An Equilibrium-Correction Model for Dynamic Network Data

David Dekker*‡

Philip Hans Franses

Erasmus University Rotterdam

David Krackhardt

INSEAD and Carnegie Mellon University

June 11th, 2001

Abstract

We propose a two-stage MRQAP to analyze dynamic network data, within the framework of an equilibrium-correction (EC) model. Extensive simulation results indicate practical relevance of our method and its improvement over standard OLS. An empirical illustration additionally shows that the EC model yields interpretable parameters, in contrast to an unrestricted dynamic model.

*Dekker, David, Philip Hans Franses, and David Krackhardt
2003 "An Equilibrium-Correction Model for Dynamic Network Data."
Journal of Mathematical Sociology, 27: 193-215.

* We would like to thank Richard Paap, Bas Donkers, Erjen van Nierop and Frans Stokman for useful comments.
‡ Corresponding author: Erasmus University Rotterdam, Tinbergen Institute, P.O. Box 1738, 3000 DR ROTTERDAM, The Netherlands, Tel: ++31 10 4088938, Fax: ++31(0)10-4089031, E-mail: ddekker@few.eur.nl

1.Introduction

In network analysis there is an increasing interest in longitudinal investigations (see for example Doreian & Stokman 1996; Feld 1997; Burt 2000). Current models for these analyses are often based on Markov Chain methods, see Leenders (1996) for overview. Although these models have proven to be useful (Snijders 2000; van de Bunt 1999), they do have some potential limitations. One such limitation is that Markov Chain methods do not make a distinction between "change" effects and "level" effects of explanatory variables. As we believe that this distinction is useful in network studies, we propose a model that explicitly incorporates "change" and "level" effects.

The model specification we propose to use is the equilibrium-correction model (EC-model), which is often used in time-series econometrics (see Greene, 2000). This model describes effects on changes in a dependent variable, which can for example be relationship strength. In this respect it mirrors models like the p^*-model (Wasserman & Pattison, 1995) and SIENNA (Snijders, 2000), which address the probability of change. A distinction is however that the EC-model explicitly incorporates effects of changes in explanatory variables over time (short-term effects) and effects of a variable that describes equilibrium relation (long-term effects). As such, we believe the EC-model to be a valuable instrument for the analysis of network dynamics.

As is well known, inference on network data based on ordinary least squares (OLS) or non-linear least squares (NLS) can lead to spurious results. Autocorrelation (serial as well as structural) may lead to underestimation of standard errors, which makes correct inference based on these estimates impossible (see Johnston & DiNardo 1996). Although the equilibrium-correction model handles serial autocorrelation, it is considered for network data it seems wise to rely on the multiple-regression quadratic assignment procedure (MRQAP) for parameter inference (Hubert & Schultz 1976; Krackhardt 1988). MRQAP is a non-parametric method, which makes no a-priori distributional assumptions.

The outline of the paper is as follows. In section 2 we first briefly discuss the equilibrium-correction model and the MRQAP approach. In section 3 we report on the extensive simulations to check if the model works in practice. In section 4 we discuss an empirical illustration. In the final section we present our conclusions.

2. Qap-ing An Equilibrium-Correction Model

In econometric time series analysis the equilibrium-correction model is often used due to some nice features. Most importantly, the model handles serial autocorrelation (which occur when observations are dependent over time), while it also gives interpretable parameters. In the following we first discuss the advantages of the EC-model. Second, we discuss the MRQAP approach which is practically relevant as network data are prone to structural autocorrelation because of the inherent row and/or column dependency between observed relations (Lincoln, 1984).

2.1 An Equilibrium-Correction Model

There are several ways to deal with serial autocorrelation in network data. Serial autocorrelation implies that the error terms ($\varepsilon_{ij,t}$) are correlated over time, for example like $\varepsilon_{ij,t}=\rho\varepsilon_{ij,t-1} + v_t$, with $0<\rho<1$, and where v_t might be distributed as $N(0, \sigma^2_v)$. In such data there is a correlation between observations in subsequent periods. In this exemplary case then we can say that data have a first-order dynamic structure. A general model to handle first-order dynamics is the so-called auto-regressive distributed lag model, ADL(1,1) model, which is given by,

$$y_{ij,t} = \beta_0 + \rho y_{ij,t-1} + \beta_1 x_{ij,t} + \beta_2 x_{ij,t-1} + e_{ij,t}. \qquad (1)$$

In this model it is assumed that $y_{ij,t}$ depends on its own past, and also on current and past explanatory variables $x_{ij,t}$. Of course, *(1)* can be extended to include more than one explanatory variable, in which case $x_{ij,t}$ denotes a vector.

