Computational efficiency of generalized variance and vector variance

Shamshuritawati Sharif, Wan Nur Syahidah Wan Yusoff, Zurni Omar, and Suzilah Ismail

Citation: AIP Conference Proceedings **1635**, 906 (2014); doi: 10.1063/1.4903690
View online: http://dx.doi.org/10.1063/1.4903690
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1635?ver=pdfcov
Published by the AIP Publishing

Computational Efficiency of Generalized Variance and Vector Variance

Shamshuritawati Sharif[a], Wan Nur Syahidah Wan Yusoff[b], Zurni Omar[c] and Suzilah Ismail[d]

[a,c,d]*School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.*
shamshurita@uum.edu.my, zurni@uum.edu.my, halizus@uum.edu.my
[b]*Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia.*
wnsyahidah@ump.edu.my

Abstract. In multivariate statistical quality control, the existing tests known as Generalized Variance (GV) and Vector Variance (VV), plays an important role in measuring process variability. In this paper, we present the computational efficiency of both tests to illustrate that their complexity as a function of dimension. From the mathematical derivation and simulation study, the computational efficiency of VV outperforms GV, particularly when the number of variables is large.

Keywords: *multivariate variability, covariance matrix, computational efficiency.*
PACS: 02.50

THE STABILITY TEST OF COVARIANCE STRUCTURE

Nowadays, testing the stability of process variability in multivariate setting is very crucial. There are many different methods discussed in the literature which are constructed based on the notion of multivariate variability measure. According to Djauhari [1], a multivariate variability measure is defined as a non-negative, real valued function of a covariance matrix such that the more scattered the population, the larger the value of that function and, conversely, the less scattered the population, the smaller the value of that function. In multivariate setting, variability is numerically represented by covariance matrix. The importance of covariance structure stability has been shown in many research areas such as medical research [2], genetic research [3], [4], [5], [6], personality research [7], financial industry [8],[1], [9] real estate industry [10] and service industry [11], [12].

In this paper, we focus on the use of the determinant of sample covariance matrix or, equivalently, sample generalized variance (GV) and vector variance (VV) as multivariate variability measures to monitor the stability of covariance structure. We can see the role of generalized variance in controlling the stability of covariance structure in [2], [1], [3], whereas the role of vector variance as multivariate variability measure has been shown in [4], [1] and [13].

Many researchers employ GV due to its simplicity in geometric interpretation and computation. However, GV requires the condition that the covariance matrix must be non-singular. In addition, it is quite cumbersome to compute covariance matrices if the data set is of high dimension. Therefore, to overcome this condition, [1] and [4] suggested using VV in controlling the stability of covariance structure.

We begin in the next section with the issue of non-singularity of covariance matrix. Later on, we present the research results on computational efficiency of GV and VV. In that section, its computational complexity and computational time of these two methods will be provided. Finally, a conclusion will be delivered in the last section.

International Conference on Quantitative Sciences and Its Applications (ICOQSIA 2014)
AIP Conf. Proc. 1635, 906-911 (2014); doi: 10.1063/1.4903690

NON SINGULARITY OF THE COVARIANCE MATRIX

The first issue is matrix singularity. The determinant of covariance matrix, i.e GV will be zero if (1) an entire row is zero, (2) two rows or column are equal, or (3) a row or column is a constant multiple of another row or column. If the determinant is zero, the matrix will be singular and does not have an inverse. However, VV does not require the condition that covariance matrix must be non-singular. The following example will clarify this statement.

Let $\Sigma = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$. Then, GV = $|2(2) - 2(2)| =$ 0 and thus the matrix Σ is a singular. However, the value of VV $= 2^2 + 2^2 + 2^2 + 2^2 =$ 16. Mathematically, we can conclude that VV is more lenient than GV.

COMPUTATIONAL EFFICIENCY

Next, the second issue is about the computational efficiency of GV and VV. In this section, the total number of operations and computation time needed to calculate both GV and VV are shown.

Total Number Operations Needed for GV and VV

In this section, we present the algorithm of GV and VV for p=2 to p=5.

(i) Case p = 2

Let $S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$. Then,

(a) GV $= (s_{11}s_{22}) - (s_{12}s_{21})$

Total number of operations needed is 3 (2 multiplications and 1 subtraction).

(b) VV= $s_{11}^2 + s_{12}^2 + s_{21}^2 + s_{22}^2$.

