Advances and Applications in Statistics
Volume 22, Number 2, 2011, Pages 119-127
Published Online: August 10, 2011
This paper is available online at http://pphmj.com/journals/adas.htm

A LINEAR MODEL TEST OF SUFFICIENCY FOR NON-ORTHOGONALITY OF MATRICES

FATHALI FIROOZI, DONALD LIEN,
SAEID MAHDAVI and LYNDA DE LA VINA

Department of Economics
University of Texas at San Antonio
U.S.A.
e-mail: ffiroozi@utsa.edu

Abstract

Matrix orthogonalities appear in a number of mathematical and statistical contexts. This study presents a sufficient condition for non-orthogonality of general matrices in stochastic contexts and provides a linear model procedure for testing the condition.

1. Introduction

Let X_1 be an $(n \times K)$ matrix and X_2 be an $(n \times L)$ matrix. In the sense of orthogonality or perpendicularity in the Euclidean space $\mathbb{R}^n$, the condition $X_1' X_2 = 0$ is equivalent to the statement that every column of X_1 is orthogonal to every column of X_2 (Halmos [6]). The general matrix orthogonality $X_1' X_2 = 0$ plays a central role in many statistical contexts, including the following. Let $y = Z\beta + \varepsilon$ be a standard regression model with more than one regressor, so y is the $(n \times 1)$ vector of observations on the dependent variable, Z is the $(n \times K)$ matrix of observations on the regressors, β is the $(K \times 1)$ vector of coefficients, and ε is the $(n \times 1)$ vector of stochastic errors. Partition matrix Z into two matrices X_1 and X_2

2010 Mathematics Subject Classification: 15, 62.

Keywords and phrases: partial correlation, regression, hypothesis testing.

Received March 29, 2011

so that $Z = [X_1 \ X_2]$ and the model $y = Z\beta + \varepsilon$ is decomposed into $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where β_1 and β_2 form the corresponding partition of β, i.e., $\beta = [\beta_1', \beta_2']'$. It can then be shown that the orthogonality condition $X_1'X_2 = 0$ is sufficient for the absence of "omitted variable bias" in the regression $y = X_2\beta_2 + \varepsilon$, i.e., the least squares estimation of $y = X_2\beta_2 + \varepsilon$ generates the same estimate for β_2 as the least squares estimation of $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ (Davidson and MacKinnon [2] and Greene [5]). Another application is in the context of instrumental variables. If X_2 is correlated with the error term ε, then X_2 is replaced with another set X_3, where the instrument X_3 satisfies the exogeniety property $X_3'\varepsilon = 0$ and the relevance property $X_3'X_2 \neq 0$. The instrument is said to be weak if $X_3'X_2$ is near zero. The implications of weak instruments have received attention in the recent literature (Godfrey [3] and Stock et al. [10]). Other versions of the general orthogonality $X_1'X_2 = 0$ appear in the study of independence and other statistical contexts (Solomon and Taylor [9], Graybill [4], Mendenhall and Scheaffer [7] and Schott [8]).

Although matrix orthogonalities appear in several stochastic contexts, the literature lacks an explicit and simple procedure for testing the validity of general matrix orthogonalities in stochastic contexts. Such a test is important since non-experimental sample data rarely satisfy an orthogonality condition exactly even if the true features satisfy the orthogonality condition. The issue also arises in experimental cases that are subject to measurement error. The underlying objective of this study is to propose a readily accessible routine for checking if two matrices of arbitrary sizes drawn from a stochastic context are in fact non-orthogonal or the deviation from orthogonality could be only due to a sampling error. The study first briefly reviews some of the rather known but core statistical properties of matrix orthogonality in Section 2. A sufficient condition for non-orthogonality is established in Section 3 and a procedure for testing the condition is proposed in Section 4. Section 5 concludes by highlighting some of the distinctive features of the proposed test in relation to those in the existing literature.

2. Orthogonality and Correlation

Let x and y be two $(n \times 1)$ vectors that represent sample observations on two variables, $\bar{x} = \left(\sum x_i\right)/n$ and $\bar{y} = \left(\sum y_i\right)/n$ be the corresponding mean values, and $(x - \bar{x})$ and $(y - \bar{y})$ be the corresponding vectors of deviations from the mean. The simple correlation between the two variables *x* and *y* is defined by $cor(x, y) = cov(x, y)/[var(x)var(y)]^{1/2}$, where $cov(x, y) = (x - \bar{x})'(y - \bar{y})/n$ is the covariance of *x* and *y* and $var(x) = cov(x, x)$ is the variance of *x*. Then it follows from the stated definitions that the orthogonality $(x - \bar{x})'(y - \bar{y}) = 0$ is equivalent to the absence of simple correlation between *x* and *y*. The following result is an immediate consequence of the stated definitions.

