HCEO WORKING PAPER SERIES

Recovering Social Networks from Panel Data: Identification, Simulations and an Application

Aureo de Paula
Imran Rasul
Pedro Souza

Working Paper

2018-013

03/2018

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Recovering Social Networks from Panel Data: Identification, Simulations and an Application*

Áureo de Paula Imran Rasul Pedro CL Souza†

March 2018

Abstract

It is almost self-evident that social interactions can determine economic behavior and outcomes. Yet, information on social ties does not exist in most publicly available and widely used datasets. We present methods to recover information on the entire structure of social networks from observational panel data that contains no information on social ties between individuals. 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 all globally identified. We describe how high-dimensional estimation techniques can be used to estimate the model based on the Adaptive Elastic Net GMM method. We showcase our method in Monte Carlo simulations using two stylized and two real world network structures. Finally, we employ our method to study tax competition across US states. We find the identified network structure of tax competition differs markedly from the common assumption of tax competition between geographically neighboring states. We analyze the identified social interactions matrix to provide novel 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 method shows how the analysis of social interactions can be usefully extended to economic realms where no network data exists.

JEL Classification: C18, C31, D85, H71.

*We gratefully acknowledge financial support from the ESRC through the Centre for the Microeconomic Analysis of Public Policy (RES-544-28-0001), the Centre for Microdata Methods and Practice (RES-589-28-0001) and the ERC (SG338187). We thank Edo Airoldi, Oriana Bandiera, Larry Blume, Yann Bramoullé, Stephane Bonhomme, Vasco Carvalho, Gary Chamberlain, Andrew Chesher, Sergio Firpo, Jean-Pierre Florens, Eric Gautier, Giacomo de Giorgi, Matthew Gentzkow, Stefan Hoderlein, Bo Honor'e, Matt Jackson, Dale Jorgensen, Christian Julliard, 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 Siegloch, Pasquale Schiraldi, Kevin Song, John Sutton, Adam Szeidl, Elie Tamer, and seminar and conference participants for valuable comments. We also thank Tim Besley and Anne Case for comments and sharing data. All errors remain our own.

†de Paula: University College London, São Paulo School of Economics, CeMMAP and IFS, a.paula@ucl.ac.uk; Rasul: University College London and IFS, i.rasul@ucl.ac.uk; Souza: PUC-Rio, pedroclsouza@econ.puc-rio.br

1 Introduction

In many economic environments, behavior and outcomes are shaped by social interactions between agents.[1] In microeconomic analysis of individual behavior, social interactions have been key to understanding outcomes as diverse as educational outcomes, the demand for financial assets, and technology adoption in low-income settings.[2] In macroeconomics, the structures of production and credit networks are important channels through which real and financial shocks propagate, or through which firms learn.[3] In political economy and public economics, ties between jurisdictions are key to understanding tax setting behavior.[4]

Underpinning all these bodies of research is some measurement of the underlying social ties between agents/nodes, be they individuals, firms or jurisdictions. However, information on social ties does not exist in most publicly available and widely used datasets. To overcome this limitation, the first generation of social interaction studies *postulated* ties based on some common observable. Classic examples include using geographic proximity as the basis on which social interactions occur, or appealing to the concept of homophily and assuming agents similar on observables socially interact with each other. A more recent wave of research has *elicited* data on social networks, predominantly in contexts of individual or household behavior. It is however increasingly recognized that both postulated and elicited networks remain imperfect solutions to the fundamental problem of missing data on social ties, because of econometric concerns that arise through either method, or simply because of the cost of collecting social networks data.[5]

Two consequences of the difficulty in collecting network data are that: (i) classes of problems in which social interactions play a critical role remain understudied, because social networks data is missing or too costly to collect; (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,

[1] Jackson *et al.* (2017) overview theoretical and empirical advances in the social networks literature. de Paula (2017) overviews advances in econometrics on identifying social interactions and related computational issues.

[2] Sacerdote (2001) present evidence of the impact of randomly assigned college peers on each others' GPA scores and other student outcomes; Bursztyn *et al.* (2014) show how financial asset purchases are causally influenced by peers through social learning and social utility; Conley and Udry (2010) show how social learning about a new technology takes place among farmers in Ghana as farmers adjust input usage to align with those of their information neighbors who were surprisingly successful in the previous period.

[3] Acemoglu *et al.* (2012) show how the propagation of microeconomic shocks through input-output linkages can lead to macroeconomic fluctuations. Chaney (2014) presents evidence on the dynamics of French firms' exports, highlighting how existing contacts are used to search for new trading partners.

[4] It has long been argued these cross-jurisdiction interactions could be driven by factor mobility (Tiebout, 1956) or by political yardstick competition (Shleifer, 1985; Besley and Case, 1994).

[5] As detailed in de Paula (2017), elicited networks are often self-reported, that can introduce error for 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.*, 2017).

using only observational panel data that itself contains *no* information on network ties. In short, our method allows the study of social interactions without data on actual social networks, and the validation of structures of social interaction where social ties have hitherto only been postulated.

We do so in an entirely standard setting in which a researcher is assumed to have panel data on individuals $i = 1, ...N$ for instances $t = 1, ..., T$. 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 instance t is denoted y_{it} and is generated according to a canonical structural model of social interactions:[6]

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

Outcome y_{it} depends on the outcome of other individuals to whom i is socially tied, y_{jt}, and possibly the characteristics of those individuals, x_{jt}. $W_{0,ij}$ measures whether and how the outcome and characteristics of j causally impact the outcome for i. As outcomes for all individuals obey equations analogous to (1), the system of equations can be written in matrix notation where the entire structure of social interactions is captured by the adjacency matrix, denoted W_0. Our approach allows for unobserved heterogeneity across individuals α_i and common shocks to all individuals α_t. This framework encompasses the classic linear-in-means specification of Manski (1993). In his terminology, ρ_0 and γ_0 capture endogenous and exogenous social effects, and α_t captures correlated effects. The distinction between endogenous and exogenous peer effects is critical, as only the former generates social multiplier effects.

Manski's seminal contribution set out the reflection problem of separately identifying endogenous social effects from exogenous and correlated effects in linear models.[7] However, it has been largely overlooked that Manski (1993) also set out another challenge, on the identification of the social network in the first place:

> "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." (Manski (1993), p. 536)

This is the problem we tackle and so expand the scope of identification beyond ρ_0, β_0 and

[6]Blume *et al.* (2015) present micro-foundations based on non-cooperative games of incomplete information for individual choice problems, that result in this estimating equation for a class of social interaction models.

[7]Manski (1993) highlights the difficulties (and potential restrictions) for identifying ρ_0, β_0 and γ_0 when all individuals interact with each other, and when this is observed by the researcher. In (1), this corresponds to $W_{0,ij} = N^{-1}$ if $i \neq j$ and $W_{0,ii} = 0$.

γ_0. Our point of departure from much of the literature is to therefore 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, ρ_0 and γ_0, are globally identified. Our identification strategy is fundamentally different from those employed elsewhere in this nascent literature. However it delivers sufficient conditions that are mild, and closely related 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).[8]

The formulation of our problem involves the estimation of a high-dimensional parameter vector including all the entries of the social interactions matrix W_0. We thus use the Adaptive Elastic Net GMM method (Caner and Zhang, 2014) to estimate the model (though other methods can also be employed).

By identifying the social interactions matrix W_0, our method allows the recovery of aggregate network characteristics such as the degree distribution, patterns of homophily, reciprocity, and clustering. Moreover, we also recover node-level statistics such as the strength of social interactions between nodes, and the centrality of nodes (as measured by their Katz-Bonacich or eigenvector centrality for example). As Jackson *et al.* (2017) and de Paula (2017) discuss, this is useful because such aggregate and node-level statistics often map back to underlying models of social interaction.[9]

We showcase our method using Monte Carlo simulations and a real world application. The simulations are based both 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 our method on the simulated panel data to recover estimates of all elements in W_0, and the endogenous and exogenous social effect parameters (ρ_0, γ_0). 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. These networks vary in size, complexity, and their aggregate and node-level features.

Despite this heterogeneity across scenarios, we find our method to perform well in all four

[8]If the known W_0 differs from the linear-in-means example, ρ_0, β_0 and γ_0 can be identified (Bramoullé *et al.*, 2009; De Giorgi *et al.*, 2010). Intuitively, identification in those cases relies on peers-of-peers that are not necessarily connected to individual i and can be used to leverage variation from exclusion restrictions in (1), or if there are groups of different sizes within which all individuals interact among 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). 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 the higher moments of ϵ_{it}. Finally, we note that (1) is a spatial autoregressive model. In that literature, W_0 is also typically assumed known. Anselin (2010) discusses this literature and reviews recent progress.

[9]For example, aggregate statistics help identify conditions under which a process of contagion leads to a persistent level of infection, or the extent to which polarized views are likely to coexist in society. Node-level statistics often relate to modelling the influences of a given agent on the behavior of others. More precisely, the various measures of centrality all capture aspects important for intermediation, for games of complements (Ballester *et al.*, 2006), learning and diffusion, the identification of key players, or the possibility to trade favors.

simulations. In a reasonable dimension of panel data T and with varying node numbers across simulations (N), we find the true network structure W_0 is well recovered. For each simulated network, the proportion of correctly identified true links is over 85% even for low $T = 25$. The proportion of true non-links (zeroes in W_0) captured correctly as zeros is over 90% even when $T = 25$. Both proportions rapidly increase with T. *A fortiori*, we are able to estimate aggregate and node-level statistics of each network, demonstrating the accurate recovery of key players in networks for example. Furthermore, 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 final part of our analysis applies our method to study a 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 years 1962-2015, $T = 53$.

We find the identified network structure of tax competition to differ markedly from the common assumption of competition between geographic neighbors. While geography is a robust determinant of tax competition, the identified economic network shows social interactions are far more spatially dispersed. The identified network has fewer edges than the geography-based network, and this is reflected in the far lower clustering coefficient in the identified network than in the geographic network (.026 versus .194). With the recovered social interactions matrix we establish, beyond geography, what covariates correlate to the existence of ties between states and the strength of those ties. We identify non-adjacent states that influence tax setting and, more broadly, we establish that some states – such as Delaware, a well known low-tax state – are especially focal in driving tax setting in other jurisdictions. Finally, we conduct simulations to assess the equilibrium propagation of tax setting shocks. We compare the general equilibrium effects obtained using our identified network to what would have been obtained assuming the postulated geographic network of interactions. We show significant differences between the two scenarios, both for the equilibrium level of state taxes, and the dispersion of taxes across states.

We use all these results to shed new light on the two main hypotheses for social interactions in tax setting across jurisdictions: factor mobility and yardstick competition.[10]

We contribute to a nascent literature on the identification of social interactions models when W_0 is not entirely known to the researcher. Closely related to our work, Blume *et al.* (2015) investigate the case when W_0 is *partially* observed. Specifically, Blume et al. (2015, Theorem 6) shows that if two individuals are *known* to not be directly connected, the parameters of interest in a model related to (1) can be identified. An alternative approach is taken in Blume et al. (2011,

[10]See Tiebout (1956), Shleifer (1985) and Besley and Case (1994).

Theorem 7): they suggest 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 (as in Blume *et al.*, 2015) or placing additional structure on W_0 (as in Blume et al. (2011)) is complementary to our approach as it reduces the number of parameters in W_0 to be retrieved.

Rose (2015) also presents related identification results for a linear model much like (1). Assuming sparsity of the neighborhood structure, Rose (2015) offers identification conditions under rank restrictions on sub-matrices of the reduced form coefficient matrix from a regression of outcomes (y_{it}) on covariates (x_{it}). 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, Theorem 6) who show identification results can be leveraged if individuals are *known* not to be connected. Our analysis shows identification does not require sparsity, and relies on plausible and intuitive conditions, whereas the auxiliary rank conditions necessary in Rose (2015) may be computationally complex to verify.

Another paper related to our work is Breza *et al.* (2017). Motivated by the concern that social network data is costly to collect, Breza *et al.* (2017) propose an alternative approach: to collect aggregated relational data instead (i.e. responses to questions of the form, 'How many of your social network have trait k?') They estimate a specific statistical link-formation model using the aggregate data to obtain Bayesian estimates for the probability distribution of the network and related statistics. We instead develop results on the identification of connection structures directly relevant to outcomes, and our results do not rely on a particular probabilistic link-formation model. However, aggregate relational data on outcome-relevant connections, or more generally, partial information on networks, may provide further restrictions to facilitate identification and estimation in our setup.

Our conclusions discuss how our approach can be modified, and assumptions weakened, to integrate in partial knowledge of W_0. We also discuss a broader agenda that considers the formation of social ties, and the next steps required to simultaneously identify models of network formation and the structure of social interactions.

The paper is organized as follows. Section 2 presents the sufficient conditions under which the social interactions matrix, endogenous and exogenous social effects are globally identified. Section 3 describes the high-dimensional estimation techniques we use for estimation, based on the Adaptive Elastic Net GMM method. Section 4 presents simulation results from stylized and real-world social networks. Section 5 presents the application of our methods to tax competition between US states. Section 6 concludes. The Appendix provides proofs, and further details related to estimation and the simulations.

2 Identification

2.1 Setup

Consider a researcher with access to panel data covering $i = 1, \ldots, N$ individuals repeatedly observed over $t = 1, \ldots, T$ instances. The aim is to use this panel data to identify a social interactions model, with no data on actual social ties being available. For expositional ease, we begin by considering identification in a slightly simpler version of the canonical model described in (1), where we adopt the subscript "0" to denote parameters generating the data, and non-subscripted parameters are generic values in the parameter space:

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

Hence for now, we drop individual-specific (α_i) and time-constant fixed effects (α_t), and assume x_{it} is a one-dimensional regressor for individual i and instance t. Later in this Section we extend the analysis to include individual-specific and time-constant fixed effects. We also later allow for multidimensional covariates $x_{k,it}$, $k = 1, \ldots, K$. In the terminology of Manski (1993), ρ_0 and γ_0 capture endogenous and exogenous social effects.

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

$$y_t = \rho_0 W_0 y_t + \beta_0 x_t + \gamma_0 W_0 x_t + \epsilon_t. \tag{3}$$

The vector of outcomes $y_t = (y_{1t}, \ldots, y_{NT})'$ assembles the individual outcomes in instance t, and the vector x_t does the same with individual characteristics. Therefore, y_t, x_t and ϵ_t have dimension $N \times 1$, the social interactions matrix W_0 is $N \times N$, and ρ_0, β_0, and γ_0 are scalar parameters. We do not make any distributional assumptions on ϵ_t beyond $\mathbb{E}(\epsilon_t | x_t) = 0$. We assume the network structure is predetermined and fixed. We treat it as a parameter to be identified and estimated.[11]

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

$$y_t = \Pi_0 x_t + \nu_t, \tag{4}$$

with $\Pi_0 = (I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0)$ and $\nu_t \equiv (I - \rho_0 W_0)^{-1} \epsilon_t$.

We wish to establish identification of the structural parameters of the model, including the social interactions matrix W_0, from the coefficients matrix Π_0. Without data on the network W_0,

[11] A related set of papers instead focusses on the distribution of networks generating the pattern in data and aims to estimate aggregate network effects. Souza (2014) offers several identification and estimation results in this spirit. In particular, he infers the network distribution within a certain class of statistical network formation models from outcome data from many groups, such as classrooms, in few time periods. We instead concentrate on estimating the set of links for one group followed over $t = 1, \ldots, T$ instances.

we treat it as an additional parameter, in this 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 in the group. These are summarized by the entries of the coefficient matrix Π_0, which, in turn, encode information about W_0 and $(\rho_0, \beta_0, \gamma_0)$.

We first set out five assumptions underpinning our main identification results. These Assumptions (A1-A5) deliver an identified set of up to two points. Three of these are entirely standard in the social interactions. A fourth is a normalization required to separately identify (ρ_0, γ_0) from W_0, and the fifth is closely related to known results on the identification of (ρ_0, γ_0) when W_0 is known (Bramoullé *et al.*, 2009).

Our first assumption explicitly states that no individual affects himself and is a standard condition in social interaction models:

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

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 $\theta = (W_{12}, \ldots, W_{N,N-1}, \rho, \gamma, \beta)' \in \mathbb{R}^m$, 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 $\Pi : \mathbb{R}^m \to \mathbb{R}^{N^2}$ define the relation between structural and reduced-form parameters:

$$\Pi(\theta) = (I - \rho W)^{-1} (\beta I + \gamma W), \tag{5}$$

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

As ϵ_t (and, consequently, ν_t) is mean-independent from x_t, $\mathbb{E}[\epsilon_t | x_t] = 0$, the matrix Π_0 can be consistently estimated as the linear projection of y_t on x_t.[12] 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).

We limit social interactions to a manageable level. This is closely related to the concept of stationarity in network models. Assumption (A2) controls the propagation of shocks and guarantees that they die as they reverberate through the network, which provides adequate stability in the system. It implies the maximum eigenvalue norm of $\rho_0 W_0$ is less than one. It also ensures $(I - \rho_0 W_0)$ is a non-singular matrix, and so 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) $\sum_{j=1}^{N} |(W_0)_{ij}| \leq 1$ for every $i = 1, \ldots, N$ and $|\rho_0| < 1$.

[12] If x_t is endogenous, i.e. $\mathbb{E}[\epsilon_t | x_t] \neq 0$, a vector of valid and relevant instrumental variables z_t may still be used to identify Π_0. In either case, identification of Π_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) being full-rank.

We next assume that network effects do not cancel out, another standard assumption.

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

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

$$\Pi(\theta_0) = \beta_0 I + (\rho_0\beta_0 + \gamma_0)\sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k. \tag{6}$$

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

Identification of the social effects parameters (ρ_0, γ_0) requires that at least one row of W_0 adds to a fixed and known number. Otherwise, ρ_0 and γ_0 cannot be separately identified from W_0. Clearly, no such condition would be required if W_0 was observed.

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

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, ρ_0 and γ_0 represent the row-sum for individual i in W_y and W_x, respectively. In line with the literature, we maintain that the same W_0 governs the structure of both endogenous (W_y) and exogenous (W_x) effects. We later discuss relaxing this assumption when more than one regressor is used.

Our final assumption provides for a specific kind of network asymmetry. We require that if the diagonal of W_0^2 is not constant, identification is achieved for the parameters of the model.

(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 ι, where ι is the $N \times 1$ vector of ones.

