Multicollinearity Robust QAP for Multiple-Regression

David Dekker*†
University of Nijmegen
David Krackhardt
Carnegie Mellon University
Tom Snijders
University of Groningen

March 30, 2003

ABSTRACT.The quadratic assignment procedures for inference on multiple-regression coefficients(MRQAP) has become popular in social network analysis. These tests have been developed to assess the sizes of a set of multiple-regression coefficients. However, research practitioners often use these tests to assess the size of *individual* multiple-regression coefficients. Although this might be a harmless extension, our our concern focuses on this practice under conditions of multicollinearity. In this paper we show analytically that different MRQAP-tests for *individual* parameter estimates are biased under multicollinearity. Subsequently, we propose a new MRQAP-test, which we call "semi-partialing" that is robust against multicollinearity. Extensive simulation results, as well as re-analysis of the classic Laumann-Marsden-Galaskiewicz(1978)data show the added value of this new "semi-partialing" method over the existing methods.

1 Introduction

In this paper we analyze the sensitivity of "multiple regression-quadratic assignment procedure (MRQAP)" tests for individual parameter estimates under conditions of multicollinearity. These tests are especially useful for models that endure problems with unspecified autocorrelation structures. As autocorrelation makes use of t-tests for inference on the parameter estimates inappropriate, MRQAP-tests aim to offer an alternative. In fields

*We would like to thank Richard Paap and Bas Donkers for useful comments.
†Correspondence: University of Nijmegen, P.O. Box 9108,6500HK, Nijmegen, The Netherlands, d.dekker@nsm.kun.nl

that use network data the MRQAP-test is often applied. Although its frequent use, some important questions about the MRQAP tests remain unanswered. One is whether the MRQAP-test is robust under multicollinearity conditions?

Here, we briefly elaborate on the background of MRQAP tests and analyze robustness of different MRQAP-tests against multicollinearity. Subsequently, we show through simulations the seriousness of multicollinearity conditions for different MRQAP-tests on linear model estimates. Also, we show empirical consequences through re-analysis of the classic Laumann-Marsden-Galaskiewicz (1978) data.

2 QAP: Non-Parametric Inference in Social Network Analysis

To assess the association between data on interpersonal relations, a family of inference tests have been developed based on the quadratic assignment procedure (QAP)(for overview see Hubert, 1987). The null-hypothesis of these tests is that the test-statistic of association equals the expected value of the test-statistic under the permutation distribution (Hubert, 1987). In other words, we test whether there is no similar pattern between the elements of the different variables.

Data on network variables typically is represented in the form of a square matrix. Based on random permutations of the rows and columns of one variable, the QAP generates a permutation distribution that is similar to the underlying distribution for which inference is drawn. Each permutation creates a random data set that is automorphic to the original data on that variable and, hence, when related to the other variable provides a random estimate of the relation between the two variables. As there are many ($n!$) possible permutations we may draw a random sample of these permutations is to generate a reference distribution (Hubert, 1987). A major advantage is that the test makes no assumptions about the distribution of parameters. Rather, the QAP constructs a reference distribution of random parameters that could have been derived from a dataset with the same structure as the dataset under evaluation.

2.1 MRQAP-test

Several authors propose to extend the bi-variate situation discussed above to the multivariate situation (e.g. Krackhardt, 1988; Hubert, 1987; Smouse, Long & Sokal, 1986). In this paragraph we discuss several possible QAP tests for multivariate regression.

Smouse et al. (1986) develop what we label as, the Y-permutation test (see also UCINET by Borgatti, Everett & Freeman, 2002). Permutation distributions for all coefficients of explanatory variables are derived from the permutation of the columns and rows of the dependent variable.

The partialing-test (Krackhardt, 1988; Smouse, et al., 1986) starts with estimation of the residual matrices of the partial regression of Y on Z, and the partial regression of X_i on Z, respectively E_{YZ} and E_{X_iZ}. Where Z represent all but the i^{th} explanatory variables. From regression analysis we know that the coefficient estimate of E_{YZ} on E_{X_iZ} equals the estimate of the i^{th} coefficient in the regression of Y on X. Where X represents the full set of explanatory variables. The $n!$ possible rearrangements of rows and columns of E_{YZ}(or E_{X_iZ}) provide a permutation distribution for the i^{th} variable's coefficient.

Another test we discern is the X-permutation test. In this test we permute one explanatory variable's (X_i) row's and column's. to generate a reference distribution for it's coefficient.

A fourth MRQAP-test, we call semi-partialing-permutation, combines X-permutation-test and the partialing-permutation test. In this test we replace X_i with E_{X_iZ}. Again, we know that the regression of Y on Z and E_{X_iZ} will give the same estimate of the i^{th} variable coefficient as the i^{th} variable coefficient in the regression of Y on X. (Note, that the other q coefficients will change). Permutations of the rows and columns of E_{X_iZ} allows to generate a permutation distribution.

2.2 Robustness against Multicollinearity

It can easily be shown that correlation between dependent variables affects LS multiple regression coefficient estimates. Also, coefficient estimates are dependent upon the correlation between the dependent and other explanatory variables. Here we refer to both types of correlation as multicollinearity.

Randomization of data by permutation of the rows and columns not only randomizes the correlation between the dependent and the variable of interest, also it disturbs the multicollinearity relations. The question is whether these disturbing effects would cause an analytic bias in MRQAP-tests discussed above.

