Ivan Jacob Agaloos Pesigan

September 2, 2023

References

Andrews: Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space **Andrews-2000**

Donald W. K. Andrews. "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space". In: *Econometrica* 68.2 (Mar. 2000), pp. 399–405. DOI: 10.1111/1468-0262.00114.

Bauer et al.: Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations **Bauer-Preacher-Gil-2006**

Daniel J. Bauer, Kristopher J. Preacher, and Karen M. Gil. "Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations". In: *Psychological Methods* 11.2 (2006), pp. 142–163. DOI: 10.1037/1082-989x.11.2.142.

Beran: The impact of the bootstrap on statistical algorithms and theory **Beran-2003**

Rudolf Beran. "The impact of the bootstrap on statistical algorithms and theory". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994972.

Abstract: Bootstrap ideas yield remarkably effective algorithms for realizing certain programs in statistics. These include the construction of (possibly simultaneous) confidences sets and tests in classical models for which exact or asymptotic distribution theory is intractable. Success of the bootstrap, in the sense of doing what is expected under a probability model for data, is not universal.

Modifications to Efron's definition of the bootstrap are needed to make the idea work for modern procedures that are not classically regular.

Boos: Introduction to the bootstrap world **Boos-2003**

Dennis D. Boos. "Introduction to the bootstrap world". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994971.

Abstract: The bootstrap has made a fundamental impact on how we carry out statistical inference in problems without analytic solutions. This fact is illustrated with examples and comments that emphasize the parametric bootstrap and hypothesis testing.

Casella: Introduction to the silver anniversary of the bootstrap **Casella-2003**

George Casella. "Introduction to the silver anniversary of the bootstrap". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994967.

G. W. Cheung et al.: Testing mediation and suppression effects of latent variables **Cheung-Lau-2007**

Gordon W. Cheung and Rebecca S. Lau. "Testing mediation and suppression effects of latent variables". In: *Organizational Research Methods* 11.2 (July 2007), pp. 296–325. DOI: 10.1177/1094428107300343.

Abstract: Because of the importance of mediation studies, researchers have been continuously searching for the best statistical test for mediation effect. The approaches that have been most commonly employed include those that use zero-order and partial correlation, hierarchical regression models, and structural equation modeling (SEM). This study extends MacKinnon and colleagues (MacKinnon, Lockwood, Hoffmann, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004, MacKinnon, Warsi, & Dwyer, 1995) works by conducting a simulation that examines the

distribution of mediation and suppression effects of latent variables with SEM, and the properties of confidence intervals developed from eight different methods. Results show that SEM provides unbiased estimates of mediation and suppression effects, and that the bias-corrected bootstrap confidence intervals perform best in testing for mediation and suppression effects. Steps to implement the recommended procedures with Amos are presented.

M. W.-L. Cheung: Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models **Cheung-2007**

Mike W.-L. Cheung. "Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models". In: *Structural Equation Modeling: A Multidisciplinary Journal* 14.2 (May 2007), pp. 227–246. DOI: 10.1080/10705510709336745.

Abstract: Mediators are variables that explain the association between an independent variable and a dependent variable. Structural equation modeling (SEM) is widely used to test models with mediating effects. This article illustrates how to construct confidence intervals (CIs) of the mediating effects for a variety of models in SEM. Specifically, mediating models with 1 mediator, 2 intermediate mediators, 2 specific mediators, and 1 mediator in 2 independent groups are illustrated. By using phantom variables (Rindskopf, 1984), a Wald CI, percentile bootstrap CI, bias-corrected bootstrap CI, and a likelihood-based CI on the mediating effect are easily constructed with some existing SEM packages, such as LISREL, Mplus, and Mx. Monte Carlo simulation studies are used to compare the coverage probabilities of these CIs. The results show that the coverage probabilities of these CIs are comparable when the mediating effect is large or when the sample size is large. However, when the mediating effect and the sample size are both small, the bootstrap CI and likelihood-based CI are preferred over the Wald CI. Extensions of this SEM approach for future research are discussed.

M. W.-L. Cheung: Comparison of methods for constructing confidence intervals of standardized indirect effects **Cheung-2009a**

Mike W.-L. Cheung. "Comparison of methods for constructing confidence intervals of standardized indirect effects". In: *Behavior Research Methods* 41.2 (May 2009), pp. 425–438. DOI: 10.3758/brm.41.2.425.

Abstract: Mediation models are often used as a means to explain the psychological mechanisms between an independent and a dependent variable in the behavioral and social sciences. A major limitation of the unstandardized indirect effect calculated from raw scores is that it cannot be interpreted as an effect-size measure. In contrast, the standardized indirect effect calculated from standardized scores can be a good candidate as a measure of effect size because it is scale invariant. In the present article, 11 methods for constructing the confidence intervals (CIs) of the standardized indirect effects were evaluated via a computer simulation. These included six Wald CIs, three bootstrap CIs, one likelihood-based CI, and the PRODCLIN CI. The results consistently showed that the percentile bootstrap, the bias-corrected bootstrap, and the likelihood-based approaches had the best coverage probability. Mplus, LISREL, and Mx syntax were included to facilitate the use of these preferred methods in applied settings. Future issues on the use of the standardized indirect effects are discussed.

M. W.-L. Cheung: Constructing approximate confidence intervals for parameters with structural equation models **Cheung-2009b**

Mike W.-L. Cheung. "Constructing approximate confidence intervals for parameters with structural equation models". In: *Structural Equation Modeling: A Multidisciplinary Journal* 16.2 (Apr. 2009), pp. 267–294. DOI: 10.1080/10705510902751291.

