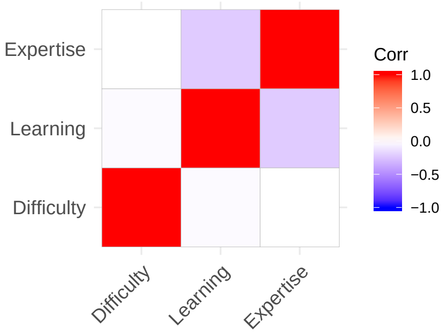
- Take some structuring of data.



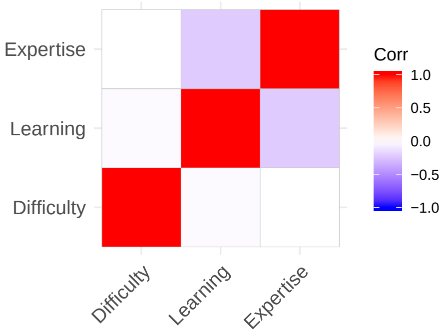
- Take some structuring of data.
- Estimate covariance matrix.



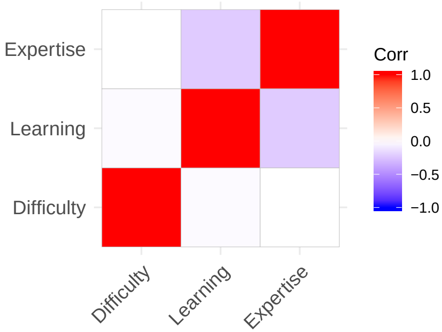
- Take some structuring of data.
- Estimate covariance matrix.
- Compare saturated model with our model – obtain estimate of distance from saturated model.

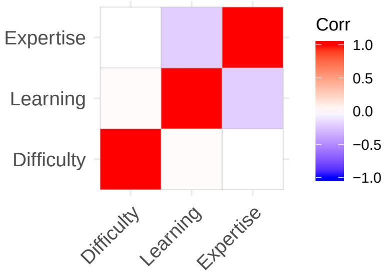


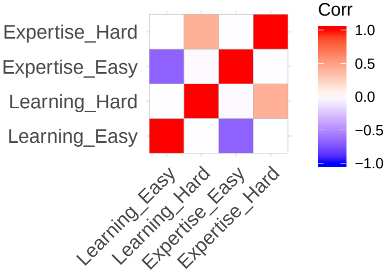
- Take some structuring of data.
- Estimate covariance matrix.
- Compare saturated model with our model – obtain estimate of distance from saturated model.
- If the true model can be represented by some covariance matrix of our data structure, we have estimate of distance from true model.

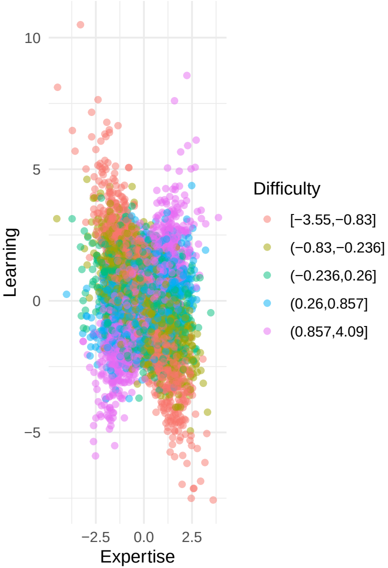


- Take some structuring of data.
- Estimate covariance matrix.
- Compare saturated model with our model – obtain estimate of distance from saturated model.
- If the true model can be represented by some covariance matrix of our data structure, we have estimate of distance from true model.
- Otherwise, we just have estimate of distance from best linear model given our arbitrary data structuring.

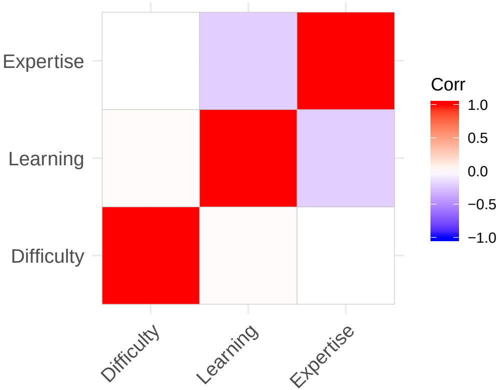



Expertise
Learning
Difficulty
Difficulty
Learning
Expertise
Corr
1.0
0.5
0.0
−0.5
−1.0

