THE ECONOMETRICS OF NETWORKS

ADVANCES IN ECONOMETRICS

Series editors: Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano, Eric Hillebrand, Daniel L. Millimet, Rodney Strachan

Previous Volumes

Volume 21	Modelling and Evaluating Treatment Effects in Econometrics – Edited by Daniel L. Millimet, Jeffrey A.Smith and Edward Vytlacil
Volume 22	Econometrics and Risk Management – Edited by Jean-Pierre Fouque, Thomas B. Fomby and Knut Solna
Volume 23	Bayesian Econometrics – Edited by Siddhartha Chib, Gary Koop, Bill Griffiths and Dek Terrell
Volume 24	Measurement Error: Consequences, Applications and Solutions – Edited by Jane Binner, David Edgerton and Thomas Elger
Volume 25	Nonparametric Econometric Methods – Edited by Qi Li and Jeffrey S. Racine
Volume 26	Maximum Simulated Likelihood Methods and Applications – Edited by R. Carter Hill and William Greene
Volume 27A	Missing Data Methods: Cross-sectional Methods and Applications – Edited by David M. Drukker
Volume 27B	Missing Data Methods: Time-series Methods and Applications – Edited by David M. Drukker
Volume 28	DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments – Edited by Nathan Balke, Fabio Canova, Fabio Milani and Mark Wynne
Volume 29	Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill
Volume 30	30th Anniversary Edition – Edited by Dek Terrell and Daniel Millmet
Volume 31	Structural Econometric Models – Edited by Eugene Choo and Matthew Shum
Volume 32	VAR Models in Macroeconomics — New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy
Volume 33	Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park
Volume 34	Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier
Volume 35	Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman
Volume 36	Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee
Volume 37	Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage and R. Kelley Pace
Volume 38	Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano
Volume 39	The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi
Volume 40A	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part A – Edited by Ivan Jeliazkov and Justin L. Tobias
Volume 40B	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part B – Edited by Ivan Jeliazkov and Justin L. Tobias
Volume 41	Essays in Honor of Cheng Hsiao – Edited by Tong Li, M. Hashem Pesaran and Dek Terrell

ADVANCES IN ECONOMETRICS VOLUME 42

THE ECONOMETRICS OF NETWORKS

EDITED BY

ÁUREO DE PAULA
University College London, UK

ELIE TAMER
Harvard University, USA

MARCEL-CRISTIAN VOIA
University of Orléans, France

emerald PUBLISHING

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2020

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83867-576-9 (Print)
ISBN: 978-1-83867-575-2 (Online)
ISBN: 978-1-83867-577-6 (Epub)

ISSN: 0731-9053 (Series)

ISOQAR
REGISTERED
Certificate Number 1985
ISO 14001

ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

INVESTOR IN PEOPLE

CONTENTS

INTRODUCTION

ECONOMETRICS OF NETWORKS

Áureo de Paula, Elie Tamer and Marcel Voia

The *Econometrics of Networks*. Volume 42 of the series Advances in Econometrics (AiE) aims at providing novel methodological and empirical research on the econometrics of networks. The volume includes both theoretical and empirical/policy papers with the specific purpose of providing an opportunity for a dialogue between academics and practitioners to better understand this new and important area of research and its role in policy discussions.

The volume is a good resource for graduate students and researchers. It includes 13 chapters covering various topics such as identification of network models, network formation, networks and spatial econometrics and applications of financial networks. One can also learn about network models with different types of interactions, sample selection in social networks, trade networks, stochastic dynamic programing in space, spatial panels, survival and networks, financial contagion, spillover effects, interconnectedness on consumer credit markets and a financial risk meter.

The book can be also a resource for data scientists and professionals from the industry as it provides a useful resource for applications, such as, for example, offering theoretical discussions about within and between groups interactions, the role unknown networks play in social interactions and the role of strong ties on trade networks.

A brief description of the chapters of the book which are grouped in four sections is presented below.

SECTION 1: IDENTIFICATION OF NETWORK MODELS

This section comprises of three chapters. The chapter by Tiziano Arduini, Eleonora Patacchini and Edoardo Rainone, "Identification and Estimation of Network Models with Heterogeneous Interactions," generalizes the standard linear-in-means model to allow for multiple types of between and within-type interactions. It extends the Bramoulle, Djebbari, and Fortin's (2009) and

Calvo-Armengol, Patacchini, and Zenou's (2009) identification conditions and Liu and Lee's (2010) estimation approach when network data are available and peer effects are heterogeneous by peer-type. The proposed methodology is inspired by Liu and Lee (2010) and Liu (2014) where the many instruments are derived from many networks (groups) observed in the sample. Differently, in the model proposed by the authors the many instruments derive from the multiple subnetwork framework. A multiple subnetwork framework does not only result in a larger number of instruments but also yields multiple approximations of the optimal instruments. The bias arising when interactions are ignored is analytically derived and evaluated in finite samples using simulation experiments. The authors show that the form of the many-instrument bias differs, though the leading order remains unchanged.

