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.