In unweighted networks, the diagonal 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 that Π_0 does not pin down θ_0. The strengthening of this hypothesis is the formal price to pay for the social interactions matrix W_0 being unknown to the researcher.[13]

[13]To see the strength of the assumption of Bramoullé *et al.* (2009) when W_0 is known, choose constants c_1, c_2,

2.2 Main Identification Results

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

$$\begin{aligned}
\Pi_0 v_{0,j} &= \beta_0 v_{0,j} + (\rho_0\beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k v_{0,j} \\
&= \left[\beta_0 + (\rho_0\beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} \lambda_{0,j}^k \right] v_{0,j} \\
&= \frac{\beta_0 + \gamma_0\lambda_{0,j}}{1 - \rho_0\lambda_{0,j}} v_{0,j}. \qquad (7)
\end{aligned}$$

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 Π_0 with associated eigenvalue $\lambda_{\Pi,j} = \frac{\beta_0+\gamma_0\lambda_{0,j}}{1-\rho_0\lambda_{0,j}}$. The fact that eigenvectors of Π_0 are also eigenvectors of W_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.[14]

Now let $\Theta \equiv \{\theta \in \mathbb{R}^m : \text{Assumptions (A1)-(A5) are satisfied}\}$ be the structural parameter space of interest. Our first theorem establishes local identification of the mapping. A parameter point θ_0 is locally identifiable if there exists a neighborhood of θ_0 containing no other θ which is observationally equivalent. Using classical results in Rothenberg (1971), we show our assumptions are sufficient to ensure that the Jacobian of Π relative to θ is non-singular, which, in turn, suffices to establish local identification.

---

and c_3 such that $c_1 I + c_2 W_0 + c_3 W_0^2 = 0$. Focusing on 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 that if Assumptions A1-A5 do not hold, one can construct examples where Π_0 does not pin down θ_0. Take, for instance, $N = 5$ with θ_0 and θ where $\beta = \beta_0 = 1$, $\rho = 1.5$, $\rho_0 = 0.5$, $\gamma = -2.5$, $\gamma_0 = 0.5$,

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

Both W and W_0 violate (A5) ($(W^2)_{kk} = (W_0^2)_{kk} = 0.5$ for any k), and ρ 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 $\Pi = (I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0) = (I - \rho W)^{-1}(\beta I + \gamma W)$. This arises 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*.

[14] 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).

Theorem 1. *Assume (A1)-(A5).* $\theta_0 \in \Theta$ *is locally identifiable.*

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 Π 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)-(A5). Then* $\Pi(\cdot)$ *is a proper mapping. Moreover, the set* $\{\theta : \Pi(\theta) = \Pi(\theta_0)\}$ *is finite.*

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 $\Theta_+ \equiv \Theta \cap \{\rho\beta + \gamma > 0\}$.[15] 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)-(A5), then for every* $\theta \in \Theta_+$ *we have* $\Pi(\theta) = \Pi(\theta_0) \Rightarrow \theta = \theta_0$. *That is,* θ_0 *is globally identified with respect to the set* Θ_+.

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)-(A5). 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.*

We now turn our attention to the problem of identifying the sign of $\rho_0\beta_0 + \gamma_0$ from the observation of Π_0. This would then allow us to establish global identification using Theorem 2. It is apparent from (6) that if $\rho_0 > 0$ and $(W_0)_{ij} \geq 0$, for all $i, j = \{1, \ldots, N\}$ the off-diagonal elements of Π_0 identify the sign of $\rho_0\beta_0 + \gamma_0$.

Corollary 3. *Assume (A1)-(A5). If* $\rho_0 > 0$ *and* $(W_0)_{ij} \geq 0$, *the model is globally identified.*

Empirical applications often suggest endogenous social interactions are positive (so that $\rho_0 > 0$), in which case global identification is fully established by Corollary 3. On the other hand, if $\rho_0 < 0$ (which would be the case if outcomes of interest were strategic substitutes, for example), ρ_0^k in (6) 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., permutable into a block-triangular matrix or, equivalently, a strongly connected social network), the model is also identifiable without further restrictions on ρ_0:

[15] The global inversion results we use are related to, but different from, those used by Komunjer (2012), Lee and Lewbel (2013) and Chiappori *et al.* (2015). Those authors use variations on a classical inversion result of Hadamard. In contrast, we employ results on the cardinality of the pre-image of a function, relying on less stringent assumptions. Specifically, while the classical Hadamard result requires that the image of the function be simply-connected (Theorem 6.2.8 of Krantz and Parks, 2013), the results we rely on do not.

Corollary 4. *Assume (A1)-(A5),* $(W_0)_{ij} \geq 0$ *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 holds if there are at least two real eigenvalues, or if ρ_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 re-write 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. 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, including in the empirical application on tax competition considered in Section 5.

2.3 Extensions

2.3.1 Individual Fixed Effects

We observe outcomes for $i = 1, \ldots, N$ individuals repeatedly through $t = 1, \ldots, T$ instances. If t corresponds to time, it is natural to think of there being unobserved heterogeneity across individuals needing to be accounted for when estimating Π_0. The structural model (2) is then,

$$y_{it} = \rho_0 \sum_{j=1}^{N} W_{0,ij} y_{jt} + \beta_0 x_{it} + \gamma_0 \sum_{j=1}^{N} W_{0,ij} x_{jt} + \alpha_i + \epsilon_{it}, \tag{8}$$

which can be written in matrix form as,

$$y_t = \rho_0 W_0 y_t + x_t \beta_0 + W_0 x_t \gamma_0 + \alpha^* + \epsilon_t, \tag{9}$$

where α^* 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 $\bar{\epsilon}_t = T^{-1} \sum_{t=1}^{T} \epsilon_t$,

$$y_t - \bar{y}_t = \rho_0 W_0 \left(y_t - \bar{y}_t\right) + \left(x_t - \bar{x}_t\right) \beta_0 + W_0 \left(x_t - \bar{x}_t\right) \gamma_0 + \epsilon_t - \bar{\epsilon}_t, \tag{10}$$

if W_0 is does not change with time. Identification from the reduced form follows from previous theorems, since Π_0 is unchanged when regressing $y_t - \bar{y}_t$ on $x_t - \bar{x}_t$.

2.3.2 Common Shocks

We now include unobserved variables common to all individuals in the network. These correlated effects α_t capture the intuition that some unobservable might change outcomes of all individuals in the same instance t, confounding the identification of social interactions. We have not placed any distributional assumption on the covariance matrix of the disturbance term. Hence, our analysis

readily incorporates correlated effects that are orthogonal to x_t. When this is not the case, one possibility is to model the corrected effects α_t explicitly. The model then becomes,

$$y_t = \rho_0 W_0 y_t + x_t \beta_0 + \gamma_0 W_0 x_t + \alpha_t \iota + \epsilon_t, \tag{11}$$

where α_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,

$$y_t = \Pi_0 x_t + \alpha_t \Pi_{01} \iota + v_t. \tag{12}$$

We propose a transformation to eliminate the correlated effects: exclude the individual-invariant α_t, subtracting the mean of the variables at a given period (global differencing). For this purpose, define $H = \frac{1}{n}\iota\iota'$. We next 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 (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 $i = 1, ...N$ we have that $\sum_{j=1,...,N}(W_0)_{ij} = 1$.

This can be written compactly as $W_0\iota = \iota$. Under row-sum normalization we then have that,

$$\begin{aligned}(I - H)\, y_t &= (I - H)\,(I - \rho_0 W_0)^{-1}\,(\beta_0 I + \gamma_0 W_0)\, x_t + (I - H)\,(I - \rho_0 W_0)^{-1}\,\epsilon_t \\ &= (I - H)\,\Pi_0 x_t + (I - H)\, v_t,\end{aligned} \tag{13}$$

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, Π_0 is identified from $\tilde{\Pi}_0$ (and, as a consequence, the previous results immediately apply).

Proposition 1. *Under row-sum normalization of* W_0, Π_0 *is identified from* $\tilde{\Pi}_0$.

Under row-sum normalization of W_0, a common group-level shock affects individuals homogeneously since $(I - \rho_0 W_0)^{-1}\alpha_t\iota = \alpha_t(I + \rho_0 W_0 + \rho_0^2 W_0^2 + \cdots)\iota = \frac{\alpha_t}{1-\rho_0}\iota$, which is a vector with no variation across entries. Consequently, global differencing eliminates correlated effects and $(I - H)\,(I - \rho_0 W_0)^{-1}\,\alpha_t \iota = (I - \rho_0 W_0)^{-1}\,\alpha_t\,(I - H)\,\iota = 0$. In the absence of row-sum normalization, global differencing does not ensure that correlated effects are eliminated. To see this, note that $(I - \rho_0 W_0)^{-1}$ is no longer row-sum normalized and, crucially, $\alpha_t(I - \rho_0 W_0)^{-1}\iota$ is not a vector with 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* $\Theta = \{\theta \in \mathbb{R}^m : \textit{Assumptions (A1)-(A5) are satisfied}\}$ *there are* N *matrices* $W_0^{(1)}, \ldots, W_0^{(N)}$ *such that* $[r_{W_0^{(1)}} \ \cdots \ r_{W_0^{(N)}}]$ *has rank* N, *then the only transformation such that* $(I - \tilde{H})(I - \rho_0 W_0)^{-1}\iota = 0$ *is* $\tilde{H} = I$.

2.3.3 Testing Row-sum Normalization

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

$$\begin{aligned}
\Pi_0\iota &= \beta_0\iota + (\rho_0\beta_0 + \gamma_0)\sum_{k=1}^{\infty}\rho_0^{k-1}W_0^k\iota \\
&= \left[\beta_0 + (\rho_0\beta_0 + \gamma_0)\sum_{k=1}^{\infty}\rho_0^{k-1}\right]\iota \\
&= \frac{\beta_0 + \gamma_0}{1 - \rho_0}\iota. \qquad (14)
\end{aligned}$$

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 Π_0 has constant row-sums, which suggests row-sum normalization is testable. In the Appendix we derive the relevant Wald test statistic to do so.

2.3.4 Multivariate Covariates

The analysis naturally expands for multivariate x_t of dimension $n \times k$. In this case, the reduced-form model (4) is,

$$y_t = \sum_{k=1}^{K}\Pi_{0,k}x_{k,t} + \nu_t, \qquad (15)$$

where $\Pi_{0,k} = (I - \rho_0 W_0)^{-1}(\beta_{0,k} + \gamma_{0,k}W_0)$, $x_{k,t}$ 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 β_0 and γ_0, respectively. The previous identification results then apply sequentially to each $\Pi_{0,k}$, $k = 1, \ldots, K$. In fact, we only then need to maintain $W_x = \gamma_0 W_0$ 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,

$$y_t = \rho_0 W_0 y_t + \sum_{k=1}^{K}\beta_{0,k}x_t + \sum_{k=1}^{K}\gamma_{0,k}W_{0,k}x_{k,t} + \epsilon_t. \qquad (16)$$

Let $W_{0,k} = W_0$ be the case for $k = 1$. Then, having identified ρ_0 and W_0 from $\Pi_{1,0}$,

$$(I - \rho_0 W_0)\Pi_{0,k} = \beta_{0,k}I + \gamma_{0,k}W_{0,k}, \qquad (17)$$

for $k = 2, \ldots, K$. 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 $k = 1, \ldots, K$, we can identify $\gamma_{0,k}$ and thus $W_{0,k}$ for every $k = 1, \ldots, K$.[16]

3 Estimation

The parameter vector to be estimated is high-dimensional: $\theta = (W_{12}, \ldots, W_{N,N-1}, \rho, \gamma, \beta)' \in \mathbb{R}^m$, 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. OLS estimation requires $m \ll T$: that there are many more time periods than parameters. This requirement is unlikely to be met in most economic applications. Instead, to estimate a large number of parameters with limited data, we need to utilize high-dimensional estimation methods. Such estimation techniques are the focus of a rapidly growing literature (Fan *et al.*, 2011).

Sparsity is a key assumption underlying all high-dimensional techniques, i.e. a large number of true model parameters are zero, although their location is unknown – and impose some penalization. In the context of social interactions, W_0 is sparse if $\tilde{N}$, the number of non-zero elements of W_0, is such that $\tilde{N} \ll T$. The exact location of the non-zeros is unknown and it may be that $N > T$. Sparsity thus corresponds to assuming that individuals influence or are influenced by a small number of others, relative to the overall size of the network. As such, sparsity is typically not a binding constraint in social networks analysis.

There are for example, multiple sparse structures in stylized networks: (i) a star: all individuals receive spillovers from the same individual; (ii) lattice: each individual is a source of spillover only to one other individual; (iii) interactions in pairs or triads or small groups, such as those described by De Giorgi *et al.* (2010); and (iv) small world networks (Watts, 1999). Prominent real world networks used in economic analysis are also sparse. For example, in individual-level elicited data from *AddHealth* on teenage friendships (defined as reciprocated nominations), the density of links is around 2% of all feasible links. In firm-level data, the density of production networks in the US is less than 1% of all feasible links (Atalay *et al.*, 2011).

Our preferred method estimates the interaction matrix in the reduced form while penalizing and imposing sparsity on the structural object W_0.[17] To accomplish this, we make use of the Adaptive Elastic Net GMM, developed by Caner and Zhang (2014). This is a two-step estimator.

[16] 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).

[17] We choose to 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 Π_0. In Appendix B.1, we show that $[\Pi_0]_{ij} = 0$ if, and only if, there are no paths between i and j in W_0, and so the pair is not connected. So sparsity in Π_0 is understood as W_0 being "sparsely connected", which is a stronger assumption than sparsity in W_0.

The first step is the solution to,

$$\tilde{\theta}(p) = (1 + p_2/T) \cdot \underset{\theta \in \mathbb{R}^p}{\arg\min} \left\{ g_{NT}(\theta)' M_T g_{NT}(\theta) + p_1 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} |W_{i,j}| + p_2 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} |W_{i,j}|^2 \right\}, \tag{18}$$

where $\theta = (W_{1,2}, \ldots, W_{N,N-1}\, \rho, \gamma, \beta)'$ 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 a GMM objective function with moment conditions based on the orthogonality between the disturbance term and the covariates: $g_{NT}(\theta) = \sum_{t=1}^{T} [x_{1t} e_t(\theta)' \ \cdots \ x_{NT} e_t(\theta)']'$, $e_t(\theta) = y_t - (I - \rho W)^{-1} (\beta I + \gamma W) x_t$.[18] There are $q \equiv N^2$ moment conditions since x_{it} is orthogonal to e_{jt}, for each $i, j = 1, \ldots, N$. Hence the GMM weight matrix M_T is of dimension $N^2 \times N^2$, symmetric, and positive definite.[19]

The fact that the penalization terms p_1 and p_2 are greater than zero is what makes (18) 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_{ij}, i.e. the sum of the strength of links, for all node-pairs. The second penalization term, $p_2 \sum_{i,j=1, i \neq j}^{N} |W_{i,j}|^2$, penalizes for 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).

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 penalization also shrinks non-zero estimates towards zero. A second (adaptive) step provides large improvements by re-weighting the penalization by the inverse of the first-step estimates:[20]

$$\hat{\theta}(p) = (1 + p_2/T) \cdot \underset{\theta \in \mathbb{R}^p}{\arg\min} \left\{ g_{NT}(\theta)' M_T g_{NT}(\theta) + p_1^* \sum_{\substack{\{i,j: \tilde{W}_{ij} \neq 0, \\ i,j=1,\ldots,N, \\ i \neq j\}}} \frac{|W_{i,j}|}{|\tilde{W}_{i,j}|^{-\gamma}} + p_2 \sum_{\substack{\{i,j: \tilde{W}_{ij} \neq 0, \\ i,j=1,\ldots,N, \\ i \neq j\}}} |W_{i,j}|^2 \right\}, \tag{19}$$

where $\tilde{W}_{i,j}$ is the (i,j)-th element of the first-step estimate of W, and we follow Caner and Zhang (2014) to set $\gamma = 2.5$. Elements $\tilde{W}_{i,j}$ estimated as zeros in the first stage are kept as zero in the second stage, because $\tilde{W}_{i,j} = 0$ implies the effective penalization is infinite. We write $p = (p_1, p_1^*, p_2)$ as the final set of penalization parameters. Conditional on p, the final estimate is $\hat{\theta}(p)$.

As in Caner and Zhang (2014, p. 35), the penalization parameters p are chosen by the BIC

[18] For expositional ease, we describe estimation in the context of the reduced form model (4), thereby abstaining from individual fixed or correlated effects.

[19] For simplicity, we use $M_T = I_{N^2 \times N^2}$. Note that if x_t is endogenous, one can also exploit moment conditions with respect to available instrumental variables.

[20] This modification is similar to the Adaptive step in the Adaptive Lasso of Zou (2006).

criterion. This balances model fit with the number of parameters included in the model.[21]

In Appendix B.2 we provide further implementation details for the procedure, 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.[22] Indeed, in Appendix B.3 we show how OLS can also be used to estimate θ if T is sufficiently large. This makes precise the benefits of penalized estimation for any given T and highlights that sparsity is not required for our identification results. OLS estimates also form the basis for the Wald test statistic used to test Assumption (A4') on row-sum normalization.

4 Simulations

4.1 Set-Up

We now apply our method and demonstrate the properties of the Adaptive Elastic Net GMM estimator, using Monte Carlo simulations under different true network structures. The simulations are based on two stylized random network structures, and two real world networks. These four networks are chosen as they vary in their size (N), complexity, and their aggregate and node-level features. All four networks are also sparse. The two stylized networks considered are:

(i) Erdos-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 (Erdos and Renyi, 1960). Such a network could be observed in practice if connections are formed independently of one another.

(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 the number of members as the other. More specifically, we assume individuals $i = 1, \ldots, \frac{N}{3}$ are affiliated to Party A and are led by individual 1; individuals $i = \frac{N}{3} + 1, \ldots, N$ are affiliated to Party B and are 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

[21] More specifically, the choice of p, which we denote as $\hat{p}$, is the one that minimizes

$$\text{BIC}(p) = \log\left[g_{NT}\left(\hat{\theta}(p)\right)' M_T g_{NT}\left(\hat{\theta}(p)\right)\right] + A\left(\hat{\theta}(p)\right) \cdot \frac{\log NT}{NT}$$

where $A\left(\hat{\theta}(p)\right)$ counts the number of non-zero coefficients among $\{W_{1,2}, \ldots, W_{N,N-1}\}$.

[22] For example, Manresa (2016) also relies on a Lasso-related methodology but restricts ρ_0 to be zero and so ignores endogenous social effects.If instrumental variables are available, Lam and Souza (2016) propose estimating (1) directly using the Adaptive Lasso and exploiting sparsity of the estimated W_0. Gautier and Rose (2016) extend the (identification-robust) Self-Tuning Instrumental Variable estimator in Gautier and Tsybakov (2014).

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.

The two real world networks we consider were elicited by researchers in very different scenarios.

(*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 (hence the sparsity of the network). 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 include several about relationships with other households in the village. We use social ties based on family relations, and focus on village 10 that is comprised of $N = 77$ 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 in the same village.

For the stylized networks (*i*) and (*ii*), we first assess the performance of the estimator for a fixed network size, $N = 30$. In the Appendix we show how performance varies with alternative network sizes. We simulate the real-world networks (*iii*) and (*iv*) using non-isolated nodes in each (so $N = 70$ and 65 respectively). We exclude isolated nodes because they do not conform with row-sum normalization (as also assumed in Bramoullé *et al.*, 2009).

Our method identifies entries in W_0 and so naturally recovers links of varying strength. It has long been recognized that link strength might both play an important role in social interactions, the seminal example being the distinct roles of strong and weak ties in job search (Granovetter, 1973). Data limitations often force researchers to postulate some ties to be weaker than others (say, based on frequency of interaction or joint characteristics of the linked individuals). This is in sharp contrast to our approach, that identifies the continuous strength of ties, $W_{0,ij}$, where $W_{0,ij} > 0$ is interpreted as node (state) j influencing outcomes in node (state) i.

