RUTGERS
School of Arts and Sciences

Department of Statistics

Takeaki Kariya
Nagoya University of Commerce and Business School, Japan

A modelling Framework for Regression with Collinearity

Tuesday, October 4, 2022
12:45 PM

Zoom Meeting: Meeting ID: 99075124232
Password: 952486

https://rutgers.zoom.us/j/99075124232?pwd=UDdPVjRncXZFcXpvbFE0OWJyMVdSUT09

Light refreshments will be served

Abstract: This study addresses a fundamental, yet overlooked, gap between the standard theory and empirical practices in the OLS regression $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}$. To fill it, introducing a new concept "accommodation", this paper formulates a novel conceptual framework for developing our own model selection process in empirical modelling for given $(\boldsymbol{y},\boldsymbol{X})$ with collinearity in $\boldsymbol{X}$. With no use of $\boldsymbol{y}$, the new process enables us to find a class of effective and collinearity-resilient models. In fact, it directly controls not only the sampling variance of each OLSE, which includes Variance Inflation Factor, but also the individual power property of each t-test on regression coefficient, which includes what we call "Power Deflation Factor" as a collinearity factor. This framework will give an ordering on the set of all the sub-models in terms of efficiency and collinearity. And to materialize our model selection process, two computational algorithms are proposed.

Consequently, it will provide an advance model-screening process and serve as an empirical platform for pre-selecting a class of effective models that well accommodate $\boldsymbol{y}$ with both collinearity and inefficiency controlled in advance. In such a class of models, we can freely use such statistical measures and procedures with use of $\boldsymbol{y}$ as OLS estimation, *t*-value, coefficient of determination, stepwise model selection, etc.

It is shown that in terms of predictive sampling variance of the *k*-th OLSE, the lower bound attains if and only if the mean of the explanatory vector $\boldsymbol{x_k}$ is 0 and $\boldsymbol{x_i}'\boldsymbol{x_k}=0$ $(j\neq k)$. Also without using $\boldsymbol{y}$, two algorithms for finding models with collinearity controlled are proposed, so that frequently used model selection procedures can be effectively used. However, in Kariya, Kurata and Hayashi (2022, JFSSA conference) since *t*-statistics are shown to be correlated, the stepwise model selection procedures are ineffective as they stand.

Bio: Professor of Nagoya University of Commerce and Business (2020-). Ph.D. in Statistics (U of Minnesota 75). In the past, Professor of Hitotsubashi U, Kyoto U, Meiji U. etc., Visiting Professor of Rutgers U, LSE, University of Chicago, etc. Published Books *Robustness of Tests* (with B.K.Sinha, Academic Press 89), *Generalized Least Squares* (with H. Kurata, Wiley 04), Asset Pricing (with R. Liu, Springer 03), etc. Published articles; The general MNOVA problem (AS 78), Transformations preserving normality and Wishart-ness (JMA 86 with Nabeya), A nonlinear version of the Gauss-Markov theorem (JASA 85), Equivariant estimation with an ancillary statistic (AS 89), etc. Japan Statistical Society Award (99). President of Japanese Association of Financial Econometrics and Engineering (93-98)