AIC and BIC for Simple Models of Severely Complex Data

UNIVERSITY OF NOTRE DAME

Ian Campbell

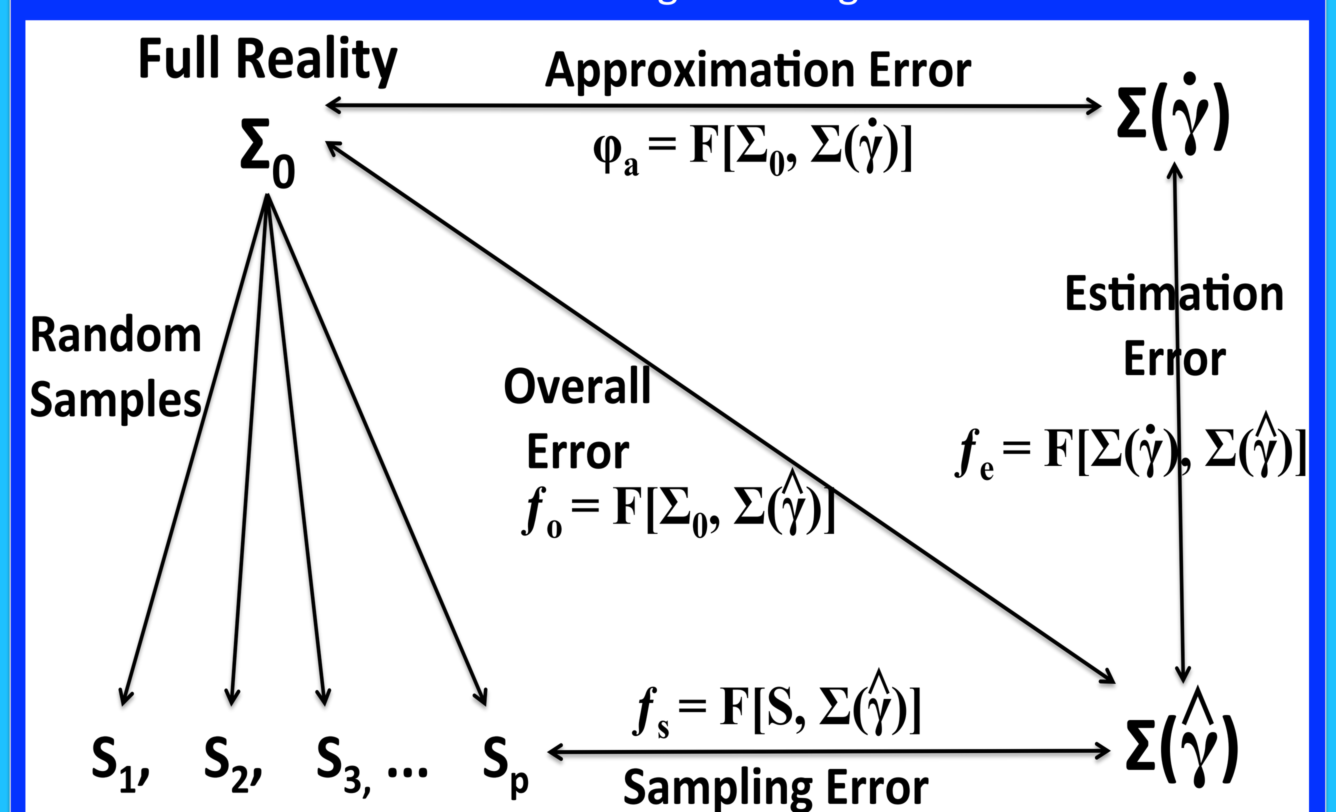
Modeling Error for Covariance Structures

For comparing non-nested models, psychologists often use the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

When modeling covariance structure, a discrepancy function measures the distance between two covariance matrices, with a common choice being the Maximum Likelihood discrepancy:

$$F_{ML}(A,B) = \log|B| - \log|A| + tr(AB^{-1}) - p$$

A useful framework for considering modeling errors:[1-4]



- Σ_0 = true population covariance matrix
- S = covariance matrix of a random sample from the population
- $\Sigma(\dot{\gamma})$ = model implied covariance matrix (MICM) calculated from fitting the model to Σ_0
- $\Sigma(\hat{\gamma})$ = MICM estimated from fitting the model to S

The Different Types of Modeling Errors

- φ_a = discrepancy between the best possible MICM and Σ_0
- f_o = discrepancy between the MICM fit to S and full reality
- f_e = discrepancy between the MICM fit to S and the MICM fit to Σ_0
- f_s = discrepancy between the MICM fit to S and S

The Complexity of Behavioral Data

Most previous literature comparing AIC and BIC's model selection performance has included both the actual data-generating process (DGP) and models more and/or less complex than the true DGP.[5,6]

This favors BIC,[6] but does not reflect common psychological settings, where we expect human behavior to be the result of many different process interacting in a complicated fashion.