To establish the performance of the estimator in capturing variation in link strength, we proceed as follows for each network. First, for each node we randomly assign one of their links to have value $W_{0,ij} = .7$. As the underlying data generating process is assumed to allow for common time effects (α_t), we then set the weight on all other links from the node to be equal and such that row-sum normalization (A4') is complied with (hence the sum of values across all links equals

one).[23] As we consider larger networks, we typically expect them to have more non-zero 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 small parameter estimates shrink to zero.[24] Finally, to aid exposition, we set a threshold value for link strength to distinguish 'strong' and 'weak' links. A strong link is defined as one for which $W_{0,ij} > .3$ and a weak link then has $W_{0,ij} \leq .3$.

Summary statistics for each network are presented in Panel A of Table 1. Following Jackson *et al.* (2017), we consider the following network-wide statistics: number of edges, number of strong and weak edges, number of reciprocated edges, clustering coefficient, number of components, and the size of the maximal component. In addition, we report the standard deviation calculated across elements of the diagonal of W_0^2. If this is zero, then the diagonal of W_0^2 is either zero or proportional to the vector of ones, and Assumption A5 would not be satisfied. We can see that for each case this statistic is well above zero.

Following Jackson *et al.* (2017), we also consider the following node-level statistics: in- and out-degree distribution (mean and standard deviation), and the three most central individuals. Overall, we see that the four networks differ markedly in their complexity, as well as their size. The networks vary in the relative importance of strong and weak ties: the Erdos-Renyi network only has strong ties, 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.

Panel data for each of the four simulations is generated as,

$$y_t = (I - \rho_0 W_0)^{-1}(x_t\beta + W_0 x_t\gamma + \alpha_t\iota + \alpha^* + \epsilon_t), \tag{20}$$

where α_t is a (scalar) time effect and α^* 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 $T = \{25, 50, 75, 100, 125, 150\}$. 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 (ϵ_t) are simulated as standard Gaussian, both generated from $N(0_N, I_{N\times N})$ distributions. As described later, we conduct a series of robustness checks to evaluate the sensitivity of the simulations to alternative parameters choices.

For each combination of parameters, we conduct $1,000$ simulation runs. On the initial 50 runs, we choose penalization parameters p that minimize the BIC criteria on a grid. This is a computationally intensive procedure, since it requires running the optimization procedure described

[23] For example, if in a given row of W_0 there are two links, one will be randomly selected to be set to .7, and the other set to .3. If there are three links one is set to .7 and the other two set to each have weight .15 to maintain row-sum normalization, and so on. For the Erdos-Renyi network, there are thus only strong ties as each node has only one link to another node.

[24] Caner and Zhang (2014) state that "local to zero coefficients should be larger than $N^{-\frac{1}{2}}$ to be differentiated from zero."

in Section 3 as many times as the number of points in the grid for p.[25] To reduce the computational burden, we do so only in the initial 50 simulation runs and consider these simulation runs as a calibration of p. For the remaining 950 iterations, the penalization parameter p is set fixed at the median p computed over the calibration runs. This only worsens the performance of the estimator, since a sub-optimal p is chosen for the majority of the iterations.

4.2 Results

We evaluate the performance of the procedure for varying panel lengths $T = \{25, 50, 75, 100, 125, 150\}$, using the following metrics. First, we 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 network 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 $MAD(\hat{\Pi}) = \frac{1}{N(N-1)} \sum_{i,j,i\neq j} |\hat{\Pi}_{ij} - \Pi_{ij,0}|$. The closer these metrics are to zero, more of the elements in the true matrix are correctly estimated. Finally, we evaluate the performance of the procedure 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 "post-Elastic Net" estimates obtained after having estimated the social interactions matrix by the Elastic Net GMM procedure. We use peers-of-peers' covariates from the estimated matrix as instrumental variables.

Figure 1 shows the simulation results. Each Panel presents a particular metric as we vary T, for the simulated networks. For example, Panel A shows that for each network, the proportion of zero entries in W_0 correctly estimated as zeros is above 90% even when exploiting a small number of time periods ($T = 25$). The proportion approaches 100% as T grows. Conversely, Panel B shows the proportion of non-zeros entries estimated as non-zeros is also high for small T, being at least 85% across networks, even for $T = 25$, and this increases as T grows. As discussed above, the Adaptive Elastic Net estimator is better in recovering true zero entries because it is a well-known feature that shrinkage estimators tend to shrink small parameters to zero.

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 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 quickly in T (we do not report the bias in $\hat{\beta}$ since it is very close to zero for every T). The fact that biases are not zero is as expected, being analogous to well-known results for autoregressive time series models.

Figure 2 provides a visual representation of the simulated and actual networks under $T = 100$ time periods. The network size is set to $N = 30$ in the two stylized networks, and as described

[25] In our simulations, we set the penalization grid to $p_1 = [0, .025, .05, .10]$, $p_1^* = [0, .025, .05, .10]$ and $p_2 = [0, .025, .05, .10]$, resulting in $4^3 = 64$ points per run.

above, $N = 70$ for the high school network, and $N = 65$ for the village household network. In comparing the simulated and true network, Figure 2 distinguishes between three types of edges: 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 2 further distinguishes between strong and weak links: strong links are shown in solid edges ($W_{0,ij} > .3$), and weak links are shown as dashed edges.

Consider first Panel A of Figure 2, comparing the simulated and true Erdos-Renyi network. We find that all zero and all non-zero links are correctly estimated. Hence all links are recovered and no edges are added to the true network (all edges are in blue). For the political party network, Panel B shows that all strong edges are correctly estimated (it also highlights the party leader nodes). 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, larger T is needed to achieve similar performance as in the other simulated networks in terms of detecting weak links. Again, we never estimate any added edges (no edges are green).

The larger real-world networks are the most complex to recover. Panel C shows that in the high school network, although strong edges are all recovered, 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: strong edges are all recovered, and here the majority of weak edges are also recovered. A relatively small share of overall edges are added or missed.

Panel B of Table 1 allows us to compare 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. As Figure 2 showed, the random Erdos-Renyi network is perfectly recovered. For the political party network, the number of recovered edges is slightly lower than the true network (38 vs. 45). This is driven by weak edges: while all the strong edges are recovered (30 out of 30), not all the weak ones are (8 vs. 15). On node-level statistics, the mean of the in- and out-degree distributions are slightly lower in the recovered network, the clustering coefficient is exactly recovered, and all three nodes with the highest out-degree are correctly captured (nodes 1, 11 and 28). The party leaders are individuals 1 and 11, so these focal individuals are both correctly recovered.

Performance in the two real world networks is also encouraging. In the high school network, all strong edges are correctly recovered, as are the majority of weak edges. However, as already noted in Figure 2, because weak edges are not well estimated in the high school network, the average in- and out- degrees are smaller in the recovered network relative to the true network. We recover two

out of the three individuals with the highest out-degree (nodes 21 and 69). Finally, in the village network, all strong edges are recovered, the majority of weak edges are recovered, the clustering coefficients are similar across recovered and true networks (.134 vs. .141) and we recover two out of the three households with the highest out-degree (nodes 16, 35, and 57).

4.3 Robustness

Appendix Table A1 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. Appendix Table A2 conducts robustness checks on the sensitivity of the estimates to parameters choices. We consider true parameters $\rho_0 = \{.1, .3, .7, .9\}$, $\gamma_0 = \{.3, .7\}$, $\beta_0 = \{.0, .8\}$. 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 ρ_0, β_0, γ_0, and q. For $\beta_0 = 0$, performance is slightly worse. This is expected as the exogenous variation from x_t no longer affects y_t directly.

Finally, to show the gains from using the Adaptive Elastic Net GMM estimator, Appendix Table A3 shows simulation results using Adaptive Lasso estimates of the interaction matrix Π_0, so estimating and penalizing the reduced-form. The Adaptive Lasso estimator performs relatively worse: the mean absolute deviation between $\hat{W}$ and W_0 is often two to three times larger than the corresponding Adaptive Elastic Net estimates. Appendix Table A4 shows the performance of the procedure based on OLS estimates of Π_0. Given OLS requires $m \ll T$, 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.

Taken together, the simulation results suggest the Adaptive Elastic Net GMM estimator is preferred over Adaptive Lasso and OLS estimators. This procedure does well in recovering true network structures, and *a fortiori*, 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 strength of correlated shocks.

5 Application: Tax Competition Between US States

We apply our methods to a long-standing issue in public economics on tax setting behavior by governors of US states (Wilson, 1999). Since the seminal empirical studies in tax competition between jurisdictions (Case *et al.*, 1989; Case *et al.*, 1993), it has been well-recognized that the definition of competing 'neighbors' is the key empirical challenge, and theory cannot completely resolve the issue. Two primary mechanisms have been argued to drive the structure of interactions across jurisdictions in tax setting behavior: 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. Indeed, a body of evidence finds that tax bases are mobile in response to tax differentials.[26] 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)).

A second mechanism for social interactions in tax setting occurs through political economy channels. In particular, yardstick competition between jurisdictions is driven by voters making comparisons between states to learn about their own politician's quality.[27] Besley and Case (1995) formalize the idea by developing a model of tax-setting in which voters use taxes set by governors in neighboring states to infer the quality of their own governor. This generates informational externalities across jurisdictions, forcing incumbents into a form of yardstick competition, where their tax setting behavior is determined by what other incumbents are doing. Yardstick competition leads naturally to the postulated interactions matrix being 'political neighbors': other states that voters make comparisons to.

This application showcases the practical use of our approach to recover social interactions in a setting in which the number of nodes and time periods is relatively low: the data covers mainland US states, $N = 48$, for years 1962-2015, $T = 53$. In this application, we can contrast the identified interactions matrix W_0 with a natural null hypothesis that states are only influenced by their geographic neighbors, W_{geo}. This is a common postulated structure for interactions in much of the empirical literature on tax competition. Moreover, with the recovered social interactions matrix we can establish, beyond geography, what predicts the existence and strength of ties between states. Finally, relative to the postulated geographic network of interactions, we conduct simulations using our economic neighbors network to assess: (i) the equilibrium propagation of tax setting shocks from any given state to all mainland US states; (ii) the general equilibrium impacts of such tax setting shocks on the average level and inequality in tax rates, as well as other statistics for tax rates based on the network of social interactions. Taken together, this body of evidence allows us to provide novel insights related to the role of factor mobility and yardstick competition in driving tax setting behavior across US states.

[26] See Hines (1996), Devereux and Griffith (1998), and Kleven *et al.* (2013, 2014).
[27] See Shleifer (1985) and Besley and Case (1995).

5.1 Data and Empirical Specification

We denote state tax liabilities for state i in year t as τ_{it}, covering state taxes collected from real per capita income, sales and corporate taxes. We measure this using a series constructed from data published annually in the Statistical Abstract of the United States. Our constructed series covers mainland US states ($N = 48$) for years 1962-2015, ($T = 53$). Our analysis therefore extends the sample used by Besley and Case (1995), that runs from 1962-1988 ($T = 26$).[28] 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 his tax program. Their modeling framework then implies a standard social interactions specification for the tax setting behavior of state governors:

$$\Delta\tau_{it} = \rho \sum_{j=1}^{N} W_{0,ij}\Delta\tau_{jt} + \gamma \sum_{j=1}^{N} W_{0,ij}x_{jt} + \beta x_{it} + \alpha_i + \alpha_t + \epsilon_{it}. \tag{21}$$

This allows for tax setting behavior to be 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 the economic/demographic characteristics of neighbors ($\sum_{j=1}^{N} W_{0,ij}x_{jt}$); (iii) state i's characteristics (x_{it}), that include income per capita, the unemployment rate, and the proportion of young and elderly. All specifications include state fixed effects and time effects (α_i, α_t) to allow for time-invariant unobserved heterogeneity across states in tax changes, and for common (macroeconomic) shocks to all states. Appendix Table A5 presents descriptive statistics for the original Besley and Case (1995) sample, and for our extended sample.

Our approach identifies the structure of social interactions among 'economic neighbors', that we denote $\hat{W}_{econ}$. At various points, it will be useful to compare the recovered structure of economic neighbors to the often postulated structure based on geographic neighbors, that we denote W_{geo}, as shown in Figure 3A for mainland US states. Moreover, much of the earlier literature focuses on endogenous social effects and ignores exogenous social effects by setting $\gamma = 0$. Our method allows us to relax this constraint and thus estimate the full typology of social effects described by Manski (1993). This is important because only endogenous social effects lead to social multipliers. In the context of tax competition, such multipliers are crucial to identify as they can lead to a race-to-the-bottom or sub-optimal level of public goods provision.[29]

To facilitate comparison with the earlier literature, after estimating the neighborhood matrix, we address the potential endogeneity of neighbors' tax changes ($\Delta\tau_{jt}$) by instrumenting for them, and thus recover consistent estimates of the endogenous social effect parameter, ρ. Following Besley

[28] Besley and Case (1995) test their political agency model using a two equation set-up: (i) on gubernatorial re-election 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 1977-88; 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.

[29] See Brennan and Buchanan (1980), Wilson (1986) and Oates and Schwab (1988).

and Case (1995), we instrument for $\Delta\tau_{jt}$ using neighbors' lagged change in income per capita, and neighbors' lagged change in unemployment rate. These instruments are essentially in the spirit of using exogenous social effects to instrument for neighbor's tax changes. However, given that our method allows us to estimate exogenous social effects ($\gamma \neq 0$), these instruments will generally be weaker when estimating the full specification in (21). We thus also follow Bramoullé *et al.* (2009) and De Giorgi *et al.* (2010), and instrument neighbors' tax changes with neighbor-of-neighbor characteristics.

5.2 Preliminary Findings

Table 2 presents our preliminary findings and comparison to Besley and Case (1995). Column 1 shows OLS estimates of (21) where the postulated social interactions matrix is based on geographic neighbors, exogenous social effects are ignored so $\gamma = 0$ 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: $\widehat{\rho}_{OLS} = .375$. Column 2 shows this to be robust to instrumenting neighbors' tax changes using the instrument set proposed by Besley and Case (1995). $\widehat{\rho}_{2SLS}$ is more than double the magnitude of $\widehat{\rho}_{OLS}$ suggesting tax setting behaviors across jurisdictions are strategic complements, and OLS estimates are heavily downward-biased.

Columns 3 and 4 replicate both specifications over the longer sample period we construct for each state's tax liability. The evidence confirms Besley and Case's (1995) finding on social interactions to be robust in a longer sample period. We again note that $\widehat{\rho}_{2SLS}$ is more than double the magnitude of $\widehat{\rho}_{OLS}$. 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 61 cents. This is similar to the headline estimate of Besley and Case (1995).[30]

5.3 Endogenous and Exogenous Social Interactions (ρ and γ)

Given this solid foundation in our extended panel data set, we proceed to apply our methods to establish whether there is evidence of endogenous and exogenous social interactions in tax setting behavior. We first focus on the estimated endogenous and exogenous social interaction parameters, and in the next subsection we detail the identified social interactions matrix, $\hat{W}_{econ}$.

Column 1 of Table 3 shows the initial estimates obtained from the Adaptive Elastic Net procedure where $\gamma = 0$. Columns 2 and 3 show the resulting OLS and 2SLS estimates for ρ: $\widehat{\rho}_{2SLS} = .641 > \widehat{\rho}_{OLS} = .378 > 0$.[31] Columns 4 to 6 estimate the full model in (21), allowing for both endogenous and exogenous social effects. Columns 5 and 6 show the OLS and 2SLS estimates

[30] Nor is the magnitude very different from earlier work examining fiscal expenditure spillovers. For example, Case *et al.* (1989) find that US state government levels of per-capita expenditures are significantly impacted by the expenditures of their neighbors, with the size of the impact being that a one dollar increase in neighbors' expenditures leads to an increase in own-state expenditures by seventy cents.

[31] We report robust standard errors and so do not adjust them for the fact that $\hat{W}_{econ}$ is estimated.

of ρ are smaller, and less precisely estimated when exogenous social effects are allowed. This is not surprising given that the instrument set is based on neighbors' characteristics. Many of these characteristics are directly controlled for in (21), thus reducing the effective variation induced by the instrument in neighbors' taxes. Hence, in Column 7, we report 2SLS estimates based on the alternative instruments using neighbor-of-neighbor characteristics. This represents our preferred specification: we find $\widehat{\rho}_{2SLS} = .608$ (with a standard error of .220). This value also accords with the requirements on ρ in Corollaries 3 and 4, needed for global identification.

In short, there is robust evidence of endogenous social interactions in tax setting behavior of governors across mainland US states. Appendix Table A6 shows the full set of exogenous social effects (so Columns 1 to 4 in Table A6 refer to the same specifications as Columns 4 to 7 in Table 3). We find there are exogenous social effects across states operating through economic neighbors' income per capita and unemployment rate. Demographic characteristics of economic neighbors to state i do not impact its tax rate.

5.4 Identified Social Interactions Matrix ($\hat{W}_{econ}$)

We now describe the identified social interactions matrix, $\hat{W}_{econ}$ from estimating the full model in (21). Figure 3B provides a graphical representation of how the structure 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 (either state i causally impacts state taxes in j, and/or *vice versa*). This comparison makes it clear whether all states geographically adjacent to i matter for its tax setting behavior and whether there are relevant non-adjacent states that influence its tax rate.

The left-hand panel of Figure 3B shows the network of geographic neighbors (whose edges are colored blue), onto which we have superimposed the edges that are *not* identified as links in W_{econ}; these dropped edges are indicated in red. This first implies that not all geographically adjacent states are relevant for tax setting behavior.

The right-hand panel of Figure 3B adds new edges identified in $\hat{W}_{econ}$ that are *not* part of W_{geo}. These represent non-adjacent states through which social interactions occur. This implies the existence of spatially dispersed social interactions between states.

As Table 4 summarizes, W_{geo} has 214 edges, while $\hat{W}_{econ}$ has only 144 edges. States are less connected than implied by postulating geographic networks. $\hat{W}_{econ}$ and W_{geo} have 79 edges in common. However, W_{geo} has 135 edges that are absent in $\hat{W}_{econ}$. Hence, while geography remains a key determinant of tax competition, the majority of geographical neighbors ($135/214 = 63\%$) are not relevant for tax setting. There are 65 edges that exist only in W_{geo}, so although there are fewer edges in $\hat{W}_{econ}$, 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} (.026 versus .194).[32]

[32]The clustering coefficient is the frequency of the number of fully connected triplets over the total number of

The broad implication is that in the context of tax setting, economic distance is imperfectly measured if we simply assume that interactions depend only on geographical distance. As detailed below, this has many implications for the economics of tax competition.

