ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis BSc$^2$ Econometrics / Economics

# Comparison of Adaptive LASSO, Adaptive Elastic Net and SCAD in the context of recovering the network structure from panel data


Name student: Veronika Soloveva

Student ID number: 450948

Supervisor: Yutao Sun

Second assessor: Maria Grith



Date final version: 12$^{th}$ of July 2020



**Abstract**

The knowledge of real-world network can largely facilitate decision making and thereby improve efficiency. In the novel study by De Paula, Rasul, and Souza (2020), the authors derive assumptions under which a network can be recovered directly from the data. Building on their results, I compare three distinct regularisation and variable selection techniques in the context of estimating network ties. In particular, Adaptive LASSO, Adaptive Elastic Net and SCAD are implemented with a GMM objective function. I find that each method has its own advantages and drawbacks and none of them is globally optimal. Moreover, I conduct a case study to investigate how corporate tax rates are interrelated in Western Europe. I conclude that there is a high level of connectivity among the countries in this area and the relations depend on a complex combination of factors.

# Table of Contents

# 1 Introduction

Everything in the world is interconnected, this makes it both beautiful and complex. The human body has different subsystems that have independent functions but all influence each other. In a society, the actions of every individual are affected by other people as well as external circumstances. The economic and political situation in one country significantly impacts the countries it has relations with. In every context, knowing the structure of a network enables understanding the group dynamics which, in turn, can largely facilitate decision making and improve efficiency.

Studying networks comprises three main types of activities: the postulation of theoretical models based on observation and knowledge, the identification of the model parameters and the data collection (Blume, Brock, Durlauf, & Jayaraman, 2015). To derive conclusions about real-world phenomena using data, it is important to be aware of possible limitations and requirements of each one of these steps. In this paper, I cover all three activities, but my main focus lies on the parameter estimation which belongs to the field of econometrics. The estimation of network models usually involves dealing with high-dimensional data sets where variables simultaneously influence each other. Therefore, it is necessary to derive assumptions under which the model parameters can be uniquely identified. An important contribution to this field was made by Manski (1993) who introduced the so-called 'Reflection problem'. This problem encompasses that network parameters are not identifiable if the researcher does not possess some prior information on relations among individuals in a given group.

Since this study, a lot of econometric research about networks has been done, most of which uses the data containing information on social ties (Sacerdote, 2001; Topa, 2001). However, as a result of high costs of data collection as well as privacy concerns, the information about the relations within a group of interest is often absent (Breza, Chandrasekhar, McCormick, & Pan, 2017). In this case, the authors often work with partial knowledge. For example, Blume et al. (2015) construct the sociomatrices by using the available information to divide the individuals into multiple neighbourhoods. Other researchers use economic theory to postulate the relations inside a group (Cassette & Paty, 2008; Redoano, 2014). A notable disadvantage of using limited information to create an adjacency matrix is that it can largely bias the results.

This is where the work of De Paula et al. (2020) makes an important contribution. They prove that, under certain assumptions, model parameters and the matrix capturing the interactions among individuals can be jointly identified from the data set. The authors demonstrate the validity of their results by estimating different simulated networks. Furthermore, they use data on taxation and other characteristics in different states of the United States of America to study the tax competition among them.

They perform the estimation by using Generalised Method of Moments (GMM) combined with diverse regularisation and variable selection techniques, among which the Adaptive Least Absolute Shrinkage and Selection Operator (LASSO) and Adaptive Elastic Net. The latter was found to be the best performing method, however, the authors acknowledge that some other methods might work better. Consequently, the question arises whether Adaptive Elastic Net is truly an optimal choice for the estimation.

