

## Articles (2020-2023)

Ivan Jacob Agaloos Pesigan

July 16, 2023

### References

**Cheung et al.: semlbci: An R package for forming likelihood-based confidence intervals for parameter estimates, correlations, indirect effects, and other derived parameters**

**Cheung-Pesigan-2023**

---

Shu Fai Cheung and Ivan Jacob Agaloos Pesigan. “semlbci: An R package for forming likelihood-based confidence intervals for parameter estimates, correlations, indirect effects, and other derived parameters”. In: *Structural Equation Modeling: A Multidisciplinary Journal* (May 2023), pp. 1–15. DOI: [10.1080/10705511.2023.2183860](https://doi.org/10.1080/10705511.2023.2183860).

Abstract: There are three common types of confidence interval (CI) in structural equation modeling (SEM): Wald-type CI, bootstrapping CI, and likelihood-based CI (LBCI). LBCI has the following advantages: (1) it has better coverage probabilities and Type I error rate compared to Wald-type CI when the sample size is finite; (2) it correctly tests the null hypothesis of a parameter based on likelihood ratio chi-square difference test; (3) it is less computationally intensive than bootstrapping CI; and (4) it is invariant to transformations. However, LBCI is not available in many popular SEM software packages. We developed an R package, semlbci, for forming LBCI for parameters in models fitted by lavaan, a popular open-source SEM package, such that researchers have more options in forming CIs for parameters in SEM. The package supports both unstandardized and standardized

estimates, derived parameters such as indirect effect, multisample models, and the robust LBCI proposed by Falk.

**Li et al.: Bayesian forecasting with a regime-switching zero-inflated multilevel poisson regression model: An application to adolescent alcohol use with spatial covariates**

**Li-Oravecz-Zhou-et-al-2022**

---

Yanling Li et al. “Bayesian forecasting with a regime-switching zero-inflated multilevel poisson regression model: An application to adolescent alcohol use with spatial covariates”. In: *Psychometrika* 87.2 (Jan. 2022), pp. 376–402. DOI: [10.1007/s11336-021-09831-9](https://doi.org/10.1007/s11336-021-09831-9).

**Abstract:** In this paper, we present and evaluate a novel Bayesian regime-switching zero-inflated multilevel Poisson (RS-ZIMLP) regression model for forecasting alcohol use dynamics. The model partitions individuals’ data into two phases, known as regimes, with: (1) a zero-inflation regime that is used to accommodate high instances of zeros (non-drinking) and (2) a multilevel Poisson regression regime in which variations in individuals’ log-transformed average rates of alcohol use are captured by means of an autoregressive process with exogenous predictors and a person-specific intercept. The times at which individuals are in each regime are unknown, but may be estimated from the data. We assume that the regime indicator follows a first-order Markov process as related to exogenous predictors of interest. The forecast performance of the proposed model was evaluated using a Monte Carlo simulation study and further demonstrated using substance use and spatial covariate data from the Colorado Online Twin Study (CoTwins). Results showed that the proposed model yielded better forecast performance compared to a baseline model which predicted all cases as non-drinking and a reduced ZIMLP model without the RS structure, as indicated by higher AUC (the area under the receiver operating characteristic (ROC) curve) scores, and lower mean absolute errors (MAEs) and root-mean-square errors (RMSEs). The improvements in forecast performance were even more pronounced when we limited the comparisons to participants who showed at least one instance of transition to drinking.

Daniel McNeish and David P. MacKinnon. “Intensive longitudinal mediation in Mplus”. In: *Psychological Methods* (Dec. 2022). DOI: [10.1037/met0000536](https://doi.org/10.1037/met0000536).

Abstract: Much of the existing longitudinal mediation literature focuses on panel data where relatively few repeated measures are collected over a relatively broad timespan. However, technological advances in data collection (e.g., smartphones, wearables) have led to a proliferation of short duration, densely collected longitudinal data in behavioral research. These intensive longitudinal data differ in structure and focus relative to traditionally collected panel data. As a result, existing methodological resources do not necessarily extend to nuances present in the recent influx of intensive longitudinal data and designs. In this tutorial, we first cover potential limitations of traditional longitudinal mediation models to accommodate unique characteristics of intensive longitudinal data. Then, we discuss how recently developed dynamic structural equation models (DSEMs) may be well-suited for mediation modeling with intensive longitudinal data and can overcome some of the limitations associated with traditional approaches. We describe four increasingly complex intensive longitudinal mediation models: (a) stationary models where the indirect effect is constant over time and people, (b) person-specific models where the indirect effect varies across people, (c) dynamic models where the indirect effect varies across time, and (d) cross-classified models where the indirect effect varies across both time and people. We apply each model to a running example featuring a mobile health intervention designed to improve health behavior of individuals with binge eating disorder. In each example, we provide annotated Mplus code and interpretation of the output to guide empirical researchers through mediation modeling with this increasingly popular type of longitudinal data.