5.5 Strength of Ties and Reciprocity

As demonstrated in the simulations, our method naturally identifies the continuous strength of ties, $W_{0,ij}$, where $W_{0,ij} > 0$ is interpreted as node (state) j influencing outcomes in node (state) i. This allows us to assess whether links are reciprocal: namely whether $W_{0,ij} > 0$ and $W_{0,ji} > 0$, and asymmetries in link strength. This is useful because recent developments in tax competition theory, using insights from the social networks literature, suggest links need not be reciprocal or of symmetric strength.[33]

Figure 4A shows the distribution of $W_{0,ij}$'s across edges in $\hat{W}_{econ}$ (conditional on $W_{0,ij} > 0$). The strength of ties between pairs of states varies greatly. The mean strength of ties is .19, that is far higher than the median strength, .085, suggesting there are many weak ties. At the other end of the distribution, the strongest 10% of ties have weight above .6.

On the reciprocity of ties, Table 4 reveals that only 29.2% of edges in $\hat{W}_{econ}$ are reciprocal (all edges in W_{geo} are reciprocal by construction). Hence, tax competition is both spatially disperse and highly asymmetric. In most cases where tax setting in state i is influenced by taxes in state j, the opposite is not true.

Panels B and C in Figure 4 illustrate this for California, indicating the strength of each tie ($\hat{W}_{econ,CA,j}$). Figure 4B shows the in-network for California: those states that influence tax setting in California. Some geographic neighbors to California influence its tax setting behavior (Nevada and Oregon), although the strength of these ties is weak. On the other hand, non-adjacent states influence California (Colorado, Maine), and these in-network ties are stronger than the geographically adjacent in-network ties. Figure 4C shows the out-network for California, again indicating each tie strength ($\hat{W}_{econ,i,CA}$): those states whose taxes are influenced by taxes in California. We see that none of the geographic neighbors to California are influenced by its tax setting behavior, whereas a number of non-adjacent states are influenced by California (including East Coast states such as Virginia, and Southern states, such as Louisiana). When states are influenced by taxes in California, these ties tend to be relatively strong ties: $\hat{W}_{econ,i,CA} \geqslant .19$ for all five in-network ties.

Of course, focusing in on the ties of a given state is only illustrative. In Section 5.7 we provide

triplets. Other metrics can also be used to provide a scalar comparison of W_{geo} and $\hat{W}_{econ}$. One way to do so is to reshape both matrices as vectors of length (48×47) and to compute their correlation. Doing so, we obtain a correlation coefficient of .322.

[33] Janeba and Osterleh (2013) present theory and evidence on asymmetries in social interactions in tax setting across jurisdictions. Their model is based on an assumption of two levels of tax competition. The first level involves local competition among geographically close neighbors. The second idea is that large/populous jurisdictions (cities) compete with other cities of which some are geographically distant. Results of a survey of political decision-makers (mayors of 1108 cities and municipalities) in the German state of Baden-Wï¿œrttemberg support their theory.

systematic evidence on the characteristics of states that predict the existence and strength of ties between them, considering the sample of all $N \times (N-1) = 48 \times 47 = 2256$ potential links that could have formed between mainland states.

To be clear, given the common time shocks α_t in (21), 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^* impacts it, so that $W_{0,ij^*} > 0$. 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. Examining this possibility directly in $\hat{W}_{econ}$, we see this occurs for five states: Minnesota, New Jersey, New Mexico, Vermont, and Wisconsin. These states have an out-degree of zero. Their tax rate appear to impact no other states.

Table 4 reports the degree distribution across all nodes (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, relative to the network of geographic neighbors, is picked up again in the degree distribution being lower for $\hat{W}_{econ}$ than for W_{geo}. In $\hat{W}_{econ}$ the dispersion of in- and out-degree networks is very different (as measured by the standard deviation), being near double for the in-degree. This asymmetry in $\hat{W}_{econ}$ further suggests that some highly focal or influential states drive tax setting behavior in other states.

Figures 5A and 5B show complete histograms for the in- and out-degree across states. The histogram on the left is for in-degree, and shows that states under $\hat{W}_{econ}$ generally have lower in-degree than under W_{geo}. The states that are influenced by the highest number of other states are Utah, Pennsylvania and Ohio. The histogram on the right for out-degree, shows the five states described above that do not impact other states (Wisconsin, Vermont, New Mexico, New Jersey and Minnesota). Delaware is an outlier influential state in its out-degree in determining tax setting in other states: as discussed later, Delaware is a well-known potential tax haven.[34]

5.6 Counterfactuals

We can contrast how shocks to tax setting in a given state propagate through mainland US states under our estimated social interactions matrix $\hat{W}_{econ}$, relative to what would have been predicted under a postulated network structure based on geographic neighbors, W_{geo}. We illustrate this by considering a scenario in which California exogenously increases the change in its taxes per capita. We set $\Delta\tau_{it}$ to correspond to an increase of 10%. We then examine the propagation of this tax setting shock on other states. Figure 4C already hints at how such propagation might start: first through the impulse responses of the out-network of California, then through the impulse responses

[34]Dyreng *et al.* (2013) find that taxes play an important role in determining whether firms locate subsidiaries in Delaware: a Delaware-based state tax avoidance strategy lowers state effective tax rates by around 1 percentage point. They also report that in June 2010, Delaware landed at the top of *National Geographic* magazine's published list of the most secretive tax havens in the world (ahead of foreign tax havens such as Luxembourg, Switzerland, and the Cayman Islands).

of the out-network of those states and so forth. This immediately suggests that as $\hat{W}_{econ}$ is spatially more dispersed than W_{geo}, the general equilibrium effect on state taxes might be very different under both scenarios. We therefore derive not only the general equilibrium impacts of tax shocks, we also discuss the implications for tax inequality under $\hat{W}_{econ}$ and the W_{geo} counterfactual. We measure the differential change in equilibrium state taxes in state j under the two scenarios using the following statistic:

$$\Upsilon_j = \log(\Delta\tau_{jt}|\hat{W}_{econ}) - \log(\Delta\tau_{jt}|W_{geo}), \tag{22}$$

so that positive (negative) values imply taxes being higher (or lower) under $\hat{W}_{econ}$ than W_{geo}.[35]

Figure 6 graphs Υ_j for each mainland US state (including for California itself, the origin of the shock). A wide discrepancy between the equilibrium state tax rates predicted under $\hat{W}_{econ}$ relative to W_{geo}: across states Υ_j varies from -3.03 to 9.61. Only in one state is Υ_j close to zero. Appendix Table A7 summarizes the general equilibrium effects under both scenarios. We see that average tax rate increases are 74% higher under $\hat{W}_{econ}$. The dispersion of tax rates across states also increases dramatically under $\hat{W}_{econ}$ relative to W_{geo}. Finally, we note that assuming interactions across states are based solely on geographic neighbors, we miss the fact that many states will have relatively small tax increases.

5.7 Factor Mobility or Yardstick Competition?

We conclude by presenting two strategies to shed light on whether factor mobility and yardstick competition drive interactions across states in tax setting behavior: (i) exploiting information in the identified social interactions matrix $\hat{W}_{econ}$; (ii) following Besley and Case (1995), using gubernatorial re-election as an indirect test of the relevance of yardstick competition.

For our first strategy, we estimate the factors correlated with the existence/strength of links between states i and j in $\hat{W}_{econ}$ using the following dyadic regression specification:

$$\hat{W}_{econ,ij} = \lambda_0 + \lambda_1 X_{ij} + \lambda_2 X_i + \lambda_3 X_j + u_{ij}. \tag{23}$$

To begin, we discretize link strength so $\hat{W}_{econ,ij} \in \{0, 1\}$ and predict the existence of a link using a linear probability model. We then estimate the correlates of link strength $\hat{W}_{econ,ij} \in [0, 1]$ using a Tobit model. The elements X_{ij}, X_i, and X_j correspond to characteristics of the pair of states (i, j), of state i, and state j, respectively. Covariates are time-averaged over the sample period, and robust standard errors are reported. The sample thus corresponds to $N \times (N-1) = 48 \times 47 = 2256$ potential ij links that could have formed.

Table 5 presents the results. Column 1 controls only for whether states i and j are geographic neighbors. This is highly predictive of a link between them. Columns 2 and 3 show that distance

[35] We calculate the counterfactual at $\hat{\rho}_{2SLS} = .608$, the endogenous effect parameter estimated in our preferred specification, Column 7 of Table 3.

between states also negatively correlates with them being linked, but that when both geographic adjacency and distances are included, it is the former that is more predictive. Hence, we control only for whether i and j are geographic neighbors in the remaining Columns.

The next set of specifications use the insight from the literature that economic neighbors are likely to be based on a mixture of similarity in geography, income per capita, and demography (Case *et al.*, 1989). Column 4 thus 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 of 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 predicts ties, whereas demographic homophily does not.

Columns 5 to 7 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).[36] 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 period 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. This corroborates earlier evidence in Figure 5B, where Delaware, Wyoming and Nevada were among the states with the highest out-degree.

The specification in Column 7 shows that with this full set of controls, geographic adjacency remains a robust predictor of the existence of links between states. However, the identified economic network highlights additional significant predictors of tax competition between stares. In particular, political homophily *reduces* the likelihood of a link, suggesting that any element of yardstick competition driving social interactions occurs when voters compare their governor to governors of the opposing political party in other states. The tax haven states appear to be especially influential in the tax setting behaviors of other states. The strong influence of these tax haven states might be especially indicative of endogenous social effects in tax setting, and may lead to a race-to-the-bottom (Brennan and Buchanan, 1980; Wilson, 1986; Oates and Schwab, 1988). Relative to these factors, the economic and demographic similarity between states play an insignificant role in determining links between states.

The final column considers the continuous link strength as an outcome and reports Tobit partial

[36]We also experimented with alternative measures of labor migration, and 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, 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).

average effect estimates. This reinforces that geography, political homophily, and tax haven status all robustly correlate to the strength of influence states tax setting has on others. Labor mobility between states does not robustly predict either the existence or strength of ties.

Our second strategy to investigate factor mobility and yardstick competition follows the intuition of Besley and Case (1995). They suggest an indirect test of the relevance of yardstick competition is that this mechanism only applies to governors not facing term limits. Therefore we compare our main effects across two subsamples: state-years in which the governor can and cannot run for reelection. The results are reported in Table 6. The 2SLS results suggest that in both samples, endogenous social interaction effects exist, although they are more precisely estimated when governors *can* run for re-election.

Taken together, our evidence suggests that both factor mobility (of both labor and capital, as measured through the influence of tax havens), and yardstick competition (occurring through comparisons to governors of the other political party), are important mechanisms driving the existence and strength of social interactions in tax setting behavior across US states.

6 Conclusion

It is almost self-evident that social interactions can determine economic behavior and outcomes. Yet, information on social ties does not exist in most publicly available and widely used datasets. The literature has tried to overcome this challenge either through *postulating* ties based on some common observable, or *eliciting* ties. It is however increasingly recognized that both postulated and elicited networks remain imperfect solutions to the fundamental problem of missing data on social ties, either because of econometric concerns that arise through either method, or the cost of collecting primary data on networks. Two consequences are that: (i) classes of problems in which social interactions play a critical role remain understudied; (ii) there is no way to validate social interactions analysis in contexts where ties are postulated.

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. In short, our method allow the study of social interactions with social networks data, and the validation of social interactions analysis where social ties have hitherto only been postulated. In the context of a canonical social interactions model, we provide sufficient conditions under which the social interactions matrix, and endogenous and exogenous social effects are globally identified. Our identification strategy is novel, and may bear fruits in other areas (see for example the proof for Theorem 2). We describe how high-dimensional estimation techniques can be used to estimate the model based on the Adaptive Elastic Net GMM method. We showcase our method in Monte Carlo simulations using two stylized and two real world networks. Finally, we employ our method to provide novel insights on the study of tax competition across US states.

Our method is immediately applicable to other classic peer effects problems where large T panel data is readily available. For example, in finance, a long-standing question has been whether CEOs and other top executives are subject to relative performance evaluation, and if so, what is the comparison set of firms/CEOs used (Edmans and Gabaix, 2016).[37] Other fields such as macroeconomics, political economy and trade are all obvious areas in which long panel data sets exist for key outcomes, and social interactions across jurisdictions/countries etc. might be key determinants of them.

Three further directions for future research are worth highlighting. 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 methods can then still be applied to complete knowledge of W_0, and this could be achieved with potentially weaker assumptions for identification, and in shorter panels. To illustrate possibilities, Appendix Figure A1 shows results from a final simulation exercise in which we assume the researcher starts with partial knowledge of W_0. We do so for the Banerjee *et al.* (2013) village 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 six metrics generally improves for given T.

Second, we have developed our approach in the context of the canonical linear social interactions model (1). This allows to build closely 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, as established in Blume *et al.* (2011) and Blume *et al.* (2015), the reflection problem is functional-form dependent and may not apply to many non-linear models. An important topic for future research is thus to extend the insights gathered here to non-linear social interaction settings.

Finally, our approach has taken the network structure as predetermined and fixed. Clearly, an important part of the social networks literature examines endogenous network formation (Jackson *et al.*, 2017; de Paula, 2017). Our analysis allows us to only hint at the issue in two ways. First, in sufficiently long panels it would be possible to examine whether the identified social interactions matrix were stable across sample splits. Second, the kind of dyadic regression analysis in Section 5 on the correlates of entries in $W_{0,ij}$ suggests factors driving link formation and dissolution. A rich avenue for future work is to address the challenge of simultaneously identifying and estimating models of network formation and social interaction.

[37] Edmans and Gabaix (2016) overview the theory and empirics of executive compensation. Applying the informativeness principle in contract theory to CEO pay suggests peer performance is informative about the degree to which firm value is due to high CEO effort or luck. In a first generation of studies, Aggarwal and Samwick (1999) and Murphy (1999) showed that CEO pay is determined by absolute, rather than relative performance. However, this conclusion has been challenged by others such as Gong *et al.* (2011) who argue these conclusions arise from identifying relative performance evaluation (RPE) based on an implicit approach, assuming a peer group (e.g. based on industry and/or size). Indeed, when Gong *et al.* (2011) study the explicit use of RPE, based on the disclosure of peer firms and performance measures mandated by the SEC in 2006, they actually find that 25% of S&P 1500 firms explicitly using RPE. Our method could provide novel evidence on the matter.

References

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80, 1977–2016.

Aggarwal, R. K. and A. A. Samwick (1999). Executive Compensation, Strategic Competition, and Relative Performance Evaluation: theory and evidence. *The Journal of Finance*, 54.

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.

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. Chandrasekhar, E. Duflo, and M. Jackson (2016). Gossip: Identifying Central Individuals in a Social Network. MIT Working Paper.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The Diffusion of Microfinance. *Science*, 341, 1236498.

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.

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. Chandrasekhar, T. McCormick, and M. Pan (2017). Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data. Harvard University Working Paper.

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 Neighborrs. *NBER Working Paper 3032*.

Case, A., H. Rosen, and J. Hines (1993). Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States. *Journal of Public Economics*, 52, 285–307.

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.

Chiappori, P.-A., I. Komunjer, and D. Kristensen (2015). Nonparametric Identification and Estimation of Transformation Models. *Journal of Econometrics*, 188, 22–39.

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.

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 Marco, G., G. Gorni, and G. Zampieri (2014). Global Inversion of Functions: an Introduction. *ArXiv:1410.7902v1*.

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.

Devereux, M. and R. Griffith (1998). Taxes and the Location of Production: Evidence from a Panel of US Multinationals. *Journal of Public Economics*, 68, 335–67.

Dyreng, S., B. Lindsey, and J. Thornbock (2013). Exploring the Role Delaware Plays as a Domestic Tax Haven. *Journal of Financial Economics*, 108, 751–72.

Edmans, A. and X. Gabaix (2016). Executive Compensation: a modern primer. *Journal of Economic literature*, 54, 1232–87.

Erdos, P. and A. Renyi (1960). On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5, 17–60.

Fan, J., J. Lv, and L. Qi (2011). Sparse High Dimensional Models in Economics. *Annual Review of Economics*, 3, 291–317.

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.

Gong, G., L. Y. Li, and J. Y. Shin (2011). Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts. *The Accounting Review*, 86, 1007–43.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 6, 1360–80.

Harris, G. and C. Martin (1987). The Roots of a Polynomial Vary Continuously as a Function of the Coefficients. *Proceedings of the American Mathematical Society*, 100, 390–2.

Hines, J. (1996). Altered States: Taxes and the Location of Foreign Direct Investment in America. *American Economic Review*, 86, 1076–94.

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.

Janeba, E. and S. Osterleh (2013). Tax and the City – A Theory of Local Tax Competition. *Journal of Public Economics*, 106, 89–100.

Kennedy, J. and R. Eberhart (1995). Particle Swarm Optimization. *Proceedings of the IEEE International Conference on Neural Networks*, IV, 1942–8.

Kleven, H., C. Landais, and E. Saez (2013). Taxation and International Mobility of Superstars: Evidence from the European Football Market. *American Economic Review*, 103, 1892–924.

——— (2014). Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark. *Quarterly Journal of Economics*, 129, 333–78.

Komunjer, I. (2012). Global Identification in Nonlinear Models with Moment Restrictions. *Econometric Theory*, 28, 719–29.

Krantz, S. G. and H. R. Parks (2013). *The Implicit Function Theorem*, Birkhauser.

Lam, C. and P. C. Souza (2016). Sparse Adjustment for Spatial Weight Matrix with Instrumental Variables. .

Lee, L.-F. (2007). Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. *Journal of Econometrics*, 60, 531–42.

Lee, S. and A. Lewbel (2013). Nonparametric Identification of Accelerated Failure Time Competing Risks Models. *Econometric Theory*, 29, 905–19.

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.

Murphy, K. J. (1999). Executive Compensation. *Handbook of Labor Economics*, 3, 2485–563.

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.

Rothenberg, T. (1971). Identification in Parametric Models. *Econometrica*, 39, 577–91.

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.

Souza, P. C. (2014). Estimating Network Effects without Network Data. PUC-Rio Working Paper.

Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64, 416–24.

Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, 105, 493–527.

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.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101, 1418–29.

Zou, H. and H. H. Zhang (2009). On the Adaptive Elastic-net with a Diverging Number of Parameters. *Ann. Statist.*, 37, 1733–51.

A Proofs

Theorem 1

Proof. The local identification result follows Rothenberg (1971). We first demonstrate that the parameter space is open which is an maintained assuption (Assumption I) in that article. The spectral radius of a matrix is a continuous function of its entries[38] and we restrict the spectral radius to lie in an open set relative to the set of non-negative real numbers (which is the admissible range for the spectral radius of a matrix). Since under a continuous function the inverse image of an open set is also open, the set of ρ and W such that ρW has spectral radius stricly less than one is open. Consequently, since $\beta, \gamma \in \mathbb{R}$, the parameter space $\Theta \subset \mathbb{R}^m$ is an open set (recall that $m = N(N-1)+3$.)