A potential drawback of *(1)* is that it may not always be easy to interpret the estimated parameters. For example, there is the possibility that β_1 and β_2 get opposite signs. One way to facilitate parameter interpretation amounts to rewrite *(1)* into the equilibrium-correction model, that is

$$y_{ij,t} - y_{ij,t-1} = \gamma_0 + \gamma_1(x_{ij,t} - x_{ij,t-1}) + \gamma_2(y_{ij,t-1} - \gamma_3 x_{ij,t-1}) + e_{ij,t}. \qquad (2)$$

It is easy to see that the parameters in *(2)* are uniquely related with those in *(1)* by $\gamma_0 = \beta_0$, $\gamma_1 = \beta_1$, $\gamma_2 = (\rho - 1)$ and $\gamma_3 = \frac{-(\beta_1 + \beta_2)}{(\rho - 1)}$.

The EC specification enables a sensible interpretation of the parameters. In the EC model, γ_1 can be interpreted as the short term effect of x on y as it captures the effect of changes of x on those of y. Furthermore, γ_3 can be interpreted as indicating the long-term equilibrium relation between y and x, while γ_2 measures the speed of adjustment of y to that long-term equilibrium.

For time series data, OLS (or NLS) yields consistent estimates of γ_1, γ_2, γ_3. However, for network data, with potential structural autocorrelation it may not. To solve this issue, Krackhardt (1988) proposes a method for parameter inference that is robust against structural autocorrelation, and this is what we discuss next.

2.2 MRQAP to Handle Structural Autocorrelation

A major problem with network data is that it is sensitive to structural autocorrelation, and hence a straightforward application of OLS might result in spurious findings (see Greene

2000; Jonston & DiNardo 1996). Structural autocorrelation may occur because row and/or column entries in a socio-matrix are dependent. Krackhardt (1988) proposes the MRQAP as an inference procedure that is robust against structural autocorrelation. The QAP entails a non-parametric test for the significance of parameter estimates. It compares OLS parameter estimates based on the original data with OLS estimates that are estimated using random data. Simultaneous permutation of the rows and columns of the dependent network data matrix generates random data with exactly the same autocorrelation structure as the original data. Repeating parameters estimation with different sets of such random data generates a distribution of estimates with which estimates based on the original data can be compared. As the expected value of the repeated estimates is zero, an original estimate that is sufficiently larger or smaller than the randomly generated coefficients can be considered to differ significantly from zero.

Krackhardt (1988) shows that the QAP is robust to structural autocorrelation in the two and three variable regression model, where this model does not involve dynamics. It remains to be seen whether this also applies to a dynamic model.

2.3 Solutions to Anticipated Problems

We anticipate some problems if we would straightforwardly apply the MRQAP to the EC model or the ADL(1,1) model. These problems primarily concern our specification of the level of serial autocorrelation in the EC-model and ADL(1,1) model, that is the ρ-parameter. The randomization of $y_{ij,t}$ has consequences for the estimation of ρ, γ_2 and γ_3 as well as of β_2 and β_3 in (1) or (2) during the QAP-procedure. In our discussion of the possible problems with MRQAP, we will indicate a randomized $y_{ij,t}$ in the MRQAP as $y^*_{ij,t}$ and also will identify parameter estimates that are generated by the MRQAP with an asterisk.

Consider again the ADL(1,1) model in (1). MRQAP seems to offer a good basis to test whether ρ is a spurious result due to structural autocorrelation. Under the null hypothesis of MRQAP, the expected value of ρ^* is zero, that is, there is no relation between $y^*_{ij,t}$ and $y^*_{ij,t-1}$. If the value of ρ would not differ from, say, at least 90% of the ρ^* that were estimated during the MRQAP, we would have no grounds to reject the null hypothesis at a 10% level. In that case we should consider that the OLS value of ρ is due to neglected structural autocorrelation or is just zero indeed.