The ultimate objective of this paper is to investigate how various estimation techniques differ in recovering network ties. For this, I carry out an extensive simulation study to analyse the performance of GMM with three distinct penalties on recovering four types of networks. In particular, the methods employed are Adaptive LASSO, Adaptive Elastic Net and Smoothly Clipped Absolute Deviation (SCAD). The networks for the study are carefully chosen such that they satisfy all the necessary assumptions, but vary in their characteristics to avoid biased conclusions. To summarise the goal, I formulate the following research question:

*What are the relative strengths and weaknesses of Adaptive LASSO, Adaptive Elastic Net and SCAD in the context of recovering the network structure from panel data?*

All models are estimated for various sample sizes by utilising the Particle Swarm Algorithm. Moreover, final estimates of the parameters are obtained employing two-stage least squares (2SLS) with peers-of-peers as instruments. Although these methods have been compared in various settings, the estimation of a whole network is a new area which is not yet sufficiently researched. Therefore, I enrich the econometric literature on high-dimensional estimation techniques by providing an overview of the qualities of the studied methods in the context of networks.

I find that all three methods have comparable performance and none of them is dominant. Adaptive Elastic Net is shown to recover the largest fraction of redundant parameters. Further, all networks can identify the majority of strong links but perform poorly in finding weaker ones. SCAD has the highest hit rate for weak links, however, this comes at the cost of a large number of falsely classified edges. Lastly, Adaptive LASSO enjoys the advantage of lowest computation time with only slightly inferior results.

After learning about the properties of each method, I conduct a case study in which I investigate the taxation in the first fifteen countries joining the European Union (henceforth Western Europe).[1] The question I aim to answer with this study is:

*How are the corporate income tax rates interrelated in Western Europe?*

To perform the analysis, I take a two-stage approach. First, I estimate the social interaction matrix using the data on statutory corporate income tax rates as well as on several social-economic characteristics of the respective countries for years 1981 up to 2018. For this, I employ the Adaptive Elastic Net and SCAD methodologies. As a second step, I use the Logit regression to investigate whether and to what extent the links depend on geographic and social-economic proximity of a pair of countries. I find that the corporate income tax rates are highly interconnected among the studied states. The results show that the connections go beyond the neighbouring countries and depend on a complex mix of various factors such as economic policy and demographics of each country.

In what follows, I first summarise the linear social interaction model and the main identification results of De Paula et al. (2020) in section 2. Next, I provide an explanation of the estimation procedure in section 3. In section 4, I describe relevant network characteristics, the selected networks as well as the procedure for data simulation. I then present the simulation results in section 5. Further, section 6 encompasses the detailed report of the case study and section 7 concludes the paper.

[1]Those are Austria (AUT), Belgium (BEL), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Ireland (IRL), Italy (ITA), Luxembourg (LUX), The Netherlands (NLD), Portugal (PRT), Spain (ESP), Sweden (SWE) and United Kingdom (GBR).

# 2 Model and identification

In this section, I introduce the linear social interactions model employed by De Paula et al. (2020) together with the assumptions for the parameter identification they derive. The starting point of the model is availability of panel data with $T$ observations on $N$ individuals. The individual outcomes, denoted by $y_{it}$, and individual characteristics, $x_{it}$, are assumed to be known for $t = 1, ..., T$ and $i = 1, ..., N$. Given those, the model is:

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

This model can be written more compactly in a matrix notation:

$$y_t = \rho_0 W_0 y_t + \beta_0 x_t + \gamma_0 W_0 x_t + \alpha^* + \alpha_t \imath + \varepsilon_t. \tag{2}$$

$W_0$ is the adjacency matrix, also called social interactions matrix, where $W_{0,ij}$ measures the effect of individual $j$ on the outcome of an individual $i$. The coefficients $\rho_0$ and $\gamma_0$ stand for the endogenous and exogenous effects respectively. Furthermore, $\beta_0$ represents the impact of an individual's own characteristics and $\alpha^*$ represents the unobserved heterogeneity. Finally, the scalar $\alpha_t$ accounts for time-specific shocks affecting each individual, $\imath$ is an $N \times 1$ vector of ones and $\varepsilon_t$ captures the individual deviations for every outcome. The subscript zero stands for the true parameter value.

The model has been presented in this form by Blume et al. (2015). It rests on the inference that all individuals choose their actions (which result in the outcomes) simultaneously and thus this is an incomplete information game. Each individual has a goal of maximising his or her own utility and so uses all available knowledge to make the best possible decision. As mentioned in the Introduction, De Paula et al. (2020) derive a set of assumptions necessary for the joint identification of network parameters and $W_0$. Since I will need to comply with these when estimating, I briefly review them below.