We have that,

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

where Δ_{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,

[38] The eigenvalues of W vary continuously with its entries since they are the solution to the associated characteristic polynomial and the roots for a polynomial are a continuous function of its coefficients (see Harris and Martin, 1987). Consequently the spectral radius is a continuous function of the entries in W.

incorporated through the restriction that $\varphi \equiv \sum_{j=1,j\neq i}^{N} W_{ij} - 1 = 0$, for the unit-normalised row i. The derivative of the restriction φ is the m-dimensional vector $\nabla'_W \equiv \frac{\partial \varphi}{\partial \theta'} = \left[e'_i \otimes \iota'_{N-1}\ 0_{1\times 3}\right]$ (where e_i is an N-dimensional vector with 1 in the ith 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.[39]

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

$$c_{W_{12}}\frac{\partial \Pi}{\partial W_{12}} + \cdots + c_{W_{N,N-1}}\frac{\partial \Pi}{\partial W_{N,N-1}} + c_\rho \frac{\partial \Pi}{\partial \rho} + c_\gamma \frac{\partial \Pi}{\partial \gamma} + c_\beta \frac{\partial \Pi}{\partial \beta} \quad = \quad 0 \tag{24}$$

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

$$\sum_{j\neq i, j=1,\ldots,n} c_{W_{ij}} \quad = \quad 0. \tag{25}$$

Premultiplying equation (24) by $(I - \rho W)$ and substituting the derivatives,

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

Define $C \equiv \sum_{i,j=1,i\neq j}^{N} c_{W_{ij}} \Delta_{ij}$. Since the spectral radius of ρW 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

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

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

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

which, upon rearrangement, yields

$$\left(\gamma + \rho\beta\right) C + c_\beta I + \left(\beta c_\rho - c_\beta \rho + c_\gamma\right) W + \left(c_\rho \gamma - \rho c_\gamma\right) W^2 \quad = \quad 0. \tag{26}$$

Because $C_{ii} = 0$ and $W_{ii} = 0$ (by A1), we have that $c_\beta + \left(c_\rho \gamma - \rho c_\gamma\right)\left(W^2\right)_{ii} = 0$ for all $i = 1, \ldots, N$.

[39] For a parameter vector to be locally identified, Rothenberg (1971) requires that the derivative matrix ∇ 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 ∇ 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.

Since by assumption A5 there isn't a constant κ such that $\text{diag}\left(W_0^2\right) = \kappa\iota$, then $c_\beta = c_\rho\gamma - \rho c_\gamma = 0$. Plugging back in (26), 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

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

Note that, by equation (25), $\sum_{j\neq i, j=1,\ldots,n} c_{W_{ij}} = 0$. 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 $c_\rho\left(\rho\beta + \gamma\right) = 0$. 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. □

Corollary 1

Proof. The parameter θ_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 Π 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 Π is a continuous function of θ, the pre-image of a compact set, which is closed, is also closed. Because $\mathcal{W}$ is bounded and $\rho \in (-1, 1)$, their corresponding coordinates in $\theta \in \Pi^{-1}(\mathcal{A})$ are bounded. Suppose the coordinates for β or γ in $\theta \in \Pi^{-1}(\mathcal{A})$ are not bounded. So one can find a sequence $(\theta_k)_{k=1}^{\infty}$ such that $|\beta_k| \to \infty$ or $|\gamma_k| \to \infty$.

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

$$\|\beta I + \gamma W\| \;\; \leq \;\; \left\|(\beta I + \gamma W)\left(I - \rho W\right)^{-1}\right\| \cdot \|I - \rho W\|\,.$$

Note that $(I - \rho W)^{-1}$ and $(\beta I + \gamma W)$ commute, and so

$$\left\|(\beta I + \gamma W)\left(I - \rho W\right)^{-1}\right\| = \left\|(I - \rho W)^{-1}\left(\beta I + \gamma W\right)\right\| = \|\Pi\|\,.$$

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 $C \in \mathbb{R}$, since $\mathcal{W}$ is bounded by assumption A2. We then have that

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

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

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

The Lagrangean for the right-hand side minimization problem is:

$$\mathcal{L}\left(W_{i1}, \ldots, W_{i,i-1}, W_{i,i+1}, \ldots, W_{iN}; \mu\right) \;=\; \sum_{j=1}^{N} W_{ij}^2 - \mu \left(\sum_j W_{ij} - 1 \right).$$

where μ is the Lagrangean multiplier for the normalisation constraint. The first-order conditions for this convex minimization problem are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial W_{ij}} &= 2W_{ij} - \mu = 0, \quad \text{for any } j \neq i \\ \frac{\partial \mathcal{L}}{\partial \mu} &= \sum_{j=1}^{N} W_{ij} - 1 = 0. \end{aligned}$$

The first equation implies that $W_{ij} = \frac{\mu}{2}$ for $j = 1, \ldots, i-1, i+1, \ldots, N$. Using the fact that $W_{ii} = 0$, the second equation implies that $\mu = 2/(N-1)$. We have then that $W_{ij} = \frac{1}{N-1}, j \neq i$ and, consequently, $\|W\|^2 \geq (N-1)\frac{1}{(N-1)^2} = \frac{1}{N-1}$. Hence, if $|\gamma_k| \to \infty$, the numerator in the lower bound for $\|\Pi\|$ above also goes to infinity. Consequently, $\mathcal{A}$ would not be compact.

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

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

The following lemmas are used in proving Theorem 2.

Lemma 1. *Assume (A1)-(A4). If $\gamma_0 = 0$, then $\theta_0 \in \Theta$ is globally identified.*

Proof. With $\gamma = 0$, we have

$$\Pi \;=\; (I - \rho W)^{-1}\beta.$$

Take a candidate (W^*, ρ^*, β^*) possibly different from (W, ρ, β) such that the models are observationally equivalent. Then

$$\Pi = (I - \rho W)^{-1}\beta = (I - \rho^* W^*)^{-1}\beta^*.$$

Note that the equation above is equivalent to

$$b_1 I + b_2 W + b_3 W^* \;=\; 0$$

with $b_1 = \beta^* - \beta$, $b_2 = -\rho\beta^*$ and $b_3 = \rho^*\beta$. This implies that $b_1 = 0$ since $\text{diag}(W) = \text{diag}(W^*) = 0$. Because $b_2 = -\rho\beta^* = -\rho\beta \neq 0$ (by A3) and $b_3 = \rho^*\beta = \rho^*\beta^* \neq 0$ (also by A3), $b_2, b_3 \neq 0$ and we can write $W^* = -\frac{b_2}{b_3}W$. Since the ith row of W and W^* sums to one, we have that $\frac{b_2}{b_3} = -1$ which implies that $\rho^*/\rho = \beta/\beta^* = 1$ and, in turn, $W = W^*$. □

Lemma 2. *Assume (A1)-(A2) and (A4)-(A5). The image of $\Pi(\cdot)$, for $\theta \in \Theta_+$, is path-connected and, therefore, connected.*

Proof. Take θ and $\theta^* \in \Theta_+$. Consider first the subvectors corresponding to the adjacency matrices W and W^*. Without loss of generality, let $1, \ldots, N$ be ordered such that $(W^2)_{11} > (W^2)_{22}$. Consider the adjacency matrix W_* corresponding to the network of directed connections $\{(1,2),(2,1)\}$ and $\{(3,4),(4,5),\ldots,(N-1,N),(N,3)\}$:

$$W_* \;=\; \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}.$$

Note that $\text{diag}(W_*^2) = (1, 1, 0, \ldots, 0)$ and this is an admissible adjacency matrix under assumptions (A1)-(A2) and (A4)-(A5). 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 $i = 1, \ldots, N$ 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)-(A5) 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 directed connections $\{(1,3),(3,1)\}$ and $\{(2,4),(4,5),\ldots,(N-1,N),(N,2)\}$:

$$W_{**} = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Like $W(t)$ above, this path can be seen to satisfy assumptions (A1)-(A2) and (A4)-(A5). Now note that a path can also be constructed between W_* and W_{**} as their convex combination also satisfies (A1)-(A2) and (A4)-(A5). 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 θ^* and $\theta \in \Theta_+$. (Note also that $|\rho(t)| < 1$ so Assumption (A2) is satisfied.) These facts taken together imply that

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

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

Theorem 2

Proof. The proof uses Corollary 1.4 in Ambrosetti and Prodi (1995, p. 46),[40] 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* $\Pi^{-1}(\overline{\Pi})$ *is constant for any* $\overline{\Pi}$ *in the image of* $\Pi(\cdot)$.

The mapping $\Pi(\theta)$ is continuous and proper (by Corollary 1), with 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 $\gamma = 0$. By Lemma 1, that cardinality is one. □

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 (6), the off-diagonal elements of $\Pi(\theta)$ are equal to the off-diagonal elements of $(\rho\beta + \gamma)\sum_{k=1}^{\infty} \rho^{k-1} W^k$, the sign of those elements identifies the sign of $\rho\beta + \gamma$. By Theorem 2, the model is identified. □

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, λ_0, corresponds to an eigenvalue of Π_0 according to:

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

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

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

If the eigenvalue λ_0 is real, $b_0 = 0$ and the corresponding λ_{Π_0} eigenvalue is also real. Differentiating $Re(\lambda_{\Pi_0})$, the real part of λ_{Π_0}, with respect to $Re(\lambda_0) = a_0$, we get:

$$\frac{\partial Re(\lambda_{\Pi_0})}{\partial a_0} = \frac{(1 - \rho_0 a_0)^2 - \rho_0^2 b_0^2}{[(1 - \rho_0 a_0)^2 + \rho_0^2 b_0^2]^2} \times (\rho_0\beta_0 + \gamma_0). \tag{27}$$

[40] Related results can be found in Ambrosetti and Prodi (1972) and de Marco *et al.* (2014)

If the eigenvalue λ_0 is real, the expression (27) becomes:

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

The fraction multiplying $\rho_0\beta_0 + \gamma_0$ is positive. If $\rho_0\beta_0 + \gamma_0 < 0$, the real eigenvalues of Π_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 smallest real eigenvalue of Π_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 Π_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 λ_{Π_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 (27) is positive. The minimum value of the numerator subject to $|\lambda_0|^2 = a_0^2 + b_0^2 \leq 1$ is given by

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

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 μ 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^*, μ^*) 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 $\mu^* = 0$, ∂b_0 implies that $b_0^* = 0$. Then ∂a_0 would 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 ∂b_0 implies that $\mu^* = \rho_0^2$ and ∂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 $1/2 - \rho_0^2$. 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

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

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

If $\rho_0\beta_0 + \gamma_0 < 0$, the real part of the eigenvalues of Π_0 is decreasing on the real part of the eigenvalues of W_0. Consequently, the eigenvector corresponding to the eigenvalue of W_0 with the largest real part will correspond to the eigenvalue of Π_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 Π_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 λ_{Π_0} eigenvalue it is associated with.

By Theorem 2, the model is identified. □

Proposition 1

Proof. From equation (7) we observed that $\Pi_0 v_j = \lambda_{\Pi_0,j} v_j$, where v_j is an eigenvector of both W_0 and Π_0 with corresponding eigenvalue $\lambda_{\Pi_0,j} = \frac{\beta_0 + \gamma_0\lambda_{0,j}}{1-\rho_0\lambda_{0,j}}$. We also have that

$$\tilde{\Pi}_0 v_j \;=\; (I-H)\Pi_0 v_j \;=\; \lambda_{\Pi_0,j}(I-H)v_j \tag{28}$$

and so $\tilde{\Pi}_0$ and Π_0 have common eigenvalues, with corresponding eigenvector $\tilde{v}_j = v_j - \bar{v}_j\iota$, where $\bar{v}_j = \frac{1}{N}\iota' v_j$, $j = 1, \ldots, N$. Since $\lambda_{\Pi_0,j}$ and $\tilde{v}_j$ are observed from $\tilde{\Pi}_0$, identification of Π_0 is equivalent to identification of $\bar{v}_j$.

To establish identification of $\bar{v}_j$, note that $W_0(\tilde{v}_j + \bar{v}_j\iota) = \lambda_j(\tilde{v}_j + \bar{v}_j\iota)$ since v_j is an eigenvector of W_0. Consider an alternative constant $\bar{v}_j^* \neq \bar{v}_j$ that satisfies the previous equation. Then

$$W_0\iota(\bar{v}_j - \bar{v}_j^*) \;=\; \lambda(\bar{v}_j - \bar{v}_j^*). \tag{29}$$

Since $W_0\iota = \iota$, v_j must satisfy $(1-\lambda_j)(\bar{v}_j - \bar{v}_j^*) = 0$. For $j = 2, \ldots, N$, $|\lambda_j| < 1$. So $\bar{v}_j = \bar{v}_j^*$ and therefore identified. For $j = 1$, it is known that $\lambda_1 = 1$ with eigenvector $v_1 = \iota$. □

Proposition 2

Proof. Under row-sum normalization and $|\rho_0| < 1$, $(I - \rho_0 W_0)^{-1}\iota = \iota + \rho_0 W_0\iota + \rho_0^2 W_0^2\iota + \cdots = \iota + \rho_0\iota + \rho_0^2\iota + \cdots = \iota\frac{1}{1-\rho_0}$, so $\Pi_{01} \equiv (I - \rho_0 W)^{-1}$ has constant row-sums. If row-sum normalization fails, Π_{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 $h_{11}^* r_{W_0,1} - h_{12} r_{W_0,2} - \cdots - h_{1N} r_{W_0,N} = 0$ 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, $W_0^{(1)}, \ldots, W_0^{(n)}$,

such that $[r_{W_0^{(1)}} \cdots r_{W_0^{(N)}}]$ has rank N, then $h_{11}^* = h_{12} = \cdots = h_{1N} = 0$. Since the same reasoning applies to all rows, $\tilde{H}$ is the trivial transformation $\tilde{H} = I$. □

B Estimation

B.1 Sparsity of W_0 and Π_0

Define $\tilde{N}$ as the number of non-zero elements of Π_0. We say that Π_0 is sparse if $\tilde{N} \ll T$. Denote the number of connected pairs in W_0 via paths of any length as $\tilde{N}_c$. We equivalently say that W_0 is "sparse connected" if $\tilde{N}_c \ll T$. We show that sparsity of Π_0 is related to sparse connectedness of W_0.

Proposition 3. *Π_0 is sparse if, and only if, the number of unconnected pairs W_0 is small.*

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

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

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, sparsity of Π_0 translates into a large number of (i, j) unconnected pairs in W_0. □

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, the sparse connectedness implies sparsity and therefore is a stronger requirement. To see this, take any arbitrary network G with $\tilde{N}(G)$ non-zero elements and $\tilde{N}_c(G)$ connected pairs. Now consider the operation of "completing" G: for every connected (i, j) pair, add a direct link between (i, j) if non-existent in G and denote the resulting matrix as $\mathcal{C}(G)$. It is clear that $\tilde{N}(G) \leq \tilde{N}(\mathcal{C}(G))$. Yet, $\tilde{N}(\mathcal{C}(G)) = \tilde{N}_c(G)$.

B.2 Adaptive Elastic Net

Caner and Zhang (2014) show that (see their Theorems 3 and 4):

Proposition 4. *Under the following assumptions:*

(i) Define $\hat{G}_{NT}(\theta) = \frac{\partial g_{NT}(\theta)}{\partial \theta'}$. Then $\sup_{\theta \in B_m} \|\hat{G}_{NT}(\theta) - G(\theta)\| \xrightarrow{p} 0$ where $G(\theta)$ is a continuous function in θ with full column rank m, B_m is compact and individual components of θ are uniformly bounded by a constant a, $0 \leq a < \infty$.

(ii) $\left\| \left[E\left(g_{NT}(\theta^0) g_{NT}(\theta^0)' \right)^{-1} - \Omega^{-1} \right] \right\|_2^2 \to 0.$

(iii) M_T *is a symmetric and positive definite matrix such that* $\|M_T - M\|_2^2 \xrightarrow{p} 0$.

(iv) $m \equiv N(N-1)+3 = (NT)^{\alpha}$, $0 < \alpha < 1$, $p/(NT) = (N(N-1)+3)/(NT) \to 0$, $q/(NT) = N^2/(NT) = N/T \to 0$ *as* $NT \to \infty$.

(v) $\lambda_1/(NT) \to 0$, $\lambda_2/(NT) \to 0$, $\lambda_1^*/(NT) \to 0$, $\frac{\lambda_1^*}{(NT)^{3+\alpha}}(NT)^{\gamma(1-\alpha)} \to \infty$, $(NT)^{1-v}\eta^2 \to \infty$, $(NT)^{1-\alpha}\eta^{\gamma} \to \infty$ *for* $\eta = \min_{ij, W_{0,ij} \neq 0} |W_{0,ij}|$, $\eta = O(N^{-p})$, $0 < p < \alpha/2$ *and* $\gamma > \frac{2+\alpha}{1-\alpha}$.

(vi) $\max_t \frac{E\left\|g_t(\theta^0)\right\|_{2+l}^{2+l}}{\sqrt{NT}} \to 0$ *for* $l > 0$ *and* $g_t(\theta) = [x_{1t}e_t(\theta)' \ \cdots \ x_{NT}e_t(\theta)']'$

then, for compact Θ *and* $\theta_0 \in int(\Theta)$,

1. $P(\{j : \hat{\theta}_j \neq 0\} = \mathcal{A}) \to 1$, *where* $\mathcal{A} = \left\{j : \theta_j^0 \neq 0\right\}$ *and* $\theta^0 = (w_0', \rho_0, \beta_0, \gamma_0)'$.
2. *Let* $\hat{\theta}_{\mathcal{A}}$ *correspond to the estimator elements corresponding to the non-zero true parameters. Taking* $M_{NT} = \hat{\Omega}^{-1}$, $K_{NT}(nT)^{-\frac{1}{2}}(\hat{\theta}_{\mathcal{A}} - \theta_{\mathcal{A}}^0) \xrightarrow{d} N(0,1)$, $K_{NT} = \varphi'\left[\hat{G}_{NT}(\hat{\theta}_{\mathcal{A}})'\hat{\Omega}^{-1}\hat{G}_{NT}(\hat{\theta}_{\mathcal{A}})\right]^{\frac{1}{2}}$, *and* φ *is an unit-norm vector.*

Condition *(i)-(iii)* are standard in the literature. In particular, *(i)* is satisfied immediately by continuity of the Cayley transform with respect to θ and corresponds to assumption 1 in Caner and Zhang (2014). Condition Conditions (ii) and (iii) correspond to assumptions 3(i) and 2 in that paper, respectively. *(iv)* sets admissible rates of growth relating N and T and matches assumption 3(ii) in Caner and Zhang (2014). Since $m = N(N-1)+3$, $\frac{m}{NT} = \frac{N-1}{T} + \frac{3}{NT} \to 0$ implies that T grows faster than N (in this setting, condition $\frac{q}{NT} \to 0$ is redundant since $q = N^2$). Fan *et al.* (2011) denote this case as "relatively high-dimensional" since sample size is required to grow faster than the number of parameters.[41] Condition *(v)* gives the rate of growth of the different penalization factors and limits the absolute value of the minimum of the non-zero element of W_0. It appears as assumptions 5(i)-(v) and 3(iii) in Caner and Zhang (2014). It is a well-known feature that shrinkage estimators, such as the Adaptive Elastic Net, shrink small parameters to zero. As Caner and Zhang (2014, pp. 33) state, "local to zero coefficients should be larger than $N^{-\frac{1}{2}}$ to be differentiated from zero". Finally, condition *(vi)* is replicates assumption 4 in Caner and Zhang (2014) which is necessary to apply the triangular central limit theorem. The authors also show finite-sample properties of the estimators above.