The chapter by Hon Ho Kwok, "Identification Methods for Social Interactions Models with Unknown Networks," develops a two-step identification method for social interaction models with unknown networks and discusses how the proposed methods are connected to the identification methods for models with known networks. In the first step, a linear regression is used to identify the reduced forms, while the second step decomposes the reduced forms to identify the primitive parameters. The proposed methods use panel data to identify networks. Two cases are considered: the sample exogenous vectors span $\mathbf{R}^n$ (long panels), and the sample exogenous vectors span a proper subspace of $\mathbf{R}^n$ (short panels). For the short panel case, in order to solve the sample covariance matrices' non-invertibility problem, the chapter proposes to represent the sample vectors with respect to a basis of a lower-dimensional space so that fewer regression coefficients are needed in the first step. This allows for the identification of some reduced form submatrices, which provide the equations necessary for identifying the primitive parameters.

The chapter by TszKin Julian Chan, "Snowball Sampling and Sample Selection in a Social Network," studies a snowball sampling method for social networks with endogenous peer selection. Snowball sampling is a sampling design which preserves the dependence structure of the network. It sequentially collects the information of vertices linked to the vertices collected in the previous iteration. The snowball samples suffer from a sample selection problem because of the endogenous peer selection. The chapter proposes a new estimation method that uses the relationship between samples in different iterations to correct for selection. In the application, the snowball samples collected from Facebook is used to estimate the proportion of users who support the Umbrella Movement in Hong Kong.

SECTION 2: NETWORK FORMATION

This section has two chapters. The chapter by Áureo de Paula, "Trade Networks and the Strength of Strong Ties," surveys the relevant literature on strategic formation of networks and uses it to motivate looking at questions related to the behavior of individuals in the presence of imperfect market institutions. In particular, the chapter is interested in how individuals devote resources to the establishment of reliable connections in order to attenuate the frictions that reduce trading and insurance opportunities by looking to answer questions such as:

When should we expect to see the appearance of such interpersonal networks as a stable support of economic transactions? Having been established as a stable phenomenon, does voluntary networking improve upon the situation in which no such connections can be established?

To answer these questions de Paula extends a trade network first suggested by an example in Jackson and Watts (2002) and builds a model which shows that the investment in strong ties often, though not always, produces stable configurations that manage to improve upon the imperfections of market institutions.

The author finds that such voluntary networks of "strong ties" can usually be sustained as a stable outcome, though examples are not hard to achieve in which no equilibrium configuration occurs.

Additionally, he finds that whenever such a structure exists, it improves general well-being over a situation in which only formal unreliable markets existed. And finally, the analysis suggests that though voluntary networking efforts are no substitute for an improvement in the reliability of formal institutions, emergence of informal insurance networks or extensive investment in connections should come as no surprise in the presence of "noisy" market institutions.

The chapter by Seth Richards-Shubik, "Application and Computation of a Flexible Class of Network Formation Models," discusses the empirical application of a class of strategic network formation models, using the approach to identification introduced by de Paula, Richards-Shubik and Tamer (2018). The chapter emphasizes the interplay between model specification and computational complexity and suggests approaches that help in improving empirical/computational tractability. Two detailed examples, one on friendship networks and another on coauthorship networks, are used to illustrate these issues and to demonstrate the performance of the approach with both simulation and empirical evidence.

The analysis shows how the utility specification impacts dimensionality. Additionally, the author shows how machine learning techniques can be used for dimension reduction in the coauthorship model to make the model computationally feasible while including a rich set of covariates. The chapter presents a more general estimation method, which expands the potential range of applications. Also, a statistical inference is provided with minimal computational burden.

SECTION 3: NETWORKS AND SPATIAL ECONOMETRICS

This section comprises of four chapters.

The chapter by Harry J. Paarsch and John Rust, "Implementing Faustmann–Marshall–Pressler at Scale: Stochastic Dynamic Programming in Space," constructs an intertemporal model of rent-maximizing behavior on the part of a timber harvester under potentially multi-dimensional risk as well as geographical heterogeneity. Subsequently, the authors use recursive methods (the method of stochastic dynamic programing) to characterize the optimal policy function which is the rent-maximizing timber-harvesting profile.

One feature of their application to forestry in the province of British Columbia is the unique and detailed information that is organized in the form of a dynamic geographical information system which helps to account for site-specific cost heterogeneity in harvesting and transportation, as well as an uneven-aged stand dynamics in timber growth and yield across space and time in the presence of stochastic lumber prices.