Remark 1. Let X_1 be an $(n \times K)$ matrix and X_2 be an $(n \times L)$ matrix. Suppose the columns of either X_1 or X_2 have zero sums. Then the orthogonality $X_1'X_2 = 0$ is equivalent to the absence of simple correlation between every column of X_1 and every column of X_2.

Consider the regression

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots \beta_{K+1} x_{K+1} + \varepsilon \tag{1}$$

with $K \geq 2$. Select a regressor x_j and refer to the rest of the regressors $\{x_2, ..., x_{j-1}, x_{j+1}, ..., x_{K+1}\}$ as the "other regressors". The partial correlation between *y* and x_j when the other regressors are partialed out is denoted by *parcor* (y, x_j) and computed as follows. Let e_y be the residuals when *y* is regressed on the other regressors and e_j be the residuals when x_j is regressed on the other regressors. The stated partial correlation is then computed as the simple correlation between e_j and e_y, i.e., $parcor(y, x_j) = cor(e_y, e_j)$. One of the central results regarding partial correlation provides an equivalent and useful computational statement for $parcor(y, x_j)$ is as follows. In the regression specified in (1) with a sample of size *n*, let $\tilde{b}_j$ be the least squares estimate of β_j, $SE(\tilde{b}_j)$ be its estimated standard error, $t_j = \tilde{b}_j / SE(\tilde{b}_j)$ be its usual *t*-statistic, and $D = n - K - 1$. It can then be shown that (Graybill [4] and Greene [5])

$$parcor(y,\ x_j) = \frac{t_j}{\sqrt{t_j^2 + D}} = \frac{\tilde{b}_j}{\sqrt{\tilde{b}_j^2 + D \cdot [SE(\tilde{b}_j)]^2}}. \tag{2}$$

The result shows that $parcor(y,\ x_j)$ and $\tilde{b}_j$ share the same sign and $parcor$ $(y,\ x_j) = 0$ if and only $\tilde{b}_j = 0.$ The following result can readily be shown.

Remark 2. Let X_1 be an $(n \times K)$ matrix and X_2 be an $(n \times L)$ matrix. Suppose the columns of either X_1 or X_2 have zero sums. Then the orthogonality $X_1' X_2 = 0$ is equivalent to the absence of partial correlation between every column of X_1 and every column of X_2 when any subset of the remaining columns in X_1 and X_2 is partialed out.

Remark 3. Let M be the $(n \times n)$ mean deviation matrix defined by $M = I - \frac{1}{n}\mathbf{11}',$ where I is the $(n \times n)$ identity matrix and $\mathbf{1}$ is the $(n \times 1)$ vector of 1's. It can readily be verified that M is symmetric, idempotent and transforms vectors into mean deviations. Thus, if v is an $(n \times 1)$ vector, then $Mv = v - \bar{v}\mathbf{1},$ where $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i,$ hence vector Mv has zero sum. Also, if $W = [w_1, ..., w_K]$ is an $(n \times K)$ matrix with its jth column denoted by $w_j,$ the mean deviation transformation of W defined by $MW = [Mw_1, ..., Mw_K]$ satisfies the zero column sum condition.

Remark 4. The zero sum condition stated in Remarks 1 and 2 above is often either present or can be obtained by mean deviation transformations. Consider the regression $y = [\mathbf{1}\ X][\alpha\ \beta']' + \varepsilon,$ where β is the slope vector. The requirement of zero column sum in $[y\ X]$ could be obtained by transforming y and X into deviations from the column means, that is, transforming y into My and X into MX, where M is the mean deviation matrix as defined in Remark 3 above. The transformation then reduces $y = [\mathbf{1}\ X][\alpha\beta']' + \varepsilon$ to $My = MX\beta + \varepsilon',$ so the constant term is eliminated. Alternatively, the mean deviation transformation could be applied only to X. These two transformations generate the required zero sum condition and, more importantly, the least squares estimate $\tilde{b}$ of the slope vector β in the stated regression $y = [\mathbf{1}\ X][\alpha\ \beta']' + \varepsilon$ is invariant under these two transformations. However, transforming only y into its mean deviations generates the required zero sum in y but will revise $\tilde{b}$.

3. A Sufficient Condition for Non-orthogonality

Assumption. Throughout this and the following sections, it is assumed that the mean deviation transformation stated in Remark 4 above has been applied in advance so that all of the matrices satisfy the zero sum condition stated in Remarks 1 and 2 above.