B.2.1 Implementation and Initial Conditions

To make our procedure robust to the choice of initial condition, we use the particle swarm algorithm. This is an optimization algorithm tailored to more aptly find global optima, which does not depend on choice of initial conditions. It works as follows. The procedure starts from a large

[41] These are sufficient conditions to ensure convergence of the parameters of the Elastic Net GMM. While it might be possible to relax some of those assumptions, we are unaware of any such work for this particular estimator.

number of initial conditions covering the parameter space, known as "particles" (Kennedy and Eberhart, 1995). Each particle is iterated independently until convergence. The algorithm returns the optimum calculated across particles.[42,43]

To ensure compliance with row-sum normalization for each row i of W, one non-zero parameter W_{i,j^*} is set to $1 - \sum_{j=1,j\neq j^*}^{N} W_{i,j}$. This avoids making use of constrained optimization routines.[44] We also impose the restriction that $\rho \geq 0$ and $W_{ij} \geq 0$ by minimizing the objective function with respect to $\tilde{\rho}$ with $\rho = \tilde{\rho}^2$ and $\tilde{W}_{ij}$ with $W_{ij} = \tilde{W}_{ij}^2$.

Optimization of (18) starts from the initial condition selected by the particle swarm algorithm and is minimized with respect to the parameters that were neither set to zero nor were chosen to ensure row-sum normalization. Estimates from the first stage are subsequently used to adjust the penalization, as in the Adaptive Elastic Net GMM objective function (19).[45] The steps above are repeated for different combinations of $p = (p_1, p_1^*, p_2)$, selected on a grid. The final estimate is the one that minimizes the BIC criterion.

B.3 OLS

For the purpose of estimation, it is convenient to write the model in the stacked form. Let $x = [x_1, \ldots, x_T]'$ be the $T \times N$ matrix of explanatory variables, $y_i = [y_{i1}, \ldots, y_{iT}]'$ be the $T \times 1$ vector of response variables for individual i and $\pi_i^0 = [\pi_{i1}^0, \ldots, \pi_{iN}^0]'$ where π_{ij}^0 is a short notation for the (i,j)-th element of Π_0. The concise model is then,

$$y_i = x\pi_i^0 + v_i \tag{30}$$

[42]We set Caner and Zhang's (2014) suggestion for the initial condition as one of those particles, with minor modifications. The authors suggest calculating the absolute value of the derivative of the unpenalized GMM objective function evaluated at zero, ∇_W, and the set parameters smaller than p_1 at zero. The rationale is that if the GMM objective function is invariant with respect to certain parameters, the Elastic Net problem achieves a corner solution (where parameters are set to zero). In our case, allowing only for positive interactions, we set to zero the elements such that $-\nabla_W \leq p_1$. All other elements of W gain equal weights such that row-sum normalization is respected. The derivative ∇_W is mechanically zero if $\rho = \gamma = 0$. So we set $\rho = .5$, given that the parameter space is bounded and $\rho \in [0, 1)$. The other parameters that enter the derivative are $\hat{\beta}$ estimated from a regression of y on x, with the full set of fixed effects, and $\gamma = 0$.

[43]We also implemented an additional five particles. Particle 2: like Particle 1 but with size proportional to the magnitude of the derivatives conditional on $-\nabla_W$ being greater than p_1; Particle 3: sets to non-zero all positive elements of $-\nabla_W$ with equal weights; Particle 4: selects 5% highest values of $-\nabla_W$, sets all others to zero, and non-zero gain equal weights; Particle 5: W obtained from the Lasso regression of y_t on the y_t of others with penalization p_1; Particle 6: W obtained from the Lasso regression of y_t on the x_t of others with penalization p_1. In all cases, weights are rescaled by row-specific constants such that row-sum normalization is complied with. The remaining 94 particles are uniformly randomly selected by the built-in MATLAB particle swarm algorithm.

[44]At each row, we pick the j^* closest to the main diagonal of W.

[45]Note that the Elastic Net penalty $p_1 \sum |W_{i,j}|$ is invariant with respect to choices of W if row-sum normalization is imposed. Yet, the penalty affects the initial selection of arguments in which $W_{i,j}$ is restricted to zero if the derivative of the objective function is smaller than p_1 in absolute value.

for each $i = 1, \ldots, N$, where also $v_i = [v_{i1}, \ldots, v_{iT}]'$. Model (30) can then be estimated equation-by-equation. Denote $\pi^0 = [\pi_1^{0\prime}, \ldots, \pi_N^{0\prime}]'$. Stacking the full set of N equations,

$$y = X\pi^0 + v \tag{31}$$

where $y = [y_1, \ldots, y_N]$, $X = I_N \otimes x$, $\pi^0 = \text{vec}\,(\Pi_0')$, and $v = [v_1, \ldots, v_N]$. If the number of individuals in the network N is fixed and much smaller than data points available, $N^2 \ll NT$, equation (31) 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,

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

where $Q_T \equiv \frac{1}{NT}X'X$, $Q \equiv p\lim_{T\to\infty} Q_T$, $\Sigma_T \equiv \frac{1}{NT}X'vv'X$ and $\Sigma \equiv p\lim_{T\to\infty} \Sigma_T$. 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 (6),

$$\begin{aligned} \Pi_0 \iota_N &= \beta_0 \iota + (\rho_0\beta_0 + \gamma_0) \sum_{k=1}^{\infty} \rho_0^{k-1} W_0^k \iota \\ &= \frac{\beta_0 + \gamma_0}{1 - \rho_0} \iota \end{aligned} \tag{32}$$

where ι_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 (32) implies that Π 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 π^0. Under the null hypothesis,

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

where $R = [I_{N-1} \otimes \iota_N'; -\iota_{N-1} \otimes \iota_N']$. The Wald statistic is $W = NT\,(R\hat{\pi})'\,(Q^{-1}\Sigma Q^{-1})^{-1}\,(R\hat{\pi}) \sim \chi^2_{N-1}$ 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,

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

where ∇_θ is the gradient of $\hat{\theta}$ with respect to $\hat{\pi}$.

Table 1: True and Recovered Networks

	Erdos-Renyi	Political Party	High school *Coleman (1964)*	Village *Banerjee et al. (2013)*
		A. True Networks		
Number of nodes	**30**	**30**	**70**	**65**
(a) Network-wide statistics				
Number of edges	30	45	366	240
Number of strong edges	30	30	70	65
Number of weak edges	0	15	296	175
Number of reciprocated edges	2	2	184	240
Clustering coefficient	-	.000	.120	.141
Number of components	12	11	3	3
Size of maximal component	10	16	68	51
Standard deviation of the diagonal of squared W	.254	.254	.167	.239
(b) Node-level statistics				
In-degree distribution	1.00 (0.00)	1.50 (.509)	5.23 (2.04)	3.69 (2.35)
Out-degree distribution	1.00 (1.05)	1.50 (2.49)	5.23 (3.64)	3.69 (2.35)
Nodes with highest out-degree	{ 7, 11, 26 }	{ 1, 11 , 28 }	{ 21, 22, 69 }	{ 16, 35, 57 }
		B. Recovered Networks		
(a) Network-wide statistics				
Number of edges	30	38	210	194
Number of strong edges	30	30	70	68
Number of weak edges	0	8	140	126
Number of reciprocated edges	2	2	184	170
Clustering coefficient	-	.000	.162	.134
Number of components	12	11	1	4
Size of maximal component	10	14	70	48
(b) Node-level statistics				
In-degree distribution	1.00 (0.00)	1.27 (.450)	3.00 (1.18)	2.99 (1.29)
Out-degree distribution	1.00 (1.05)	1.27 (1.76)	3.00 (1.02)	2.99 (1.15)
Nodes with highest out-degree	{ 7, 11, 26 }	{ 1, 11, 28 }	{ 21, 48, 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 networks statistic are calculated over the average network across simulations with T=100.

Table 2: 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**	**(3) OLS**	**(4) 2SLS**
Geographic Neighbors' Tax Change (t - [t-2])	.375***	.868***	.271***	.642***
	(.120)	(.273)	(.075)	(.152)
Period	1962-1988	1962-1988	1962-2015	1962-2015
First Stage (F-stat)		6.267		27.320
Controls	Yes	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,296	1,248	2,592	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 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 neighbors' tax change is instrumented by lagged neighbor's state income per capita and unemployment rate. At the foot of Columns 2 and 4 we report the p-value on the F-statistic from the first stage of the null hypothesis that instruments are jointly equal to zero. 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 3: 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) Initial	(2) OLS	(3) 2SLS: IVs are Characteristics of Neighbors	(4) Initial	(5) OLS	(6) 2SLS: IVs are Characteristics of Neighbors	(7) 2SLS: IVs are Characteristics of Neighbors-of Neighbors
Economic Neighbors' Tax Change (t - [t-2])	.886	.378***	.641***	.645	.145**	.332*	.608***
		(.061)	(.060)		(.072)	(.199)	(.220)
Period		1962-2015				1962-2015	
First Stage (F-stat)			19.353			9.571	10.480
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,952	2,952	2,544	2,952	2,952	2,544	2,592

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 state i's total taxes per capita in year t. We allow for exogenous social effects in Columns 4 to 7. In subsequent OLS and IV regressions, the economic neighbors' effect is calculated as the weighted average of economic neighbors' variables. OLS regressions estimates are shown in Column 2, 3 and 5. Column 3 and 6 show the 2SLS regression where each geographic neighbors' tax change is instrumented by lagged neighbor's state income per capita and unemployment rate. Column 7 shows a 2SLS regression where each geographic neighbors' tax change is instrumented by lagged neighbor-of-neighbor's state income per capita and unemployment rate. At the foot of Columns 3, 6 and 7 we report the p-value on the F-statistic from the first stage of the null hypothesis that instruments are jointly equal to zero. 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 4: Geographic Versus Economic Neighbor Networks

	Geographic Neighbor Network	Economic Neighbor Network
Number of Edges	214	144
Edges in Both Networks	79	79
Edges in W-geo only	135	
Edges in W-econ only		65
Clustering	.1936	.0259
Reciprocated Edges	100%	29.17%
Degree Distribution Across Nodes (states)		
out-degree	4.458 (1.597)	3.000 (1.185)
in-degree		3.000 (2.073)

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 5: Predicting Links to Economic Neighbors

Columns 1-7: Linear Probability Model; Column 8: Tobit
Dependent variable (Cols 1-7): =1 if Economic Link Between States Identified
Dependent variable (Col 8): =Weighted Link Between States
Coefficient estimates, standard errors in parentheses

	Geography			Economic and Demographic Homophyly	Labor Mobility	Political Homophyly	Tax Havens	Tobit, Partial Avg Effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Geographic Neighbor	.699***		.701***	.701***	.698***	.698***	.697***	.068***
	(.030)		(.032)	(.030)	(.031)	(.031)	(.031)	(.006)
Distance		-.453***	-.008					
		(.033)	(.024)					
Distance sq.		.0949***	.003					
		(.007)	(.006)					
GDP Homophyly				2.409**	2.369*	2.296*	1.046	.322
				(1.183)	(1.186)	(1.193)	(1.150)	(.302)
Demographic Homophyly				.222	.235	.241	.256	.077
				(.226)	(.226)	(.228)	(.225)	(.067)
Net Migration					.044*	.044*	-0.032	0.001
					(.025)	(.025)	(.025)	(.002)
Political Homophyly						-.057	-.083**	-.025*
						(.042)	(.042)	(.014)
Tax Haven Sender							.107***	.021***
							(.024)	(.005)
Adjusted R-squared	0.427	0.152	0.427	0.428	0.429	0.429	0.440	-
Observations	2,256	2,256	2,256	2,256	2,256	2,256	2,256	2,256

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The specifications in Columns 1-7 are cross-sectional linear probabilities models where the dependent variable is equal to 1 if two states are linked, and zero otherwise. In Column 8 the dependent variable is the weighted link between states. Column 8 reports the partial average effects from a Tobit model. 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 homophyly is the absolute difference of states' GDP per capita. Demographic homophyly 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 based on individuals tax returns (Source: Internal Revenue Service, https://www.irs.gov/statistics/soi-tax-stats-migration-data). Political homophyly 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 in parentheses.

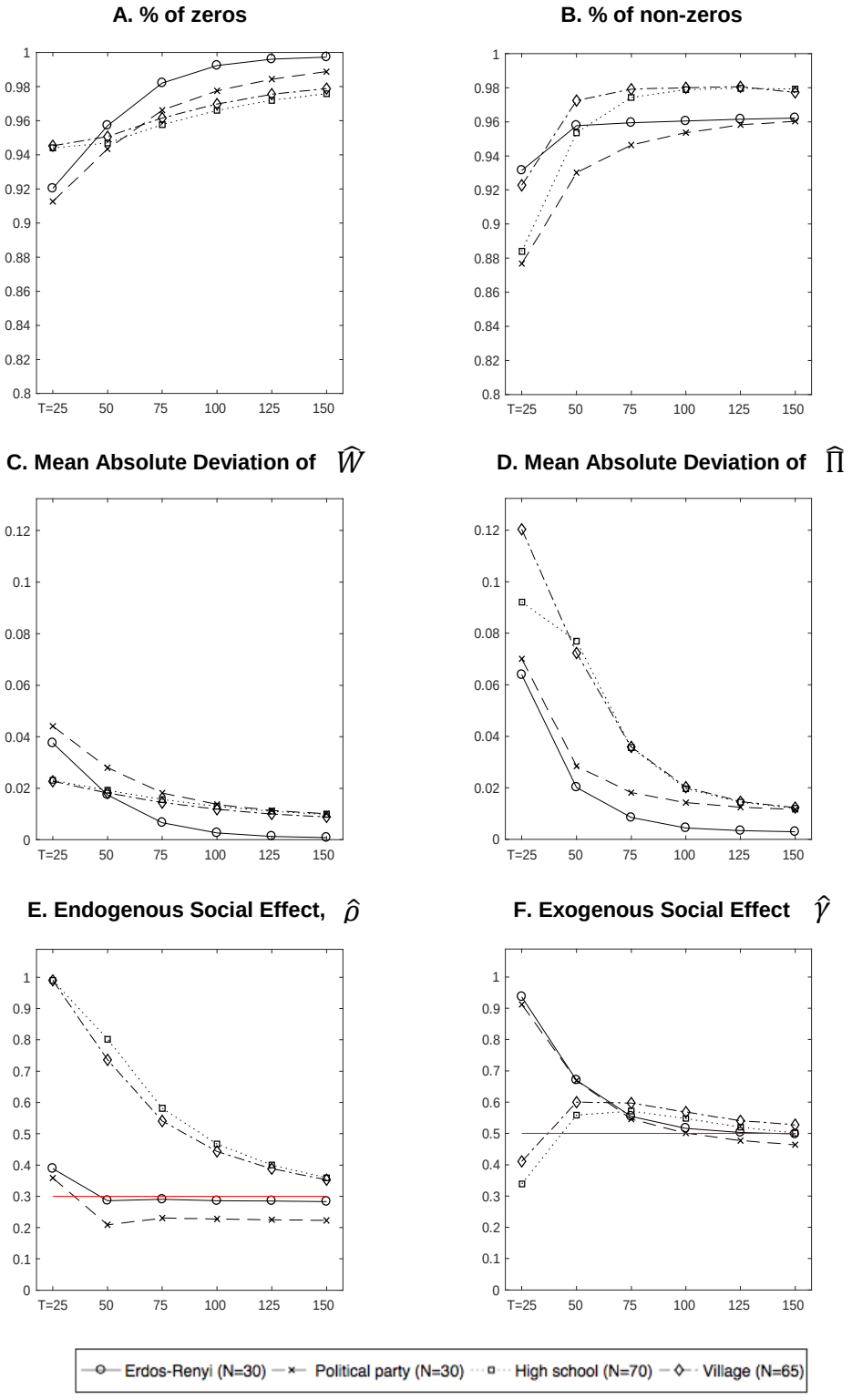
Table 6: Gubernatorial Term Limits

Dependent variable: Change in per capital income and corporate taxes
Coefficient estimates, standard errors in parentheses
IVs: Characteristics of Neighbors-of-Neighbors

	All Governors		Governor Cannot Run for Re-election		Governor Can Run for Re-election	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS
Economic Neighbors' tax change (t - [t-2])	.145**	.608***	.016	.937*	.182**	.543**
	(.072)	(.220)	(.105)	(.534)	(.084)	(.237)
Period	1962-2015		1962-2015		1962-2015	
First Stage (F-stat)		10.480		2.835		10.120
Controls	Yes	Yes	Yes	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,592	2,592	640	640	1,917	1,917

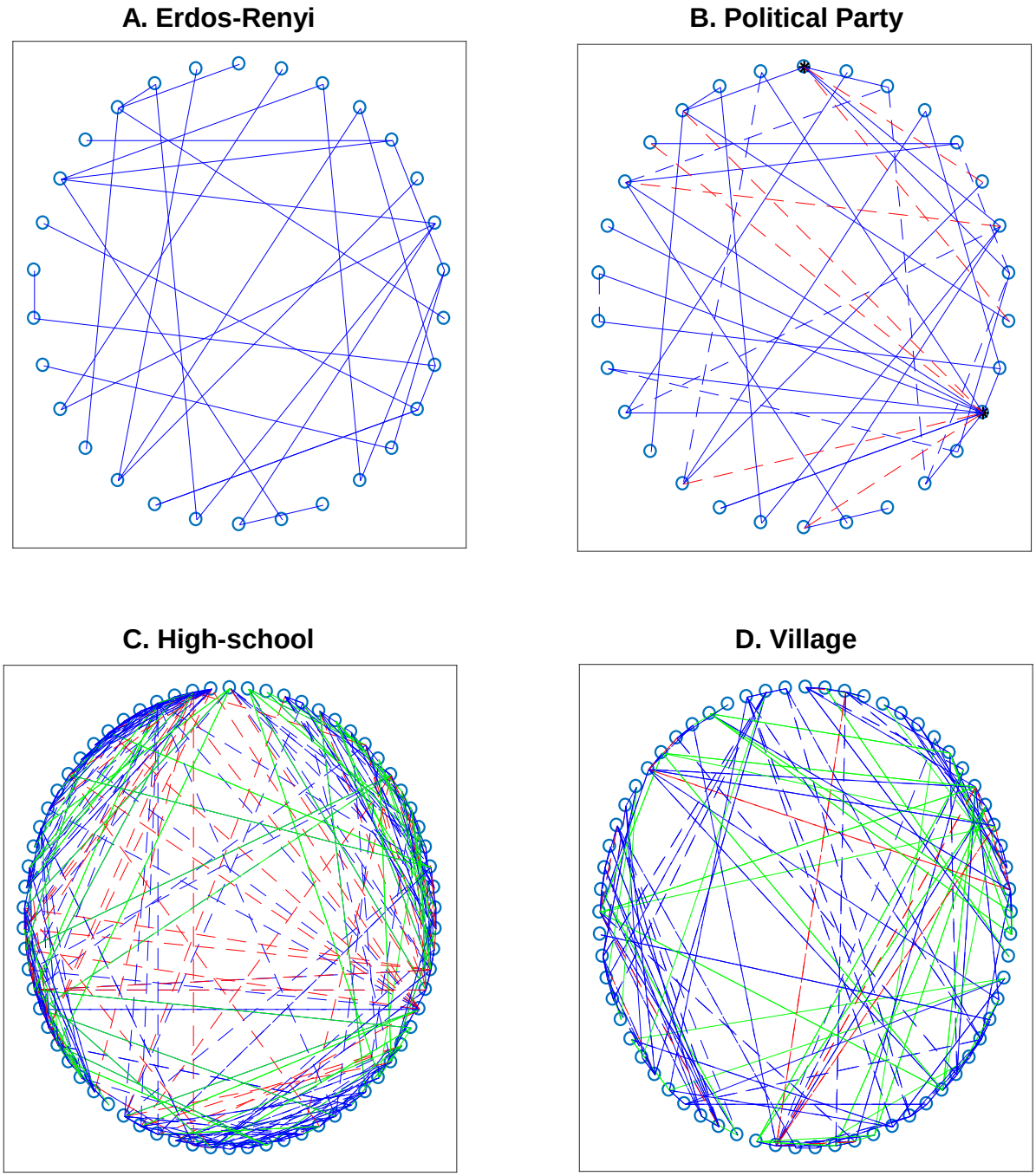
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample in Columns 1 and 2 covers 48 mainland US states running from 1962 to 2015. In Columns 3 and 4 we use the subsample of state-years in which the governor that faced term limits in the subsequent gubernatorial election. In Columns 5 and 6 we use the subsample of state-years in which the governor did not face term limits in the subsequent gubernatorial election, and so could run for reelection. The dependent variable is state i's total taxes per capita in year t. We first estimate our procedure which outputs parameters and the network of economic neighbors. We penalize geographic neighbors throughout and also allow for exogenous social effects. OLS regressions estimates are shown in Columns 1, 3 and 5. Columns 2, 4 and 6 show a 2SLS regression where each geographic neighbors' tax change is instrumented by lagged neighbor-of-neighbor's state income per capita and unemployment rate. At the foot of Columns 2, 4 and 6 we report the p-value on the F-statistic from the first stage of the null hypothesis that instruments are jointly equal to zero. 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.