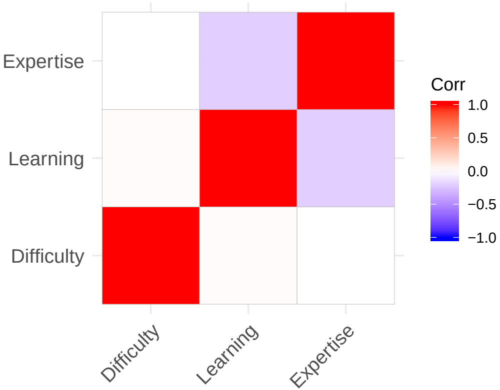

Expertise_Hard
Expertise_Easy
Learning_Hard
Learning_Easy
Learning_Easy
Learning_Hard
Expertise_Easy
Expertise_Hard
Corr
1.0
0.5
0.0
−0.5
−1.0


10
5
0
−5
Learning
−2.5
0.0
2.5
Expertise
Difficulty
[−3.55,−0.83]
(−0.83,−0.236]
(−0.236,0.26]
(0.26,0.857]
(0.857,4.09]

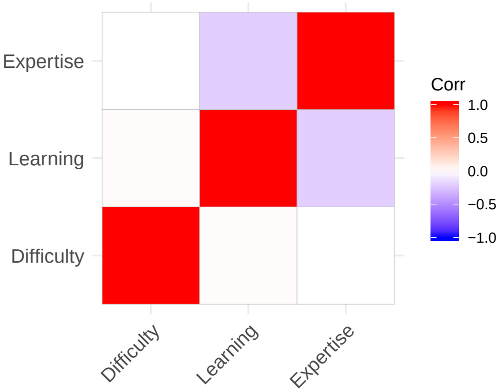
- Easy to obtain ‘perfect’ fit wrt saturated covariance.



- Easy to obtain ‘perfect’ fit wrt saturated covariance.
- Saturated covariance is very free in some regards, very limited in others, highly dependent on data structure – not a great metric.



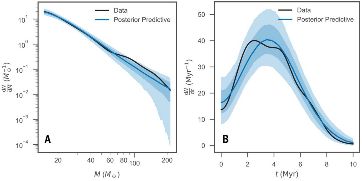
- Easy to obtain ‘perfect’ fit wrt saturated covariance.
- Saturated covariance is very free in some regards, very limited in others, highly dependent on data structure – not a great metric.
- Can we do better?

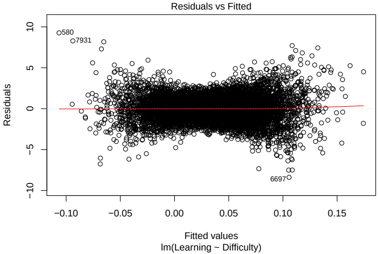


Alternative approaches:

Alternative approaches:

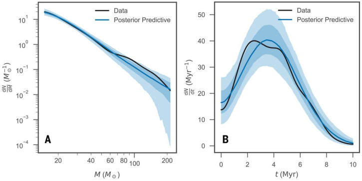
- Posterior predictive checks:

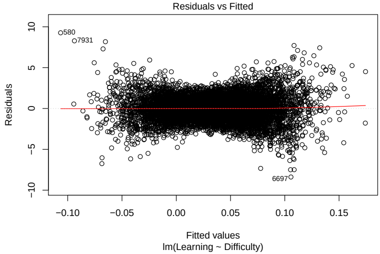




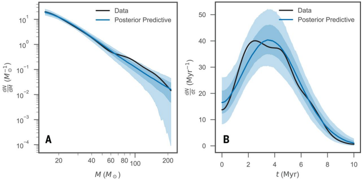
Alternative approaches:

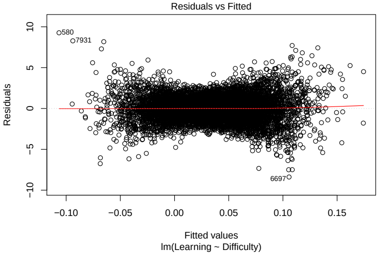
- Posterior predictive checks:
    - Decide on quantities of interest





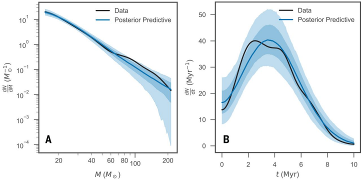
- Posterior predictive checks:
  - Decide on quantities of interest
  - Compare quantities from empirical data to randomly generated data from model.

