

SENSITIVITY OF MRQAP TESTS TO
COLLINEARITY AND AUTOCORRELATION
CONDITIONS

DAVID DEKKER¹

RADBOUD UNIVERSITY NIJMEGEN

DAVID KRACKHARDT

CARNEGIE MELLON UNIVERSITY

TOM A. B. SNIJDERS

UNIVERSITY OF GRONINGEN

¹Correspondence should be sent to e-mail: d.dekker@fm.ru.nl, phone: ++31243615462,
fax: ++31242611933

Abstract

MRQAP (Multiple Regression Quadratic Assignment Procedure) tests are permutation tests for linear multiple regression model coefficients for data organized in square matrices of relatedness among n objects. Such a data structure is typical in social network studies, where variables indicate some type of relation between a given set of actors. We present a new permutation method (called “double semi-partialing”, or DSP) that complements the family of extant approaches to MRQAP tests. We assess the statistical bias (type I error rate) and statistical power of the set of five methods, including DSP, across a variety of conditions of network autocorrelation, of spuriousness (size of confounder effect), and of skewness in the data. These conditions are explored across three assumed data distributions: normal, gamma, and negative binomial. We find that the Freedman-Lane method and the DSP method are the most robust against a wide array of these conditions, with a slight advantage given to DSP over the Freedman-Lane approach under some cases. We also find that all five methods perform better if the test statistic is pivotal. Finally, we find limitations of usefulness for MRQAP tests: All tests degrade under simultaneous conditions of extreme skewness and high spuriousness for gamma and negative binomial distributions.

Key words: Keywords: MRQAP, Mantel tests, Permutation tests, Social Networks, Network Autocorrelation, Collinearity, Dyadic data

Introduction

The problem of statistical analysis of social network data is an old one (Proctor, 1969). Such data are dyadic in the sense that for a basic set of objects O_1, \dots, O_n , the observed variables refer to pairs of objects and therefore are doubly indexed, e.g., the variable Y_{ij} refers to the way in which object O_i is related to object O_j . Depending on the application Y_{ij} can be continuous (e.g. trade between nations) or discrete (e.g. binary in trust or no trust between individuals, or count as in frequency of linking between web pages). It has long been understood that the lack of independence of dyadic observations creates a severe barrier to reasonable interpretations of statistical tests (Laumann & Pappi, 1976; Laumann, Marsden, & Galaskiewicz, 1977). Indeed, using standard OLS models on dyadic data with a moderate amount of structural autocorrelation (lack of independence among observations within the rows and within the columns of network data) biases the estimates of the second moments to such an extent that it is not uncommon for type I errors of t -statistics to exceed 50% (Krackhardt, 1988), rendering the significance tests unusable for all practical purposes.

Within the past twenty years, two substantially different approaches to solving this problem have been proffered: linear models tested with the aid of the Mantel test (Mantel, 1967), which is often referred to as the quadratic assignment procedure (QAP) in social network studies, and exponential random graph models. The QAP approach (Mantel, 1967; Hubert, 1987; Krackhardt, 1987) provides a specific type of permutation test which keeps intact the dyadic data structure under the permutations. The principle of this test can be applied to many kinds of models, but it is

usually applied to linear models for data treated as continuous. Through a series of Monte Carlo simulations, Krackhardt (1988) showed that parameters in an OLS model of network autocorrelated data could be tested using a multiple regression extension of the QAP test, called MRQAP. Exponential random graph models (Holland & Leinhardt, 1981; Frank & Strauss, 1986; Wasserman & Pattison, 1996; Snijders, Pattison, Robins, & Handcock, 2006), on the other hand, are a family of models for dichotomous or other discrete network data where the focus is on modeling the specific network-related dependence structure. The present paper is in the line of the first approach.

Krackhardt's MRQAP approach had an appeal of simplicity and accessibility and at the time it was said that it "... takes an exciting step forward in the analysis of network relations" (Pattison, 1988). It has been frequently used to test models in network research (e.g., S. P. Borgatti & Cross, 2003; Gibbons & Olk, 2003; Sorenson & Stuart, 2001; Mizruchi, 1990). However, the simulations used in Krackhardt (1988) to justify this approach were confined to a quite restricted set of conditions, namely that there is no correlation among the independent variables nor a correlation in the population between the dependent variable and *any* of the independent variables (that is, the null hypothesis was independence among all variables). More recently, literature on permutation tests has shown that collinearity² might cause problems (e.g., Anderson & Robinson, 2001). However, in this literature the problem of the typical kind of dependence of matrix data is usually not addressed and no specific

²Here we use the term collinearity to indicate any type of correlation between the independent variables used in an analysis.

attention is given to QAP-like tests (for an exception, see Legendre, 2000).

In this paper we study what the literature on permutation tests in linear models has to offer for MRQAP tests for social network data. We assess the problems that might occur due to collinearity in MRQAP-like tests and whether the different proposed permutation tests are robust against sender and receiver dependence as encountered in social network research. Based on prior results in the literature on permutation tests in linear models, we conjecture about the performance of different MRQAP approaches under various collinearity, skewness, and autocorrelation conditions. Furthermore, we develop a new approach to the MRQAP test that addresses some of the shortcomings of previously used versions of MRQAP. Finally, we evaluate these conjectures through an extensive simulation study under parameter values under the null as well as alternative hypotheses, that are typical for data sets encountered in practice in social network analysis for numerical relational variables. We show how various reasonable permutation methods, all consistent with the spirit of the originally proposed MRQAP tests, and including our new method, vary considerably in their ability to provide statistical tests with correct type I error rates and good power properties. The conclusion is that for data that do not have a strongly skewed or granular discrete distribution, and for which a linear model is appropriate, some – but not all – of the MRQAP approaches can perform very well in practice.

QAP Permutation Test for Regression Coefficients

Background

The QAP approach to studying dyadic data in general has been widely used by researchers in a broad array of disciplines, such as statistics (Oden & Sokal, 1992), biology (Legendre, 2000), and psychology (Hubert, 1987). The approach was first suggested by the statistician Mantel (1967) to address the epidemiological question of whether the distribution of diseases is geographically clustered. It was Hubert (1987) who adopted the term “quadratic assignment” and who found a vast array of applications for this type of test.

Since these early days, many statisticians have uncovered a variety of techniques for performing these permutation tests and concomitant advantages and disadvantages for different data situations (see Legendre, 2000, for good review on distance data) (see Anderson & Legendre, 1999; Anderson & Robinson, 2001, for good studies on permutation tests in ordinary regression). However, the vast majority of these studies have focused on the type of autocorrelation proposed in the original Mantel study — that is, observations are thought to be autocorrelated to the extent that they are geographically proximal. In Legendre, Lapointe, and Casgrain (1994), autocorrelation was phylogenetic and not geographic.

The focus of this paper, by contrast, is network data that are commonly considered to have a different kind of autocorrelation structure. In dyadic data, observations are autocorrelated if they are in the same row (emanate from the same sender) or in the same column (go to the same receiver) (e.g., Holland & Leinhardt, 1981). We wish to assess whether these insights into the properties of the QAP permutation tests hold up

under conditions of network autocorrelated data.

Bivariate QAP

The MRQAP test was developed as an extension of the bivariate QAP model. Mantel (1967) first proposed the quadratic assignment procedure (QAP) to assess association between spatial and temporal distance data using a regression approach. These data can be represented in square $n \times n$ matrices where the elements reflect Euclidian distances between numerical attributes of n objects (O_j , where $j = 1, \dots, n$) and the diagonal is structurally zero. Application of the QAP method subsequently was proposed for other structurally autocorrelated data such as social network data where square matrices reflect relations between individuals (e.g., Baker & Hubert, 1981; Krackhardt, 1987). Matrices that reflect social networks are specific in that dyadic relations are often asymmetric. This is usually not the case with the distance matrices analyzed in other fields using the Mantel test, although it happens from time to time.

Assume we wish to use QAP to assess statistical significance of some measure of association Γ_{YX} between distance or network variables Y and X , where both Y and X are $n \times n$ matrices referring to a set of n objects O_1, \dots, O_n , and where the (i, j) elements of the matrices refer to aspects of how objects i and j are related. The objects are points in space for distance data, and network vertices (e.g., social actors) for network data. The data matrices could be symmetric or non-symmetric, depending on the situation; this distinction is of no concern to the discussion given here. The association coefficient Γ_{YX} could be, e.g., the elementwise Pearson product moment correlation between X and Y (to calculate association

measures variables are reshaped into vectors). For example, (Kadushin, 1995) shows how French financial elite friendship ties are correlated with (dis)similarity on several attribute variables, such as political preference, educational institute, and club membership.