Figure 1: Simulation Results, Adaptive Elastic Net GMM



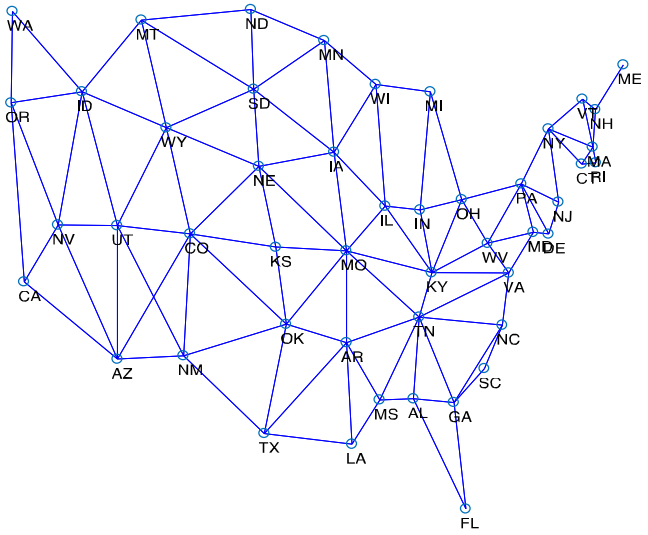
Notes: These simulation results are based on the Adaptive Elastic Net GMM algorithm, with penalization parameters chosen by BIC, under various true networks and time periods T=25, 50, 100, 125 and 150. In all cases, 1000 Monte Carlo iterations were performed. The true parameters are rho-0=.3, beta-0=.4 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 2: Simulated and True Networks



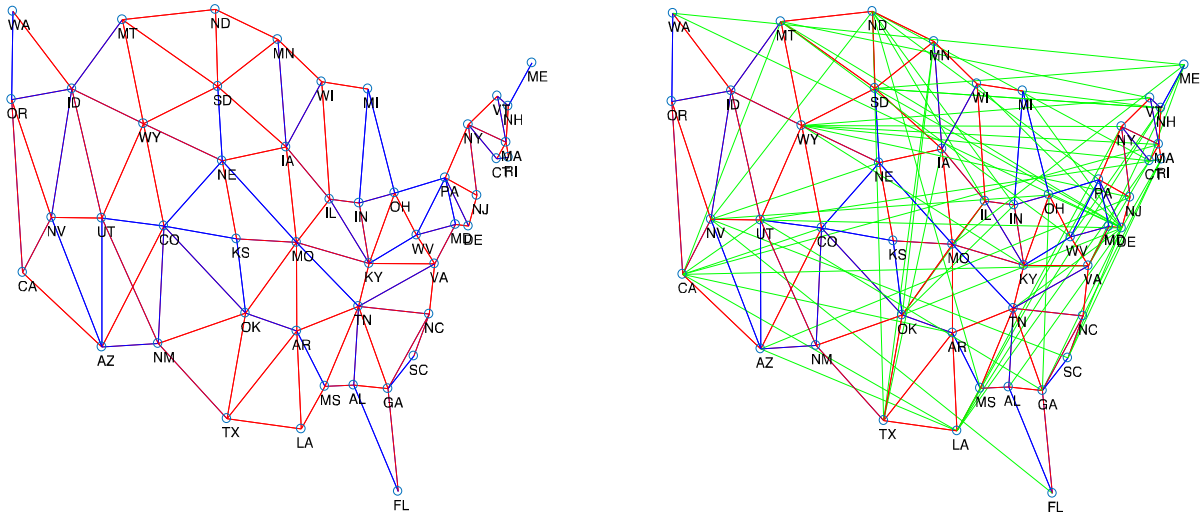
Notes: These simulation results are based on the Elastic Net algorithm, with penalization parameters chosen by BIC, under various true networks and time periods T=50, 100 and 150. In the two stylized networks (Erdos-Renyi and political party), we set N=30, and the real world networks, the high school 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 solid edges (whose strength is greater than or equal to .3), and weak links are shown as dashed edges.

Figure 3A: Network Graph of US States, Geographic Neighbors



Notes: Figure 3A 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).

Figure 3B: Network Graph of US States, Identified Economic Neighbors

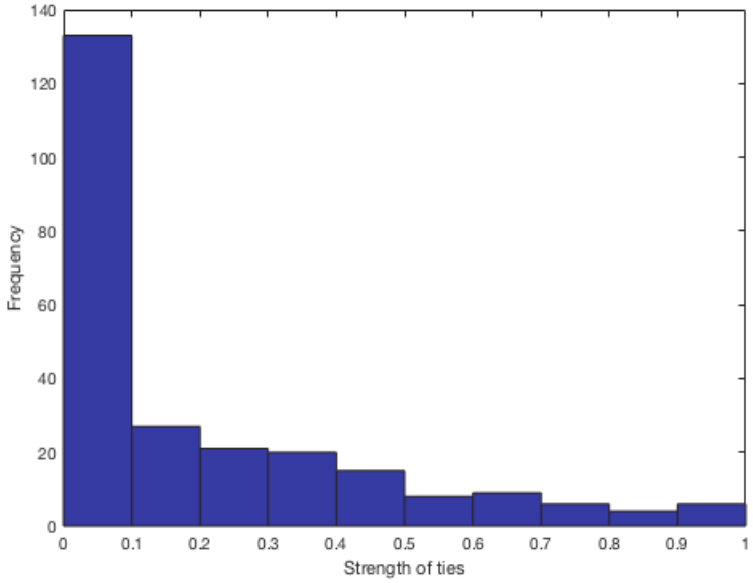


Geographic network edges
Removed (geographic) edges in economic network
New edges added in economic networks

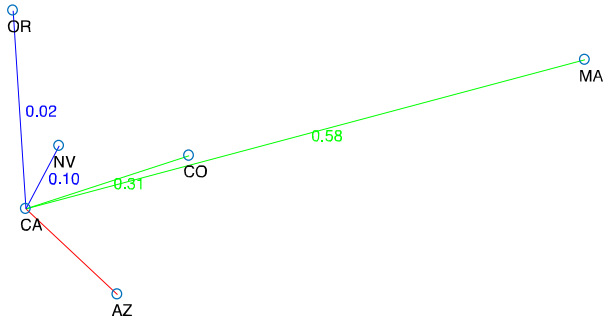
Notes: Figure 3B represents the continental United States (N=48). The economic network is derived from our preferred specification, where we penalize geographic neighbors to states, and allow for exogenous social effects. A blue edge is drawn between a pair of states if they are geographic neighbors 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 read 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 4: Strength of Ties and Reciprocity

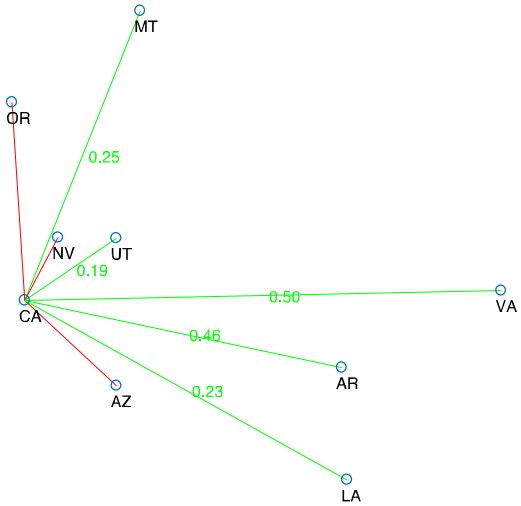
Panel A: Histogram of Strength of Ties, Conditional on $W_{0,ij}>0$



Panel B: In-network for California



Panel C: Out-network for California

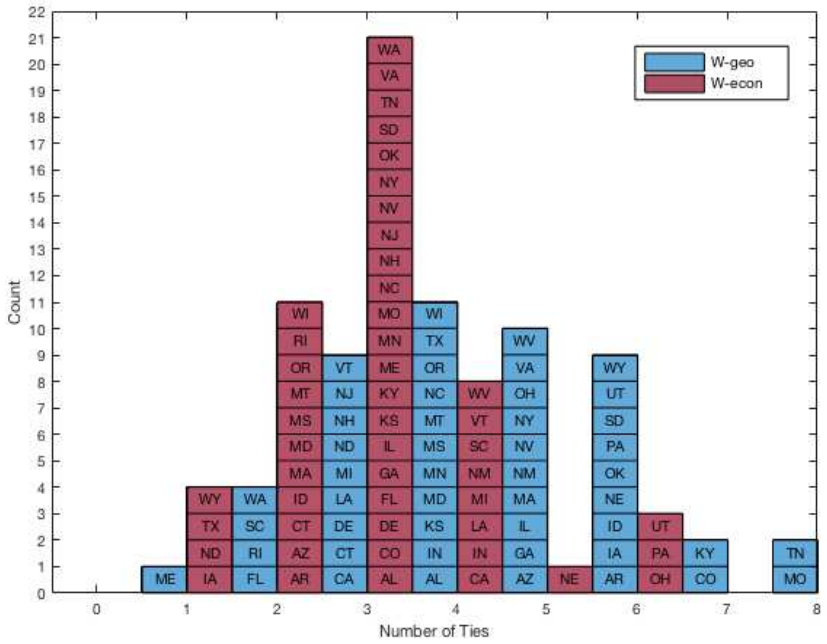


Geographic network edges
Removed (geographic) edges in economic network
New edges added in economic networks

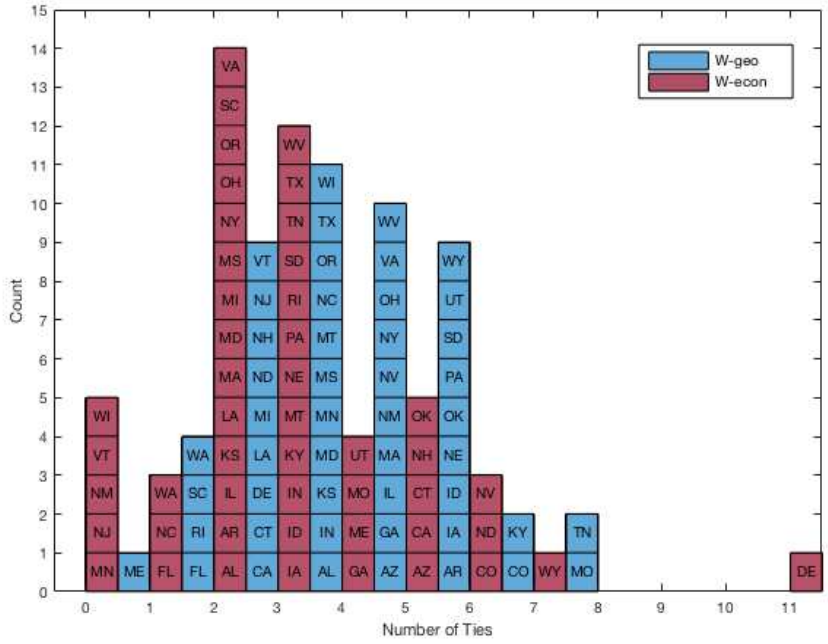
Notes: Panel A is the histogram of ties in the economic network, conditional on non-zero ties. Panels B and C show the in-network and out-network of California as derived from our preferred specification, where we penalize geographic neighbors to states, and allow for exogenous social effects. The in-network are the states that determine tax setting in California. The out-network is the states in which taxes are set in direct response to those in California. 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. State abbreviations are as used by US Post Office (http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf).

Figure 5: In- and Out-degree Distribution

Panel A: In-degree distribution



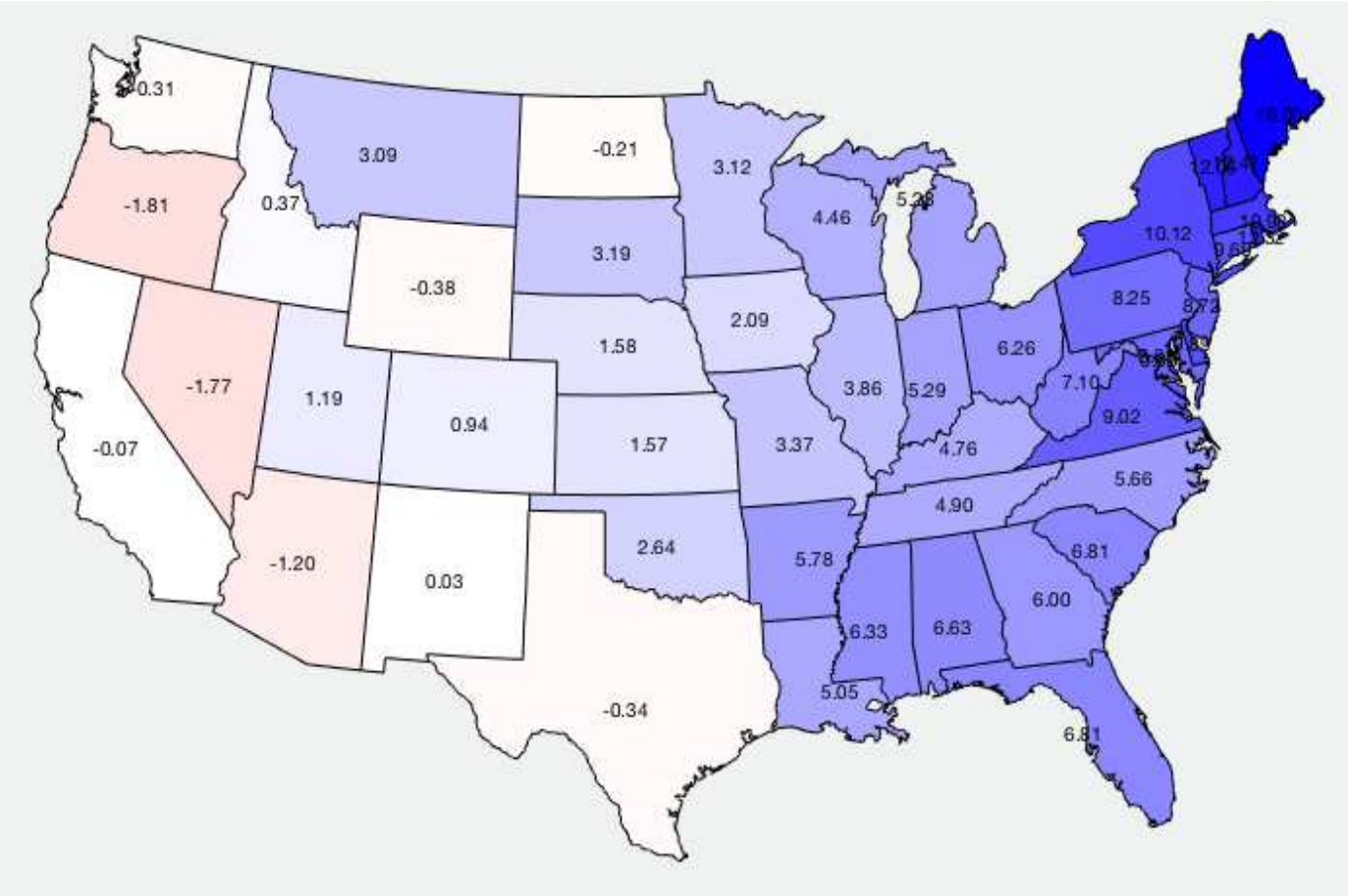
Panel B: Out-degree distribution



Notes: In-degree distribution (Panel A) and out-degree distribution (Panel B). Distribution calculated from geographic neighbors' network (W-geo) in blue. Distribution calculated from economic neighbor's network in (W-econ) in red. State abbreviations are as used by US Post Office (http://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf).

Figure 6: General Equilibrium Impacts of CA Tax Rise Shocks

State's Reaction to 10 p.p. increase in CA taxes



Log(equilibrium taxes under W-econ) - Log(equilibrium taxes under W-geo)

Positive values indicate higher equilibrium taxes under Economic neighbors than geographic neighbors

Negative values indicate low equilibrium taxes under Economic neighbors than geographic neighbors

Notes: This shows the equilibrium impulse responses in taxes set in each state as a result of California increasing its tax change by 10%. This is as derived from our preferred specification, where we penalize geographic neighbors to states, and allow for exogenous social effects. 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.

