

# Identifying network ties from panel data: Theory and an application to tax competition

---

Áureo de Paula  
Imran Rasul  
Pedro CL Souza

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP21/23

# Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition\*

Áureo de Paula

Imran Rasul

Pedro CL Souza<sup>†</sup>

October 2023

## Abstract

Social interactions determine many economic behaviors, but information on social ties does not exist in most publicly available and widely used datasets. We present results on the identification of social networks from observational panel data that contains no information on social ties between agents. In the context of a canonical social interactions model, we provide sufficient conditions under which the social interactions matrix, endogenous and exogenous social effect parameters are globally identified if networks are constant over time. We also provide an extension of the method for time-varying networks. We then describe how high-dimensional estimation techniques can be used to estimate the interactions model based on the Adaptive Elastic Net Generalized Method of Moments. We employ the method to study tax competition across US states. The identified social interactions matrix implies that tax competition differs markedly from the common assumption of competition between geographically neighboring states, providing further insights into the long-standing debate on the relative roles of factor mobility and yardstick competition in driving tax setting behavior across states. Most broadly, our identification and application show that the analysis of social interactions can be extended to economic realms where no network data exists.

*JEL Classification: C31, D85, H71.*

---

\*We gratefully acknowledge financial support from the ESRC through the Centre for the Microeconomic Analysis of Public Policy (ES/T014334/1), the Centre for Microdata Methods and Practice (RES-589-28-0001) and the Large Research Grant ES/P008909/1 and from the ERC (SG338187). We thank Edo Airolidi, Luis Alvarez, Michele Aquaro, Oriana Bandiera, Larry Blume, Yann Bramoullé, Stephane Bonhomme, Vasco Carvalho, Gary Chamberlain, Andrew Chesher, Christian Dustmann, Sérgio Firpo, Jean-Pierre Florens, Eric Gautier, Giacomo de Giorgi, Matthew Gentzkow, Stefan Hoderlein, Bo Honoré, Matt Jackson, Dale Jorgensen, Christian Juliard, Maximilian Kasy, Miles Kimball, Thibaut Lamadon, Simon Sokbae Lee, Arthur Lewbel, Tong Li, Xiadong Liu, Elena Manresa, Charles Manski, Marcelo Medeiros, Angelo Mele, Francesca Molinari, Pepe Montiel, Andrea Moro, Whitney Newey, Ariel Pakes, Eleonora Pattachini, Michele Pelizzari, Martin Pesendorfer, Christiern Rose, Adam Rosen, Bernard Salanie, Olivier Scaillet, Sebastien Sieglösch, Pasquale Schiraldi, Tymon Sloczynski, Kevin Song, John Sutton, Adam Szeidl, Thiago Tachibana, Elie Tamer, and seminar and conference participants for valuable comments. We also thank Tim Besley and Anne Case for comments and sharing data. Daniel Barbosa provided outstanding research assistance. A previous version of this paper was circulated as “Recovering Social Networks from Panel Data: Identification, Simulations and an Application.” All errors remain our own. Codes are available on Zenodo (<https://zenodo.org/XXX>) and github (<https://github.com/YYY>) repositories.

<sup>†</sup>de Paula: University College London, CeMMAP and IFS, a.paula@ucl.ac.uk; Rasul: University College London and IFS, i.rasul@ucl.ac.uk; Souza: Queen Mary University, p.souza@qmul.ac.uk.

# 1 Introduction

In many economic environments, behavior is shaped by social interactions between agents. In individual decision problems, social interactions have been key to understanding outcomes as diverse as educational test scores, the demand for financial assets, and technology adoption (Sacerdote, 2001; Bursztyn *et al.*, 2014; Conley and Udry, 2010). In macroeconomics, the structure of firms' production and credit networks propagate shocks, or help firms to learn (Acemoglu *et al.*, 2012; Chaney, 2014). In political economy and public economics, ties between jurisdictions are key to understanding tax setting behavior (Tiebout, 1956; Shleifer, 1985; Besley and Case, 1994).

Underpinning all these bodies of research is some measurement of the underlying social ties between agents. However, information on social ties does not exist in most publicly available and widely used datasets. To overcome this limitation, studies of social interaction either *postulate* ties based on common observables or homophily, or *elicit* data on networks. However, it is increasingly recognized that postulated and elicited networks remain imperfect solutions to the fundamental problem of missing data on social ties, because of econometric concerns that arise with either method, or simply because of the cost of collecting network data.<sup>1</sup>

Two consequences are that (i) the classes of problems in which social interactions occur are understudied, because social networks data is missing or too costly to collect; and (ii) there is no way to validate social interactions analysis in contexts where ties are postulated. In this paper, we tackle this challenge by deriving sufficient conditions under which global identification of the *entire structure* of social networks is obtained, using only observational panel data that itself contains *no* information on network ties. Our identification results allow the study of social interactions without data on social networks, and the validation of structures of social interaction where social ties have hitherto been postulated. The recovered networks are economically meaningful to explain the effects under study, since they are entirely estimated from the data itself, and not driven by *ex ante* assumptions on how individuals interact.

A researcher is assumed to have panel data on individuals  $i = 1, \dots, N$  for instances  $t = 1, \dots, T$ . An instance refers to a specific observation for  $i$  and need not correspond to a time period (for example, if  $i$  refers to a firm,  $t$  could refer to market  $t$ ). The outcome of interest for individual  $i$  in

---

<sup>1</sup>As detailed in de Paula (2017), elicited networks are often self-reported and can introduce error to the outcome of interest. Network data can be censored if only a limited number of links can feasibly be reported. Incomplete survey coverage of nodes in a network may lead to biased aggregate network statistics. Chandrasekhar and Lewis (2016) show that even when nodes are randomly sampled from a network, partial sampling leads to non-classical measurement error and biased estimation. Collecting social network data is also a time- and resource-intensive process. In response to these concerns, a nascent strand of literature explores cost-effective alternatives to full elicitation to recover aggregate network statistics (Breza *et al.*, 2020).

instance  $t$  is  $y_{it}$  and is generated according to a canonical structural model of social interactions:<sup>2</sup>

$$y_{it} = \rho_0 \sum_{j=1}^N W_{0,ij} y_{jt} + \beta_0 x_{it} + \gamma_0 \sum_{j=1}^N W_{0,ij} x_{jt} + \alpha_i + \alpha_t + \epsilon_{it}. \quad (1)$$

Outcome  $y_{it}$  depends on the outcomes of other individuals to whom  $i$  is socially tied,  $y_{jt}$ , and  $x_{jt}$  includes characteristics of those individuals.<sup>3</sup>  $W_{0,ij}$  measures how the outcome and characteristics of  $j$  causally impact the outcome for  $i$ . The network is initially assumed to be fixed over time, and we later provide an extension of the method for time-varying networks. As outcomes for all individuals obey equations analogous to (1), the system of equations can be written in matrix notation, where the structure of interactions is captured by the adjacency matrix, denoted by  $W_0$ . Our approach allows for unobserved heterogeneity across individuals  $\alpha_i$  and common shocks to individuals  $\alpha_t$ . This framework encompasses a classic linear-in-means specification as in Manski (1993). In his terminology,  $\rho_0$  and  $\gamma_0$  capture endogenous and exogenous social effects, and  $\alpha_t$  captures correlated effects. The distinction between endogenous and exogenous peer effects is critical, as only the former generates social multiplier effects. In line with the literature, we maintain that the same  $W_0$  governs the structure of both endogenous and exogenous effects. We later discuss relaxing this assumption when more than one regressor is used.

Manski’s seminal contribution set out the reflection problem of separately identifying endogenous, exogenous, and correlated effects in linear models. However, it has been somewhat overlooked that he also set out another challenge in the identification of the social network in the first place.<sup>4</sup> This is the problem we tackle, and thus, we expand the scope of identification beyond  $\rho_0$ ,  $\beta_0$ , and  $\gamma_0$ . Our point of departure from much of the literature is therefore to presume  $W_0$  is *entirely unknown* to the researcher. We derive sufficient conditions under which all the entries in  $W_0$ , and the endogenous and exogenous social effect parameters,  $\rho_0$  and  $\gamma_0$ , are globally identified from “reduced form” parameters. By identifying the social interactions matrix  $W_0$ , our results allow the recovery of aggregate network characteristics, such as the degree distribution and patterns of homophily, as well as node-level statistics such as the strength of social interactions between nodes, and the centrality of nodes. Such aggregate and node-level statistics often map back to underlying models of social interaction (Ballester *et al.*, 2006; Jackson *et al.*, 2017; de Paula, 2017).

---

<sup>2</sup>Blume *et al.* (2015) present micro-foundations for this estimating equation based on non-cooperative games of incomplete information for individual choice problems.

<sup>3</sup>In the case in which  $t$  is considered to be a time period,  $x_{it}$  may also include lagged values of  $y_{it}$ .

<sup>4</sup>Manski (1993) highlights difficulties (and potential restrictions) in identifying  $\rho_0$ ,  $\beta_0$  and  $\gamma_0$  when *all* individuals interact with each other, and when this is observed by the researcher. In (1), this corresponds to  $W_{0,ij} = N^{-1}$ , for  $i, j = 1, \dots, N$ . At the same time, he states (p. 536), “I have presumed that researchers know how individuals form reference groups and that individuals correctly perceive the mean outcomes experienced by their supposed reference groups. There is substantial reason to question these assumptions (...) If researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether observed behavior can be used to infer these unknowns (...) The conclusion to be drawn is that informed specification of reference groups is a necessary prelude to analysis of social effects.”

Our identification strategy is new and fundamentally different from those employed elsewhere in the literature and does not rely on requirements about network sparsity. However, it delivers sufficient conditions that are mild and relate to existing results on the identification of social effects parameters when  $W_0$  is known (Bramoullé *et al.*, 2009; De Giorgi *et al.*, 2010; Blume *et al.*, 2015). The intuition for our identification result is simple: model (1) has  $N^2$  reduced-form parameters, and there are  $N(N - 1) + 3$  structural unknowns (as no unit affects itself, so  $W_{0,ii} = 0$ ). So there are more equations than unknowns if  $N \geq 2$ , and we demonstrate those can be solved for the parameters of interest under the assumptions we invoke. Our identification result is also useful in other estimation contexts, such as when a researcher has partial knowledge of  $W_0$ ,<sup>5</sup> or in navigating between priors on reduced-form and structural parameters in a Bayesian framework (see, e.g., Gefang *et al.*, 2023), thus avoiding issues the raised by Kline and Tamer (2016).

Global identification is a necessary requirement for consistency of extremum estimators such as those based on the GMM (Hansen 1982; Newey and McFadden 1994). Our identification analysis provides primitives for this condition. To estimate the model, we employ the Adaptive Elastic Net GMM method (Caner and Zhang, 2014), as this allows us to deal with a potentially high-dimensional parameter vector (in comparison to the time dimension in the data) including all the entries of the social interactions matrix  $W_0$ , although other estimation protocols may also be entertained (e.g. using Bayesian methods or *a priori* information).<sup>6</sup>

We showcase the method using Monte Carlo simulations based on stylized random network structures as well as real-world networks. In each case, we take a fixed network structure  $W_0$  and simulate panel data as if the data generating process were given by (1). We then apply the method to the simulated panel data to recover estimates of all elements in  $W_0$ , as well as the endogenous and exogenous social effect parameters  $(\rho_0, \gamma_0)$ . The networks considered vary in size, complexity, and their aggregate and node-level features. In small samples, we find that the majority of links are identified even for  $T = 5$ , and the proportion of true non-links (zeros in  $W_0$ ) captured correctly as zeros is over 85% even when  $T = 5$ . Of course, there are important limitations to the use of the method in small- $T$  cases. Biases are expected and manifest themselves in two ways. First, weak links can be shrunk to zero, and the strength of strong edges can be overestimated. Second, the estimates of  $\rho$  and  $\gamma$  can suffer from small-sample bias, being analogous to well-known results for autoregressive time series models. Both properties rapidly improve with  $T$ . For instance, biases in the estimation of endogenous and exogenous effects parameters  $(\hat{\rho}, \hat{\gamma})$  fall quickly with  $T$  and are close to zero for large sample sizes. The endogenous and exogenous social effects are also

---

<sup>5</sup>One such example is the nascent literature of Aggregate Relational Data (ARD) as in Breza *et al.* (2020). Another possibility is that individuals are known to belong to subgroups, so  $W_0$  is block diagonal.

<sup>6</sup>The Elastic Net was introduced by Zou and Hastie (2005) in part to circumvent difficulties faced by alternative estimation protocols (e.g., LASSO) when the number of parameters,  $p$ , exceeds the number of observations,  $n$  (where  $p$  and  $n$  follow the notation in that paper). Whereas the theoretical results on the large-sample properties of elastic net estimators usually have not exploited sparsity, several articles have demonstrated their performance in data scenarios where this occurs. In Section 3, we provide an informal discussion on the performance in our context.

correctly captured as  $T$  increases. *A fortiori*, we estimate aggregate and node-level statistics of each network, demonstrating the accurate recovery of key players in networks, for example.

In the final part of our analysis, we apply the method to shed new light on a classic real-world social interactions problem: tax competition between US states. The literatures in political economy and public economics have long recognized the behavior of state governors might be influenced by decisions made in “neighboring” states. The typical empirical approach has been to postulate the relevant neighbors as being geographically contiguous states. Our approach allows us to infer the set of “economic” neighbors determining social interactions in tax setting behavior from panel data on outcomes and covariates alone. In this application, the panel data dimensions cover mainland US states,  $N = 48$ , for the years 1962-2015,  $T = 53$ .

The identified network structure of tax competition differs markedly from the assumption of competition between geographic neighbors. The identified economic network has fewer edges, and we identify non-adjacent states that influence tax setting behaviors. Differences in the structure of the identified economic and geography-based networks are reflected in the far lower clustering coefficient in the former (.042 versus .419). With the recovered social interactions matrix we establish, beyond geography, which covariates correlate to the existence of ties between states and so shed new light on hypotheses for social interactions in tax setting: factor mobility and yardstick competition (Tiebout, 1956; Shleifer, 1985; Besley and Case, 1994). The identified network highlights significant predictors of tax competition between states beyond distance: political homophily *reduces* the likelihood of a link, suggesting any yardstick competition driving social interactions occurs when voters compare their governor to those of the opposing party in other states. Tax haven states appear to be less influential in tax setting behaviors, easing concerns over a race-to-the-bottom in tax setting. Labor mobility between states does not robustly predict the existence of economic ties between states in tax setting behavior.

Given the relatively long study period in this application, at a final stage of analysis we extend our method to allow the strength of social interactions in tax competition ( $\rho_0, \gamma_0$ ) and the structure of links in the economic network ( $W_0$ ) to vary over time as we change the weight placed on observations from any given time period. We document the gradual increase in strength of social interactions over time, and the changing nature of the network of interactions. We utilize these findings to conduct counterfactual simulations of the general equilibrium propagation of tax shocks from a given state to all other mainland US states, and how these general equilibrium effects of the same policy shock vary as we place weight on observations later in our study period.

Our paper contributes to the literature on the identification of social interactions models. The first generation of papers studied the case where  $W_0$  is known, so only the endogenous and exogenous social effects parameters needed to be identified. It is now established that if the known  $W_0$  differs from the linear-in-means example where all units are linked with equal weights,  $\rho_0$  and  $\gamma_0$  can be identified (Bramoullé *et al.*, 2009; De Giorgi *et al.*, 2010). Intuitively, identification in those cases can use peers-of-peers, are not necessarily connected to individual  $i$  and can be used

to leverage variation from exclusion restrictions in (1), or can use groups of different sizes within which all individuals interact with each other (Lee, 2007). Bramoullé *et al.* (2009) show these conditions are met if  $I$ ,  $W_0$ , and  $W_0^2$  are linearly independent, which is shown to hold generically by Blume *et al.* (2015). However, as made precise in Section 2, the linear algebraic arguments employed by Bramoullé *et al.* (2009) or Blume *et al.* (2015) do not apply when  $W_0$  is unobserved, and other arguments have to be used instead.<sup>7</sup>

Blume *et al.* (2015) investigate the case when  $W_0$  is *partially* observed and show that if two individuals are *known* not to be directly connected, the parameters of interest in a model related to (1) can be identified. Blume *et al.* (2011) take an alternative approach: suggesting a parameterization of  $W_0$  according to a pre-specified distance between nodes. We do not impose such restrictions, but note that partial observability of  $W_0$  or placing additional structure on  $W_0$  is complementary to our approach, as it reduces the number of parameters in  $W_0$  to be retrieved. Bonaldi *et al.* (2015) and Manresa (2016) estimate models like (1) when  $W_0$  is not observed, but where  $\rho_0$  is set to zero so there are no endogenous social effects. They use sparsity-inducing methods from the statistics literature, but the presence of  $\rho_0$  in our case complicates identification because it introduces issues of simultaneity that we address.<sup>8</sup>

Rose (2015) also presents related identification results for linear models like (1), assuming the sparsity of the neighborhood structure. Intuitively, given two observationally equivalent systems, sparsity guarantees the existence of pairs that are not connected in either. Since observationally equivalent systems are linked via the reduced-form coefficient matrix, this pair allows one to identify certain parameters in the model. Having identified those parameters, Rose (2015) shows that one can proceed to identify other aspects of the structure (see also Gautier and Rose, 2016). This is related to the ideas in Blume *et al.* (2015), who show identification results can be leveraged if individuals are *known* not to be connected. Our main identification results do *not* rely on properties of sparse networks, and make use of plausible and intuitive conditions, whereas the auxiliary rank conditions necessary may be computationally complex to verify. More recently, Lewbel *et al.* (2022) propose an estimation strategy for the parameters  $\rho_0$ ,  $\beta_0$ , and  $\gamma_0$  of model (1) in the absence of network links if many different groups can be observed. Battaglini *et al.* (2022) estimate a structural model specifically for the case of unobserved social connections in the US Congress.

Finally, in the statistics literature, Lam and Souza (2020) study the penalized estimation of model (1) when  $W_0$  is not observed, assuming the model and social interactions are identified. The statistical literature on graphical models has investigated the estimation of neighborhoods

---

<sup>7</sup>Alternative identification approaches when  $W_0$  is known focus on higher moments (variances and covariances across individuals) of outcomes (de Paula, 2017) and rely on additional restrictions on higher moments of  $\epsilon_{it}$ . Note that (1) is a spatial autoregressive model. In that literature,  $W_0$  is also typically assumed to be known (Anselin, 2010).

<sup>8</sup>Manresa (2016) allows for unit-specific  $\beta_0$  parameters. While in many applications those are taken to be homogeneous, we also discuss extensions on how heterogeneity in those parameters can be handled when  $\rho_0 \neq 0$  in Appendix B.

defined by the covariance structure of the random variables at hand (Meinshausen and Buhlmann, 2006). This corresponds to a model where  $y_t = (I - \rho_0 W_0)^{-1} \epsilon_t$  is jointly normal (abstracting from covariates). On a graph with  $N$  nodes corresponding to the variables in the model, an edge between two nodes (variables)  $i$  and  $j$  is absent when these two variables are conditionally independent given the other nodes. In the model above, the inverse covariance matrix is  $(I - \rho_0 W_0)^\top \Sigma_\epsilon^{-1} (I - \rho_0 W_0)$ , where  $\Sigma_\epsilon$  is the variance covariance structure for  $\epsilon_t$ . The discovery of zero entries in this matrix is not equivalent to the identification of  $W_0$  and involves  $\Sigma_\epsilon$  (as do identification strategies using higher moments when  $W_0$  is known).<sup>9</sup>

We build on these papers by studying the problem where  $W_0$  is potentially entirely unknown to the researcher. In so doing, we open up the study of social interactions to realms where social network data does not exist. In our case, we consider the definition of the network as the one that mediates, together with the variables  $x_{it}$ , the outcome process  $y_{it}$  according to Equation (1). The identified network may be a combination of elicited types of social interactions – such as friendship formation, lending and borrowing relations, links with relatives – or different from elicited data, as long as the links are relevant in determining the outcomes. In our case, and in line with the literature, the network ties  $W_{ij}$  are considered to be deterministic parameters or predetermined. Alternatively, the networks are assumed to be the outcome of a stochastic process, such as the latent space model (Hoff *et al.*, 2002; Breza *et al.*, 2020) or Exponential Random Graphs models (Holland and Leinhardt, 1981).

Our conclusions discuss how our approach can be modified, and assumptions weakened, to integrate partial knowledge of  $W_0$ . We discuss further applications and the steps required to simultaneously identify models of network formation and the structure of social interactions. The practical use of our proposed method has already been demonstrated in applications. For example, Fetzer *et al.* (2021) study the impact on conflict of the transition of security responsibilities between international and Afghan forces. Our proposed method is used to control for violation of SUTVA-type hypotheses that might occur because of spillover and displacement effects of insurgent forces across districts. Since the pattern of displacement is unobserved – and, in fact, insurgents have incentives to obfuscate their strategy – the current method is applied to fully recover the network and bound the effects of the end of the military occupation on conflict.<sup>10</sup>

We proceed as follows. Section 2 presents our core result: the sufficient conditions under which the social interactions matrix, endogenous and exogenous social effects are globally identified. Section 3 describes the high-dimensional techniques used for estimation based on the Adaptive Elastic Net GMM method and presents simulation results from stylized and real-world networks. Section 4 applies our methods to study tax competition between US states. Section 5 concludes. The Appendix provides proofs and further details on estimation and simulations.

---

<sup>9</sup>Meinshausen and Buhlmann (2006)’s and Lam and Souza (2020)’s neighborhood estimates rely on (penalized) regressions of  $y_{it}$  on  $y_{1t}, \dots, y_{i-1,t}, y_{i+1,t}, \dots, y_{N,t}$ , which do not address the endogeneity in estimating  $W_0$ .

<sup>10</sup>Zhou (2019) applies our identification results, focusing on unobserved networks with grouped heterogeneity, to suggest a nonlinear least squares procedure for estimation on a single network observation.



## 2 Identification

### 2.1 Setup

Consider a researcher with panel data covering  $i = 1, \dots, N$  individuals repeatedly observed over  $t = 1, \dots, T$  instances. The number of individuals  $N$  in the network is fixed but potentially large. The aim is to use this data to identify a social interactions model with no data on actual social ties. For expositional ease, we first consider identification in a simpler version of the canonical model in (1), where we drop individual-specific ( $\alpha_i$ ) and time-constant fixed effects ( $\alpha_t$ ) and assume  $x_{it}$  is a one-dimensional regressor for individual  $i$  and instance  $t$ . We later extend the analysis to include individual-specific, time-constant fixed effects and allow for multidimensional covariates  $x_{k,it}$ ,  $k = 1, \dots, K$ . We adopt the subscript “0” to denote parameters generating the data, and non-subscripted parameters are generic values in the parameter space:

$$y_{it} = \rho_0 \sum_{j=1}^N W_{0,ij} y_{jt} + \beta_0 x_{it} + \gamma_0 \sum_{j=1}^N W_{0,ij} x_{jt} + \epsilon_{it}. \quad (2)$$

As the outcomes for all individuals  $i = 1, \dots, N$  obey equations analogous to (2), the system of equations can be more compactly written in matrix notation as:

$$y_t = \rho_0 W_0 y_t + \beta_0 x_t + \gamma_0 W_0 x_t + \epsilon_t. \quad (3)$$

The vector of outcomes  $y_t = (y_{1t}, \dots, y_{Nt})'$  assembles the individual outcomes in instance  $t$ ; the vector  $x_t$  does the same with individual characteristics.  $y_t$ ,  $x_t$ , and  $\epsilon_t$  have dimension  $N \times 1$ , the social interactions matrix  $W_0$  is  $N \times N$ , and  $\rho_0$ ,  $\beta_0$ , and  $\gamma_0$  are scalar parameters. We do not make any distributional assumptions on  $\epsilon_t$  beyond  $\mathbb{E}(\epsilon_t | x_t) = 0$  (or  $\mathbb{E}(\epsilon_t | z_t) = 0$  for an appropriate instrumental variable  $z_t$  if  $x_t$  is endogenous). We assume the network structure is predetermined and constant, and that the number of individuals  $N$  is fixed and repeated. In reality, networks may evolve over time. We thus later expand the method for dynamic network cases. The network structure  $W_0$  is a parameter to be identified and estimated.

The social interaction model (3) has been widely studied (Manski, 1993; Manresa, 2016; and Blume *et al.*, 2015, among many others), but it is also restrictive in at least two senses. First, we consider  $W_{ij}$  to be fixed and predetermined, and not through models of strategic network formation (Jackson and Wolinsky, 1996; de Paula *et al.*, 2018) or of stochastic nature, as in the class of Exponential Random Graphs (Holland and Leinhardt, 1981) or Latent Distance models (Hoff *et al.*, 2002; Breza *et al.*, 2020). If there is feedback between outcome determination and link formation, and especially if this involves unobservables, it would be important to model network formation more explicitly.

A regression of outcomes on covariates corresponds, then, to the reduced form for (3),

$$y_t = \Pi_0 x_t + \nu_t, \quad (4)$$

with  $\Pi_0 = (I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0)$  and  $\nu_t \equiv (I - \rho_0 W_0)^{-1} \epsilon_t$ . If  $W_0$  is observed, Bramoullé *et al.* (2009) note that a structure  $(\rho, \beta, \gamma)$  that is observationally equivalent to  $(\rho_0, \beta_0, \gamma_0)$  is such that  $(I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0) = (I - \rho W_0)^{-1}(\beta I + \gamma W_0)$ . This can be written as a linear equation in  $I, W_0$ , and  $W_0^2$ , and identification is established if those matrices are linearly independent. If  $W_0$  is not observed, the putative unobserved structure comprises  $W_0$ , and an observationally equivalent parameter vector will instead satisfy  $(I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0) = (I - \rho W)^{-1}(\beta I + \gamma W)$ . Following the strategy in Bramoullé *et al.* (2009) would lead to an equation in  $I, W, W_0$ , and  $WW_0$ , so the insights obtained in that paper do *not* carry over to the case we study when  $W_0$  is unknown.

We establish identification of the structural parameters of the model, including the social interactions matrix  $W_0$ , from the coefficients matrix  $\Pi_0$ . Without data on the network  $W_0$ , we treat it as an additional parameter in an otherwise standard model relating outcomes and covariates. Our identification strategy relies on how changes in covariates  $x_{it}$  reverberate through the system and impact  $y_{it}$ , as well as outcomes for other individuals. These are summarized by the entries of the coefficient matrix  $\Pi_0$ , which, in turn, encode information about  $W_0$  and  $(\rho_0, \beta_0, \gamma_0)$ . A non-zero partial effect of  $x_{it}$  on  $y_{jt}$  indicates the existence of direct *or* indirect links between  $i$  and  $j$ . When  $\rho_0 = 0$  (and  $\Pi_0 = \beta_0 I + \gamma_0 W_0$ ), only direct links produce such a correlation. When  $\rho \neq 0$ , both direct and indirect connections may generate a non-zero response, but distant connections will lead to a lower response. Our results formally determine sufficient conditions to precisely disentangle these forces.

We set out six assumptions underpinning our main identification results. Three of these are entirely standard. A fourth is a normalization required to separately identify  $(\rho_0, \gamma_0)$  from  $W_0$ , and the fifth is closely related to known results on the identification of  $(\rho_0, \gamma_0)$  when  $W_0$  is known (Bramoullé *et al.*, 2009). The sixth assumption pertains to the relation between the nature of repeated multiple observations of the outcome and covariates and restrictions on the stability of  $W$ . These Assumptions (A1-A6) deliver an identified set of up to two points.

Our first assumption explicitly states that no individuals affect themselves and is a standard condition in social interaction models:

$$(A1) \quad (W_0)_{ii} = 0, \quad i = 1, \dots, N.$$

Assumption (A1) rules out applications with self-influence. For example, Input-Output matrices typically feature  $(W_0)_{ii} > 0$ , as firms tend to source from other firms in the same industry. With Assumption (A1), we can omit elements on the diagonal of  $W_0$  from the parameter space. We thus can denote a generic parameter vector as  $\theta = (W_{12}, \dots, W_{N,N-1}, \rho, \gamma, \beta)' \in \mathbb{R}^m$ , where  $m = N(N-1) + 3$ , and  $W_{ij}$  is the  $(i, j)$ -th element of  $W$ . Reduced-form parameters can be tied

back to the structural model (3) by letting  $\Pi : \mathbb{R}^m \rightarrow \mathbb{R}^{N^2}$  define the relation between structural and reduced-form parameters:

$$\Pi(\theta) = (I - \rho W)^{-1} (\beta I + \gamma W),$$

where  $\theta \in \mathbb{R}^m$ , and  $\Pi_0 \equiv \Pi(\theta_0)$ .

As  $\epsilon_t$  (and, consequently,  $\nu_t$ ) is mean-independent from  $x_t$ ,  $\mathbb{E}[\epsilon_t|x_t] = 0$ , the matrix  $\Pi_0$  can be identified as the linear projection of  $y_t$  on  $x_t$ . We do not impose additional distributional assumptions on the disturbance term, except for conditions that allow us to identify the reduced-form parameters in (4). If  $x_t$  is endogenous, i.e.,  $\mathbb{E}[\epsilon_t|x_t] \neq 0$ , a vector of instrumental variables  $z_t$  may still be used to identify  $\Pi_0$ . In either case, identification of  $\Pi_0$  requires variation of the regressor across individuals  $i$  and through instances  $t$ . In other words, either  $\mathbb{E}[x_t x_t']$  (if exogeneity holds) or  $\mathbb{E}[x_t z_t']$  (otherwise) is full-rank.

Our next assumption controls the propagation of shocks and guarantees that they die as they reverberate through the network. This provides adequate stability and is related to the concept of stationarity in network models. It implies the maximum eigenvalue norm of  $\rho_0 W_0$  is less than one and ensures  $(I - \rho_0 W_0)$  is a non-singular matrix. As the variance of  $y_t$  exists, the transformation  $\Pi(\theta_0)$  is well-defined, and the Neumann expansion  $(I - \rho_0 W_0)^{-1} = \sum_{j=0}^{\infty} (\rho_0 W_0)^j$  is appropriate.

$$(A2) \quad \sum_{j=1}^N |\rho_0(W_0)_{ij}| < 1 \text{ for every } i = 1, \dots, N, \quad \|W_0\| < C \text{ for some positive } C \in \mathbb{R} \text{ and } |\rho_0| < 1.$$

We next assume that network effects do not cancel out, another standard assumption. As we will show, this assumption rules out the pathological case in which endogenous and exogenous effects exactly cancel each other out:

$$(A3) \quad \beta_0 \rho_0 + \gamma_0 \neq 0.$$

The need for this assumption can be shown by expanding the expression for  $\Pi(\theta_0)$ , which is possible by (A2):

$$\Pi(\theta_0) = \beta_0 I + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k. \quad (5)$$

If Assumption (A3) were violated,  $\beta_0 \rho_0 + \gamma_0 = 0$  and  $\Pi_0 = \beta_0 I$ , so the endogenous and exogenous effects would balance each other out, and network effects would be altogether eliminated in the reduced form.<sup>11</sup>

Identification of the social effects parameters  $(\rho_0, \gamma_0)$  requires that at least one row of  $W_0$  adds to a fixed and known number. Otherwise,  $\rho_0$  and  $\gamma_0$  cannot be separately identified from  $W_0$ . Clearly, no such condition would be required if  $W_0$  were observed.

---

<sup>11</sup>One important case is when networks do not determine outcomes, which we interpret as  $\rho_0 = \gamma_0 = 0$  or with  $W_0$  representing the empty network. From equation (5), it is clear that if  $\Pi(\theta_0)$  is *not* diagonal with constant entries, then it must be that  $(\rho_0 \beta_0 + \gamma_0) \neq 0$ , which implies that  $\rho_0 \neq 0$  or  $\gamma_0 \neq 0$ , and also that  $W_0$  is non-empty. Taken together, this suggests that the observation that  $\Pi(\theta_0)$  is not diagonal is sufficient to ensure that network effects are present and Assumption (A3) is not violated.

(A4) There is an  $i$  such that  $\sum_{j=1,\dots,N}(W_0)_{ij} = 1$ .

Letting  $W_y \equiv \rho_0 W_0$  and  $W_x \equiv \gamma_0 W_0$  denote the matrices that summarize the influence of peers' outcomes (the endogenous social effects) and characteristics on one's outcome (the exogenous social effects), respectively, the assumption above can be seen as a normalization. In this case,  $\rho_0$  and  $\gamma_0$  represent the row-sum for individual  $i$  in  $W_y$  and  $W_x$ , respectively.<sup>12</sup>

The fifth assumption allows for a specific kind of network asymmetry. We require the diagonal of  $W_0^2$  not to be constant as one of our sufficient conditions for identification.

(A5) There exists  $l, k$  such that  $(W_0^2)_{ll} \neq (W_0^2)_{kk}$ , i.e., the diagonal of  $W_0^2$  is not proportional to  $\iota$ , where  $\iota$  is the  $N \times 1$  vector of ones.

In unweighted networks, the diagonal of the square of the social interactions matrix captures the number of reciprocated links for each individual or, in the case of undirected networks, the popularity of those individuals. Assumption (A5) hence intuitively suggests differential popularity across individuals in the social network.

