

Department of Economics

On Testing for Diagonality of Large Dimensional Covariance Matrices

George Kapetanios

Working Paper No. 526

October 2004

ISSN 1473-0278



On Testing for Diagonality of Large Dimensional Covariance Matrices

George Kapetanios*
Queen Mary, University of London

October 13, 2004

Abstract

Datasets in a variety of disciplines require methods where both the sample size and the dataset dimensionality are allowed to be large. This framework is drastically different from the classical asymptotic framework where the number of observations is allowed to be large but the dimensionality of the dataset remains fixed. This paper proposes a new test of diagonality for large dimensional covariance matrices. The test is based on the work of John (1971) and Ledoit and Wolf (2002) among others. The theoretical properties of the test are discussed. A Monte Carlo study of the small sample properties of the test indicate that it behaves well under the null hypothesis and has superior power properties compared to an existing test of diagonality for large datasets.

Keywords: Panel Data, Large Sample Covariance Matrix, Maximum Eigenvalue
JEL Codes: C12, C15, C23

1 Introduction

The emergence of large multivariate datasets in a variety of disciplines necessitates the use of statistical methods appropriate for an asymptotic framework where both the number of variables and the number of observations tend to infinity. Examples of such large datasets emerge in disciplines as diverse as finance and molecular biology.

A problem that arises frequently in statistical analysis of multivariate datasets concerns the diagonality of covariance matrices. For example, the analysis of panel data in econometrics usually assumes that the error terms of the regression models of every unit in the panel are not contemporaneously correlated. Whereas, the problem of testing whether a covariance matrix is either equal or proportional to the identity matrix has received some attention in the case where the dimensionality of the covariance matrix is large (see Ledoit

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

and Wolf (2002)) , the problem of testing for diagonality has not. Clearly this is an equally relevant problem for empirical work. Only under quite restrictive conditions is it reasonable to hypothesise that every series in a dataset has the same variance.

This note aims to address this issue. We adapt the framework adopted by Ledoit and Wolf (2002) to construct a new test of diagonality of large dimensional covariance matrices. The note is organised as follows: Section 2 discusses the new test. Section 3 presents results on a Monte Carlo study. Finally, Section 4 concludes.

2 Theory

Let $X_N = [x_{ij}]$ be a $T \times N$ matrix of i.i.d. random variables that are normally distributed with mean vector μ_N and covariance matrix $\Sigma_N = [\sigma_{ij}]$. Let $\lambda_{1,N}, \dots, \lambda_{N,N}$ denote the eigenvalues of Σ_N . The dimensionality, N , and sample size, T , are increasing integer functions of some index k such that $\lim_{k \rightarrow \infty} N(k) = \infty$, $\lim_{k \rightarrow \infty} T(k) = \infty$ and $\lim_{k \rightarrow \infty} N(k)/T(k) = c \in (0, \infty)$. Dependence of N and T on k will be suppressed in what follows. It is assumed that the average eigenvalue given by $\alpha = 1/N \sum_{i=1}^N \lambda_{i,N}$ and the variance of the eigenvalues $\delta^2 = \sum_{i=1}^N (\lambda_{i,N} - \alpha)^2 / N$ are independent of the index k , where, $\alpha > 0$. Further, it is assumed that

$$\sum_{i=1}^N \frac{\lambda_{i,N}^j}{N} \rightarrow d_j < \infty$$

for $j = 3, 4$. Corresponding to the population covariance, the sample covariance matrix is denoted by $S_N = [s_{ij,N}]$ whose elements are given by

$$s_{ij,N} = \frac{1}{T} \sum_{i=1}^T (x_{ij} - m_{j,N})^2$$

where $m_{j,N} = 1/T \sum_{i=1}^T x_{ij}$. The test we suggest is derived from a test of sphericity of Σ_N as discussed in Muirhead (1982), Anderson (1982), John (1971) and Ledoit and Wolf (2002). The test statistic for the sphericity test is given by

$$U = \frac{1}{N} \text{tr} \left[\left(\frac{S_N}{(1/N) \text{tr}(S_N)} - I \right)^2 \right]$$

Using results on the eigenvalues of large dimensional covariance matrices Ledoit and Wolf (2002) show that

$$TU - N \xrightarrow{d} N(1, 4) \tag{1}$$

as N, T jointly tend to infinity.