Their framework is a powerful tool, and by using it one can conduct a variety of different policy experiments. First, the authors use geography both in the planar sense and in the three-dimensional sense. Second, they consider site-specific heterogeneity both on the cost side in terms of harvesting and transportation and on the growth and yield side in terms of heterogeneous stands of timber. Third, they use best-practice biological methods to model the dynamics of uneven-aged forest growth and yield. Fourth, in the past, economists have typically demonstrated their methods by solving simple, prototypical examples in closed-form or they have imposed conditions sufficient to sign comparative static predictions. Alternatively, in this chapter the authors use recent developments in computational methods to solve numerically for the optimal policy function. Finally, they compared their optimal harvesting policy with the harvests that have occurred during the past eight years, or so, finding striking and significant differences.

The chapter by Heng Chen and Matthew Strathearn, "A Spatial Panel Model of Bank Branches in Canada," empirically analyzes the spatial bank branch network in Canada. The authors study the market structure (both industrial and geographic concentrations) via its own or adjacent postal areas. The empirical framework considers branch density (the ratio of the total number of branches to area size) by employing a spatial two-way fixed effects model.

The addition of geographic concentration, as measured by the average distance to the closest bank branch, allows the authors to reflect on the degree of spatial clustering among bank branches in a given postal area. By controlling for both industrial and geographic concentrations the authors can capture not only the degree of competitiveness in a given area, but also how well the area is serviced in terms of travel distance to the nearest branch.

The chapter finds that there are no effects associated with market structure but that there are strong spatial within and nearby effects associated with the socio-economic variables. In addition, the chapter studies the effect of spatial competition from rival banks and finds that large and small banks tend to avoid markets dominated by their competitors.

The chapter by Sophia Ding and Peter H. Egger, "Full-information Bayesian Estimation of Cross-sectional Sample Selection Models," proposes an approach to estimate cross-sectional sample selection models, where the shocks on the units of observation feature some interdependence through spatial or network autocorrelation.

There is research that aims at addressing these two issues in conjunction. The authors are showing that previous Bayesian algorithms such as the ones developed by Dogan and Taspinar (2018) do not allow for the latent variables (and the associated random shocks) to affect the latent outcome of the units whose outcome is observed. This may lead to biased parameters, in particular, of the covariances of the disturbances between the equations. This chapter improves on the prior Bayesian work on this subject by proposing a modified approach toward sampling using the multivariate-truncated, cross-sectionally dependent latent variable of the selection equation. The chapter outlines the model and implementation approach and provides simulation results documenting improved performance.

The chapter by Diego Rojas, Juan Estrada, Kim P. Huynh and David T. Jacho-Chavez, "Survival Analysis of Banknote Circulation: Fitness, Network Structure, and Machine Learning," utilizes machine learning techniques to study the distribution network patterns of over 900 million banknotes using an administrative data set from the Bank of Canada's Currency Information Management System. The data contain information regarding the printing date, physical fitness and where and when these banknotes return to the Bank of Canada's distribution centers. Having constructed networks at the region and at the financial institution (those requesting as well as depositing banknotes) level, the authors use a *K*-prototypes clustering unsupervised machine learning algorithm to classify notes into different types. This information is then used when fitting a hazard model to explain how long a banknote stays in circulation. The results show that their denominations, and not fitness measures, are the main determinants of a banknote duration in circulation after controlling for the network structure.

SECTION 4: APPLICATIONS OF FINANCIAL NETWORKS

The last section comprises of four chapters and provides applications to financial networks.

The chapter by Pablo Estrada and Leonardo Sánchez-Aragón, "Financial Contagion in Cross-holdings Networks: the case of Ecuador," applies a financial contagion model proposed by Elliott, Golub, and Jackson (2014) to a cross-shareholding network of firms in Ecuador where the nodes are the firms and the links are the cross-shareholdings among firms. A novel data set provided by SUPERCIAS is used in the analysis.

The financial contagion model uses a network of financial interdependencies among firms in a dependency matrix where each element represents the cross-shareholding. The authors study how a negative shock that affects one firm propagates through the network and generates a cascade of failures. The results show that the Ecuadorian market exhibits low levels of diversification and integration, which means that the effects of cascades cannot be amplified throughout the network. Low integration implies the presence of weak links in the network. Results also show the presence of a giant weakly connected component (40% of the total firms) because diversification is moderate suggesting that cascade effects are still weak.

Furthermore, a sensitivity analysis is conducted to determine which parameters contribute to firm's failure. When allowed the threshold, the failure cost, and the drop market value to vary, only two waves of contagion are noticeable. It was also found that two firms coming from the finance and trade industry cause the highest contagion and when a shock affects an entire industry there are more firm failures from trade and manufacturing industries than other industries. The results can be relevant for policymakers as they are better able to monitor the market and anticipate future losses.

The chapter by Edoardo Rainone, "Estimating Spillover Effects with Bilateral Outcomes," is concerned with the estimation of spillover effects when outcomes arise as a consequence of bilateral interactions instead from individual actions, in other words the analysis refers to network effects when outcomes are generated on links and not on nodes.