Similarly, we could analyze the β_2 and β_3 parameters in the ADL(1,1,) model, but here also problems could arise. Note again that there is no relation between $y^*_{ij,t}$ and $y_{ij,t-1}$ (the expected value of ρ^* is zero). However, there is a relation between $y^*_{ij,t}$ and $L(y_{ij,t-1})$, where $L(.)$ represents the randomization function that describes the permutation of rows and columns that created $y^*_{ij,t}$. This relation implies that serial autocorrelation did not disappear, but that it does not have a first-order structure anymore. Actually, the serial autocorrelation in the data has taken a form that can best be interpreted as a form of structural autocorrelation. In the MRQAP the serial autocorrelation that was controlled for in the original model, has become uncontrolled structural autocorrelation. As such during an MRQAP, the level of serial autocorrelation (ρ) affects the estimation of the other parameters. This has strong consequences for the usefulness of the benchmark distribution of β_2 and β_3 that was generated by the MRQAP.

A consequence of this increase in the level of structural autocorrelation is that the variation in the size of the estimates of the parameters increases (recall that neglected autocorrelation decreases the efficiency of parameter estimates). As ρ does not correct for serial autocorrelation anymore, the estimates of the other parameters would increasingly differ from zero for increasing levels of serial autocorrelation. This would make the MRQAP a too conservative test, because the range that captures, say, 90% of the values of β_2^* and β_3^* becomes broader.

To solve the above problems, we advocate the use of a two-stage quadratic assignment procedure (TS MRQAP). To see whether ρ captures structural or serial autocorrelation, we apply MRQAP as would be done for non-dynamic multiple regression models. Hence, we simultaneously randomize i and j of $y_{ij,t}$ to generate random data with the same structural autocorrelation as $y_{ij,t}$. In the second stage, we not only randomize $y_{ij,t}$, but also $y_{ij,t-1}$ such that the relation between $y_{ij,t}^*$ and $y_{ij,t-1}^*$ still involves ρ^*. When applying MRQAP, we then explicitly control for serial autocorrelation, which allows the assessment of whether the other parameter estimates are spurious due to neglected structural autocorrelation.

With regard to γ_3 in the EC-model (model (2)), a final remark has to be made. As $\rho<1$, when ρ becomes larger (and ρ-1 thus becomes smaller), $\gamma_3 = \frac{-(\beta_2 + \beta_3)}{(\rho - 1)}$, would go to infinity when ρ approaches 1. The TS MRQAP may then give too liberal results for γ_3, especially when ρ is large. To counter this outcome we need to control for ρ when testing the null hypotheses that $\gamma_3=0$. As γ_3 is zero when $\beta_2 + \beta_3 = 0$, it suffices to test whether this condition holds.

3 Simulations

In this section we present some simulations to see whether TS MRQAP, as we described in the previous section, works in practice. These simulations would indicate whether a TS MRQAP analysis of the ADL(1,1) and the EC-model is robust against structural autocorrelation.

3.1 Data Generating Process

As is done in Krackhardt (1988), we generate random data with varying levels of structural and serial autocorrelation on a dependent variable ($y_{ij,t}$) and a single independent variable ($x_{ij,t}$). This data generating process (DGP) implies that there is neither a short-term nor a long term relation between x and y. We estimate the parameters for the two period ADL(1,1) model in *(1)* and the associated EC-model in *(2)*, with the following data:

$$y_{ij,t} = K_R\zeta_{yi,t} + K_C\zeta_{yj,t} + K_B\zeta_{yij,t} + \rho(y_{ij,t-1}) \tag{3}$$

$$x_{ij,t} = K_R\zeta_{xi,t} + K_C\zeta_{xj,t} + K_B\zeta_{xij,t} \tag{4}$$

where K_R and K_C represent the levels of structural autocorrelation in respectively the rows and columns of the matrix and ρ is the serial autocorrelation parameter. The $\zeta_{xi,t}, \zeta_{xj,t}, \zeta_{xij,t}, \zeta_{yi,t}, \zeta_{yj,t}$, and $\zeta_{yij,t}$ are randomly distributed gaussian variables (*N(0,1)*). The autocorrelations take values between $0 < K_B \leq 1, K_R = 1 - K_B$, $K_R = K_C$ and $0 < \rho < 1$, with steps of .05. Thus, 441 combinations of structural and serial autocorrelation values have been evaluated.

3.2 Tests

In the simulations we record the percentage of rejections (based on 1000 runs) of the (true) null hypotheses, that is, that there are no short-term and long-term relations between dependent and explanatory variables. As both the dependent and independent variables are random, we would expect to find no relations between them. On the other hand, we would expect the relation between the dependent (y_t) and lagged dependent (y_{t-1}) to be as large as ρ. Therefore we only test the null hypothesis (ρ=0).