This assumption is related to the network asymmetry condition proposed elsewhere, such as in Bramoullé *et al.* (2009). They show that when  $W_0$  is known, the structural model (2) is identified if  $I$ ,  $W_0$ , and  $W_0^2$  are linearly independent. Given the remaining assumptions, this condition is satisfied if (A5) is satisfied, but the converse is *not* true: one can construct examples in which  $I$ ,  $W_0$ , and  $W_0^2$  are linearly independent when  $W_0^2$  has a constant diagonal, so  $\Pi_0$  does not pin down  $\theta_0$ . See Example 1 in Appendix A. The strengthening of this hypothesis is the formal price to pay for the social interactions matrix  $W_0$  being unknown to the researcher.

Before proceeding to our formal results, we provide a very simple illustration to shed light on how the assumptions above come together to provide identification. Suppose the observed reduced-form matrix is,

$$\Pi_0 = \frac{1}{455} \begin{bmatrix} 275 & 310 & 0 \\ 310 & 275 & 0 \\ 0 & 0 & 182 \end{bmatrix},$$

and that, following (A4), the first row is normalized to one. From the third row and column of  $\Pi_0$ , we see there is no path of any length connecting the individual in row 3 to or from those in rows 1 or 2, since her outcome is not affected by their covariates and their outcomes are not affected by her covariates. In other words, individual 3, is isolated and  $(W_0)_{13} = (W_0)_{23} = (W_0)_{31} = (W_0)_{32} = 0$ . On the other hand, individuals 1 and 2 cannot be isolated, as their covariates are correlated with the other individual's outcome, reflecting (A5).<sup>13</sup> Due to the row-sum normalization of the first

<sup>12</sup>Alternatively, one could normalize  $\rho^* = 1$  and rescale the network accordingly. In this case,  $W^* = \rho_0 W_0$  would be identified instead. Also,  $W_x = \frac{\gamma_0}{\rho_0} W^*$  so  $\gamma_0$  would be identified relative to  $\rho_0$ .  $W_y$  and  $W_x$  would be unchanged.

<sup>13</sup>If on the other hand,  $(W_0)_{ij} = 0.5, i \neq j$  in violation of (A5), and all agents were connected, the model would not be identified.

row,  $(W_0)_{12} = 1$ . Using (A3), it can be seen that  $W_0$  is symmetric if  $\Pi_0$  is symmetric. We thus find that  $(W_0)_{21} = 1$ . This and (A1) map all elements of  $W_0$ , and thus,

$$W_0 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

As the third individual is isolated, she will only be affected by her exogenous  $x_i$  and not by endogenous or exogenous peer effects. Hence, the (3,3) element of  $\Pi_0$  is equal to  $\beta_0 = \frac{182}{455} = .4$ . To find  $\rho_0$ , note that  $(I - \rho_0 W_0)\Pi_0 = \beta_0 I + \gamma_0 W_0$ . Hence, focusing on the (1,1) elements of the matrices above, we find that  $\frac{275}{455} - \rho_0 \frac{310}{455} = .4$ , implying  $\rho_0 = .3$  (complying with (A2)). Finally,  $\gamma_0$  is identified from entry (1,2), giving  $\gamma_0 = \frac{310}{455} - .3 \frac{275}{455} = .5$ .

Our final assumption articulates the need for a constant network  $W_0$  observed over multiple instances of  $y_t$  and  $x_t$ :

(A6)  $y_t$  and  $x_t$  are observed for individuals  $i = 1, \dots, N$ , and instances  $t = 1, \dots, T$ , and the network  $W_0$  does not depend on  $t$

Here, “instances” can refer to time but also to settings in which the same units are observed over multiple episodes. For example, if  $i$  are firms, then  $t$  can be segmented markets in which they operate. For simplicity, we refer to an instance as a time period. If  $\Pi_0$  is known, the main identification result we articulate below will state that  $W_0$ ,  $\rho_0$ ,  $\beta_0$  and  $\gamma_0$  are globally identified. However, in practice,  $\Pi_0$  is rarely observed and thus all quantities need to be estimated. For this purpose, when  $\Pi_0$  is not known, multiple observations of  $y_t$  and  $x_t$  with a constant  $W_0$  are required to implement the estimator. We expand on estimation requirements in Section 3.1.

Importantly, the main identification results (for a given  $\Pi_0$ ) could, in principle, be applied for each time period  $t$ . That is, one can write a version of Equation (3) as

$$y_t = \rho_0 W_{0t} y_t + \beta_0 x_t + \gamma_0 W_{0t} x_t + \epsilon_t,$$

where  $W_{0t}$  is time-varying and, consequently, the reduced-form interaction matrix  $\Pi_{0t} = (I - \rho_0 W_{0t})^{-1}(I\beta_0 + W_{0t}\gamma_0)$  is also time-varying. If the reduced-form matrices were known, the identification results we develop below could be applied to the reduced-form element by element for each  $\Pi_{0t}$ . Again, one rarely observes  $\Pi_{0t}$ , for all  $t \in [1, T]$ . This observation will motivate an extension of the method, presented in Section 2.3.3, where Assumption (A6) is relaxed and  $W_0$  is allowed to vary with  $t$ .

In an extension, we allow the network  $W_{0t}$  to vary over time and introduce kernel weights. Akin to the nonparametric regression  $Y = f(X) + \epsilon$ ,  $f$  is identified if  $\mathbb{E}(\epsilon|X) = 0$ , and it is possible to estimate  $f$  using neighboring observations if  $f$  is sufficiently smooth or varies slowly. Similar considerations extend to varying-coefficient models and, in particular, time-varying coefficient

models where local stability conditions as those discussed in Dahlhaus (2012) are usually invoked (see also Hastie and Tibshirani (1993) and their Example (e)).

## 2.2 Main Identification Results

Under the assumptions above, we can begin to identify parameters related to the network. These results are then useful for our main identification theorems. Let  $\lambda_{0j}$  denote an eigenvalue of  $W_0$  with corresponding eigenvector  $v_{0,j}$  for  $j = 1, \dots, N$ . Assumptions (A2) and (A3) allow us to identify the eigenvectors of  $W_0$  directly from the reduced form. As  $|\rho_0| < 1$ :

$$\begin{aligned} \Pi_0 v_{0,j} &= \beta_0 v_{0,j} + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k v_{0,j} \\ &= \left[ \beta_0 + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} \lambda_{0,j}^k \right] v_{0,j} \\ &= \frac{\beta_0 + \gamma_0 \lambda_{0,j}}{1 - \rho_0 \lambda_{0,j}} v_{0,j}. \end{aligned} \tag{6}$$

The infinite sum converges as  $|\rho_0 \lambda_{0,j}| < 1$  by (A2). The equation above implies that  $v_{0,j}$  is also an eigenvector of  $\Pi_0$  with the associated eigenvalue  $\lambda_{\Pi,j} = \frac{\beta_0 + \gamma_0 \lambda_{0,j}}{1 - \rho_0 \lambda_{0,j}}$ . The fact that eigenvectors of  $W_0$  are also eigenvectors of  $\Pi_0$  has a useful implication: eigencentralities may be identified from the reduced form, even when  $W_0$  is not identified. As detailed in de Paula (2017) and Jackson *et al.* (2017), such eigencentralities often play an important role in empirical work as they allow a mapping back to underlying models of social interaction.<sup>14</sup>

Now let  $\Theta \equiv \{\theta \in \mathbb{R}^m : \text{Assumptions (A1)-(A6) are satisfied}\}$  be the structural parameter space of interest. Our identification argument is structured as follows: a) we first establish local identification of the mapping  $\Pi(\theta)$  using classical results on the rank of the gradient of Rothenberg, 1971 (Theorem 1); b) we then show that  $\Pi(\theta)$  is proper (Corollary 1); and c) has a connected image (Lemma 2, in the Appendix); d) allowing us to state the cardinality of the pre-image  $\Pi^{-1}(\bar{\Pi})$  is constant for any  $\bar{\Pi}$  in the image of  $\Pi(\cdot)$ , and that the cardinality is at most 2 (Theorem 2). We then provide additional conditions to narrow the identified set to a singleton (Corollaries 2-4).

We now formally present our results. Our first theorem establishes local identification of the mapping. A parameter point  $\theta_0$  is locally identifiable if there exists a neighborhood of  $\theta_0$  containing no other  $\theta$  which is observationally equivalent. Using classical results in Rothenberg (1971), we show that our assumptions are sufficient to ensure that the Jacobian of  $\Pi$  relative to  $\theta$  is non-singular, which, in turn, suffices to establish local identification.

**Theorem 1.** *Assume (A1)-(A6).  $\theta_0 \in \Theta$  is locally identified.*

<sup>14</sup>To identify the eigencentralities, we identify the eigenvector that corresponds to the dominant eigenvalue. If  $W_0$  is non-negative and irreducible, this is the (unique) eigenvector with strictly positive entries, by the Perron-Frobenius theorem for non-negative matrices (see Horn and Johnson, 2013, p. 534).

An immediate consequence of local identification is that the set  $\{\theta \in \Theta : \Pi(\theta) = \Pi(\theta_0)\}$  is discrete (i.e., its elements are isolated points). The following corollary establishes that  $\Pi$  is a proper function, i.e. the inverse image  $\Pi^{-1}(K)$  of any compact set  $K \subset \mathbb{R}^{N^2}$  is also compact (Krantz and Parks, 2013, p. 124). Since it is discrete, the identified set must be finite.

**Corollary 1.** *Assume (A1)-(A6). Then  $\Pi(\cdot)$  is a proper mapping. Moreover, the set  $\{\theta : \Pi(\theta) = \Pi(\theta_0)\}$  has a finite number of elements.*

Under additional assumptions, the identified set is at most a singleton in each of the partitioning sets  $\Theta_- \equiv \Theta \cap \{\rho\beta + \gamma < 0\}$  and  $\Theta_+ \equiv \Theta \cap \{\rho\beta + \gamma > 0\}$ .<sup>15</sup>

Since  $\Theta = \Theta_- \cup \Theta_+$ , if the sign of  $\rho_0\beta_0 + \gamma_0$  is unknown, the identified set contains, at most, two elements. In the theorem that follows, we show global identification only for  $\theta \in \Theta_+$ , since arguments are mirrored for  $\theta \in \Theta_-$ .

**Theorem 2.** *Assume (A1)-(A6). Then for every  $\theta \in \Theta_+$ , we have  $\Pi(\theta) = \Pi(\theta_0) \Rightarrow \theta = \theta_0$ . That is,  $\theta_0$  is globally identified with respect to the set  $\Theta_+$ .*

Similar arguments apply if Theorem 2 instead were to be restricted to  $\theta \in \Theta_-$ . The proof of the corollary below is immediate and therefore omitted.

**Corollary 2.** *Assume (A1)-(A6). If  $\rho_0\beta_0 + \gamma_0 > 0$ , then the identified set contains at most one element, and similarly if  $\rho_0\beta_0 + \gamma_0 < 0$ . Hence, if the sign of  $\rho_0\beta_0 + \gamma_0$  is unknown, the identified set contains, at most, two elements.*<sup>16</sup>

We now turn our attention to the problem of identifying the sign of  $\rho_0\beta_0 + \gamma_0$  from the observation of  $\Pi_0$ . This would then allow us to establish global identification using Theorem 2. It is apparent from (5) that if  $\rho_0 > 0$  and  $(W_0)_{ij} \geq 0$ , for all  $i, j = \{1, \dots, N\}$ , the off-diagonal elements of  $\Pi_0$  identify the sign of  $\rho_0\beta_0 + \gamma_0$ .

**Corollary 3.** *Assume (A1)-(A6). If  $\rho_0 > 0$  and  $(W_0)_{ij} \geq 0$ , the model is globally identified.*

Real-world applications often suggest endogenous social interactions are positive ( $\rho_0 > 0$ ), in which case global identification is fully established by Corollary 3. On the other hand, if  $\rho_0 < 0$  (e.g., if outcomes are strategic substitutes),  $\rho_0^k$  in (5) alternates signs with  $k$ , and the off-diagonal elements no longer carry the sign of  $\rho_0\beta_0 + \gamma_0$ . Nonetheless, if  $W_0$  is non-negative and irreducible (i.e., not permutable into a block-triangular matrix or, equivalently, a strongly connected social network), the model is also identifiable without further restrictions on  $\rho_0$ :

<sup>15</sup>The global inversion results we use are related to, but different from, variations on a classic inversion result of Hadamard that has been used in the literature. In contrast, we employ results on the cardinality of the pre-image of a function, relying on less stringent assumptions. While the Hadamard result requires the image of the function to be simply-connected (Theorem 6.2.8 of Krantz and Parks, 2013), the results we rely on do not.

<sup>16</sup>Under some special conditions, the mirror image of  $\theta_0$  can be characterized from equation (5). If  $-W_0$  satisfies Assumption (A4), we may set  $\rho^* = -\rho_0$ ,  $\beta^* = \beta_0$ ,  $\gamma^* = -\gamma_0$  and  $W^* = -W_0$ . Then,  $\rho_0\beta_0 + \gamma_0 = -(\rho^*\beta^* + \gamma^*)$ . Also note that  $\sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k = -\sum_{k=1}^{\infty} (\rho^*)^{k-1} (W^*)^k$ , so  $(\rho_0\beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k = (\rho^*\beta^* + \gamma^*) \sum_{k=1}^{\infty} (\rho^*)^{k-1} (W^*)^k$ . It follows that  $\Pi(\theta_0) = \Pi(\theta^*)$ , where  $\theta^* = (\rho^*, \beta^*, \gamma^*, W^*)$ .

**Corollary 4.** *Assume (A1)-(A6),  $(W_0)_{ij} \geq 0$  and  $W_0$  is irreducible. If  $W_0$  has at least two real eigenvalues or  $|\rho_0| < \sqrt{2}/2$ , then the model is globally identified.*

Corollary 4 requires that  $W_0$  be irreducible, i.e., that it is not permutable into a block upper-triangular matrix. In the context of directed graphs, this is similar to requiring that the matrix be strongly connected, that is, that any node can be reached from any other node. The corollary then rules out cases when the network is not connected, for example, if there are two disjoint groups (with no connection across groups), or a star network pointing from the center towards the edges. The corollary holds if there are at least two real eigenvalues, or if  $\rho_0$  is appropriately bounded. Since  $W_0$  is non-negative, it has at least one real eigenvalue by the Perron-Frobenius theorem. If  $W_0$  is symmetric, for example, its eigenvalues are all real, and Corollary 4 holds. It also holds if  $(W_0)_{ij} \leq 0$ , as we can rewrite the model as  $\rho W_0 = -\rho|W_0|$ , where  $|W_0|$  is the matrix whose entries are the absolute values of the entries in  $W_0$ . However, Corollary 4 rules out cases that mix positive  $(W_0)_{ij} \geq 0$  and negative interactions  $(W_0)_{ij} \leq 0$ . In any case, the bound on  $|\rho_0|$  is sufficient and holds in most (if not all) empirical estimates we are aware of obtained from either elicited or postulated networks, and in our application on tax competition.

## 2.3 Extensions

We present three extensions of the method for individual fixed effects, common shocks, and time-varying  $W$ . Appendix B describes extensions for multivariate covariates and heterogeneous  $\beta_0$ .

### 2.3.1 Individual Fixed Effects

We observe outcomes for  $i = 1, \dots, N$  individuals repeatedly through  $t = 1, \dots, T$  instances. If  $t$  corresponds to time, it is natural to think of there being unobserved heterogeneity across individuals,  $\alpha_i$ , to be accounted for when estimating  $\Pi_0$ . The structural model (2) is then,

$$y_{it} = \rho_0 \sum_{j=1}^N W_{0,ij} y_{jt} + \beta_0 x_{it} + \gamma_0 \sum_{j=1}^N W_{0,ij} x_{jt} + \alpha_i + \epsilon_{it},$$

which can be written in matrix form as

$$y_t = \rho_0 W_0 y_t + x_t \beta_0 + W_0 x_t \gamma_0 + \alpha^* + \epsilon_t,$$

where  $\alpha^*$  is the vector of fixed effects. Individual-specific and time-constant fixed effects can be eliminated using the standard subtraction of individual time averages. Defining  $\bar{y}_t = T^{-1} \sum_{t=1}^T y_t$ ,  $\bar{x}_t = T^{-1} \sum_{t=1}^T x_t$ , and  $\bar{\epsilon}_t = T^{-1} \sum_{t=1}^T \epsilon_t$ ,

$$y_t - \bar{y}_t = \rho_0 W_0 (y_t - \bar{y}_t) + (x_t - \bar{x}_t) \beta_0 + W_0 (x_t - \bar{x}_t) \gamma_0 + \epsilon_t - \bar{\epsilon}_t,$$



if  $W_0$  does not change with time. Identification from the reduced form follows from previous theorems, since  $\Pi_0$  is unchanged when regressing  $y_t - \bar{y}_t$  on  $x_t - \bar{x}_t$ .<sup>17</sup>

### 2.3.2 Common Shocks

We next allow for unobserved common shocks to all individuals in the network in the same instance  $t$ . Such correlated effects  $\alpha_t$  can confound the identification of social interactions. As we have not placed any distributional assumption on the covariance matrix of the disturbance term, our analysis readily incorporates correlated effects that are orthogonal to  $x_t$ . When this is not the case, one possibility is to model the correlated effects  $\alpha_t$  explicitly. The model then is,

$$y_t = \rho_0 W_0 y_t + x_t \beta_0 + \gamma_0 W_0 x_t + \alpha_t \iota + \epsilon_t,$$

where  $\alpha_t$  is a scalar capturing shocks in the network common to all individuals. Let  $\Pi_{01} = (I - \rho_0 W_0)^{-1}$  and  $\Pi_{02} = (\beta_0 I + \gamma_0 W_0)$  such that  $\Pi_0 = \Pi_{01} \Pi_{02}$ . The reduced-form model is

$$y_t = \Pi_0 x_t + \alpha_t \Pi_{01} \iota + v_t.$$

We propose a transformation to eliminate the correlated effects: exclude the individual-invariant  $\alpha_t$ , subtracting the mean of the variables in a given period (global differencing). For this purpose, define  $H = \frac{1}{n} \iota \iota'$ . We note that in empirical and theoretical work, it is customary to strengthen Assumption (A4) and require that *all* rows of  $W_0$  sum to one if no individual is isolated (see for example Blume *et al.*, 2015). This strengthened assumption is usually referred to as row-sum normalization, and is stated below:

(A4') For all  $i = 1, \dots, N$ , we have that  $\sum_{j=1, \dots, N} (W_0)_{ij} = 1$ .

This can be written compactly as  $W_0 \iota = \iota$ . In this case,  $W_0$  can be interpreted as the normalized adjacency matrix. Under row-sum normalization we have that,

$$\begin{aligned} (I - H) y_t &= (I - H) (I - \rho_0 W_0)^{-1} (\beta_0 I + \gamma_0 W_0) x_t + (I - H) (I - \rho_0 W_0)^{-1} \epsilon_t \\ &= (I - H) \Pi_0 x_t + (I - H) v_t, \end{aligned}$$

because  $(I - H) (I - \rho_0 W_0)^{-1} \alpha_t \iota = 0$  if Assumption (A4') holds. It then follows that  $\tilde{\Pi}_0 = (I - H) \Pi_0$  is identified. The next proposition shows that, under row-sum normalization of  $W_0$ ,  $\Pi_0$  is identified from  $\tilde{\Pi}_0$  (and, as a consequence, the previous results immediately apply).

**Proposition 1.** *If  $W_0$  is non-negative, irreducible, and row-sum normalized,  $\Pi_0$  is identified from  $\tilde{\Pi}_0$ .*

---

<sup>17</sup>As is the case in panel data, this would require strict exogeneity ( $\mathbb{E}[\epsilon_s | x_t] = 0$  for any  $s$  and  $t$ ) or predetermined errors ( $\mathbb{E}[\epsilon_s | x_t] = 0$  for  $s \geq t$ ) so that the matrix  $\Pi_0$  can be consistently estimated.

Under row-sum normalization of  $W_0$ , a common group-level shock affects individuals homogeneously since  $(I - \rho_0 W_0)^{-1} \alpha_t \iota = \alpha_t (I + \rho_0 W_0 + \rho_0^2 W_0^2 + \dots) \iota = \frac{\alpha_t}{1 - \rho_0} \iota$ , which is a vector with no variation across entries. Consequently, global differencing eliminates correlated effects and  $(I - H)(I - \rho_0 W_0)^{-1} \alpha_t \iota = (I - \rho_0 W_0)^{-1} \alpha_t (I - H) \iota = 0$ . Absent row-sum normalization, global differencing does not ensure correlated effects are eliminated. To see this, note that  $(I - \rho_0 W_0)^{-1}$  is no longer row-sum normalized and  $\alpha_t (I - \rho_0 W_0)^{-1} \iota$  does not have constant entries.

The next proposition makes this point formally: that the stronger Assumption (A4') is *necessary* to eliminate group-level shocks by showing it is not possible to construct a data transformation that eliminates group effects in the absence of row-sum normalization.

**Proposition 2.** *Define  $r_{W_0} = (I - \rho_0 W_0)^{-1} \iota$ . If in space  $\Theta = \{\theta \in \mathbb{R}^m : \text{Assumptions (A1)-(A6) are satisfied}\}$ , there are  $N$  matrices  $W_0^{(1)}, \dots, W_0^{(N)}$  such that  $[r_{W_0^{(1)}} \dots r_{W_0^{(N)}}]$  has rank  $N$ , then the only transformation such that  $(I - \tilde{H})(I - \rho_0 W_0)^{-1} \iota = 0$  is  $\tilde{H} = I$ .*

It is useful to be able to test for row-sum normalization (A4') as it enables common shocks to be accounted for in the social interactions model. This is possible as

$$\begin{aligned} \Pi_0 \iota &= \beta_0 \iota + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k \iota \\ &= \left[ \beta_0 + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} \right] \iota \\ &= \frac{\beta_0 + \gamma_0}{1 - \rho_0} \iota. \end{aligned} \tag{7}$$

The last equality follows from the observation that, under row-normalization of  $W_0$ ,  $W_0^k \iota = W_0 \iota = \iota$ ,  $k > 0$ . This implies  $\Pi_0$  has constant row-sums, which suggests row-sum normalization is testable. In the Appendix, we derive a Wald test statistic to do so.<sup>18</sup>

### 2.3.3 Time-varying $W$

We now relax Assumption (A6), which states that  $W_0$  does not vary across the time periods  $t = 1, \dots, T$ . The version of Equation (3) with time-varying network is

$$y_t = \rho_0 W_{0t} y_t + \beta_0 x_t + \gamma_0 W_{0t} x_t + \epsilon_t,$$

with the reduced-form matrix  $\Pi_{0t} = (I - \rho_0 W_{0t})^{-1} (I \beta_0 + W_{0t} \gamma_0)$ . We note that the identification results developed in Subsection 2.2 can, in principle, be applied element by element to each  $\Pi_{0t}$ , leading to the identification of a time-varying  $W_{0t}$  (and, potentially, of the parameters  $\rho_0$ ,  $\beta_0$  and  $\gamma_0$ ).

<sup>18</sup>For ease of explanation, in the Appendix, we derive the test under the asymptotic distribution of the OLS estimator. The test generally holds with minor adjustments for estimators with known asymptotic distributions.

In practice, implementing any estimation strategy with a time-varying  $\Pi_{0t}$  (or  $W_{0t}$ ) is not feasible using only observation from the single time period  $t$ . We instead adopt a kernel-weighted version. Define period-specific weights  $\omega_t$ , and consider the transformed data  $\tilde{y}_s = \omega_s(t)y_s$  and  $\tilde{x}_s = \omega_s(t)x_s$ ,  $s = 1, \dots, T$ . Evidently, uniform weights  $\omega_s = 1$ ,  $s = 1, \dots, T$  are equivalent to the strategy not considering time-varying networks, and assuming that the networks are fixed within those windows. Alternatively, one could estimate  $W_t$  in time windows by setting  $\omega_t = 1[\underline{t} \leq t \leq \bar{t}]$ , where  $\underline{t}$  and  $\bar{t}$  are the start and end of the time window for which  $W_t$  is estimated. In this case, the minimum effective window length  $\bar{t} - \underline{t}$  can be computed as we discuss in Section 3.1. In the context of DSGE models with time-varying parameters, Kapetanios *et al.* (2019) suggests a Gaussian kernel with positive weights throughout the entire sample. As in nonparametric regression with smooth kernel weights, it also assumes that the network evolves slowly over time. We further discuss this strategy in the estimation section, and it is implemented in the empirical application section below.

### 3 Implementation

We now transition from our core identification results to their practical implementation. In practice, Ordinary Least Squares (OLS) can only be used to estimate  $\theta$  if  $T \gg N$ , which is in practice unlikely to be met, as this is a high-dimensional problem. Our preferred approach makes use of penalized estimation techniques that can be used for any given  $T$ . More specifically, we make use of the Adaptive Elastic Net GMM (Caner and Zhang, 2014), which is based on the penalized GMM objective function. Given the identification results presented in Section 2, the population moments used in forming the GMM objective function will be satisfied at the true parameter vector.

After setting out the estimation procedure, we showcase the method using Monte Carlo simulations based on stylized and real-world network structures. In each case, we take a fixed network structure  $W_0$ , and simulate panel data as if the data generating process were given by the model in (1). We apply the method to the simulated panel data to recover estimates of all elements in  $W_0$ , as well as the endogenous and exogenous social effect parameters.

#### 3.1 Estimation

The parameter vector to be estimated is high-dimensional:  $\theta = (W_{12}, \dots, W_{N,N-1}, \rho, \gamma, \beta)' \in \mathbb{R}^m$ , where  $m = N(N-1) + 3$  and  $W_{ij}$  is the  $(i, j)$ -th element of the  $N \times N$  social interactions matrix  $W_0$ . To be clear, in a network with  $N$  individuals, there are  $N(N-1)$  potential interactions because an individual could interact with everyone else but herself (which would violate Assumption A1). As a consequence, even with a modest  $N$ , there are many more parameters to estimate, and  $m$  is large. For example, a network with  $N = 50$  implies more than 2,000 parameters to estimate. While we consider  $N$  (and thus  $m$ ) fixed, we still refer to  $\theta$  as high-dimensional. OLS estimation requires  $m \ll NT (\Rightarrow N \ll T)$ , so many more time periods than individuals: a requirement often

met in finance data sets (van Vliet, 2018) or in other fields (see, e.g., Section 4.2 in Rothenhäusler *et al.*, 2015). Instead, to estimate a large number of parameters with limited data, we utilize high-dimensional estimation methods, which are the focus of a rapidly growing literature.

Sparsity is a key assumption underlying many high-dimensional estimation techniques. In the context of social interactions, we say that  $W_0$  is sparse if  $\tilde{m}$ , the number of non-zero elements of  $W_0$ , is such that  $\tilde{m} \ll NT$ . The notion of sparsity thus depends on the number of time periods. Sparsity corresponds to assuming that individuals influence or are influenced by a small number of others, relative to the overall size of the potential network and the time horizon in the data. As such, sparsity is typically *not* a binding constraint in social networks analysis.<sup>19</sup>

In the estimation of sparse models, the “effective number of parameters” (or “effective degrees of freedom”) relates to the number of variables with non-zero estimated coefficients (Tibshirani and Taylor, 2012). In the context of the current social network model, this is equivalent to  $m$  parameters, where  $m = dN(N - 1) + K$  and  $d$  is the network density defined as  $\tilde{m}/(N(N - 1))$ . The Adaptive Elastic Net GMM estimator presented by Caner and Zhang (2014) converges at a rate of  $\sqrt{NT/\tilde{m}} = \sqrt{NT/[dN(N - 1) + K]} = O(\sqrt{T/(dN)})$  (see remark 7 in Caner and Zhang, 2014). Hence the quality of the large sample results relies on a comparison between  $T$  and  $dN$ . In line with this, we thus require  $NT \gg dN(N - 1) + K$ . For example, in the high-school network of Coleman (1964) that is part of our simulation exercise,  $N = 70$  and  $d = 0.076$ . Assuming  $K = 3$ ,  $N(N - 1) \times d + 3 = 370.1$ .<sup>20</sup>

Finally, to reiterate, our identification results themselves do *not* depend on the sparsity of networks. In particular, Assumptions (A1)-(A6) *do not* impose restrictions on the number of links in  $W_0$ , or  $\tilde{m}$ .<sup>21</sup> The identification results presented in Section 2 apply more broadly and irrespective of the estimation procedure.

Our preferred approach estimates the interaction matrix in the reduced form while penalizing and imposing sparsity on the structural object  $W_0$ . We impose sparsity and penalization in the structural-form matrix  $W_0$  because this is a weaker requirement than imposing sparsity and penalization in the reduced-form matrix  $\Pi_0$ .<sup>22</sup> To do so, we use the Adaptive Elastic Net GMM (Caner

---

<sup>19</sup>Common stylized networks are sparse, such as the star, lattice (each individual is a source of spillover only to one other individual), or interactions in pairs, triads or small groups (De Giorgi *et al.*, 2010). Real-world economic networks are also sparse. The sparsity in *AddHealth* friendship network is around 98%. Sparsity of the production networks in the US is above 99% (Atalay *et al.*, 2011).

<sup>20</sup>As pointed out by a referee, variation in  $x$  will also matter for estimation precision. This is reflected in the asymptotic distribution for this estimator, shown later in this subsection.

<sup>21</sup>If  $N \rightarrow \infty$ , Assumption (A2) would imply vanishing  $(W_0)_{ij}$  entries. As highlighted previously, we consider  $N$  to be fixed, in line with many practical applications. Furthermore, Assumption (A2) is used to represent inverse matrices as Neumann series in our identification results. What is necessary for this to hold is that a sub-multiplicative norm on  $\rho W$  be less than one. Here we use a specific norm (i.e., the maximum row-sum norm), but other (induced) norms are also possible (i.e., the 2-norm or the 1-norm) (see Horn and Johnson, 2013, Chapter 5.6).

<sup>22</sup>Note that even if  $W$  is sparse,  $\Pi$  may not be sparse. In Appendix C.1, we show that  $[\Pi_0]_{ij} = 0$  if, and only if, there are no paths between  $i$  and  $j$  in  $W_0$ , so the pair is not connected. So, sparsity in  $\Pi_0$  is understood as  $W_0$  being “sparsely connected”, which is a stronger assumption than sparsity in  $W_0$ .

and Zhang, 2014), which is based on the penalized GMM objective function,

$$G_{NT}(\theta, p) \equiv g_{NT}(\theta)' M_T g_{NT}(\theta) + p_1 \sum_{\substack{i,j=1 \\ i \neq j}}^N |W_{i,j}| + p_2 \sum_{\substack{i,j=1 \\ i \neq j}}^N |W_{i,j}|^2 \quad (8)$$

where  $\theta = (W_{1,2}, \dots, W_{N,N-1}, \rho, \gamma, \beta)'$  with dimension  $m = N(N-1) + 3$ , and  $p_1$  and  $p_2$  are the penalization terms. The term  $g_{NT}(\theta)' M_T g_{NT}(\theta)$  is the unpenalized GMM objective function with moment conditions based on orthogonality between the structural disturbance term and the covariates:  $g_{NT}(\theta) = \sum_{t=1}^T [x_{1t}e_t(\theta)' \cdots x_{Nt}e_t(\theta)']'$ ,  $e_t(\theta) = y_t - \rho W y_t - \beta x_t - \gamma W x_t$ . There are  $q \equiv N^2$  moment conditions since  $x_{it}$  is orthogonal to  $e_{jt}$  for each  $i, j = 1, \dots, N$ . Hence, the GMM weight matrix  $M_T$  is of dimension  $N^2 \times N^2$ , symmetric, and positive definite. For simplicity, we use  $M_T = I_{N^2 \times N^2}$ . Note that if  $x_t$  is econometrically endogenous, one can also exploit moment conditions with respect to available instrumental variables.<sup>23</sup> Given the identification results presented in Section 2, if  $\theta \neq \theta_0$  and does not belong to the identified set, then  $\Pi(\theta) \neq \Pi(\theta_0)$ . Consequently, the populational version of the GMM objective function is uniquely minimized at the true parameter vector  $\theta_0$ .

The penalization terms in Equation (8) are what makes this different from a standard GMM problem. The first term,  $p_1 \sum_{i,j=1, i \neq j}^N |W_{i,j}|$ , penalizes the sum of the absolute values of  $W_{i,j}$ , i.e., the sum of the strength of links, for all node-pairs. Depending on the choice of  $p_1$ , some  $W_{i,j}$ 's will be estimated as exact zeros. A larger share of parameters will be estimated as zeros if  $p_1$  increases. The second term,  $p_2 \sum_{i,j=1, i \neq j}^N |W_{i,j}|^2$ , penalizes the sum of the square of the parameters. This term has been shown to provide better model-selection properties, especially when explanatory variables are correlated (Zou and Zhang, 2009). The first-stage estimate is

$$\tilde{\theta}(p) = (1 + p_2/T) \cdot \arg \min_{\theta \in \mathbb{R}^m} G_{NT}(\theta, p) \quad (9)$$

where  $(1 + p_2/T)$  is a bias-correction term also used by Caner and Zhang (2014).