With the diffusion of over-the-counter (OTC) platforms and the advances in the economic theory related to networks, the chapter emphasizes the importance of assessing network effects with link-based outcomes. A link-based spatial autoregressive (SAR) model is proposed together with identification conditions and a two step least square (2SLS) estimation procedure. The author shows analytically and with Monte Carlo simulations that using a standard node-based SAR, which models nodes' instead of links' outcomes, produces misleading results when the data generating process (DGP) is link-based. The methodology is illustrated using real data from an interbank network. The results highlight that under conditions that are often met in OTC markets, modeling nodes' outcomes can lead to biased results and misleading policy implications.

The chapter by Anson T. Y. Ho, "Interconnectedness through the Lens of Consumer Credit Markets," looks at the interconnectedness between financial institutions (FIs) through the lens of consumer credits. Financial systemic risk is often assessed by the interconnectedness of FIs in terms of cross ownership, overlapping investment portfolios, interbank credit exposures and other factors. Using detailed consumer credit data in Canada, this chapter constructs a novel banking network to measure FIs' interconnectedness in consumer credit markets. Results show that FIs on average are more connected to each other over the sample period, when the interconnectedness measure increases by 21% from 2014 to 2019.

The FIs with more diversified portfolios are also more connected in the network. Participation in mortgage markets has strong positive influence on FIs' connectedness, because FIs with mortgage operations have more similar portfolios to the large FIs. Findings in this chapter highlight the importance of FIs exposure to the household sector, which may have important implications on systemic risk and the risk of multiple stress incidences across FIs. Measuring the connectedness in consumer lending networks is the first step in quantifying the potential systemic risk generated by FIs' consumer lending operations. Deeper understanding on how FIs finance their consumer lending is required to further translate the connectedness in consumer lending network into systemic risk measures.

Finally, the chapter by Andrija Mihoci, Michael Althof, Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, "FRM Financial Risk Meter," proposes a systemic risk measure that accounts for links and mutual dependencies between financial institutions utilizing tail event information. The proposed Financial Risk Meter (FRM) is based on a Lasso quantile regression and it is designed to capture tail event co-movements. The FRM focus lies on understanding active data sets characteristics and the interdependencies in a network topology.

The focus of the chapter is on two selected FRM indices, namely FRM@Americas and FRM@Europe for the equity markets, and SRM@EuroArea as an application to the asset class of government bonds. Augmenting them, for example, by simultaneously checking varieties of quantiles of FRM components, one can monitor economic activity and network dynamics, and suggest further improvements in portfolio risk management.

The chapter's findings are: first, FRM correlates positively with other measures of systemic risk and peaks around crises; second, a detailed inspection of the active set across time allows to detect the network's nodes presenting the highest risk of spillover and third, FRM is shown to predict upcoming recession periods and serves as a leading indicator for systemic risk in a variety of world regions, the US and the EU market.

Therefore, the FRM can be viewed as an early recession indicator that can help to detect distressed areas in the financial system network consisting of banks and non-banks, and thereby can help prevent spillovers into the wider financial industry. Finally, FRM can measure tail event risk, accounts for network dynamics characteristics and offers a flexible risk measuring platform.

In practice, FRM can be applied to the return time series of selected financial institutions and macroeconomic risk factors.

REFERENCES

Bramoulle, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*, 41–55.

Calvo-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, *76*(4), 1239–1267.

Chan, T. J. (2020). Snowball sampling and sample selection in a social network. In Advances in Econometrics (Vol. 42).

Chen, H., & Strathearn, M. (2020). A spatial panel model of bank branches in Canada. In Advances in Econometrics (Vol. 42).

de Paula, A. (2020). Trade networks and the strength of strong ties. In Advances in Econometrics (Vol. 42).

de Paula, A., Richards-Shubik, S., & Tamer. E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, *86*(1), 263–288.

Ding, S., & Egger, P. (2020). Full-information Bayesian estimation of cross-sectional sample selection models. In Advances in Econometrics (Vol. 42).

Dogan, O., & Taspinar, S. (2018). Bayesian inference in spatial sample selection models. *Oxford Bulletin of Economics and Statistics*, *80*(1), 90–121.

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, *104*(10), 3115–3153.

Estrada, P., & Sanchez-Aragon, L. P. (2020). Financial contagion in cross-holdings networks: The case of Ecuador. In Advances in Econometrics (Vol. 42).

Ho, A. T. Y. (2020). Interconnectedness through the lens of consumer credit markets. Advances in Econometrics (Vol. 42).

Jackson, M., & Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, *106*(2), 265–295.

Kwok, H. H. (2020). Identification methods for social interactions models with unknown networks. In Advances in Econometrics (Vol. 42).

