

## Estimation of the covariance matrix based on two types of the forward search algorithm

---

ALEŠ TOMAN

Faculty of Economics  
University of Ljubljana (Slovenia)  
ales.toman@ef.uni-lj.si

Multivariate normal distribution is a crucial assumption in many statistical models and estimation of the population covariance matrix is usually the first key step in modelling multivariate data. The standard sample covariance matrix is commonly used but it gives unreliable estimates if the normal distribution does not fit the data.

Forward search algorithm is an iterative and graphical method for data exploration and robust parameter estimation [1]. The algorithm orders the data according to their distances from the underlying distribution or model. The nearest observations form the initial basic set, which is very robust. The basic set is then increased in size step by step until all observations are included. Parameter estimates and different statistics are computed with the basic sets of increasing sizes.

We will demonstrate the use of the forward search algorithm in the context of robust covariance matrix and confirmatory factor model estimation [2,3]. The former uses Cook's distance to measure the influence of an observation and the later uses observational residuals. The comparison of the two methods applied to real and simulated data sets will show that outliers and influential observations can be model specific. Model-based robust techniques should therefore be emphasized.

### References

- [1] A. C. Atkinson, M. Riani, A. Cerioli, *Exploring multivariate data with the forward search* Springer-Verlag, New York, 2004.
- [2] W.-Y. Poon, Y.-K. Wong, *A forward search procedure for identifying influential observations in the estimation of a covariance matrix*, Structural Equation Modeling **11** (2004) 357–374.
- [3] A. Toman, *Robust confirmatory factor analysis based on the forward search algorithm*, Statistical Papers **55** (2014) 233–252.