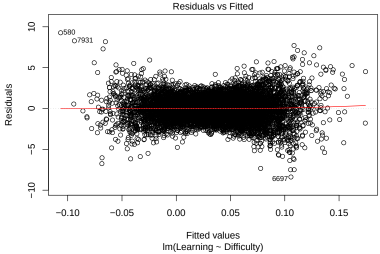




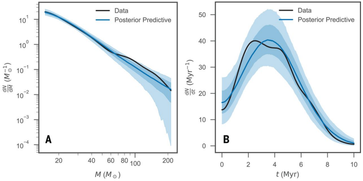
Alternative approaches:

- Posterior predictive checks:
  - Decide on quantities of interest
  - Compare quantities from empirical data to randomly generated data from model.
  - Useful but not very general – requires re-thinking for each model.

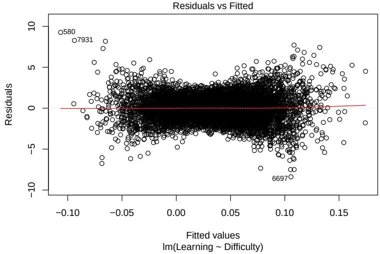




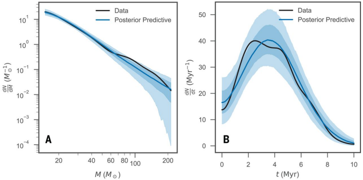
- Posterior predictive checks:
  - Decide on quantities of interest
  - Compare quantities from empirical data to randomly generated data from model.
  - Useful but not very general – requires re-thinking for each model.



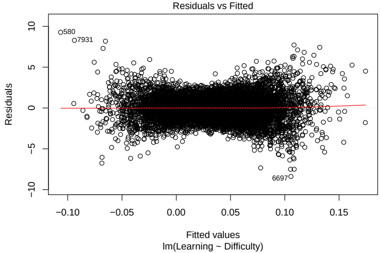
- Residual checks:



- Posterior predictive checks:
  - Decide on quantities of interest
  - Compare quantities from empirical data to randomly generated data from model.
  - Useful but not very general – requires re-thinking for each model.

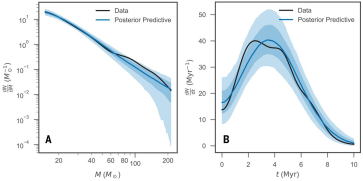


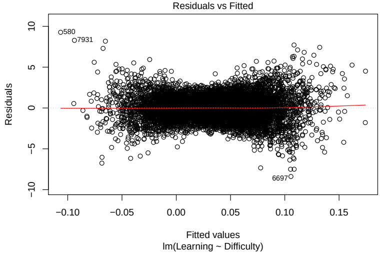
- Residual checks:
  - Visually check residuals from fit for patterns, some specific tests.



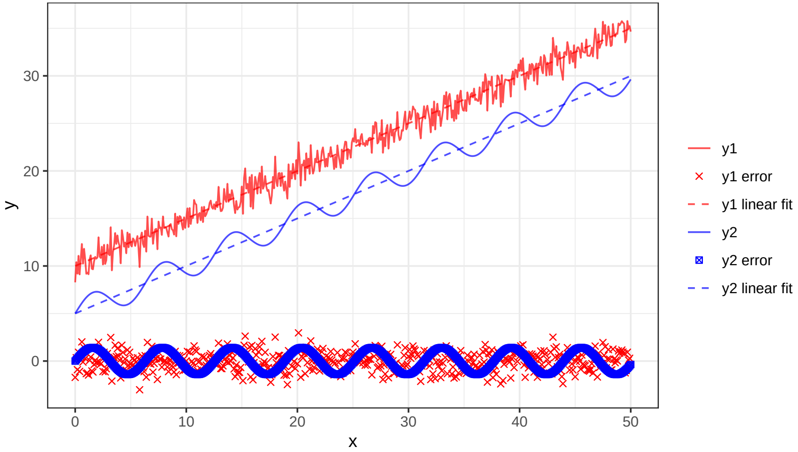
- Posterior predictive checks:
  - Decide on quantities of interest
  - Compare quantities from empirical data to randomly generated data from model.
  - Useful but not very general – requires re-thinking for each model.

- Residual checks:
  - Visually check residuals from fit for patterns, some specific tests.
  - Useful, good guide to improving model, but: non-general, too much of an art, typically univariate.

