

An Instant of Performance Criteria for Outlier Identification

B. Onoghojobi

Department of Statistics, Federal University of Technology, Owerri, Nigeria

Abstract: Problem statement: Different approach to deal with criteria for multivariate outlier identification procedures when it was studentized was presented. **Approach:** In this respect, associated unpractical dynamic linear model and outlier identification were considered. **Results:** A resultant studentized Becker outlier version for multivariate outlier was obtained, as an improvement on the original model. **Conclusion:** The studentized Becker multivariate outlier incorporated the modeling change in an evolving extension.

Key words: Multivariate outlier, studentized Becker, dynamic linear, outlier identification

INTRODUCTION

Statistics has it that due to potential influence of outliers empirical moments are highly subject to bias. Outliers are defined as observation (s) in a data set which appears to be inconsistent with the remainder of that data set. While identification of outliers is often thought of as a means of eliminating observation (s) from data set due to disturbance. A number of methods on performance criteria for multivariate outlier identification procedure are available in the literature.

These include methods of identify all outliers in one step; generalizing the univariate concept of outlier identifiers (Davies and Gather, 1993), as well as the identification rules (Becker and Gather, 1999) and the robust method for outlier identification (Gather and Becker, 1997). However, the above research fails to capture the outlier in dynamic liner model with Markov switching. The general ideas of studentized score method are well established in statistics and Econometrics (Cai *et al.*, 1998, Koenker, 1981; Yang and Tse, 2008; Lyon and Tsai, 1996; Deng and Perron, 2008; Silvia and Cribbari-Neto, 2004). Studentized gene network state space model is a tool for estimating. Time-dependent gene network from time series microarray data (Onoghojobi, 2009).

In this present study we attempt to overcome the difficulties associated with capturing the outlier in dynamic linear model by reformulating the performance criteria for multivariate outlier identification procedure as a studentized Becker outlier identifier and solving same by the usual appropriate procedure.

MATERIALS AND METHODS

The tools and logic for this research are similar to the ressearch of (Becker and Gather, 1999) in terms of

identification rules and the concept of studentizing score test is basin on (Yang and Tse, 2008).

Becker outlier identifiers: Generally identification of outlier means to eliminate observations from a data set to avoid disturbance or further analyses. There are various methods for detecting outlier in multivariate data. But the idea of Davies and Gather (1993) concentrate on the procedures which identify all outlier in one step and generalizing the univariate concept of outlier identification.

Outlier region is defined as region of low probability of the assumed underlying distribution: such as, an outlier is seen as an observation in unlikely position relative to the assumed underlying distribution:

$$x \in \text{out}(\phi, \mu, \Sigma) := \{ \underline{x} \in \mathbf{R}^t : (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) > \chi_{r;1-\phi} \} \quad (1)$$

is called ϕ outlier with respect to the multivariate normal distribution $N(\mu, \Sigma)$.

Hence, the task of estimating the unknown outlier region $\text{out}(N(\phi, \mu, \Sigma))$ is equivalent to the aim of outlier identification. The detection rules of this type here, is so called outlier identifier:

$$(\underline{x}_N, \phi_N) := \left\{ \begin{array}{l} \underline{x} \in \mathbf{R}^t := (\underline{x} - \underline{U})^T \\ Y^{-1}(\underline{x} - \underline{U}) \geq a(r, N, \phi_N) \end{array} \right\} \quad (2)$$

Where:

\underline{U} and Y = Estimator for $\underline{\mu}$ and Σ

a = Normalizing constant

\tilde{X}_N = A sample of size N , the value of ϕ_N is chosen with respect to the sample size

An observation located in the region OR is then detect as outlier.