The model from equation 2 can be simplified by excluding individual fixed effects and common shocks. In this case, there are five assumptions to be satisfied. The first one states that an individual does not affect himself:

$$A1: W_{0,ii} = 0, \; i = 1,...,N.$$

The next assumption makes sure the effect of each shock vanishes with time:

*A2:* $\sum_{j=1}^{N} \mid \rho_0 W_{0,ij} \mid < 1$ *for every* $i = 1,\ldots N$, $\| W_0 \| < C$, $C$ *positive,* $C \in \mathbb{R}$ *and* $\mid \rho_0 \mid < 1$.

The third assumption prevents cancelling out of endogenous and exogenous network effects:

*A3:* $\beta_0\rho_0 + \gamma_0 \neq 0$.

Furthermore, a normalisation is necessary to enable joint identification of $\rho_0$ and $\gamma_0$:

*A4: There exists an* $i$ *such that* $\sum_{j=1}^{N} W_{0,ij} = 1$.

Lastly, assumption A5 requires that a network is asymmetric, meaning that individuals differ in popularity:

*A5: There exist* $l$ *and* $k$ *such that* $W^2_{0,ll} \neq W^2_{0,kk}$.

Given these assumptions and provided $\rho_0 > 0$ and $W_{0,ij} \geqslant 0$, the model has a unique global solution.[2]

To incorporate the unobserved heterogeneity, for the individual characteristics and outcomes the mean over time should be subtracted. This results in replacing $x_{it}$ by $x^*_{it} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$ and $y_{it}$ by $y^*_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$. Then, the model can be written in a reduced form as follows:

$$y^*_t = (I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0)x^*_t + (I - \rho_0 W_0)^{-1}\alpha_t \imath + (I - \rho_0 W_0)^{-1}\varepsilon^*_t. \tag{3}$$

To account for common shocks, assumption A4 needs to be modified:

*A4':* $\sum_{j=1}^{N} W_{0,ij} = 1$ *for all* $i = 1, ..., N$.

Next, the global differences should be taken by pre-multiplying by $(I - H)$, where $H = \frac{1}{N}(\imath\imath')$. If Assumption A4' holds, $(I - H)(I - \rho_0 W_0)^{-1}\alpha_t \imath = 0$ and the model becomes:

$$(I - H)y^*_t = (I - H)(I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0)x^*_t + \nu_t. \tag{4}$$

This is the model I work with in the remainder of the paper unless specified otherwise.

[2]Stated in Corollary 3 of De Paula et al. (2020).

# 3 Estimation

I start this section by introducing the GMM objective function. Then I describe three variable selection procedures which are employed and compared in this paper. Lastly, I explain how the optimisation is performed.

## 3.1 Problem description

The ultimate goal of the estimation is to recover the unknown parameter vector $\theta = (W_{1,2}, ..., W_{N,N-1}, \rho, \gamma, \beta)'$ from the observed characteristics and outcomes of the individuals within a group. What makes the estimation challenging is that while there are only $NT$ observations, there are $m = N(N-2)+3$ parameters.[3] At the same time, since the true value of the majority of the parameters is usually zero, the density of the networks of interest is low. Hence, some high-dimensional estimation technique which can simultaneously select and estimate the variables is required. An appropriate solution is using a penalty function.

Therefore, to obtain the parameters from the data, I minimise the penalised GMM function. In this paper, the GMM objective function takes the following form:

$$G_{NT}(\theta) = g_{NT}(\theta)' M_T g_{NT} \tag{5}$$

It is based on the moment conditions which assume independence of covariates and error terms. Here, $g_{NT}(\theta) = (I-H)\sum_{t=1}^{T}[x_{1t}^{*}\nu_t(\theta)' ... x_{Nt}^{*}\nu_t(\theta)']'$ with residuals $\nu_t(\theta) = (I-H)y_t^{*} - (I-H)(I-\rho W)^{-1}(\beta I + \gamma W)x_t^{*}$. Furthermore, $M_T$ is a weight matrix and it is set to $I_{N^2 \times N^2}$ for simplicity. To ensure that $\rho \geqslant 0$ and $W \geqslant 0$, $\rho$ and $W$ are replaced for the minimisation by $\tilde{\rho}^2$ and $\tilde{W}^2$ respectively.