In the QAP we randomly permute the n objects O_j to obtain a new order of objects $O_{\pi(j)}$, the permutation being denoted by π . For variable Y we calculate the distances or network variables between all objects in the order $O_{\pi(j)}$; the matrix obtained is called $\pi(Y)$. This is equivalent to applying the same permutation to the row and column indices of Y . Since the permuted structure is isomorphic to the original, all the structural features of the original matrix are retained, except for those referring to the order of the objects. The result is that we generate a random data set $\pi(Y)$ with the same row-column interdependence of observations as there is in Y .

Under the null hypothesis of no association between Y and X , the observed association $\Gamma_{\pi(Y)X}$ between $\pi(Y)$ and X is a random association that was drawn from the same underlying distribution as the actually observed association Γ_{YX} . Now we can simulate this distribution by repeating the permutations of the data and after each permutation calculating $\Gamma_{\pi(Y)X}$. This generates a reference distribution of values for Γ under the set of all $n!$ permutations. Comparing the observed Γ_{YX} to this reference distribution provides a distribution-free way of testing the null hypothesis that X and Y are independent, under the model assumption that the distribution of the variables arranged in X and Y is permutationally invariant under permutation of the objects O_1, \dots, O_n . In practice, it suffices to use a relatively small random sample (1000 or 10000) from the set of all $n!$ permutations to approximate this reference

distribution (see also Pitman, 1937; Mantel, 1967; Jackson & Somers, 1989).

Hence, the QAP provides a permutation- or randomization-based nonparametric test of dependence between two square matrix variables of the same size, which may represent, e.g., distances or similarities between objects, or a relation in a group of social actors. The QAP also directly provides the p -value calculated as the relative frequency of the values of the statistic, in the null distribution, that are larger than or equal to the observed value.

Validity of randomization tests hinges on the exchangeability of values of variables under permutation. In this context, the objects generating the rows and columns are exchangeable. Of all the $(n(n - 1))!$ possible permutations of the matrix elements not considering the diagonal, only those $n!$ permutations are used that respect the structure imposed by these objects, and this keeps intact the autocorrelation structure of the observations arranged in the square matrix.

MRQAP Approaches

In Multiple Regression - QAP (MRQAP), the study of the dependence between matrices X and Y is compounded because there are other variables Z for which the association between X and Y must be controlled. Multiple regression is often used for hypothesis testing based on dyadic data in social sciences. E.g., in their analysis of predicting the volume of international trade Krempel and Plümper (2003) a distance measure, the gross domestic product of countries and dummy variables that indicate whether two countries share a border and have a coast on the same ocean. On a micro level, Krackhardt and Kilduff (1990) study the effect of

friendship ties on cultural agreement among individuals and control for tenure differences and formal organizational position.

Here, all variables are considered to be mean-centred over all observations excluding the diagonal values, which are ignored. Overviews and comparisons for vector data are given by Anderson and Legendre (1999) as well as Anderson and Robinson (2001), for analysis of variance (ANOVA) by Anderson and Ter Braak (2003), and for distance matrices by Legendre (2000). The approach taken is the multiple linear regression approach: it is assumed that Y depends linearly on Z , and it is tested whether there is an additional (linear) dependence on X . It should be noted that in all cases, for vector and matrix data alike, linear dependence between variables is proposed as an approximation that is convenient and, hopefully, a good representation of reality. In the case of real-valued numerical data, the approximation often is quite good, and it may be expected that there is sufficient robustness for small deviations from linearity. For binary and other coarsely grained discrete data, however, linear models are less appropriate and other models for spatial or network data such as, e.g., those proposed by Snijders et al. (2006) or Heagerty and Lele (1998) may be used.

The basic linear model for square matrix data considered here is

$$Y = \beta X + Z\gamma + E \tag{1}$$

where Y , X , and E are $n \times n$ matrices, β is a scalar, Z is a $n \times n \times q$ array, and γ is $q \times 1$. The diagonals of the matrices are always ignored. The null hypothesis is

$$H_0 : \beta = 0 . \tag{2}$$

The variables Z and X are not assumed to be independent. Specifically,

we assume between these variables the linear model

$$X = \delta Z + V, \quad (3)$$

where V is a $n \times n$ matrix. The situation $\delta \neq 0$ will be called collinearity. The nonparametric approach to square matrix data means here that the residuals associated with the n objects, are exchangeable, or equivalently, the matrices E and V are invariant under permutations of rows and columns simultaneously by the same permutation. Whenever the term *permutation* is used, it will be assumed that this permutation acts on rows and columns simultaneously and in the same way, as describe above for the QAP.

It will be convenient to use the following notation. The OLS estimator of β –the partial regression coefficient– for the vectorized data is denoted by

$$\hat{\beta} = \hat{\beta}(Y, X | Z). \quad (4)$$

We use the notation $\pi(Y)$ for the data matrix Y where both rows and columns are rearranged by the permutation π .

As stated earlier, the permutation principle can be applied to various test statistics. The QAP was explained above using as test statistic an association coefficient such as the elementwise correlation coefficient between the two matrices X and Y . Consider first the bivariate case where the third variable Z is absent from the model. Then the dependence between X and Y can be tested by the correlation coefficient r in a permutation distribution. The following statistics can be used:

1. the correlation coefficient $r = r(X, Y)$,

2. the t transform $t = \sqrt{n(n-1) - 2r/\sqrt{1-r^2}}$, which has an exact t null distribution only if E is normally distributed and there is no network autocorrelation, i.e., all elements of the matrices E are independent,
3. or the estimated regression coefficient

$$\hat{\beta} = \frac{s_Y}{s_X} r \quad (5)$$

where the standard deviations refer to the vectorized data.

In the bivariate QAP permutation test the estimated regression coefficient is a constant multiple of the correlation coefficient since the two standard deviations remain constant under the permutations. Therefore, these three test statistics will lead to the same results for the bivariate QAP. For the MRQAP approaches where more complicated statistics must be used to control for the dependence on the third variable Z , the equivalence between these three possibilities no longer holds. Consequently, in the MRQAP case, we cannot assume that these three apparently equivalent measures of association will behave the same way under random permutations. Indeed, as we will show, they do not.

Different methods of data permutation have been proposed. They can be classified as “raw data” permutation methods and “residual” methods, as will be explained below. Some of these methods, when applied to some statistics may not have desirable properties. For example, Kennedy and Cade (1996) show that the method of permuting the raw Y -matrix yields substantial biased results with the use of a non-pivotal statistic. A statistic is defined to be pivotal when the distribution of the statistic under the null hypothesis is independent of the “nuisance” parameters; in our case, the nuisance parameters may be taken to be the residual variance and the

regression coefficient δ in (3). When considering the null hypothesis $\beta = 0$ under permutational invariance (rather than independence) of the residuals, the nuisance parameters also include the type of dependence (autocorrelation) of the residuals E . The partial regression coefficient is not pivotal. Thus, the Kennedy and Cade (1996) result suggests that it would be inappropriate to use the regression coefficient as a test statistic under the permutation of the raw Y -matrix. Furthermore, Anderson and Robinson (2001) show that although most “residual” methods asymptotically give the same results, they differ in the ability to approximate an exact test in small samples. It can be concluded from this literature that in some procedures it is clearly better to use a pivotal statistic; in other cases there might be no important difference. We will explore the sensitivity of each of the MRQAP methods by subjecting each to both pivotal statistics (under the null hypothesis) and non-pivotal statistics in our test below.

Raw Data Methods: X and Y Permutations

Intuitively attractive approaches to MRQAP for the three matrix case (equation 1), are two simple extensions of the bivariate case: the X -permutation and Y -permutation procedures (e.g., Oja, 1987; Smouse, Long, & Sokal, 1986). In the former, we permute the X matrix associated to the tested multiple-regression coefficient and subsequently estimate the multiple regression coefficients for the permuted data sets to generate its reference distribution. In the latter we permute Y .

The H_0 for the permutation of X is that Y is not related to X although it could be related to Z . Now, permuting X destroys the

relationship between X and Z and thereby violates the ancillarity principle (Welch, 1990; Braak, 1992; Anderson & Legendre, 1999), which states that the dependence between these independent variables should be kept intact in the permutational procedure. It could be expected that this lack of respect for the dependence between X and Z will lead to incorrect type I error rates in cases of collinearity.

