An Extension of Yoshida, Imoto, Higuchi and Miyano Result

Onoghojobi B.
Department of Statistics, University of Ibadan, Nigeria

Abstract: Problem statement: Different approaches to deal with dynamic model when it is studentized are presented. Approach: In this respect, the recursive formula for calculating state space in the canonical form. Results: The asymptotic distribution of their test under the linear system for the gene network and a studentized-dynamic linear model with Markov denoising switching for estimating time-dependent gene network structure were presented. A resultant studentized version for gene network state space model was obtained, as an improvement on the original model. Conclusion/Recommendations: The dynamic model with Markov switching for estimating time-dependent gene regulatory networks handled the problem of modeling change in an evolving time series. The studentized version incorporated the modeling change and test for heteroscedasticity.

Key words: Studentized-dynamic, Markov-denoising switching, studentized approach, recursive formula

INTRODUCTION

In this study we develop a new result for application of importance studentized method to state space with Markov switching for estimating gene regulatory network from time series microarray experiments. A draw back of dynamic linear model with Markov switching for estimating time series-dependent gene network structure $s^{\theta^{\omega}}$ or (7) is that, it is crucially dependent on the assumption that $\varepsilon_i$ is normally distributed. The general ideas of importance studentized method are well established in statistics and econometrics Cai, Hurvich and Tsai [2], Koenker [10], Yang and Tse [14], Lyon and Tsai [11], Deng and Perron [4] and Silvia, Ferrari and Cribari-Neto [5], following a robustification suggested by Bickel [1] we proposed a studentized version $s^{\omega}$.

Yoshida, Imoto and Higuchi [15] proposed a simple method of estimating time-dependent gene network from time series microarray data by dynamic linear models with Markov switching and the state space version was later explored by Yamaguchi and Higuchi [13] in which the state space model was represented by linear system for the gene network.

A state space model consists of a state equation and an observation equation. The state equation models the process of the state while the observation equation links the observations to the states. A standard linear state space model can be written as:

$$X_t = F X_{t-1} + \eta_t$$  \hspace{1cm} (1)

$$Y_t = T X_t + \varepsilon_t$$  \hspace{1cm} (2)

Where:

- $X_t = \text{The observation vector}$.
- $F_t = \text{The state transaction matrix}$ and
- $T_t = \text{Observation matrix}$

It is assumed that $\eta_t$ and $\varepsilon_t$ are independent and identically distributed, with $\eta_t \sim N(0, \alpha)$ and $\varepsilon_t \sim N(0, \beta)$. The system matrices ($F_t$, $T_t$) and the covariance matrices ($\alpha$, $\beta$) may contain unknown parameter $\lambda$ and $\delta$ respectively.

The stationary linear problem is when $F_t$ and $T_t$ are constant matrices solved by Wiener [12] and Kailath [6] and the non-stationary case was solved by Dacheng, Xu-Feng and Hulin [5], Kalman [7], Kalman and Bucy [9], Kalman [8], Yamaguchi and Higuchi [13] and Yoshida, Imoto and Higuchi [15] stated that if the true network structure underlying the data changes at certain point, the fitting of the usual dynamic linear model fail to estimate the structure of gene network, hence the efficient information from the data cannot be obtained. They therefore proposed a dynamic linear model with Markov switching time-dependent gene network structure for time series gene expression data, without taken into consideration the effect of studentizing the model as a test for heteroscedasticity. To solve this problem, we propose a studentized dynamic linear
model with Markov switching for estimating time gene
network from time series gene expression data
\( s(\theta) \). The new modification are necessary for
conducting test for heteroscedasticity, estimation of
parameters and their signal extraction of the state
vector.

The study is organized as follows: we introduce the
state model with Markov switching in details and derive
a new score test to handle the state space model.

**MATERIALS AND METHODS**

**Gene network state space model for linear system:**
The tools and logic for this research are similar to the
work of Yamaguchi and Higuchi\(^ {13} \) and Yoshida, Imoto
and Higuchi\(^ {15} \), when dealing with estimating Time
Series Dependent Gene Networks from Time Series
Microarray Data. Also Yang and Tse\(^ {14} \) for the tools
and process of studentizing test. The state space model
with Markov switching for estimating time dependent
gene regulatory networks from time series microarray
experiment is derived from a vector \( r \) observed random
variables \( Y_t \) defined in (2) and \( X_t \) in (1).