Implementing the numerical optimization embedded in Equation (9) is computationally challenging, as  $m = N(N-1) + 3$  may entail a large number of function arguments. We instead implement the following modification to use fast Least-Angle Regression (LARS) algorithms (Efron *et al.*, 2004). For any given  $\rho$ ,  $\beta$ , and  $\gamma$ , the expression for  $e_t(\theta)$  is linear in  $W$ :

$$e_t(\theta) = y_t - x_t \beta - W(\rho y_t + x_t \gamma) = \tilde{y}_{it}(\beta) - W \tilde{x}_t(\rho, \gamma)$$

where  $\tilde{y}_{it}(\beta) \equiv y_t - x_t \beta$  and  $\tilde{x}_t(\rho, \gamma) \equiv \rho y_t + x_t \gamma$  and, following the strategy above, is instrumented

---

<sup>23</sup>For expositional ease, we describe estimation in the context of the reduced-form model (4), thereby abstaining from individual fixed or correlated effects. As the GMM estimator uses moments between the structural disturbance terms and covariates, this endogeneity is built into the estimation procedure.

with  $x_t$ . This motivates a two-step optimization routine:

$$\min_{\theta \in \Theta = \Theta_1 \times \Theta_2} G_{NT}(\theta, p) = \min_{(\rho, \beta, \gamma) \in \Theta_1} \left[ \min_{W_{ij} \in \Theta_2} G_{NT}(\theta, p) \right],$$

where the expression in brackets has a computationally efficient solution through the LARS algorithm. The numerical optimization is then subsequently conducted over the parameter space of  $(\rho, \beta, \gamma)$  only. We also impose row-sum normalization. Details of the implementation are expanded in Appendix Subsection C.2.

A second (adaptive) step provides improvements by re-weighting the penalization by the inverse of the first-step estimates (Zou, 2006):

$$\tilde{\theta}^*(p) = (1 + p_2/T) \cdot \arg \min_{\theta \in \Theta} \left\{ g_{NT}(\theta)' M_T g_{NT}(\theta) + p_1^* \sum_{\substack{\{i,j:\tilde{W}_{ij} \neq 0, \\ i,j=1,\dots,N, \\ i \neq j\}}} \frac{|W_{i,j}|}{|\tilde{W}_{i,j}|^c} + p_2 \sum_{\substack{\{i,j:\tilde{W}_{ij} \neq 0, \\ i,j=1,\dots,N, \\ i \neq j\}}} |W_{i,j}|^2 \right\}, \quad (10)$$

where  $\tilde{W}_{i,j}$  is the  $(i, j)$ -th element of the first-step estimate of  $W$ . We follow Caner and Zhang (2014) and set  $c = 2.5$ . If  $\tilde{W}_{i,j} < 0.05$ , we set  $\tilde{W}_{i,j} = 0.05$ . This ensures that the second-stage estimates can be non-zero even if the first-stage estimates were zero or small. The computational improvement – described above for the first-stage estimator – is also applied in the adaptive stage.

As a third and final step, we fix the support of  $\tilde{\theta}^*(p)$ ,  $\mathcal{S} = \{\rho, \beta, \gamma\} \cup \{W_{ij} : \tilde{W}_{ij}^* \neq 0\}$  and estimate the final parameters without penalization. This takes as arguments only the elements of  $\tilde{\theta}^*(p)$  that were estimated as non-zero in the adaptive step. In essence, this step boils down to a standard GMM approach,

$$\hat{\theta}_{\mathcal{S}}(p) = \arg \min_{\theta \in \mathcal{S}} \{g_{NT}(\theta)' M_T g_{NT}(\theta)\}. \quad (11)$$

Importantly, Caner and Zhang (2014) show that the third-step estimator is asymptotically normal, with a known and easy-to-compute distribution,

$$\delta' \left[ (\hat{G}' M_T \hat{G})^{-1} \cdot (\hat{G}' M_T \Omega M_T \hat{G}) \cdot (\hat{G}' M_T \hat{G})^{-1} \right]^{1/2} \cdot \sqrt{NT/\tilde{m}} \cdot (\hat{\theta}_{\mathcal{S}} - \theta_0) \xrightarrow{d} N(0, 1),$$

where  $\hat{G} \equiv \hat{G}(\hat{\theta}) = \nabla g_{NT}(\theta)$  and  $\Omega \equiv E[g_{NT}(\theta)g_{NT}(\theta)']$ .<sup>24</sup> This allows us to conduct hypothesis testing and inference on the  $\rho, \beta, \gamma$  and the non-zero elements of  $W$ .

We write  $p = (p_1, p_1^*, p_2)$  as the final set of penalization parameters. Conditional on  $p$ , the estimate of the procedure is  $\hat{\theta}(p)$ . As in Caner and Zhang (2014, p. 35), the penalization parameters

<sup>24</sup>This applies in the case of small  $p_2$ . In the case of large  $p_2$ , the asymptotic distribution is pre-multiplied by  $K_n = \frac{I + p_2 [\hat{G}(\hat{\theta})' \hat{\Omega}^{-1} \hat{G}(\hat{\theta})]^{-1}}{1 + p_2/NT}$ . See Theorem 4 of Caner and Zhang (2014).

$p$  are chosen by the BIC criterion. This balances model fit with the number of parameters included in the model.<sup>25</sup>

In Appendix C.2, we provide further implementation details, including the choice of initial conditions. Of course, other estimation methods are available, and our identification results do not hinge on any particular estimator. Our aim is to demonstrate the practical feasibility of using the Adaptive Elastic Net estimator rather than claim it is the optimal estimator.<sup>26</sup>

## 3.2 Simulations

We showcase the method using Monte Carlo simulations. We describe the simulation procedures, results, and robustness checks in more detail in Appendix D.1. Here, we just provide a brief overview to highlight how well the method works to recover social networks even in relatively short panels.

For each simulated network, we take a fixed network structure  $W_0$  and simulate panel data as if the data generating process were given by (1). We then apply the method to the simulated panel data to recover estimates of all elements in  $W_0$ , as well as the endogenous and exogenous social effect parameters  $(\rho_0, \gamma_0)$ . Our result identifies entries in  $W_0$  and so naturally recovers links of varying strength. It is long recognized that link strength might play an important role in social interactions (Granovetter, 1973). Data limitations often force researchers to postulate some ties to be weaker than others (say, based on interaction frequency). In contrast, our approach identifies the continuous strength of ties,  $W_{0,ij}$ , where  $W_{0,ij} > 0$  implies node  $j$  influences node  $i$ .

The stylized networks we consider are a random network and a political party network in which two groups of nodes each cluster around a central node. The real-world networks we consider are the high-school friendship network in Coleman (1964) from a small high school in Illinois, and one of the village networks elicited in Banerjee *et al.* (2013) from rural Karnataka, India.

Summary statistics for each network are presented in Panel A of Table A1. The four networks differ in their size, complexity, and the relative importance of strong and weak ties. For example, the Erdős-Renyi network only has strong ties, while the political party network has twice as many strong as weak ties. For the real-world networks, the mean out-degree distributions are higher, so the majority of ties are weak, with the high school network having around 80% of its edges being weak ties. All four networks are also sparse.

---

<sup>25</sup>Following Caner and Zhang, 2014, the choice of  $p$ , which we denote as  $\hat{p}$ , is the one that minimizes

$$\text{BIC}(p) = \log \left[ g_{NT}(\hat{\theta}(p))' M_T g_{NT}(\hat{\theta}(p)) \right] + A(\hat{\theta}(p)) \cdot \frac{\log T}{T}$$

where  $A(\hat{\theta}(p))$  counts the number of non-zero coefficients among  $\{W_{1,2}, \dots, W_{N,N-1}\}$ , and larger than a numerical tolerance, which we set at  $10^{-5}$ . See also Zou *et al.* (2007).

<sup>26</sup>See the alternative approaches of Gautier and Tsybakov (2014), Manresa (2016), Lam and Souza (2016), and Gautier and Rose (2016).

For the stylized networks, we assess the performance of the estimator for a fixed network size,  $N = 30$ . We simulate the real-world networks using non-isolated nodes in each (so  $N = 70$  and  $65$  respectively).<sup>27</sup>

We evaluate the procedure over varying panel lengths (starting from short panels with  $T = 5$ ), using various metrics. Given our core contribution is to identify the social interactions matrix, we first examine the proportion of true zero entries in  $W_0$  estimated as zeros and the proportion of true non-zero entries estimated as non-zeros. A global perspective of the proximity between the true and estimated networks can be inferred from their average absolute distance between elements. This is the mean absolute deviation of  $\hat{W}$  and  $\hat{\Pi}$  relative to their true values, defined as  $MAD(\hat{W}) = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} |\hat{W}_{ij} - W_{ij,0}|$  and  $MAD(\hat{\Pi}) = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} |\hat{\Pi}_{ij} - \Pi_{ij,0}|$ . As these metrics are closer to zero, more of the elements in the true matrix are correctly estimated. Finally, we evaluate the procedure’s performance using averaged estimates of the endogenous and exogenous social effect parameters,  $\hat{\rho}$  and  $\hat{\gamma}$ . In keeping with the estimation strategy in our empirical application, we report unpenalised GMM.

### 3.3 Results

Figure A1 shows the simulation results as evaluated using the six metrics described above. Panel A shows that for each network, the proportion of zero entries in  $W_0$  correctly estimated as zeros is above 95% even when  $T = 5$ . The proportion approaches 100% as  $T$  grows. Conversely, Panel B shows the proportion of strong non-zero entries estimated as non-zeros (defined as larger than 0.3) is also high for a small  $T$ . It is above 70% from  $T = 5$  for the Erdős-Renyi network, being at least 85% across networks for  $T = 25$ , and increasing as  $T$  grows. As discussed above, the Adaptive Elastic Net estimator may only recover strong edges well, and not necessarily the weaker ones, due to the well-known issue with shrinkage estimators that they tend to shrink small parameters to zero. We return to this issue below.

Panels C and D show that for each simulated network, the mean absolute deviation between estimated and true networks for  $\hat{W}$  and  $\hat{\Pi}$  falls quickly with  $T$  and is close to zero for large sample sizes. Finally, Panels E and F show that biases in the endogenous and exogenous social effects parameters,  $\hat{\rho}$  and  $\hat{\gamma}$ , also fall in  $T$  (we do not report the bias in  $\hat{\beta}$  since it is close to zero for all  $T$ ). The fact that biases are not zero is as expected for a small  $T$ , being analogous to well-known results for autoregressive time series models.<sup>28</sup>

Figure A2 shows that, as  $T$  increases, the procedure detects weaker links. The figure also shows that, with low sample sizes, weak edges are generally not detected. This pattern is consistent with the well-known fact that small parameters are likely shrunk to zero due to the penalization (Belloni and Chernozhukov, 2011). The absence of weak edges also implies that the strength of strong edges

<sup>27</sup>Like Bramoullé *et al.* (2009), we exclude isolated nodes because they do not conform to row-sum normalization.

<sup>28</sup>The bias in spatial auto-regressive models with a small number of observations *even when the network is observed* is similarly documented by Smith (2009), Neuman and Mizruchi (2010), Wang *et al.* (2014), and others.



may also be over-estimated, since rows are normalized to one. In Panel A, we show the distribution of the estimates of  $\hat{W}_{ij}$ , with  $T = 25$  and for the high-school network. We show the distribution for the five most common values of  $W_{0,ij}$ . We find that most edges weaker than .5 are not detected; edges with a strength of .75 are substantially more likely to be estimated as non-zeros. When they are detected as non-zeros, they are more likely to be over-estimated. When we estimate  $W$  with  $T = 150$ , Panel B shows that virtually all edges with strength greater than .5 are estimated as non-zeros, and most edges with strength .375 are also detected. We further see a more continuous distribution of estimates of edge strength. Only edges smaller than .25 are not detected. Panels C and D show a similar conclusion for the village network.

Figure A3 shows the simulated and actual networks under  $T = 100$  time periods. The network size is set to  $N = 30$  in the two stylized networks,  $N = 70$  for the high school network, and  $N = 65$  for the village household network. In comparing the simulated and true networks, Figure A3 distinguishes between kept edges, added edges, and removed edges. Kept edges are depicted in blue: these links are estimated as non-zero in at least 5% of the iterations, and are also non-zero in the true network. Added edges are depicted in green: these links are estimated as non-zero in at least 5% of the iterations but the edge is zero in the true network. Removed edges are depicted in red: these links are estimated as zero in at least 5% of the iterations but are non-zero in the true network. Figure A3 further distinguishes between strong and weak links: strong links are shown as solid edges ( $W_{0,ij} > .3$ ), and weak links are shown as dashed edges.

Panel A of Figure A3 compares the simulated and true Erdős-Renyi networks. All links are recovered. For the political party network, Panel B shows that all strong edges are correctly estimated. However, around half the weak edges are recovered (blue dashed edges), with the others being missed (red dashed edges). As discussed above, this is not surprising given that shrinkage estimators force small non-zero parameters to zero. Hence, a larger  $T$  is needed to achieve similar performance to the other simulated networks in terms of detecting weak links. For the more complex and larger real-world networks, Panel C shows that in the high-school network, the strong edges are all recovered. However, around half the weak edges are missing (red dashed edges), and there are a relatively small number of added edges (green edges): these amount to 87 edges, or approximately 1.9% of the 4,534 zero entries in the true high-school network. A similar pattern of results is seen in the village network in Panel D: the strong edges are all recovered, and here the majority of weak edges are also recovered.

Panel B of Table A1 compares the network- and node-level statistics calculated from the recovered social interactions matrix  $\hat{W}$  to those in Panel A from the true interactions matrix  $W_0$ . The random Erdős-Renyi network is perfectly recovered. For the political party network, the number of recovered edges is slightly lower than in the true network (41 vs. 45), and all edges are classified as strong. The mean of the in- and out-degree distributions are slightly lower in the recovered network, and all three nodes with the highest out-degree are correctly captured (nodes 1, 11, and 28), include both party leaders (individuals 1 and 11). We then move to discussing the performance in

the two real-world networks. In the high-school network, 30% of all edges are correctly recovered, and they are all strong edges. As already noted in Figure A2, weak edges are not well estimated in the high-school network. This draws two main consequences. First, the average in- and out-degrees are smaller in the recovered network relative to the true network. Second, we over-estimate the number of strong edges (61 vs. 113). This is a downside of row-sum normalization: because some weak edges get estimated as zeros, the non-zeros are over-estimated so that the row adds to one. We do, however, recover all three individuals with the highest out-degree. Finally, in the village network, half the edges are recovered. The same phenomena of underestimating weak and overestimating strong edges are again observed. We again recover the three households with the highest out-degree (nodes 16, 35, and 57).

In the Appendix, we show the robustness of the simulation results to (i) varying network sizes and (ii) alternative parameter choices and richening the structure of shocks across nodes. We also demonstrate the gains from using the Adaptive Elastic Net GMM estimator over alternative estimators, such as the Adaptive Lasso and OLS.

## 4 Application: Tax Competition between US States

We apply our results to shed new light on a classic social interactions problem: tax competition between US states (Wilson, 1999). Defining competing “neighbors” remains the central empirical challenge in this literature. Theory provides some guidance on the issue through two mechanisms driving interactions across jurisdictions: factor mobility and yardstick competition.

On factor mobility, Tiebout (1956) first argued that labor and capital can move in response to differential tax rates across jurisdictions. Factor mobility leads naturally to the postulated social interactions matrix being (i) geographic neighbors, given labor mobility, and (ii) jurisdictions with similar economic or demographic characteristics, given capital mobility (Case *et al.*, 1989).

Yardstick competition is driven by voters making comparisons between states to learn about their own politician’s quality (Shleifer, 1985). Besley and Case (1995) formalize the idea in a model where voters use taxes set by governors in other states to infer their own governor’s quality. Yardstick competition leads naturally to the postulated interactions matrix being “political neighbors”: states that voters make comparisons to.

In this application, the number of nodes and time periods is relatively low: the data covers mainland US states,  $N = 48$ , for the years 1962-2015,  $T = 53$ . Our approach identifies the structure of social interactions among “economic neighbors”, denoted  $W_{econ}$ . We contrast this against a null that state taxes are influenced by geographic neighbors,  $W_{geo}$ , as shown in Panel A of Figure 1A. With  $W_{econ}$  recovered, we can establish, beyond geography, what predicts the strength of ties between states and provide fresh insights on drivers of tax competition.

Before using the real data, we confirm the estimator’s performance when the true network is  $W_{geo}$  in simulated settings. In line with the findings of the previous section, Panel B of Figure 1A

shows that (i) the procedure recovers strong edges frequently (more specifically, 89% of the true strong edges are recovered) and (ii) performance deteriorates when recovering weak edges (72%). In all cases, the estimator does not add edges not in the true network. This suggests recovered economic links that deviate from geographic links may indeed carry signal, while weak links may not get detected. Finally, the estimator for  $\rho$  and  $\gamma$  may show some downward bias with the sample sizes in the application, consistent with the simulations in Appendix Figure A1.

## 4.1 Data and Empirical Specification

We denote state tax liabilities for state  $i$  in year  $t$  as  $\tau_{it}$ , covering state taxes collected from real per capita income, sales, and corporate taxes. We extend the sample used by Besley and Case (1995), that runs from 1962-1988 ( $T = 26$ ).<sup>29</sup> The outcome considered,  $\Delta\tau_{it}$ , is the change in tax liabilities between years  $t$  and  $(t - 2)$  because it might take a governor more than a year to implement a tax program. Their model implies a standard social interactions specification for the tax setting behavior of state governors:

$$\Delta\tau_{it} = \rho_0 \sum_{j=1}^N W_{0,ij} \Delta\tau_{jt} + \sum_{k=1}^K \sum_{j=1}^N W_{0,ij} x_{jkt} \gamma_{0,k} + \sum_{k=1}^K \beta_{0,k} x_{ikt} + \alpha_i + \alpha_t + \epsilon_{it}, \quad (12)$$

where  $k = 1, \dots, K$  are the covariates for state  $i$  in period  $t$ . Tax setting behavior is determined by (i) endogenous social effects arising through neighbors' tax changes ( $\sum_{j=1}^N W_{0,ij} \Delta\tau_{jt}$ ); (ii) exogenous social effects arising through neighbors' characteristics ( $\sum_{j=1}^N W_{0,ij} x_{jkt}$ ); and (iii) state  $i$ 's characteristics ( $x_{ikt}$ ), including income per capita, the unemployment rate, and the proportions of young and elderly in the state's population. All specifications include state and time effects ( $\alpha_i$ ,  $\alpha_t$ ). Due to the inclusion of the time effects  $\alpha_t$ , we normalize the rows of  $W_{econ}$  to one. Table A6 presents descriptive statistics for the Besley and Case (1995) sample and our extended sample.

Much of the earlier literature on tax competition has focused on endogenous social effects and ignored exogenous social effects by setting  $\gamma = 0$ . Our identification result allows us to relax this restriction and estimate the full typology of social effects described by Manski (1993). This is important because only endogenous social effects lead to social multipliers from tax competition, and they are crucial to identify as they can lead to a race-to-the-bottom or suboptimal public goods provision (Brennan and Buchanan, 1980; Wilson, 1986; Oates and Schwab, 1988).

---

<sup>29</sup>Besley and Case (1995) test their political agency model using a two-equation set-up: (i) on gubernatorial reelection probabilities and (ii) on tax setting. Our application focuses on the latter because this represents a social interaction problem. They use two tax series: (i) TAXSIM data (from the NBER), which runs from 1977-1988 and (ii) state tax liabilities series constructed from data published annually in the Statistical Abstract of the US, that runs from 1962-1988. All their results are robust to either series. We extend the second series.

## 4.2 Preliminary Findings

Table 1 presents our preliminary findings and comparison to Besley and Case (1995). Throughout this section, we refer to “OLS estimates” as the estimates of the main equation (12) when  $W_0$  is postulated as  $W_{geo}$  or  $W_{econ}$  and  $\rho_0$ ,  $\gamma_{0,k}$ , and  $\beta_{0,k}$  are estimated by OLS.<sup>30</sup> Column 1 shows those estimates where the postulated social interactions matrix is based on geographic neighbors, exogenous social effects are ignored and the panel includes all 48 mainland states but runs only from 1962-1988 as in Besley and Case (1995). Social interactions influence gubernatorial tax setting behavior:  $\hat{\rho}_{OLS} = .375$ . Column 2 shows this to be robust to instrumenting neighbors’ tax changes using the instrument set proposed by Besley and Case (1995): namely, instrumenting for  $\Delta\tau_{jt}$  using geographic neighbors’ lagged changes in per capita income and unemployment rates. These instruments are in the spirit of using exogenous social effects to instrument for neighbors’ tax changes.  $\hat{\rho}_{2SLS}$  is more than double the magnitude of  $\hat{\rho}_{OLS}$ , suggesting tax setting behaviors across jurisdictions are strategic complements.

Columns 3 and 4 replicate both specifications over the longer sample, confirming Besley and Case’s (1995) finding on social interactions to be robust.  $\hat{\rho}_{2SLS}$  is again more than double  $\hat{\rho}_{OLS}$ . The result in Column 4 implies that for every dollar increase in the average tax rates among geographic neighbors, a state increases its own taxes by 64 cents. This is similar to the headline estimate of Besley and Case (1995).<sup>31</sup>

## 4.3 Endogenous and Exogenous Social Interactions

We now move beyond much of the earlier literature to first establish whether there are endogenous and exogenous social interactions in tax setting. We first focus on the endogenous and exogenous social interaction parameters, and in the next subsection, we detail the identified social interactions matrix,  $\hat{W}_{econ}$ . To do so, we need to modify slightly how we instrument for neighbors’ tax changes: the instrument set proposed by Besley and Case (1995) based on geographic neighbors’ characteristics will generally be weaker when estimating the full specification in (12) because the instruments are now directly controlled for in (12). We use an Adaptive Elastic Net GMM approach, which instruments neighbors’ tax changes with the characteristics of all other states. With the inclusion of endogenous and exogenous social effects, this represents our preferred approach.

Columns 1 and 2 of Table 2 show OLS and GMM estimates for  $\rho$  obtained from the Adaptive Elastic Net procedure, where we still set  $\gamma = 0$  but use our preferred instrument set:  $\hat{\rho}_{GMM} = .709 >$

---

<sup>30</sup>We postulate that  $W$  is  $W_{econ}$  obtained by running the procedure in Section 3, retrieving  $\hat{W}$ , and re-running model (12) with  $W = \hat{W}$ . For such OLS estimates, we use robust standard errors and ignore the sampling uncertainty in the estimated  $W_{econ}$ .

<sup>31</sup>Nor is the magnitude very different from earlier work examining fiscal expenditure spillovers. For example, Case *et al.* (1989) find that US state governments’ levels of per-capita expenditure are significantly impacted by the expenditures of their neighbors, with a one-dollar increase in neighbors’ expenditures leading to a seventy-cent increase in own-state expenditures.

$\hat{\rho}_{OLS} = .649$ . Columns 3 and 4 estimate the full model in (12). Relative to when exogenous social effects are assumed away ( $\gamma = 0$ ), the OLS and GMM estimates of  $\rho$  are smaller, but we continue to find robust evidence of endogenous social interactions in tax setting. The specification in Column 4 represents our preferred one:  $\hat{\rho}_{GMM} = .452$  (with a standard error of .132). This value meets the requirements on  $\rho$  in Corollaries 3 and 4 for global identification.<sup>32</sup>

#### 4.4 Identified Social Interactions Matrix

Figure 1B shows how the structures of economic ( $\hat{W}_{econ}$ ) and geographic networks ( $W_{geo}$ ) differ, where connected edges imply that two states are linked in at least one direction (state  $i$  causally impacts state taxes in  $j$ , and/or *vice versa*). This comparison makes clear whether all states geographically adjacent to  $i$  matter for its tax setting behavior and whether there are non-adjacent states that influence its tax rate.

The left-hand panel of Figure 1B shows the network of geographic neighbors (whose edges are colored blue), onto which we superimpose edges *not* identified as links in  $W_{econ}$ ; dropped edges are in red. The vast majority of geographically adjacent states are irrelevant for tax setting behavior. The right-hand panel of Figure 1B adds new edges identified in  $\hat{W}_{econ}$  that are *not* part of  $W_{geo}$ ; these added edges are in green and represent non-geographically adjacent states through which social interactions occur. For tax-setting behavior, economic distance is imperfectly measured if we simply assume interactions depend only on physical distance.

Table 3 summarizes the comparison between  $W_{geo}$  and  $\hat{W}_{econ}$ .  $W_{geo}$  has 214 edges, while  $\hat{W}_{econ}$  has only 49.  $\hat{W}_{econ}$  and  $W_{geo}$  have 9 edges in common. Hence, the vast majority of geographical neighbors ( $205/214 = 96\%$ ) are not relevant for tax setting.  $\hat{W}_{econ}$  has 40 edges that are absent in  $W_{geo}$ , and the identified social interactions are more spatially dispersed than under the assumption of geographic networks. This is reflected in the far lower clustering coefficient in  $\hat{W}_{econ}$  than in  $W_{geo}$  (.042 versus .419).<sup>33</sup>

#### 4.5 Links and Reciprocity

Our estimation strategy identifies the continuous strength of links,  $W_{0,ij}$ , where  $W_{0,ij} > 0$  is interpreted as state  $j$  influencing outcomes in state  $i$ . This is useful because recent developments in tax competition theory, using insights from the social networks literature, suggest links need not be reciprocal (Janeba and Osterleh, 2013).

Table 3 reveals that only 12.2% of edges in  $\hat{W}_{econ}$  are reciprocal (all edges in  $W_{geo}$  are reciprocal by definition). Hence, tax competition is both spatially disperse and asymmetric. In most cases

<sup>32</sup>Table A7 shows the full set of exogenous social effects (so Columns 1 and 2 refer to the same specifications as Columns 3 and 4 in Table 2). Exogenous social effects operate through economic neighbors' unemployment rate, demographic characteristics, and their governor's age.

<sup>33</sup>The clustering coefficient is the frequency of the number of fully connected triplets over the total number of triplets.

where tax setting in state  $i$  is influenced by taxes in state  $j$ , the opposite is not true.

Given common time shocks  $\alpha_t$  in (12), row-sum normalization is required and ensures  $\sum_j W_{0,ij} = 1$ . Hence, for every state  $i$ , there will be at least one economic neighbor state  $j^*$  that impacts it, so  $W_{0,ij^*} > 0$ . This just reiterates that social interactions matter. On the other hand, our procedure imposes no restriction on the derived columns of  $\hat{W}_{econ}$ . It could be that a state does not affect any other state. To see this in more detail, the final rows of Table 3 report the degree distribution across states, splitting for in-networks and out-networks. In  $W_{geo}$ , the in-degree is by construction equal to the out-degree, as all ties are reciprocal. The greater sparsity of the network of economic neighbors is reflected in the degree distribution being lower for  $\hat{W}_{econ}$  than  $W_{geo}$ . In  $\hat{W}_{econ}$ , the dispersion of in- and out-degree networks is very different (as measured by the standard deviation), being nearly nine times higher for the in-degree. Hence one reason for so few reciprocal ties being in the economic network is that out-degree network ties are rarely also in-degree ties.

This asymmetry in  $\hat{W}_{econ}$  further suggests that some highly influential states drive tax setting behavior in other states. To see which states these are, Figure 2 shows a histogram for the number of out-degree links from states. Twenty states have an out-degree of zero, so their tax rates have no direct impact on any other state's tax setting behavior. The most influential states in terms of the highest out-degree are Alabama (directly impacts tax setting behavior in five other states) and South Carolina, Pennsylvania, and Montana (which each directly impact tax setting behavior in four other states). Taking South Carolina as an example, the four states that it directly impacts include its geographic neighbor, Georgia, as well as non-geographic neighbors Missouri, Montana, and Virginia.

## 4.6 Factor Mobility or Yardstick Competition?

We use  $\hat{W}_{econ}$  to shed light on the roles of factor mobility and yardstick competition in driving tax competition. To do so, we estimate the factors correlated with the existence of links between states  $i$  and  $j$  in  $\hat{W}_{econ}$  relative to  $\hat{W}_{geo}$ . For state pairs with non-zero links in either  $\hat{W}_{econ}$  or  $\hat{W}_{geo}$ , we define a dummy outcome  $\hat{W}_{econ,ij} = 1$  if a link between states  $i$  and  $j$  is estimated under  $\hat{W}_{econ}$  and  $\hat{W}_{econ,ij} = 0$  if a link between states  $i$  and  $j$  exists under  $\hat{W}_{geo}$  but not under  $\hat{W}_{econ}$ . We examine correlates of links using the following dyadic regression:

$$\hat{W}_{econ,ij} = \lambda_0 + \lambda_1 X_{ij} + \lambda_2 X_i + \lambda_3 X_j + u_{ij}, \quad (13)$$

estimated using a linear probability model. The elements  $X_{ij}$ ,  $X_i$ , and  $X_j$  correspond to characteristics of the pair of states  $(i, j)$ , state  $i$ , and state  $j$ , respectively. Covariates are time-averaged over the sample period, and robust standard errors are reported.

Table 4 presents the dyadic regression results. Column 1 controls only for the distance between states  $i$  and  $j$ : this is highly predictive of an economic link between them. This reflects that the

economic network of state  $i$  often comprises states are in the same region, but not necessarily contiguous to state  $i$ . Column 2 adds two  $X_{ij}$  covariates to capture the economic and demographic homophily between states  $i$  and  $j$ . GDP homophily is the absolute difference in the states' GDP per capita. Demographic homophily is the absolute difference in the share of young people (aged 5-17) plus the absolute difference of the share of elderly people (aged 65+) across the states. GDP homophily does not predict economic ties, whereas demographic homophily does.

Columns 3-5 then sequentially add in several sets of controls. For labor mobility, we use net state-to-state migration data to control for the net migration flow of individuals from state  $i$  to state  $i$  (defined as the flow from  $i$  to  $j$  minus the flow from  $j$  to  $i$ ).<sup>34</sup> We then add a political homophily variable between states. For any given year, this is set to one if a pair of states have governors of the same political party. As this is time-averaged over our sample, this element captures the share of the sample in which the states have governors of the same party. Lastly, we include whether state  $j$  is considered a tax haven (and so might have disproportionate influence on other states). Based on Findley *et al.* (2012), the following states are coded as tax havens: Nevada, Delaware, Montana, South Dakota, Wyoming, and New York.

Column 5 shows that with this full set of controls, distance remains a robust predictor of the existence of economic links between states. However, the identified economic network highlights additional significant predictors of tax competition between states: political homophily *reduces* the likelihood of a link, suggesting any yardstick competition driving social interactions occurs when voters compare their governor to those of the opposing party in other states. Tax haven states appear to be especially less influential in the tax setting behaviors of other states. This mirrors what was observed in Figure 2, where some of the prominent tax havens – Nevada, Delaware, and New York – were all identified to have zero out-degree links. The relatively weak influence of tax haven states eases concerns over a race-to-the-bottom in tax setting behaviors.

Column 6 controls for state  $i$  and state  $j$  fixed effects. This reinforces the idea that distance and political homophily correlate to the strength of influence states tax setting has on others (the tax haven dummy cannot be separately identified in this specification). Labor mobility between states does not robustly predict the existence of economic ties.

## 4.7 Dynamics

As in our identification result, our empirical approach has taken the network structure as fixed over the entire sample period. In the context of tax competition over our study period, this might be a strong assumption. We examine the issue in more detail by allowing the estimated  $W_{econ}$  matrix

---

<sup>34</sup>We also experimented with alternative measures of labor migration, and the results were qualitatively the same. State-to-state migration data are based on year-to-year address changes reported on individual income tax returns filed with the IRS. The data cover filing years 1991 through 2015 and include the number of returns filed, which approximates the number of households that migrated, and the number of personal exemptions claimed, which approximates the number of individuals who migrated. The data are available at <https://www.irs.gov/statistics/soi-tax-stats-migration-data> (accessed September 2017).

to vary over time by changing the weight placed on observations from any given time period. More precisely, for any given time period  $t$ , we weight observations using a Gaussian kernel with its center varying period-by-period from 1962 to 2015. The variance of the kernel is set such that 75% of the weight is given to the first half of the data (i.e., pre-1988) when the kernel is centered in 1962. Figure A4 shows the kernel employed as we vary its center: the solid kernel is centered in 1962, the start of our sample – when we place the most weight on observations from 1962. The static case considered previously is akin to using a uniform kernel over all periods. This kernel weighting approach is outlined in Section 2.3.3. We fully describe the algorithm in Appendix Section C.2.

We begin by considering time-varying estimates of the endogenous and exogenous social interaction parameters from the full model in (12). The results for the endogenous social interactions parameter are shown in Figure 3, where the shaded areas are the 95% confidence intervals of the period-by-period estimates. Panel A shows that OLS estimates of  $\hat{\rho}$  drift up over time, so the strength of endogenous interactions increases from around .35 in the late 1960s to .50 by the 2010s. In all periods, we can reject the null that the endogenous social effect is zero. Recall the earlier static estimate was  $\hat{\rho}_{OLS} = .375$ .

Panel B shows the estimated endogenous social effect when we use GMM based on the characteristics of all other states as IVs. This also drifts up from around .35 in the late 1960s to .50 by the 2010s. In the majority of periods, we can reject the null of no endogenous social effect.<sup>35</sup>

Figure 4 shows the evolution of  $\hat{W}_{econ}$  over time as we center the kernel on different periods, following the same color-coding as in Figure 1. In all periods, geography-based edges play little role, and over time, the economic network becomes denser. This highlights not only that economic networks for tax competition always differ starkly from geography-based networks, but that the nature of economic networks relevant to tax competition has changed steadily over time.