**Nüst et al.: The Rockerverse: Packages and applications for containerisation with R**  
**Nust-Eddelbuettel-Bennett-et-al-2020**

---

Daniel Nüst et al. “The Rockerverse: Packages and applications for containerisation with R”. In: *The R Journal* 12.1 (2020), p. 437. DOI: [10.32614/rj-2020-007](https://doi.org/10.32614/rj-2020-007).

Abstract: The Rocker Project provides widely used Docker images for R across different application scenarios. This article surveys downstream projects that build upon the Rocker Project images and presents the current state of R packages for managing Docker images and controlling containers. These use cases cover diverse topics such as package development, reproducible research, collaborative work, cloud-based data processing, and production deployment of services. The variety of applications demonstrates the power of the Rocker Project specifically and containerisation in general. Across the diverse ways to use containers, we identified common themes: reproducible environments, scalability and efficiency, and portability across clouds. We conclude that the current growth and diversification of use cases is likely to continue its positive impact, but see the need for consolidating the Rockerverse ecosystem of packages, developing common practices for applications, and exploring alternative containerisation software.

**Pesigan et al.: betaDelta and betaSandwich: Confidence intervals for standardized regression coefficients in R**  
**Pesigan-Sun-Cheung-2023**

---

Ivan Jacob Agaloos Pesigan, Rong Wei Sun, and Shu Fai Cheung. “betaDelta and betaSandwich: Confidence intervals for standardized regression coefficients in R”. In: *Multivariate Behavioral Research* (Apr. 2023), pp. 1–4. DOI: [10.1080/00273171.2023.2201277](https://doi.org/10.1080/00273171.2023.2201277).

Abstract: The multivariate delta method was used by Yuan and Chan to estimate standard errors and confidence intervals for standardized regression coefficients. Jones and Waller extended the earlier work to situations where data are nonnormal by utilizing Browne’s asymptotic distribution-free (ADF) theory. Furthermore, Dudgeon developed standard errors and confidence intervals, employ-

ing heteroskedasticity-consistent (HC) estimators, that are robust to nonnormality with better performance in smaller sample sizes compared to Jones and Waller’s ADF technique. Despite these advancements, empirical research has been slow to adopt these methodologies. This can be a result of the dearth of user-friendly software programs to put these techniques to use. We present the betaDelta and the betaSandwich packages in the R statistical software environment in this manuscript. Both the normal-theory approach and the ADF approach put forth by Yuan and Chan and Jones and Waller are implemented by the betaDelta package. The HC approach proposed by Dudgeon is implemented by the betaSandwich package. The use of the packages is demonstrated with an empirical example. We think the packages will enable applied researchers to accurately assess the sampling variability of standardized regression coefficients.

**Wang et al.: Investigating the impact of the time interval selection on autoregressive mediation modeling: Result interpretations, effect reporting, and temporal designs**

**Wang-Zhang-2020**

---

Lijuan Wang and Qian Zhang. “Investigating the impact of the time interval selection on autoregressive mediation modeling: Result interpretations, effect reporting, and temporal designs”. In: *Psychological Methods* 25.3 (June 2020), pp. 271–291. DOI: [10.1037/met0000235](https://doi.org/10.1037/met0000235).

Abstract: This study investigates the impact of the time interval (the time passed between 2 consecutive measurements) selection on autoregressive mediation modeling (AMM). For a widely used autoregressive mediation model, via analytical derivations, we explained why and how the conventionally reported time-specific coefficient estimates (e.g.,  $\hat{a}\hat{b}$  and  $\hat{c}'$ ) and inference results in AMM provide limited information and can arrive in even misleading conclusions about direct and indirect effects over time. Furthermore, under the stationarity assumption, we proposed an approach to calculate the overall direct and indirect effect estimates over time and the time lag lengths at which they reach maxima, using AMM results. The derivation results revealed that the overall direct and indirect effect curves are asymptotically invariant to the time interval selection, under stationarity. With finite samples and thus sampling errors and potential computing problems, however, our sim-

ulation results revealed that the overall indirect effect curves were better recovered when the time interval is selected to be closer to half of the time lag length at which the overall indirect effect reaches its maximum. An R function and an R Shiny app were developed to obtain the overall direct and indirect effect curves over time and facilitate the time interval selection using AMM results. Our findings provide another look at the connections between AMM and continuous time mediation modeling and the connections are discussed.