Abstract: Confidence intervals (CIs) for parameters are usually constructed based on the estimated standard errors. These are known as Wald CIs. This article argues that likelihood-based CIs (CIs based on likelihood ratio statistics) are often preferred to Wald CIs. It shows how the likelihood-

based CIs and the Wald CIs for many statistics and psychometric indexes can be constructed with the use of phantom variables (Rindskopf, 1984) in some of the current structural equation modeling (SEM) packages. The procedures to form CIs for the differences in correlation coefficients, squared multiple correlations, indirect effects, coefficient alphas, and reliability estimates are illustrated. A simulation study on the Pearson correlation is used to demonstrate the advantages of the likelihood-based CI over the Wald CI. Issues arising from this SEM approach and extensions of this approach are discussed.

Cribari-Neto: Asymptotic inference under heteroskedasticity of unknown form

CribariNeto-2004

Francisco Cribari-Neto. "Asymptotic inference under heteroskedasticity of unknown form". In: *Computational Statistics & Data Analysis* 45.2 (Mar. 2004), pp. 215–233. DOI: 10.1016/s0167-9473(02)00366-3.

Abstract: We focus on the finite-sample behavior of heteroskedasticity-consistent covariance matrix estimators and associated quasi-t tests. The estimator most commonly used is that proposed by Halbert White. Its finite-sample behavior under both homoskedasticity and heteroskedasticity is analyzed using Monte Carlo methods. We also consider two other consistent estimators, namely: the HC3 estimator, which is an approximation to the jackknife estimator, and the weighted bootstrap estimator. Additionally, we evaluate the finite-sample behavior of two bootstrap quasi-t tests: the test based on a single bootstrapping scheme and the test based on a double, nested bootstrapping scheme. The latter is very computer-intensive, but proves to work well in small samples. Finally, we propose a new estimator, which we call HC4; it is tailored to take into account the effect of leverage points in the design matrix on associated quasi-t tests.

Cribari-Neto et al.: A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model **CribariNeto-daSilva-2010**

Francisco Cribari-Neto and Wilton Bernardino da Silva. "A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model". In: *AStA Advances in Statistical Analysis* 95.2 (Nov. 2010), pp. 129–146. DOI: 10.1007/s10182-010-0141-2.

Abstract: The assumption that all random errors in the linear regression model share the same variance (homoskedasticity) is often violated in practice. The ordinary least squares estimator of the vector of regression parameters remains unbiased, consistent and asymptotically normal under unequal error variances. Many practitioners then choose to base their inferences on such an estimator. The usual practice is to couple it with an asymptotically valid estimation of its covariance matrix, and then carry out hypothesis tests that are valid under heteroskedasticity of unknown form. We use numerical integration methods to compute the exact null distributions of some quasi-t test statistics, and propose a new covariance matrix estimator. The numerical results favor testing inference based on the estimator we propose.

Cribari-Neto et al.: Inference under heteroskedasticity and leveraged data **CribariNeto-Souza-Vasconcellos-2007**

Francisco Cribari-Neto, Tatiene C. Souza, and Klaus L. P. Vasconcellos. "Inference under heteroskedasticity and leveraged data". In: *Communications in Statistics - Theory and Methods* 36.10 (Aug. 2007), pp. 1877–1888. DOI: 10.1080/03610920601126589.

Abstract: We evaluate the finite-sample behavior of different heteros-ke-das-ticity-consistent covariance matrix estimators, under both constant and unequal error variances. We consider the estimator proposed by Halbert White (HC0), and also its variants known as HC2, HC3, and HC4; the latter was recently proposed by Cribari-Neto (2004). We propose a new covariance matrix estimator: HC5. It is the first consistent estimator to explicitly take into account the effect that the maximal

leverage has on the associated inference. Our numerical results show that quasi-t inference based on HC5 is typically more reliable than inference based on other covariance matrix estimators.

Cribari-Neto et al.: Errata: Inference under heteroskedasticity and leveraged data, Communications in Statistics, Theory and Methods, 36, 1877–1888, 2007

CribariNeto-Souza-Vasconcellos-2008

Francisco Cribari-Neto, Tatiene C. Souza, and Klaus L. P. Vasconcellos. "Errata: Inference under heteroskedasticity and leveraged data, Communications in Statistics, Theory and Methods, 36, 1877–1888, 2007". In: *Communications in Statistics - Theory and Methods* 37.20 (Sept. 2008), pp. 3329–3330. DOI: 10.1080/03610920802109210.

Davison et al.: Recent developments in bootstrap methodology

Davison-Hinkley-Young-2003

Anthony Christopher Davison, David Victor Hinkley, and George Alastair Young. "Recent developments in bootstrap methodology". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994969.

Abstract: Ever since its introduction, the bootstrap has provided both a powerful set of solutions for practical statisticians, and a rich source of theoretical and methodological problems for statistics. In this article, some recent developments in bootstrap methodology are reviewed and discussed. After a brief introduction to the bootstrap, we consider the following topics at varying levels of detail: the use of bootstrapping for highly accurate parametric inference; theoretical properties of nonparametric bootstrapping with unequal probabilities; subsampling and the m out of n bootstrap; bootstrap failures and remedies for superefficient estimators; recent topics in significance testing; bootstrap improvements of unstable classifiers and resampling for dependent data. The treatment is telegraphic rather than exhaustive.

