ON THE SAMPLING DISTRIBUTIONS OF THE ML ESTIMATORS IN NETWORK EFFECT MODELS

Michele La Rocca¹, Giovanni C. Porzio ², Maria Prosperina Vitale ¹ and Patrick Doreian ³

¹ Department of Economics and Statistics, University of Salerno, Italy (e-mail: larocca@unisa.it, mvitale@unisa.it)
² Department of Economics and Law, University of Cassino and Southern Lazio, Italy (e-mail: porzio@unicas.it)
³ Faculty of Social Sciences, University of Ljubljana, Slovenia; Department of Sociology, University of Pittsburgh, US (e-mail: pitpat@pitt.edu)

ABSTRACT: The present contribution aims at describing the finite sample behaviour of ML estimators in network autocorrelation models, a class of auto-regressive models used to study the effect of networks on a variable of interest when the data points are interdependent. More specifically, through an extensive simulation study, it is investigated whether, and the conditions under which, the ML estimators are asymptotically normally distributed.

KEYWORDS: autocorrelation parameter, maximum likelihood estimation, network density, network autocorrelation model.

1 Introduction

The present contribution aims at describing the finite sample behaviour of ML estimators in network autocorrelation models (Doreian, 1980; Dow et al., 1982), a class of auto-regressive models used to study the effect of networks on a variable of interest when the data points are interdependent. Two models are available within this class: the network effects model and network disturbances model. In the first case, interdependencies between actors are modelled through the inclusion of an autocorrelation parameter in the dependent term, while in the second interdependencies are included in the disturbance term. Within this work, focus is given to the network effects model. It allows individual outcome to be directly associated with neighbours’ levels of outcome.

More formally, let \( y \) be a \((n \times 1)\) \( n \)-vector of values of a dependent (endogenous) variable for \( n \) individuals making up a network, let \( X \) represents the \((n \times p)\) matrix of values for the \( n \) individuals on \( p \) covariates (including an unit
vector for the intercept term), and let $W$ be the $(n \times n)$ network weight matrix whose elements, $w_{ij}$, measure the influence actor $j$ has on actor $i$. The network effects model is defined as:

$$y = \rho Wy + X\beta + \epsilon$$

where $\beta$ is a $(p \times 1)$ vector of regression parameters, $\rho$ is the network autocorrelation parameter referred as the strength of social influence in a network, and error terms $\epsilon$ are assumed to be normally distributed with zero means and equal variances, $\epsilon \sim (0, \sigma^2 I)$.

This class of models represents a popular tool for conducting social network analysis. First adopted to describe social influence mechanism (Marsden & Friedkin, 1993), it has been recently applied in many fields in social sciences (Dow & Eff, 2008; Franzese et al., 2012; Zhang et al., 2013; Vitale et al., 2015).

As for the inferential properties of the estimators, it is known that the maximum likelihood estimator of the autocorrelation parameter has a finite sample negative bias, the amount of which is positively related to the network density (Mizruchi & Neuman, 2008). This bias does not depend on network size, number of exogenous variables in the model, and whether the network weight matrix $W$ was normalized or reported in raw form. The bias also does not depend on structured networks, although it is especially pronounced at extremely low-density levels in the star network (Neuman & Mizruchi, 2010). In addition, Wang et al. (2014) show that the network autocorrelation model well controls for Type I error rates, that the statistical power is a nonlinear function of $\rho$ and the network size, and that the network density and structured networks have little impact on statistical power. With respect to this latter aspect, Farber et al. (2009) have showed that the average degree of a random network impacts the power of tests.

Within this framework, the present work aims to examine both the whole finite sample distribution of the ML estimators of the autocorrelation parameter and the regression parameters. More specifically, through an extensive simulation study, it is investigated whether, and the conditions under which, the ML estimators are asymptotically normally distributed. The finite sample distributions are evaluated with respect to the network density and topology, the distribution of error terms, and the strength of the autocorrelation parameter. In brief, our study provide answers to the following research questions: 1) how the whole sampling distributions of the network effect and the regression coefficient estimators look like; 2) how non-normally distributed errors
modify the sampling distributions; 3) to which extent confidence intervals appropriately work for both the network effect and the regression parameters.

2 First findings

The whole finite sampling distribution of the ML estimators was assessed through a Monte Carlo study by considering: i) different network densities $\Delta$, ii) different positive network effect sizes $\rho$, iii) different network topologies, and iv) normal and non-normal errors $\varepsilon$. More specifically, $\Delta$ was set to $0.05 \leq \Delta \leq 0.70$; $\rho$ is positive with four values $\{0.1, 0.3, 0.6, 0.9\}$, from low to high network effect size; two well-known topologies are taken into account, scale-free mechanism and small world configuration, beyond the E-R random graph; $\varepsilon$ distribution is normal, lognormal, and mixture. We set the sample size to 100 units and we performed 1000 MC replications. In the estimated model the covariates were normal distributed $X \sim N(0,1)$, and the network weight matrix $W$ was row normalized, randomly generated at each run. 432 simulation scenarios were then considered given by: $12 \times 4 \times 3$ network topology $\times 3$ $\varepsilon$ distribution.

The simulation study showed that for high values of the autocorrelation parameter and network density, the sampling distribution of $\hat{\rho}$ (Fig. 1) is negative biased and quite strongly asymmetric, while the sampling distribution of $\hat{\beta}$ coefficients is positive biased and asymmetric for $\hat{\beta}_0$, and unbiased and with heavy tails in presence of non-normal error for $\hat{\beta}_k$. In details, the estimator $\hat{\rho}$ not only contains a systematic negative bias, confirming the previous results discussed in the related literature, but its distribution is typically non-normal and asymmetric as well. This occurs not only in randomly generated networks but in well-established network structures. Furthermore, the non-normality and the asymmetry is not confined strictly to networks with high density. At least in small world networks, these features exist also for very low levels of density. All that has an impact on confidence intervals for $\rho$.

Proper methods could be introduced and studied to deal with the bias of autocorrelation parameter. Particularly, bootstrap based distributions could be more accurate with respect to the asymptotic normal distribution and should be preferred taking into account different network density and network effect size.

References

DOREIAN, P. 1980. Linear Models with Spatially Distributed Data: Spatial
Disturbances or Spatial Effects? Sociology Methods & Research, 9, 29–60.