Total number of operations needed is 7 (4 multiplications and 3 additions).

(ii) Case p = 3

Let $S = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix}$. Then,

(a) GV $= s_{11}(s_{22}s_{33} - s_{23}s_{32}) - s_{12}(s_{33}s_{21} - s_{23}s_{31}) + s_{13}(s_{32}s_{21} - s_{22}s_{31})$

Total number operations needed are 14 (9 multiplications, 1 addition and 4 subtractions).

(b) VV $= s_{11}^2 + s_{12}^2 + s_{13}^2 + s_{21}^2 + s_{22}^2 + s_{23}^2 + s_{31}^2 + s_{32}^2 + s_{33}^2$.

Total number operations needed are 17 (9 multiplications, 8 additions).

(iii) Case $p = 4$

Let $S = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{pmatrix}$. Then,

(a)
$$\begin{aligned} \text{GV} = {} & s_{11}\big((s_{22}s_{33}s_{44} - s_{22}s_{34}s_{43}) - (s_{23}s_{32}s_{44} - s_{23}s_{34}s_{42}) + (s_{24}s_{32}s_{43} - s_{24}s_{33}s_{42})\big) - \\ & s_{12}\big((s_{21}s_{33}s_{44} - s_{21}s_{34}s_{43}) - (s_{23}s_{31}s_{44} - s_{23}s_{34}s_{41}) + (s_{24}s_{31}s_{43} - s_{24}s_{33}s_{41})\big) + \\ & s_{13}\big((s_{21}s_{32}s_{44} - s_{21}s_{34}s_{42}) - (s_{22}s_{31}s_{44} - s_{22}s_{34}s_{41}) + (s_{24}s_{31}s_{42} - s_{24}s_{32}s_{41})\big) - \\ & s_{14}\big((s_{21}s_{32}s_{43} - s_{21}s_{33}s_{42}) - (s_{22}s_{31}s_{43} - s_{22}s_{33}s_{41}) + (s_{23}s_{31}s_{42} - s_{23}s_{32}s_{41})\big) \end{aligned}$$

Total number operations needed are 75 (52 multiplications, 5 additions and 18 subtractions).

(b)
$$\text{VV} = s_{11}^2 + s_{12}^2 + s_{13}^2 + s_{14}^2 + s_{21}^2 + s_{22}^2 + s_{23}^2 + s_{24}^2 + s_{31}^2 + s_{32}^2 + s_{33}^2 + s_{34}^2 + s_{41}^2 + s_{42}^2 + s_{43}^2 + s_{44}^2$$

Total number operations needed are 31 (16 multiplications, 15 additions).

(iv) Case $p = 5$

Let $S = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} \end{pmatrix}$. Then

(a) GV =

$$s_{11}\begin{pmatrix} (s_{22}s_{33}s_{44}s_{55} - s_{22}s_{33}s_{45}s_{54}) - (s_{22}s_{34}s_{43}s_{55} - s_{22}s_{34}s_{45}s_{53}) + (s_{22}s_{35}s_{43}s_{54} - s_{22}s_{35}s_{44}s_{53}) - \\ (s_{23}s_{32}s_{44}s_{55} - s_{23}s_{32}s_{45}s_{54}) + (s_{23}s_{34}s_{42}s_{55} - s_{23}s_{34}s_{45}s_{52}) - (s_{23}s_{35}s_{42}s_{54} - s_{23}s_{35}s_{44}s_{52}) + \\ (s_{24}s_{32}s_{43}s_{55} - s_{24}s_{32}s_{45}s_{53}) - (s_{24}s_{33}s_{42}s_{55} - s_{24}s_{33}s_{45}s_{52}) + (s_{24}s_{35}s_{42}s_{53} - s_{24}s_{35}s_{43}s_{52}) - \\ (s_{25}s_{32}s_{43}s_{54} - s_{25}s_{32}s_{44}s_{53}) + (s_{25}s_{33}s_{42}s_{54} - s_{25}s_{33}s_{44}s_{52}) - (s_{25}s_{34}s_{42}s_{53} - s_{25}s_{34}s_{43}s_{52}) \end{pmatrix} -$$