Liu, X., & Lee, L. F. (2010). GMM estimation of social interaction models with centrality. *Journal of Econometrics*, *159*, 99–115.
Liu, X. (2014). Identification and efficient estimation of simultaneous equations network models. *Journal of Business & Economic Statistics*, *32*(4), 516–536.
Mihoci, A., Althof, M., Chen, C. Y.-H., & Hardle, W. K. (2020). FRM financial risk meter. In Advances in Econometrics (Vol. 42).
Paarsch, H. J., & Rust, J. (2020). Implementing Faustmann–Marshall–Pressler at scale: Stochastic dynamic programming in space. In Advances in Econometrics (Vol. 42).
Rainone, E. (2020). Estimating spillover effects in OTC networks. In Advances in Econometrics (Vol. 42).
Richards-Shubik, S. (2020). Application and computation of a flexible class of network formation models. In Advances in Econometrics (Vol. 42).
Rojas, D., Estrada, J., Huynh, K. P., & Jacho-Chavez, D. T. P. (2020). Survival analysis of banknote circulation: Fitness, network structure, and machine learning. In Advances in Econometrics (Vol. 42).

SECTION 1

IDENTIFICATION OF NETWORK MODELS

This page intentionally left blank

CHAPTER 1

IDENTIFICATION AND ESTIMATION OF NETWORK MODELS WITH HETEROGENEOUS INTERACTIONS

Tiziano Arduini, Eleonora Patacchini and Edoardo Rainone

ABSTRACT

The authors generalize the standard linear-in-means model to allow for multiple types with between and within-type interactions. The authors provide a set of identification conditions of peer effects and consider a two-stage least squares estimation approach. Large sample properties of the proposed estimators are derived. Their performance in finite samples is investigated using Monte Carlo simulations.

Keywords: Networks; heterogeneous peer effects; spatial autoregressive model; two-stage least squares; efficiency;

JEL classifications: C13; C21; D62

The Econometrics of Networks
Advances in Econometrics, Volume 42, 3–25

ISSN: 0731-9053/doi:10.1108/S0731-905320200000042006

1. INTRODUCTION

Interaction among agents can be modeled in several ways. When the exact topology of connections is known, one possibility is to consider the peer effects that stem from the given network structure. There is a large and growing literature on peer effects in economics using network data.[1] A popular model employed in empirical work is the linear-in-means model (Manski, 1993). Three assumptions underlie this statistical model: (i) the network is exogenous; (ii) the effects of all peers are equal; and (iii) peer status is measured without error.

This chapter is concerned with the specification and estimation of a peer effects model when *Assumption* (ii) is removed. *Assumptions* (i) *and* (iii) are maintained.[2] Specifically, we consider a model of peer effects where different types of peers are allowed to exert a different influence and where social interactions are different between and within types.

We extend the conventional identification conditions of network models (Bramoullé, Djebbari, & Fortin, 2009; Calvó-Armengol, Patacchini, & Zenou, 2009) to multiple endogenous variables and multiple adjacency matrices. We propose efficient two-stage least squares (2SLS) estimators using instruments based on the multiple reduced form approximations. We show that the standard IV approximation (Kelejian & Prucha, 1998, 1999; Liu & Lee, 2010) involves a very large number of IVs, even if we use a low degree approximation of the optimal instruments.[3] For this reason, we consider many-instrument asymptotics (Bekker, 1994) allowing the number of IVs to increase with the sample size. We also propose a many-instrument bias-correction procedure. Simulation experiments show that the bias-corrected estimator performs well in finite samples. Finally, we investigate the bias occurring when the interaction structure is misspecified. We derive analytically the bias that occurs when only within-type peer effects are considered, that is, when interactions between types are at work but ignored by the econometrician. We then use a simulation experiment to evaluate this bias in finite samples.

Our framework is a generalization of the model proposed by Arduini, Patacchini, and Rainone (2019) to study treatment effects with heterogeneous externalities. In their model connections are defined by membership in a given group and all agents within the group are connected. The availability of information on the precise structure of interactions between agents of different types offers alternative identification conditions based on the sparsity of interactions within or between agent-types.[4]

There is a long tradition in spatial econometrics looking at spatial autoregressive models with multiple endogenous variables (see Kelejian & Prucha, 2004). In the spatial econometrics context, however, the adjacency matrix is the same for all endogenous variables. The presence of different adjacency matrices provides alternative possibilities to identify the model that have not been explored so far.

Our methodology is inspired by Liu and Lee (2010) and Liu (2014). Differently from their approach where the many instruments derive from the many networks (groups) observed in the sample, in our model the many instruments derive from the multiple subnetwork framework. A multiple subnetwork framework does not only result in an increasing number of instruments but also yields multiple approximations of the optimal instruments. We show that the form of the many-instrument bias differs, though the leading order remains unchanged.[5]

This chapter is organized as follows. The next section introduces the econometric model. Identification conditions are given in Section 3, and in Section 4 we consider 2SLS estimation for the model. Section 5 investigates the bias occurring when the interaction structure is misspecified. Section 6 concludes.