Efron: Second thoughts on the bootstrap **Efron-2003**

Bradley Efron. "Second thoughts on the bootstrap". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994968.

Abstract: This brief review article is appearing in the issue of Statistical Science that marks the 25th anniversary of the bootstrap. It concerns some of the theoretical and methodological aspects of the bootstrap and how they might influence future work in statistics.

Ernst et al.: Utilizing a quantile function approach to obtain exact bootstrap solutions Ernst-Hutson-2003

Michael D. Ernst and Alan D. Hutson. "Utilizing a quantile function approach to obtain exact bootstrap solutions". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994978.

Abstract: The popularity of the bootstrap is due in part to its wide applicability and the ease of implementing resampling procedures on modern computers. But careful reading of Efron (1979) will show that at its heart, the bootstrap is a "plug-in" procedure that involves calculating a functional $\theta\left(\hat{F}\right)$ from an estimate of the c.d.f. F. Resampling becomes invaluable when, as is often the case, $\theta\left(\hat{F}\right)$ cannot be calculated explicitly. We discuss some situations where working with the sample quantile function, $\hat{Q}$, rather than $\hat{F}$, can lead to explicit (exact) solutions to $\theta\left(\hat{F}\right)$.

Fritz et al.: Required sample size to detect the mediated effect Fritz-MacKinnon-2007

Matthew S. Fritz and David P. MacKinnon. "Required sample size to detect the mediated effect". In: *Psychological Science* 18.3 (Mar. 2007), pp. 233–239. DOI: 10.1111/j.1467-9280.2007.01882.x.

Abstract: Mediation models are widely used, and there are many tests of the mediated effect. One of the most common questions that researchers have when planning mediation studies is, "How many subjects do I need to achieve adequate power when testing for mediation?" This article presents the

necessary sample sizes for six of the most common and the most recommended tests of mediation for various combinations of parameters, to provide a guide for researchers when designing studies or applying for grants.

Graham et al.: How many imputations are really needed? Some practical clarifications of multiple imputation theory **Graham-Olchowski-Gilreath-2007**

John W. Graham, Allison E. Olchowski, and Tamika D. Gilreath. "How many imputations are really needed? Some practical clarifications of multiple imputation theory". In: *Prevention Science* 8.3 (June 2007), pp. 206–213. DOI: 10.1007/s11121-007-0070-9.

Abstract: Multiple imputation (MI) and full information maximum likelihood (FIML) are the two most common approaches to missing data analysis. In theory, MI and FIML are equivalent when identical models are tested using the same variables, and when m, the number of imputations performed with MI, approaches infinity. However, it is important to know how many imputations are necessary before MI and FIML are sufficiently equivalent in ways that are important to prevention scientists. MI theory suggests that small values of m, even on the order of three to five imputations, yield excellent results. Previous guidelines for sufficient m are based on relative efficiency, which involves the fraction of missing information (γ) for the parameter being estimated, and m. In the present study, we used a Monte Carlo simulation to test MI models across several scenarios in which γ and m were varied. Standard errors and p-values for the regression coefficient of interest varied as a function of m, but not at the same rate as relative efficiency. Most importantly, statistical power for small effect sizes diminished as m became smaller, and the rate of this power falloff was much greater than predicted by changes in relative efficiency. Based our findings, we recommend that researchers using MI should perform many more imputations than previously considered sufficient. These recommendations are based on γ, and take into consideration one's tolerance for a preventable power falloff (compared to FIML) due to using too few imputations.

Hall: A short prehistory of the bootstrap **Hall-2003**

Peter Hall. "A short prehistory of the bootstrap". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994970.

Abstract: The contemporary development of bootstrap methods, from the time of Efron's early articles to the present day, is well documented and widely appreciated. Likewise, the relationship of bootstrap techniques to certain early work on permutation testing, the jackknife and cross-validation is well understood. Less known, however, are the connections of the bootstrap to research on survey sampling for spatial data in the first half of the last century or to work from the 1940s to the 1970s on subsampling and resampling. In a selective way, some of these early linkages will be explored, giving emphasis to developments with which the statistics community tends to be less familiar. Particular attention will be paid to the work of P. C. Mahalanobis, whose development in the 1930s and 1940s of moving-block sampling methods for spatial data has a range of interesting features, and to contributions of other scientists who, during the next 40 years, developed half-sampling, subsampling and resampling methods.

Hatemi-J: A new method to choose optimal lag order in stable and unstable VAR models **HatemiJ-2003**

Abdulnasser Hatemi-J. "A new method to choose optimal lag order in stable and unstable VAR models". In: *Applied Economics Letters* 10.3 (Feb. 2003), pp. 135–137. DOI: 10.1080/1350485022000041050.

Abstract: A crucial aspect of empirical research based on the vector autoregressive (VAR) model is the choice of the lag order, since all inference in the VAR model is based on the chosen lag order. Here, a new information criterion is introduced for this purpose. The conducted Monte Carlo simulation experiments show that this new information criterion performs well in picking the true lag order in stable as well as unstable VAR models.

Hatemi-J: Multivariate tests for autocorrelation in the stable and unstable VAR models **HatemiJ-2004**

Abdulnasser Hatemi-J. "Multivariate tests for autocorrelation in the stable and unstable VAR models". In: *Economic Modelling* 21.4 (July 2004), pp. 661–683. DOI: 10.1016/j.econmod.2003.09.005.