The H_0 for the permutation of Y is that Y is not related to X and Z taken together. It hence is a test for the hypothesis that both β and γ are 0, not a test of β alone (controlling for a possibly non-null effect of Z). However, various researchers have been using the test for the latter purpose. This may explain that the literature on permutation tests has over the years presented contradictory advice on the use of the Y -permutation approach to MRQAP. Although Smouse et al. (1986) suggest this method is not the best approach, it has been applied in many studies. In social network analysis it has been applied perhaps unwittingly by many researchers, because it was the only approach to MRQAP implemented in the main network analysis software package UCINET, versions 3 to 6.1 (S. Borgatti, Everett, & Freeman, 2002).

Manly (1997) argued for the Y -permutation, but this approach was criticized by Kennedy and Cade (1996), especially if a non-pivotal statistic is used. Anderson and Robinson (2001) found that the asymptotic significance level of the Y -permutation test is indeed equal to the nominal level when an asymptotically pivotal statistic is used. An asymptotically pivotal statistic does not depend on any unknown parameters and thereby adjusts for any nuisance parameters that are not of interest to the test (Anderson & Robinson, 2001). The partial correlation coefficient and the corresponding t -statistics are examples of asymptotically pivotal statistics,

independent of first and second moment population parameters. They are monotonic to each other, and can therefore be used equivalently for a significance test of individual multiple regression coefficients.

The following shows the effects of collinearity (a linear relation between X and Z corresponding to $\delta \neq 0$ in (3)) when one uses the partial regression coefficient as reference statistic. The following relation holds between the partial regression coefficient $b_{yx.z}$ and the partial correlation coefficient $r_{yx.z}$:

$$b_{yx.z} = r_{yx.z} \frac{\sqrt{1 - r_{yz}^2}}{\sqrt{1 - r_{zx}^2}} \frac{S.D.(y)}{S.D.(x)}. \quad (6)$$

Under Y -permutation the variable Y is replaced with $\pi(Y)$. The partial regression coefficient under permutation is

$$b_{\pi(y)x.z} = r_{\pi(y)x.z} \frac{\sqrt{1 - r_{\pi(y)z}^2}}{\sqrt{1 - r_{zx}^2}} \frac{S.D.(y)}{S.D.(x)}. \quad (7)$$

Under random permutations, $E(r_{\pi(y)z})$ will be approximately 0, but this is not necessarily true for the non-permuted data. Now, if $\rho_{yz} \neq 0$ the values $r_{\pi(y)z}$ under random permutations will tend to be lower, and consequently the reference distribution consisting of the values of (7) will tend to be larger in absolute value. This will lead to a conservative test with an unnecessarily low power.

Similarly, when we use the multiple regression coefficient under X -permutation the reference values are

$$b_{y\pi(x).z} = r_{y\pi(x).z} \frac{\sqrt{1 - r_{yz}^2}}{\sqrt{1 - r_{z\pi(x)}^2}} \frac{S.D.(y)}{S.D.(x)}. \quad (8)$$

Under permutation, $E(r_{z\pi(x)})$ is approximately 0, which again is not necessarily true under the null-hypothesis ($\beta_{yx.z} = 0$). If $\rho_{zx} \neq 0$ the reference values will tend to be too small in absolute value, leading to a liberal test which is undesirable.

It was proved by (Anderson & Robinson, 2001) that for permutation tests on i.i.d. data, the asymptotic permutational distribution of the partial correlation coefficient, which is the correlation between the residuals of Y and X , under permutations of Y , is standard normal under the null hypothesis. This implies that the Y -permutation approach gives asymptotically a test with the correct type I error rate when a pivotal statistic is used.

Residual Permutation Methods

Next to the X -permutation and Y -permutation approaches, various approaches have been proposed in the literature that are based on fitting a regression model, calculating the residuals, and permuting these residuals. These residuals are estimates of E . If the values of E would be known exactly then an exact test could be based on these, but the fact that only approximations to E are available leads to a diversity of solutions, none of which are exact (Anderson & Robinson, 2001).

The Freedman-Lane Approach to Permutations

First, we follow the procedure proposed for the usual linear model by Freedman and Lane (1983) and discussed further by various other authors (cf. Anderson & Robinson, 2001). In our case the structure is more complicated because we are dealing with square arrays expressing

relational data, but the basic approach can be similar.

Now consider the general case that we do have the third variable Z in the model. The null hypothesis is that there is no effect of X on Y ; all variation in Y can be explained by Z , or is modelled by E . The proposal by Freedman and Lane (1983) is, first, to estimate the effect of Z alone on Y , yielding the residuals

$$\hat{\epsilon}_{YZ} = Y - \hat{\gamma}Z \quad (9)$$

where $\hat{\gamma}$ is the OLS regression parameter estimate for the reduced model

$$Y = \gamma Z + E . \quad (10)$$

The residuals are equal to

$$\hat{\epsilon}_{YZ} = (\gamma - \hat{\gamma})Z + E , \quad (11)$$

which, under H_0 , will be close to E if n is large and the model is well-specified. Second, they propose to permute the residuals and calculate a new hypothetical Y -value made of the fitted portion of the regression equation plus the permuted residuals:

$$Y^\pi = \hat{\gamma}Z + \pi(\hat{\epsilon}_{YZ}) , \quad (12)$$

which is different from $\pi(Y)$ introduced earlier. For each of the $n!$ permutations π , the pivotal F -statistic is calculated where Y^π is regressed on X controlling for Z , i.e., the model

$$Y^\pi = \beta X + \gamma Z + E , \quad (13)$$

is applied. This statistic may be called $F(Y^\pi, X | Z)$. For a one-dimensional X -variable, this is a constant multiple of the squared

t -statistic. The result proved by Freedman and Lane in the usual linear model is that for large n and normal error, the permutation distribution of this statistic is approximately a multiple of chi squared, and the same as the large sample distribution of the usual parametric F test. For permutation testing, the important thing is that this asymptotic distribution does not depend on the unknown δ or γ , and that under $H_0 : \beta = 0$, it is the same as the asymptotic distribution that would be obtained if the true E would be used instead of $\hat{\epsilon}_{YZ}$. This implies that the permutation test of the null hypothesis defined by (1), (2), and the assumption of permutational invariance over E has approximately for large n the correct level of significance. This result is subject to some assumptions which say basically that there are no outliers in the Z or Y variables.

If Z is absent from the model, then the two procedures presented in the preceding subsection are also equivalent to this procedure, because permuting X with constant Y or permuting Y with constant X is equivalent if there is no Z variable.

Comparing the F test on the residuals with formula (13), and using (12) and (11), we have

$$\begin{aligned}
\hat{\beta}(Y^\pi, X \mid Z) &= \hat{\beta}(\hat{\gamma}Z + \pi(\hat{\epsilon}_{YZ}), X \mid Z) \\
&= \hat{\beta}(\pi(\hat{\epsilon}_{YZ}), X \mid Z) \\
&= \hat{\beta}(\pi((\gamma - \hat{\gamma})Z), X \mid Z) + \hat{\beta}(\pi(E), X \mid Z) \quad (14)
\end{aligned}$$

and in the last line, the second term has a permutation invariant distribution while the first term is negligible when n is large and therefore $(\gamma - \hat{\gamma})$ is small. This suggests that the Freedman-Lane approach of

permuting residuals can be used not only for asymptotically pivotal statistics but also when using the partial regression coefficient as a test statistic. The same procedure for pivotal statistics was proposed by Anderson and Legendre (1999) for vector data and Legendre (2000) for matrix data.

This method, which we call the ‘Freedman-Lane Semi-Partialing’ (*FLSP*) approach, might also be termed *Y*-conditional-semi-partialing regression approach, since the effect of Z is partialled out from Y and then the resulting Y residuals are permuted and regressed on X and Z . Note that in the process Z is entered in the regression twice. The effect of Z need not be also partialled out from X , since Z is used in the regression anyway and X and Z are not permuted with respect to each other. Furthermore, putting in $\hat{\gamma}Z$ as is done in (12) does not change the F statistic because Z is being controlled for in model (13). Following Anderson and Robinson (2001) we could also use the partial correlation $r(\hat{\epsilon}_{YZ}, X | Z)$ as a test statistic, and test it in the permutational distribution $r(\pi(\hat{\epsilon}_{YZ}), X | Z)$. This will also be approximately the same as using the F test as proposed by Freedman and Lane (cf. the ‘first idea’ in their proof, which implies that asymptotically, $F(Y^\pi, X | Z)$ and $r^2(Y^\pi, X | Z)$ are functions of each other).