$$s_{12}\begin{pmatrix} (s_{21}s_{33}s_{44}s_{55} - s_{21}s_{33}s_{45}s_{54}) - (s_{21}s_{34}s_{43}s_{55} - s_{21}s_{34}s_{45}s_{53}) + (s_{21}s_{35}s_{43}s_{54} - s_{21}s_{35}s_{44}s_{53}) - \\ (s_{23}s_{31}s_{44}s_{55} - s_{23}s_{31}s_{45}s_{54}) + (s_{23}s_{34}s_{41}s_{55} - s_{23}s_{34}s_{45}s_{51}) - (s_{23}s_{35}s_{41}s_{54} - s_{23}s_{35}s_{44}s_{51}) + \\ (s_{24}s_{31}s_{43}s_{55} - s_{24}s_{31}s_{45}s_{53}) - (s_{24}s_{33}s_{41}s_{55} - s_{24}s_{33}s_{45}s_{51}) + (s_{24}s_{35}s_{41}s_{53} - s_{24}s_{35}s_{43}s_{51}) - \\ (s_{25}s_{31}s_{43}s_{54} - s_{25}s_{31}s_{44}s_{53}) + (s_{25}s_{33}s_{41}s_{54} - s_{25}s_{33}s_{44}s_{51}) - (s_{25}s_{34}s_{41}s_{53} - s_{25}s_{34}s_{43}s_{51}) \end{pmatrix} +$$

$$s_{13}\begin{pmatrix} (s_{21}s_{32}s_{44}s_{55} - s_{21}s_{32}s_{45}s_{54}) - (s_{21}s_{34}s_{42}s_{55} - s_{21}s_{34}s_{45}s_{52}) + (s_{21}s_{35}s_{42}s_{52} - s_{21}s_{35}s_{44}s_{52}) - \\ (s_{22}s_{31}s_{44}s_{55} - s_{22}s_{31}s_{45}s_{54}) + (s_{22}s_{34}s_{41}s_{55} - s_{22}s_{34}s_{45}s_{51}) - (s_{22}s_{35}s_{41}s_{54} - s_{22}s_{35}s_{44}s_{51}) + \\ (s_{24}s_{31}s_{42}s_{55} - s_{24}s_{31}s_{45}s_{52}) - (s_{24}s_{32}s_{41}s_{55} - s_{24}s_{32}s_{45}s_{51}) + (s_{24}s_{35}s_{41}s_{52} - s_{24}s_{35}s_{42}s_{51}) - \\ (s_{25}s_{31}s_{42}s_{54} - s_{25}s_{31}s_{44}s_{52}) + (s_{25}s_{32}s_{41}s_{54} - s_{25}s_{32}s_{44}s_{51}) - (s_{25}s_{34}s_{41}s_{52} - s_{25}s_{34}s_{42}s_{51}) \end{pmatrix} -$$

$$s_{14}\begin{pmatrix}(s_{21}s_{32}s_{43}s_{55}-s_{21}s_{32}s_{45}s_{53})-(s_{21}s_{33}s_{42}s_{55}-s_{21}s_{33}s_{45}s_{52})+(s_{21}s_{35}s_{42}s_{53}-s_{21}s_{35}s_{43}s_{52})-\\(s_{22}s_{31}s_{43}s_{55}-s_{22}s_{31}s_{45}s_{53})+(s_{22}s_{33}s_{41}s_{55}-s_{22}s_{33}s_{45}s_{51})-(s_{22}s_{35}s_{41}s_{53}-s_{22}s_{35}s_{43}s_{51})+\\(s_{23}s_{31}s_{42}s_{55}-s_{23}s_{31}s_{45}s_{52})-(s_{23}s_{32}s_{41}s_{55}-s_{23}s_{32}s_{45}s_{51})+(s_{23}s_{35}s_{41}s_{52}-s_{23}s_{35}s_{42}s_{51})-\\(s_{25}s_{31}s_{42}s_{53}-s_{25}s_{31}s_{43}s_{52})+(s_{25}s_{32}s_{41}s_{53}-s_{25}s_{32}s_{43}s_{51})-(s_{25}s_{33}s_{41}s_{52}-s_{25}s_{33}s_{42}s_{51})\end{pmatrix}+$$