All inference of the parameters in the EC-model can be done on the basis of the ADL(1,1) model. An advantage of this model is that it is linear in the parameters. From the ADL(1,1) parameter estimates we derive the parameter values and standard errors of the EC-model parameters (see Greene 2000, pp.118-120). We determine the robustness against autocorrelation as the degree to which the t-test and TS MRQAP-test reject the null hypotheses of no significant effects at the α =0.10 level. We expect for TS MRQAP that the rejection rate of the null hypotheses to be α on average (see Krackhardt 1988).

3.3 Simulation Results

Figures 1a to 3c and table 1a and 1b summarize our simulation results. First, figure 1a shows us that the TS MRQAP analysis of ρ is robust against structural autocorrelation. With increasing levels of structural autocorrelation, the number of rejections based on the MRQAP-test remains 10% when indeed there is no serial autocorrelation. As expected we see that the t-test is not robust against structural autocorrelation (see Figure 1b). this graph indicates that the t-test based rejection rate of the null-hypothesis that ρ=0 increases as structural autocorrelation increases.

Insert figure 1a and 1b about here

Secondly, table 1a shows that regular MRQAP is too conservative, because the rejection rate goes to zero in the analysis of β_2. These results are similar for γ_2 and β_3 and we therefore do not report those results. When we control for serial autocorrelation, as we do in the TS MRQAP analysis, results are satisfactory (see table 1b). Furthermore, figure 2a shows us that TS MRQAP analysis of β_2 (and γ_2 and β_3) is robust against structural autocorrelation, without becoming a test that is too conservative. And, as expected, figure 2b shows that the t-test of β_2 (and γ_2 and β_3) is not robust against structural autocorrelation.

Insert tables 1a and 1b about here

Insert figures 2a and 2b about here

Figure 3a shows that when we do not control for ρ the TS MRQAP-analysis of γ_3 $(=\frac{-(\beta_2+\beta_3)}{(\rho-1)})$ is not robust against increasing levels of serial autocorrelation. When the structural autocorrelation is indeed zero, the TS MRQAP-analysis rejects the null-hypothesis that $\gamma_3=0$ more often with increasing ρ. However, as discussed above, to test whether $\gamma_3=0$ it is sufficient to test that $\beta_2 + \beta_3 = 0$. From figure 3b it becomes clear that TS MRQAP-analysis of this condition is robust against structural autocorrelation. Figure 3c again shows that the t-test of $\gamma_3=0$ is not robust against structural autocorrelation.

Insert figures 3a, 3b, and 3c about here

To summarise our simulation results, it seems that TS MRQAP has excellence performance, and it is more reliable than the OLS-based t-statistics.

4. An empirical illustration: Consistent Accuracy

To illustrate the usefulness of EC-models we present an example in which we analyze both ADL(1,1) and EC-models. In this example, we focus on accuracy of social structural perception. In the example we show that indeed the ADL(1,1)-model may give results that have a difficult interpretation, while the interpretation of the EC-model is much more straightforward. First, we will give a short background on the importance of accuracy studies and we discuss the value of a longitudinal study on accuracy. Subsequently, we discuss the data after which we show some results.

4.1 Accuracy of Perceptions

Krackhardt (1990) shows that individuals that accurately perceive the structure of relationships, of which they are a part, positively affects the power they hold in that network. Casciaro (1998) suggests that accurate perceptions may not only affect the individual's ability to get what he/she wants, but also that they have consequences for groups and organizations. Those individuals who perceive the social structure, which defines the access to resources, more accuratly are better able to obtain the resources which are needed for groups and organizations (Burt 1992).

Several studies have shown that degree centrality in networks affect individuals accuracy of perceived networks (Casciaro 1998; Bondonio 1998). Degree centrality is measured as the number of people that have a direct relationship with a focal individual. In this illustration we focus on the effects of indegree centrality and outdegree centrality. The indegree is the number of relationships that a focal individual receives, while the outdegree is the number of relationships that originate from that focal individual.

Centrality indicates the potential for communication in which an actor could be involved (Freeman 1979). More involvement in the communication in the network could have two effects on perception accuracy. First, a central individual receives more information about the structure of the network. Or better, such an individual receives information on the perceptions about the network structure of more other individuals in the network. This effect of centrality is especially captured by outdegree of advice request relationships. Secondly, the perceptions of a more central individual are more dominant in the network. More individuals will take notice of the perceptions of a central individual and therefore his\her perceptions are more likely to become dominant. This effect of centrality would be especially captured by the indegree of advice request networks.