Abstract: This study investigates the size and power properties of three multivariate tests for autocorrelation, namely portmanteau test, Lagrange multiplier (LM) test and Rao F-test, in the stable and unstable vector autoregressive (VAR) models, with and without autoregressive conditional heteroscedasticity (ARCH) using Monte Carlo experiments. Many combinations of parameters are used in the simulations to cover a wide range of situations in order to make the results more representative. The results of conducted simulations show that all three tests perform relatively well in stable VAR models without ARCH. In unstable VAR models the portmanteau test exhibits serious size distortions. LM and Rao tests perform well in unstable VAR models without ARCH. These results are true, irrespective of sample size or order of autocorrelation. Another clear result that the simulations show is that none of the tests have the correct size when ARCH is present irrespective of VAR models being stable or unstable and regardless of the sample size or order of autocorrelation. The portmanteau test appears to have slightly better power properties than the LM test in almost all scenarios.

Hayes: Beyond Baron and Kenny: Statistical mediation analysis in the new millennium **Hayes-2009**

Andrew F. Hayes. "Beyond Baron and Kenny: Statistical mediation analysis in the new millennium". In: *Communication Monographs* 76.4 (Dec. 2009), pp. 408–420. DOI: 10.1080/03637750903310360.

Abstract: Understanding communication processes is the goal of most communication researchers. Rarely are we satisfied merely ascertaining whether messages have an effect on some outcome of focus in a specific context. Instead, we seek to understand how such effects come to be. What kinds of

causal sequences does exposure to a message initiate? What are the causal pathways through which a message exerts its effect? And what role does communication play in the transmission of the effects of other variables over time and space? Numerous communication models attempt to describe the mechanism through which messages or other communication-related variables transmit their effects or intervene between two other variables in a causal model. The communication literature is replete with tests of such models. Over the years, methods used to test such process models have grown in sophistication. An example includes the rise of structural equation modeling (SEM), which allows investigators to examine how well a process model that links some focal variable X to some outcome Y through one or more intervening pathways fits the observed data. Yet frequently, the analytical choices communication researchers make when testing intervening variables models are out of step with advances made in the statistical methods literature. My goal here is to update the field on some of these new advances. While at it, I challenge some conventional wisdom and nudge the field toward a more modern way of thinking about the analysis of intervening variable effects.

Hayes et al.: Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation **Hayes-Cai-2007**

Andrew F. Hayes and Li Cai. "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation". In: *Behavior Research Methods* 39.4 (Nov. 2007), pp. 709–722. DOI: 10.3758/bf03192961.

Holmes: Bootstrapping phylogenetic trees: Theory and methods **Holmes-2003a**

Susan Holmes. "Bootstrapping phylogenetic trees: Theory and methods". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994979.

Abstract: This is a survey of the use of the bootstrap in the area of systematic and evolutionary biology. I present the current usage by biologists of the bootstrap as a tool both for making inferences

and for evaluating robustness, and propose a framework for thinking about these problems in terms of mathematical statistics.

Holmes: Bradley Efron: A conversation with good friends **Holmes-2003b**

Susan Holmes. "Bradley Efron: A conversation with good friends". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994981.

Abstract: Bradley Efron is Professor of Statistics and Biostatistics at Stanford University. He works on a combination of theoretical and applied topics, including empirical Bayes, survival analysis, exponential families, bootstrap and jackknife methods and confidence intervals. Most of his applied work has originated in biomedical consulting projects at the Stanford Medical School, mixed in with a few papers concerning astronomy and physics. Even his theoretical papers usually begin with specific applied problems. All three of the interviewers here have been close scientific collaborators. Brad was born in St. Paul, Minnestora, May 1938, to Esther and Miles Efron, Jewish-Russian immigrants. A Merit Scholarship, in the program's inaugural year, brought him to Caltech, graduating in Mathematics in 1960. He arrived at Stanford that Fall, eventually gaining his Ph.D., under the direction of Rupert Miller and Herb Solomon, in the Statistics Department, whose faculty also included Charles Stein, Herman Chernoff, Manny Parzen, Lincoln Moses and Ingram Olkin. Brad has lived at Stanford since 1960, with sabbaticals at Harvard, Imperial College and Berkeley. He has held several administrative positions in the university: Chair of Statistics, Associate Dean of Science, Chairman of the University Advisory Board and Chair of the Faculty Senate. He is currently Chair of the Undergraduate Program in Applied Mathematics. Honors include doctorates from Chicago, Madrid and Oslo, a MacArthur Prize Fellowship, membership in the National Academy of Sciences and the American Academy of Arts and Sciences, fellowship in the IMS and ASA, the Wilks Medal, Parzen Prize, the newly inaugurated Rao Prize and the outstanding statistician award from the Chicago ASA chapter. He has been the Rietz, Wald, and Fisher lecturers and holds the Max H. Stein endowed chair as Professor of Humanities and Sciences at Stanford. Professional

service includes Theory and Methods Editor of JASA and President of the IMS. Currently he is President-Elect of the American Statistical Association, becoming President in 2004.

Horowitz: The bootstrap in econometrics **Horowitz-2003**

Joel L. Horowitz. "The bootstrap in econometrics". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994976.

Abstract: This paper presents examples of problems in estimation and hypothesis testing that demonstrate the use and performance of the bootstrap in econometric settings. The examples are illustrated with two empirical applications. The paper concludes with a discussion of topics on which further research is needed.