Double Semi-Partialing

We now introduce a new approximately exact test that follows a similar logic as the *FLSP* and hence complements the set of similar existing permutation tests. Whereas the *FLSP* approach conditions the test statistic under permutation on the relationship between Y and Z , we

could follow an analogous reasoning and condition it on the relationship between X and Z . Although this is a logical analogue of the Freedman & Lane approach, it seems not to have been proposed before.

Similar to (9) we define

$$\hat{\epsilon}_{XZ} = X - \hat{\delta}Z \quad (15)$$

where $\hat{\delta}$ is the OLS estimate for the model

$$X = \delta Z + V . \quad (16)$$

In the new method these residuals are permuted and the model

$$Y = \beta \pi(\hat{\epsilon}_{XZ}) + \gamma Z + E \quad (17)$$

is used to obtain reference values for the test statistic. The rationale is here that under the null hypothesis $\beta = 0$, the reference model (17) for Y is the same as the original model (1), and if the estimation error $\hat{\delta} - \delta$ is negligible, the permutational invariance assumption for V implies that

$$\pi(\hat{\epsilon}_{XZ}) = \pi((\delta - \hat{\delta})Z + V) \quad (18)$$

has the same distribution as V .

This approach has some parallels to the Freedman & Lane approach that focuses on generating a permuted Y . More, specifically Y^π , which as is shown in (14), is equal to using $\pi(\hat{\epsilon}_{YZ})$ (see also Anderson & Legendre, 1999). Instead of generating Y^π this procedure focusses on creating a X^π , or rather, similar to what Anderson and Legendre (1999) propose to simplify things, focusses on $\pi(\hat{\epsilon}_{XZ})$.

We refer to this new approach as *Double-Semi-Partialing (DSP)* regression since the effect of Z is partialled out of X and then the resulting

residuals $\hat{\epsilon}_{XZ}$ are permuted and entered in a regression of Y on both $\hat{\epsilon}_{XZ}$ and Z . As such Z enters the regression twice, hence the “double”. Where the *FLSP*-approach minimizes the effect of ancillary variables and the dependent variable under permutation, the *DSP*-approach minimizes the correlation between the focal variable and the control variables under permutation. Both methods therefore respect the ancillarity principle as they condition on the nuisance statistics. Simulations below will have to determine which of these two methods minimizes the effect of the “nuisance” parameters under permutation.

Partialing

Many other approaches could be considered with respect to what is partialled out and what is permuted. One of the possibilities is to estimate residuals E under the full model (1) rather than the reduced model (10), as proposed by Braak (1992). However, Anderson and Robinson (2001) obtained for this method power properties that were inferior to some other methods, which was confirmed by some of our own simulations (not further reported below). Therefore this approach is not further elaborated here.

Another approach proposed in the MRQAP literature is called partialing (Smouse et al., 1986; Krackhardt, 1988; Kennedy, 1995). This approach tests $r(\hat{\epsilon}_{YZ}, \hat{\epsilon}_{XZ})$ in the distribution of reference values $r(\hat{\epsilon}_{YZ}, \pi(\hat{\epsilon}_{XZ}))$. The partialing approach was found by Anderson and Legendre (1999) to be incorrect in the case of multiple regression, and also found to be incorrect by Legendre (2000) for the partial Mantel test.

Anderson and Legendre (1999) review many different permutation proposals for the usual linear model (with vector data). These authors

conclude that the Freedman-Lane proposal (here called *FLSP*) applied to the partial correlation coefficient corresponds best to the test based on the unobserved true residuals E ; where ‘best’ refers to the actual significance level as well as the power of the test. For the partialing approach, which they call the Kennedy approach in reference to Kennedy (1995), they show that this test will always result in a higher level of type I errors than the Freedman-Lane approach.

In the remainder of this paper we analyze how different MRQAP tests on partial regression coefficients perform in terms of accuracy and power when we impose different collinearity and autocorrelation conditions. Also, we compare the consequences of using non-pivotal statistics in “raw data” permutation methods. This extends the simulation study of MRQAP tests for matrix data carried out by Legendre (2000), who considered distributions representing distance matrices, whereas we specifically focus on dependence structures that are typical for network data.

Method

This section presents a simulation study using model specifications consistent with the type of network data to which MRQAP approaches have been applied in the literature (e.g., Nelson, 1989; Krackhardt & Kilduff, 1999; S. P. Borgatti & Cross, 2003). Autocorrelations in the sense of within-row and within-column correlations range from nil to strong. Normally distributed data are considered as well as skewed and discrete data.

For getting simulated matrices with correlations within rows as well as within columns, and with different kinds of non-normal data, it is convenient to use the formula

$$M_{ij} = R_i + C_j + E_{ij} \quad (19)$$

where the row contributions R_i , the column contributions C_j , and the cell contributions E_{ij} are independent ($M = X, Y, Z$). When the row and column contributions have the same variance $\text{var}(R_i) = \text{var}(C_j) = \sigma_1^2$ and the cell contributions have variance $\text{var}(E_{ij}) = \sigma_2^2$, then the within-row as well as the within-column autocorrelation is $\rho_a = \sigma_1^2 / (2\sigma_1^2 + \sigma_2^2)$.

In the simulation we matrices are of dimensions 10×10 . Hence, since diagonal values are ignored, simulations are based on 90 dyad level observations. The following other choices are made for the simulations:

1. 4 levels of autocorrelation, $\rho_a = .000, .150, .300, .450$;
2. Normal distributions, Gamma distributions (*skewed*), Negative Binomial distributions (*skewed and discrete*);
3. 3 levels of skewness (g), respectively (.250, .750, .950).

The reason for studying the effect of skewness of the simulated data on the performances of the tests is that in fact network data is often skewed.

Here we use results based on cumulants to derive the skewness of the sum of R_i, C_j, E_{ij} in equation (19). Similar to moments, cumulants are a set of descriptive constants of distributions (Stuart & Ord, 2005, p.86).

The skewness of the sum of two independent random variables can be derived relative easily, because the third cumulant equals the third central moment and the cumulant of a sum of random variables is the sum of the cumulants of those variables (Stuart & Ord, 2005, p.89; p.156). Hence, the sum of third central moments of two independent random variables equals the third central moment of the sum of these variables.

Here we use $\langle \rangle$ -brackets as expectations operator. Now, let

$$\mu_3(q) = \langle q - \langle q \rangle \rangle^3 \quad (20)$$

$$\mu_3(r) = \langle r - \langle r \rangle \rangle^3 \quad (21)$$

respectively, be the third central moments of q and r than

$$\mu_3(q + r) = \mu_3(q) + \mu_3(r) \quad (22)$$

Define skewness as

$$\gamma_s = \frac{\mu_3(s)}{\mu_2(s)^{1\frac{1}{2}}} \quad (23)$$

where the second central moment of z is, $\langle z - \langle z \rangle \rangle^2 = \sigma_z^2$, is the variance of variable z . For $s = q + r$ we get

$$\gamma_{q+r} = \frac{\mu_3(q) + \mu_3(r)}{(\mu_2(q) + \mu_2(r))^{1\frac{1}{2}}}. \quad (24)$$

We want to express the numerator in equation (24) in terms of γ_q and γ_r . After some algebra we can derive

$$\gamma_{q+r} = \frac{\gamma_q \sigma_q^3 + \gamma_r \sigma_r^3}{(\sigma_q^2 + \sigma_r^2)^{1\frac{1}{2}}} \quad (25)$$

Given a desired level of skewness for M_{ij} now allows us to choose the levels of skewness for R_i , C_i and E_{ij} . Recall that we already choose levels of variance to control the levels of autocorrelation in the data.

Also, we impose different levels of correlation between Y and X implied by the joint dependence (“collinearity”) of X and Y on Z . More specifically, in this case

$$r_{YX} = r_{YZ}r_{ZX} \tag{26}$$

We enforce that $r_{YZ} = r_{ZX}$, hence $\sqrt{r_{YX}} = r_{YZ} = r_{ZX}$. The value of r_{YX} is a spurious correlation between Y and X , reflecting an apparent relationship between Y and X that is in fact induced by the third variable Z . When we control for Z in a multiple regression setting the bivariate effect of X on Y disappears. In the following, “spuriousness” hence indicates the correlation between Y and X due to the presence of Z .