Research Question

How well do AIC and BIC perform at selecting their respecitve target models when the true data generating process is vastly more complex than any of the candidate models under consideration?

Simulation Fitting 5 Candidate Models to Data from a Complex Process

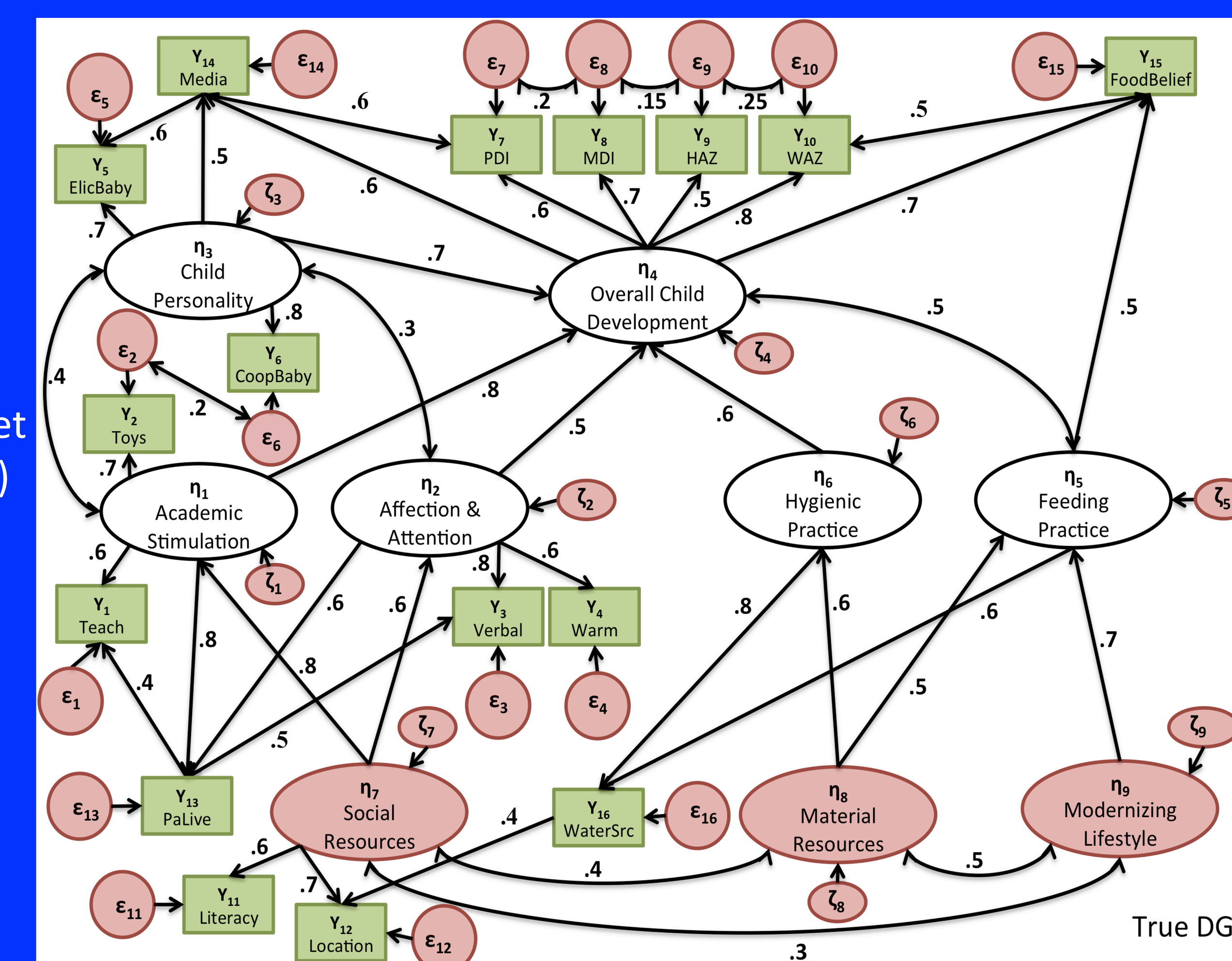
Data was simulated from the displayed complex path diagram.[7]

- Exogenous variables are in red
- Manifest variables are in green
- Endogenous factors are in white

Five candidate models were fit to a subset of manifest vars (y_1, y_3, y_8, y_9, y_{10}, y_{11}, y_{14})

1. A 2-factor mediation model (p = 17)
2. A correlated 2-factor model (p = 17)
3. A bi-factor model (p = 24)
4. A one factor model (p = 15)
5. A 2-factor model with cross-loadings (p = 23)

All of these candidate models are vastly less complex than the true DGP



Results below show each model's Error of Approximation (φ_a) and Expected BIC (EBIC) and AIC (EAIC) at each sample size, as well as the selection rate from 1,000 samples at each N.