As we mentioned in the introduction, testing for sphericity may be restrictive in some empirical applications where there is no reason to hypothesise that different variables have equal variances. To circumvent this problem we suggest a new hypothesis given by

$$H_0 : \sigma_{ij} = 0, \quad i \neq j$$

A first step towards a test of this hypothesis is to normalise all variables such that they have unit variance by dividing every observation by the estimated standard deviation of the variable to which it pertains. This amounts to restricting all the diagonal elements of S_N to be equal to one. Then, the test based on the test statistic U is applied on the transformed data. We denote the test statistic when applied to transformed data as U^* . In what follows we analyse the asymptotic distribution of U^* .

It is easy to see that

$$U = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{s_{ij}}{\frac{1}{N} \sum_{i=1}^N s_{ii}} - \delta_{ij} \right)^2 \quad (2)$$

where $\delta_{ij} = I(\{i = j\})$ and $I(\{.\})$ is the indicator function taking the value 1 if the event $\{.\}$ occurs and zero otherwise. Normalising the data to have variance equal to 1, does not affect the asymptotic behaviour of the terms in the summation in (2) for which $i \neq j$. But, clearly all terms for which $i = j$, will be equal to zero. We need to determine the effect of the lack of those terms to the behaviour of U^* . To do this we examine

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{s_{ii}}{\frac{1}{N} \sum_{i=1}^N s_{ii}} - 1 \right)^2$$

where s_{ii} are obtained from unnormalised data assuming that $\sigma_{ii} = 1$. We first note that

$$\left(\frac{s_{ii}}{\frac{1}{N} \sum_{i=1}^N s_{ii}} - 1 \right) - (s_{ii} - 1) = o_p(1)$$

and hence we can examine

$$\frac{1}{N} \sum_{i=1}^N (s_{ii} - 1)^2$$

instead. Under the assumption of normality we know that

$$\sqrt{T}(s_{ii} - 1) \xrightarrow{d} N(0, 2)$$

Thus, by the law of large number for i.i.d. observations we have that

$$\frac{T}{N} \sum_{i=1}^N (s_{ii} - 1)^2 \xrightarrow{p} 2$$

Table 1: Experiments A

N/T	50	100	200	500	1000	50	100	200	500	1000
	U^*					CD				
50	0.04	0.04	0.05	0.04	0.04	0.04	0.05	0.05	0.06	0.05
100	0.04	0.03	0.04	0.05	0.06	0.05	0.06	0.05	0.05	0.06
200	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.05	0.06	0.06

By Proposition 3 of Ledoit and Wolf (2002) we know that (1) holds. Hence, it also follows that

$$TU^* - N \xrightarrow{d} N(-1, 4)$$

A few remarks are in order. Firstly, we note that only the expectation of the asymptotic distribution changes and not its variance. Secondly, we note that the denominator of the main argument of U , given by $(1/N)tr(S_N)$ disappears after the transformation. This result in the test statistic being the same as the statistic V in Ledoit and Wolf (2002). Ledoit and Wolf (2002) diagnose a power problem for this statistic in the case where $a = \frac{1-c}{1+c}$. However, this is not problem in our case since in our case $\alpha = 1$, and hence the problem occurs only for $c = 0$. which is not a permissible value in the setup we consider. We examine the small sample properties of this new test in the next section.

3 Monte Carlo Study

3.1 Monte Carlo Design

We wish to investigate the size and power of the new test. For the size experiments we let x_{ij} be i.i.d. $N(0, 1)$ random variables independent across N and T . We consider a variety of values for N, T . These are $N = 50, 100, 200$ and $T = 50, 100, 200, 500, 1000$. We consider all possible combinations of these values. We put more emphasis on larger values of T as these might be more relevant for some applications in macroeconomic panel data where the problem of diagonality has been repeatedly encountered, (see e.g., Chang (2002) for work that addresses the problem in the context of panel unit root test). These are referred to as Experiments A. Results are reported in Table 1. As is clear, the new test is well behaved under the null hypothesis.