To represent the collinearity conditions we use a 3×3 correlation matrix T that represents the correlations between Y , X , and Z . Furthermore, we vectorize and concatenate our square data matrices into the data matrix D of dimension $n(n - 1) \times 3$. A Cholesky decomposition of T provides the weights for the construction of the correlated data set DT . From this correlated data set DT of dimension $(n(n - 1) \times 3)$ we reconstruct the 3 square $(n \times n)$ variable matrices Y , X , and Z . In our simulations the collinearity conditions expressed by ρ_{XZ} increase in 3 steps by .3 from 0 to .9. In the case of integer values in the negative binomial count data sets, we round the values of Y , X , and Z to the nearest integer.

In the “type I error”-simulations we record the percentage of rejections (based on 1000 runs per step) of the (true) null hypotheses, where there is no relation between Y and X (occurrences of type I error). We assess robustness of MRQAP tests by evaluating results at a significance level of $\alpha = .10$ (similar to Krackhardt, 1988). To assess the consistency of the different tests based on the Monte Carlo simulations we use the lower and upper boundaries of the 99% expected outcome interval (Edgington, 1995).

For the power study we generated data for which the null hypothesis

was false, i.e. $\beta \neq 0$. Data were generated in a similar fashion as in the “type I error”- study described above. The difference was that we set the correlation as

$$r_{YX} = r_{YZ}r_{ZX} + .1 \quad (27)$$

where again $r_{YZ} = r_{ZX}$.

In the power study we thus vary the level of spuriousness given an extra correlation of .1 between Y and X , and we vary the levels of autocorrelation in the data.

Results

The consequences for type I error of using the non-pivotal statistic $\hat{\beta}$ in the “raw data” methods are shown in Figures 1a to 1b. They show how increasing spuriousness due to collinearity of Y and X with a third variable Z harms the accuracy of these MRQAP tests. The MRQAP test using $\hat{\beta}$ with the Y -permutation will almost never reject the null-hypothesis when spuriousness increases (see Figure 1a). This should translate in reduced power when there is an effect. In contrast, the MRQAP test that uses $\hat{\beta}$ with the X -permutation almost always rejects the null hypothesis (see Figure 1b). Both tests are therefore uninformative. Figures 2a and 2b show that under these conditions pivotal statistics do give results close to the expected α level of .10 for the same set of permutations on the same data. These results corroborate the conclusions of Legendre (2000) for matrix and those of Legendre and Anderson (1999) and Anderson and Robinson (2001) for i.i.d. (vector) data.

Figures 1 and 2 about here

Note that in Figures 1 and 2 the results seem to be independent of the level of autocorrelation. A close inspection of our results for the other methods (partialing, *FLSP*, *DSP*) shows similar results (see Figures 1c to 1e and 2c to 2e). This confirms and extends the results of Krackhardt (1988) as it shows that statistical bias of MRQAP tests is minimal under varying conditions of autocorrelation and spuriousness - but only when a pivotal statistic, such as t , is used.

However, the methods do differ in their ability to approximate the exact test. The 99% expected outcome interval for 1000 runs and $\alpha = .10$ is $(.075 - .125)$ (Edgington, 1995). If rejection proportions fall outside this interval they significantly differ from the nominal value $\alpha = .10$ and this indicates a biased reference distribution.

The results for Normal data (Figure 2) show that all values remain within the expected outcome interval. Remarkable is the fact that this is also true for the “raw data” methods. Despite the good performance of “raw data” methods, overall these methods typically give less robust results compared to the semi-partialling methods for a non-pivotal statistic (see Figure 1). Therefore, in the remainder we will not further focus our results on the “raw data” methods. Partialing gives type I error rates that are always larger than or equal to the results of *FLSP* (all data points in Figure 2c are higher than comparable points in Figure 2e). This is consistent with the findings of Anderson and Robinson (2001). Furthermore, the results show that *DSP* is often closer to the expected type I error level than partialing (for Normal data the average of absolute deviation from .10-level over 16 simulated autocorrelation-collinearity conditions are .008 and .013 respectively). Because *FLSP* has been analytically shown to be superior to partialling (Anderson & Robinson,

2001) we subsequently focus our results on the differences between *FLSP* and *DSP*.

Figures 3 and 4 about here

Simultaneously high levels of skewness and spuriousness harm the MRQAP tests. This can be seen from the increasing type I error rates in Figures 3 and 4 as both increase as skewness and spuriousness increases. Figure 3 shows that Gamma data with $g = .250$ do not generate conditions that cause a serious problem for *FLSP* and *DSP*. Indeed these results are comparable to the results generated with normal data. However, for high levels of skewness ($g = .75$ and $g = .95$) we see that rejection rates structurally cross the upper bound of .125 when spuriousness is also high ($r_{YX} > .30$). Overall, results for the Gamma-distributions and Negative Binomial distributions are similar, although the various MRQAP tests perform slightly less well for the latter (compare Figures 3 and 4).

The results of the power study with normal data are summarized in Figure 5 (similar results were obtained for the skewed distributions). The results show that the power of the different methods depends heavily on both the levels of spuriousness and of autocorrelation in the data. Figure 5 shows that autocorrelation decreases power, while spuriousness enhances power given that formula (27) is used to define the point under the alternative hypothesis where power is studied. The fact that autocorrelation decreases power results from the fact that autocorrelated data contain less statistical information that can be utilized in this analysis. The results show that for both types of data, when autocorrelation becomes excessive, power of the test deteriorates heavily.

Figure 5 about here

As would be expected partialing has the highest apparent power, which is due to its inflated rate of type I error. Because the rate of type I error is substantially higher than the significance level, the resulting tests are invalid (see Edgington, 1995, p.13); thus they will not be considered further as a viable option here. The other methods give similar results, although the “residual” methods perform somewhat better than the “raw data” methods, as we expected. The power of the tests is lower for the skewed data than for the normally distributed data.

Discussion and Conclusion

Many studies in many different fields employ MRQAP for analysis of network data in the linear model framework. It is a popular approach mainly because it both utilizes the benefits of multiple regression analysis and offers a solution to problems of autocorrelation that prohibit the use of standard statistical tests. However, the results obtained with MRQAP analyses are sensitive to particular methods and options implemented in standard software packages. In particular, the “raw data” permutation methods in this study that use non-pivotal statistics performed so poorly that they should be avoided, as rejection rates either rise far beyond acceptable levels or become extremely low. In the permutation test literature, these difficulties have been recognized (e.g., Anderson & Robinson, 2001). Our analysis expands on this literature and adds to it more extensive insights into the problems of the different approaches in the case of structurally autocorrelated data such as encountered in network analysis. Also it adds insights in how combinations of spuriousness and skewness affect the tests.

Despite the findings of Legendre (2000) that outliers inflate type I

error for “raw data” methods (specifically Y -permutation), some might insist that our results suggest that Y -permutation on a pivotal statistic is a good choice, especially for Gaussian data. This method would be the most efficient implementation option for software developers, because it simultaneously generates results for all coefficients.

However, simple reliance on a pivotal statistic presents its own problems, since it is not always determinable whether a statistic is pivotal in any given situation. A t -statistic and a partial correlation coefficient are pivotal statistics for multivariate normal distributions; but, when data stem from other distributions where higher moments are informative, these statistics are no longer guaranteed to be pivotal. Furthermore, the autocorrelation structures of network data imply that these statistics may no longer be pivotal under the null hypothesis of row and column permutational invariance considered in our paper.

The non-pivotal statistics in “raw data” methods may be statistically biased because of ancillary variables’ effects. In the simulations presented here the “residual” methods do not suffer from these problems. As partialling is always inferior to *FLSP*, two recommendable methods remain, i.e. *DSP* and *FLSP*.

We must emphasize that this study is limited to conditions where the data conform to a general linear model, especially models where the dependent variable is continuous or at least countable integers (Poisson-like). In particular, we do not know the behavior of the different MRQAP approaches for models wherein the data are binary or ordinal, as is often found in social network data (see Snijders et al. (2006) for alternative models for these cases). It remains for further research to show what MRQAP approaches, possibly vastly different from those discussed

here, have to offer for these types of models.

Taking all of these considerations into account, we suggest that in general the safest recommendation for researchers testing linear models of continuous or skewed network count data employ MRQAP analysis using a *DSP* method with a *t*-statistic (or similarly judged pivotal statistic). But, given the varieties of types, distributions, constraints and conditions of data gathered in the network area, perhaps the best advice is that researchers simulate their own data structures and use the method demonstrated to be most robust in their particular problem and circumstances.

