2018年6月15日　統計数理研究所　オープンハウス

Analysis of variance for multivariate time series

Hideaki Nagahata* and Masanobu Taniguchi (Risk Analysis Research Center, Institute of Statistical Mathematics(*) / Waseda University)

Classical ANOVA works well only for multivariate time series with no autocorrelation. But new ANOVA works well for multivariate time series.

For independent observations, Analysis of variance (ANOVA) has been enoughly tailored. Recently there has been much demand for ANOVA of dependent observations in many fields. For example it is important to analyze differences among industry averages of financial data. However ANOVA for dependent observations has been immature. In this paper, we study ANOVA for dependent observations. Specifically, we show the asymptotics of classical tests proposed for independent observations and give a sufficient condition for them to be asymptotically χ^2 distributed. If the sufficient condition is not satisfied, we suggest a likelihood ratio test based on Whittle likelihood and derive an asymptotic χ^2 distribution of our test.

1 Theoretical results

1.1 Setting

Let p vetor-valued series $\boldsymbol{X}_{i1},\dots \boldsymbol{X}_{in_i}$ be generated from

$$\boldsymbol{X}_{it}=\boldsymbol{\mu}+\boldsymbol{\alpha}_i+\boldsymbol{\epsilon}_{it},\ t=1,\dots,n_i,\ i=1,\dots,q,$$

where

- $\boldsymbol{\epsilon}_i\equiv\{\boldsymbol{\epsilon}_{it};\ t=1,\dots,n_i\}$, $i=1,\dots,q$, are stationary with mean $\boldsymbol{0}$, autocovariance matrix $\boldsymbol{\Gamma}(\cdot)$ and spectral density matrix $\boldsymbol{f}(\lambda)$,
- $\{\boldsymbol{\epsilon}_{it};\ t=1,\dots,n_i\}$, $i=1,\dots,q$, are mutually independent.
- $\{\boldsymbol{\epsilon}_{it}\}$ is generated from the generalized linear process:

$$\boldsymbol{\epsilon}_{it}=\sum_{j=0}^{\infty}\boldsymbol{A}(j)\boldsymbol{\eta}_i(t-j),\ \sum_{j=0}^{\infty}\|\boldsymbol{A}(j)\|<\infty,$$

where $\boldsymbol{\eta}_i(t)\overset{i.i.d.}{\sim}(\boldsymbol{0},\boldsymbol{G})$, and $\boldsymbol{A}(j)$s are $p\times p$ constant matrices.

Consider the problem of testing

$$H:\boldsymbol{\alpha}_1=\cdots=\boldsymbol{\alpha}_q.$$

1.2 Classical method and result

- For independent observations, the following likelihood ratio test (1), Lawley-Hotelling test (2), and Bartlett-Nanda-Pillai test (3) have been proposed:

$$LR\equiv -n\log\{|\hat{\mathcal{S}}_E|/|\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H|\},\quad(1)$$
$$LH\equiv n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\},\quad(2)$$
$$BNP\equiv n\mathrm{tr}\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H)^{-1},\quad(3)$$

where

$$\hat{\mathcal{S}}_H\equiv\sum_{i=1}^{q}n_i(\hat{\boldsymbol{X}}_{i\cdot}-\hat{\boldsymbol{X}}_{\cdot\cdot})(\hat{\boldsymbol{X}}_{i\cdot}-\hat{\boldsymbol{X}}_{\cdot\cdot})',$$

$$\hat{\mathcal{S}}_E\equiv\sum_{i=1}^{q}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})'.$$

Assumption 1. $\det\{\boldsymbol{f}(0)\}>0$.

Assumption 2 (Uncorrelated disturbance)**.**

$$\boldsymbol{\Gamma}(j)=\boldsymbol{0}\ \textit{for all}\ j\neq 0.$$

Theorem 1. *Suppose Assumptions 1-2, and that $\{\boldsymbol{\epsilon}_{it}\}$ has the forth-order cumulant. Then, under the null hypothesis H, the tests LR, LH, and $BNP\xrightarrow{d}\chi^2_{p(q-1)}$.*

1.3 New method and result

- For dependent observations, we propose the following Whittle likelihood test (4):

$$WLR\equiv 2\{l(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\alpha}}_i)-l(\hat{\boldsymbol{\mu}},\boldsymbol{0})\}\quad(4)$$
$$=\sum_{i=1}^{q}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_{i\cdot}'\{2\pi\boldsymbol{f}(0)\}^{-1}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_{i\cdot},\quad(5)$$

where

– Whittle's approximation to the likelihood function:

$$l(\boldsymbol{\mu},\alpha_i)\equiv-\frac{1}{2}\sum_{i=1}^{q}\sum_{s=0}^{n_i-1}\mathrm{tr}\left\{\boldsymbol{I}_i(\lambda_s)\boldsymbol{f}(\lambda_s)^{-1}\right\},\lambda_s=2\pi s/n_i,$$
$$\boldsymbol{I}_i(\lambda)\equiv\tfrac{1}{2\pi n_i}\left\{\textstyle\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\mu}-\boldsymbol{\alpha}_i)e^{i\lambda t}\right\}\left\{\textstyle\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\boldsymbol{\mu}-\boldsymbol{\alpha}_i)e^{i\lambda u}\right\}^*,$$