The intended condition emerges from the following theorem.

Theorem 5. *Consider the regression model* $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, *where y is* $(n \times 1)$, X_1 *is* $(n \times K)$, X_2 *is* $(n \times L)$, *and the* $(n \times 1)$ *vector of error terms* (ε) *satisfies the usual assumptions of the classical regression model. In particular, adopt the following assumptions*: (i) $[X_1\ X_2]$ *has full column rank*, (ii) *there is no endogeniety problem so that* $E[[X_1\ X_2]'\varepsilon] = 0$, *and* (iii) $E(X_1'X_1) = P_1$ *and* $E(X_2'X_2) = P_2$, *where* P_1 *and* P_2 *are positive definite matrices. Then* $\beta_2 = 0$ *if* $E[y\ X_1]'X_2 = 0$.

Proof. Premultiplying $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ by X_2' and then taking expectation lead to

$$E(X_2'y) = E(X_2'X_1)\beta_1 + E(X_2'X_2)\beta_2 + E(X_2'\varepsilon).$$

It follows then from the assumptions of the theorem that

$$\beta_2 = P_2^{-1}[E(X_2'y) - E(X_2'X_1)\beta_1]. \tag{3}$$

Hence, if $E(X_2'y) = 0$ and $E(X_2'X_1) = 0$, which are equivalent to $E[y\ X_1]'X_2 = 0$, then $\beta_2 = 0$. □

Remark 6. It follows from Theorem 5 that the nonzero condition $\beta_2 \neq 0$ is a sufficient condition for the non-orthogonality $E[y\ X_1]'X_2 \neq 0$. The condition is not necessary since $\beta_2 = 0$ does not imply $E[y\ X_1]'X_2 = 0$.

4. A Test of Non-orthogonality

Theorem 5 and Remark 6 suggest a regression-based and testable sufficient condition for non-orthogonality of matrices. To elaborate, consider the problem of testing the non-orthogonality $E(Z'X_2) \neq 0$, where Z is $(n \times K)$, X_2 is $(n \times L)$, and $[Z \;\; X_2]$ has full column rank. Let y be an arbitrary column of Z, and X_1 be the $[n \times (K-1)]$ matrix that contains the remaining columns in Z. There is no loss in the generality of $Z'X_2 = 0$ if Z is stated as $Z = [y \;\; X_1]$. Consider then the regression $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where the error term (ε) satisfies the usual assumptions of the classical regression. It follows from Theorem 5 and Remark 6 that the non-orthogonality condition $E(Z'X_2) \neq 0$ is implied by the nonzero condition $\beta_2 \neq 0$. Testing the latter is the standard F-test of linear equality restrictions on the regression parameters under the classical assumptions (Graybill [4], Mendenhall and Scheaffer [7] and Greene [4]). Thus, rejection of the null hypothesis H_0 in

$$\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

on the stated regression $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ verifies the non-orthogonality $E(Z'X_2) \neq 0$.

The stated result is independent of the decomposition $Z = [y \;\; X_1]$ of matrix Z. To show this independence, suppose $Z = [\hat{y} \;\; \hat{X}_1]$ is another decomposition of matrix Z. The orthogonality $E(Z'X_2) = 0$, then implies $E(X_2'\hat{y}) = 0$ and $E(X_2'\hat{X}_1) = 0$. Thus, by the result in (3) with y replace by $\hat{y}$, and X_1 replaced by $\hat{X}_1$, the implication $\beta_2 = 0$ continues to hold. Therefore, $\beta_2 \neq 0$ implies $E(Z'X_2) \neq 0$, regardless of the choice of column vector y from matrix Z to form the decomposition $Z = [y \;\; X_1]$ and regression $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$.

The procedure can be summarized as follows. Decompose the observed matrix $[Z \;\; X_2]$ into $[Z \;\; X_2] = [y \;\; X_1 \;\; X_2]$ as described above and let e be the standard

$(n \times 1)$ vector of residuals that result from the least squares estimation of the unrestricted model