The state space model \( Y_t \) in its canonical form
implicitly represents a relationship \( r \) genes across time
point see Yoshida, Imoto, Higuchi\(^ {15} \). The method is \( Y_t \).

Let \( (F_t, T_t, D_t) \) denote the model parameters
depending on the time index, where:

\[
D_t = \text{diag} \{ d_1, \ldots, d_r \} \text{ of } \varepsilon_t \text{ such that:}
\]

\[
D_t^{-1/2} T_t = L_t \Lambda_t U_t
\]

Where:

\( L_t = \) The matrix of \( p \) orthogonal vector length \( r \)
\( \Lambda_t = \) The diagonal matrix that contain \( p \) singular
values and \( U_t \) is a \( p \times p \) orthogonal matrix

The in verse of (3) is given as:

\[
T_t^{-1} = U_t \Lambda_t^{-1} L_t^{-1}
\]

From Eq. 1 and 4:

\[
T_t^{01} (Y_t - \varepsilon_t) = (T_t X_t) T_t^{01}
\]

\[
T_t^{01} (Y_t - \varepsilon_t) = T_t T_t^{01} X_t
\]

\[
T_t^{01} D_t^{-1/2} (Y_t - \varepsilon_t) = X_t
\]

\[
T_t^{01} D_t^{-1/2} (Y_t - \varepsilon_t) = X_t
\]

where, \( T_t T_t^{01} = D_t^{1/2} \).

\[
T_t^{01} D_t^{-1/2} (Y_t - \varepsilon_t) = F_t X_{t-1} + \eta_t
\]

where, \( X_{t-1} = T_t^{01} D_t^{-1/2} (Y_{t-1} - \varepsilon_{t-1}) + \eta_t \).

From Eq. 5 we have:

\[
T_t^{01} D_t^{-1/2} (Y_t - \varepsilon_t) = F_t T_t^{01} D_t^{-1/2} (Y_{t-1} - \varepsilon_{t-1}) + \eta_t
\]

\( F_t \) is known as the state transition matrix which
captures the intensity of the interaction. By introduction
of an interaction matrix \( M_t \) such that \( M_t = D_t^{-1/2} T_t F_t T_t^{01} \)
the state space model represent the linear system for the
gene network as:

\[
D_t^{-1/2} (Y_t - \varepsilon_t) = M_t D_t^{-1/2} (Y_{t-1} - \varepsilon_{t-1}) + D_t^{-1/2} T_t \eta_t
\]

\section*{RESULTS AND DISCUSSION}

**Result of Studentized gene network state space
model for linear system:** The main analytical result of
this study was obtained in (13) as shown below. Let \( s(\theta) \) denote (7) which can be rewritten as a score test:

\[
s(\theta) = T_t^{01} (Y_t - \varepsilon_t) T_t
\]

For details prove contact the author.

We begin by establishing a standard score test
from (8). Hence, suppose that:

\[
s(\theta) = T_t^{01} V T_t
\]

where, \( V = (Y_t - \varepsilon_t) \).

Using the ideas of Cai \textit{et al.}\(^ {2} \), Koenter\(^ {10} \), Yang
and Tse\(^ {14} \), Deng and Perron\(^ {4} \) and Silvia \textit{et al.}\(^ {5} \), let:

\[
s(\omega) = T_t^{01} V T_t = s(\omega) 2 \sigma_t^2
\]

where, \( s(\omega) \) is a standard score test Lyon and Tsa\(^ {2} \),
therefore:

\[
s(\omega) = \frac{T_t^{01} V T_t}{2 \sigma_t^2}
\]
Koenter\textsuperscript{10} proposed studentizing the statistic $s\left(\hat{\omega}\right)$ by replacing $2\alpha_0^i$ given in Eq. 11 with $
abla \phi = \sum \left(\hat{e}_n - \sigma_0^2\right) / n$. The resulting studentized score test can be obtained by substituting $\phi$ into Eq. 11 with regards to Eq. 10:

$$s\left(\hat{\omega}\right)_i = \frac{s\left(\hat{\theta}\right)}{\phi}$$  \hspace{1cm} (12)

Hence:

$$s\left(\hat{\omega}\right)_i = s\left(\hat{\omega}\right)_i / \phi$$  \hspace{1cm} (13)

Where:

$s\left(\hat{\theta}\right) = \text{The original score test}$

$s\left(\hat{\omega}\right)_i = \text{The standard score test}$

$s\left(\hat{\omega}\right)_i = \text{The proposed studentized score test result.}$

CONCLUSION

The studentized test result in (13), has eliminated the draw back in the dynamic linear model with Markov Switching for estimating Time Series Dependent Gene structure in (7), the point raised by the studentized gene network might plausibly be considered common knowledge, except for the fact that proposing tests for heteroscedasticity based on Gaussian assumption for Markov switching are still being investigated.