Kauermann et al.: A note on the efficiency of sandwich covariance matrix estimation **Kauermann-Carroll-2001**

Göran Kauermann and Raymond J. Carroll. "A note on the efficiency of sandwich covariance matrix estimation". In: *Journal of the American Statistical Association* 96.456 (Dec. 2001), pp. 1387–1396. DOI: 10.1198/016214501753382309.

Abstract: The sandwich estimator, also known as robust covariance matrix estimator, heteroscedasticity-consistent covariance matrix estimate, or empirical covariance matrix estimator, has achieved increasing use in the econometric literature as well as with the growing popularity of generalized estimating equations. Its virtue is that it provides consistent estimates of the covariance matrix for parameter estimates even when the fitted parametric model fails to hold or is not even specified. Surprisingly though, there has been little discussion of properties of the sandwich method other than consistency. We investigate the sandwich estimator in quasi-likelihood models asymptotically, and in the linear case analytically. We show that under certain circumstances when the quasi-likelihood model is correct, the sandwich estimate is often far more variable than the usual parametric variance estimate. The increased variance is a fixed feature of the method and the price that one pays

to obtain consistency even when the parametric model fails or when there is heteroscedasticity. We show that the additional variability directly affects the coverage probability of confidence intervals constructed from sandwich variance estimates. In fact, the use of sandwich variance estimates combined with t-distribution quantiles gives confidence intervals with coverage probability falling below the nominal value. We propose an adjustment to compensate for this fact.

Lahiri: On the impact of bootstrap in survey sampling and small-area estimation **Lahiri-2003**

Partha Lahiri. "On the impact of bootstrap in survey sampling and small-area estimation". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994975.

Abstract: Development of valid bootstrap procedures has been a challenging problem for survey samplers for the last two decades. This is due to the fact that in surveys we constantly face various complex issues such as complex correlation structure induced by the survey design, weighting, imputation, small-area estimation, among others. In this paper, we critically review various bootstrap methods developed to deal with these challenging issues. We discuss two applications where the bootstrap has been found to be effective.

Lele: Impact of bootstrap on the estimating functions **Lele-2003**

Subhash R. Lele. "Impact of bootstrap on the estimating functions". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994973.

Abstract: Estimating functions form an attractive statistical methodology because of their dependence on only a few features of the underlying probabilistic structure. They also put a premium on developing methods that obtain model-robust confidence intervals. Bootstrap and jackknife ideas can be fruitfully used toward this purpose. Another important area in which bootstrap has proved its use is in the context of detecting the problem of multiple roots and searching for the consistent

root of an estimating function. In this article, I review, compare and contrast various approaches for bootstrapping estimating functions.

Long et al.: Using heteroscedasticity consistent standard errors in the linear regression model **Long-Ervin-2000**

J. Scott Long and Laurie H. Ervin. "Using heteroscedasticity consistent standard errors in the linear regression model". In: *The American Statistician* 54.3 (Aug. 2000), pp. 217–224. DOI: 10.1080/00031305.2000.10474549.

MacKinnon et al.: Distribution of the product confidence limits for the indirect effect: Program PRODCLIN **MacKinnon-Fritz-Williams-etal-2007**

David P. MacKinnon, Matthew S. Fritz, et al. "Distribution of the product confidence limits for the indirect effect: Program PRODCLIN". In: *Behavior Research Methods* 39.3 (Aug. 2007), pp. 384–389. DOI: 10.3758/bf03193007.

Abstract: This article describes a program, PRODCLIN (distribution of the PRODuct Confidence Limits for INdirect effects), written for SAS, SPSS, and R, that computes confidence limits for the product of two normal random variables. The program is important because it can be used to obtain more accurate confidence limits for the indirect effect, as demonstrated in several recent articles (MacKinnon, Lockwood, & Williams, 2004; Pituch, Whittaker, & Stapleton, 2005). Tests of the significance of and confidence limits for indirect effects based on the distribution of the product method have more accurate Type I error rates and more power than other, more commonly used tests. Values for the two paths involved in the indirect effect and their standard errors are entered in the PRODCLIN program, and distribution of the product confidence limits are computed. Several examples are used to illustrate the PRODCLIN program. The PRODCLIN programs in rich text format may be downloaded from www.psychonomic.org/archive.

MacKinnon et al.: A comparison of methods to test mediation and other intervening variable effects **MacKinnon-Lockwood-Hoffman-etal-2002**

David P. MacKinnon, Chondra M. Lockwood, Jeanne M. Hoffman, et al. "A comparison of methods to test mediation and other intervening variable effects". In: *Psychological Methods* 7.1 (2002), pp. 83–104. DOI: 10.1037/1082-989x.7.1.83.

Abstract: A Monte Carlo study compared 14 methods to test the statistical significance of the intervening variable effect. An intervening variable (mediator) transmits the effect of an independent variable to a dependent variable. The commonly used R. M. Baron and D. A. Kenny (1986) approach has low statistical power. Two methods based on the distribution of the product and 2 difference-in-coefficients methods have the most accurate Type I error rates and greatest statistical power except in 1 important case in which Type I error rates are too high. The best balance of Type I error and statistical power across all cases is the test of the joint significance of the two effects comprising the intervening variable effect.

MacKinnon et al.: Confidence limits for the indirect effect: Distribution of the product and resampling methods **MacKinnon-Lockwood-Williams-2004**

David P. MacKinnon, Chondra M. Lockwood, and Jason Williams. "Confidence limits for the indirect effect: Distribution of the product and resampling methods". In: *Multivariate Behavioral Research* 39.1 (Jan. 2004), pp. 99–128. DOI: 10.1207/s15327906mbr3901_4.