References

- Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62(3), 271-303.
- Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, 43(1), 75-88.
- Anderson, M. J., & Ter Braak, C. J. F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, 73(2), 85-113.
- Baker, F., & Hubert, L. (1981). The analysis of social interaction data. *Sociological Methods and Research*, 9, 339-361.
- Borgatti, S., Everett, M., & Freeman, L. (2002). *Ucinet for windows: Software for social network analysis*. Harvard Analytic Technologies.
- Borgatti, S. P., & Cross, R. (2003). A relational view of information

- seeking and learning in social networks. *Management Science*, 49(4), 432-445.
- Braak, C. ter. (1992). Permutation versus bootstrap significance tests in multiple regression and anova. In K.-H. Jckel, G. Rothe, & W. Siedler (Eds.), *Bootstrapping and related techniques*. Berlin: Springer-Verlag.
- Edgington, E. (1995). *Randomization tests. 3rd edition*. New York: Marcel Dekker, Inc.
- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832-842.
- Freedman, D., & Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics*, 1, 292-298.
- Gibbons, D., & Olk, P. M. (2003). Individual and structural origins of friendship and social position among professionals. *Journal of Personality and Social Psychology*, 84(2), 340-351.
- Heagerty, P. J., & Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93, 1099-1123.
- Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability-distributions for directed-graphs. *Journal of the American Statistical Association*, 76(373), 33-50.
- Hubert, L. J. (1987). *Assignment methods in combinatorial data analysis* (Vol. 73). New York and Basel: Marcel Dekker, INC.
- Jackson, D. A., & Somers, K. M. (1989). Are probability estimates from the permutation model of mantel's test stable? *Canadian Journal of Zoology*, 67, 766-769.

- Kadushin, C. (1995). Friendships among the french financial elite. *American Sociological Review*, 60(2), 201-221.
- Kennedy, P. E. (1995). Randomization tests in econometrics. *Journal of Business & Economic Statistics*, 13(1), 85-94.
- Kennedy, P. E., & Cade, B. S. (1996). Randomization tests for multiple regression. *Communications in Statistics-Simulation and Computation*, 25(4), 923-936.
- Krackhardt, D. (1987). Qap partialling as a test of spuriousness. *Social Networks*, 9, 171-186.
- Krackhardt, D. (1988). Predicting with networks - nonparametric multiple-regression analysis of dyadic data. *Social Networks*, 10(4), 359-381.
- Krackhardt, D., & Kilduff, M. (1990). Friendship patterns and culture: The control of organizational diversity. *American Anthropologist*, 92, 142-154.
- Krackhardt, D., & Kilduff, M. (1999). Whether close or far: Perceptions of balance in friendship networks in organizations. *Journal of Personality and Social Psychology*, 76, 770-782.
- Krempel, L., & Plümper, T. (2003). Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualization. *Journal of Social Structure*, 4(1).
- Laumann, E. O., Marsden, P. V., & Galaskiewicz, J. (1977). Community-elite influence structures: Extension of a network approach. *American Journal of Sociology*, 83(3), 594-631.
- Laumann, E. O., & Pappi, F. U. (1976). *Networks of collective action: A perspective on community influence systems*. New York: Academic

Press.

- Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial mantel tests. *Journal of Statistical Computation and Simulation*, 67(1), 37-73.
- Legendre, P., & Anderson, M. (1999). Distance-based redundancy analysis: testing multi-species responses in multi-factorial ecological experiments. *Ecological Monographs*, 69, 1-24.
- Legendre, P., Lapointe, F.-J., & Casgrain. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, 48, 1487-1499.
- Manly, B. (1997). *Randomization, bootstrap, and monte carlo methods in biology*. (2nd ed.). London: Chapman & Hall.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209-220.
- Mizruchi, M. S. (1990). Capitalist collective action - competition, cooperation, and conflict in the coal industry - bowman,jr. *American Journal of Sociology*, 96(2), 473-475.
- Nelson, R. E. (1989). The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal*, 32, 377-401.
- Oden, N. L., & Sokal, R. R. (1992). An investigation of 3-matrix permutation tests. *Journal of Classification*, 9(2), 275-290.
- Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Australian Journal of Statistics*, 29, 81-100.
- Pattison, P. E. (1988). Network models: Some comments on papers in this special issue. *Social Networks*, 10(4), 383-411.
- Pitman, E. (1937). Significance tests which may be applied to samples

- from any population. *Royal Statistical Society, Supplement, 4*, 119-130.
- Proctor, C. H. (1969). Analyzing pair data and point data on social relationships, attitudes and background characteristics of costa rican census bureau employees. *Social Statistics Section, Proceedings of the American Statistical Association*, 457-465.
- Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Mutiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology*, 35(4), 627-632.
- Snijders, T., Pattison, P., Robins, G., & Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology*, *in press*.
- Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, 106(6), 1546-1588.
- Stuart, A., & Ord, J. K. (2005). *Kendall's advanced theory of statistics. volume 1: Distribution theory*. London: Hodder Arnold.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks .1. an introduction to markov graphs and p. *Psychometrika*, 61(3), 401-425.
- Welch, W. (1990). Construction of permutation tests. *Journal of the American Statistical Association*, 85, 693-698.

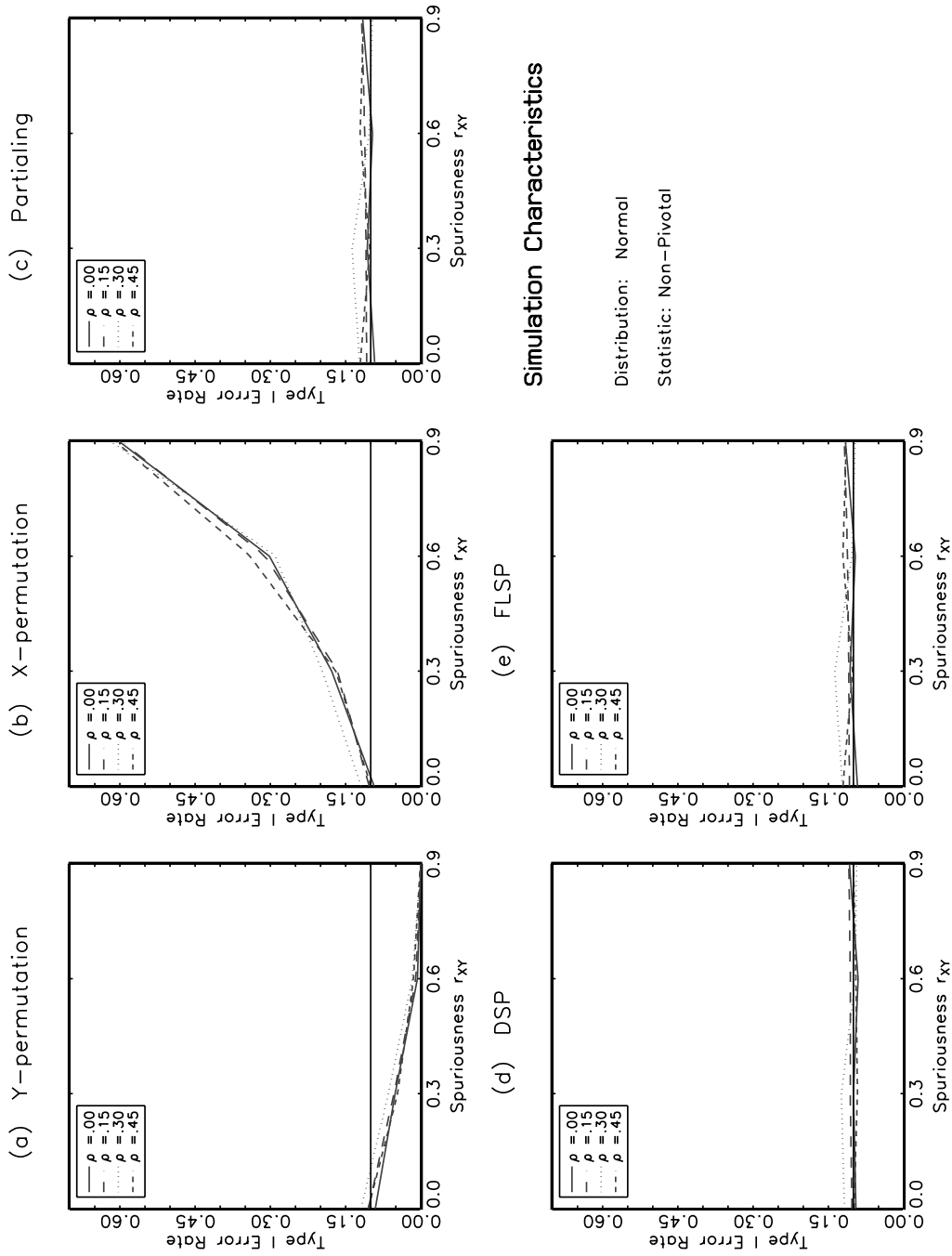


FIGURE 1.