Table A1: Simulation Results, Adaptive Elastic Net GMM

	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
% True Zeroes	.954	.976	.973	.957	.992	.997	.947	.965	.979	.950	.977	.979	.943	.978	.989	.943	.961	.976
	(.033)	(.031)	(.031)	(.009)	(.004)	(.002)	(.007)	(.006)	(.005)	(.031)	(.025)	(.024)	(.009)	(.006)	(.005)	(.007)	(.006)	(.006)
% True Non-Zeroes	.899	.919	.924	.958	.960	.962	.977	.978	.977	.914	.925	.932	.930	.954	.960	.970	.977	.977
	(.051)	(.032)	(.026)	(.017)	(.016)	(.011)	(.007)	(.007)	(.008)	(.038)	(.022)	(.009)	(.037)	(.023)	(.016)	(.015)	(.009)	(.007)
$MAD(\widehat{W})$	.021	.007	.004	.017	.003	.001	.020	.011	.006	.032	.019	.017	.028	.014	.010	.023	.014	.009
	(.014)	(.011)	(.008)	(.004)	(.001)	(.001)	(.002)	(.001)	(.001)	(.011)	(.007)	(.004)	(.004)	(.002)	(.002)	(.002)	(.001)	(.001)
$MAD(\widehat{\Pi})$	.025	.012	.009	.020	.004	.003	.054	.016	.007	.034	.021	.019	.028	.014	.012	.065	.019	.009
	(.012)	(.008)	(.006)	(.004)	(.002)	(.001)	(.018)	(.004)	(.002)	(.009)	(.006)	(.003)	(.004)	(.002)	(.001)	(.028)	(.004)	(.001)
$\widehat{\rho}$	.262	.270	.276	.286	.286	.283	.667	.398	.270	.235	.245	.241	.209	.228	.223	.700	.383	.242
	(.069)	(.044)	(.038)	(.079)	(.026)	(.022)	(.079)	(.050)	(.050)	(.078)	(.049)	(.038)	(.084)	(.035)	(.029)	(.100)	(.069)	(.051)
$\widehat{\beta}$	.403	.399	.400	.405	.400	.400	.380	.399	.401	.405	.400	.397	.404	.399	.398	.380	.400	.401
	(.039)	(.028)	(.022)	(.028)	(.018)	(.015)	(.025)	(.015)	(.012)	(.040)	(.028)	(.022)	(.029)	(.019)	(.015)	(.024)	(.015)	(.012)
$\widehat{\gamma}$	.577	.521	.507	.670	.516	.498	.748	.686	.595	.550	.481	.459	.669	.501	.463	.686	.666	.566
	(.094)	(.059)	(.046)	(.054)	(.022)	(.018)	(.118)	(.056)	(.028)	(.093)	(.060)	(.050)	(.060)	(.027)	(.021)	(.128)	(.063)	(.032)

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 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 A2: Simulation Results, Adaptive Elastic Net GMM, Alternative Parameters

A. Erdos-Renyi

	ρ_0				β_0		γ_0		q			
	.1	**.5**	**.7**	**.9**	**.0**	**.8**	**.3**	**.7**	**.3**	**.5**	**.8**	**1.0**
% True Zeroes	.986	.994	.991	.979	.986	.987	.974	.997	.996	.997	.997	.997
	(.005)	(.004)	(.004)	(.005)	(.005)	(.007)	(.006)	(.002)	(.002)	(.003)	(.002)	(.002)
% True Non-Zeroes	.951	.963	.963	.956	.806	.967	.961	.961	.963	.962	.961	.959
	(.028)	(.011)	(.011)	(.017)	(.096)	.000	(.015)	(.016)	(.012)	(.013)	(.013)	(.018)
$MAD(\widehat{W})$	.005	.002	.003	.007	.014	.004	.011	.001	.001	.001	.000	.000
	(.002)	(.001)	(.001)	(.002)	(.005)	(.002)	(.003)	(.001)	(.001)	(.001)	(.001)	(.001)
$MAD(\widehat{\Pi})$	.005	.006	.013	.077	.013	.007	.009	.004	.003	.003	.002	.002
	(.002)	(.002)	(.005)	(.051)	(.004)	(.003)	(.002)	(.002)	(.001)	(.001)	(.001)	(.001)
$\widehat{\rho}$	.081	.487	.709	.917	.318	.244	.287	.286	.285	.285	.288	.292
	(.034)	(.027)	(.025)	(.012)	(.047)	(.076)	(.036)	(.023)	(.023)	(.020)	(.017)	(.009)
$\widehat{\beta}$	.401	.399	.397	.364	(.008)	.801	.402	.399	.401	.400	.400	.400
	(.018)	(.019)	(.019)	(.023)	(.025)	(.018)	(.019)	(.019)	(.016)	(.012)	(.009)	(.001)
$\widehat{\gamma}$	.532	.512	.517	.509	.432	.552	.376	.698	.503	.495	.493	.492
	(.023)	(.022)	(.024)	(.028)	(.070)	(.033)	(.027)	(.021)	(.018)	(.015)	(.012)	(.007)

B. Political party

	ρ_0				β_0		γ_0		q			
	.1	**.5**	**.7**	**.9**	**.0**	**.8**	**.3**	**.7**	**.3**	**.5**	**.8**	**1.0**
% True Zeroes	.971	.983	.982	.966	.978	.977	.959	.986	.985	.992	.996	.997
	(.008)	(.005)	(.006)	(.007)	(.008)	(.007)	(.007)	(.005)	(.004)	(.004)	(.004)	(.003)
% True Non-Zeroes	.772	.813	.836	.864	.469	.856	.741	.834	.803	.808	.816	.832
	(.045)	(.036)	(.034)	(.031)	(.147)	(.035)	(.044)	(.036)	(.037)	(.035)	(.034)	(.027)
$MAD(\widehat{W})$	.017	.012	.012	.018	.029	.012	.023	.010	.011	.009	.008	.007
	(.003)	(.002)	(.002)	(.003)	(.007)	(.002)	(.003)	(.002)	(.002)	(.001)	(.001)	(.001)
$MAD(\widehat{\Pi})$	.011	.021	.043	.208	.021	.017	.016	.016	.013	.011	.010	.009
	(.002)	(.002)	(.004)	(.026)	(.004)	(.003)	(.002)	(.002)	(.002)	(.001)	(.001)	(.001)
$\widehat{\rho}$	.046	.403	.598	.847	.298	.171	.221	.225	.224	.222	.218	.218
	(.038)	(.031)	(.040)	(.040)	(.071)	(.067)	(.048)	(.030)	(.030)	(.028)	(.025)	(.018)
$\widehat{\beta}$	.400	.399	.398	.380	-.015	.801	.402	.400	.399	.398	.398	.398
	(.019)	(.020)	(.020)	(.026)	(.029)	(.020)	(.020)	(.019)	(.016)	(.014)	(.009)	(.003)
$\widehat{\gamma}$	.519	.489	.488	.469	.305	.548	.391	.638	.474	.454	.440	.431
	(.030)	(.027)	(.031)	(.050)	(.117)	(.038)	(.036)	(.025)	(.023)	(.019)	(.015)	(.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 A3: Simulation Results, Adaptive Lasso

	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
% True Zeroes	.728	.798	.842	.737	.848	.878	.845	.846	.913	.730	.799	.840	.737	.849	.880	.846	.846	.921
	(.023)	(.024)	(.022)	(.012)	(.008)	(.008)	(.005)	(.006)	(.027)	(.024)	(.025)	(.023)	(.013)	(.009)	(.009)	(.006)	(.006)	(.008)
% True Non-Zeroes	.996	1.000	1.000	.497	1.000	1.000	.494	.496	.930	.992	1.000	1.000	.507	1.000	1.000	.498	.501	.996
	(.039)	(.000)	(.000)	(.096)	(.001)	(.000)	(.071)	(.074)	(.177)	(.023)	(.002)	(.000)	(.105)	(.004)	(.001)	(.074)	(.076)	(.042)
$MAD(\widehat{W})$	.071	.050	.039	.064	.039	.033	.039	.039	.027	.070	.050	.039	.063	.038	.032	.039	.038	.025
	(.005)	(.005)	(.004)	(.002)	(.002)	(.002)	(.000)	(.002)	(.005)	(.005)	(.004)	(.004)	(.002)	(.002)	(.002)	(.000)	(.002)	(.001)
$MAD(\widehat{\Pi})$	.092	.056	.042	.084	.049	.036	.071	.045	.032	.093	.057	.044	.084	.050	.036	.071	.045	.033
	(.008)	(.005)	(.004)	(.005)	(.003)	(.002)	(.004)	(.003)	(.002)	(.008)	(.005)	(.003)	(.005)	(.003)	(.002)	(.004)	(.002)	(.002)
$\widehat{\rho}$	.962	.815	.535	.998	.988	.965	1.000	.993	.995	.970	.783	.512	.998	.993	.979	1.000	.994	.995
	(.112)	(.186)	(.231)	(.040)	(.066)	(.095)	(.000)	(.075)	(.054)	(.106)	(.199)	(.236)	(.041)	(.050)	(.077)	(.000)	(.069)	(.054)
$\widehat{\beta}$	.131	.285	.330	.000	.158	.254	.000	.000	.121	.144	.292	.336	.000	.177	.259	.000	.000	.167
	(.079)	(.049)	(.038)	(.000)	(.053)	(.030)	(.000)	(.000)	(.066)	(.081)	(.049)	(.039)	(.000)	(.047)	(.030)	(.000)	(.000)	(.044)
$\widehat{\gamma}$	.996	.998	.968	.000	1.000	.999	.000	.000	.863	1.000	.995	.942	.000	1.000	.999	.000	.000	.992
	(.063)	(.015)	(.051)	(.000)	(.000)	(.007)	(.000)	(.000)	(.344)	(.000)	(.023)	(.066)	(.000)	(.000)	(.014)	(.000)	(.000)	(.089)

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 A4: 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	.825	.878	.911	.884	.928	.958	.936	.966	.979	.824	.882	.916	.886	.932	.959	.940	.967	.979
	(.018)	(.020)	(.021)	(.007)	(.007)	(.006)	(.019)	(.003)	(.002)	(.020)	(.021)	(.020)	(.007)	(.007)	(.005)	(.003)	(.003)	(.002)
% True Non-Zeroes	1.000	1.000	1.000	1.000	1.000	1.000	.977	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)	(.107)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
$MAD(\widehat{W})$	.039	.032	.028	.031	.025	.022	.025	.021	.019	.037	.030	.025	.030	.024	.021	.024	.020	.018
	(.003)	(.004)	(.004)	(.001)	(.001)	(.001)	(.000)	(.000)	(.000)	(.003)	(.003)	(.003)	(.001)	(.001)	(.001)	(.000)	(.000)	(.000)
$MAD(\widehat{\Pi})$	.038	.027	.022	.039	.027	.022	.040	.027	.022	.038	.027	.022	.038	.027	.022	.039	.027	.022
	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.000)	(.000)	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.000)	(.000)
$\widehat{\rho}$	.488	.381	.362	1.000	.981	.617	1.000	1.000	1.000	.465	.396	.382	1.000	.953	.590	1.000	1.000	1.000
	(.114)	(.117)	(.108)	(.000)	(.051)	(.064)	(.000)	(.000)	(.000)	(.108)	(.104)	(.074)	(.000)	(.075)	(.062)	(.000)	(.000)	(.000)
$\widehat{\beta}$	.394	.398	.399	.343	.372	.400	.256	.350	.353	.397	.401	.399	.346	.382	.400	.280	.350	.350
	(.017)	(.009)	(.006)	(.017)	(.025)	(.003)	(.064)	(.002)	(.013)	(.022)	(.014)	(.005)	(.013)	(.024)	(.000)	(.029)	(.002)	(.000)
$\widehat{\gamma}$	.964	.871	.809	1.000	1.000	.998	.956	1.000	1.000	.955	.841	.769	1.000	1.000	.995	1.000	1.000	1.000
	(.046)	(.079)	(.083)	(.000)	(.000)	(.009)	(.205)	(.000)	(.000)	(.053)	(.068)	(.046)	(.000)	(.000)	(.015)	(.000)	(.000)	(.000)

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 A5: Summary Statistics, Tax Competition Application

	Obs	Mean	SD	Min	q25	Median	q75	Max
				A. Besley and Case sample (1962-1988)				
State total tax per capita	1296	.371	.266	.036	.145	.300	.530	1.345
State income per capita	1296	9.951	2.130	4.105	8.585	9.919	11.375	18.808
Unemployment rate	1296	5.885	2.242	1.800	4.200	5.500	7.000	18.000
Proportion of young	1296	.234	.033	.160	.210	.240	.260	.310
Proportion of elderly	1296	.106	.020	.040	.090	.110	.120	.190
State governor's age	1296	51.088	7.441	33.000	45.000	50.000	56.000	73.000
Governor term limit dummy	1296	.258	.438	.000	.000	.000	1.000	1.000
				B. Extended sample (1962-2014)				
State total tax per capita	2688	0.983	0.803	0.036	0.037	0.813	1.557	4.298
State income per capita	2736	13.268	4.016	4.147	10.348	12.960	15.879	27.974
Unemployment rate	2688	5.764	2.026	1.800	4.300	5.400	6.800	17.800
Proportion of young	2688	0.236	0.033	0.170	0.210	0.230	0.260	0.340
Proportion of elderly	2688	0.117	0.023	0.050	0.100	0.120	0.130	0.190
State governor's age	2736	53.557	8.134	33.000	47.000	53.000	59.000	78.000
Governor term limit dummy	2638	0.249	0.433	0.000	0.000	0.000	0.000	1.000

Notes: Summary statistics of variables (in levels) used in subsequent regressions. Besley and Case sample runs from 1962 to 1988 and extended sample until 2014. 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. Governor term limit dummy is equal to 1 if governor faces term limits in the current mandate. Data sources: State total tax per capita, Census of Governments (1972, 1977, 1982, 1987, 1992-2016) and Annual Survey of Goverment Finances (all other years); Steta 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 A6: Exogenous Social Effects

Dependent variable: Change in per capital income and corporate taxes
Coefficient estimates, standard errors in parentheses

	(1) Initial	(2) OLS	(3) 2SLS: IVs are Characteristics of Neighbors	(4) 2SLS: IVs are Characteristics of Neighbors-of Neighbors
Economic Neighbors' tax change (t - [t-2])	.645	.145**	.332*	.608***
		(.072)	(.199)	(.220)
Economic Neighbors' income per capita	.090	.098***	.091***	.080***
		(.011)	(.012)	(.014)
Economic Neighbors' unemployment rate	37.200	9.899***	11.780***	13.714***
		(3.443)	(2.856)	(3.022)
Economic Neighbors' population aged 5-17	1378.1	376.2	478.5	596.6
		(399.0)	(414.2)	(401.7)
Economic Neighbors' population aged 65+	-4304.5	-842.8	-769.7*	-641.3
		(504.3)	(450.2)	(468.6)
Economic Neighbors' governor age	-2.158	-0.311	-0.293	-0.263
		(.281)	(.285)	(.294)
Period			1962-2015	
First Stage (F-stat)			9.571	10.480
Controls	Yes	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,952	2,952	2,544	2,592

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 state i's total taxes per capita in year t. In all Columns, we penalize geographic neighbors in all Columns and allow for exogenous social effects. In OLS and IV regressions, the economic neighbors' effect is calculated as the weighted average of economic neighbors' variables. OLS regressions estimates are shown in Column 2. Column 3 shows the 2SLS regression where each geographic neighbors' tax change is instrumented by lagged neighbor's state income per capita and unemployment rate. Column 4 shows a 2SLS regression where each geographic neighbors' tax change is instrumented by lagged neighbor-of-neighbor's state income per capita and unemployment rate. At the foot of Columns 3 and 4 we report the p-value on the F-statistic from the first stage of the null hypothesis that instruments are jointly equal to zero. 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.

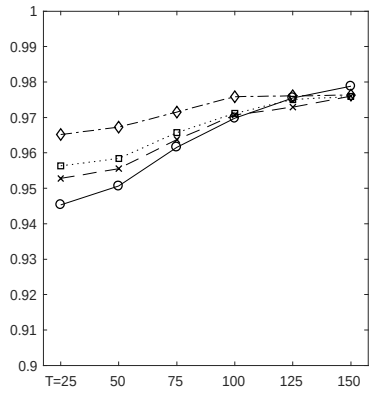
Table A7: General Equilibrium Impacts of California Tax Rise

	Geographic Neighbor Network	Economic Neighbor Network	Ratio
Average tax increase	0.0038	0.0066	1.74
Variance tax increase	0.0160	0.0153	0.96
Tax dispersion	0.0053	0.0141	2.66
States with tax increase	48	48	1.00
States with tax increase > 0.05%	11	44	4.00
States with tax increase > 0.5%	5	11	2.20
States with tax increase > 1%	4	8	2.00
States with tax increase > 2.5%	1	3	3.00
States with tax increase > 5%	1	1	1.00

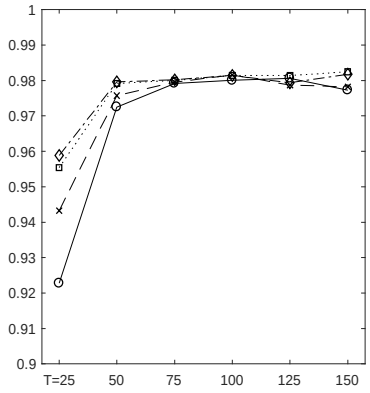
Notes: This shows the equilibrium impulse responses in taxes set in each state as a result of California increasing its tax change by 10%. The rho coefficient is derived from our preferred specification to estimate the economic network, where we penalize geographic neighbors to states, and allow for 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.

Figure A1: Simulation Results, Adaptive Elastic Net GMM
Partial Knowledge of W_0

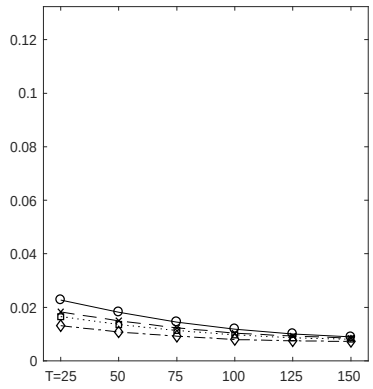
A. % of zeros



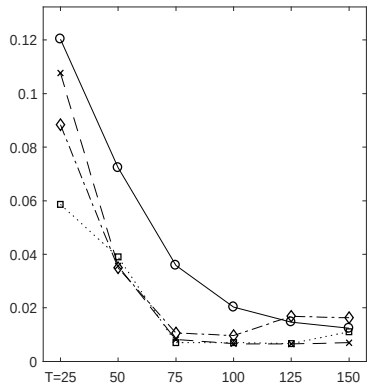
B. % of non-zeros



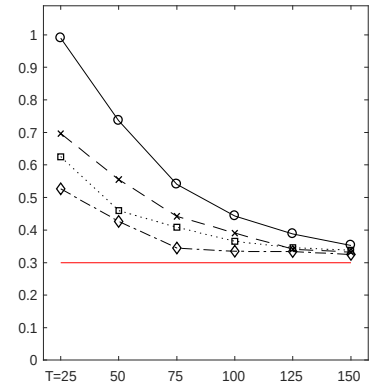
C. Mean Absolute Deviation of $\widehat{W}$



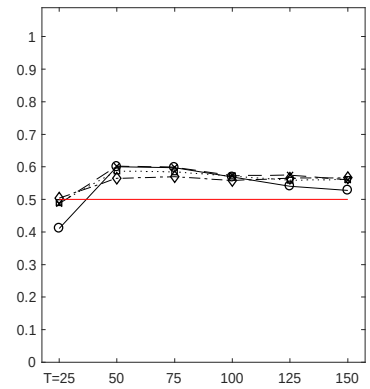
D. Mean Absolute Deviation of $\widehat{\Pi}$



E. Endogenous Social Effect, $\hat{\rho}$



F. Exogenous Social Effect $\hat{\gamma}$



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=.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.