Figure 5 shows how the features of  $\hat{W}_{econ}$  evolve as we place greater weight from early to later periods. For each statistic, we plot the period-by-period estimate when we center the kernel in any given period. The resulting smoothed estimates are then shown. To ease exposition, networks edges with  $W_{0,ij} < 1/47$  are removed. This cutoff is chosen as, in theory, states can only link at maximum with 47 other states. Panel A shows the share of edges that are kept from the previous estimate (centered in the previous period). We see relatively high stability in  $\hat{W}_{econ}$  with the smoothed estimate suggesting more than 60% of edges always being kept from one estimate to the next, with this stability increasing from the late 1980s.

Panel B shows how the overlap between  $\hat{W}_{econ}$  and  $W_{geo}$  varies over time, as measured by the share of edges that are only present in  $\hat{W}_{econ}$ . There is little overlap between the two networks over the entire sample. The smoothed estimate suggests that at least 80% of identified edges in  $\hat{W}_{econ}$  are never in  $W_{geo}$ . The divergence between economic and geographic neighbors becomes starker from the mid-1980s onwards.

---

<sup>35</sup>Standard errors are estimated without imposing the restriction that parameters vary slowly over time, and fluctuations across periods reflect variations in the network across periods.



Panels C and D show how the clustering and reciprocity of links in  $\hat{W}_{econ}$  vary as we shift the weight to later observations. Clustering of  $\hat{W}_{econ}$  increases from the 1960s through to the early 2000s. Thereafter, social interactions in tax competition become sparser. We also observe a reversal in the extent to which social interactions are reciprocal, with reciprocity rising to a peak in the early 1980s – when 20% of ties were reciprocal – and slowly falling thereafter.

Taken together, the results suggest the nature of tax competition between US states has changed over time through two mechanisms: (i) the strength of endogenous social interactions ( $\hat{\rho}$ ) has increased over time and (ii) the network of states interacted with ( $\hat{W}_{econ}$ ) varies over time. This has important implications for policy evaluation: the same intervention might have different spillover effects if implemented at different moments in time due to the evolution of  $\hat{\rho}$  and  $\hat{W}_{econ}$ . We consider this next using counterfactual simulations.

## 4.8 Counterfactuals

We use a counterfactual exercise to contrast how shocks to tax setting in a given state propagate under  $\hat{W}_{econ}$ , relative to what would have been predicted under  $W_{geo}$ . We do so for both static and dynamic estimates of  $\hat{W}_{econ}$ . We focus on South Carolina (SC), a state with one of the highest out-degree, as shown in Figure 2. We consider a scenario in which SC exogenously increases its taxes per capita by 10%. We measure the differential change in equilibrium state taxes in state  $j$  under the two network structures using the following statistic:

$$\Upsilon_j = \log(\Delta\tau_{jt}|\hat{W}_{econ}) - \log(\Delta\tau_{jt}|W_{geo}), \quad (14)$$

so that positive (negative) values imply equilibrium taxes being higher (lower) under  $\hat{W}_{econ}$ .<sup>36</sup>

Starting with the static case, Panel A of Figure 6 shows for each mainland US state the spillover effects through the economic network of tax competition. This highlights positive spillovers on tax rates in many states that are not geographic neighbors of SC. Panel B graphs  $\Upsilon_j$  to make precise how spillovers derived from  $\hat{W}_{econ}$  diverge from those predicted under  $W_{geo}$ . In 26 states,  $\Upsilon_j$  is smaller than .01% because both networks predict negligible spillovers to those states. In the remaining 22 mainland states, there is a wide discrepancy between the equilibrium state tax rates predicted under  $\hat{W}_{econ}$  relative to  $W_{geo}$ :  $\Upsilon_j$  varies from  $-1$  to  $4.03$ . The long-run effect in SC itself is also higher under  $\hat{W}_{econ}$  than under  $W_{geo}$ . The former states that given feedback effects, the long-run increase in tax rates in SC from a 10% increase is 11.4%, while the geographically based network implies a smaller equilibrium increase of 10.3%.

As  $\hat{W}_{econ}$  is spatially more dispersed than  $W_{geo}$ , the general equilibrium effects are different under the two network structures. Table 5 summarizes the general equilibrium implications for tax inequality under  $\hat{W}_{econ}$  and the  $W_{geo}$  counterfactual. The average tax rate increase under

---

<sup>36</sup>For  $W_{geo}$ , we calculate the counterfactual at  $\hat{\rho}_{GMM} = .452$ , the endogenous effect parameter estimated in our preferred specification, Column 4 of Table 2.

$\hat{W}_{econ}$  is three times that estimated under  $W_{geo}$ . Moreover, the dispersion of tax rates across states increases under  $\hat{W}_{econ}$  relative to  $W_{geo}$ . Finally, assuming interactions are based solely on geographic neighbors, we miss the fact that many states have relatively small tax increases.

We can repeat the exercise using the dynamically estimated economic network. Throughout, we calculate the general equilibrium effects of the *same* policy experiment: SC increasing its taxes per capita by 10%. These general equilibrium effects vary over the sample period because the strength of social interactions in tax competition vary ( $\hat{\rho}_{GMM}$ ), as shown in Figure 3, and identified economic neighbors vary over time ( $\hat{W}_{econ}$ ), as shown in Figure 4. The results are summarized in Figure 7. Placing weight on the early or later part of the sample generates similar changes in average tax rates and their variance in general equilibrium – with both being lower than simulated under the static model. Placing more weight on the middle of the sample period generates higher changes in average tax rates and their variance in general equilibrium.

The differential general equilibrium impacts found as we place different weights across sample observations links to recent discussions on the external validity of internally valid causal impacts based on micro-evidence. While the earlier literature has emphasized the potential interaction of treatment effects with aggregate shocks (Rosenzweig and Udry, 2020) or how behavioral responses to social insurance policies vary over the business cycle (Kroft and Notowidigdo, 2016), our analysis provides another explanation for the changing impacts of policies where social interactions determine behavior: changes in the strength of social interactions and the network of economic interactions.

## 5 Discussion

In a canonical social interactions model, we provide sufficient conditions under which the social interactions matrix, and endogenous and exogenous social effects are all globally identified, even absent information on social links. Our identification strategy is novel and may bear fruit in other areas. The method is immediately applicable to other classic social interactions problems, but where data on social links is either missing or partial. In fields such as macroeconomics, political economy, and trade, there are core areas of research where social interactions across jurisdictions/countries etc. drive key outcomes, panel data exist over many periods, and the number of nodes is relatively fixed. Moreover, while our discussion and application have focused on a continuous policy response (state taxes), our methods can also be applied to the extensive margin of policy adoption and diffusion. Such diffusion models might generate network interactions where some states influence the later adoption of economic and social policies in other jurisdictions. This issue is studied by DellaVigna and Kim (2022) in the context of US state policies – they examine the diffusion of over 700 policies in the past 70 years. Their work also suggests the nature of interactions across states has changed: while geographic proximity is a good predictor of policy diffusion, they also find that since 2000, political alignment across states has become the strongest

predictor of diffusion.

In finance, high-frequency panel data is readily available and relevant for the study of core research questions. For example, a long-standing question has been whether CEOs are subject to relative performance evaluation, and if so, what is the comparison set of firms/CEOs used (Edmans and Gabaix, 2016). More generally, our method can be readily applied to a large class of economic questions around contagion, risk, and the fragility of economic and environmental systems. For example, since the financial crisis of 2008, it has become clear that linkages between actors such as firms or banks are complex and often hidden, yet because endogenous network interactions cause feedback loops and have multiplier effects, they can have enormous implications for the evolution of a financial crisis or the propagation of supply shocks in aggregate. Identifying such synchronicity is a critical first step to putting in place policies to reduce the fragility of economic systems (van Vliet, 2018; Elliott and Golub, 2022; Goldstein *et al.*, 2022).

Advances in the availability of administrative data, data from social media, mobile technologies, and online economic transactions all offer new possibilities to identify social interactions with long panels or high-frequency data collection, where data on social ties will typically be missing.

Three further directions for future research are of note. First, under partial observability of  $W_0$  (as in Blume *et al.*, 2015), the number of parameters in  $W_0$  to be retrieved falls quickly. Our approach can then still be applied to complete knowledge of  $W_0$ , such as if Aggregate Relational Data is available, and this could be achieved with potentially weaker assumptions for identification, and in even shorter panels. To illustrate possibilities, Figure A5 shows results from a final simulation exercise in which we assume the researcher starts with partial knowledge of  $W_0$ . We do so for Banerjee *et al.* (2013) village family network, showing simulation results for scenarios in which the researcher knows the social ties of the three (five, ten) households with the highest out-degree. For comparison, we also show the earlier simulation results when  $W_0$  is entirely unknown. This clearly illustrates that with partial knowledge of the social network, performance on all metrics improves rapidly for any given  $T$ .

Second, we have developed our approach in the context of the canonical linear social interactions model (1). This builds on Manski (1993) when  $W_0$  is known to the researcher, and the reflection problem is the main challenge in identifying endogenous and exogenous social effects. However, the reflection problem is functional-form dependent and may not apply to many non-linear models (Blume *et al.*, 2011, Blume *et al.*, 2015). An important topic for future research is to extend the analysis to non-linear social interaction settings. Relatedly, the canonical social interaction model assumes that the same  $W_0$  governs the endogenous and exogenous channels. Despite the relaxation we propose in Section 2.3.3, we see this as a limitation of the current method, and future research is needed to allow for a fully flexible approach.

Finally, an important part of the social networks literature examines endogenous network formation (Jackson *et al.*, 2017; de Paula, 2017). Our analysis allows us to begin probing the issue in two ways. First, the kind of dyadic regression analysis in Section 4 on the correlates of entries

in  $W_{0,ij}$  suggests factors driving link formation and dissolution. Second, this leads naturally to a broad agenda going forward, to address the challenge of simultaneously identifying and estimating time varying models of network formation and social interaction, all in cases where data on social networks is not required.

## References

- ACEMOGLU, D., V. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80, 1977–2016.
- ANSELIN, L. (2010). Thirty Years of Spatial Econometrics. *Papers in Regional Science*, 89, 3–25.
- ATALAY, E., A. HORTACSU, J. ROBERTS, AND C. SYVERSON (2011). Network Structure of Production. *Proceedings of the American Mathematical Society*, 108, 5199–202.
- BALLESTER, C., A. CALVO-ARMENDOL, AND Y. ZENOU (2006). Who’s Who in Networks. Wanted: The Key Player. *Econometrica*, 74, 1403–17.
- BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013). The Diffusion of Microfinance. *Science*, 341, 1236498.
- BATTAGLINI, M., E. PATACCINI, AND E. RAINONE (2022). Endogenous social interactions with unobserved networks. *The Review of Economic Studies*, 89, 1694–1747.
- BELLONI, A. AND V. CHERNOZHUKOV (2011). High Dimensional Sparse Models: an Introduction. Tech. Rep. 1106.5242, arXiv.org Collection, <http://arxiv.org/abs/1106.5242>.
- BESLEY, T. AND A. CASE (1994). Unnatural Experiments? Estimating the Incidence of Endogenous Policies. *NBER Working Paper 4956*.
- (1995). Incumbent Behavior: Vote-seeking, Tax-setting, and Yardstick Competition. *American Economic Review*, 85, 25–45.
- BLUME, L., W. A. BROCK, S. N. DURLAUF, AND Y. IOANNIDES (2011). Identification of Social Interactions. in *Handbook of Social Economics*, ed. by J. Behabib, A. Bisin, and M. O. Jackson, Noth-Holland, vol. 1B.
- BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND R. JAYARAMAN (2015). Linear Social Interactions Models. *Journal of Political Economy*, 123, 444–96.
- BONALDI, P., A. HORTACSU, AND J. KASTL (2015). An Empirical Analysis of Funding Costs Spillovers in the EURO-zone with Application to Systemic Risk. Princeton University Working Paper.
- BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009). Identification of Peer Effects Through Social Networks. *Journal of Econometrics*, 150, 41–55.
- BRENNAN, G. AND J. BUCHANAN (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*, Cambridge University Press.

- BREZA, E., A. G. CHANDRASEKHAR, T. H. MCCORMICK, AND M. PAN (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110, 2454–84.
- BURSZTYN, L., F. EDERER, B. FERMAN, AND N. YUCHTMAN (2014). Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions. *Econometrica*, 82, 1273–301.
- CANER, M. AND H. H. ZHANG (2014). Adaptive Elastic Net for Generalized Method of Moments. *Journal of Business and Economic Statistics*, 32, 30–47.
- CASE, A., J. R. HINES, AND H. S. ROSEN (1989). Copycatting: Fiscal Policies of States and Their Neighbors. *NBER Working Paper 3032*.
- CHANDRASEKHAR, A. AND R. LEWIS (2016). Econometrics of Sampled Networks. Working Paper.
- CHANEY, T. (2014). The Network Structure of International Trade. *American Economic Review*, 104, 3600–34.
- COLEMAN, J. S. (1964). *Introduction to Mathematical Sociology*, London Free Press Glencoe.
- CONLEY, T. G. AND C. R. UDRY (2010). Learning About a New Technology: Pineapple in Ghana. *American Economic Review*, 100, 35–69.
- DAHLHAUS, R. (2012). 13 - Locally Stationary Processes. in *Time Series Analysis: Methods and Applications*, ed. by T. Subba Rao, S. Subba Rao, and C. Rao, Elsevier, vol. 30 of *Handbook of Statistics*, 351–413.
- DE GIORGI, G., M. PELLIZZARI, AND S. REDAELLI (2010). Identification of Social Interactions through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics*, 2, 241–75.
- DE PAULA, A. (2017). Econometrics of Network Models. in *Advances in Economics and Econometrics: Theory and Applications*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press.
- DE PAULA, A., E. TAMER, AND S. RICHARDS-SHUBIK (2018). Identifying Preferences in Networks with Bounded Degree. *Econometrica*, 86, 263–288.
- DELLAVIGNA, S. AND W. KIM (2022). Policy Diffusion and Polarization across US States. *National Bureau of Economic Research*.
- EDMANS, A. AND X. GABAIX (2016). Executive Compensation: a modern primer. *Journal of Economic literature*, 54, 1232–87.
- EFRON, B., T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI (2004). Least angle regression. *The Annals of statistics*, 32, 407–499.
- ELLIOTT, M. AND B. GOLUB (2022). Networks and Economic Fragility. *Annual Review of Economics*, 14, 665–696.

- FETZER, T., P. C. SOUZA, O. VANDEN EYNDE, AND A. L. WRIGHT (2021). Security transitions. *American Economic Review*, 111, 2275–2308.
- FINDLEY, M., D. NIELSON, AND J. SHARMAN (2012). Global Shell Games: Testing Money Launderers’ and Terrorist Financiers’ Access to Shell Companies. Griffith University Working Paper.
- GAUTIER, E. AND C. ROSE (2016). Inference in Social Effects when the Network is Sparse and Unknown. (in preparation).
- GAUTIER, E. AND A. TSYBAKOV (2014). High-Dimensional Instrumental Variables Regression and Confidence Sets. Working Paper CREST.
- GEFANG, D., S. G. HALL, AND G. S. TAVLAS (2023). Identifying Spatial Interdependence in Panel Data with Large N and Small T. .
- GOLDSTEIN, I., A. KOPYTOV, L. SHEN, AND H. XIANG (2022). Synchronicity and Fragility. *Manuscript*.
- GRANOVETTER, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 6, 1360–80.
- HASTIE, T. AND R. TIBSHIRANI (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 757–796.
- HOFF, P. D., A. E. RAFTERY, AND M. S. HANDCOCK (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97, 1090–1098.
- HOLLAND, P. W. AND S. LEINHARDT (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76, 33–50.
- HORN, R. A. AND C. R. JOHNSON (2013). *Matrix Analysis*, Cambridge University Press.
- JACKSON, M., B. ROGERS, AND Y. ZENOU (2017). The Economic Consequences of Social Network Structure. *Journal of Economic Literature*, 55, 49–95.
- JACKSON, M. AND A. WOLINSKY (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71, 44–74.
- JANEBA, E. AND S. OSTERLEH (2013). Tax and the City – A Theory of Local Tax Competition. *Journal of Public Economics*, 106, 89–100.
- KAPETANIOS, G., R. M. MASOLO, K. PETROVA, AND M. WALDRON (2019). A time-varying parameter structural model of the UK economy. *Journal of Economic Dynamics and Control*, 106, 103705.
- KLINE, B. AND E. TAMER (2016). Bayesian Inference in a Class of Partially Identified Models. *Quantitative Economics*, 7, 329–366.
- KRANTZ, S. G. AND H. R. PARKS (2013). *The Implicit Function Theorem*, Birkhauser.
- KROFT, K. AND M. J. NOTOWIDIGDO (2016). Should unemployment insurance vary with the unemployment rate? Theory and evidence. *The Review of Economic Studies*, 83, 1092–1124.

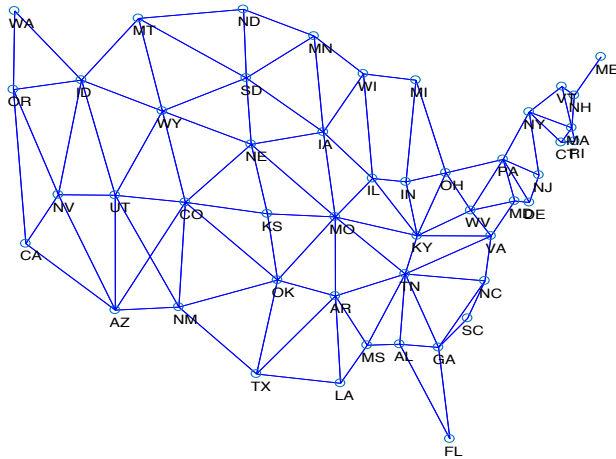
- LAM, C. AND P. C. SOUZA (2016). Detection and Estimation of Block Structure in Spatial Weight Matrix. *Econometric Reviews*, 35, 1347–1376.
- (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 38, 693–710.
- LEE, L.-F. (2007). Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. *Journal of Econometrics*, 60, 531–42.
- LEWBEL, A., X. QU, AND X. TAN (2022). Social Networks with Unobserved Links. *Journal of Political Economy* (forthcoming).
- MANRESA, E. (2016). Estimating the Structure of Social Interactions Using Panel Data. Manuscript.
- MANSKI, C. F. (1993). Identification of Endogenous Social Effects: the reflection problem. *The Review of Economic Studies*, 60, 531–42.
- MEINSHAUSEN, N. AND P. BUHLMANN (2006). High-Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, 34, 1436–1462.
- NEUMAN, E. J. AND M. S. MIZRUCHI (2010). Structure and bias in the network autocorrelation model. *Social Networks*, 32, 290–300.
- OATES, W. AND R. SCHWAB (1988). Economic Competition Among Jurisdictions: Efficiency-enhancing or Distortion-inducing?. *Journal of Public Economics*, 35, 333–54.
- ROSE, C. (2015). Essays in Applied Microeconometrics. Ph.D. thesis, University of Bristol.
- ROSENZWEIG, M. R. AND C. UDRY (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies*, 87, 343–381.
- ROTHENBERG, T. (1971). Identification in Parametric Models. *Econometrica*, 39, 577–91.
- ROTHENHÄUSLER, D., C. HEINZE, J. PETERS, AND N. MEINSHAUSEN (2015). BACKSHIFT: Learning causal cyclic graphs from unknown shift interventions. in *Advances in Neural Information Processing Systems*, 1513–1521.
- SACERDOTE, B. (2001). Peer Effects with Random Assingment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116, 681–704.
- SHLEIFER, A. (1985). A Theory of Yardstick Competition. *Rand Journal of Economics*, 16.
- SMITH, T. E. (2009). Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, 41, 307–332.
- TIBSHIRANI, R. J. AND J. TAYLOR (2012). Degrees of freedom in lasso problems. *The Annals of Statistics*, 40, 1198–1232.
- TIEBOUT, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64, 416–24.

- VAN VLIET, W. (2018). Connections as Jumps: Estimating Financial Interconnectedness from Market Data. Manuscript.
- WANG, W., E. J. NEUMAN, AND D. A. NEWMAN (2014). Statistical power of the social network autocorrelation model. *Social Networks*, 38, 88–99.
- WILSON, J. (1986). A Theory of Interregional Tax Competition. *Journal of Urban Economics*, 19, 296–315.
- (1999). Theories of Tax Competition. *National Tax Journal*, 52, 269–304.
- ZHOU, W. (2019). A Network Social Interaction Model with Heterogeneous Links. *Economics Letters*, 180, 50–53.
- ZOU, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101, 1418–29.
- ZOU, H. AND T. HASTIE (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.
- ZOU, H., T. HASTIE, AND R. TIBSHIRANI (2007). On the “Degrees of Freedom” of the LASSO. *The Annals of Statistics*, 35, 2173–2192.
- ZOU, H. AND H. H. ZHANG (2009). On the Adaptive Elastic-net with a Diverging Number of Parameters. *Ann. Statist.*, 37, 1733–51.

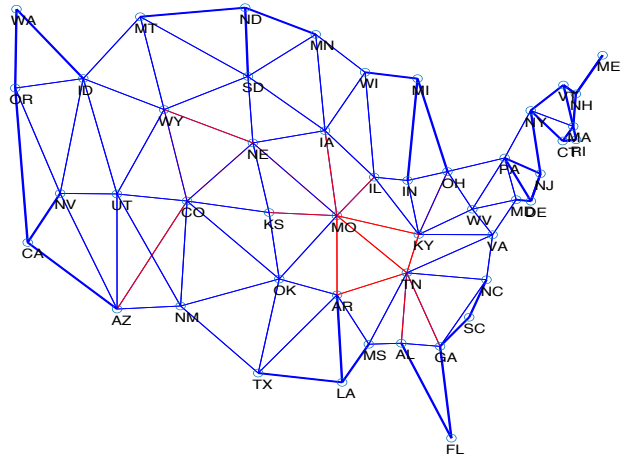


**Figure 1A: Network Graph of US States, Geographic Neighbors**

**A. Geographic Network**



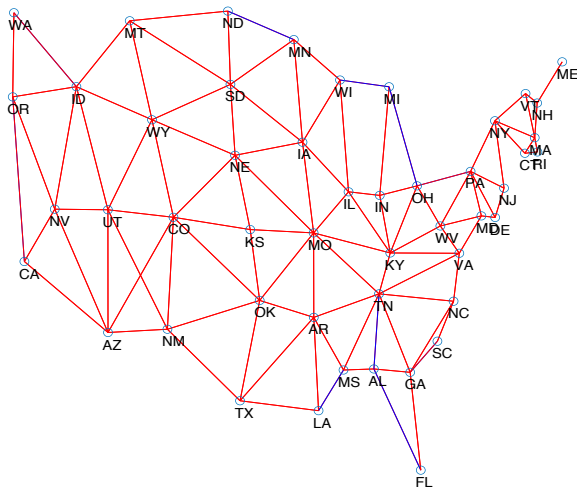
**B. Geographic Network (recovered in simulations)**



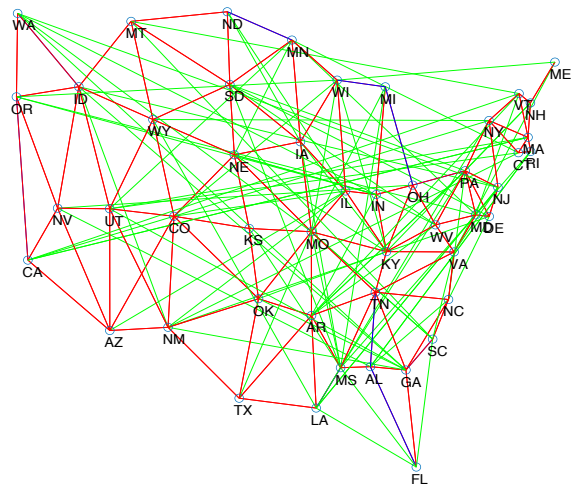
**Notes:** Panel A represents the continental US states (N=48). An edge is drawn between a pair of states if they share a geographic border. State abbreviations are as used by US Post Office (<http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf>). Panel B is the outcome of a simulation exercise where the true network is the geographic network. 1,000 Monte Carlo iterations were performed. The true parameters are  $\rho_0=0.3$ ,  $\beta_0=0.4$  and  $\gamma_0=0.5$ . All specifications include time and node fixed effects. Kept edges are depicted in blue: these links are estimated as non-zero in at least 5% of the iterations and are also non-zero in the true network. Added edges are depicted in green: these links are estimated as non-zero in at least 5% of the iterations but the edge is zero in the true network. Removed edges are depicted in red: these links are estimated as zero in at least 5% of the iterations but are non-zero in the true network. The figure further distinguish between strong and weak links: strong links are shown in thick edges (with strength is greater than or equal to .3).

**Figure 1B: Network Graph of US States, Identified Economic Neighbors**

**A. Economic Network (kept and removed edges only)**



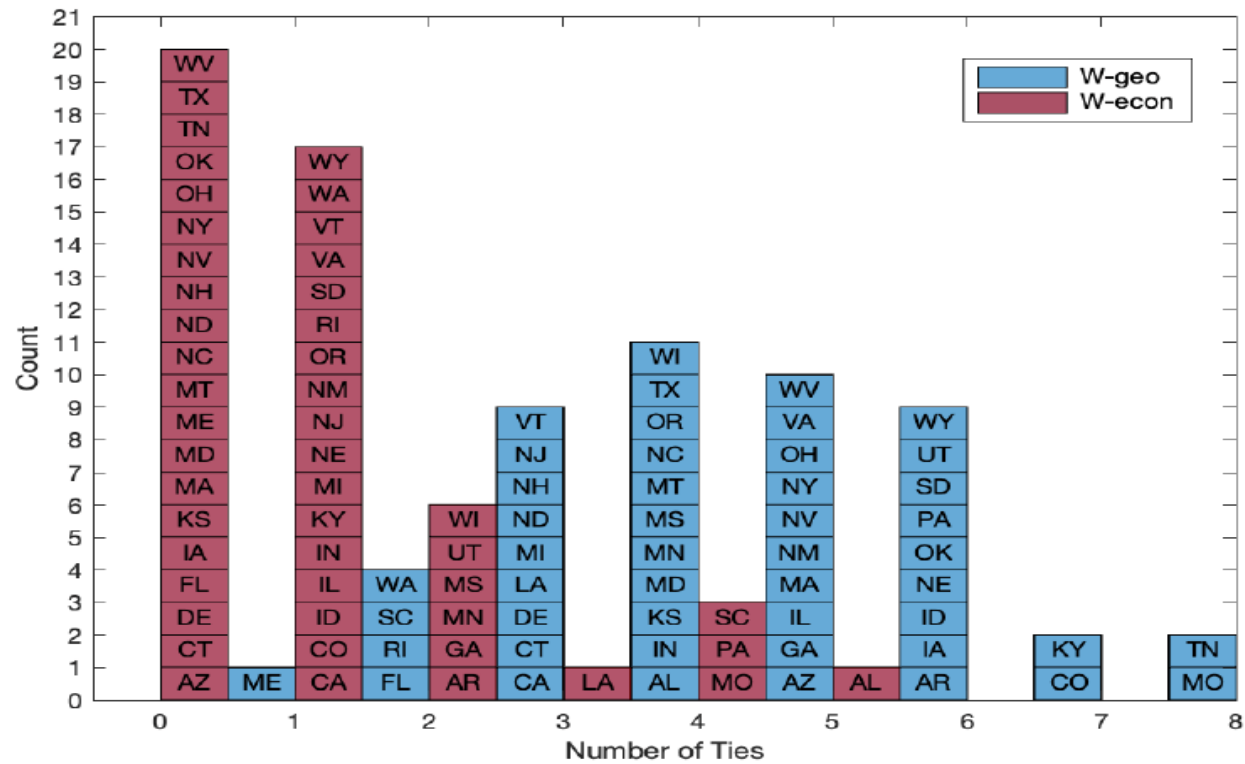
**B. Economic Network (all edges)**



**Geographic network edges**  
**Removed (geographic) edges in economic network**  
**New edges added in economic networks**

**Notes:** Figure 1B represents the continental United States (N=48). The economic network is derived from our preferred specification, where we penalize geographic neighbors to states, and allow for exogenous social effects. A blue edge is drawn between a pair of states if they are geographic neighbors and were estimated as connected. A red edge is drawn between a pair of states if they are geographic neighbors but were not estimated as connected. A green edge is drawn between a pair of states if they are not geographic neighbors and were estimated connected. The left hand side graph just shows red and blue edges. The right hand side shows all three types of edges. State abbreviations are as used by US Post Office (<http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf>).

## Figure 2: Out-degree Distribution

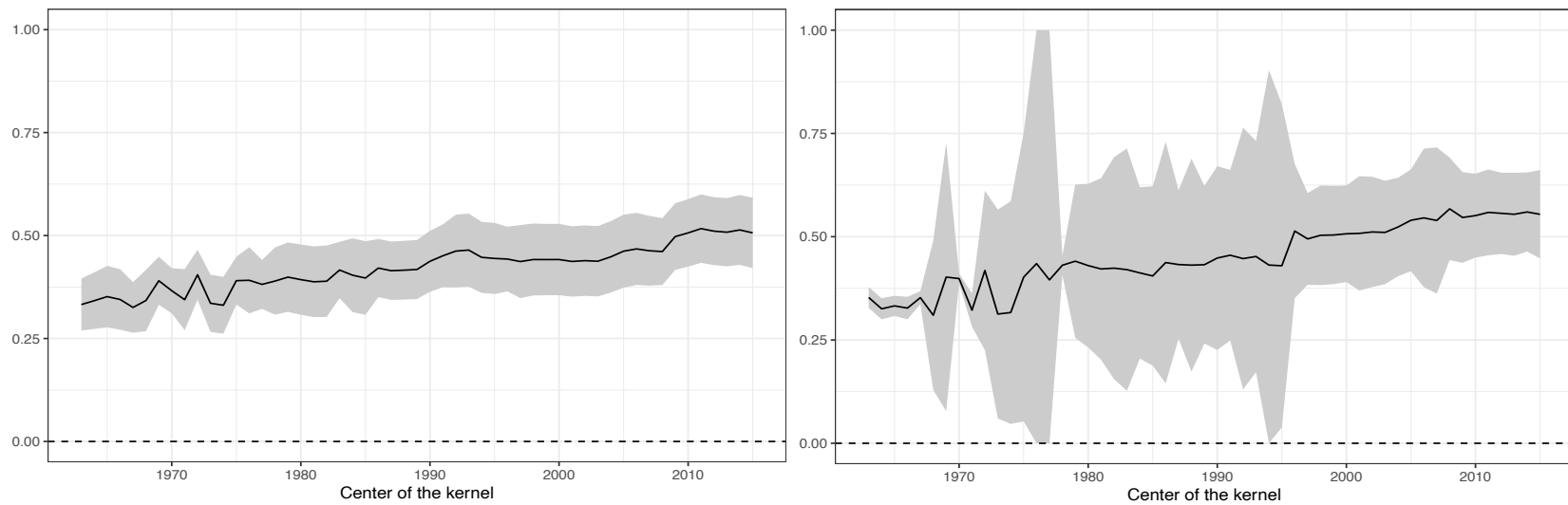


**Notes:** The out-degree distribution calculated from geographic neighbor's network (W-geo) is shown in blue. The distribution calculated from economic neighbor's network in (W-econ) is shown in red. State abbreviations are as used by US Post Office (<http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf>).

**Figure 3: Dynamic Endogenous Social Effects ( $\rho$ )**

**A. OLS**

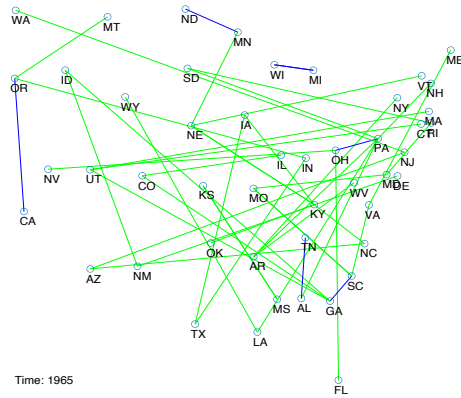
**B. GMM**



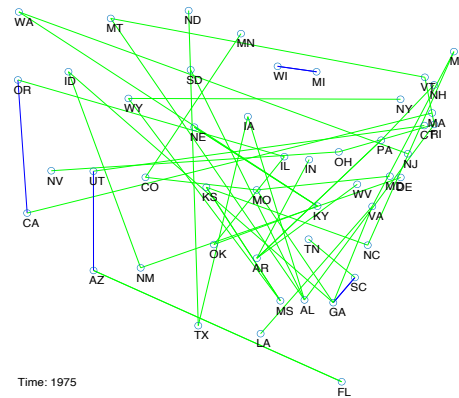
**Notes:** Panel A shows the OLS estimates of the endogenous social effects estimate ( $\rho$ ) with a Gaussian kernel with center varying period-by-period from 1962 to 2015. The variance of the kernel is set such that 75% of the weight is given to the first half of the data (i.e. pre-1988) with the kernel centered at 1962. Panel B shows the GMM estimates. Shaded areas are the 95% confidence intervals of the period-by-period estimates. Robust standard errors are shown in Panel A and standard errors based on the Caner and Zhang (2014) procedure are shown in Panel B.