## 3.2 Variable selection

According to Fan and Li (2001), a good penalty should possess three main qualities: a stable model selection, sparse solution and unbiased estimates for larger coefficients. Furthermore, an important quality is the oracle property which means that a technique "performs

[3]Only N(N-2) parameters of $W_0$ need to be estimated since in each row $i$ of $W$ one entry next to the main diagonal, $W_{i,j^*}$, is set to 1 - $\sum_{j=1,j\neq j^*}^{N} W_{i,j}$ to satisfy the assumption A4'.

as well as if the true underlying model were given in advance" (Zou, 2006). Next, three estimation methods used in this study are described and compared in terms of these properties.

### 3.2.1 Adaptive LASSO

The first estimation method is Adaptive LASSO. It consists of two steps. The estimate of the first step is equivalent to that of LASSO:

$$\hat{\theta}_l(p_1) = \underset{\theta \in \mathbb{R}^m}{\arg\min} \left\{ G_{NT}(\theta) + p_1 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mid W_{ij} \mid \right\}. \tag{6}$$

Although LASSO performs well in shrinking the insignificant coefficients to zero, it produces biased estimates for large coefficients (Fan & Li, 2001). For this reason, an adaptive step was proposed by Zou (2006). This step introduces coefficient-specific weights which use the adjacency matrix estimate of the first step, $\hat{W}_l$, and thereby achieves the oracle property. The resulting Adaptive LASSO is defined as:

$$\hat{\theta}_{al}(p_1^*) = \underset{\theta \in \mathbb{R}^m}{\arg\min} \left\{ G_{NT}(\theta) + p_1^* \sum_{\substack{i,j:\hat{W}_{l;ij} \neq 0, \\ i,j=1,\dots N, \\ i \neq j}} \frac{\mid W_{ij} \mid}{\mid \hat{W}_{l;ij} \mid^{\gamma}} \right\}. \tag{7}$$

Similarly to Caner and Zhang (2014), I set the exponent $\gamma$ in equation 7 to 2.5 as it has been shown to achieve satisfactory results in diverse settings.

### 3.2.2 Adaptive Elastic Net

Next to biased coefficients, a drawback of LASSO is that it performs poorly if the explanatory variables are highly correlated. To improve this, Zou and Hastie (2005) suggest a method called Elastic Net which extends LASSO with quadratic penalty of Ridge regression. Furthermore, Zou and Zhang (2009) combine the two extensions in the Adaptive Elastic Net and prove that it simultaneously satisfies the oracle property and successfully handles collinearity. Later, Caner and Zhang (2014) showed that using the GMM objective function instead of Least Squares does not affect the desired properties. Hence, the second technique I use is the Adaptive Elastic Net whose first step estimate is calculated as follows:

$$\hat{\theta}_{en}(p_1, p_2) = (1 + p_2/T) \cdot \underset{\theta \in \mathbb{R}^m}{\arg\min} \left\{ G_{NT}(\theta) + p_1 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mid W_{ij} \mid + p_2 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mid W_{ij} \mid^2 \right\}. \tag{8}$$

The adaptive step estimate is then given by:

$$\hat{\theta}_{aen}(p_1^*, p_2) = (1+p_2/T) \cdot \underset{\theta \in \mathbb{R}^m}{\arg\min} \left\{ G_{NT}(\theta) + p_1^* \sum_{\substack{i,j:\hat{W}_{en;ij} \neq 0, \\ i,j=1,\ldots N, \\ i \neq j}} \frac{\mid W_{ij} \mid}{\mid \hat{W}_{en;ij} \mid^{\gamma}} + p_2 \sum_{\substack{i,j:\hat{W}_{en;ij} \neq 0, \\ i,j=1,\ldots N, \\ i \neq j}}^{N} \mid W_{ij} \mid^2 \right\}. \tag{9}$$

In both steps, the term $(1 + p_2/T)$ adjusts for the bias. Moreover, also here $\gamma$ is set to 2.5.