Abstract: The most commonly used method to test an indirect effect is to divide the estimate of the indirect effect by its standard error and compare the resulting z statistic with a critical value from the standard normal distribution. Confidence limits for the indirect effect are also typically based on critical values from the standard normal distribution. This article uses a simulation study to demonstrate that confidence limits are imbalanced because the distribution of the indirect effect is normal only in special cases. Two alternatives for improving the performance of confidence limits for the indirect effect are evaluated: (a) a method based on the distribution of the product of

two normal random variables, and (b) resampling methods. In Study 1, confidence limits based on the distribution of the product are more accurate than methods based on an assumed normal distribution but confidence limits are still imbalanced. Study 2 demonstrates that more accurate confidence limits are obtained using resampling methods, with the bias-corrected bootstrap the best method overall.

Peugh et al.: Missing data in educational research: A review of reporting practices and suggestions for improvement **Peugh-Enders-2004**

James L. Peugh and Craig K. Enders. "Missing data in educational research: A review of reporting practices and suggestions for improvement". In: *Review of Educational Research* 74.4 (Dec. 2004), pp. 525–556. DOI: 10.3102/00346543074004525.

Abstract: Missing data analyses have received considerable recent attention in the methodological literature, and two "modern" methods, multiple imputation and maximum likelihood estimation, are recommended. The goals of this article are to (a) provide an overview of missing-data theory, maximum likelihood estimation, and multiple imputation; (b) conduct a methodological review of missing-data reporting practices in 23 applied research journals; and (c) provide a demonstration of multiple imputation and maximum likelihood estimation using the Longitudinal Study of American Youth data. The results indicated that explicit discussions of missing data increased substantially between 1999 and 2003, but the use of maximum likelihood estimation or multiple imputation was rare; the studies relied almost exclusively on listwise and pairwise deletion.

Politis: The impact of bootstrap methods on time series analysis **Politis-2003**

Dimitris N. Politis. "The impact of bootstrap methods on time series analysis". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994977.

Abstract: Sparked by Efron's seminal paper, the decade of the 1980s was a period of active research on bootstrap methods for independent data–mainly i.i.d. or regression set-ups. By contrast, in the

1990s much research was directed towards resampling dependent data, for example, time series and random fields. Consequently, the availability of valid nonparametric inference procedures based on resampling and/or subsampling has freed practitioners from the necessity of resorting to simplifying assumptions such as normality or linearity that may be misleading.

Preacher et al.: SPSS and SAS procedures for estimating indirect effects in simple mediation models **Preacher-Hayes-2004**

Kristopher J. Preacher and Andrew F. Hayes. "SPSS and SAS procedures for estimating indirect effects in simple mediation models". In: *Behavior Research Methods, Instruments, & Computers* 36.4 (Nov. 2004), pp. 717–731. DOI: 10.3758/bf03206553.

Abstract: Researchers often conduct mediation analysis in order to indirectly assess the effect of a proposed cause on some outcome through a proposed mediator. The utility of mediation analysis stems from its ability to go beyond the merely descriptive to a more functional understanding of the relationships among variables. A necessary component of mediation is a statistically and practically significant indirect effect. Although mediation hypotheses are frequently explored in psychological research, formal significance tests of indirect effects are rarely conducted. After a brief overview of mediation, we argue the importance of directly testing the significance of indirect effects and provide SPSS and SAS macros that facilitate estimation of the indirect effect with a normal theory approach and a bootstrap approach to obtaining confidence intervals, as well as the traditional approach advocated by Baron and Kenny (1986). We hope that this discussion and the macros will enhance the frequency of formal mediation tests in the psychology literature. Electronic copies of these macros may be downloaded from the Psychonomic Society's Web archive at www.psychonomic.org/archive/.

Preacher et al.: Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models **Preacher-Hayes-2008**

Kristopher J. Preacher and Andrew F. Hayes. "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models". In: *Behavior Research Methods* 40.3 (Aug. 2008), pp. 879–891. DOI: 10.3758/brm.40.3.879.

Abstract: Hypotheses involving mediation are common in the behavioral sciences. Mediation exists when a predictor affects a dependent variable indirectly through at least one intervening variable, or mediator. Methods to assess mediation involving multiple simultaneous mediators have received little attention in the methodological literature despite a clear need. We provide an overview of simple and multiple mediation and explore three approaches that can be used to investigate indirect processes, as well as methods for contrasting two or more mediators within a single model. We present an illustrative example, assessing and contrasting potential mediators of the relationship between the helpfulness of socialization agents and job satisfaction. We also provide SAS and SPSS macros, as well as Mplus and LISREL syntax, to facilitate the use of these methods in applications.

Raghunathan et al.: A multivariate technique for multiply imputing missing values using a sequence of regression models **Raghunathan-Lepkowski-Hoewyk-etal-2001**

Trivellore E. Raghunathan et al. "A multivariate technique for multiply imputing missing values using a sequence of regression models". In: *Survey Methodology* 27.1 (2001), pp. 85–95.