2. THE NETWORK MODEL WITH HETEROGENEOUS PEER EFFECTS

Suppose the n observations in the data are partitioned into $\bar{r}$ networks, with n_r agents in the rth network. For the rth network, let

$$Y_r = \phi G_r Y_r + X_r \beta + G_r X_r \gamma + \varepsilon_r, \tag{1}$$

where $Y_r = (y_{1r}, \ldots, y_{nr})'$ is an n_r-dimensional vector of outcomes, $G_r = [g_{ij,r}]$ is an $n_r \times n_r$ adjacency matrix, $g_{ij,r}$ is equal to 1 if i and j are connected, 0 otherwise. X_r is a $n_r \times p$ matrix of exogenous variables capturing individual characteristics. For ease of presentation, we assume G_r and X_r are non-stochastic. $\varepsilon_r = (\varepsilon_{1r}, \ldots, \varepsilon_{nr})'$ is a vector of errors whose elements are i.i.d. with 0 mean and variance σ_r^2 for all i.[6]

For model (1), ϕ represents *the endogenous effect*, where an agent's choice/outcome may depend on those of his/her peers on the same activity, and γ represents *the contextual effect*, where an agent's choice/outcome may depend on the exogenous characteristics of his/her peers.

Let us suppose there is a finite number of types of agents in the population. For simplicity, let us consider two types a and b. A and B are thus sets that partition $\{1,\ldots, n\}$. The n_a and n_b denote the cardinalities of A and B. Let $X_r^* = (X_r, G_r X_r)$ and $\beta = (\beta', \gamma')'$.

Let us now define $Y_r = \left(Y_{a,r}', Y_{b,r}'\right)'$, $X = \left(X_{a,r}', X_{b,r}'\right)'$, and $G_r = \begin{bmatrix} G_{a,r} & G_{ab,r} \\ G_{ba,r} & G_{b,r} \end{bmatrix}$, where $G_{a,r}$ is formed only among nodes of type A and $G_{ab,r}$ keeps trace of links from b to a.[7] Regularity conditions are listed in Appendix 1. Model (1) can be generalized in the following way:

$$Y_{a,r} = \phi_a G_{a,r} Y_{a,r} + \phi_{ab} G_{ab,r} Y_{b,r} + X_{a,r}^* \beta_a^* + G_{ab,r} X_{b,r} \gamma_{ab} + \varepsilon_{a,r}, \tag{2}$$

$$Y_{b,r} = \phi_b G_{b,r} Y_{b,r} + \phi_{ba} G_{ba,r} Y_{a,r} + X_{b,r}^* \beta_b^* + G_{ba,r} X_{a,r} \gamma_{ba} + \varepsilon_{b,r}, \tag{3}$$

where $\beta_a^* = \left(\beta_a', \gamma_a'\right)'$, $X_{a,r}^* = \left(X_{a,r}, G_{a,r} X_{a,r}\right)$, $X_{b,r}^* = \left(X_{b,r}, G_{b,r} X_{b,r}\right) \beta_b^* = \left(\beta_b', \gamma_b'\right)'$, and $\varepsilon_{a,r}$ and $\varepsilon_{b,r}$ are i.i.d. errors with variance σ_a^2 and σ_b^2, respectively. Let us suppose for simplicity that $\sigma_a^2 = \sigma_b^2 = \sigma^2$. If we stack up equations (2) to (3) and restrict the endogenous effect parameters of the two equations to be the same (i.e., $\phi_a = \phi_b = \phi_{ab} = \phi_{ba}$), then we obtain model (1). Let us define the following matrices

$$A_m\delta_a = X^*_{a,r}\beta^*_a + G_{ab,r}X_{b,r}\gamma_{ab} + \varepsilon_{a,r},$$
$$B_m\delta_b = X^*_{b,r}\beta^*_b + G_{ba,r}X_{a,r}\gamma_{ba} + \varepsilon_{b,r},$$

where $A_m = (X_{a,r},\ G_{ab,r}X_{b,r},\ \varepsilon_{a,r})$, $\delta_a = \left(\beta^{*\prime}_a, \gamma'_{ab}, 1\right)'$, $B_m = (X_{b,r},\ G_{ba,r}\ X_{a,r},\ \varepsilon_{b,r})$ and $\delta_b = \left(\beta^{*\prime}_b, \gamma'_{ba}, 1\right)$. By plugging $Y_{b,r}$ in equation (2) we have

$$\begin{aligned} Y_{a,r} &= \phi_a G_a Y_{a,r} + \phi_{ab} G_{ab,r} Y_r \left(J_{b,r}\left(\phi_{ba} G_{ba,r} Y_{a,r} + B_m\delta_b\right)\right) + A_m\delta_a \\ &= \left(\phi_a G_{a,r} + \phi_{ab}\phi_{ba} C_{a,r}\right) Y_a + \phi_{ab} G_{ab,r} J_{b,r} B_m \delta_b + A_m\delta_a \end{aligned} \tag{4}$$