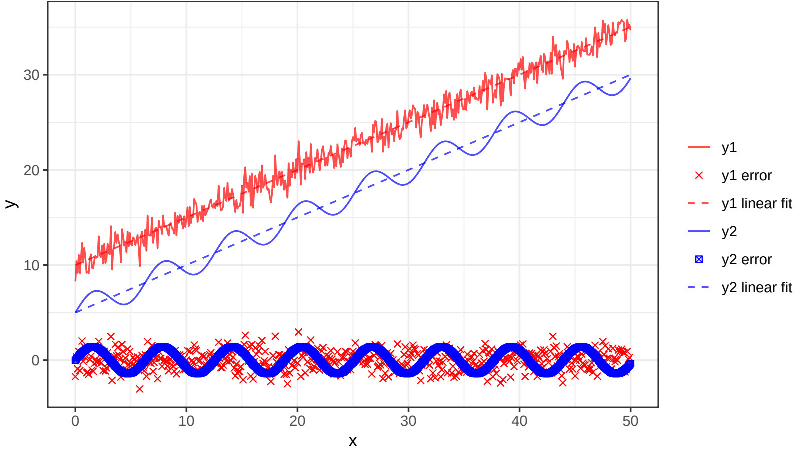




- If we knew and fit the true / best model, residuals will be random – contain no information wrt our data.



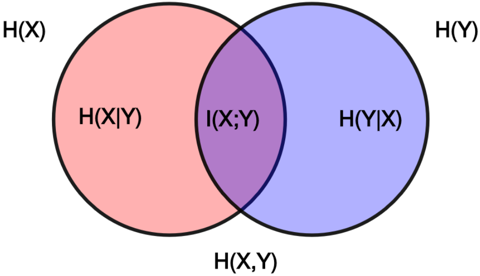
- If we knew and fit the true / best model, residuals will be random – contain no information wrt our data.
- Therefore, information that our residuals do contain, can be used to quantify distance from best possible model.



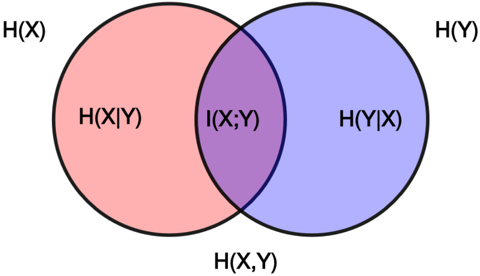
- Would be easy if we knew the true model!

- Would be easy if we knew the true model!
- Need some general function to detect and quantify structure in data.

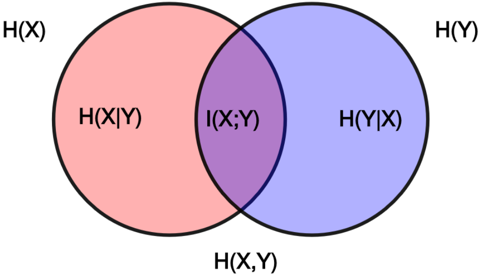
- Shannon differential entropy, H, is a scale dependent measure of information content of a variable – negative expectation of log probability.



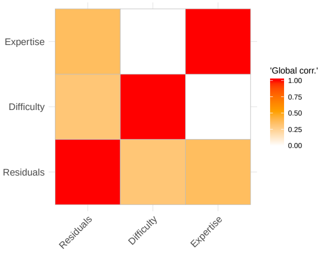
- Shannon differential entropy, H, is a scale dependent measure of information content of a variable – negative expectation of log probability.
- Mutual information, $I(X, Y) = H(X) + H(Y) - H(X, Y)$, quantifies information shared by X and Y.

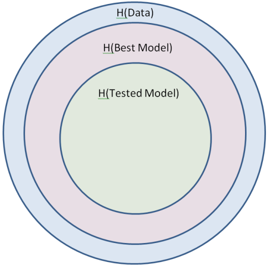


- Shannon differential entropy, H, is a scale dependent measure of information content of a variable – negative expectation of log probability.
- Mutual information, $I(X,Y) = H(X) + H(Y) - H(X,Y)$, quantifies information shared by X and Y.
- Conditional entropy, $H(X|Y) = H(X) - I(X,Y)$, quantifies information unique to X with respect to Y.