#### 3.2.3 SCAD

The third method used is SCAD. It possesses all the desired properties of a good penalty and is most commonly defined in form of its derivative. For $i \neq j$ it can be written as:

$$J'_{\lambda}(\mid W_{ij} \mid) = \lambda \left\{ I(\mid W_{ij} \mid \leq \lambda) + \frac{(a\lambda - \mid W_{ij} \mid)_+}{(a-1)\lambda)} I(\mid W_{ij} \mid > \lambda) \right\}, \tag{10}$$

with $a > 2$ and $\lambda > 0$ being the tuning parameters. Despite great statistical properties, the SCAD penalty is non-convex and is therefore challenging to estimate. The first two proposed solutions use Local Quadratic Approximation (Fan & Li, 2001) and Local Linear Approximation (Zou & Li, 2008). More efficient approaches were derived by Kim, Choi, and Oh (2008) and Wu and Liu (2009). The former uses Coupling of the Concave Convex Procedure (CCCP) and the latter the Difference Convex Algorithm (DCA). These two are essentially the same so I focus on describing the CCCP which is used for the estimation.

The SCAD penalty can be decomposed as a sum of convex and concave functions such that $J'_{\lambda}(\mid W_{ij} \mid) = J'_{\lambda,1}(\mid W_{ij} \mid) + J'_{\lambda,2}(\mid W_{ij} \mid)$ with:

$$\begin{cases} J'_{\lambda,1}(\mid W_{ij} \mid) = \lambda, \\ J'_{\lambda,2}(\mid W_{ij} \mid) = -\lambda(1 - \frac{(a\lambda - \mid W_{ij} \mid)_+}{(a-1)\lambda)}) I(\mid W_{ij} \mid > \lambda), \end{cases} \tag{11}$$

and with $(t)_+ = tI(t > 0)$. Furthermore, for $J_{\lambda,2}(W_{ij})$ a linear approximation can be used. The parameters of SCAD-penalised GMM can then be obtained by iterative estimation:

$$\hat{\theta}_{scad,k+1}(\lambda) = \underset{\theta \in \mathbb{R}^m}{\arg\min} \left\{ G_{NT}(\theta) + \frac{1}{NT} \lambda \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mid W_{i,j} \mid + \frac{1}{NT} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} J'_{\lambda,2}(\mid \hat{W}_{scad,k;ij} \mid) W_{ij} \right\}. \tag{12}$$

I only perform three iterations since through preliminary testing I find that in most cases this is enough for the estimates to converge.

## 3.3 Implementation

### 3.3.1 Particle swarm algorithm

Similarly to De Paula et al. (2020), I choose Particle Swarm Algorithm for the optimisation of all three estimation methods. In this algorithm, multiple particles (i.e. initial conditions) independently search for the optimal solution which increases the probability of finding the global minimum. Next to that, this algorithm does not use derivatives which is convenient for functions that are non-linear in their parameters. Since these advantages come at the cost of efficiency, I make several simplifications as compared to De Paula et al. (2020).

First of all, to ensure the feasibility of an extensive simulation study, I only perform 100, instead of 1000, simulation rounds. Then, I make use of the $'MaxTime'$ option in the build-in *particleswarm* function in $MATLAB$ and restrict the maximum estimation time to 150 seconds for networks with $N = 30$, to 175 seconds for a network with $N = 45$ and to 200 seconds for a network with $N = 65$. Finally, I set the lower bound to zero and the upper bound to one on the parameters as this information is incorporated in the assumptions and helps to reduce the search space. Furthermore, I only use five, instead of 100, initial particles, of which the first three are calculated from the data.[4]

*Particle 1:* The parameters corresponding to the entries of $W$ are obtained from the LASSO regression of $y_t$ on $y_t$ of the other individuals. As LASSO penalty, $p_1$ is used for Adaptive LASSO and Adaptive Elastic Net. For SCAD this penalty is fixed at 0.15. The estimated parameters get weights proportional to their magnitude such that the row-sum normalisation is respected. To initialise $\beta$, I use an estimate of the Ordinary Least Squares regression of $y$ on $x$. Finally, I set $\gamma$=0 and $\rho$=0.5 which ensures that Assumption A3 holds.