$$s_{15}\begin{pmatrix}(s_{21}s_{32}s_{43}s_{54}-s_{21}s_{32}s_{44}s_{53})-(s_{21}s_{33}s_{42}s_{54}-s_{21}s_{33}s_{44}s_{52})+(s_{21}s_{34}s_{42}s_{53}-s_{21}s_{34}s_{43}s_{52})-\\(s_{22}s_{31}s_{43}s_{54}-s_{22}s_{31}s_{44}s_{53})+(s_{22}s_{33}s_{41}s_{54}-s_{22}s_{33}s_{44}s_{51})-(s_{22}s_{34}s_{41}s_{53}-s_{22}s_{34}s_{43}s_{51})+\\(s_{23}s_{31}s_{42}s_{54}-s_{23}s_{31}s_{44}s_{52})-(s_{23}s_{32}s_{41}s_{54}-s_{23}s_{32}s_{44}s_{51})+(s_{23}s_{34}s_{41}s_{52}-s_{23}s_{34}s_{42}s_{51})-\\(s_{24}s_{31}s_{42}s_{53}-s_{24}s_{31}s_{43}s_{52})+(s_{24}s_{32}s_{41}s_{53}-s_{24}s_{32}s_{43}s_{51})-(s_{24}s_{33}s_{41}s_{52}-s_{24}s_{33}s_{42}s_{51})\end{pmatrix}$$

Total number operations needed are 484 (365 multiplications, 27 additions and 92 subtractions).

(b) $$\mathrm{VV}=s_{11}^2+s_{12}^2+s_{13}^2+s_{14}^2+s_{15}^2+s_{21}^2+s_{22}^2+s_{23}^2+s_{24}^2+s_{25}^2+s_{31}^2+s_{32}^2+s_{33}^2+ s_{34}^2+s_{35}^2+s_{41}^2+s_{42}^2+s_{43}^2+s_{44}^2+s_{45}^2+s_{51}^2+s_{52}^2+s_{53}^2+s_{54}^2+s_{55}^2$$

Total number operations needed are 49 (25 multiplications, 24 additions).

The above results reveal that the higher dimension of data, the greater the number of operations needed to compute GV if compared to VV.

Computational Time of Generalized Variance and Vector Variance

In this section, we present the computational time of GV and VV for p=2, p=3, p=4, and p=5. The aim of our simulation study is to illustrate that computational efficiency as a function of the dimension. The simulation is implemented using the MATLAB R2009a programming and the experiments were performed on Intel Core2 Duo with 2.40GHz-based machine having 3063MB of main memory. Table 1 displays the comparison on time consuming in computation of GV and VV at sample size of 10, 20 and 30. We demonstrate this information by using a plot in Figure 1 – Figure 3.

TABLE [1]. Computational time of GV and VV for p=2 to p=5 at different sample size, n

n	***p***	**GV**	**VV**
10	p=2	0.10	1.00
	p=3	0.10	2.00
	p=4	1200.00	40.00
	p=5	1700.00	42.00
20	p=2	0.20	1.40
	p=3	0.25	2.30
	p=4	1500.00	50.00
	p=5	3100.00	52.00
30	p=2	0.25	2.00
	p=3	0.38	2.80
	p=4	1600.00	50.00
	p=5	3200.00	58.00