Type I error rates of different MRQAP approaches applied to $\hat{\beta}$, a non-pivotal statistic, under varying conditions of spuriousness and autocorrelation, and normally distributed data. Horizontal reference line for $\alpha = .10$.

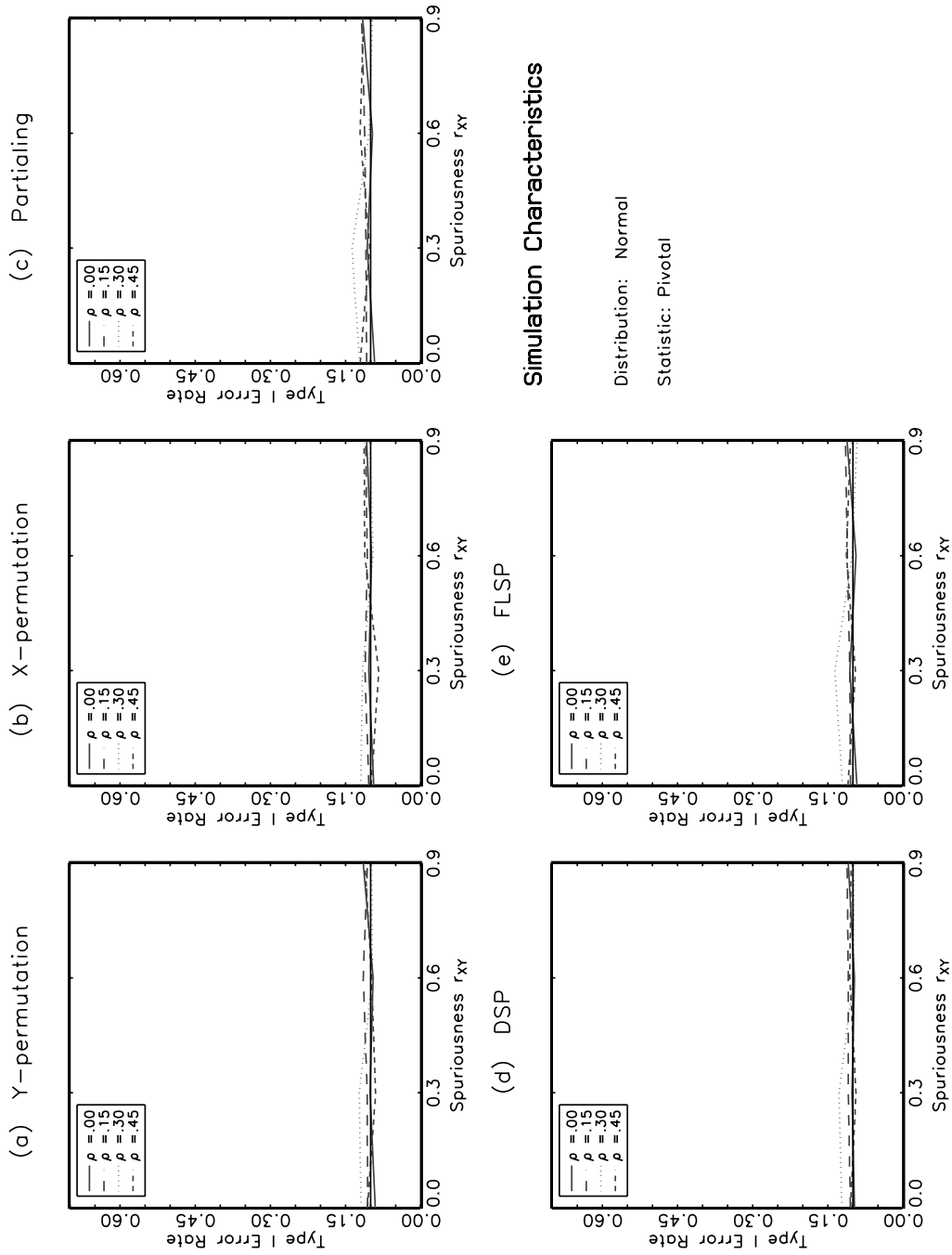


FIGURE 2.

Type I error rates of different MRQAP approaches applied to t_β , a pivotal statistic, under varying conditions of spuriousness (r_{XY}) and autocorrelation (ρ), and normally distributed data. Horizontal reference line for $\alpha = .10$.

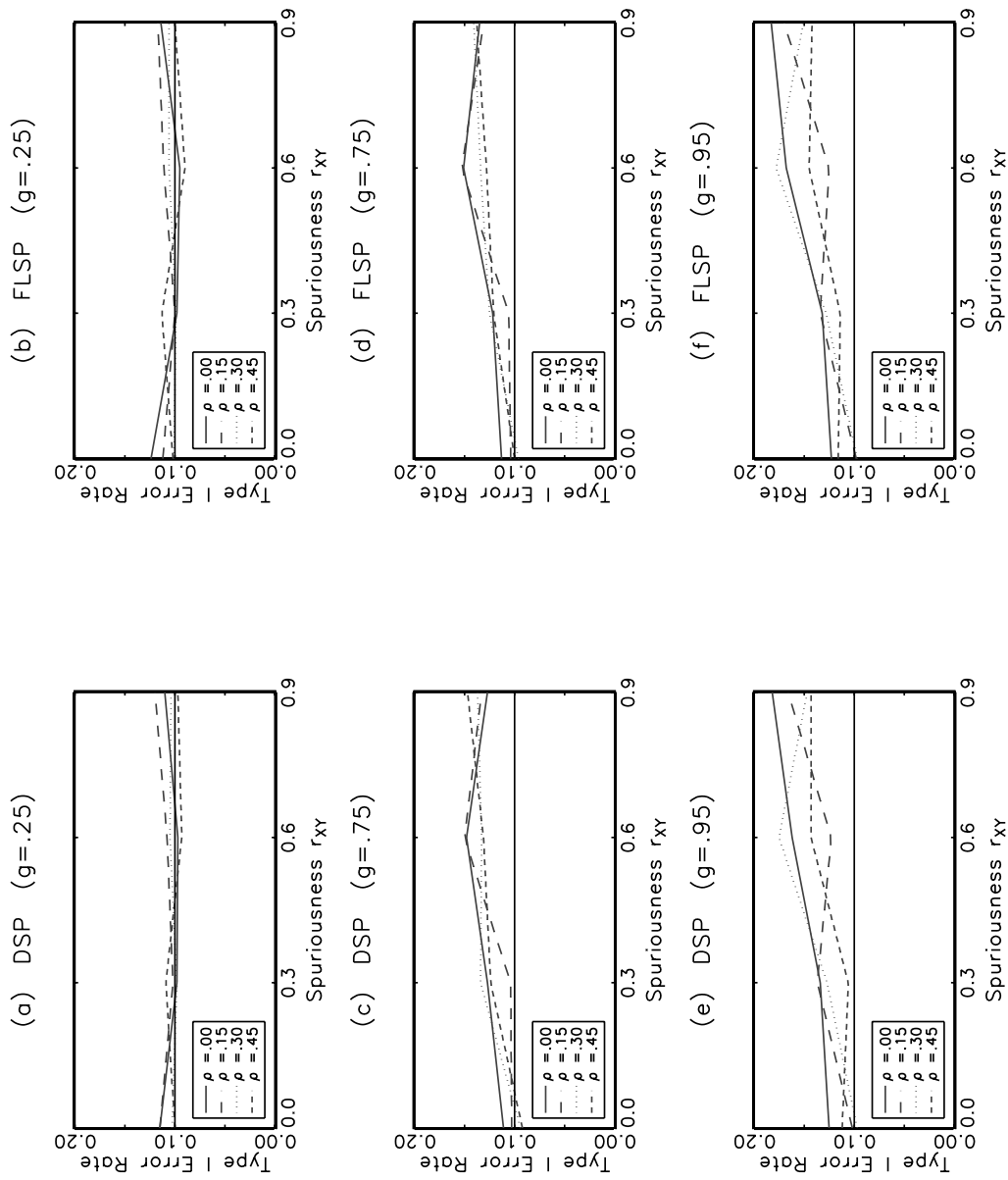


FIGURE 3.