**Figure 4: Dynamic Network Graph of US States, Identified Economic Neighbors**

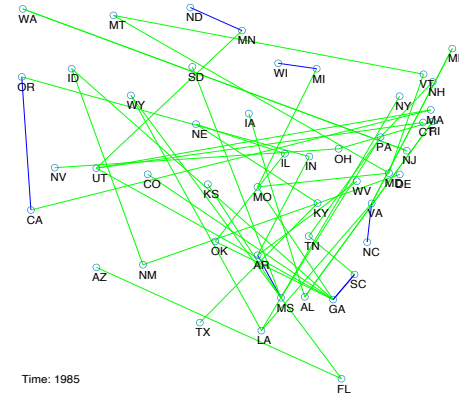
**Centered in 1965**



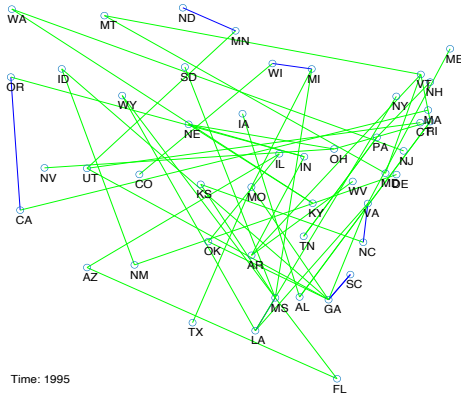
**Centered in 1975**



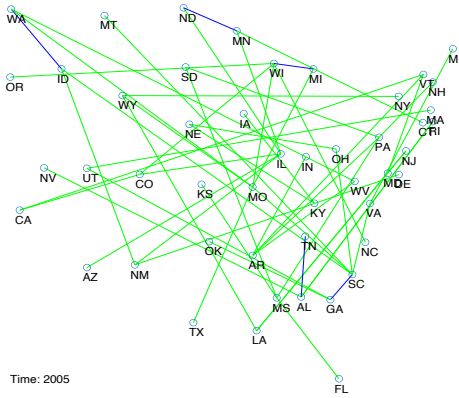
**Centered in 1985**



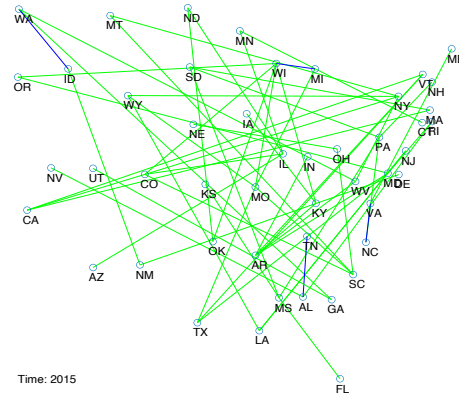
**Centered in 1995**



**Centered in 2005**



**Centered in 2015**

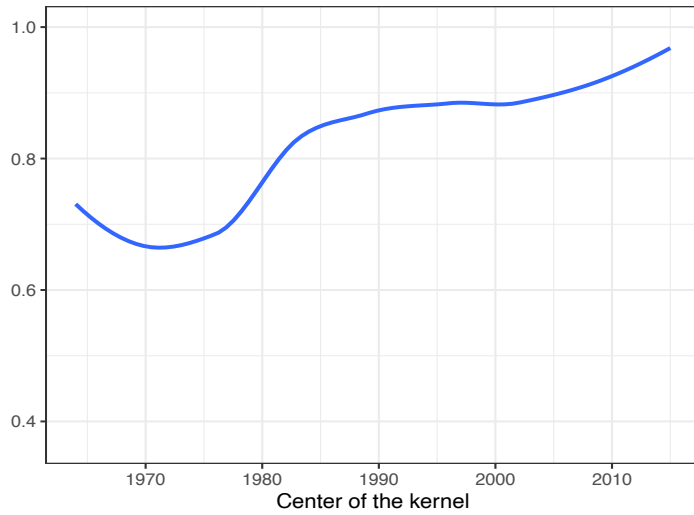


Geographic network edges  
 New edges added in economic networks

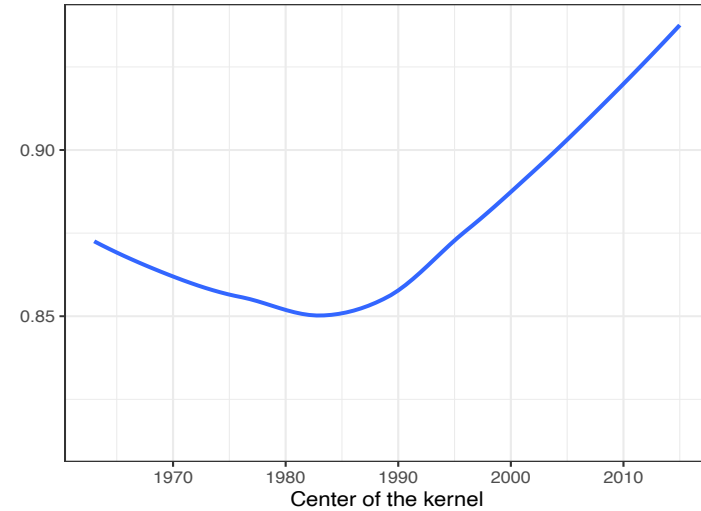
**Notes:** Each Figure represents the continental United States (N=48). The economic network is derived from our preferred specification, where we allow for exogenous social effects. A blue edge is drawn between a pair of states if they are geographic neighbors and were estimated as connected. A green edge is drawn between a pair of states if they are not geographic neighbors and were estimated connected. State abbreviations are as used by US Post Office (<http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf>).

## Figure 5: Dynamic Network Graph of US States, Identified Economic Neighbors

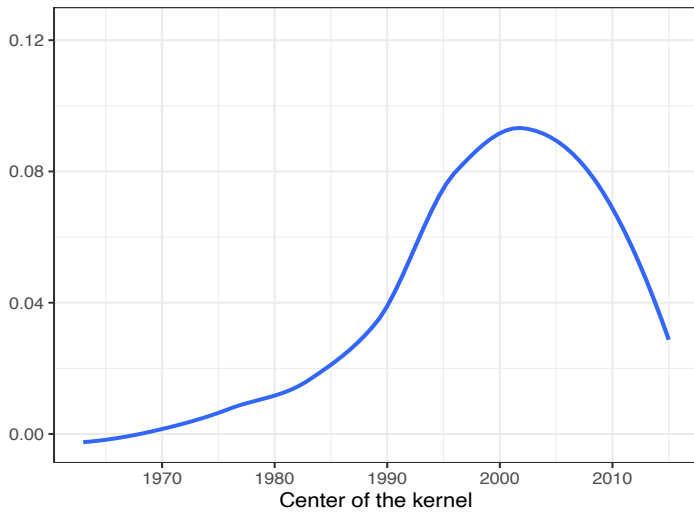
### A. Share of edges kept from previous network



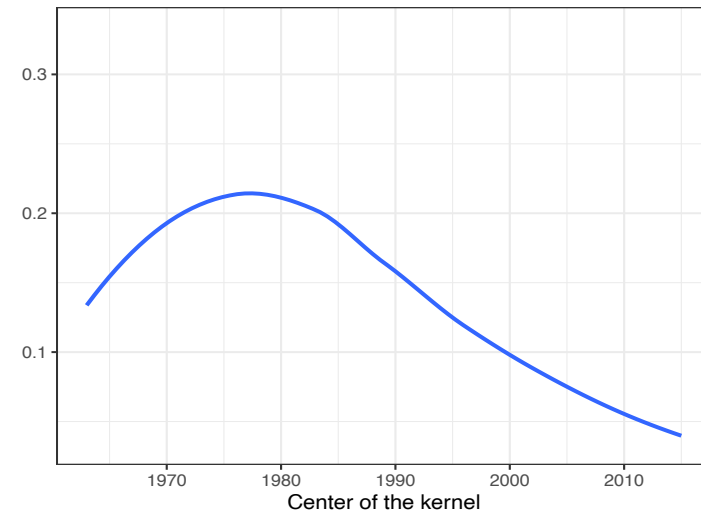
### B. Share of edges in W-econ only



### C. Clustering

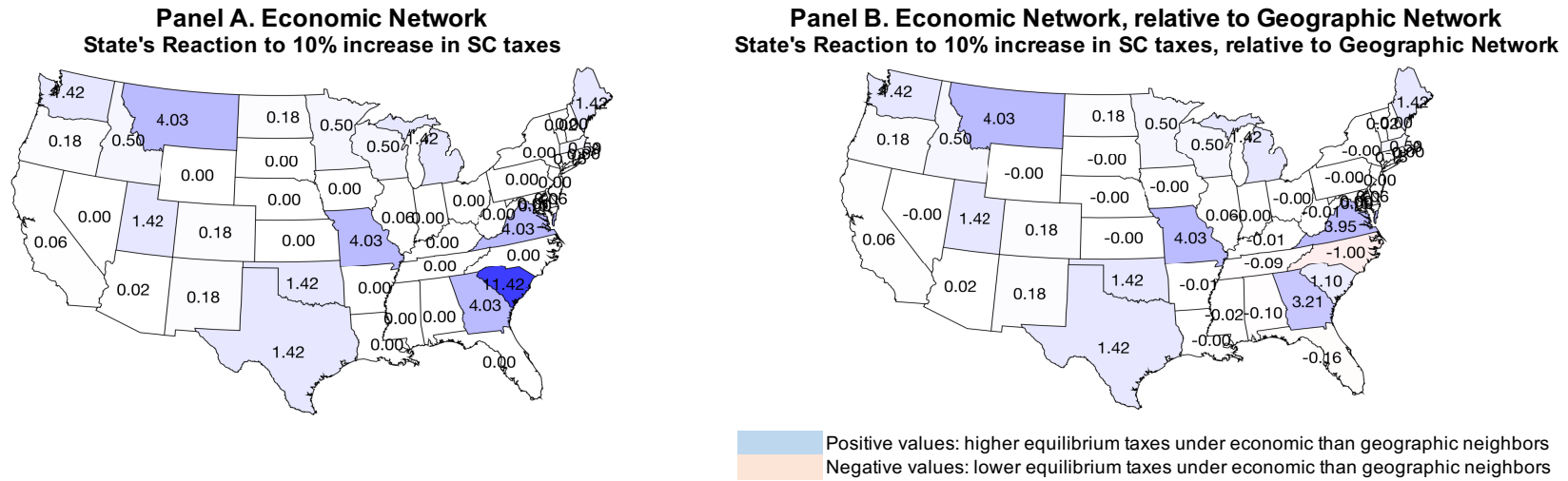


### D. Reciprocity



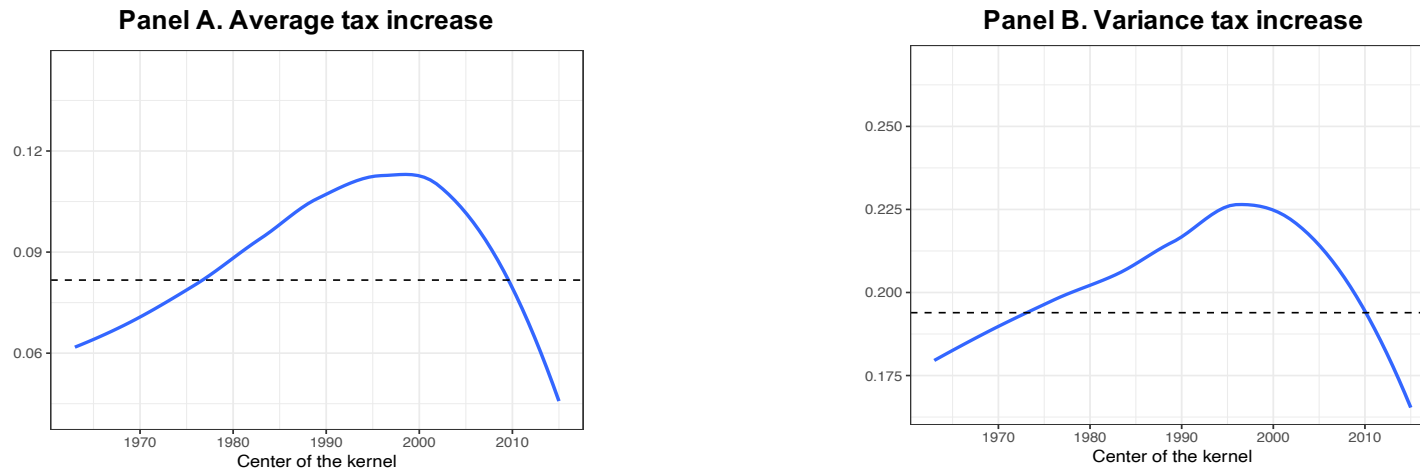
**Notes:** The panels show various statistics of the estimated economic networks using a Gaussian kernel with its center varying period-by-period from 1962 to 2015. The variance of the kernel is set such that 75% of the weight is given to the first half of the data (i.e. pre-1988) with the kernel centered at 1962. Panel A shows the share of edges kept from the network in the previous period. Panel B shows the share of edges in W-econ only (compared to W-geo). Panel C shows the clustering coefficient of the estimated economic networks. Panel D shows the share of reciprocated edges in the economic network. In all networks, edges smaller than  $1/47$  are removed. This cutoff is chosen as, in theory, states can only link at maximum with 47 other states. The blue line shows the smoothed estimates across time periods.

**Figure 6: General Equilibrium Impacts of South Carolina Tax Rises**



**Notes:** This shows the equilibrium impulse responses in taxes set in each state as a result of South Carolina increasing its tax change by 10%. We compare these derived tax changes under the identified economic network structure, relative to that assumed under a geographic neighbors structure. We graph the log change in equilibrium taxes under economic neighbors, minus the log change in equilibrium taxes under geographic neighbors. Positive values (red shaded) states indicate higher equilibrium taxes under economic neighbors than geographic neighbors, and negative values (blue shaded) states indicate lower equilibrium taxes under economic neighbors than geographic neighbors.

**Figure 7: Dynamic General Equilibrium Impacts of SC Tax Rise Shocks**



**Notes:** The panels show equilibrium impulse responses in taxes set in each state as a result of South Carolina increasing its tax change by 10%. These are based on estimated economic networks using a Gaussian kernel with its center varying period-by-period from 1962 to 2015. The variance of the kernel is set such that 75% of the weight is given to the first half of the data (i.e. pre-1988) with the kernel centered at 1962. Panel A shows the average tax increase for the kernel centered in each period (in gray dots) and the smoothed line (in blue). The horizontal dashed line corresponds to the average tax increase if the network was considered to be static. Panel B shows a similar construction for the variance of tax increase across states.

## Table 1: Geographic Neighbors

Dependent variable: Change in per capita income and corporate taxes

Coefficient estimates, standard errors in parentheses

	Besley and Case (1995) Sample		Extended Sample	
	(1) OLS	(2) 2SLS: IVs are Characteristics of Geographic Neighbors	(3) OLS	(4) 2SLS: IVs are Characteristics of Geographic Neighbors
<b>Geographic Neighbors' Tax Change (t - [t-2])</b>	.375*** (.080)	.868*** (.359)	.271*** (.050)	.642*** (.200)
<b>Period</b>	1962-1988	1962-1988	1962-2015	1962-2015
<b>First Stage (F-stat)</b>		14.2		24.2
<b>Controls</b>	Yes	Yes	Yes	Yes
<b>State and Year Fixed Effects</b>	Yes	Yes	Yes	Yes
<b>Observations</b>	1,296	1,248	2,544	2,544

**Notes:** \*\*\* denotes significance at 1%, \*\* at 5%, and \* at 10%. In all specifications, a pair of states are considered neighbors if they share a geographic border. The sample covers 48 mainland US states. In Columns 1 and 2 the sample runs from 1962 to 1988 (as in Besley and Case (1995)). In Columns 3 and 4 the sample is extended to run from 1962 to 2015. The dependent variable is the change in state i's total taxes per capita in year t. OLS regressions estimates are shown in Columns 1 and 3. Columns 2 and 4 show 2SLS regressions where each geographic neighbor's tax change is instrumented by lagged neighbor's state income per capita and unemployment rate. All regressions control for state i's income per capita in 1982 US dollars, state i's unemployment rate, the proportion of young (aged 5-17) and elderly (aged 65+) in state i's population, and the state governor's age. All specifications include state and time fixed effects. With the exception of governor's age, all variables are differenced between period t and period t-2. Robust standard errors are reported in parentheses.

## Table 2: Economic Neighbors

Dependent variable: Change in per capita income and corporate taxes

Coefficient estimates, standard errors in parentheses

	No Exogenous Social Effects		Exogenous Social Effects	
	(1) OLS	(2) GMM	(3) OLS	(4) GMM
<b>Economic Neighbors' Tax Change (t - [t-2])</b>	.649*** (.047)	.709*** (.035)	.402*** (.044)	.452*** (.132)
<b>Period</b>	1962-2015		1962-2015	
<b>Controls</b>	Yes	Yes	Yes	Yes
<b>State and Year Fixed Effects</b>	Yes	Yes	Yes	Yes
<b>Observations</b>	2,544	2,544	2,544	2,544

**Notes:** \*\*\* denotes significance at 1%, \*\* at 5%, and \* at 10%. The sample covers 48 mainland US states running from 1962 to 2015. The dependent variable is the change in state *i*'s total taxes per capita in year *t*. We allow for exogenous social effects in Columns 3 and 4. OLS regressions estimates are shown in Columns 1 and 3. Columns 2 and 4 show the GMM estimates where each economic neighbors' tax change is instrumented by the characteristics of all states. All regressions control for state *i*'s income per capita in 1982 US dollars, state *i*'s unemployment rate, the proportion of young (aged 5-17) and elderly (aged 65+) in state *i*'s population, and the state governor's age. All specifications include state and time fixed effects. With the exception of governor's age, all variables are differenced between period *t* and period *t-2*. Columns 1 and 3 report robust standard errors in parentheses. Columns 2 and 4 report standard errors adopting the procedure described in Caner and Zhang (2014).



### Table 3: Geographic Versus Economic Networks

	Geographic Network	Economic Network
Number of Edges	214	49
Edges in Both Networks	9	9
Edges in W-geo only	205	
Edges in W-econ only		40
Clustering	.419	.042
Reciprocated Edges	100%	12.2%
<b>Degree Distribution Across Nodes (states)</b>		
out-degree		1.021 (0.144)
in-degree	4.458 (1.597)	1.021 (1.246)

**Notes:** This compares statistics derived from the geographic network of US states to those from the estimated economic network among US states. The number of edges, edges in both networks, edges in W-geo only, edges in W-econ only, counts the number of edges in those categories. Reciprocated edges is the frequency of in-edges that are reciprocated by out-edges (by construction, this is 100% for geographic networks). The clustering coefficient is the frequency of the number of fully connected triplets over the total number of triplets. The degree distribution across nodes counts the average number of connections (standard deviation in parentheses): we show this separately for in-degree and out-degree (by construction, these are identical for geographic networks).

**Table 4: Predicting Links to Economic Neighbors**

Linear Probability Model

Dependent variable = 1 if Economic Link Between States Identified, = 0 if geographically linked

Robust standard errors in parentheses

	Distance	Economic and Demographic Homophily	Labor Mobility	Yardstick Competition	Tax Havens	Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Distance</b>	.890*** (.081)	.921*** (.082)	.921*** (.082)	.940*** (.091)	.940*** (.091)	1.287*** (.120)
<b>Distance sq.</b>	-.135*** (.025)	-.139*** (.024)	-.139*** (.025)	-.144*** (.027)	-.145*** (.027)	-.255*** (.039)
<b>GDP Homophily</b>		-.063 (.078)	-.063 (.079)	-.083 (.082)	-.092 (.085)	-.219 (.348)
<b>Demographic Homophily</b>		-1.745*** (.552)	-1.745*** (.554)	-1.047* (.605)	-.960 (.604)	.579 (1.240)
<b>Net Migration</b>			-.033 (.603)	-.020 (.577)	-.185 (.612)	-.039 (1.48)
<b>Political Homophily</b>				-.337*** (.120)	-.321*** (.119)	-.287* (.155)
<b>Tax Haven</b>					-.093** (.036)	
<b>Origin and destination FE</b>	No	No	No	No	No	Yes
<b>Adjusted R-squared</b>	.664	.664	.664	.651	.657	.831
<b>Observations</b>	254	254	254	212	212	212

**Notes:** \*\*\* denotes significance at 1%, \*\* at 5%, and \* at 10%. The specifications in all Columns are cross-sectional linear probability models where the dependent variable is equal to 1 if an economic link between states is identified, and zero if a geographic link exists between the states. A pair of states is considered a first-degree geographic neighbor if they share a border. Distance and distance squared are calculated from the centroids of states' capital cities. GDP homophily is the absolute difference of states' GDP per capita. Demographic homophily is the absolute difference of share of young (aged 5-17) plus the absolute difference of the share of elderly in states' population (aged 65+). Net migration (in millions) based on individuals tax returns (Source: Internal Revenue Service, <https://www.irs.gov/statistics/soi-tax-stats-migration-data>). Political homophily is equal to one if a pair of states have governors of same party at given year. Nevada, Delaware, Montana, South Dakota, Wyoming and New York are considered tax haven states. Time averages are taken for all explanatory variables. Robust standard errors are shown in parentheses.

**Table 5: General Equilibrium Impacts of South Carolina Tax Rises**

	<b>Geographic Neighbor Network</b>	<b>Economic Neighbor Network</b>	<b>Ratio</b>
<b>Average tax increase</b>	0.03	0.08	3.10
<b>Variance tax increase</b>	0.15	0.19	1.30
<b>Tax dispersion</b>	0.01	0.32	34.41
<b>States with tax increase &gt; 0.05%</b>	12	26	2.17
<b>States with tax increase &gt; 0.5%</b>	7	23	3.29
<b>States with tax increase &gt; 1%</b>	4	20	5.00
<b>States with tax increase &gt; 2.5%</b>	3	15	5.00
<b>States with tax increase &gt; 5%</b>	3	15	5.00

**Notes:** This shows the equilibrium impulse responses in taxes set in each state as a result of South Carolina increasing its tax change by 10%. The rho coefficient is derived from our preferred specification to estimate the economic network where we allow for exogenous social effects (based on a sample of 48 mainland US states running from 1962 to 2015). We compare these derived tax changes under the identified economic network structure, relative to that assumed under a geographic neighbors structure. The final Column shows the ratio of the same statistic derived under each network.

# Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition

## Online Appendix

Áureo de Paula

Imran Rasul

Pedro CL Souza\*

October 2023

### A Proofs

#### Example 1

To see how the assumption of Bramoullé *et al.* (2009) grants identification when  $W_0$  is known, choose constants  $c_1$ ,  $c_2$ , and  $c_3$  such that  $c_1I + c_2W_0 + c_3W_0^2 = 0$ . Focusing on the diagonal elements of this condition, we see that if the diagonal of  $W_0^2$  is not proportional to the diagonal of  $I$ , then  $c_1 = c_3 = 0$  because  $\text{diag}(W_0) = 0$ . It follows that  $c_2 = 0$  if at least one (off-diagonal) element of  $W_0$  is non-zero. However, the converse is not true, so if Assumptions A1-A6 do not hold, one can construct examples where  $\Pi_0$  does not pin down  $\theta_0$ . Take, for instance,  $N = 5$  with  $\theta_0$  and  $\theta$  where  $\beta = \beta_0 = 1$ ,  $\rho = 1.5$ ,  $\rho_0 = 0.5$ ,  $\gamma = -2.5$ , and  $\gamma_0 = 0.5$ :

$$W_0 = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix}.$$

Both  $W$  and  $W_0$  violate (A5) ( $(W^2)_{kk} = (W_0^2)_{kk} = 0.5$  for any  $k$ ), and  $\rho$  violates (A2). Nonetheless,  $I$ ,  $W_0$ , and  $W_0^2$  are linearly independent and, likewise, so are  $I$ ,  $W$ , and  $W^2$ . In this case, *both* parameter sets produce  $\Pi = (I - \rho_0W_0)^{-1}(\beta_0I + \gamma_0W_0) = (I - \rho W)^{-1}(\beta I + \gamma W)$ . This arises

---

\*de Paula: University College London, CeMMAP and IFS, a.paula@ucl.ac.uk; Rasul: University College London and IFS, i.rasul@ucl.ac.uk; Souza: Queen Mary University, p.souza@qmul.ac.uk.

even as  $W$  and  $W_0$  represent very different network structures: any pair connected under  $W$  is not connected under  $W_0$  and *vice versa*.

### Theorem 1

*Proof.* The local identification result follows Rothenberg (1971). Under the assumptions in our model, the parameter space  $\Theta \subset \mathbb{R}^m$  is an open set (recall that  $m = N(N - 1) + 3$ ). This corresponds to Assumption I in Rothenberg (1971). We have that

$$\begin{aligned}\frac{\partial \Pi}{\partial W_{ij}} &= \rho(I - \rho W)^{-1} \Delta_{ij} (I - \rho W)^{-1} (\beta I + \gamma W) + (I - \rho W)^{-1} \gamma \Delta_{ij} \\ \frac{\partial \Pi}{\partial \rho} &= (I - \rho W)^{-1} W (I - \rho W)^{-1} (\beta I + \gamma W) \\ \frac{\partial \Pi}{\partial \gamma} &= (I - \rho W)^{-1} W \\ \frac{\partial \Pi}{\partial \beta} &= (I - \rho W)^{-1},\end{aligned}$$

where  $\Delta_{ij}$  is the  $N \times N$  matrix with 1 in the  $(i, j)$ -th position and zero elsewhere. Write the  $N^2 \times m$  derivative matrix  $\nabla_{\Pi} \equiv \frac{\partial \text{vec}(\Pi)}{\partial \theta'}$ . By assumption, row  $i$  in matrix  $W$  sums up to one, incorporated through the restriction that  $\varphi \equiv \sum_{j=1, j \neq i}^N W_{ij} - 1 = 0$  for the unit-normalized row  $i$ . The derivative of the restriction  $\varphi$  is the  $m$ -dimensional vector  $\nabla'_W \equiv \frac{\partial \varphi}{\partial \theta'} = [e'_i \otimes \iota'_{N-1} \ 0_{1 \times 3}]$  (where  $e_i$  is an  $N$ -dimensional vector with 1 in the  $i$ th component and zero otherwise). Following Theorem 6 of Rothenberg (1971), the structural parameters  $\theta \in \Theta$  are locally identified if, and only if, the matrix  $\nabla \equiv [\nabla'_{\Pi} \ \nabla'_W]'$  has rank  $m$ .<sup>1</sup>

If  $\nabla$  does not have rank  $m$ , there is a nonzero vector  $\mathbf{c} \equiv (c_{W_{12}}, \dots, c_{W_{N, N-1}}, c_{\rho}, c_{\gamma}, c_{\beta})'$  such that  $\nabla \cdot \mathbf{c} = 0$ . This implies that

$$c_{W_{12}} \frac{\partial \Pi}{\partial W_{12}} + \dots + c_{W_{N, N-1}} \frac{\partial \Pi}{\partial W_{N, N-1}} + c_{\rho} \frac{\partial \Pi}{\partial \rho} + c_{\gamma} \frac{\partial \Pi}{\partial \gamma} + c_{\beta} \frac{\partial \Pi}{\partial \beta} = 0 \quad (1)$$

and, for the unit-normalized row  $i$  (see A4),

$$\sum_{j \neq i, j=1, \dots, n} c_{W_{ij}} = 0. \quad (2)$$

---

<sup>1</sup>For a parameter vector to be locally identified, Rothenberg (1971) requires that the derivative matrix  $\nabla$  have rank  $m$  at that point and that this vector be (rank-)regular. A (rank-)regular point of the parameter space is one for which there is a neighborhood where the rank of  $\nabla$  is constant (see Definition 4 in Rothenberg, 1971). Because we show that the derivative matrix has rank  $m$  at every point in the parameter space, this also guarantees that every point in the parameter space is (rank-)regular.

Pre-multiplying equation (1) by  $(I - \rho W)$  and substituting the derivatives,

$$\begin{aligned} & \sum_{i,j=1,i \neq j}^N c_{W_{ij}} [\rho \Delta_{ij} (I - \rho W)^{-1} (\beta I + \gamma W) + \gamma \Delta_{ij}] + \\ & + c_\rho W (I - \rho W)^{-1} (\beta I + \gamma W) + c_\gamma W + c_\beta I = 0. \end{aligned}$$

Define  $C \equiv \sum_{i,j=1,i \neq j}^N c_{W_{ij}} \Delta_{ij}$ . Since the spectral radius of  $\rho W$  is strictly less than one by A2, one can show (by representing  $(I - \rho W)^{-1}$  as a Neumann series, for instance) that  $(\beta I + \gamma W)$  and  $(I - \rho W)^{-1}$  commute. Then, the expression above is equivalent to

$$\rho C (\beta I + \gamma W) (I - \rho W)^{-1} + \gamma C + c_\rho W (\beta I + \gamma W) (I - \rho W)^{-1} + c_\gamma W + c_\beta I = 0.$$

Post-multiplying by  $(I - \rho W)$ , we obtain

$$\rho C (\beta I + \gamma W) + \gamma C (I - \rho W) + c_\rho W (\beta I + \gamma W) + c_\gamma W (I - \rho W) + c_\beta (I - \rho W) = 0$$

which, upon rearrangement, yields

$$(\gamma + \rho\beta) C + c_\beta I + (\beta c_\rho - c_\beta \rho + c_\gamma) W + (c_\rho \gamma - \rho c_\gamma) W^2 = 0. \quad (3)$$

Because  $C_{ii} = 0$  and  $W_{ii} = 0$  (by A1), we have that  $c_\beta + (c_\rho \gamma - \rho c_\gamma) (W^2)_{ii} = 0$  for all  $i = 1, \dots, N$ . Since by Assumption (A5) there isn't a constant  $\kappa$  such that  $\text{diag}(W_0^2) = \kappa \iota$ , then  $c_\beta = c_\rho \gamma - \rho c_\gamma = 0$ . Plugging back in (3), we obtain

$$(\gamma + \rho\beta) C + (\beta c_\rho + c_\gamma) W = 0.$$

which implies that  $C = -\frac{\beta c_\rho + c_\gamma}{\gamma + \rho\beta} W$ , since  $\gamma + \rho\beta \neq 0$  by Assumption (A3). Taking the sum of the elements in row  $i$ , we get

$$(\gamma + \rho\beta) \sum_{j \neq i, j=1, \dots, n} c_{W_{ij}} + (\beta c_\rho + c_\gamma) = 0.$$

Note that, by equation (2),  $\sum_{j \neq i, j=1, \dots, n} c_{W_{ij}} = 0$ . So,  $\beta c_\rho + c_\gamma = 0$  and  $C = -\frac{\beta c_\rho + c_\gamma}{\gamma + \rho\beta} W = 0$ . This implies that  $c_{W_{ij}} = 0$  for any  $i$  and  $j$ . Combining  $\beta c_\rho + c_\gamma = 0$  with  $c_\rho \gamma - \rho c_\gamma = 0$  obtained above, we get that  $c_\rho (\rho\beta + \gamma) = 0$ . Since  $\rho\beta + \gamma \neq 0$ , then  $c_\rho = 0$ . Given that  $\beta c_\rho + c_\gamma = 0$ , it follows that  $c_\gamma = 0$ . This shows that  $\theta \in \Theta$  is locally identified.  $\square$

### Corollary 1

*Proof.* The parameter  $\theta_0$  being locally identified (see Theorem 1) implies that the set  $\{\theta : \Pi(\theta) = \Pi(\theta_0)\}$  is discrete. If it is also compact, then the set is finite. To establish that, we now show that

$\Pi$  is a proper function: the inverse image  $\Pi^{-1}(K)$  of any compact set  $K \subset \mathbb{R}^m$  is also compact (see Krantz and Parks, 2013, p. 124).

Let  $\mathcal{A}$  be a compact set in the space of  $N \times N$  real matrices. Since it is a compact set in a finite dimensional space, it is closed and bounded. Since  $\Pi$  is a continuous function of  $\theta$ , the pre-image of a compact set, which is closed, is also closed. Because  $\|W\|$  is bounded and  $\rho \in (-1, 1)$ , their corresponding coordinates in  $\theta \in \Pi^{-1}(\mathcal{A})$  are bounded. Suppose the coordinates for  $\beta$  or  $\gamma$  in  $\theta \in \Pi^{-1}(\mathcal{A})$  are not bounded. One can find a sequence  $(\theta_k)_{k=1}^{\infty}$  such that  $|\beta_k| \rightarrow \infty$  or  $|\gamma_k| \rightarrow \infty$ .