MRQAP-tests, test the null-hypothesis, H0:

$$\beta_i = \sum_{k=1}^{n!} \frac{\pi_k(\beta_i)}{n!} = E(\pi_k(\beta_i))$$

in the model,

$$Y = f(\beta_i, X_i, \gamma, Z, E) \tag{1}$$

where Y, X_i, and E are $n \times n$ matrices, Z is a $n \times n \times q$ matrix, β_i is a real number, and γ is a $q \times 1$ vector of real numbers. The probability distributions of the matrices are permutation invariant, i.e., for all of the $n!$ possible permutations, π_k, of the n rows and columns of Y,

$$L(Y) = L(\pi_k(Y)) \tag{2}$$

where $k = 1, ..., n!$. Furthermore, under the null-hypothesis of no relation between X_i and Y, we can rewrite 1 as

$$Y = f(\gamma, Z, E) \tag{3}$$

We may express any estimator of β_i controlling for the effects of Z on Y as

$$\beta_i = \beta_i(Y, X_i, Z) = \beta_i(f(\gamma, Z, E), X_i, Z) \tag{4}$$

which is permutation invariant,

$$\beta_i(f(\gamma, Z, E), X_i, Z) = \beta_i(\pi_k(f(\gamma, Z, E)), \pi_k(X_i), \pi_k(Z)) \tag{5}$$

Now assume, no multicollinearity, i.e. $Cov(E, X_i) = 0$, and $\gamma = 0$, where E are the residuals from the LS regression in 1. Hence, under the null hypothesis and the permutation invariance of β_i, we have

$$L(\beta_i(f(\gamma, Z, E), X_i, Z)) = L(\beta_i(\pi_k(f(\gamma, Z, E)), X_i, Z) \tag{6}$$

$$L(\beta_i(f(\gamma, Z, E), X_i, Z)) = L(\beta_i(f(\gamma, Z, E), \pi_k(X_i), Z) \tag{7}$$

Showing that Y-permutation and X-permutation are unbiased under these conditions, respectively. Analogous, we can show that partialing and semi-partialing are unbiased tests.

However, so far we assume no multicollinearity. First, let us consider what happens if $Cov(E, X_i) \neq 0$, i.e. correlation between X_i and a linear combination of Z. In this case we can rewrite 3 as

$$Y = f(\gamma, Z, X_i, U) \tag{8}$$

where U is the residual matrix. However, substituting 8 in 4 under X-permutation results in

$$L(\beta_i(f(\gamma, Z, X_i, U), X_i, Z)) \neq L(\beta_i(f(\gamma, Z, X_i, U), \pi_k(X_i), Z) \tag{9}$$

Because, the right-hand-side is dependent on more non-random variables than the left-hand-side of 9.

Under Y-permutation in cases where there is a "third variable" effect Smouse et al. (1986) mention that problems may occur. It is easily shown that these problems are analog to those that occur under X-permutation when there is correlation between the independent variables. Also, we can show that the partialing method becomes biased when both multicollinearity conditions occur.

Under semi-partialing we can rewrite 4 as,

$$\beta_i = \beta_i(f(\gamma, Z, E), E_{X_iZ}, Z) \tag{10}$$

where E_{X_iZ} is the residual of the regression of X_i on Z. Because we know E_{X_iZ} is independent from Z by construction from 7 we know that semi-partialing will be unbiased under multicollinearity conditions.

3 Simulation Results and LMG-Data

The simulation results show the sensitivity of MRQAP-tests for multicollinearity conditions. Especially, results show that X-permutation creates an upward bias, while Y-permutation is downward biased. Partialing-permutation (Smouse et al., 1986; Krackhardt, 1988), only shows small upward bias in the simulation results. We further inquire the sources of this limited bias for the linear model. Consistent with our expectations formulated in section 3, the results for semi-partialing show robustness under all multicollinearity conditions. We will show how these different tests affect the results of the classic LMG-data.

4 Conclusion

As most studies assess type 1 error to determine effect size, X-permutation would have serious consequences for theory testing under conditions of multicollinearity. Also, Y-permutation-test is biased, although downward in cases of third variable effects. This effectively makes the Y-permutation-test unusable to assess the null hypothesis we consider here. When we were to assess the null hypotheses that all variables have no effect, the Y-permutation-test could be of value. However, the downward bias does raise questions about the power of that test, especially in cases with more than two explanatory variables. This is a serious issue that deserves attention, because many studies have published results on the basis of this test. With regard to partialing-permutation the multicollinearity conditions do not seem to cause serious problems. We were only able to detect a slight upward bias. However, it is not clear which null-hypothesis could be assessed with a partialing-test. Finally, the semi-partialing-test is analytically unbiased as confirmed by the simulation results. We propose that this test might be a suitable alternative to t-tests, in case of data with structural autocorrelation. We do emphasize that these results are not yet confirmed for other than Gaussian data.

References

Borgatti, S.P., M.G. Everett, and L.C. Freeman (2002), Ucinet for Windows: Software for Social Network Analysis.: Harvard Analytic Technologies.

Hubert, Lawrence J. (1987), Assignment Methods in Combinatorial Data Analysis. New York and Basel: Marcel Dekker, INC.

Krackhardt, D. (1988), "Predicting with Networks - Nonparametric Multiple-Regression Analysis of Dyadic Data," Social Networks, 10 (4), 359-81.

Smouse, Peter E., Jeffery C. Long, and Robert R. Sokal (1986), "Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence," Systematic Zoology, 35 (4), 627-32.