Candidate Model	φ_a	N = 50 EBIC | EAIC BIC & AIC Rate	N = 100 EBIC | EAIC BIC & AIC Rate	N =400 EBIC | EAIC BIC & AIC Rate	N = 800 EBIC | EAIC BIC & AIC Rate	N = 1200 EBIC | EAIC BIC & AIC Rate
1	.1512	84.92 | 52.41 .22 | .19	104.26 | 59.97 .19 | .11	173.20 | 105.34 .10 | .03	245.48 | 165.84 .02 | .01	312.87 | 226.34 .02 | .01
2	.0842	81.63 | 49.13 .42 | .40	97.63 | 53.34 .49 | .39	146.47 | 78.61 .45 | .24	191.95 | 112.31 .35 | .19	232.54 | 146.01 .29 | .19
3	.1248	104.01 | 58.12 .00 | .05	126.88 | 64.36 .00 | .10	197.61 | 101.81 .10 | .28	264.18 | 151.74 .20 | .33	323.84 | 201.68 .27 | .37
4	.5508	98.67 | 69.99 .36 | .24	136.61 | 97.53 .30 | .19	322.64 | 262.76 .13 | .06	553.35 | 483.08 .14 | .07	779.74 | 703.39 .09 | .04
5	.0284	96.37 | 52.39 .01 | .13	113.73 | 53.81 .02 | .21	154.13 | 62.33 .22 | .39	181.44 | 73.69 .29 | .40	202.12 | 85.05 .33 | .39

Discussion

Answering the Research Question:

1. At small sample sizes, BIC fails to target the model with the lowest error of approximation.
2. Eventually, when N is large enough, BIC correctly targets the model with the lowest error of approximation.
3. AIC incorporates sample size to target the model with the lowest overall error.
   - This target model will change as sample size increases and estimation error decreases
   - Overall error combines approximation and estimation error[2,4]
4. Even when AIC and BIC are targeting their correct models, large model selection uncertainty exists.
   - The selection rate for the correct model never exceeded 50%
   - Very large samples are needed to ensure AIC and BIC not only target their correct models but also reliably select then in any given sample

The Behavior of AIC and BIC as N Increases

In a finite sample AIC targets the model with lowest overall error,[6] while BIC consistently targets the model with lowest approximation error.[8]

However, "sufficiently large" N is required for BIC's statistical consistency.[4] In sample sizes seen in psychology, Expected BIC (EBIC) does not always target the model with the lowest error of approximation.

As N increases, the difference between overall error and approximation error decreases because estimation error shrinks.[2] Thus, in this simulation AIC began to target the model with the lowest approximation error as this became the model with the lowest overall error, too.

In this simulation, AIC began to target the model with the lowest approximation error more quickly than BIC.

The Significance of the Problem

Model selection is an important part of data analysis. Previous simulations have recommended BIC over AIC due to BIC's consistency.[5,8] However, the N required for BIC to reach its consistency can be very large, especially when the true DGP is much more complex than any considered model.

AIC may be a better choice due to its ability to incorporate sample size and how quickly estimation error shrinks with increasing N relative to the large sample sizes needed for BIC to reach consistency.

Literature Cited

1. Linhart, H., & Zucchini, W. (1986). Model selection. New York, NY: Wiley.
2. Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512–519.
3. Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. Multivariate Behavioral Research, 48, 28-56.
4. Lubke, G. H., Campbell, I., McArtor, D., Miller, P., Luningham, J., van den Berg, S. M. (2016). Assessing model selection uncertainty using a bootstrap approach: An update. Structural Equation Modeling 1-16.
5. Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007) Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14, 4.
6. Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research, 33, 2.
7. Path diagram loosely based on a model from United Nations (1995). Human Development Report 1995. New York, NY: Oxford University Press.
8. Kass, R. E. & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. Journal of the American Statistical Association, 90, 431.

Acknowledgments

I would like to thank everyone at ND who has helped me increase my knowledge while in graduate school and with this project in particular.

Thank you especially to:

Gitta Lubke Dan McArtor
Patrick Miller Justin Luningham

Genetics and Statistical Learning Lab, Department of Psychology, University of Notre Dame