where $J_{b,r} = \left(I - \phi_b G_{b,r}\right)^{-1} = \sum_{\kappa=1}^{\infty}\left(\phi_b G_{b,r}\right)^{\kappa}$ provided that $\| \phi_b G_{b,r} \|_\infty < 1$, where $\| \cdot \|_\infty$ is the row-sum matrix norm. The *ij*th element of $J_{b,r}$ sums all *k*-distance paths from *j* to *i* when $i, j \in B$ scaling them by ϕ^k_b and $C_{a,r} = G_{ab,r}J_{b,r}G_{ba,r}$.[8] Therefore the reduced form of model (2) is

$$Y_{a,r} = M_{a,r}\left(\phi_{ab} G_{ab,r} J_{b,r} B_m \delta_b + A_m\delta_a\right) \tag{5}$$

where $M_{a,r} = (I - \phi_a G_{a,r} - \phi_{ab}\phi_{ba} C_{a,r})^{-1}$.[9] A sufficient condition for the non-singularity of $\left(I - \phi_a G_{a,r} - \phi_{ab}\phi_{ba} C_{a,r}\right)$ is $\| \phi_a G_{a,r} \|_\infty + \| \phi_{ab}\phi_{ba} C_{a,r} \|_\infty \leq 1$. This condition also implies that $M_{a,r}$ is uniformly bounded in absolute value.[10]

We note that: (i) we present an aggregate model specification (i.e., *G* which multiplies *y* in model (1) is not row-normalized), but the approach also applies to an average model (i.e., when *G* which multiplies *y* in model (1) is row-normalized)[11]; (ii) our model specification has two types, but all the assumptions, propositions and proofs can be naturally extended to a finite number of types; (iii) we consider a single network, but the approach can be easily extended to the case of multiple networks (i.e., a network with several components) with the addition of network fixed effects in the model specification[12]; and (iv) we can also add a heterogeneous spatial lag in the error term $\varepsilon_a = \rho_a W_a \varepsilon_a + \rho_{ab} W_{ab} \varepsilon_b$.[13]

3. IDENTIFICATION

For notational convenience, from now on we omit the network observation index. Let us define $Z_a = \left(G_a Y_a, G_{ab,r} Y_b, X^*_a, G_{ab} X_b\right)$. Endogenous effects in equation (2) are identified if $E(Z_a)$ has full column rank for large *n*.[14] In this section, we find sufficient conditions for $E(Z_a)$ to have full column rank.[15] The detailed proof is given in Appendix 3. Here, identification means that a consistent estimator of the parameters of model (2) exists.

Proposition 1. Let X_a and X_b have full column rank. If M_a, M_b, J_a and J_b are invertible,[16] then $E(Z_a)$ has full column rank in the following cases

1. (a) i. $\beta_a\phi_a + \gamma_a \neq 0$,
   ii. I_a, G_a and G_a^2 are linearly independent.
   [and]
   (b) i. $\beta_b\phi_b + \gamma_b \neq 0$,
   ii. G_{ab} and $G_{ab}G_b$ are linearly independent.
   [or]
2. (a) i. $\gamma_{ab} \neq 0$,
   ii. G_{ab} and G_aG_{ab} are linearly independent.
   [and]
   (b) i. $\gamma_{ba} \neq 0$,
   ii. I_a, G_a and $G_{ab}G_{ba}$ are linearly independent.

Proposition 1 generalizes the identification conditions in Bramoullé et al. (2009). Note that conditions (*1a*) are exactly the same identification conditions found by Bramoullé et al. (2009) in the case of homogeneous effects (i.e., only one type). *Proposition 1* here is more general as it provides alternative possibilities. When more than one type is considered we do not need linear independence of a particular set of matrices – we have multiple sufficient conditions. Even if I_a, G_a and G_a^2 are linearly dependent we can still identify ϕ_a, and the other parameters, relying on linear independence of network paths passing through type B nodes.[17] The set of adjacency matrices' combinations can be represented as a *Tree-indexed Markov chain* – the parameters can be identified because of the multiple branches of the tree (see Appendix 3). Obviously, if G_a, G_{ba}, G_{ab} and G_b are complete and consequently all products among them are linearly dependent, then the parameters of the model remain not identified. However, if type A nodes are in a complete network, but the matrices representing between-type interactions are sparse (i.e., G_{ab} and G_{ba} are not complete), then identification can be achieved and ϕ_a can be estimated even if G_a is complete. Systems in panels (b) and (c) in Fig. 1 can be