Abstract: This article describes and evaluates a procedure for imputing missing values for a relatively complex data structure when the data are missing at random. The imputations are obtained by fitting a sequence of regression models and drawing values from the corresponding predictive distributions. The types of regression models used are linear, logistic, Poisson, generalized logit or a mixture of these depending on the type of variable being imputed. Two additional common features in the imputation process are incorporated: restriction to a relevant subpopulation for some variables and logical bounds or constraints for the imputed values. The restrictions involve

subsetting the sample individuals that satisfy certain criteria while fitting the regression models. The bounds involve drawing values from a truncated predictive distribution. The development of this method was partly motivated by the analysis of two data sets which are used as illustrations. The sequential regression procedure is applied to perform multiple imputation analysis for the two applied problems. The sampling properties of inferences from multiply imputed data sets created using the sequential regression method are evaluated through simulated data sets.

Schafer et al.: Missing data: Our view of the state of the art **Schafer-Graham-2002**

Joseph L. Schafer and John W. Graham. "Missing data: Our view of the state of the art". In: *Psychological Methods* 7.2 (2002), pp. 147–177. DOI: 10.1037/1082-989x.7.2.147.

Abstract: Statistical procedures for missing data have vastly improved, yet misconception and unsound practice still abound. The authors frame the missing-data problem, review methods, offer advice, and raise issues that remain unresolved. They clear up common misunderstandings regarding the missing at random (MAR) concept. They summarize the evidence against older procedures and, with few exceptions, discourage their use. They present, in both technical and practical language, 2 general approaches that come highly recommended: maximum likelihood (ML) and Bayesian multiple imputation (MI). Newer developments are discussed, including some for dealing with missing data that are not MAR. Although not yet in the mainstream, these procedures may eventually extend the ML and MI methods that currently represent the state of the art.

Serlin: Testing for robustness in Monte Carlo studies **Serlin-2000**

Ronald C. Serlin. "Testing for robustness in Monte Carlo studies". In: *Psychological Methods* 5.2 (2000), pp. 230–240. DOI: 10.1037/1082-989x.5.2.230.

Abstract: Monte Carlo studies provide the information needed to help researchers select appropriate analytical procedures under design conditions in which the underlying assumptions of the procedures are not met. In Monte Carlo studies, the 2 errors that one could commit involve (a) concluding

that a statistical procedure is robust when it is not or (b) concluding that it is not robust when it is. In previous attempts to apply standard statistical design principles to Monte Carlo studies, the less severe of these errors has been wrongly designated the Type I error. In this article, a method is presented for controlling the appropriate Type I error rate; the determination of the number of iterations required in a Monte Carlo study to achieve desired power is described; and a confidence interval for a test's true Type I error rate is derived. A robustness criterion is also proposed that is a compromise between W. G. Cochran's (1952) and J. V. Bradley's (1978) criteria.

Shao: Impact of the bootstrap on sample surveys **Shao-2003**

Jun Shao. "Impact of the bootstrap on sample surveys". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994974.

Abstract: This article discusses the impact of the bootstrap on sample surveys and introduces some of the main developments of the bootstrap methodology for sample surveys in the last twenty five years.

Shrout et al.: Mediation in experimental and nonexperimental studies: New procedures and recommendations **Shrout-Bolger-2002**

Patrick E. Shrout and Niall Bolger. "Mediation in experimental and nonexperimental studies: New procedures and recommendations". In: *Psychological Methods* 7.4 (2002), pp. 422–445. DOI: 10.1037/1082-989x.7.4.422.

Abstract: Mediation is said to occur when a causal effect of some variable X on an outcome Y is explained by some intervening variable M. The authors recommend that with small to moderate samples, bootstrap methods (B. Efron & R. Tibshirani, 1993) be used to assess mediation. Bootstrap tests are powerful because they detect that the sampling distribution of the mediated effect is skewed away from 0. They argue that R. M. Baron and D. A. Kenny's (1986) recommendation of first testing the $X \to Y$ association for statistical significance should not be a requirement when there is a priori

belief that the effect size is small or suppression is a possibility. Empirical examples and computer setups for bootstrap analyses are provided.

P. S. Soltis et al.: Applying the Bootstrap in Phylogeny Reconstruction

Soltis-Soltis-2003

Pamela S. Soltis and Douglas E. Soltis. "Applying the Bootstrap in Phylogeny Reconstruction". In: *Statistical Science* 18.2 (May 2003). DOI: 10.1214/ss/1063994980.

Abstract: With the increasing emphasis in biology on reconstruction of phylogenetic trees, questions have arisen as to how confident one should be in a given phylogenetic tree and how support for phylogenetic trees should be measured. Felsenstein suggested that bootstrapping be applied across characters of a taxon-by-character data matrix to produce replicate "bootstrap data sets," each of which is then analyzed phylogenetically, with a consensus tree constructed to summarize the results of all replicates. The proportion of trees/replicates in which a grouping is recovered is presented as a measure of support for that group. Bootstrapping has become a common feature of phylogenetic analysis. However, the interpretation of bootstrap values remains open to discussion, and phylogeneticists have used these values in multiple ways. The usefulness of phylogenetic bootstrapping is potentially limited by a number of features, such as the size of the data matrix and the underlying assumptions of the phylogeny reconstruction program. Recent studies have explored the application of bootstrapping to large data sets and the relative performance of bootstrapping and jackknifing.

Taylor et al.: Tests of the three-path mediated effect **Taylor-MacKinnon-Tein-2007**

Aaron B. Taylor, David P. MacKinnon, and Jenn-Yun Tein. "Tests of the three-path mediated effect". In: *Organizational Research Methods* 11.2 (July 2007), pp. 241–269. DOI: 10.1177/1094428107300344.