Denote the Frobenius norm of the matrix  $A$  as  $\|A\|$ . By the submultiplicative property  $\|AB\| \leq \|A\| \cdot \|B\|$ ,

$$\|\beta I + \gamma W\| \leq \|I - \rho W\| \cdot \|(I - \rho W)^{-1}(\beta I + \gamma W)\| = \|I - \rho W\| \cdot \|\Pi\|.$$

It follows that

$$\frac{\|\beta I + \gamma W\|}{\|I - \rho W\|} \leq \|\Pi\|.$$

Given  $W$  has zero main diagonal,

$$\|\beta I + \gamma W\|^2 = \beta^2 \|I\|^2 + \gamma^2 \|W\|^2 = \beta^2 N + \gamma^2 \|W\|^2.$$

Also,  $\|I - \rho W\|^2 = N + \rho^2 \|W\|^2 \leq N + \rho^2 C$ , for some constant  $C \in \mathbb{R}$  by Assumption (A2). We then have that

$$\frac{\sqrt{\beta^2 N + \gamma^2 \|W\|^2}}{\sqrt{N + \rho^2 C}} \leq \|\Pi\|.$$

By Assumption (A2), the denominator above is bounded. Hence,  $|\beta_k| \rightarrow \infty \Rightarrow \|\Pi(\theta_k)\| \rightarrow \infty$ . We now use the fact that  $\sum_j W_{ij} = 1$  to show that there is a lower bound on  $\|W\|^2$ , and so  $|\gamma_k| \rightarrow \infty \Rightarrow \|\Pi(\theta_k)\| \rightarrow \infty$ . To see this, note that

$$\min_{\text{s.t. } \sum_j W_{ij}=1} \|W\|^2 \geq \min_{\text{s.t. } \sum_j W_{ij}=1} \sum_{j=1}^N W_{ij}^2.$$

Since the objective function in above is convex, it can be shown that it is minimized at  $W_{ij} = \frac{1}{N-1}, j \neq i$  and, consequently,  $\|W\|^2 \geq (N-1) \frac{1}{(N-1)^2} = \frac{1}{N-1}$ . Hence, if  $|\gamma_k| \rightarrow \infty$ , the numerator in the lower bound for  $\|\Pi\|$  above also goes to infinity. Consequently,  $\mathcal{A}$  is not compact.

Therefore, if  $\mathcal{A}$  is compact, the coordinates in  $\theta \in \Pi^{-1}(\mathcal{A})$  corresponding to  $\beta$  and  $\gamma$  are also bounded. Hence,  $\Pi^{-1}(\mathcal{A})$  is bounded (and closed). Consequently, it is compact.

For a given reduced-form parameter matrix  $\Pi$ , the set  $\{\theta : \Pi(\theta) = \Pi(\theta_0)\}$  is then compact. Since it is also discrete, it is finite.  $\square$

The following lemmas are used in proving Theorem 2.

**Lemma 1.** *Assume (A1)-(A6). If  $\gamma_0 = 0$ ,  $W_0$  is such that  $(W_0)_{1,2} = (W_0)_{2,1} = 1$  and  $(W_0)_{ij} = 0$  otherwise, with  $\rho_0 \neq 0$  and  $\beta_0 \neq 0$ , then  $\theta_0 \in \Theta$  is identified.*

*Proof.* Take  $\theta = (W_{12}, \dots, W_{N,N-1}, \rho, \gamma, \beta) \in \Theta$  possibly different from  $\theta_0$  such that the models are observationally equivalent, so  $\Pi_0 = \Pi$ . Then,

$$(I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0) = (I - \rho W)^{-1}(\beta I + \gamma W).$$

Since  $\gamma_0 = 0$  and  $(I - \rho W)^{-1}$  and  $(\beta I + \gamma W)$  commute (see the proof for Theorem 1), it follows that

$$\Pi_0 = \Pi \Leftrightarrow \beta_0(I - \rho_0 W_0)^{-1} = (\beta I + \gamma W)(I - \rho W)^{-1}$$

or, equivalently,

$$\beta_0(I - \rho W) = (I - \rho_0 W_0)(\beta I + \gamma W).$$

This last equation implies that

$$(\beta_0 - \beta)I - (\gamma + \beta_0 \rho)W + \rho_0 \beta W_0 + \rho_0 \gamma W_0 W = 0. \quad (4)$$

We first note that  $(W_0 W)_{N,N} = 0$  since  $(W_0)_{N,i} = 0$  for any  $1 \leq i \leq N$  and, by Assumption (A1),  $(W)_{N,N} = (W_0)_{N,N} = 0$ . So  $\beta_0 = \beta$ . Taking elements  $(i, j)$  such that  $i \geq 3$  and  $i \neq j$  in equation (4), and using the fact that  $\beta_0 = \beta$ , we find that  $-(\gamma + \beta_0 \rho)(W)_{ij} = -(\gamma + \beta \rho)(W)_{ij} = 0$  for any  $(i, j)$  such that  $i \geq 3$  and  $i \neq j$ . By Assumption (A3),  $\gamma + \beta \rho \neq 0$  and it follows that  $(W)_{ij} = 0$  for any  $(i, j)$  such that  $i \geq 3$  and  $i \neq j$ . In fact, since  $(W)_{i,i} = 0$  by Assumption (A1), we get that  $(W)_{ij} = 0$  for any  $(i, j)$  such that  $i \geq 3$ .

Using Assumption (A1) and since  $\beta_0 = \beta$ , elements (1, 1) and (2, 2) in equation (4) imply that  $\rho_0 \gamma (W)_{2,1} = \rho_0 \gamma (W)_{1,2} = 0$ . Given that  $\rho_0 \neq 0$ , we get that  $\gamma (W)_{2,1} = \gamma (W)_{1,2} = 0$ . From element (1, 2) in equation (4), we find that  $-(\gamma + \beta_0 \rho)(W)_{1,2} + \rho_0 \beta = 0$  or, equivalently,  $(\rho_0 - \rho (W)_{1,2})\beta_0 - \gamma (W)_{1,2} = 0$ . Given that  $\gamma (W)_{1,2} = 0$  and  $\beta_0 \neq 0$ , it must be that  $\rho_0 - \rho (W)_{1,2} = 0$ . Making the analogous argument for element (2, 1), we would also obtain that  $\rho_0 - \rho (W)_{2,1} = 0$ .

If both  $(W)_{1,2}$  and  $(W)_{2,1}$  are equal to zero, using the fact that  $W_{ij} = 0$  for any  $(i, j)$  such that  $i \geq 3$ , we would then obtain that  $W^2$  is equal to a matrix of zeros, which contradicts Assumption (A5). Hence,  $(W)_{1,2} \neq 0$  or  $(W)_{2,1} \neq 0$ . If  $(W)_{1,2} \neq 0$ , using the fact that  $\gamma (W)_{1,2} = 0$ , we get that  $\gamma = 0$ . Equivalently, if  $(W)_{2,1} \neq 0$ , and using the fact that  $\gamma (W)_{2,1} = 0$ , we again get that  $\gamma = 0$ . So, in either case,  $\gamma = \gamma_0 = 0$ .

Taking element (1,  $j$ ) in equation (4), with  $j \geq 3$ , we get that  $-(\gamma + \rho \beta_0)W_{1,j} + \gamma \rho_0 (W)_{2,j} = -\rho \beta_0 W_{1,j} = 0$ . Similarly, element (2,  $j$ ), with  $j \geq 3$  implies that  $-(\gamma + \rho \beta_0)W_{2,j} + \gamma \rho_0 (W)_{1,j} =$



$-\rho\beta_0W_{2,j} = 0$ . Then, from  $-\rho\beta_0(W)_{1,j} = -\rho\beta_0(W)_{2,j} = 0$  for  $j \geq 3$ , it follows that  $-\rho(W)_{1,j} = -\rho(W)_{2,j} = 0$  since  $\beta_0 \neq 0$ .

From  $\rho_0 - \rho(W)_{1,2} = 0$ , if  $(W)_{1,2} \neq 0$ , we get that  $\rho = \rho_0/(W)_{1,2} \neq 0$ . Equivalently, if  $(W)_{2,1} \neq 0$ , we get that  $\rho = \rho_0/(W)_{2,1} \neq 0$ . Since  $(W)_{1,2} \neq 0$  or  $(W)_{2,1} \neq 0$ , we obtain that  $\rho \neq 0$ . Then, because  $-\rho(W)_{1,j} = \rho(W)_{2,j} = 0$  for  $j \geq 3$ , we have that  $(W)_{1j} = (W)_{2j} = 0$  for  $j \geq 3$ .

Given that  $\rho_0 - \rho(W)_{1,2} = 0$ ,  $\rho_0 - \rho(W)_{2,1} = 0$  and  $\rho \neq 0$ , we obtain that  $(W)_{1,2} = (W)_{2,1} = \frac{\rho_0}{\rho}$ . Since  $(W)_{1,j} = 0$  for  $j \neq 2$ ,  $(W)_{2,j} = 0$  for  $j \neq 1$  and  $(W)_{ij} = 0$  for  $i \geq 3$ , by Assumption (A5) we get that  $(W)_{1,2} = (W)_{2,1} = 1$  and  $\rho = \rho_0$ . Hence,  $((W)_{1,2}, \dots, (W)_{N,N-1}, \rho, \gamma, \beta) = ((W_0)_{1,2}, \dots, (W_0)_{N,N-1}, \rho_0, \gamma_0, \beta_0)$ .  $\square$

**Lemma 2.** *Assume (A1)-(A2) and (A4)-(A6). The image of  $\Pi(\cdot)$ , for  $\theta \in \Theta_+$ , is path-connected and, therefore, connected.*

*Proof.* Take  $\theta$  and  $\theta^* \in \Theta_+$ . Consider first the subvectors corresponding to the adjacency matrices  $W$  and  $W^*$ . Without loss of generality, let  $1, \dots, N$  be ordered such that  $(W^2)_{11} > (W^2)_{22}$ . Consider the adjacency matrix  $W_*$  corresponding to the network of unweighted directed connections  $\{(1, 2), (2, 1)\}$  and  $\{(3, 4), (4, 5), \dots, (N-1, N), (N, 3)\}$ . Note that  $\text{diag}(W_*^2) = (1, 1, 0, \dots, 0)$  and this is an admissible adjacency matrix under assumptions (A1)-(A2) and (A4)-(A6). We first show that  $W$  is path-connected to  $W_*$ .

Consider the path given by

$$W(t) = tW_* + (1-t)W$$

which implies that

$$\begin{aligned} (W(t)^2)_{11} &= (1-t)^2(W^2)_{11} + t^2 + (1-t)t(W_{12} + W_{21}) \\ (W(t)^2)_{22} &= (1-t)^2(W^2)_{22} + t^2 + (1-t)t(W_{12} + W_{21}). \end{aligned}$$

Since  $(W(t)^2)_{11} - (W(t)^2)_{22} = (1-t)^2[(W^2)_{11} - (W^2)_{22}] > 0$  for  $t \in [0, 1)$  and  $W(1) = W_*$ , (A5) is satisfied for any matrix  $W(t)$  such that  $t \in [0, 1]$ . Since all rows in  $W_*$  sum to one and  $(W_*)_{ii} = 0$  for any  $i$ , it is straightforward to see that  $W(t)$  also satisfies (A1) and (A4). Finally,  $\sum_{j=1}^N |W_{ij}(t)| \leq t \sum_{j=1}^N |(W_*)_{ij}| + (1-t) \sum_{j=1}^N |W_{ij}| \leq 1$  for every  $i = 1, \dots, N$ , and  $W(t)$  satisfies Assumption (A2).

If  $W^*$  is such that  $(W^{*2})_{11} \neq (W^{*2})_{22}$ , the convex combination of  $W^*$  and  $W_*$  is also seen to satisfy (A1)-(A2) and (A4)-(A6), and a path between  $W$  and  $W^*$  can be constructed via  $W_*$ . If, on the other hand  $(W^{*2})_{11} = (W^{*2})_{22}$ , suppose without loss of generality that  $(W^{*2})_{11} \neq (W^{*2})_{33}$ . In this case, one can construct a path between  $W^*$  and  $W_{**}$  where  $W_{**}$  represents the network of unweighted directed connections  $\{(1, 3), (3, 1)\}$  and  $\{(2, 4), (4, 5), \dots, (N-1, N), (N, 2)\}$ .

Like  $W(t)$  above, this path can be seen to satisfy assumptions (A1)-(A2) and (A4)-(A6). Now note that a path can also be constructed between  $W_*$  and  $W_{**}$ , as their convex combination also satisfies (A1)-(A2) and (A4)-(A6). For example, note that  $\hat{W}(t) = tW_* + (1-t)W_{**}$  is such that  $(\hat{W}(t)^2)_{11} = t^2 + (1-t)^2$  and  $(\hat{W}(t)^2)_{NN} = 0$ , so  $(\hat{W}(t)^2)_{11} - (\hat{W}(t)^2)_{NN} > 0$  for any  $t \in (0, 1)$  and both  $\hat{W}(0)$  and  $\hat{W}(1)$  satisfy (A5). Hence, we can construct a path  $W(t)$  between  $W$  and  $W^*$  through  $W_*$  and  $W_{**}$ .

Furthermore,  $\rho(t) = t\rho^* + (1-t)\rho$ ,  $\beta(t) = (t\rho^*\beta^* + (1-t)\rho\beta)/(t\rho^* + (1-t)\rho)$ ,  $\gamma(t) = t\gamma^* + (1-t)\gamma$  are such that

$$f(t) \equiv \rho(t)\beta(t) + \gamma(t) = t(\rho^*\beta^* + \gamma^*) + (1-t)(\rho\beta + \gamma) > 0,$$

since  $\theta^*$  and  $\theta \in \Theta_+$ . (Note also that  $|\rho(t)| < 1$ , so Assumption (A2) is satisfied.) These facts taken together imply that

$$\theta(t) \equiv (W(t)_{12}, \dots, W(t)_{N,N-1}, \rho(t), \gamma(t), \beta(t)) \in \Theta_+.$$

That is,  $\Theta_+$  is path-connected and therefore connected. Since  $\Pi(\cdot)$  is continuous on  $\Theta_+$ ,  $\Pi(\Theta_+)$  is connected.  $\square$

## Theorem 2

*Proof.* The proof uses Corollary 1.4 in Ambrosetti and Prodi (1995, p. 46),<sup>2</sup> which we reproduce here with our notation for convenience: *Suppose the function  $\Pi(\cdot)$  is continuous, proper, and locally invertible with a connected image. Then, the cardinality of  $\Pi^{-1}(\bar{\Pi})$  is constant for any  $\bar{\Pi}$  in the image of  $\Pi(\cdot)$ .*

The mapping  $\Pi(\theta)$  is continuous and proper (by Corollary 1), with a connected image (Lemma 2), and non-singular Jacobian at any point (as per the proof for Theorem 1), which guarantees local invertibility. Following Corollary 1.4 in Ambrosetti and Prodi (1995, p. 46) reproduced above, we obtain that the cardinality of the pre-image of  $\Pi(\theta)$  is finite and constant. Take  $\theta \in \Theta_+$  such that  $\gamma = 0$ ,  $(W)_{1,2} = (W)_{2,1} = 1$  and  $(W)_{i,j} = 0$  otherwise, with  $\rho \neq 0$  and  $\beta \neq 0$ . By Lemma 1, that cardinality is one.  $\square$

## Corollary 3

*Proof.* Since  $\rho \in (0, 1)$  and  $W_{ij} \geq 0$ ,  $\sum_{k=1}^{\infty} \rho^{k-1} W^k$  is a non-negative matrix. By (5), the off-diagonal elements of  $\Pi(\theta)$  are equal to the off-diagonal elements of  $(\rho\beta + \gamma) \sum_{k=1}^{\infty} \rho^{k-1} W^k$ , and the sign of those elements identifies the sign of  $\rho\beta + \gamma$ . By Theorem 2, the model is identified.

<sup>2</sup>Related results can be found in Ambrosetti and Prodi (1972) and de Marco *et al.* (2014).

□

### Corollary 4

*Proof.* Since  $W_0$  is non-negative and irreducible, there is a real eigenvalue equal to the spectral radius of  $W_0$  corresponding to the unique eigenvector whose entries can be chosen to be strictly positive (i.e., all the entries share the same sign). A generic eigenvalue of  $W_0$ ,  $\lambda_0$ , corresponds to an eigenvalue of  $\Pi_0$  according to:

$$\lambda_{\Pi_0} = \beta_0 + (\rho_0\beta_0 + \gamma_0) \frac{\lambda_0}{1 - \rho_0\lambda_0}$$

If  $\lambda_0 = a_0 + b_0i$  where  $a_0, b_0 \in \mathbb{R}$  and  $i = \sqrt{-1}$ , then

$$\lambda_{\Pi_0} = \beta_0 + (\rho_0\beta_0 + \gamma_0) \frac{a_0(1 - \rho_0a_0) - \rho_0b_0^2}{(1 - \rho_0a_0)^2 + \rho_0^2b_0^2} + (\rho_0\beta_0 + \gamma_0) \frac{b_0}{(1 - \rho_0a_0)^2 + \rho_0^2b_0^2}i.$$

If the eigenvalue  $\lambda_0$  is real,  $b_0 = 0$  and the corresponding eigenvalue  $\lambda_{\Pi_0}$  is also real. Differentiating  $Re(\lambda_{\Pi_0})$ , the real part of  $\lambda_{\Pi_0}$ , with respect to  $Re(\lambda_0) = a_0$ , we get

$$\frac{\partial Re(\lambda_{\Pi_0})}{\partial a_0} = \frac{(1 - \rho_0a_0)^2 - \rho_0^2b_0^2}{[(1 - \rho_0a_0)^2 + \rho_0^2b_0^2]^2} \times (\rho_0\beta_0 + \gamma_0). \quad (5)$$

If the eigenvalue  $\lambda_0$  is real, expression (5) becomes:

$$\frac{\partial Re(\lambda_{\Pi_0})}{\partial a_0} = \frac{\partial \lambda_{\Pi_0}}{\partial a_0} = \frac{1}{(1 - \rho_0a_0)^2} \times (\rho_0\beta_0 + \gamma_0).$$

The fraction multiplying  $\rho_0\beta_0 + \gamma_0$  is positive. If  $\rho_0\beta_0 + \gamma_0 < 0$ , the real eigenvalues of  $\Pi_0$  are decreasing on the real eigenvalues of  $W_0$ . Consequently, the eigenvector corresponding to the largest (real) eigenvalue of  $W_0$  will be associated with the smallest real eigenvalue of  $\Pi_0$ . If, on the other hand,  $\rho_0\beta_0 + \gamma_0 > 0$ , the eigenvector corresponding to the largest real eigenvalue of  $W_0$  will correspond to the largest real eigenvalue of  $\Pi_0$ . Since that eigenvector is the unique eigenvector that can be chosen to have strictly positive entries, the sign of  $\rho_0\beta_0 + \gamma_0$  is identified by the  $\lambda_{\Pi_0}$  eigenvalue it is associated with and whether it is the largest or smallest real eigenvalue. By Theorem 2, the model is identified.

If there is only one real eigenvalue, note that the denominator in the fraction in (5) is positive. The minimum value of the numerator subject to  $|\lambda_0|^2 = a_0^2 + b_0^2 \leq 1$  is given by

$$\min_{a_0, b_0} (1 - \rho_0a_0)^2 - \rho_0^2b_0^2 \text{ s.t. } a_0^2 + b_0^2 \leq 1.$$

The Lagrangean for this minimization problem is given by

$$\mathcal{L}(a_0, b_0; \mu) = (1 - \rho_0 a_0)^2 - \rho^2 b_0^2 + \mu(a_0^2 + b_0^2 - 1)$$

where  $\mu$  is the Lagrange multiplier associated with the constraint  $a_0^2 + b_0^2 \leq 1$ . The Kuhn-Tucker necessary conditions for the solution  $(a_0^*, b_0^*, \mu^*)$  of this problem are given by

$$\begin{aligned} (\partial a_0 : ) \quad & \rho_0(1 - \rho_0 a_0^*) - \mu^* a_0^* = 0 \\ (\partial b_0 : ) \quad & (\rho_0^2 - \mu^*) b_0^* = 0 \\ & \mu^* (a_0^{*2} + b_0^{*2} - 1) = 0 \\ & a_0^{*2} + b_0^{*2} \leq 1 \text{ and } \mu^* \geq 0. \end{aligned}$$

Let  $\rho_0 \neq 0$ . (Otherwise, the objective function above is equal to one irrespective of  $a_0$  or  $b_0$ , and the partial derivative is  $\rho_0 \beta_0 + \gamma_0$ ). If  $\mu^* = 0$ ,  $\partial b_0$  implies that  $b_0^* = 0$ . Then,  $\partial a_0$  will have  $a_0^* = \rho_0^{-1}$ , which violates  $a_0^{*2} + b_0^{*2} \leq 1$ .

Hence, a solution should have  $\mu^* > 0$ . In this case, there are two possibilities:  $b_0^* = 0$  or  $b_0^* \neq 0$ . If  $b_0^* \neq 0$ , condition  $\partial b_0$  implies that  $\mu^* = \rho_0^2$ , and  $\partial a_0$  then gives  $a_0^* = (2\rho_0)^{-1}$ . Because the constraint is binding,  $b_0^{*2} = 1 - (4\rho_0^2)^{-1}$ . In this case,  $a_0^{*2} \leq 1$  and  $b_0^{*2} \geq 0$  requires that  $|\rho_0| \geq 1/2$ . The value of the minimised objective function in this case is  $1/2 - \rho_0^2$ . This is positive if  $|\rho_0| < \sqrt{2}/2$ .

The other possibility is to have  $b_0 = 0$ . Because the constraint is binding,  $a_0 = 1$  and the objective function takes the value  $(1 - \rho_0)^2 > 0$ . Since  $(1 - \rho_0)^2 - 1/2 + \rho_0^2 = 2\rho_0^2 - 2\rho_0 + 1/2 \geq 0$ , this solution is dominated by the previous one when  $|\rho_0| \geq 1/2$ .

Consequently, the fraction multiplying  $\rho_0 \beta_0 + \gamma_0$  is non-negative, and it can be ascertained that

$$\text{sgn} \left[ \frac{\partial \text{Re}(\lambda_{\Pi_0})}{\partial a_0} \right] = \text{sgn}[\rho_0 \beta_0 + \gamma_0]$$

as long as  $|\rho_0| < \sqrt{2}/2$ .

If  $\rho_0 \beta_0 + \gamma_0 < 0$ , the real part of the eigenvalues of  $\Pi_0$  is decreasing on the real part of the eigenvalues of  $W_0$ . Since the dominant eigenvalue for  $W_0$  will be real and thus the one with the largest real part, we only need to focus on the real part of the eigenvalues. Consequently, the eigenvector corresponding to the eigenvalue of  $W_0$  with the largest real part will correspond to the eigenvalue of  $\Pi_0$  with the smallest real part. If, on the other hand,  $\rho_0 \beta_0 + \gamma_0 > 0$ , the eigenvector corresponding to the eigenvalue of  $W_0$  with the largest real part will correspond to the eigenvalue of  $\Pi_0$  with the largest real part. Since that eigenvector is the unique eigenvector that can be chosen to have strictly positive entries, the sign of  $\rho_0 \beta_0 + \gamma_0$  is identified by the  $\lambda_{\Pi_0}$  eigenvalue it is associated with.

By Theorem 2, the model is identified. □

## Proposition 1

*Proof.* From equation (6), we observed that  $\Pi_0 v_j = \lambda_{\Pi_0, j} v_j$ , where  $v_j$  is an eigenvector of both  $W_0$  and  $\Pi_0$  with the corresponding eigenvalue  $\lambda_{\Pi_0, j} = \frac{\beta_0 + \gamma_0 \lambda_{0, j}}{1 - \rho_0 \lambda_{0, j}}$ . Defining  $c$  as the row-sum of  $\Pi_0$ , we also have that

$$\begin{aligned} \tilde{\Pi}_0(I - H)v_j &= (I - H)\Pi_0(I - H)v_j = (I - H)\Pi_0 v_j - (I - H)\Pi_0 H v_j \\ &= \lambda_{\Pi_0, j}(I - H)v_j - (I - H)cH v_j = \lambda_{\Pi_0, j}(I - H)v_j - (H - H^2)cv_j \\ &= \lambda_{\Pi_0, j}(I - H)v_j - (H - H)cv_j = \lambda_{\Pi_0, j}(I - H)v_j, \end{aligned}$$

where the third equality is obtained from  $\Pi_0 H = cH$ , where  $c = \frac{\beta_0 + \gamma_0}{1 - \rho_0}$ , and the fifth equality holds since  $H$  is idempotent. So,  $\tilde{\Pi}_0$  and  $\Pi_0$  have common eigenvalues, with the corresponding eigenvector  $\tilde{v}_j = v_j - \bar{v}_j \iota$  for  $\tilde{\Pi}_0$ , where  $\bar{v}_j = \frac{1}{N} \iota' v_j$ ,  $j = 1, \dots, N$ .

If  $\Pi_0$  is diagonalizable, since  $\lambda_{\Pi_0, j}$  and  $\tilde{v}_j$  are observed from  $\tilde{\Pi}_0$ , identification of  $\Pi_0$  is equivalent to identification of  $\bar{v}_j$ . In order to handle non-diagonalizable matrices, we relate the Jordan canonical forms for  $\Pi_0$  and  $\tilde{\Pi}_0$ . (If  $\Pi_0$  is diagonalizable, the Jordan canonical form will coincide with its eigendecomposition.) In this case,  $\Pi_0 = MJM^{-1}$ , where  $J$  is the Jordan form and  $M$  is a matrix comprised of ordinary and generalized eigenvectors. The generalized eigenvectors can be obtained recursively from ordinary eigenvectors.<sup>3</sup> We have established above that if  $v_j$  is an ordinary eigenvector for  $\Pi_0$ ,  $\tilde{v}_j = v_j - \bar{v}_j \iota$  is an ordinary eigenvector for  $\tilde{\Pi}_0$ , corresponding to the same eigenvalue. If this eigenvalue has multiplicity larger than one, let  $x_j$  be the initial generalized eigenvector for  $\Pi_0$ , obtained as  $\Pi_0 x_j = \lambda_{\Pi_0, j} x_j + v_j$ . Then, notice that

$$\begin{aligned} \tilde{\Pi}_0(I - H)x_j &= (I - H)\Pi_0(I - H)x_j = (I - H)\Pi_0 x_j - (I - H)\Pi_0 H x_j \\ &= \lambda_{\Pi_0, j}(I - H)x_j + (I - H)v_j - (I - H)cH x_j = \lambda_{\Pi_0, j}(I - H)x_j + (I - H)v_j, \end{aligned}$$

where we reproduce similar steps as above. This implies that  $\tilde{x}_j = x_j - \bar{x}_j \iota$  is a generalized eigenvector for  $\tilde{\Pi}_0$  obtained from the ordinary eigenvector  $\tilde{v}_j$ . If additional generalized eigenvectors are needed for the Jordan block, they can be obtained recursively from  $x_j$  and  $\tilde{x}_j$ . Notice also that both  $\Pi_0$  and  $\tilde{\Pi}_0$  will have the same Jordan form  $J$  since their eigenvalues are the same. Hence, to establish identification, it suffices to establish identification of  $\bar{v}_j$  and  $\bar{x}_j$ .

To establish identification of  $\bar{v}_j$ , note that  $\Pi_0(\tilde{v}_j + \bar{v}_j \iota) = \lambda_{\Pi_0, j}(\tilde{v}_j + \bar{v}_j \iota)$  since  $v_j$  is an eigenvector of  $\Pi_0$ . Consider an alternative constant  $\bar{v}_j^* \neq \bar{v}_j$  that satisfies the previous equation. Then,

$$\Pi_0 \iota(\bar{v}_j - \bar{v}_j^*) = \lambda_{\Pi_0, j} \iota(\bar{v}_j - \bar{v}_j^*).$$

---

<sup>3</sup>“The search for the Jordan form of  $A$  becomes a search for these strings of vectors, each one headed by an eigenvector. For each  $i$ , either  $Ax_i = \lambda_i x_i$  or  $Ax_i = \lambda_i x_i + x_{i-1}$ . The vectors  $x_i$  go into the columns of  $M$ , and each string produces a single block in  $J$ .” (Strang, 2006)

Since  $\Pi_0 \iota = c$ ,  $v_j$  must satisfy  $(c - \lambda_{\Pi_0, j})(\bar{v}_j - \bar{v}_j^*) = 0$ . For  $j = 2, \dots, N$ ,  $|\lambda_{0, j}| < 1$  which implies that  $c \neq \lambda_{\Pi_0, j}$ . So,  $\bar{v}_j = \bar{v}_j^*$  and therefore is identified. For  $j = 1$ ,  $\lambda_{\Pi_0, 1} = c$  with eigenvector  $v_1 = \iota$  if  $W_0$  is non-negative and irreducible, since it corresponds to  $\lambda_{0, 1} = 1$ , which is a simple eigenvalue with both algebraic and geometric multiplicity equal to one by the Perron-Frobenius theorem.

Now consider  $x_j$  and assume, without loss, that it is a generalized eigenvector obtained from  $v_j$  as above. In this case,  $\Pi_0(\tilde{x}_j + \bar{x}_j \iota) = \lambda_{\Pi_0, j}(\tilde{x}_j + \bar{x}_j \iota) + (\tilde{v}_j + \bar{v}_j \iota)$ . Since we have established the identification  $\bar{v}_j$  above, consider an alternative constant  $\bar{x}_j^* \neq \bar{x}_j$  that satisfies the previous equation. Then, similar arguments as those used above allow one to obtain identification for  $\bar{x}_j$  and analogously for possible successive generalized eigenvectors. This then allows us to reconstruct the Jordan decomposition for  $\Pi_0$  from the Jordan decomposition of  $\tilde{\Pi}_0$ .  $\square$

## Proposition 2

*Proof.* Under row-sum normalization and  $|\rho_0| < 1$ ,  $(I - \rho_0 W_0)^{-1} \iota = \iota + \rho_0 W_0 \iota + \rho_0^2 W_0^2 \iota + \dots = \iota + \rho_0 \iota + \rho_0^2 \iota + \dots = \iota \frac{1}{1 - \rho_0}$ , so  $\Pi_{01} \equiv (I - \rho_0 W)^{-1}$  has constant row-sums. If row-sum normalization fails,  $\Pi_{01}$  may not have constant row-sums. Define  $h_{ij}$  as the  $(ij)$ -th element of  $\tilde{H}$ . The first row of the system  $(I - \tilde{H})(I - \rho_0 W)^{-1} \iota = (I - \tilde{H})r_{W_0} = 0$  is  $h_{11}^* r_{W_0, 1} - h_{12} r_{W_0, 2} - \dots - h_{1N} r_{W_0, N} = 0$  where  $h_{11}^* = 1 - h_{11}$  and  $r_{W_0, l}$  is the  $l$ -th element of  $r_{W_0}$ . If there are  $N$  possible  $W_0, W_0^{(1)}, \dots, W_0^{(n)}$ , such that  $[r_{W_0^{(1)}} \dots r_{W_0^{(N)}}]$  has rank  $N$ , then  $h_{11}^* = h_{12} = \dots = h_{1N} = 0$ . Since the same reasoning applies to all rows,  $\tilde{H}$  is the trivial transformation  $\tilde{H} = I$ .  $\square$

## B Extensions

### Extension: Multivariate Covariates

Allowing for multivariate  $x_t$  of dimension  $n \times k$ , the reduced-form model (4) is,

$$y_t = \sum_{k=1}^K \Pi_{0, k} x_{k, t} + \nu_t,$$

where  $\Pi_{0, k} = (I - \rho_0 W_0)^{-1} (\beta_{0, k} I + \gamma_{0, k} W_0)$ ,  $x_{k, t}$  refers to the  $k$ -th column of  $x_t$ , and  $\beta_{0, k}$  and  $\gamma_{0, k}$  select the  $k$ -th element of  $K$ -dimensional  $\beta_0$  and  $\gamma_0$ , respectively. The previous identification results then apply element by element to each  $\Pi_{0, k}$ ,  $k = 1, \dots, K$ . In fact, we only then need to maintain  $W_x = \gamma_0 W_0$  and  $\gamma_0 \neq 0$  for one covariate. It is therefore possible to allow the structure of endogenous and exogenous social effects to differ for  $K - 1$  of the covariates. With  $K$  covariates, equation (3) is,

$$y_t = \rho_0 W_0 y_t + \sum_{k=1}^K \beta_{0, k} x_{k, t} + \sum_{k=1}^K \gamma_{0, k} W_{0, k} x_{k, t} + \epsilon_t.$$

Let  $W_{0,k} = W_0$  be the case for  $k = 1$ . Then, having identified  $\rho_0$  and  $W_0$  from  $\Pi_{0,1}$ ,

$$(I - \rho_0 W_0)\Pi_{0,k} = \beta_{0,k}I + \gamma_{0,k}W_{0,k},$$

for  $k = 2, \dots, K$ . The parameter  $\beta_{0,k}$  then corresponds to the diagonal elements of  $(I - \rho_0 W_0)\Pi_{0,k}$  and the off-diagonal entries correspond to the off-diagonal elements of  $\gamma_{0,k}W_{0,k}$ . If Assumption (A4) holds for every  $k = 1, \dots, K$ , we can identify  $\gamma_{0,k}$  and thus  $W_{0,k}$  for every  $k = 1, \dots, K$ .<sup>4</sup>

### Extension: Heterogeneous $\beta_0$

While many applications assume  $\beta_0$  to be homogeneous across individual units, we here consider possible avenues allowing for heterogeneous coefficients. In a slight abuse of notation, consider for this subsection  $\beta_0$  in equation system (3) to be  $\text{diag}(\beta_{01}, \dots, \beta_{0N})_{N \times N}$ . Instead of a homogeneous scalar,  $\beta_0$  is a diagonal matrix with the individual-specific coefficients  $\beta_{01}, \dots, \beta_{0N}$  along its diagonal. When  $\rho_0 = 0$  as in Manresa (2016),  $\Pi_0 = \beta_0 + \gamma_0 W_0$ . In this case, under Assumption (A1),  $\beta_0$  is identified from the diagonal elements in  $\Pi_0$  and  $\gamma_0 W_0$  is identified from its off-diagonal elements.