*Particle 2:* Similarly to particle 1, but now with LASSO regression of $y_t$ on $x_t$ of others.[5]

*Particle 3:* I sum the estimates from the LASSO regression of $y_t$ on $y_t$ of others and from the LASSO regression of $y_t$ on $x_t$ of other. All other computations are similar to particle 1.

In addition, the second and third iterations of SCAD use estimates from the previous step as an initial condition. The remaining particles are generated randomly by the build-in option of the *particleswarm* function.

[4] The particles based on the gradient are omitted due to their computational intensity.

[5] This particle is also used as the initial estimate of $W_{ij}$ for the first iteration of SCAD.

### 3.3.2 Choice of tuning parameters

For all three methods described above, the estimates strongly depend on the regularisation parameters. However, choosing those individually in each simulation run or making use of cross-validation would again heavily affect the evaluation time. Therefore, for Adaptive LASSO and Adaptive Elastic Net, the selection of the tuning parameter is only carried out for the first five simulation runs. In the subsequent 95 runs, the median of the optimal penalty vectors is used. The selection is done by evaluating the model for all possible $p$'s on the grid given by $\{0.15,\ 0.2,\ 0.225\}$. This results in 9 possible permutations for Adaptive LASSO and 27 for Adaptive Elastic Net. Note that this grid differs from De Paula et al. (2020) who define it as $\{0,\ 0.025,\ 0.05,\ 0.1\}$. This choice is made after observing that increasing the penalty leads to a higher number of zeroes estimated correctly as well as a general performance improvement. For each of the first five runs, the optimal penalty vector is the one minimising the BIC criterion defined as:

$$BIC(p) = log\big[g_{NT}(\hat{\theta}(p))' M_T g_{NT}(\hat{\theta}(p))\big] + A(\hat{\theta}(p)) \cdot \frac{logT}{T}, \tag{13}$$

with $A(\hat{\theta}(p)) = \sum_{i,j:\hat{W}_{ij}\neq 0} \hat{W}_{ij}$. Because the objective of this paper is estimating the parameters, the BIC criterion is preferred over other information criteria based on its asymptotic consistency (Nishii, 1984).

For SCAD, a similar approach to parameter tuning is used. To start with, I follow the convention and set $a$=3.7 as suggested by Fan and Li (2001). As only one parameter needs to be chosen, for the first 10 out of 100 simulation runs the estimation is done for all $\lambda$ in the set $\{3,\ 2.5,\ 2,\ 1,\ 0.5\}$. The optimum per round is selected based on BIC (equation 13 with $\lambda$ instead of $p$) and the median of the ten optimal values is used for the remaining runs.

### 3.3.3 Evaluation of results

Following De Paula et al. (2020), I calculate the *"post-"* $\beta_0$, $\gamma_0$ and $\rho_0$. For this, I implement 2SLS with peers-of-peers covariates as instruments using the estimated $\hat{W}$. I describe the procedure in detail in Appendix A.

To compare the different estimation methods, I calculate the mean absolute deviation of the social interaction matrix, defined as $MAD(\hat{W}) = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} (\hat{W}_{ij} - W_{0,ij})$. Furthermore, I separately compute the proportion of correctly recovered zero and non-zero entries of $W_0$. I also report the number of correctly classified strong and weak links. For a deeper analysis, the network characteristics described in next section are presented.

# 4 Networks and data simulation

As real-world networks differ a lot in their structure, it is important to pick diverse networks for testing the performance of the proposed estimation methods. In this section, I first introduce the relevant network terminology and characteristics, then I describe the four selected networks and subsequently explain how the data will be simulated.