A major issue concerns the power of the test. Given the fact that U is the locally most powerful invariance test for sphericity as proven by Ledoit and Wolf (2002) one would expect good power properties for this test. As a comparison, we consider the test proposed by Pesaran (2004) for testing for diagonality. The test statistic and its asymptotic distribution

for this test are given by

$$\sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right) \xrightarrow{d} N(0, 1)$$

where $\hat{\rho}_{ij}$ is the estimated correlation between the i -th and j -th variables. This test will be referred to as CD . The above asymptotic result is obtained under joint N, T asymptotics. Note that this test is a modification of the test developed by Breusch and Pagan (1980) which used squares of $\hat{\rho}_{ij}$ instead. The behaviour of that test, however, had been explored only for T asymptotics keeping N fixed.

Initial experimentation suggests that both tests are quite powerful and, hence, interest focuses on modest departures from the null hypothesis. We consider three different such departures. All consider tridiagonal forms for Σ_N . The first set of power experiments referred to as Experiments B, specify the off-diagonal elements of Σ_N to be equal to $\sigma_{ii-1} = \sigma_{ii+1} = 0.05, 0.1, \dots, 0.5$. The second set of power experiments referred to as Experiments C, specify the off-diagonal elements of Σ_N to be given by $\sigma_{ii-1} = \sigma_{ii+1} \sim U(0, \sigma)$ where $\sigma = 0.05, 0.1, \dots, 0.5$. Finally, the third set of power experiments referred to as Experiments D, specify the off-diagonal elements of Σ_N to be given by $\sigma_{ii-1} = \sigma_{ii+1} \sim U(-\sigma, \sigma)$ where $\sigma = 0.05, 0.1, \dots, 0.5$. This final set of experiments consider the rather more realistic case where correlations exist between variables but are on average close to zero. Results for these three sets of experiments are presented in Tables 2-4.

3.2 Monte Carlo Results

Results make interesting reading. As we noted earlier, both tests have well behaved behaviour under the null hypothesis. Moving on to the more interesting issue of power, we see a number of patterns emerging. If we were to rank the sets of experiments in terms of the extent of their departure from the null hypothesis, we would rank experiments B and D as more distant than experiments C. Firstly, we note that the U^* test obeys this ranking being more powerful for experiments D, followed by B and then by C. The CD test is least powerful for experiments D. The reason for this is that the CD test is not well equipped to pick up departures from the null hypothesis where the correlations between variables maybe different from zero but are close to zero on average. This is because the test is based on $\hat{\rho}_{ij}$ rather than, say, $\hat{\rho}_{ij}^2$. However, this choice is understandable given the problematic behaviour of a test based on $\hat{\rho}_{ij}^2$ as N grows, as noted by Pesaran (2004).

Moving on to a comparison between the two tests, we see that in a majority of cases

the U^* test dominates. There are instances where the CD test has higher power, especially when σ and T are small, but these cases are, very much, in the minority.

4 Conclusion

Datasets in a variety of disciplines require methods where both the sample size and the dataset dimensionality are allowed to be large. This framework is drastically different from the classical asymptotic framework where the number of observations is allowed to be large but the dimensionality of the dataset remains fixed.

This paper proposes a new test of diagonality for large dimensional covariance matrices. The test is based on the work of John (1971) and Ledoit and Wolf (2002) among others. The theoretical properties of the test are discussed. A Monte Carlo study of the small sample properties of the test indicate that it behaves well under the null hypothesis and has superior power properties compared to an existing test of diagonality for large datasets.

References

- ANDERSON, T. W. (1982): *An Introduction Multivariate Statistical Analysis*. Wiley.
- BREUSCH, T. S., AND A. R. PAGAN (1980): “The LM Test and its Applications to Model Specifications in Econometrics,” *Review of Economic Studies*, 47, 239–253.
- CHANG, Y. (2002): “Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependence,” *Journal of Econometrics*, 110, 261–292.
- JOHN, S. (1971): “Some Optimal Multivariate Tests,” *Biometrika*, 58, 123–127.
- LEDOIT, O., AND M. WOLF (2002): “Some Hypotheses Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size,” *Annals of Statistics*, 30(4), 1081–1102.
- MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. Wiley.
- PESARAN, M. H. (2004): “General Diagnostic Tests for Cross-Sectional Dependence in Panels,” *Mimeo, University of Cambridge*.