– from $\frac{\partial l(\boldsymbol{\mu},\boldsymbol{0})}{\partial\boldsymbol{\mu}}=\boldsymbol{0}$, $\frac{\partial l(\boldsymbol{\mu},\alpha_i)}{\partial\boldsymbol{\mu}}=\boldsymbol{0}$, $\frac{\partial l(\boldsymbol{\mu},\alpha_i)}{\partial\boldsymbol{\alpha}_i}=\boldsymbol{0}$,

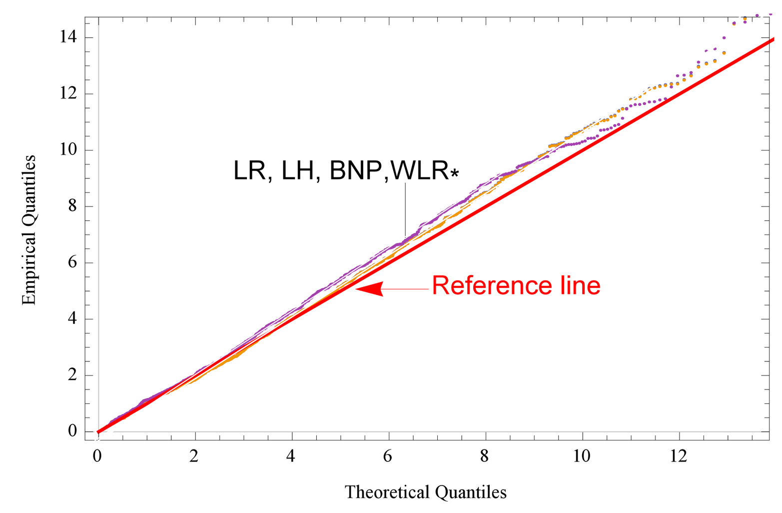
$$\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}\equiv\frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}\boldsymbol{X}_{it},$$
$$\boldsymbol{\alpha}_i=\hat{\boldsymbol{\alpha}}_i\equiv\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}}).$$

Theorem 2. *Suppose Assumption 1. Then, under the null hypothesis H, the test $WLR\xrightarrow{d}\chi^2_{p(q-1)}$, without Assumption 2.*

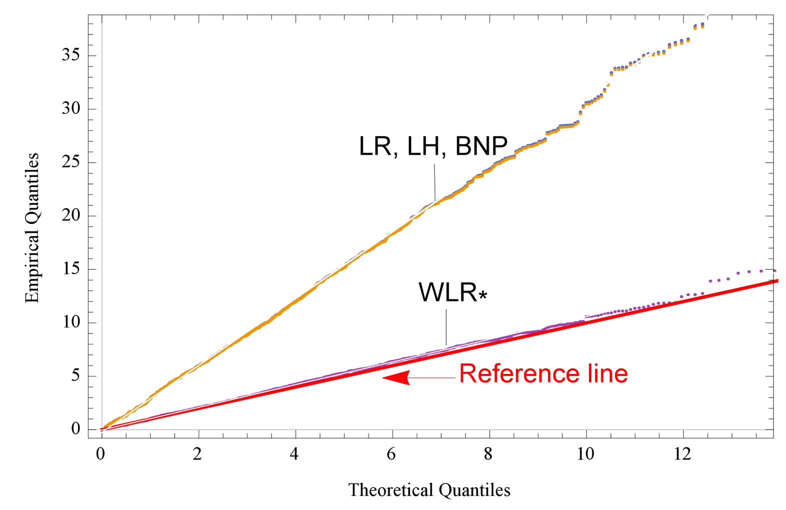
Remark 1. *Under some regularity conditions, as n_i, $i=1,\dots,q$, tend to ∞, $\hat{\boldsymbol{f}}_i(\lambda)\xrightarrow{p}\boldsymbol{f}(\lambda)$ for $i=1,\cdots,q$. So Using this, we can replace $\boldsymbol{f}(0)$ in (5) by $\hat{\boldsymbol{f}}_i(0)$:*

$$WLR^*=\sum_{i=1}^{q}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\hat{\boldsymbol{f}}_i(0)\right\}^{-1}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i.$$

2 Simulation results



(a) Uncorrelated observations from *DCC-GARCH*(1, 1)



(b) Dependent observations from $VAR(1)$

Figure 1: Q-Q plot whose theoretical quantiles are given by $\chi^2_{p(q-1)}$, and these empirical quantiles are calculated by LR, LH, BNP, and WLR^*.

3 Application to financial data

- We apply LR, LH, BNP, and WLR to the daily log data of some stocks.
- Data: This data set consists of three groups with 2 dimensions and about 2500 - 5000 cell lines.
  - 3 groups, (i) electric appliance, (ii) film, and (iii) financial industries.
  - Each industry includes 2 companies as (i)NEC & TOSHIBA, (ii) TOEI & TOHO, (iii)MUFJ & SMBC.
- Property of Data: very law S.C.F. among the 3 industries, and high S.A.C.F. of 2 companies in each industry.

→ All of the tests reject hypothesis H, and their P-values are all around 1.0.

⇒ But only WLR^* is valid because of high autocorrelation of data.