## 4.1 Network characteristics

The standard network representation is a graph with $N$ nodes and edges connecting them. The adjacency matrix $W_0$ is another representation where an entry $W_{0,ij}$ stands for an edge going from individual $i$ to individual $j$. I only work with weighted directed networks, meaning that edges are of different strengths and $W_{0,ij} \neq 0$ does not imply that $W_{0,ji} \neq 0$. This type of networks is most representative of social and economic relations. The directed ties are naturally present in seller-buyer relationships (Kranton & Minehart, 2001) and supply chain systems (Carvalho, Nirei, Saito, & Tahbaz-Salehi, 2016). Furthermore, the weak ties often serve as a connection between multiple smaller well-established networks and thereby extend microenvironments to macroenvironments (Granovetter, 1977).

The first informative feature of a network system is its density. This feature represents the level of connectedness and is calculated as the fraction of total edges over total possible edges. Further, networks may contain different number of components, i.e. independently connected subgraphs (Jackson, Rogers, & Zenou, 2017). For directed graphs, a distinction between weak and strong components is made. To find weak components, all edges are seen as undirected and two nodes belong to the same component if there is a path connecting them. On the other hand, in case of a strong link, for each pair of nodes there exists a path connecting them in both directions. In what follows, I only report the number of strong components for each network and therefore refer to those simply as "components". Lastly, a relevant measure is the clustering coefficient. It is computed as $(3 \cdot Number\ of\ triangles)/(Total\ number\ of\ triples)$ and describes to what extent the location of links is correlated (Watts & Strogatz, 1998). For the exact definition of triangles and triples I refer to Newman, Watts, and Strogatz (2002).

## 4.2 Selected networks

The first network used for this simulation study is the famous Erdos-Renyi network (Erdős & Rényi, 1960). It is constructed by randomly setting one of the entries in each row of $W_0$ equal to one such that each individual is only influenced by one other individual and not him/herself. For $N = 30$, it has a relatively low density (3.45%). Despite being a convenient choice for its simplicity, the Erdos-Renyi network does not capture many aspects present in real-world networks. It is not scale-free, i.e. the distribution of links does not follow the power law under which there exists a small number of nodes with an exceptionally high in- and/or out-degree. Also, the Erdos-Renyi network is not a small-world network as it does not have a high clustering coefficient and an average path between two nodes is long.

An attempt to improve with respect to the scale-free property is the Political Party network. In this network, the nodes are assumed to belong to two political parties, the first $\frac{N}{3}$ nodes to Party A and the remaining $2\frac{N}{3}$ to Party B, with the individuals 1 and $\frac{N}{3}+1$ being the party leaders. Each leader directly affects half of the party members and in each row there is an additional random link. As a consequence, the two nodes representing the party leaders have a significantly higher out-degree. To make sure the Assumption A4' is complied with, the link from party leader always gets value 0.7 and the additional link in the corresponding row is set to 0.3. Given $N = 30$, the density of this network is 5.17%.

The third network is based on the true interaction matrix which captures the family relations of the households in a rural village in India. The village selected is the village 10 of the original study by Banerjee, Chandrasekhar, Duflo, and Jackson (2013a) and consists of 77 households. The data set is retrieved from Harvard Dataverse (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013b). Further, some adjustments need to be made. First, the 12 isolated households are removed which results in a network with $N = 65$ and density of 5.77%. Then, the binary ties are transformed such that they have different strength and the row-sum normalisation is satisfied. This is done in the following way. Assume there are $l$ 'ones' in a row $i$ of $W_0$. One of these positive entries is randomly selected and set to 0.7. The remaining $l-1$ entries each get the value $0.3/(l-1)$. This network is larger, denser and has a higher clustering coefficient than the previous two and is therefore a suitable choice to get insights into the performance of estimation methods.

The last network is the Watts–Strogatz network. Although the links in this network do not follow the power law, it possesses the small-world property (Watts & Strogatz, 1998). In this network, the $N$ nodes are located in form of a ring and every node is connected to the $K$ nearest nodes on each side. Moreover, the parameter $\beta$ is the probability that an edge gets rewired and hereby introduces some randomness into the model.[6] I construct a Watts-Strogatz network with $N = 45$, $K = 3$ and $\beta = 0.05$ by using the function made by The MathWorks (2015). It is important to set $\beta > 0$ as otherwise the assumption A5 would be violated. I then assign the strength to each edge in the same way as for the Village network. With a density of 13.64% and clustering coefficient 0.2812, this network is significantly more complex than the other ones and hence it is a valuable contribution.