Table 2: Experiments B

σ	N/T	U^*					CD				
		50	100	200	500	1000	50	100	200	500	1000
0.05	50	0.03	0.05	0.09	0.21	0.67	0.10	0.13	0.16	0.34	0.55
	100	0.05	0.05	0.07	0.23	0.66	0.10	0.12	0.18	0.33	0.58
	200	0.05	0.06	0.09	0.23	0.69	0.10	0.13	0.23	0.36	0.55
0.10	50	0.08	0.18	0.47	1.00	1.00	0.19	0.32	0.47	0.83	0.97
	100	0.08	0.17	0.49	1.00	1.00	0.23	0.30	0.46	0.82	0.98
	200	0.07	0.16	0.51	1.00	1.00	0.19	0.30	0.50	0.84	0.99
0.15	50	0.19	0.52	0.97	1.00	1.00	0.32	0.52	0.76	0.99	1.00
	100	0.18	0.56	0.99	1.00	1.00	0.32	0.54	0.76	0.98	1.00
	200	0.18	0.58	0.99	1.00	1.00	0.33	0.53	0.77	0.98	1.00
0.20	50	0.49	0.96	1.00	1.00	1.00	0.45	0.71	0.94	1.00	1.00
	100	0.45	0.96	1.00	1.00	1.00	0.48	0.74	0.94	1.00	1.00
	200	0.47	0.96	1.00	1.00	1.00	0.49	0.73	0.94	1.00	1.00
0.25	50	0.79	1.00	1.00	1.00	1.00	0.62	0.84	0.98	1.00	1.00
	100	0.81	1.00	1.00	1.00	1.00	0.58	0.84	0.98	1.00	1.00
	200	0.82	1.00	1.00	1.00	1.00	0.61	0.86	0.99	1.00	1.00
0.30	50	0.97	1.00	1.00	1.00	1.00	0.72	0.92	1.00	1.00	1.00
	100	0.97	1.00	1.00	1.00	1.00	0.70	0.92	1.00	1.00	1.00
	200	0.97	1.00	1.00	1.00	1.00	0.70	0.92	1.00	1.00	1.00
0.35	50	0.99	1.00	1.00	1.00	1.00	0.79	0.96	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.80	0.97	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.81	0.96	1.00	1.00	1.00
0.40	50	1.00	1.00	1.00	1.00	1.00	0.85	0.99	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.86	0.98	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.86	0.98	1.00	1.00	1.00
0.45	50	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.92	0.99	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00
0.50	50	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00

Table 3: Experiments C

σ	N/T	50	100	200	500	1000	50	100	200	500	1000
		U^*					CD				
0.05	50	0.04	0.04	0.06	0.08	0.13	0.07	0.08	0.09	0.13	0.20
	100	0.04	0.03	0.05	0.07	0.13	0.08	0.08	0.10	0.14	0.21
	200	0.04	0.05	0.04	0.07	0.13	0.06	0.08	0.09	0.13	0.20
0.10	50	0.04	0.06	0.10	0.37	0.86	0.09	0.14	0.17	0.34	0.54
	100	0.05	0.07	0.10	0.39	0.89	0.10	0.12	0.19	0.37	0.56
	200	0.05	0.06	0.10	0.35	0.90	0.10	0.12	0.19	0.38	0.59
0.15	50	0.07	0.11	0.30	0.91	1.00	0.14	0.19	0.33	0.61	0.87
	100	0.08	0.11	0.33	0.94	1.00	0.13	0.19	0.33	0.64	0.89
	200	0.08	0.12	0.31	0.95	1.00	0.16	0.21	0.33	0.61	0.87
0.20	50	0.10	0.28	0.68	1.00	1.00	0.20	0.31	0.48	0.81	0.97
	100	0.10	0.24	0.70	1.00	1.00	0.20	0.29	0.51	0.82	0.98
	200	0.09	0.23	0.71	1.00	1.00	0.20	0.31	0.49	0.85	0.97
0.25	50	0.16	0.47	0.95	1.00	1.00	0.23	0.41	0.63	0.93	1.00
	100	0.16	0.50	0.96	1.00	1.00	0.28	0.41	0.63	0.94	1.00
	200	0.16	0.50	0.98	1.00	1.00	0.26	0.41	0.63	0.94	1.00
0.30	50	0.29	0.74	1.00	1.00	1.00	0.30	0.51	0.76	0.98	1.00
	100	0.30	0.79	1.00	1.00	1.00	0.34	0.53	0.76	0.98	1.00
	200	0.32	0.82	1.00	1.00	1.00	0.34	0.53	0.78	0.99	1.00
0.35	50	0.45	0.94	1.00	1.00	1.00	0.40	0.60	0.86	1.00	1.00
	100	0.49	0.96	1.00	1.00	1.00	0.38	0.61	0.87	0.99	1.00
	200	0.49	0.95	1.00	1.00	1.00	0.38	0.64	0.85	0.99	1.00
0.40	50	0.68	0.99	1.00	1.00	1.00	0.45	0.70	0.91	1.00	1.00
	100	0.67	0.99	1.00	1.00	1.00	0.49	0.74	0.92	1.00	1.00
	200	0.70	0.99	1.00	1.00	1.00	0.49	0.70	0.94	1.00	1.00
0.45	50	0.80	1.00	1.00	1.00	1.00	0.51	0.78	0.96	1.00	1.00
	100	0.85	1.00	1.00	1.00	1.00	0.54	0.78	0.96	1.00	1.00
	200	0.87	1.00	1.00	1.00	1.00	0.55	0.79	0.96	1.00	1.00
0.50	50	0.94	1.00	1.00	1.00	1.00	0.60	0.85	0.97	1.00	1.00
	100	0.94	1.00	1.00	1.00	1.00	0.60	0.82	0.98	1.00	1.00
	200	0.95	1.00	1.00	1.00	1.00	0.64	0.83	0.98	1.00	1.00