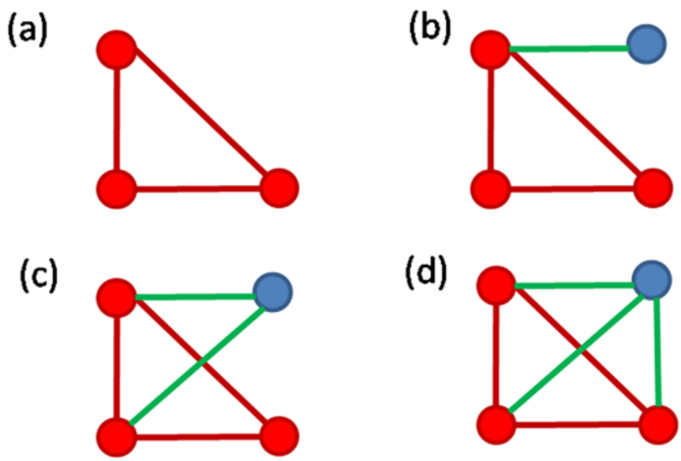

Fig. 1. Identification with Heterogeneous Nodes.

identified because the adjacency matrix of type B nodes (blue nodes in Fig. 1) is sparse, whereas systems in panels (a) and (d) cannot. The additional parameters' restrictions (condition (*1b, 2a* or *2b*)) are due to an additional vector in the full rank condition (i.e., $E(G_{ab}y_b)$).

Proposition 1 has a natural interpretation in terms of instrumental variables. A multiple type framework adds an extra layer of exclusion restrictions. In fact, multiple sets of matrices provide additional instruments. The intuition is that when we distinguish nodes in different types, a higher number of possible network intransitivities are formed. Appendix 2 provides technical details on the connection between identification in a single type model and a multiple type one.

4. THE 2SLS ESTIMATOR

We consider 2SLS estimation for the model in the spirit of Liu and Lee (2010). Following the standard technique used in spatial econometrics literature, we have the following optimal instruments from the two (symmetric) reduced forms

$$E(G_a y_a) = G_a\left(M_a\left(\phi_{ab}G_{ab}J_b E(B_m)\delta_b + E(A_m)\delta_a\right)\right) \tag{6}$$

$$E(G_{ab} y_b) = G_{ab}\left(M_b\left(\phi_{ba}G_{ba}J_a E(A_m)\delta_b + E(B_m)\delta_a\right)\right) \tag{7}$$

Recalling that $Z_a = [G_a y_a,\ G_{ab}y_b,\ E(A_m)]$ is a $n \times (k+2)$ matrix, we have $f_a = E(Z_a) = [E(G_a y_a),\ E(G_{ab}y_b),\ E(A_m)]$. Therefore, from equations (6) and (7) we have

$$Z_a = f_a + \upsilon_a = f_a + \left[\left(\phi_{ab}S_a G_{ab}J_b\varepsilon_a + S_a\varepsilon_a\right),\left(\phi_{ba}S_{ab}G_{ba}J_a\varepsilon_a + S_{ab}\varepsilon_b\right)\right][e_1,e_2]',$$

where e_1 is a first unit vector of dimension $(k+2)$, $S_a = G_a M_a$ and $S_{ab} = G_{ab}M_b$. These instruments are infeasible given the embedded unknown parameters. f_a can be considered a linear combination of IVs in $H_\infty^* = \left(S_a\left(G_{ab}J_bE(B_m),\ E(A_m)\right), S_{ab}\left(G_{ba}J_aE(A_m),E(B_m)\right),E(A_m)\right)$. Furthermore, since $S_a = G_a M_a$ and $S_{ab} = G_{ab}M_b$ provided that $\|\phi_a G_a\|_\infty + \|\phi_{ab}\phi_{ba}C_a\|_\infty \le 1$ and $\|\phi_b G_b\|_\infty < 1$, we have $S_a = G_a\sum_{j=0}^{\infty}(\phi_a G_a + \phi_{ab}\phi_{ba}C_a)^j = G_a\sum_{j=0}^{\infty}\left(\phi_a G_a + \phi_{ab}\phi_{ba}G_{ab}\sum_{j=0}^{\infty}\left(\phi_b^j G_b^j\right)G_{ba}\right)^j$. The same approximation holds for S_{ab}. It follows that

$$C_a = G_{ab}J_bG_{ab} = G_{ab}(\sum_{j=0}^{\infty}\phi_b^j G_b^j)G_{ba} = G_{ab}(\sum_{j=0}^{\infty}\phi_b^j G_b^j + (\phi_b G_b)^{p+1}J_b)G_{ba}.$$

This implies $\| C_a\sum_{j=0}^{p}\phi_b^j G_b^j \|_\infty \le \|(\phi_b G_b)^{p+1}\|_\infty \| C_a \|_\infty = o(1)\ \ as\ \ p\to\infty.$

$$S_a = G_aM_a = G_a\sum_{j=0}^{\infty}(\phi_aG_a + \phi_{ab}\phi_{ba}C_a)^j = G_a\left[\sum_{j=0}^{p}(\phi_aG_a + \phi_{ab}\phi_{ba}C_a)^j + (\phi_aG_a + \phi_{ab}\phi_{ba}C_a)^{p+1}S_a\right]\rightarrow$$