

modeLLtest: An R Package for Unbiased Model Comparison using Cross Validation

Shana Scogin¹, Sarah Petersen², Jeffrey J. Harden¹, and Bruce A. Desmarais³

¹ University of Notre Dame, Department of Political Science ² University of Notre Dame, Department of Mathematics ³ Pennsylvania State University, Department of Political Science

DOI: [10.21105/joss.01542](https://doi.org/10.21105/joss.01542)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 20 June 2019

Published: 01 September 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Selection among statistical models describing the same process is a crucial methodological step in quantitative research. Because different estimators for the same process are commonly available, researchers have developed tests for selecting among models. The R package `modeLLtest` implements a variety of such tests using leave-one-out cross-validation (LOOCV) to adjudicate among estimation methods. LOOCV accounts for outliers in data by comparing results across samples in which one observation has been left out (Diebold & Mariano, 2002).

The R package `modeLLtest` has three main functionalities. It implements an unbiased comparison of two linear parametric, non-nested models (Desmarais & Harden, 2014; Harden & Desmarais, 2011), a test comparing two estimations of the Cox proportional hazards model - the conventional partial likelihood maximization (PLM) and the iteratively reweighted robust model (IRR) (Desmarais & Harden, 2012), and a set of more general functions for comparison of arbitrary models. In addition to making these comparison tests easily accessible in an R package, we improve the running speed of selected functions by making use of C++ to perform LOOCV before pulling the output into R. As a result, these tests now run in a reasonable amount of time on large data sets. Researchers can use the fast LOOCV results as inputs to alternative and new model testing procedures.

Unbiased linear model comparison

When models are parametric and non-nested, the null hypothesis that two models fit the data equally well has often been tested using methods introduced by Vuong (1989) and Clarke (2001, 2003, 2007). These methods compare the Kullback-Leibler Divergence (KLD) of the two models from the true model that generated the data. However, more recently, Desmarais and Harden (2014) showed the estimate of the KLD used by Vuong, the individual log-likelihood contributions, is biased. They also showed the Clarke test to be inconsistent when applied to the difference in KLDs. As a solution, Desmarais and Harden (2014) derived a test based on LOOCV contributions, creating an unbiased KLD estimate. The implementation of Desmarais and Harden's test comparing two methods of estimation is a major component of `modeLLtest`. A primary example of the test's use within this context is in the adjudication among ordinary least squares (OLS) regression, robust regression, and the less-commonly used median regression (MR) (Hao & Naiman, 2007; Harden & Desmarais, 2011; Koenker, 2005). For researchers looking to make more informed decisions regarding model fit rather than relying on conventionally-used methods such as OLS alone, `modeLLtest` significantly improves the ease of choosing among estimation methods.

Cox proportional hazards model

The Cox model is a partial parametric model that does not make assumptions about the baseline hazard. The conventional partial likelihood maximization (PLM) and iteratively reweighted robust (IRR) estimation methods represent two alternatives for parameter estimation. Either could be optimal based on characteristics of the data. Plausible deviations from correct specification and operationalization, which can be caused by issues such as measurement error or omitted variables, can produce substantial bias when the Cox model is estimated by PLM. However, using the IRR estimator in the absence of these problems can result in a loss of efficiency. For this reason, Desmarais and Harden (2012) developed a novel cross-validated median fit (CVMF) test for selecting between the two estimators. The R package `modeLLtest` provides implementation of this test with original contributions which include options to specify methods other than Efron approximation (e.g., Breslow approximation), specify the truncation or trimming level for the robust estimator, adjust fitting weights, and run the test on only a subset of data.

General tools

In addition to providing the above tests, `modeLLtest` also contains tools for model comparison in a more general setting. Given any model specification and estimation method as inputs, `modeLLtest` can create a vector of LOOCV log-likelihood estimates. These vectors subsequently can be fed to a function that performs the unbiased Johnson t-test (see also Desmarais & Harden, 2014; Johnson, 1978) on the two vectors of log-likelihoods in order to select between models.

Functionality and use

The R package `modeLLtest` is designed to be used by both researchers and students looking to compare models or create LOOCV log-likelihood estimates for a given model. The ease and speed of implementing selected tests are significantly improved through the use of C++ to obtain LOOCV log-likelihood estimates. Functions can be tested using data from real-world analyses from Joshi and Mason (2008) and Golder (2010).

Future work

Future work will include providing implementation for additional models to be used in the CVDM and CVLL functions.

Acknowledgements

We thank Andrew Decker for his helpful comments and advice.

References

- Clarke, K. A. (2001). Testing Nonnested Models of International Relations: Reevaluating Realism. *American Journal of Political Science*, 45(3), 724–744. doi:[10.2307/2669248](https://doi.org/10.2307/2669248)
- Clarke, K. A. (2003). Nonparametric Model Discrimination in International Relations. *Journal of Conflict Resolution*, 47(1), 72–93. doi:[10.1177/0022002702239512](https://doi.org/10.1177/0022002702239512)

- Clarke, K. A. (2007). A simple distribution-free test for nonnested model selection. *Political Analysis*, 15(3), 347–363. doi:[10.1093/pan/mpm004](https://doi.org/10.1093/pan/mpm004)
- Desmarais, B. A., & Harden, J. J. (2012). Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model. *Political Analysis*, 20, 113–135. doi:[10.1093/pan/mpr042](https://doi.org/10.1093/pan/mpr042)
- Desmarais, B. A., & Harden, J. J. (2014). An unbiased model comparison test using cross-validation. *Quality & Quantity*, 48(4), 2155–2173. doi:[10.1007/s11135-013-9884-7](https://doi.org/10.1007/s11135-013-9884-7)
- Diebold, F. X., & Mariano, R. S. (2002). Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model Political Analysis. *Journal of Business & Economic Statistics*, 20, 134–144. doi:[10.1198/073500102753410444](https://doi.org/10.1198/073500102753410444)
- Golder, S. N. (2010). Bargaining delays in the government formation process. *Comparative Political Studies*, 43(1), 3–32. doi:[10.1177/0010414009341714](https://doi.org/10.1177/0010414009341714)
- Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. SAGE.
- Harden, J. J., & Desmarais, B. A. (2011). Linear Models with Outliers: Choosing between Conditional-Mean and Conditional-Median Methods. *State Politics & Policy Quarterly*, 11(4), 371–389. doi:[10.1177/1532440011408929](https://doi.org/10.1177/1532440011408929)
- Johnson, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association*, 73(363), 536–544. doi:[10.1080/01621459.1978.10480051](https://doi.org/10.1080/01621459.1978.10480051)
- Joshi, M., & Mason, T. D. (2008). Between democracy and revolution: Peasant support for insurgency versus democracy in nepal. *Journal of Peace Research*, 45(6), 765–782. doi:[10.1177/0022343308096155](https://doi.org/10.1177/0022343308096155)
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. doi:[10.1017/CBO9780511754098](https://doi.org/10.1017/CBO9780511754098)
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2), 307–333. doi:[10.2307/1912557](https://doi.org/10.2307/1912557)