Studentized Becker outlier identifier main result:
Let $S(\gamma)$ denote (1) which can be rewritten as a score test (Onoghojobi, 2009):

$$\underline{x} \in \text{out}(\phi, \mu, \Sigma) := \{ \underline{x} \in \mathbb{R}^r : v^b \Delta V > \chi_{r,1-\phi} \} \quad (3)$$

Where:

$$V^b = (\underline{x} - \underline{\mu})^T, \Delta = \Sigma^{-1} \text{ and } V = (\underline{x} - \underline{\mu})$$

Using the basic principal (Cai *et al.*, 1998; Koenker, 1981; Yang and Tse, 2008; Deng and Perron, 2008; Silvia and Cribbarri-Neto, 2004) let:

$$S(\hat{\gamma}) = V^b \Delta V = S(\hat{\xi}) 2\sigma_0^4 \quad (4)$$

where, $S(\hat{\xi})$ is the standard score test (Lyon and Tsai, 1996), hence:

$$S(\hat{\xi}) = \frac{V^b \Delta V}{2\sigma_0^4} \quad (5)$$

Koenker (1981) proposed studentizing $S(\hat{\xi})$ by replacing $2\sigma_0^4$ in equation (5) with $\theta = \Sigma[\hat{e}_w^2 - \sigma_0^4] / n$. The studentized score test can be obtained by substituting ϕ in Eq. 5 with regard to Eq. 4:

$$S(\hat{\xi})_s = 2\delta_0^4 S(\hat{\xi}) \quad (6)$$

Hence:

$$S(\hat{\xi})_s = S(\hat{\gamma}) / \theta \quad (7)$$

Where:

$S(\hat{\gamma})$ = The original score test

$S(\hat{\xi})$ = The original score test

$S(\hat{\xi})_s$ = The proposed studentized score test

RESULTS AND DISCUSSION

In order to compare outlier identifier using studentized score test, we can write the established studentized Becker outlier for (1) and (2) as:

$$x \in \text{out}(\phi, \mu, \Sigma) := \{ \underline{x} \in \mathbb{R}^r : S(\gamma)\theta^{-1} > \chi_{r,1-\phi} \} \quad (8)$$

$$\text{OR } (X_n, \phi_N) := \{ \underline{x} \in \mathbb{R}^r : S(\omega)\theta^{-1} \geq a(r, N; \phi_N) \} \quad (9)$$

where, $S(w)$ is an equivalent for $S(\gamma)$ in (2).

A goal instant is the follow-up of Claudia Becker performance criteria for multivariate outlier identification procedures.

CONCLUSION

The studentized Becker outlier technique has been shown to provide better robust identifier of outlier than the conventional performance criteria for multivariate outlier identification procedures.

REFERENCES

- Becker, C. and U. Gather, 1999. The masking breakdown point of multivariate outlier identification rules. *J. Am. Stat. Assoc.*, 94: 947-955. <http://www.jstor.org/pss/2670009>
- Cai, Z., C.M. Hurvch and C.L. Tsai, 1998. Score tests for heteroscedasticity in wavelet regression. *Biometrika*, 85: 229-234. <http://www.jstor.org/pss/2337322>
- Davies, P.L. and U. Gather, 1993. The identification of multiple outliers invited paper with discussion and rejoinder. *J. Am. Stat. Assoc.*, 88: 782-801.
- Deng, A. and P. Perron, 2008. The limit distribution of Cusum of square test under general mixing conditions. *Econ. Theory*, 24: 809-822. http://ideas.repec.org/a/cup/etheor/v24y2008i03p809-822_08.html
- Gather, U. and C. Becker, 1997. Outlier Identification and Robust Method. In: *Handbook of Statistics 15: Robust Inference*, Maddala, G.S. and C.R. Rao (Eds.). Elsevier, Amsterdam, pp: 123-143.
- Koenker, R., 1981. A note on studentizing a test for heteroscedasticity. *J. Econ.*, 17: 107-112. DOI: 10.1016/0304-4076(81)90062-2
- Lyon, J.D. and C.L. Ysai, 1996. A comparison of tests for heteroscedasticity. *J. R. Stat. Soc. Ser. D*, 45: 337-349. <http://www.jstor.org/pss/2988471>
- Onoghojobi, B., 2009. An extension of Yoshida, Imoto, Higuchi and Miyano result. *J. Math. Stat.*, 5: 283-286. <http://www.scipub.org/fulltext/jms2/jms254283-286.pdf>
- Silvia, L.P. and F. Cribbarri-Neto, 2004. Beta regression for modeling rates and proportion. *J. Applied Stat.*, 31: 799-815. DOI: 10.1080/0266476042000214501
- Yang, Z. and Y.K. Tse, 2008. Generalized L.M. tests for functional form and heteroscedascity. *Econ. J.*, 11: 349-376. DOI: 10.1111/j.1368-423X.2008.00242.x