With multiple covariates, as long as the coefficients are homogeneous for one of the covariates, one can also identify heterogeneous coefficients on the remaining covariates, as done in the previous subsection. For example, let there be  $K$  covariates and  $\beta_{0,k} = \text{diag}(\beta_{01,k}, \dots, \beta_{0N,k})$  for  $k = 1, \dots, N$ . Suppose  $\beta_{01,k} = \dots = \beta_{0N,k}$  for one of these covariates, and let  $k = 1$  without loss of generality. Having identified  $\rho_0$  and  $W_0$  from  $\Pi_{0,1}$ ,

$$(I - \rho_0 W_0)\Pi_{0,k} = \beta_{0,k} + \gamma_{0,k}W_{0,k},$$

for  $k = 2, \dots, K$ . Then, under Assumption (A1),  $\beta_{0,k}$  is identified from the diagonal elements in  $(I - \rho_0 W_0)\Pi_{0,k}$  and  $\gamma_{0,k}W_{0,k}$  is identified from its off-diagonal elements for  $k = 2, \dots, K - 1$ .

Alternatively, when  $\gamma_0 = 0$ , one can apply traditional simultaneous equation methods to attain identification. For example, let  $B \equiv [(I - \rho_0 W_0)', -(\beta_0 + \gamma_0 W)']'_{2N \times N}$  and  $R_{(N-1) \times 2N} = [0_{(N-1) \times (N+1)} \ I_{N-1}]$ . The restriction that  $\gamma_0 = 0$  in the first equation in equation system (3) can then be expressed as  $RB_{\cdot,1} = 0_{(N-1) \times 1}$ , where  $B_{\cdot,1}$  is the first column in  $B$ . The rank condition for

---

<sup>4</sup>If  $x$  is a scalar,  $\Pi = (I - W_y)^{-1}(\beta I + W_x)$ , where we absorb  $\rho$  and  $\gamma$  into the neighborhood matrices for simplicity. Supposing both  $W_y$  and  $W_x$  have zero diagonals (Assumption (A1)) one has  $2N(N - 1) + 1$  structural parameters against  $N^2$  elements in  $\Pi$ . Since  $2N(N - 1) + 1 > N^2$  for  $N > 1$  one would not be able to identify the parameters of interest without further information. Blume *et al.* (2015) also study the case in which the social structure mediating endogenous and exogenous social effects might differ. When  $W_x$  is known and there is partial knowledge of the endogenous social interaction matrix  $W_0$ , they show that the parameters of the model can be identified (their Theorem 6). Analogously, when there are enough unconnected nodes in each of the social interaction matrices represented by  $W_x$  and  $W_0$ , and the identity of those nodes is known, identification is also (generically) possible (their Theorem 7).

the identification of the first equation is then given by,

$$RB = \begin{bmatrix} 0 & -\beta_{0,2} & \dots & 0 \\ & \dots & & \\ 0 & 0 & \dots & -\beta_{0,N} \end{bmatrix}$$

having rank equal to  $N - 1$  (see Theorem 9.2 in Wooldridge, 2002). This will be the case if  $\beta_{0,2}, \dots, \beta_{0,N} \neq 0$ . Intuitively, this guarantees that individual specific covariates are valid instrumental variables for their outcomes. Hence, if  $\beta_{0,1}, \dots, \beta_{0,N}$  are each different from zero, identification from  $\Pi_0$  is obtained.

More generally, because there are  $N^2$  equations corresponding to the entries in  $\Pi_0$  and, allowing for heterogeneity in  $\beta_0$  and imposing assumptions (A1)-(A6), there are  $N^2 + 1$  parameters, further restrictions (like row-sum normalization) would be necessary. We conjecture that adequate restrictions would deliver positive identification results, but we focus on the more conventional setting with homogeneous  $\beta_0$ .

## C Estimation

### C.1 Sparsity of $W_0$ and $\Pi_0$

Define  $\tilde{M}$  as the number of nonzero elements of  $\Pi_0$ . We say that  $\Pi_0$  is sparse if  $\tilde{M} \ll NT$ . Denote the number of connected pairs in  $W_0$  via paths of any length as  $\tilde{m}_c$ . We equivalently say that  $W_0$  is “sparsely connected” if  $\tilde{m}_c \ll NT$ . We show that the sparsity of  $\Pi_0$  is related to the sparse connectedness of  $W_0$ .

**Proposition 3.**  $\Pi_0$  is sparse if, and only if, the number of unconnected pairs  $W_0$  is large.

*Proof.* For  $|\rho_0| < 1$ , we have that

$$\Pi_0 = \beta_0 I + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k.$$

Given that  $\rho_0 \beta_0 + \gamma_0 \neq 0$ , it follows directly that  $[\Pi_0]_{ij} = 0$  if, and only if, there are no paths between  $i$  and  $j$  in  $W_0$ . Therefore, the sparsity of  $\Pi_0$  translates into a large number of  $(i, j)$  unconnected pairs in  $W_0$ .  $\square$

On the one hand, sparsity does not imply sparse connectedness. A circular graph is clearly sparse, but all nodes connect with all other nodes through a path of length at most  $\frac{N}{2}$ . On the other hand, sparse connectedness implies sparsity and therefore is a stronger requirement. To see this, take any arbitrary network  $G$  with  $\tilde{m}(G)$  non-zero elements and  $\tilde{m}_c(G)$  connected pairs. Now consider the operation of “completing”  $G$ : for every connected  $(i, j)$  pair, add a direct



link between  $(i, j)$  if nonexistent in  $G$  and denote the resulting matrix as  $\mathcal{C}(G)$ . It is clear that  $\tilde{m}(G) \leq \tilde{m}(\mathcal{C}(G))$ . Yet,  $\tilde{m}(\mathcal{C}(G)) = \tilde{m}_c(G)$ .

## C.2 Adaptive Elastic Net

In this section, we detail the algorithm, the computational steps, and the property of the estimator as derived in Caner and Zhang (2014). For any given choice of  $p = (p_1, p_1^*, p_2)$ , the algorithm is composed of three main steps pertaining to the estimation of the Elastic Net (step 5 below), the Adaptive Elastic Net version (step 6), and the Unpenalised Post-GMM estimator (step 7). Other steps deal with the selection of the initial conditions and other details of the implementation. The runtime for a given set of penalization parameters (that is, steps 1-7 below) is expected to be around 2-3 mins for  $N = 30$  and 15-20 mins for  $N = 70$ . The increase in the computational time is due to the fact that the number of parameters to estimate grows at an  $N^2$  rate.<sup>5</sup>

1. For any choice of penalization parameters  $p = (p_1, p_1^*, p_2)$ , run steps 2-7 below.
2. Data is standardised (subtracting the mean and dividing by the standard error), and individual and time effects are removed.
  - a. For the dynamic versions, use time weights  $\omega_t$ , which are chosen by the practitioner and applied after standardization.
3. Initial conditions are set at  $(\rho, \beta, \gamma) = (0.5, \hat{\beta}_{ols}, 0)$ , where  $\hat{\beta}_{ols}$  is the OLS regression of  $y_{it}$  on  $x_{it}$  after demeaning.
4. The derivatives of the objective function with respect to all  $w_{ij}$  are computed, and set to  $w_{ij} = 0$  for the derivatives that are smaller than a small number  $\eta$ . This number can be chosen to be equal to  $p_1$  following Caner and Zhang (2014). To be conservative, we choose  $\eta = \min(p_1, 0.01)$ . Those  $w_{ij}$  are set at zero for all the algorithm that follow.
5. *Elastic Net*. We devise an algorithm in the following manner.
  - a. For any given  $(\rho, \beta, \gamma)$ , we write,

$$\underbrace{y_t - x_t\beta}_{\equiv \tilde{y}_{it}} = W \underbrace{(\rho y_t + x_t\gamma)}_{\equiv \tilde{x}_t} + \epsilon_t,$$

which is endogenous due to the mechanical dependence of  $\tilde{x}_t$  on  $y_t$ , so it is instrumented with  $x_t$ . This expression is implemented through a fast Least-Angle Regression algorithm (LARS) from Caner *et al.* (2018) (Section 4.2) that extends the original LARS

---

<sup>5</sup>Benchmarked on a 2020 Macbook Pro with an M1 processor and 16gb RAM.

algorithm to the GMM case. We obtain  $\hat{W}$  for any given  $(\rho, \beta, \gamma)$ . At this stage, only impose that  $|w_{ij}| \leq 1$  and, in particular, do not impose row-sum normalization.

- b. Use the L-BFGS-B algorithm to minimize the objective function and obtain  $(\tilde{\rho}, \tilde{\beta}, \tilde{\gamma})$ , and then compute  $\tilde{W}$  from step 5a.
6. *Adaptive Elastic Net*. Use  $\tilde{W}$  to compute the adaptive weights. Following Caner and Zhang (2014), we select the adaptive weight penalty as  $w_{ij}^{-\kappa}$  if  $w_{ij} > 0.05$  where  $\kappa = 2.5$ . If  $w_{ij} < 0.05$ , we set  $0.05^{-\kappa}$ . This ensures that the second-stage estimates can be nonzero even if the first-stage estimates were zero or small.
    - a. We then run a version of the algorithm described in step 5a, with the difference that we impose row-sum normalization. Due to this, we can no longer use a restriction-free LARS algorithm. We instead use CVXR, a convex optimization routine package in R.
    - b. Similar to step 5b, we use the L-BFGS-B package to solve for  $(\hat{\rho}, \hat{\beta}, \hat{\gamma})$  and compute the associated  $\hat{W}$ .
  7. *Unpenalised Post-GMM*. Re-estimate  $(\rho, \beta, \gamma)$  and  $W$  on the support estimated in the previous step. In other words, the zero elements of  $\hat{W}$  are set to zero, i.e., the GMM objective function is estimated under the restriction that  $\{W_{ij} = 0 : \hat{W}_{ij} = 0\}$  and setting the penalization to zero. Asymptotic standard errors are also computed at this step.
  8. Re-compute steps 1-6 on a grid for  $(p_1, p_1^*, p_2)$  and choose the penalization parameter by BIC.

We implement the following modifications and adjustments of the algorithm for the empirical analysis presented in Section 4. Steps 4' and 8' provide more stability to the period-by-period estimates by avoiding small-sample biases in the estimation of the derivative and of the information criteria.

- 2a'. We use uniform weights  $\omega_t = 1$  and a Gaussian kernel with a center varying period by period from  $t = 1963$  to 2015, i.e., ranging from the beginning to the end of the sample. The variance of the kernel is set such that 75% of the weight is given to the first half of the data (pre-1988), when the center of the kernel is at the beginning of the sample at  $t = 1963$ .
- 4'. We compute the derivative under uniform weights, store it, and input the derivative for each dynamic-network version. This is computationally efficient (given that the numerical approximation of the derivative is intensive) and ensures that the derivative is computed from the full sample.
- 8'. The grid for the dynamic-network versions is set around the penalization parameters chosen under the uniform kernel. More specifically, we run the procedure for the penalization parameters for one point in all directions in the grid and select the dynamic-network penalization parameters by BIC.

### C.3 OLS

For the purpose of estimation, it is convenient to write the model in the stacked form. Let  $x = [x_1, \dots, x_T]'$  be the  $T \times N$  matrix of explanatory variables,  $y_i = [y_{i1}, \dots, y_{iT}]'$  be the  $T \times 1$  vector of response variables for individual  $i$ , and  $\pi_i^0 = [\pi_{i1}^0, \dots, \pi_{iN}^0]'$ , where  $\pi_{ij}^0$  is a short notation for the  $(i, j)$ -th element of  $\Pi_0$ . The concise model is then

$$y_i = x\pi_i^0 + v_i \quad (6)$$

for each  $i = 1, \dots, N$ , where also  $v_i = [v_{i1}, \dots, v_{iT}]'$ . Model (6) can then be estimated equation by equation. Denote  $\pi^0 = [\pi_1^0, \dots, \pi_N^0]'$ . Stacking the full set of  $N$  equations,

$$y = X\pi^0 + v \quad (7)$$

where  $y = [y_1, \dots, y_N]$ ,  $X = I_N \otimes x$ ,  $\pi^0 = \text{vec}(\Pi_0')$ , and  $v = [v_1, \dots, v_N]$ . If the number of individuals in the network  $N$  is fixed and much smaller than the number of data points available,  $N^2 \ll NT$ , equation (7) can be estimated via ordinary least squares (OLS). Under suitable regularity conditions, the OLS estimator  $\hat{\pi} = (X'X)^{-1} X'y$  is asymptotically distributed,

$$\sqrt{NT}(\hat{\pi} - \pi^0) \xrightarrow{d} \mathcal{N}(0, Q^{-1}\Sigma Q^{-1})$$

where  $Q_T \equiv \frac{1}{NT}X'X$ ,  $Q \equiv p \lim_{T \rightarrow \infty} Q_T$ ,  $\Sigma_T \equiv \frac{1}{NT}X'vv'X$ , and  $\Sigma \equiv p \lim_{T \rightarrow \infty} \Sigma_T$ . The proof is standard and omitted here. As noted above, in typical applications, it is customary to row-sum normalize matrix  $W$ . If no individual is isolated, one obtains that, by equation (5),

$$\begin{aligned} \Pi_0 \iota_N &= \beta_0 \iota + (\rho_0 \beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k \iota \\ &= \frac{\beta_0 + \gamma_0}{1 - \rho_0} \iota \end{aligned} \quad (8)$$

where  $\iota_N$  is the  $N$ -length vector of ones. The last equality follows from the observation that, under row-normalization of  $W_0$ ,  $W^k \iota = W \iota = \iota$ ,  $k > 0$ . Equation (8) implies that  $\Pi_0$  has constant row-sums, which implies that row-sum normalization is, in principle, testable. This suggests a simple Wald statistic applied to the estimates of  $\pi^0$ . Under the null hypothesis,

$$\sqrt{NT}R\hat{\pi} \xrightarrow{d} \mathcal{N}(0, RQ^{-1}\Sigma Q^{-1}R')$$

where  $R = [I_{N-1} \otimes \iota'_N; -\iota_{N-1} \otimes \iota'_N]$ . The Wald statistic is  $W = NT(R\hat{\pi})'(Q^{-1}\Sigma Q^{-1})^{-1}(R\hat{\pi}) \sim \chi_{N-1}^2$ , which is a convenient expression for testing row-sum normalization of  $W_0$ . We also note

that the asymptotic distribution of  $\hat{\theta}$  can be immediately obtained by the Delta Method,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \nabla_{\theta}'^{-1} Q^{-1} \Sigma Q^{-1} \nabla_{\theta})$$

where  $\nabla_{\theta}$  is the gradient of  $\hat{\theta}$  with respect to  $\hat{\pi}$ . We note that the derivation of the Wald statistic for testing the row-sum normalization and the asymptotic distribution of  $\hat{\theta}$  does not depend on the OLS implementation, and can be easily adjusted for any estimator for which the asymptotic distribution is known.

## D Simulations

### D.1 Set-up

The simulations are based on two stylized random network structures and two real-world networks. These networks vary in their size, complexity, and aggregate and node-level features. All four networks are also sparse. Networks (i) and (ii) are stylized, while (iii) and (iv) are based on real data:

- (i) Erdős-Renyi network: We randomly pick exactly one element in each row of  $W_0$  and set that element to 1. This is a random graph with in-degree equal to 1 for every individual (Erdős and Renyi, 1960). Such a network could be observed in practice if connections were formed independently of one another. With  $N = 30$ , the resulting density of links is 3.45%.
- (ii) Political party network: There are two parties, each with a party leader. The leader directly affects the behavior of half the party members. We assume that one party has twice members as the other. More specifically, we assume individuals  $i = 1, \dots, \frac{N}{3}$  are affiliated with Party A and led by individual 1; individuals  $i = \frac{N}{3} + 1, \dots, N$  are affiliated with Party B and led by individual  $\frac{N}{3} + 1$ . This difference in party size allows us to evaluate our ability to recover and identify central leaders, even in the smaller party. To test the procedure further, we add one random link per row to represent ties that are not determined by links to the party leader. We simulate this network for various choices of  $N$ . If  $N$  is not a multiple of three, we round  $\frac{N}{3}$  to the nearest integer. With  $N = 30$ , this network has a density of 5.17%.
- (iii) Coleman’s (1964) high school friendship network survey: In 1957–8, students in a small high school in Illinois were asked to name “fellows that they go around with most often.” A link was considered if the student nominated a peer in either survey wave. The full network has  $N = 73$  nodes, of which 70 are non-isolated and so have at least one link to another student. On average, students named just over five friendship peers. This network has a density of 7.58%. Furthermore, the in-degree distribution shows that most individuals received a small number of links, while a small number received many peer nominations.

(iv) Banerjee *et al.*'s (2013) village network survey: These authors conducted a census of households in 75 villages in rural Karnataka, India, and survey questions included several about relationships with other households in the village. To begin with, we use social ties based on family relations (later examining insurance networks). We focus on village 10, which comprises  $N = 77$  households and so similar in size to network (iii). In this village, there are 65 non-isolated households, with at least one family link to another household. This network has a density of 5.07%.

We simulate the real-world networks (iii) and (iv) using the non-isolated nodes in each (so  $N = 70$  and 65 respectively). We proceed as follows for each network. First, for each node, we randomly assign one of its links to have three times the strength of other links. As the underlying data generating process is assumed to allow for common time effects ( $\alpha_t$ ), we then set the weight of all the node's other links to be equal and such that row-sum normalization (A4') is complied with. For example, if in a given row of  $W_0$  there are two links, one will be randomly selected to be set to .75, and the other will be set to .25. If there are three links, one is set to .6 and the other two are set to each have weight .2 to maintain row-sum normalization, and so on. For the Erdős-Renyi network, there are thus only strong ties, as each node has only one link to another node.

As we consider larger networks, we typically expect them to have more nonzero entries in each row of  $W_0$ , but row-sum normalization means that each weaker link will be of lower value. This works *against* the detection of weaker links using estimation methods involving penalization because they impose that small-parameter estimates shrink to zero.<sup>6</sup> Finally, to aid exposition, we set a threshold value for link strength to distinguish “strong” and “weak” links. A strong (weak) link is defined as one for which  $W_{0,ij} > (<=) .3$ .

Panel data for each of the four simulations is generated as

$$y_t = (I - \rho_0 W_0)^{-1} (x_t \beta + W_0 x_t \gamma + \alpha_t \iota + \alpha^* + \epsilon_t),$$

where  $\alpha_t$  is a (scalar) time effect and  $\alpha^*$  is a  $N \times 1$  vector of fixed effects, drawn respectively from  $N(1, 1)$  and  $N(\iota, I_{N \times N})$  distributions. We consider  $T = \{5, 10, 15, 25, 50, 75, 100, 125, 150\}$ . The true parameters are set to  $\rho_0 = .3$ ,  $\beta_0 = .4$ , and  $\gamma_0 = .5$  (thus satisfying Assumption A3). The exogenous variable ( $x_t$ ) and error term ( $\epsilon_t$ ) are simulated as standard Gaussian, both generated from  $N(0_N, I_{N \times N})$  distributions. This is similar to the variance terms set in other papers, e.g., Lee (2004). We later conduct a series of robustness checks to evaluate the sensitivity of the simulations to alternative parameter choices and the presence of common and individual-level shocks. For each combination of parameters, we conduct 1,000 simulation runs.

---

<sup>6</sup>Caner and Zhang (2014) state that “local to zero coefficients should be larger than  $T^{-\frac{1}{2}}$  to be differentiated from zero.”

## D.2 Robustness

Table A2 presents results for the recovered stylized networks under varying network sizes,  $N = \{15, 30, 50\}$ . Differences between the true and estimated networks are fairly constant as  $N$  increases: even for small  $N = 15$ , a large proportion of zeros and non-zeros are correctly estimated. In all cases, biases in  $\hat{\rho}$  and  $\hat{\gamma}$  decrease with larger  $T$ .

Table A3 conducts robustness checks on the sensitivity of the estimates to parameter choices for the stylized networks. We consider the true parameters  $\rho_0 = \{.1, .3, .7, .9\}$ ,  $\gamma_0 = \{.3, .7\}$ , and  $\beta_0 = \{.2, .6\}$ . We also introduce a common shock in the disturbance variance-covariance matrix by varying  $q$  in

$$\epsilon_t \sim N \left( 0, \begin{bmatrix} 1 & q & \cdots & q \\ q & 1 & \cdots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \cdots & 1 \end{bmatrix} \right)$$

where we consider  $q = \{.3, .5, .8, 1\}$ . We find the procedure to be robust to the true values of  $\rho_0$ ,  $\beta_0$ ,  $\gamma_0$ , and  $q$ . The method performs well in all scenarios.

The next set of robustness checks demonstrates the gains from using the Adaptive Elastic Net GMM estimator over alternative estimators. Table A4 shows simulation results using Adaptive Lasso estimates of the interaction matrix  $\Pi_0$ , so estimating and penalizing the reduced-form. The Adaptive Lasso estimator performs worse, and increased sample sizes are necessary to achieve similar performance compared to the Adaptive Elastic Net GMM. Table A5 then shows the performance of the procedure based on OLS estimates of  $\Pi_0$ . Given OLS requires  $m \ll NT$ , we use a time dimension ten times larger,  $T = \{500, 1000, 1500\}$ , and still find a deterioration in performance compared to the Adaptive Elastic Net GMM estimator.<sup>7</sup>

Taken together, these robustness checks suggest the Adaptive Elastic Net GMM estimator is preferred over Adaptive Lasso and OLS estimators. As discussed in the text, the procedure recovers true strong links. In finite samples, weak links can be detected as zeros due to the shrinkage estimator employed. In turn, row-sum normalization may imply that the strength of strong edges is over-estimated. We also showed that *a fortiori*, the procedure can recover network- and node-level statistics. It does so in networks that vary in size and complexity, and as the underlying social interactions model varies in the strength of endogenous and exogenous social effects, and the structure of shocks.

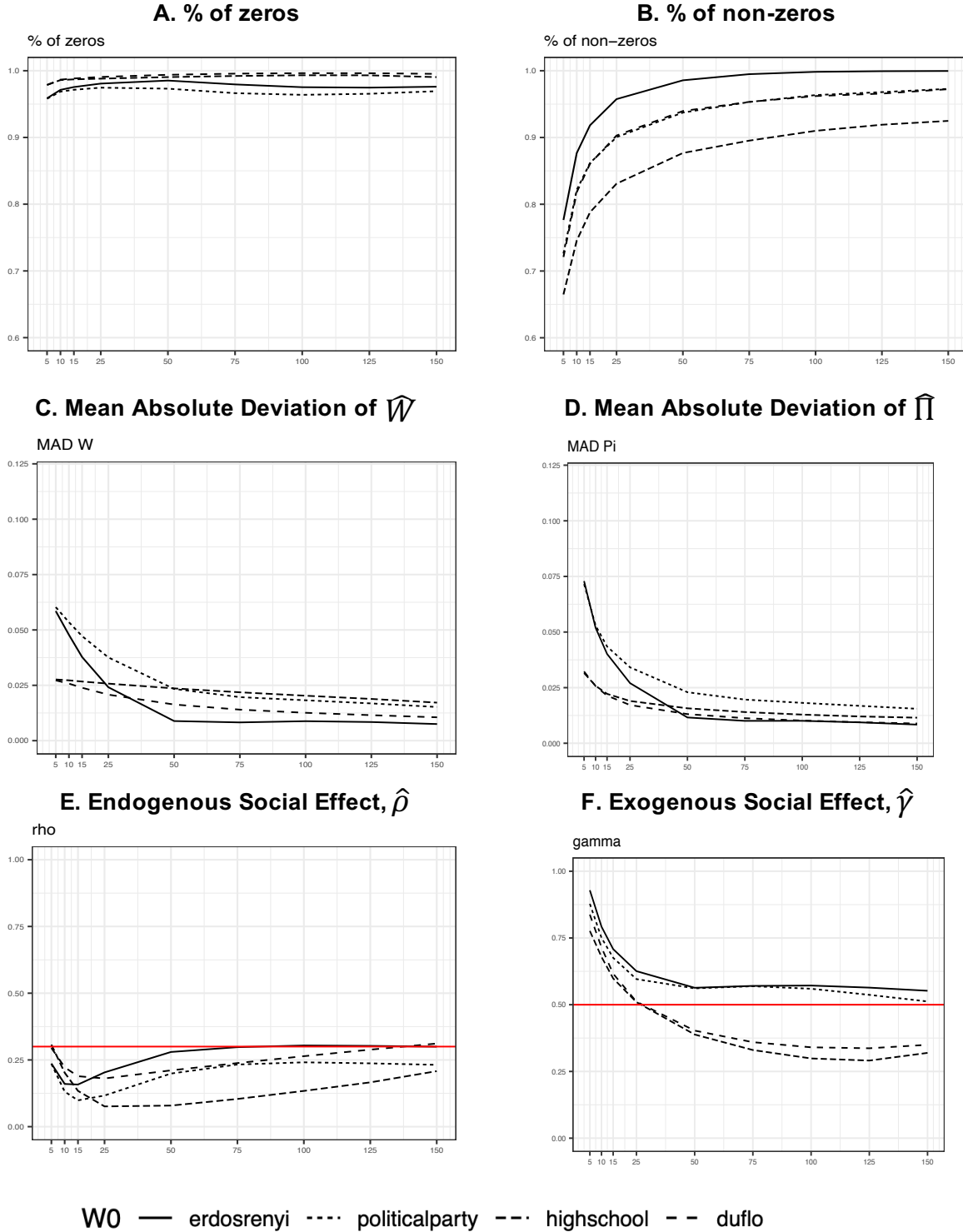
---

<sup>7</sup>As opposed to the penalized estimates, all OLS estimates are different from zero. We compute the “% True Zeroes” as the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05.

## References

- AMBROSETTI, A. AND G. PRODI (1972). On the Inversion of Some Differentiable Mappings with Singularities between Banach Spaces. *Annali di Matematica Pura ed Applicata*, 93, 231–46.
- (1995). *A Primer of Nonlinear Analysis*, Cambridge University Press.
- BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013). The Diffusion of Microfinance. *Science*, 341, 1236498.
- BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND R. JAYARAMAN (2015). Linear Social Interactions Models. *Journal of Political Economy*, 123, 444–96.
- BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009). Identification of Peer Effects Through Social Networks. *Journal of Econometrics*, 150, 41–55.
- CANER, M., X. HAN, AND Y. LEE (2018). Adaptive Elastic Net GMM Estimation With Many Invalid Moment Conditions: Simultaneous Model and Moment Selection. *Journal of Business and Economic Statistics*, 36, 24–46.
- CANER, M. AND H. H. ZHANG (2014). Adaptive Elastic Net for Generalized Method of Moments. *Journal of Business and Economic Statistics*, 32, 30–47.
- COLEMAN, J. S. (1964). *Introduction to Mathematical Sociology*, London Free Press Glencoe.
- DE MARCO, G., G. GORNI, AND G. ZAMPIERI (2014). Global Inversion of Functions: an Introduction. *ArXiv:1410.7902v1*.
- ERDÖS, P. AND A. RENYI (1960). On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5, 17–60.
- KRANTZ, S. G. AND H. R. PARKS (2013). *The Implicit Function Theorem*, Birkhauser.
- LEE, L.-F. (2004). Asumptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica*, 72, 25.
- MANRESA, E. (2016). Estimating the Structure of Social Interactions Using Panel Data. Manuscript.
- ROTHENBERG, T. (1971). Identification in Parametric Models. *Econometrica*, 39, 577–91.
- STRANG, G. (2006). *Linear algebra and its applications*, Belmont, CA Thomson, Brooks/Cole.
- WOOLDRIDGE, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

# Figure A1: Simulation Results, Adaptive Elastic Net GMM



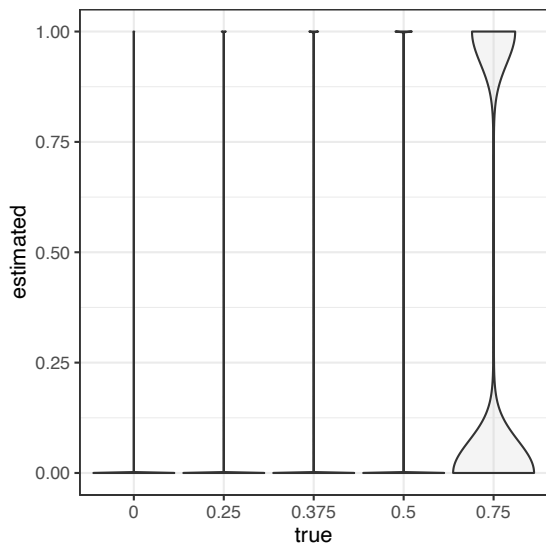
**Notes:** These simulation results are based on the Unpenalised GMM algorithm, with penalization parameters chosen by BIC, under various true networks and time periods  $T=25, 50, 75, 100, 125$  and  $150$ . In all cases, 1000 Monte Carlo iterations were performed. The true parameters are  $\rho=0.3$ ,  $\beta=0.4$  and  $\gamma=0.5$ . In Panel A, the % of zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. In Panel B, the % of non-zeros refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. In Panels C and D, the Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix  $W$  and the reduced form matrix respectively. In Panels E and F, the true parameter values are marked in the horizontal red lines. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects.



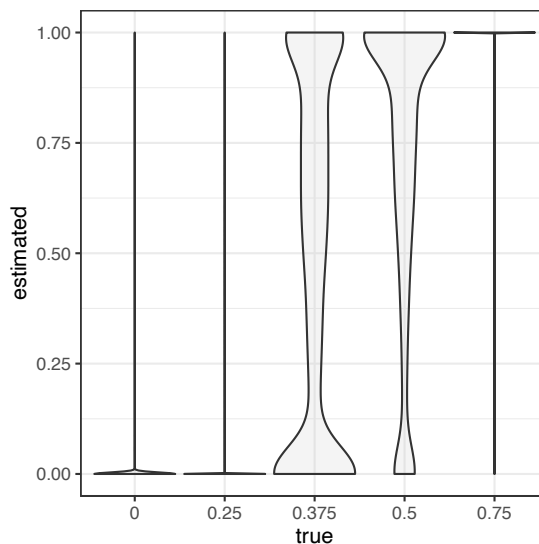
**Figure A2: Simulation Results, Adaptive Elastic Net GMM, Distribution of Main Estimates**

**High-School Friendship Network**

**A. Distribution of estimates of W, T=25**

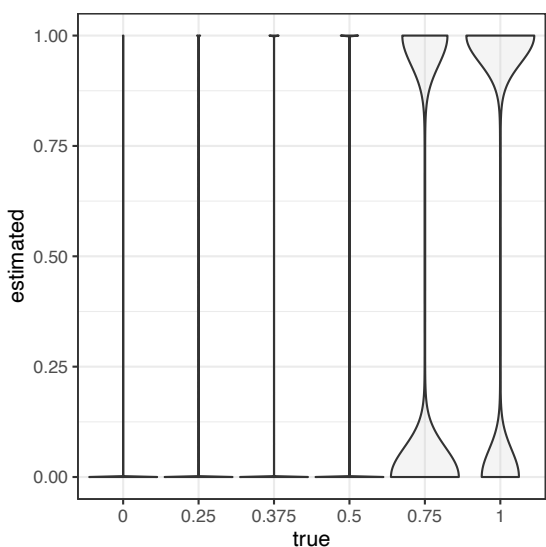


**B. Distribution of estimates of W, T=150**

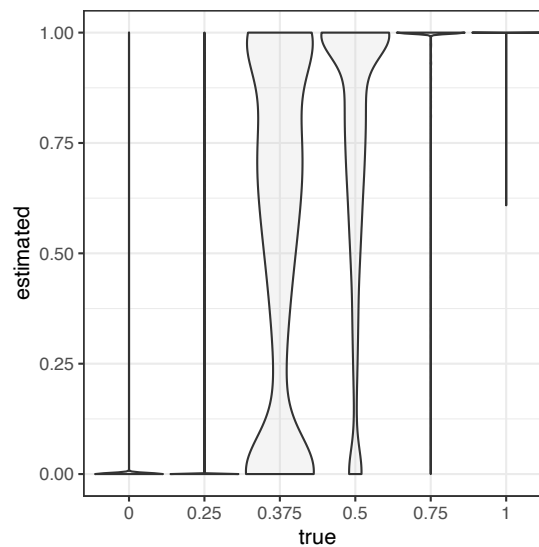


**Village Network**

**C. Distribution of estimates of W, T=25**



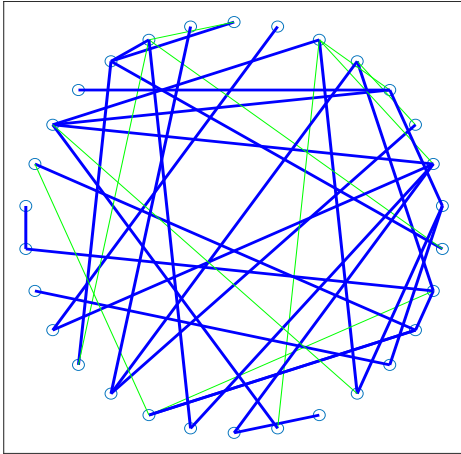
**D. Distribution of estimates of W, T=150**



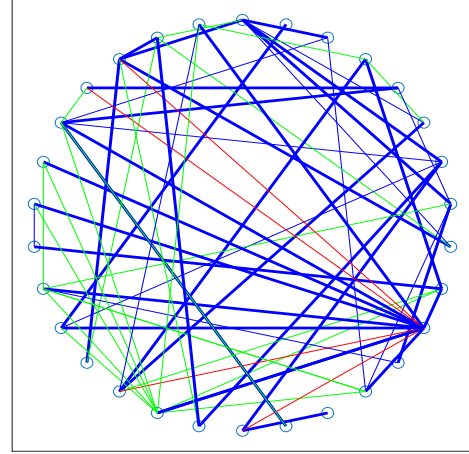
**NOTES.** These simulation results are based on the unpenalised GMM algorithm, with penalization parameters chosen by BIC, under the high-school network. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are  $\rho=0.3$ ,  $\beta=0.4$  and  $\gamma=0.5$ . Panels A and C shows the distribution of the estimates of W by their true value and T=25. We show the distribution of the estimated elements of W in five distinct values of W true values. Panels B and D shows the corresponding distribution for T=150. All specifications include time and node fixed effects.