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = [X_1 \ \ X_2]\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} + \varepsilon, \tag{4}$$

where β_1 and β_2 are the $[(K-1)\times 1]$ and $(L \times 1)$ parameter vectors. Thus $e = y - (X_1\tilde{b}_1 + X_2\tilde{b}_2)$, where $(\tilde{b}_1, \tilde{b}_2)$ is the least squares estimate of (β_1, β_2) in (4). Let e_* be the $(n \times 1)$ vector of residuals that emerge from the least squares estimation of the restricted model

$$y = X_1\beta_1 + \varepsilon. \tag{5}$$

Thus $e_* = y - X_1\hat{b}_1$, where $\hat{b}_1$ is the least squares estimate of β_1 in (5). The inequality $e'e \le e_*'e_*$, then holds and the F-statistic for the stated test is

$$F = \frac{d_2(e_*'e_* - e'e)}{d_1(e'e)}, \tag{6}$$

where $d_1 = L$ is the number of restrictions in $\beta_2 = 0$ and $d_2 = n - (K + L - 1)$ is the degree of freedom in the unrestricted regression. Under the null hypothesis, the sampling distribution of the stated F-statistic in (6) is the F-distribution $F(d_1, d_2)$. At a given type I error level α, the non-orthogonality $E(Z'X_2) \neq 0$ is supported if the computed F-statistic in (6) exceeds the cut-off value $F_\alpha(d_1, d_2)$; otherwise, the test is inconclusive with respect to the orthogonality of Z and X_2.

For small and moderate sizes of n and for typical values of α, the cut-off values $F_\alpha(d_1, d_2)$ are readily available from the standard statistical tables. For large values of n, the distribution $F(d_1, d_2)$ is approximated by its asymptotic version, which is the chi-squared distribution $\chi^2(d_1)$ with cut-off values $\chi^2_\alpha(d_1)$ that are also readily available from the standard tables for certain values of d_1 and α. There are also a number of statistical software packages that readily compute the stated statistic and cut-off values. While the stated F distribution requires the normality assumption for the regression error term (ε), the stated χ^2 distribution continues to hold asymptotically without the normality assumption (Greene [5]).

5. Concluding Remarks

This study suggested a regression-based sufficient condition for non-orthogonality of matrices and proposed a procedure for testing the condition. In relation to the existing literature, the area of multivariate analysis (Anderson [1]) suggests certain tests of independence of two sets of normal variates. The normality assumption adopted in this literature allows equivalence between independence and zero correlation as well as application of the maximum likelihood method. To briefly highlight some of the comparative features, let $X = [Z \;\; X_2]$ be a set of $(K + L)$ variates where Z contains K variates and X_2 contains L variates. Suppose X has a joint normal distribution with mean zero and the $[(K + L) \times (K + L)]$ symmetric covariance matrix Σ. Decompose Σ as

$$\Sigma = E(X'X) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11} = E(Z'Z)$ is $(K \times K)$ and $\Sigma_{22} = E(X_2'X_2)$ is $(L \times L)$. Given a random sample of n observations on $X = [Z \;\; X_2]$, it follows from $E(Z'X_2) = \Sigma_{12}$ that testing the orthogonality hypothesis $E(Z'X_2) = 0$ is equivalent to testing the independence hypothesis $\Sigma_{12} = 0$ under the stated normality. The multivariate literature (Anderson [1]) provides a likelihood ratio procedure for testing the latter hypothesis with a likelihood ratio and a sampling distribution that are both rather complicated for computation. The regression-based approach suggested in the present study is distinct in the sense that it not only presents a rather simple routine for the testing but also provides flexibility with respect to the distributional assumption adopted for the error term (ε) in the corresponding regression. Any of the widely available regression software packages can quickly perform the test suggested in the present study. The statistical properties of the proposed test are readily defined by those of the classical F-test for testing joint linear restrictions in the linear models. However, the stated advantages of the proposed test come at the cost of its limitation, namely, the proposed procedure tests only a sufficient condition for non-orthogonality of matrices and is inconclusive if the test fails to support the non-orthogonality. The proposed regression-based procedure is still significant in light of the fact that the literature lacks a general procedure that possesses the stated features and is void of the stated limitation.

References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 2003.

[2] R. Davidson and J. MacKinnon, Econometric Theory and Methods, Oxford University Press, 2004.

[3] L. Godfrey, Instrument relevance in multivariate linear models, Review of Economics and Statistics 81 (1999), 550-552.

[4] F. A. Graybill, Theory and Application of the Linear Model, Duxbury Press, 2000.

[5] W. Greene, Econometric Analysis, Prentice-Hall, 2008.

[6] P. R. Halmos, Finite-dimensional Vector Spaces, Springer, 1993.

[7] W. Mendenhall and R. Scheaffer, Mathematical Statistics with Applications, Duxbury Press, 2001.

[8] J. Schott, Matrix Analysis for Statistics, Wiley, 2005.

[9] P. Solomon and J. Taylor, Orthogonality and transformations in variance component models, Biometrika 86(2) (1999), 289-300.

[10] J. Stock, J. Wright and M. Yogo, A survey of weak instruments and weak identification in generalized method of moments, J. Business Econom. Statist. 20 (2002), 518-529.