Table 4: Experiments D

σ	N/T	50	100	200	500	1000	50	100	200	500	1000
		U^*					CD				
0.05	50	0.05	0.08	0.22	0.75	1.00	0.13	0.20	0.31	0.62	0.88
	100	0.06	0.08	0.19	0.81	1.00	0.15	0.20	0.35	0.62	0.88
	200	0.07	0.10	0.21	0.80	1.00	0.15	0.21	0.33	0.64	0.88
0.10	50	0.20	0.55	0.98	1.00	1.00	0.32	0.51	0.76	0.99	1.00
	100	0.22	0.61	0.99	1.00	1.00	0.32	0.53	0.78	0.99	1.00
	200	0.23	0.62	0.99	1.00	1.00	0.35	0.52	0.77	0.99	1.00
0.15	50	0.63	0.99	1.00	1.00	1.00	0.53	0.79	0.95	1.00	1.00
	100	0.69	1.00	1.00	1.00	1.00	0.52	0.79	0.95	1.00	1.00
	200	0.66	1.00	1.00	1.00	1.00	0.54	0.79	0.97	1.00	1.00
0.20	50	0.97	1.00	1.00	1.00	1.00	0.70	0.92	0.99	1.00	1.00
	100	0.98	1.00	1.00	1.00	1.00	0.73	0.92	0.99	1.00	1.00
	200	0.99	1.00	1.00	1.00	1.00	0.72	0.93	1.00	1.00	1.00
0.25	50	1.00	1.00	1.00	1.00	1.00	0.84	0.98	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.86	0.98	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.83	0.98	1.00	1.00	1.00
0.30	50	1.00	1.00	1.00	1.00	1.00	0.90	0.99	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.90	0.99	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.91	0.99	1.00	1.00	1.00
0.35	50	1.00	1.00	1.00	1.00	1.00	0.89	0.98	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.88	0.98	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.89	0.99	1.00	1.00	1.00
0.40	50	1.00	1.00	1.00	1.00	1.00	0.84	0.98	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.84	0.98	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.84	0.98	1.00	1.00	1.00
0.45	50	1.00	1.00	1.00	1.00	1.00	0.78	0.97	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.81	0.94	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.76	0.97	1.00	1.00	1.00
0.50	50	1.00	1.00	1.00	1.00	1.00	0.68	0.91	0.99	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	0.69	0.89	0.99	1.00	1.00
	200	1.00	1.00	1.00	1.00	1.00	0.68	0.89	0.99	1.00	1.00

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2004 George Kapetanios
All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**