## Figure A3: Simulated and True Networks

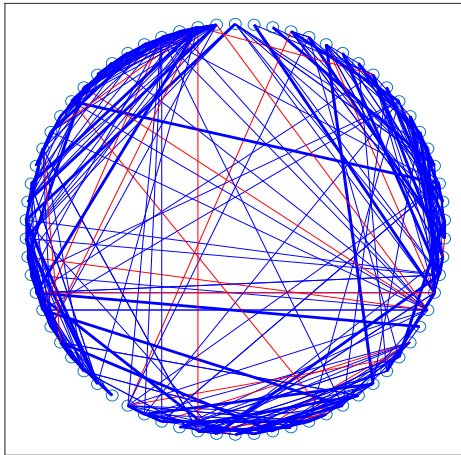
A. Erdos-Renyi



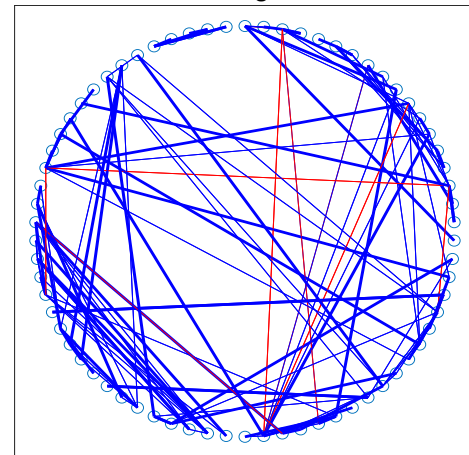
B. Political Party



C. High-school

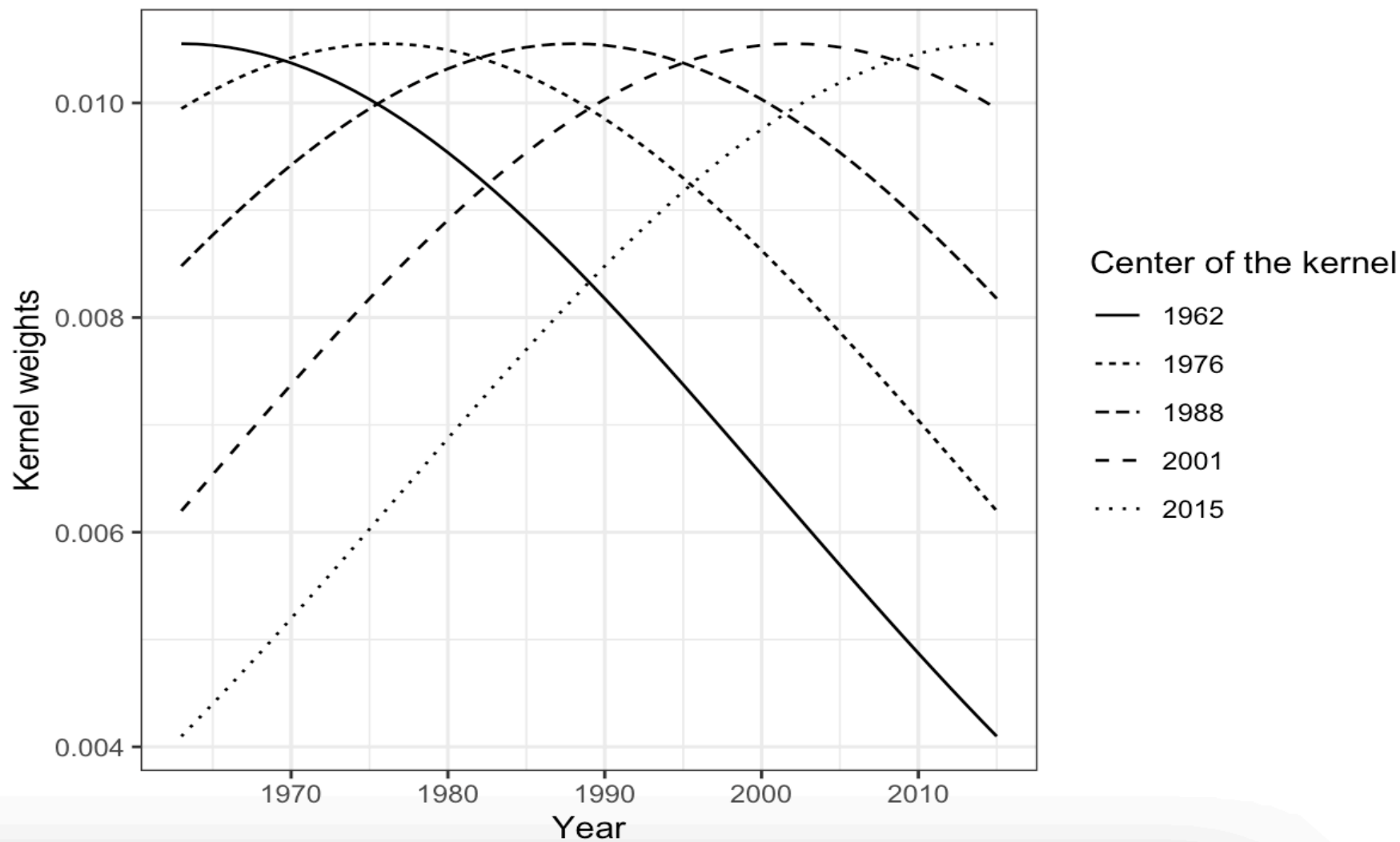


D. Village

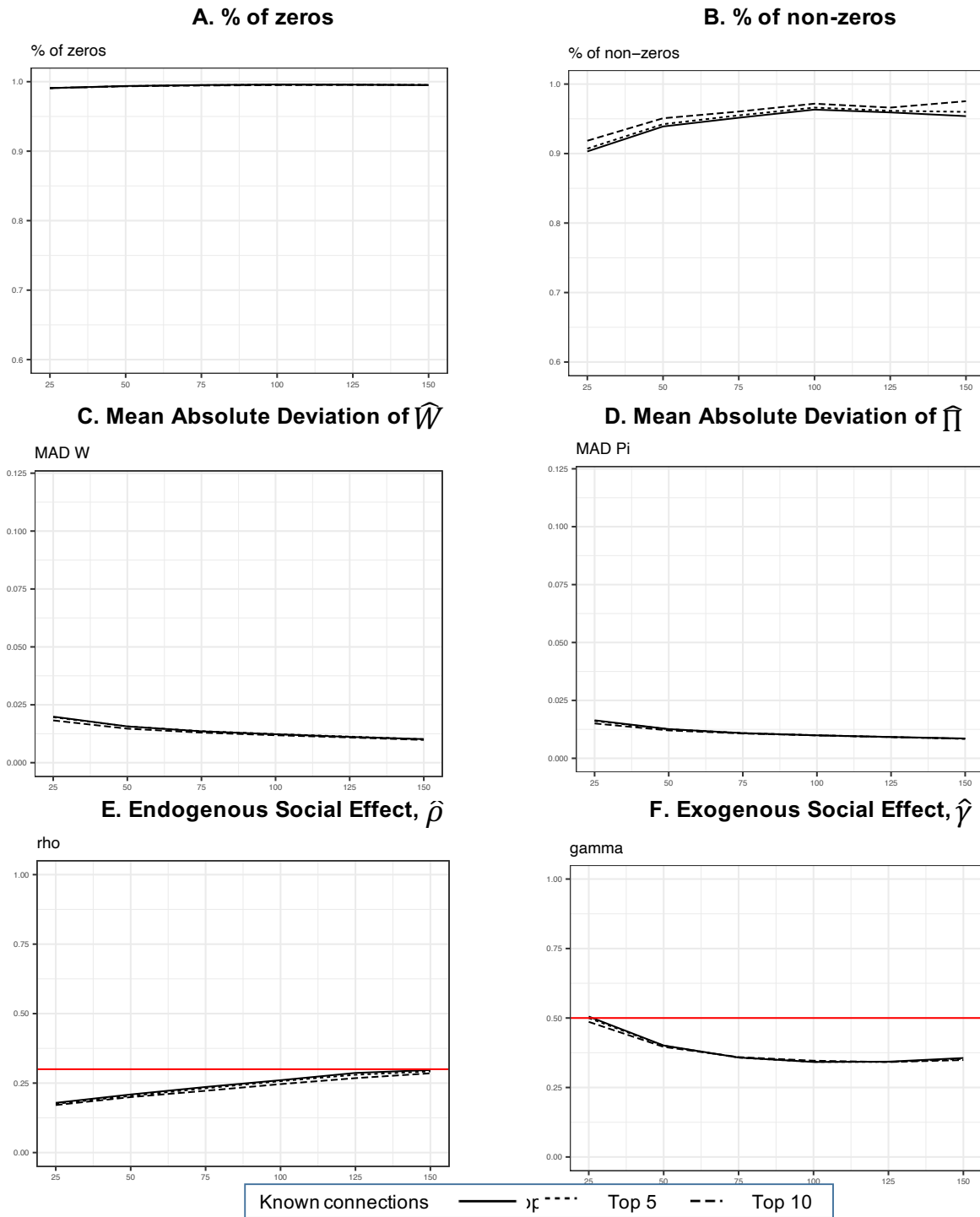


**Notes:** These simulation results are based on the Unpenalised GMM algorithm, with penalization parameters chosen by BIC, under various true networks and time periods  $T=50, 100$  and  $150$ . In the two stylized networks (Erdos-Renyi and political party), we set  $N=30$ , and the real world networks, the high school friendship and village network are based on  $N=65$  and  $70$  non-isolated nodes respectively. Party leaders in the political party network are marked in black in Panel B. In all cases, 1,000 Monte Carlo iterations were performed. The true parameters are  $\rho_0=.3$ ,  $\beta_0=.4$  and  $\gamma_0=.5$ . All specifications include time and node fixed effects. Kept edges are depicted in blue: these links are estimated as non-zero in at least 5% of the iterations and are also non-zero in the true network. Added edges are depicted in green: these links are estimated as non-zero in at least 5% of the iterations but the edge is zero in the true network. Removed edges are depicted in red: these links are estimated as zero in at least 5% of the iterations but are non-zero in the true network. The figures further distinguish between strong and weak links: strong links are shown in thick edges (whose strength is greater than or equal to .3), and weak links are shown as thin edges.

**Figure A4: Kernel Used for Dynamic Specifications**



## Figure A5: Simulation Results, Adaptive Elastic Net GMM Partial Knowledge of $W_0$ , Village Network



**Notes:** These simulation results are based on the Banerjee et al. (2013) village network, using the Adaptive Elastic Net GMM algorithm, with penalization parameters chosen by BIC, under various assumptions about knowledge of the true network and time periods  $T=25, 50, 100, 125$  and  $150$ . The "Village" case refers to the simulation implemented without knowledge about the true network. "Village (top 3)" refers to the case where all connections of the three households with highest out-degrees are assumed to be known. "Village (top 5)" and "Village (top 103)" are analogously defined. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are  $\rho_0=0.3$ ,  $\beta_0=0.4$  and  $\gamma_0=0.5$ . In Panel A, the % of zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. In Panel B, the % of non-zeros refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. In Panels C and D, the Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix  $W$  and the reduced form matrix respectively. In Panels E and F, the true parameter values are marked in the horizontal red lines. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects.

**Table A1: True and Recovered Networks**

**Erdos-Renyi**    **Political Party**    **High school**    **Village**  
*Coleman (1964)*    *Banerjee et al. (2013)*

**A. True Networks**

	<b>30</b>	<b>30</b>	<b>70</b>	<b>65</b>
<b>Number of nodes</b>	<b>30</b>	<b>30</b>	<b>70</b>	<b>65</b>
<b>(a) Network-wide statistics</b>				
<i>Number of edges</i>	30	45	366	240
<i>Number of strong edges</i>	30	30	61	65
<i>Number of weak edges</i>	0	15	305	180
<i>Number of reciprocated edges</i>	2	2	184	242
<i>Clustering coefficient</i>	.000	.037	.404	.327
<i>Number of components</i>	2	1	1	4
<i>Size of maximal component</i>	20	30	70	51
<i>Standard deviation of the diagonal of squared W</i>	.254	.254	.084	.205
<b>(b) Node-level statistics</b>				
<i>In-degree distribution</i>	1.00 (1.05)	1.50 (2.49)	5.23 (3.64)	3.50 (2.39)
<i>Out-degree distribution</i>	1.00 (0.00)	1.50 (0.51)	5.23 (2.04)	3.50 (2.37)
<i>Nodes with highest out-degree</i>	{ 7, 11, 26 }	{ 1, 11, 28 }	{ 21, 22, 69 }	{ 16, 35, 57 }

**B. Recovered Networks**

<b>(a) Network-wide statistics</b>				
<i>Number of edges</i>	30	41	112	124
<i>Number of strong edges</i>	30	41	112	124
<i>Number of weak edges</i>	0	0	0	0
<i>Number of reciprocated edges</i>	2	2	24	84
<i>Clustering coefficient</i>	.000	.027	.330	.441
<i>Number of components</i>	2	1	2	7
<i>Size of maximal component</i>	20	30	68	42
<b>(b) Node-level statistics</b>				
<i>In-degree distribution</i>	1.00 (1.05)	1.37 (1.88)	1.61 (1.44)	1.77 (1.37)
<i>Out-degree distribution</i>	1.00 (0.00)	1.37 (0.49)	1.61 (0.82)	1.77 (1.05)
<i>Nodes with highest out-degree</i>	{ 7, 11, 26 }	{ 1, 11, 28 }	{ 21, 22, 69 }	{ 16, 35, 57 }

**Notes:** Panel A refers to the true networks. Panel B refers to the recovered networks. In each Panel, the summary statistics are divided into network-wide and node-level statistics. Strong edges are defined as those with strength greater than or equal to .3. For the in-degree and out-degree distribution, the mean is shown and the standard deviation is in parentheses. The nodes with the highest out-degree are those with the greatest influence on others, and are calculated as the column-sum of the social interaction matrix. The recovered network statistics are calculated over the average network across simulations with T=100.

**Table A2: Simulation Results, Adaptive Elastic Net GMM, Alternative Network Sizes**

	A. Erdos-Renyi									B. Political party								
	N = 15			N = 30			N = 50			N = 15			N = 30			N = 50		
	T=50	100	150	T=50	100	150	T=50	100	150	T=50	100	150	T=50	100	150	T=50	100	150
<b>% True Zeroes</b>	.945	.960	.974	.985	.975	.976	.997	.997	.991	.939	.958	.975	.973	.964	.969	.993	.993	.985
	(.016)	(.014)	(.011)	(.006)	(.006)	(.005)	(.001)	(.001)	(.003)	(.017)	(.015)	(.012)	(.007)	(.007)	(.006)	(.002)	(.002)	(.003)
<b>% True Non-Zeroes</b>	.973	.996	1.000	.986	.998	1.000	.993	.999	1.000	.949	.980	.993	.937	.964	.973	.974	.989	.995
	(.045)	(.017)	(.004)	(.023)	(.007)	(.004)	(.013)	(.004)	(.001)	(.063)	(.038)	(.022)	(.048)	(.037)	(.032)	(.025)	(.016)	(.012)
<b>MAD(<math>\widehat{W}</math>)</b>	.027	.014	.008	.009	.009	.008	.004	.001	.003	.037	.024	.019	.023	.018	.015	.013	.007	.007
	(.009)	(.006)	(.004)	(.004)	(.002)	(.002)	(.002)	(.001)	(.001)	(.008)	(.005)	(.004)	(.004)	(.003)	(.002)	(.002)	(.001)	(.001)
<b>MAD(<math>\widehat{\Pi}</math>)</b>	.030	.017	.011	.012	.010	.008	.005	.002	.003	.038	.026	.021	.023	.018	.016	.012	.007	.008
	(.008)	(.005)	(.003)	(.004)	(.002)	(.002)	(.002)	(.001)	(.001)	(.007)	(.005)	(.003)	(.004)	(.002)	(.002)	(.002)	(.001)	(.001)
<b><math>\widehat{\rho}</math></b>	.270	.281	.282	.280	.304	.300	.279	.298	.309	.242	.250	.246	.199	.241	.232	.205	.250	.272
	(.070)	(.046)	(.037)	(.039)	(.030)	(.025)	(.026)	(.018)	(.017)	(.082)	(.053)	(.043)	(.056)	(.042)	(.036)	(.033)	(.023)	(.022)
<b><math>\widehat{\beta}</math></b>	.409	.405	.403	.403	.402	.402	.402	.400	.400	.411	.404	.399	.404	.402	.401	.404	.400	.401
	(.043)	(.029)	(.024)	(.030)	(.020)	(.017)	(.024)	(.016)	(.013)	(.046)	(.030)	(.025)	(.031)	(.021)	(.017)	(.025)	(.016)	(.013)
<b><math>\widehat{\gamma}</math></b>	.618	.549	.518	.563	.572	.552	.519	.505	.529	.593	.508	.471	.561	.560	.512	.483	.469	.513
	(.071)	(.045)	(.031)	(.043)	(.034)	(.027)	(.025)	(.019)	(.019)	(.079)	(.051)	(.037)	(.057)	(.042)	(.037)	(.028)	(.023)	(.024)

**Notes:** These simulation results are based on the Unpenalised GMM algorithm, with penalization parameters chosen by BIC, under various true networks, network sizes and time periods T=50, 100 and 150. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are rho=0.3, beta=0.4 and gamma=0.5. The % of true zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. The % of true non-zeroes refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. The Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix W and the reduced form matrix respectively. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects. Standard errors across iterations are in parentheses.

**Table A3: Simulation Results, Adaptive Elastic Net GMM, Alternative Parameters**

	A. Erdos-Renyi											B. Political party												
	$\rho_0$				$\beta_0$		$\gamma_0$		q			$\rho_0$				$\beta_0$		$\gamma_0$		q				
	.1	.5	.7	.9	.2	.6	.3	.7	.3	.5	.8	1.0	.1	.5	.7	.9	.2	.6	.3	.7	.3	.5	.8	1.0
% True Zeroes	.967 (.006)	.987 (.004)	.995 (.003)	.985 (.006)	.993 (.005)	.979 (.006)	.964 (.007)	.985 (.004)	.981 (.005)	.987 (.004)	.998 (.002)	1.000 (.000)	.963 (.007)	.978 (.006)	.986 (.005)	.989 (.004)	.990 (.005)	.966 (.007)	.959 (.007)	.978 (.006)	.971 (.006)	.976 (.006)	.991 (.004)	.999 (.001)
% True Non-Zeroes	.999 (.004)	.998 (.008)	.970 (.039)	.437 (.131)	.978 (.032)	1.000 (.000)	.989 (.020)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	.966 (.037)	.976 (.033)	.969 (.042)	.791 (.102)	.901 (.071)	.992 (.017)	.910 (.058)	.990 (.021)	.991 (.020)	.997 (.010)	1.000 (.000)	1.000 (.000)
$MAD(\widehat{W})$	.012 (.002)	.004 (.001)	.003 (.003)	.031 (.006)	.004 (.003)	.006 (.002)	.015 (.003)	.004 (.001)	.005 (.002)	.003 (.001)	.000 (.000)	.000 (.000)	.018 (.002)	.013 (.002)	.011 (.002)	.020 (.005)	.014 (.003)	.014 (.002)	.023 (.003)	.011 (.002)	.013 (.002)	.011 (.001)	.007 (.001)	.005 (.001)
$MAD(\widehat{\Pi})$	.008 (.001)	.008 (.002)	.015 (.009)	.219 (.056)	.005 (.003)	.008 (.002)	.011 (.002)	.007 (.002)	.006 (.001)	.004 (.001)	.002 (.001)	.001 (.000)	.012 (.001)	.022 (.003)	.046 (.005)	.232 (.014)	.013 (.003)	.016 (.002)	.015 (.002)	.017 (.002)	.014 (.002)	.012 (.001)	.009 (.001)	.008 (.001)
$\widehat{\rho}$	.091 (.029)	.496 (.020)	.683 (.020)	.811 (.108)	.278 (.026)	.302 (.023)	.302 (.035)	.296 (.020)	.299 (.021)	.297 (.018)	.291 (.012)	.292 (.006)	.052 (.035)	.401 (.034)	.555 (.033)	.692 (.046)	.203 (.033)	.249 (.035)	.222 (.047)	.231 (.029)	.240 (.033)	.244 (.027)	.239 (.020)	.238 (.014)
$\widehat{\beta}$	.403 (.017)	.400 (.017)	.398 (.017)	.482 (.090)	.208 (.016)	.602 (.017)	.404 (.017)	.401 (.017)	.401 (.014)	.401 (.012)	.400 (.007)	.399 (.001)	.401 (.017)	.399 (.017)	.398 (.018)	.406 (.027)	.210 (.015)	.601 (.017)	.403 (.017)	.399 (.017)	.401 (.014)	.400 (.012)	.399 (.008)	.398 (.003)
$\widehat{\gamma}$	.589 (.030)	.515 (.022)	.487 (.022)	.452 (.300)	.506 (.025)	.545 (.028)	.395 (.035)	.720 (.023)	.530 (.021)	.510 (.016)	.485 (.010)	.481 (.005)	.539 (.039)	.476 (.032)	.434 (.031)	.362 (.065)	.429 (.034)	.538 (.037)	.379 (.041)	.656 (.034)	.502 (.029)	.486 (.023)	.448 (.014)	.436 (.010)

**Notes:** These simulation results are based on the Adaptive Elastic Net GMM algorithm, with penalization parameters chosen by BIC, under various true networks, network sizes, time periods T=100 and parameter values. In all cases, 1000 Monte Carlo iterations were performed. The % of true zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. The % of true non-zeroes refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. The Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix W and the reduced form matrix respectively. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects. Standard errors across iterations are in parentheses.

**Table A4: Simulation Results, Adaptive Lasso**

	A. Erdos-Renyi									B. Political party								
	N = 15			N = 30			N = 50			N = 15			N = 30			N = 50		
	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500
% True Zeroes	.935 (.052)	.959 (.043)	.974 (.034)	.975 (.009)	.987 (.005)	.991 (.005)	.979 (.012)	.988 (.007)	.992 (.005)	.957 (.015)	.969 (.013)	.974 (.013)	.974 (.019)	.983 (.011)	.986 (.009)	.982 (.010)	.990 (.004)	.992 (.003)
% True Non-Zeroes	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)	1.000 (.000)
$MAD(\widehat{W})$	.073 (.108)	.058 (.098)	.040 (.075)	.013 (.011)	.010 (.003)	.011 (.015)	.009 (.023)	.008 (.028)	.007 (.026)	.031 (.010)	.031 (.008)	.031 (.007)	.020 (.029)	.019 (.020)	.018 (.016)	.009 (.013)	.007 (.002)	.008 (.002)
$MAD(\widehat{\Pi})$	.250 (.451)	.152 (.363)	.079 (.263)	.007 (.004)	.006 (.001)	.005 (.004)	.008 (.029)	.006 (.029)	.003 (.006)	.018 (.004)	.015 (.003)	.015 (.002)	.017 (.060)	.010 (.022)	.008 (.010)	.007 (.020)	.004 (.000)	.004 (.000)
$\widehat{\rho}$	.429 (.277)	.379 (.224)	.332 (.162)	.225 (.097)	.267 (.043)	.277 (.037)	.144 (.121)	.208 (.100)	.227 (.066)	.272 (.064)	.274 (.043)	.277 (.033)	.243 (.132)	.259 (.077)	.259 (.055)	.109 (.111)	.172 (.086)	.200 (.069)
$\widehat{\beta}$	.358 (.190)	.365 (.051)	.370 (.045)	.374 (.012)	.378 (.008)	.379 (.007)	.379 (.012)	.384 (.008)	.385 (.006)	.371 (.020)	.374 (.014)	.376 (.011)	.370 (.013)	.375 (.008)	.377 (.007)	.376 (.013)	.381 (.007)	.383 (.005)
$\widehat{\gamma}$	.373 (.438)	.384 (.283)	.404 (.139)	.525 (.111)	.462 (.049)	.445 (.040)	.625 (.113)	.533 (.087)	.501 (.069)	.445 (.062)	.422 (.040)	.415 (.031)	.465 (.115)	.419 (.067)	.411 (.052)	.601 (.112)	.507 (.090)	.471 (.072)

**Notes:** These simulation results are based on the Adaptive Lasso algorithm, with penalization parameters chosen by BIC, under various true networks, network sizes and time periods T=50, 100 and 150. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are rho=0.3, beta=0.4 and gamma=0.5. The % of true zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. The % of true non-zeroes refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. The Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix W and the reduced form matrix respectively. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects. Standard errors across iterations are in parentheses.



**Table A5: Simulation Results, OLS**

	A. Erdos-Renyi									B. Political party								
	N = 15			N = 30			N = 50			N = 15			N = 30			N = 50		
	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500	T=500	1000	1500
% True Zeroes	.748	.823	.873	.774	.841	.882	.796	.852	.892	.752	.828	.877	.775	.842	.883	.795	.854	.893
	(.050)	(.034)	(.023)	(.033)	(.020)	(.013)	(.044)	(.023)	(.012)	(.031)	(.023)	(.021)	(.031)	(.019)	(.022)	(.039)	(.025)	(.013)
% True Non-Zeroes	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
$MAD(\widehat{W})$	.078	.049	.037	.059	.043	.035	.054	.041	.033	.062	.043	.035	.058	.042	.035	.053	.040	.033
	(.070)	(.037)	(.018)	(.023)	(.015)	(.002)	(.019)	(.016)	(.001)	(.039)	(.004)	(.003)	(.026)	(.011)	(.015)	(.023)	(.017)	(.001)
$MAD(\widehat{\Pi})$	.134	.051	.028	.054	.030	.024	.058	.030	.023	.054	.029	.023	.052	.031	.026	.051	.030	.023
	(.306)	(.157)	(.078)	(.080)	(.025)	(.025)	(.073)	(.023)	(.001)	(.124)	(.035)	(.002)	(.074)	(.035)	(.037)	(.055)	(.025)	(.001)
$\widehat{\rho}$	.342	.313	.303	.233	.249	.270	.269	.255	.263	.303	.302	.301	.218	.252	.284	.238	.247	.262
	(.223)	(.103)	(.057)	(.167)	(.109)	(.086)	(.191)	(.092)	(.063)	(.102)	(.048)	(.035)	(.172)	(.120)	(.088)	(.203)	(.112)	(.077)
$\widehat{\beta}$	.433	.402	.396	.401	.393	.396	.483	.392	.395	.389	.395	.396	.388	.393	.395	.427	.392	.395
	(.460)	(.183)	(.017)	(.240)	(.008)	(.007)	(.607)	(.010)	(.005)	(.028)	(.014)	(.010)	(.110)	(.008)	(.007)	(.440)	(.013)	(.005)
$\widehat{\gamma}$	.488	.507	.516	.687	.620	.574	.602	.641	.600	.568	.522	.518	.719	.618	.559	.719	.649	.602
	(.635)	(.267)	(.065)	(.405)	(.142)	(.106)	(.822)	(.156)	(.084)	(.231)	(.052)	(.042)	(.286)	(.150)	(.103)	(.555)	(.138)	(.103)

**Notes:** These simulation results are based on OLS estimates, under various true networks, network sizes and time periods T=500, 1000 and 1500. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are rho=0.3, beta=0.4 and gamma=0.5. The % of true zeroes refers to the proportion of true zero elements in the social interaction matrix that are estimated as smaller than .05. The % of true non-zeroes refers to the proportion of true elements greater than .3 in the social interaction matrix that are estimated as non-zeros. The Mean Absolute Deviations are the mean absolute error of the estimated network compared to the true network for the social interaction matrix W and the reduced form matrix respectively. The recovered parameter are the estimated parameters averaged across iterations. All specifications include time and node fixed effects. Standard errors across iterations are in parentheses.

**Table A6: Summary Statistics, Tax Competition Application**

	Obs	Mean	SD	Min	q25	Median	q75	Max
<b>A. Besley and Case sample (1962-1988)</b>								
State total tax per capita	1,248	0.371	0.266	0.036	0.145	0.300	0.530	1.345
State income per capita	1,248	10.617	2.248	4.452	9.120	10.587	12.141	19.409
Unemployment rate	1,248	5.868	2.209	1.800	4.200	5.500	7.000	17.800
Proportion of young	1,248	0.258	0.029	0.180	0.240	0.260	0.280	0.340
Proportion of elderly	1,248	0.106	0.020	0.050	0.090	0.110	0.120	0.180
State governor's age	1,248	51.088	7.441	33.000	45.000	50.000	56.000	73.000
<b>B. Extended sample (1962-2015)</b>								
State total tax per capita	2,544	1.016	0.798	0.036	0.300	0.858	1.579	4.298
State income per capita	2,544	13.394	3.850	4.452	10.564	13.025	15.866	27.974
Unemployment rate	2,544	5.754	2.037	1.800	4.300	5.400	6.800	17.800
Proportion of young	2,544	0.235	0.033	0.170	0.210	0.230	0.260	0.340
Proportion of elderly	2,544	0.118	0.023	0.050	0.100	0.120	0.130	0.190
State governor's age	2,544	53.538	8.049	33.000	47.000	53.000	59.000	78.000

**Notes:** Summary statistics of variables (in levels) used in subsequent regressions. Besley and Case sample runs from 1962 to 1988 and extended sample until 2015. State total tax per capita is the sum of sales, income and corporation tax in thousands of 1982 US dollars. State income per capita in thousands of 1982 US dollars. Proportion of young is the proportion of the population between 5 and 17 years. Proportion of elderly is the proportion of the population aged 65 or older. State governor's age in years. Data sources: State total tax per capita, Census of Governments (1972, 1977, 1982, 1987, 1992-2016) and Annual Survey of Government Finances (all other years); State income per capita, Bureau of Economic Analysis; Unemployment rate, Bureau of Labor Statistics; Proportion of young (aged 5-17) and elderly (aged 65+), Census Population & Housing Data; State governor's age and political variables manually sourced from individual governor's webpages on Wikipedia.

## Table A7: Exogenous Social Effects

Dependent variable: Change in per capital income and corporate taxes

Coefficient estimates, standard errors in parentheses

	(1) OLS	(2) GMM
Economic Neighbors' tax change (t - [t-2])	.402*** (.044)	.452*** (.132)
Economic Neighbors' income per capita	.041*** (.007)	.043 (.066)
Economic Neighbors' unemployment rate	1.062 (1.826)	17.401*** (.026)
Economic Neighbors' population aged 5-17	790.8*** (303.5)	3253.387*** (.032)
Economic Neighbors' population aged 65+	-1155.7*** (328.9)	3712.6*** (.022)
Economic Neighbors' governor age	-0.274 (.264)	2.358*** (.035)
<b>Period</b>		1962-2015
<b>First Stage (F-stat)</b>		10.5
<b>Controls</b>	Yes	Yes
<b>State and Year Fixed Effects</b>	Yes	Yes
<b>Observations</b>	2,544	2,544

**Notes:** \*\*\* denotes significance at 1%, \*\* at 5%, and \* at 10%. The sample covers 48 mainland US states running from 1962 to 2015. The dependent variable is the change in state *i*'s total taxes per capita in year *t*. In the OLS and GMM regressions, the economic neighbors' effect is calculated as the weighted average of economic neighbors-of-neighbors variables. Column 2 shows the GMM estimates where each economic neighbors' tax change is instrumented by the characteristics of all states. All regressions control for state *i*'s income per capita in 1982 US dollars, state *i*'s unemployment rate, the proportion of young (aged 5-17) and elderly (aged 65+) in state *i*'s population, and the state governor's age. All specifications include state and time fixed effects. With the exception of governor's age, all variables are differenced between period *t* and period *t*-2. Column 1 reports robust standard errors in parentheses. Column 2 reports standard errors adopting the procedure described in Caner and Zhang (2014).