The dynamic model with Markov switching for estimating time-dependent gene regulatory networks handled the problem of modeling change in an evolving time series. The studentized version incorporates the modeling change and test for heteroscedasticity. In this study, we established the studentized version of the dynamic system generated from the dynamic system itself using the technique based on Gibbs sampling. It is an improvement on the state space model with Markov switching for estimating time-dependent gene-regulatory networks from time series microarray experiments. We conclude that the proposed studentised gene network appear to be useful, flexible and accurate framework for dynamic model with Markov switching.

REFERENCES

1. Bickel, P.J., 1978. Using residuals robustly 1: Tests for heteroscedasticity, nonlinearity. Ann. Stat., 6: 266-291. http://projecteuclid.org/DpubS?service=UI&version=1.0&verb=Display&handle=euclid.aos/1176344124
2. Cai, Z., C.M. Hurvich and C. Tsai, 1998. Score tests for heteroscedasticity in wavelet regression. Biometrika, 85: 229-234. DOI: 10.1093/biomet/85.1.229.
3. Dacheng, L., N. Xu-Feng and W. Hulin, 2006. Mixed-effects state space models for analysis of longitudinal dynamic systems. www.urmc.rochester.edu/smd/biostat/people/faculty/wu.html
4. Deng, A. and P. Perron, 2006. The limit distribution of CUSUM of square test under general mixing conditions. Unpublished manuscript Department of Economics, Boston University. http://ideas.repec.org/p/bos/macppr/wp2006-004.html
5. Silvia, L., P. Ferraei and F. Cribarri-Neto, 2004 Beta regression for modeling rates and Proportions. J. Applied Stat., 31: 799-815, 2004. http://ideas.repec.org/a/taf/japsta/v31y2004i7p799-815.html
6. Kailath, T., 1981. Lectures on Wiener and Kalman Filtering. Springer Verlag Publishers, UK., ISBN: 038781664X, pp: 187.
7. Kalman, R.E., 1960. A new approach to linear filtering and prediction theory. J. Basic Eng. Trans. AMSE, Ser. D., 82: 95-108. http://citeseer.comp.nus.edu.sg/context/57600/0
8. Kalman, R.E. 1963. Mathematical description of linear dynamical systems. SIAM J. Control, 1: 152-192. http://scitation.aip.org/content/scitation/servlet/GetServlet?prog=normal&id=JSCDOC000001000002000152000001&idtype=cips&gifs=yes
9. Kalman, R.E. and R.S. Bucy, 1961. New results in linear filtering and prediction theory. Trans. AMSE Ser. D. J. Basic Eng., 83: 95-108. http://citeseer.comp.nus.edu.sg/context/57600/0
10. Koenker, R., 1981. A note on studentizing a test for heteroscedasticity. J. Econometr., 17: 107-112. http://ideas.repec.org/a/eee/econom/v17y1981i1p107-112.html
11. Lyon, J. and C.L. Tsai, 1996. A comparison of tests for heteroscedasticity. Statistician, 45: 337-349. http://cat.inist.fr/?aModele=afficheN&cpsidt=3250373
12. Wiener, N., 1949. The Extrapolation, Interpolation and Smoothing of Stationary time Series with Engineering Applications: With Engineering Applications. The MIT Press, ISBN: 026223002X, pp: 163.

13. Yamaguchi, R. and T. Higuchi, 2006. State space Approach with maximum likelihood principle to identify the system of generating time-course gene expression data yeast. Int. J. Data Mining Bioinform., 1: 77-87. DOI: 10.1504/IJDMB.2006.009922

14. Yang, Z. and Y. Tse, 2008. Generalized LM tests for functional form and heteroscedasticity. Econometr. J., 11: 349-376. DOI: 10.1111/J.1368-423X.2008.00242.X.

15. Yoshida, R., S. Imoto and T. Higuchi, 2005. Estimating time dependent gene networks from time series microarray data by dynamic linear models with Markov switching. Proceeding 4th Computation Systems Bioinformatics, Aug. 08-11, IEEE Computer Society Washington DC., USA., pp: 289-298. DOI: 10.1109/CSB.2005.32