If centrality indeed enhances perceptual accuracy it should do so over time. For example, changes of centrality should be reflected in enhanced or diminished accuracy. In our illustration, we study whether centrality influences the accuracy of social structural perceptions over time. In other words we study whether centrality affects consistency in perception accuracy.

In this illustration accuracy implies a minimum deviation from a certain reference or benchmark. Krackhardt (1987) defines the locally aggregated structure and the consensus structure as two of such references for perceived social structure.

In the locally aggregated structure (LAS), whether a tie exists between two people in a dyad depends on what the two people claim about the relationship. While several rules for combining such local information can be used, in this case we use the Intersection (LAS-I) rule for such a determination. That is, a tie exists from person A to person B if and only if both A and B agree that the tie exists from A to B. Another reference for accuracy is the consensus structure (CS). In this structure a relationship exists if a majority of individuals (more than 50%) perceive the relationship to exist. We measure the accuracy of individual *k's* perceptions as the absolute deviation of individual *k*'s perceptions from these references (LAS and CS).

Different accuracies may be determined. Examples are the accuracy of individual *k* concerning the entire network (Krackhardt 1987) or the accuracy of individual *k* concerning the relationships of each individual in the network (Bondonio 1998). To keep things simple in our illustration, we focus on the perceptions of individual *k*'s own direct relationships.

The ADL(1,1) model and the EC model both have different dependent variables. In our illustration the dependent variable in the ADL(1,1) model is the accuracy of individual *k* on R_{kj} in period t. Given that our data is dichotomous, the value of this variable is always one or zero as can be seen in table 2a.

*** Insert table 2a about here***

The ADL(1,1) models in our example specify the effects of previous accuracy, current centrality and previous centrality on future accuracy. A problem with the ADL(1,1) specification could be that current centrality and previous centrality have opposite effects. It would then be difficult to understand the effects of centrality. We therefore rely on the EC model. In our illustration the EC-model assumes an effect induced by the levels of centrality and an effect of change in the level of centrality. These are different effects, with substantively different meanings.

A consequence is that the dependent variable in the EC-models differs from that of the ADL(1,1) models. In the EC-model the dependent variable is the change in accuracy or the instability of accuracy. Table 2b shows that there are three possible values for change in accuracy when data are dichotomous. The value is zero if no change occurs either because *k* remains accurate or inaccurate. The value becomes positive when an individual becomes more inaccurate and the value becomes negative when an individual becomes more accurate.

Insert table 2b about here

In our empirical analysis we investigate four different models since we aim to distinguish between LAS and CS accuracy and between ADL(1,1) and EC-models. In each model we look at the effects of three types of indegree and three types of outdegree. These different types are respectively based on the CS, LAS and the structure as perceived by each individual personally (the slices of the cognitive social structure). The network we study is an advice request network.

4.2 Data

We collected data on a group of 13 individuals on perceived advice request relationships over two periods. Hence we study 156 changes in accuracy. The data setting is similar to that described in Krackhardt & Porter (1985, 1986). The individuals in the network are employees of a big fast food chain. Employees are subject to standard rules that apply throughout the chain. For example, they have to ware prescribed uniforms. Most of the employees are high school kids that work to earn some spending money. Furthermore, working at that specific restaurant comes with social status, because it is a popular hangout place for high school kids. This data that was collected in the beginning of the 1980's was not been presented in Krackhardt & Porter (1985, 1986). The reason was that those papers focused on turnover as a dependent variable and in this branch there was no turn-over between the two periods.

4.3 Empirical Results

Tables 3a and 3b show the results of our empirical analysis, where the dependent variables are respectively, LAS-based accuracy and change in LAS-based accuracy. Table 3a immediately shows an interpretation difficulty with the ADL(1,1) model. It shows that the indegree that individuals perceive themselves to have now and in a previous period (Indegree Slice t = -.03, p=.02 and Indegree Slice t-1 = .03, p=.02) are negatively and positively related to LAS-based accuracy respectively. This would mean that his/her partners confirm the current perceptions of an individual, while the previous perceptions are not confirmed. On the other hand, in model 3b, we see that the change in accuracy is affected by the change in the perceived indegree (Indegree Slice Δ = -.03, p=.02) and not the level of perceived indegree (Indegree Slice Δ = -.00, p=.39).