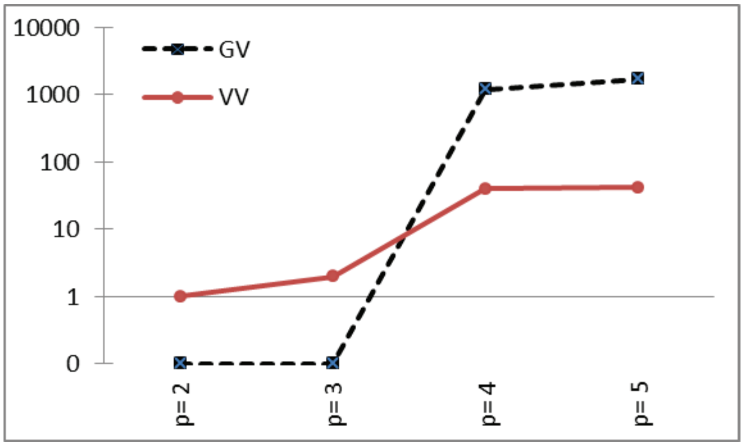

FIGURE 1. Computational time of sample size, *n=10*

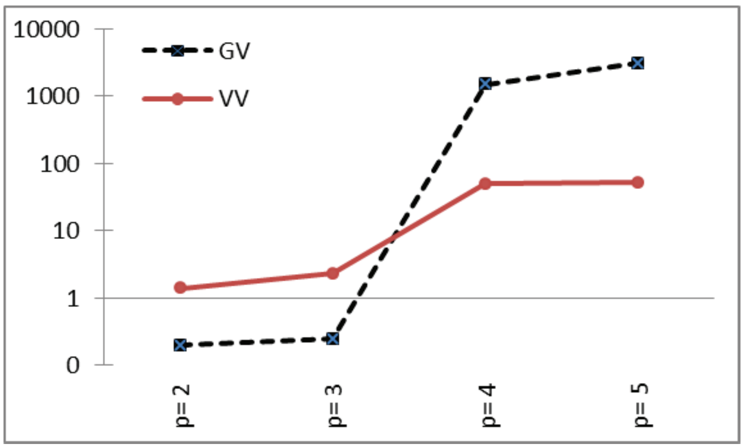

FIGURE 2. Computational time of sample size, *n=20*

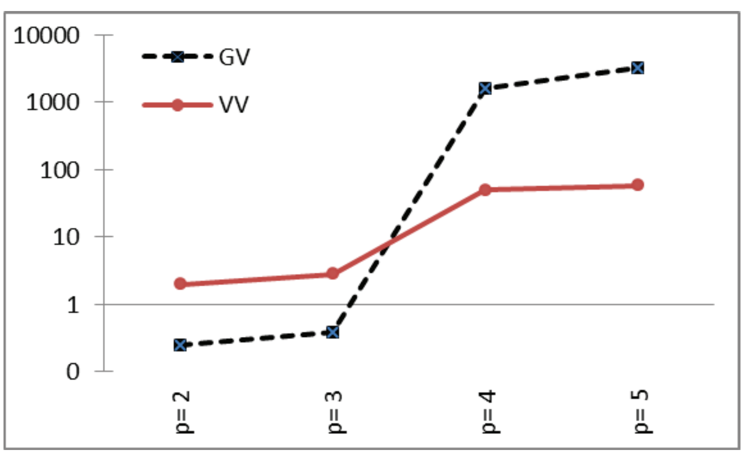

FIGURE 3. Computational time of sample size, *n=30*

For each different sample size and different number of variables, we calculated both statistics, and the running times (in microsecond) were recorded. Figure 1 – Figure 3 presents the running time mentioned. Interestingly, the plot of the running time increases directly to the number of variables. Based on the figures, GV performs better than VV when $p= 2$ and $p=3$. However, as p increases to 4 and 5, the time consumed by GV is higher than VV. The similar trend was also discovered for sample size of 10, 20 and 30.

CONCLUSION

From this analysis, we can conclude that VV can operate with both non-singular and singular covariance matrix. However, the computation of GV will be more cumbersome when dealing with high dimension data set. In terms of total number of operations needed and computation time, VV seems to be superior to GV.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support received from the Government of Malaysia, Universiti Utara Malaysia and Universiti Malaysia Pahang.

REFERENCES

1. M.A. Djauhariand I. Mohamad, *SouthAfrican Journal of Industrial Engineering*, **21(2)**, 207-215 (2010).
2. S. J. Arnold, R. Burger, P. A. Holenhole, C. A. Beverley, and A. G. Jones, *Evolution,* **62**,2451-2461 (2008)
3. K. Emil, *BMC Medical Research Methodology*, **2**, 1-14,(2002)
4. F. Eroukhmanoff, *Evol*. **36**, 323-326(2009).
5. J. Schafer, and K. Strimmer, *Statistical Applications in Genetics and Molecular Biology*, **4 (1)**, 1-29(2005).
6. S. J. Steppan, P. C. Phillips and D. Houle, *Trends Ecol.* **17**, 320-327(2002).
7. F. J. Finch and S. G. West, *Journal of Research in Personality*, **31**, 439-485(1997).
8. F.Schindler, *Journal of Property Investment and Finance*, **27 (6)**, 579-592 (2009).
9. L. Stephen, *Journal of Property Investment & Finance*, **24(5)**, 434-442 (2006).
10. P. M. A. Eichholtz, *Journal of Real Estate Research*, **11 (2)**, 149-158(1996).
11. K.C.B. Roes and D. Dorr,*International Journal of Quality Science*, **2 (3)**, 149-166(1997).
12. M. Wood, *International Journal of Service Industry Management*, **5 (4)**, 53-68(1994).
13. M. A.Djauhari, M. Mashuri and D. E. Herwindiati, *Communications in Statistics-Theory andMethod*, **37**, 1742-1754(2008).