A framework for heteroskedasticity-robust specification and misspecification testing functions for linear models in R

Achim Zeileis Giovanni Millo

28th February 2006

Specification search strategies in econometric regression modelling are based on zero-restrictions testing on a maintained model (or, in the case of non-nested models' comparison, on an encompassing one). Also in the model validation stage many diagnostic tests may be seen as restriction testing on an auxiliary model derived from the maintained one. Thus, software implementation of both restriction tests and of restrictions-based diagnostic tests may rely on the same *computing engine*. We focus first on robustness of the latter under deviations from the classical normal linear regression model, then on a flexible implementation framework giving rise to a number of functions for mainstream tests.

Heteroskedasticity, a frequent concern in econometrics and most notably in cross-sectional data, invalidates standard restriction testing procedures. Asymptotic tests based on heteroskedasticity-consistent (HC) estimates of the covariance matrix (see White, [3]) are consistent w.r.t. test size, but have poor small-sample properties unless appropriately corrected. Versions of HC restriction tests reliable for use in small samples have been available for about twenty years by now, together with experimental evidence on their performance in an empirical setting and recommendations on their use (McKinnon and White, [2]). Unfortunately, as Long and Ervin ([1]) find out, 15 years later that advice went largely unheeded by practitioners, and the situation is unlikely to have changed much since. These suboptimal habits may be rooted in the unavailability of the appropriate versions of test procedures in many statistical packages, at least without *ad hoc* programming. We discuss the implementation of a range of appropriate testing functions in package `lmtest`.

Base R provides for `lm` objects: a `summary()` method performing partial t-tests and an `anova()` method carrying out F-tests for nested model comparison. Unfortunately, variance-covariance matrix estimates other than the standard (i.e. assuming spherical errors) cannot be plugged in. The functions `coeftest()` and `waldtest()` overcome this problem allowing to plug in estimators, e.g. from package `sandwich`, which provides a general framework for specifying HC and HAC estimators in linear regression models (Zeileis, [4]). In addition, `waldtest()` implements several convenience options for specifying the models to be compared. The computational tools for tests that are based on testing a zero-restriction on an auxiliary model can in turn reuse `coeftest()` and `waldtest()`.

This modular implementation of the general framework allows the researcher

to choose his computing tool at every step of the process consistently with the theory. Moreover, the approach can be easily extended reusing the components for other tests.

We briefly present some montecarlo evidence assessing the performance of a range of Wald, LM and LR zero-restriction tests in small samples in a variety of settings characterized by different degrees of heteroskedasticity and departure from normality in the error terms, designed as in Long and Ervin ([1]). We show that in small samples substantial size bias, usually in the sense of overrejection, may affect some of the HC test versions. The implementation is based on just three functions, the standard `waldtest()` and two slight modifications for LR and LM testing, capable of reproducing all the relevant versions of these test statistics present in the literature by adjusting the optional parameters. We conclude with a sketch of the work in progress.

References

[1] J.S. Long and L. Ervin. Using heteroskedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54:217–224, 2000.

[2] J.G. MacKinnon and H. White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29:305–325, 1985.

[3] H. White. A heteroskedasticity consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, 48:817–838, 1980.

[4] A. Zeileis. Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17, 2004.