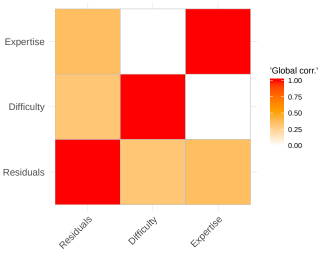


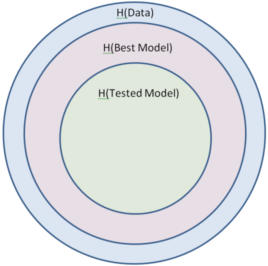
- Mutual information (nonlinear covariance) between residuals, expectations, and any additional covariates.



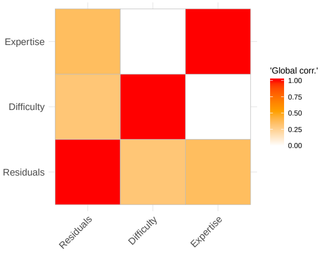


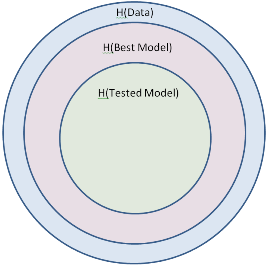
- Mutual information (nonlinear covariance) between residuals, expectations, and any additional covariates.
- Output:



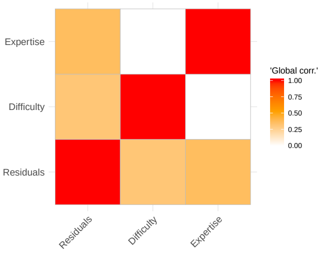


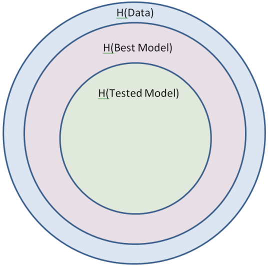
- Mutual information (nonlinear covariance) between residuals, expectations, and any additional covariates.
- Output:
  - Proportion of residual information able to be predicted from the data – model ‘wrongness’.





- Mutual information (nonlinear covariance) between residuals, expectations, and any additional covariates.
- Output:
  - Proportion of residual information able to be predicted from the data – model ‘wrongness’.
  - Where there is shared information between data and residuals – model can be improved.





- Accurate entropy estimates of some ‘ideal’ model, based on the original data, are not an easy problem, but:

- Accurate entropy estimates of some ‘ideal’ model, based on the original data, are not an easy problem, but:
  - We can base estimates from the residuals instead, leveraging the initial model.

- Accurate entropy estimates of some ‘ideal’ model, based on the original data, are not an easy problem, but:
  - We can base estimates from the residuals instead, leveraging the initial model.
  - Relative differences are sufficient to provide some guide as to goodness of fit.

- Accurate entropy estimates of some ‘ideal’ model, based on the original data, are not an easy problem, but:
  - We can base estimates from the residuals instead, leveraging the initial model.
  - Relative differences are sufficient to provide some guide as to goodness of fit.
- Overfitting can be handled either by overfitting baseline estimates in a similar fashion, or cross validation.

- Residual fit measure gets worse as informative covariates are included – because best possible model gets better.

- Residual fit measure gets worse as informative covariates are included – because best possible model gets better.
- Doesn't distinguish mean / variance or directionality – further diagnostics probably helpful.

- Residual fit measure gets worse as informative covariates are included – because best possible model gets better.
- Doesn't distinguish mean / variance or directionality – further diagnostics probably helpful.
- Current plan: R software to provide fit / diagnostics for arbitrary statistical model.

- Model fit is important if our parameters are to mean anything.

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...
- But are not great at telling us how close we are to the best model.

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...
- But are not great at telling us how close we are to the best model.
- Simplifying and quantifying residual analysis, for multivariate systems, hopefully:

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...
- But are not great at telling us how close we are to the best model.
- Simplifying and quantifying residual analysis, for multivariate systems, hopefully:
    - Makes model improvement easier.

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...
- But are not great at telling us how close we are to the best model.
- Simplifying and quantifying residual analysis, for multivariate systems, hopefully:
    - Makes model improvement easier.
    - Makes it more obvious when results are from poor models.

- Model fit is important if our parameters are to mean anything.
- Current metrics are very good in some regards...
- But are not great at telling us how close we are to the best model.
- Simplifying and quantifying residual analysis, for multivariate systems, hopefully:
    - Makes model improvement easier.
    - Makes it more obvious when results are from poor models.
    - Helps bridge some of the performance divide between prediction oriented and explanation oriented approaches.