Type I error rates of different MRQAP approaches applied to t_{β} , under varying conditions of spuriousness (r_{XY}), autocorrelation (ρ) and skewness (g), and gamma distributed data. Horizontal reference line for $\alpha = .10$.

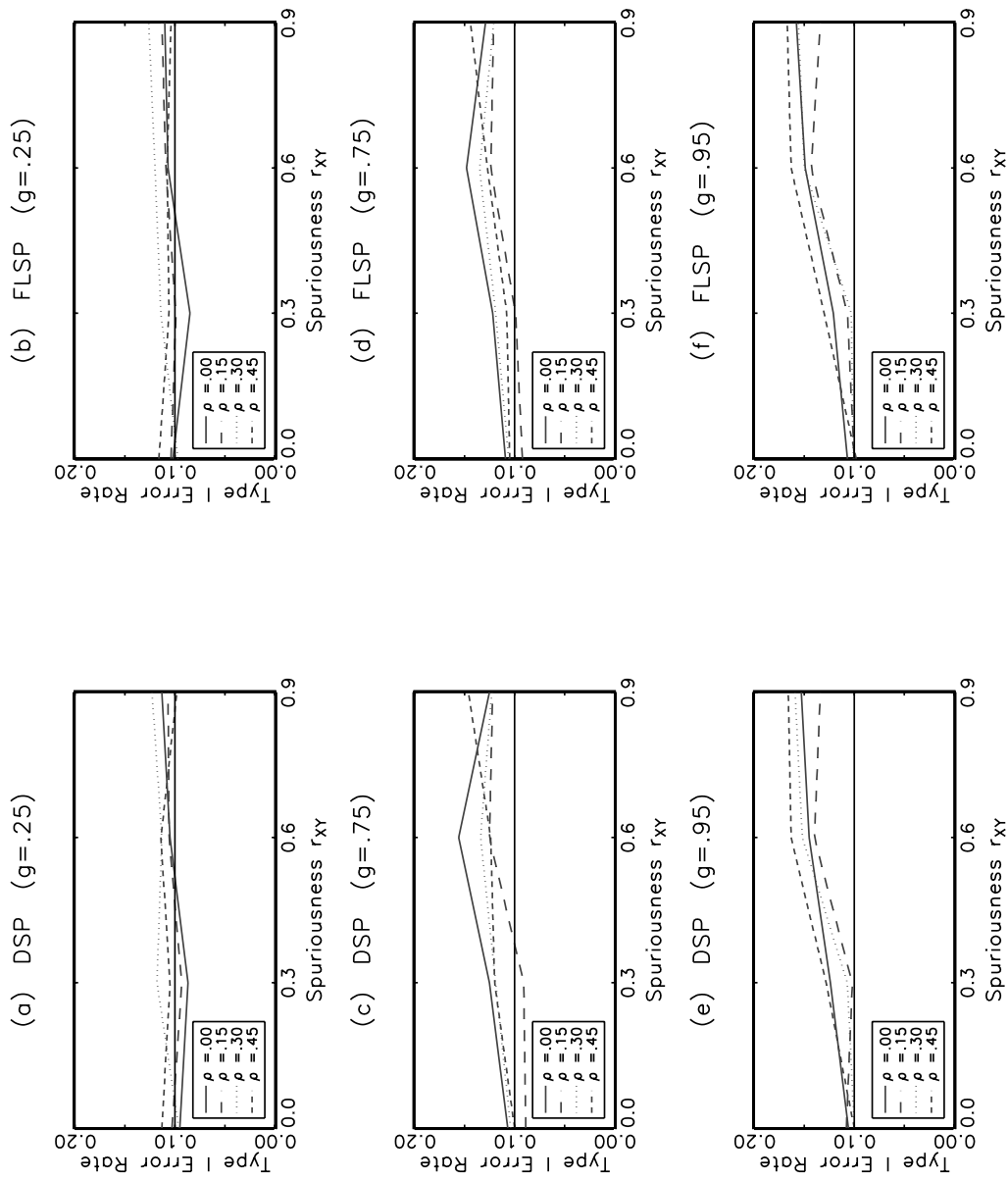


FIGURE 4.

Type I error rates of different MRQAP approaches applied to t_{β} , under varying conditions of spuriousness (r_{XY}), autocorrelation (ρ) and skewness (g), and negative binomial distributed data. Horizontal reference line for $\alpha = .10$.

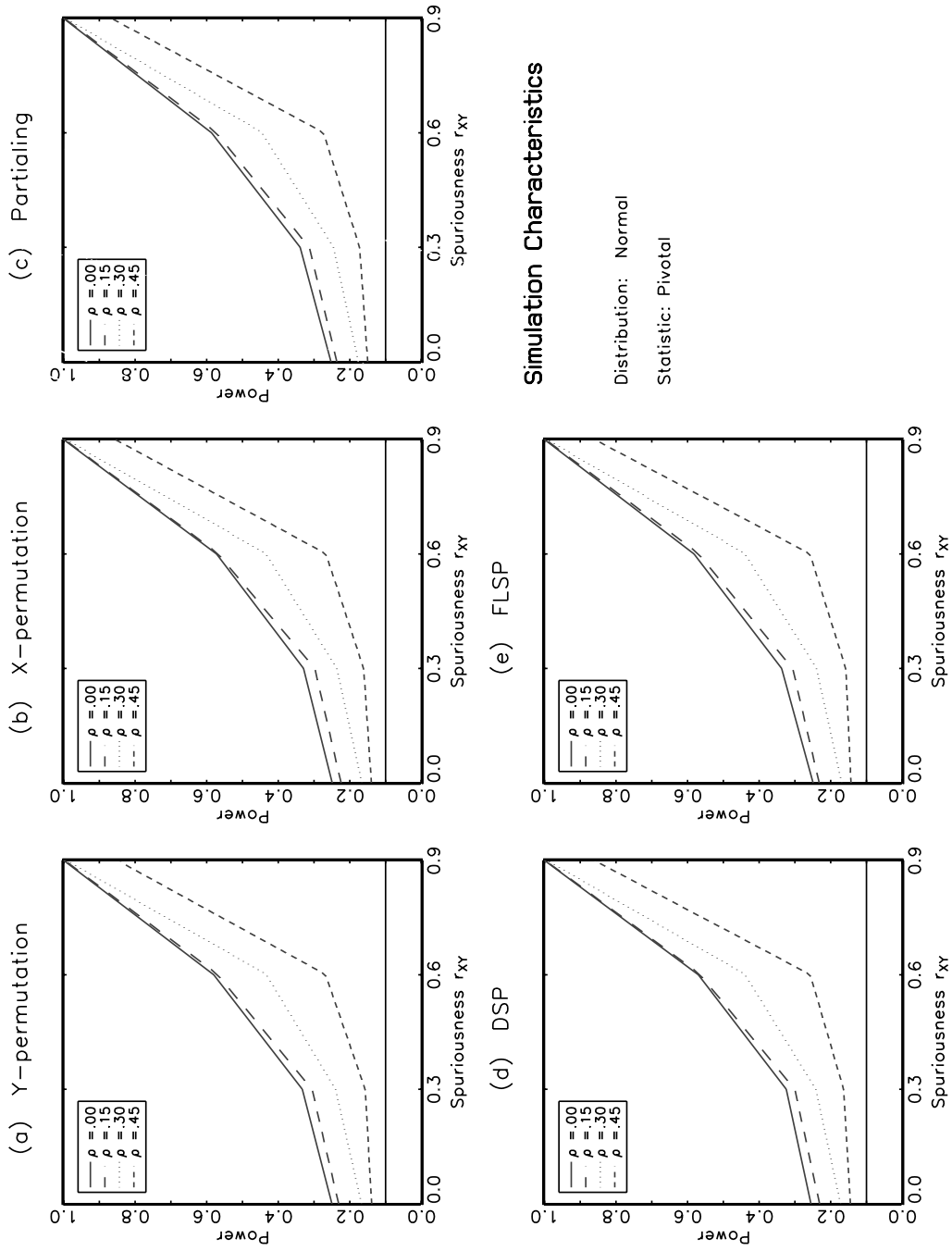


FIGURE 5.

Power of different MRQAP approaches applied to t_{β} , under varying conditions of spuriousness and autocorrelation, and normally distributed data. Horizontal reference line for $\alpha = .10$.