Abstract: In a three-path mediational model, two mediators intervene in a series between an independent and a dependent variable. Methods of testing for mediation in such a model are generalized from the more often used single-mediator model. Six such methods are introduced and compared

in a Monte Carlo study in terms of their Type I error, power, and coverage. Based on its results, the joint significance test is preferred when only a hypothesis test is of interest. The percentile bootstrap and bias-corrected bootstrap are preferred when a confidence interval on the mediated effect is desired, with the latter having more power but also slightly inflated Type I error in some conditions.

van Buuren et al.: Fully conditional specification in multivariate imputation

vanBuuren-Brand-GroothuisOudshoorn-etal-2006

Stef van Buuren et al. "Fully conditional specification in multivariate imputation". In: *Journal of Statistical Computation and Simulation* 76.12 (Dec. 2006), pp. 1049–1064. DOI: 10.1080/10629360600810434.

Abstract: The use of the Gibbs sampler with fully conditionally specified models, where the distribution of each variable given the other variables is the starting point, has become a popular method to create imputations in incomplete multivariate data. The theoretical weakness of this approach is that the specified conditional densities can be incompatible, and therefore the stationary distribution to which the Gibbs sampler attempts to converge may not exist. This study investigates practical consequences of this problem by means of simulation. Missing data are created under four different missing data mechanisms. Attention is given to the statistical behavior under compatible and incompatible models. The results indicate that multiple imputation produces essentially unbiased estimates with appropriate coverage in the simple cases investigated, even for the incompatible models. Of particular interest is that these results were produced using only five Gibbs iterations starting from a simple draw from observed marginal distributions. It thus appears that, despite the theoretical weaknesses, the actual performance of conditional model specification for multivariate imputation can be quite good, and therefore deserves further study.

Yuan et al.: Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data **Yuan-Bentler-2000**

Ke-Hai Yuan and Peter M. Bentler. "Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data". In: *Sociological Methodology* 30.1 (Aug. 2000), pp. 165–200. DOI: 10.1111/0081-1750.00078.

Abstract: Survey and longitudinal studies in the social and behavioral sciences generally contain missing data. Mean and covariance structure models play an important role in analyzing such data. Two promising methods for dealing with missing data are a direct maximum-likelihood and a two-stage approach based on the unstructured mean and covariance estimates obtained by the EM-algorithm. Typical assumptions under these two methods are ignorable nonresponse and normality of data. However, data sets in social and behavioral sciences are seldom normal, and experience with these procedures indicates that normal theory based methods for nonnormal data very often lead to incorrect model evaluations. By dropping the normal distribution assumption, we develop more accurate procedures for model inference. Based on the theory of generalized estimating equations, a way to obtain consistent standard errors of the two-stage estimates is given. The asymptotic efficiencies of different estimators are compared under various assumptions. We also propose a minimum chi-square approach and show that the estimator obtained by this approach is asymptotically at least as efficient as the two likelihood-based estimators for either normal or nonnormal data. The major contribution of this paper is that for each estimator, we give a test statistic whose asymptotic distribution is chisquare as long as the underlying sampling distribution enjoys finite fourth-order moments. We also give a characterization for each of the two likelihood ratio test statistics when the underlying distribution is nonnormal. Modifications to the likelihood ratio statistics are also given. Our working assumption is that the missing data mechanism is missing completely at random. Examples and Monte Carlo studies indicate that, for commonly encountered nonnormal distributions, the procedures developed in this paper are quite reliable even for samples with missing data that are missing at random.

Zeileis: Econometric computing with HC and HAC covariance matrix estimators

Zeileis-2004

Achim Zeileis. "Econometric computing with HC and HAC covariance matrix estimators". In: *Journal of Statistical Software* 11.10 (2004). DOI: 10.18637/jss.v011.i10.

Abstract: Data described by econometric models typically contains autocorrelation and/or heteroskedasticity of unknown form and for inference in such models it is essential to use covariance matrix estimators that can consistently estimate the covariance of the model parameters. Hence, suitable heteroskedasticity consistent (HC) and heteroskedasticity and autocorrelation consistent (HAC) estimators have been receiving attention in the econometric literature over the last 20 years. To apply these estimators in practice, an implementation is needed that preferably translates the conceptual properties of the underlying theoretical frameworks into computational tools. In this paper, such an implementation in the package sandwich in the R system for statistical computing is described and it is shown how the suggested functions provide reusable components that build on readily existing functionality and how they can be integrated easily into new inferential procedures or applications. The toolbox contained in sandwich is extremely flexible and comprehensive, including specific functions for the most important HC and HAC estimators from the econometric literature. Several real-world data sets are used to illustrate how the functionality can be integrated into applications.

Zeileis: Object-oriented computation of sandwich estimators **Zeileis-2006**

Achim Zeileis. "Object-oriented computation of sandwich estimators". In: *Journal of Statistical Software* 16.9 (Aug. 2006). DOI: 10.18637/jss.v016.i09.

Abstract: Sandwich covariance matrix estimators are a popular tool in applied regression modeling for performing inference that is robust to certain types of model misspecification. Suitable implementations are available in the R system for statistical computing for certain model fitting functions

only (in particular lm()), but not for other standard regression functions, such as glm(), nls(), or survreg(). Therefore, conceptual tools and their translation to computational tools in the package sandwich are discussed, enabling the computation of sandwich estimators in general parametric models. Object orientation can be achieved by providing a few extractor functions' most importantly for the empirical estimating functions' from which various types of sandwich estimators can be computed.