The first three networks are the same as in De Paula et al. (2020). Note that some features of the stylised networks may slightly differ from the original paper due to the randomness involved. Table 1 summarises the characteristics of all networks and also contains the standard deviation across the diagonal elements of $W_0^2$ to demonstrate that the chosen networks satisfy Assumption A5.[7] Moreover, Figure 2 in Appendix B shows their graphical representation.

## 4.3 Data generation

In each simulation round, the data is generated for a given T. This is done in the same way for each network. First, the variables $\alpha_t \sim \mathrm{N}(1,1)$, $\alpha^* \sim \mathrm{N}(\imath_N, I_{N\times N})$, $x_t \sim \mathrm{N}(0_N, I_{N\times N})$ and $\varepsilon_t \sim \mathrm{N}(0_N, I_{N\times N})$ are drawn from the corresponding distributions. Subsequently, these variables, the $W_0$ of the corresponding network as well as the true parameters $\rho_0$=0.3, $\beta_0$=0.4 and $\gamma_0$=0.5 are used to generate the outcomes $y_t$. The simulations are conducted for $T =$

[6] The additional parameter is defined as $\beta$ here to follow the conventional notation. However, for the remainder of the text, $\beta$ is the parameter of individual's characteristics in social interactions model.

[7] A5 does not hold if this value is zero.

$\{50,100,150\}$ for each of the four networks.

Table 1: *Characteristics of the selected networks*

| Characteristic | Erdos-Renyi | Political party | Village | Watts–Strogatz |
|---|---|---|---|---|
| Density | 3.45% | 5.17% | 5.77% | 13.64% |
| Total edges | 30 | 45 | 240 | 270 |
| Strong edges[a] | 30 | 30 | 65 | 45 |
| Weak edges | 0 | 15 | 175 | 225 |
| Reciprocal edges | 2 | 1 | 120 | 135 |
| In-degree distribution[b] | 1.0000 (0.0000) | 1.5000 (0.5085) | 3.6923 (2.3513) | 6.0000 (0.4767) |
| Out-degree distribution | 1.0000 (1.0828) | 1.5000 (2.1616) | 3.6923 (2.3513) | 6.0000 (0.4767) |
| Highest out-degree[c] | {29,28,5} | {11,1,5} | {35,23,57} | {19,26,34} |
| Number of components | 23 | 27 | 3 | 1 |
| Size largest component | 6 | 4 | 51 | 45 |
| Clustering coefficient | - | 0 | 0.0998 | 0.2812 |
| St. Dev. diagonal of $W_0^2$ | 0.3457 | 0.0761 | 0.1935 | 0.1545 |

[a] An edge is define as strong if its weight is >0.3. [b] For in- and out-degree distribution the mean and standard deviation in parenthesis are reported. [c] Three nodes with highest out-degree.

# 5 Results

The results are summarised in the following tables. Tables 2 up to 5 present the performance metrics for each of the four networks. Appendix B contains tables with the estimates of the peers-of-peers regressions (Tables 13, 14, 15 and 16) and with the various characteristics of the estimated networks (Tables 17, 18, 19 and 20). While the performance metrics give a concise overview, the characteristics of the recovered networks provide additional valuable insights which help understanding the origin of the results and the specifics of each method.

Table 2: *Simulation results for Erdos-Renyi network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| MAD($\hat{W}$) | 0.0976 | 0.0497 | 0.0226 | 0.0728 | 0.0434 | 0.0257 | 0.0608 | 0.0233 | 0.0154 |
| | (0.0968) | (0.0449) | (0.0132) | (0.0675) | (0.0391) | (0.0286) | (0.0519) | (0.0046) | (0.0036) |
| true zeros[a] | 0.9061 | 0.9705 | 0.9581 | 0.9600 | 0.9702 | 0.9788 | 0.9259 | 0.9470 | 0.9581 |
| | (0.0886) | (0.0228) | (0.0097) | (0.0337) | (0.0328) | (0.0158) | (0.0248) | (0.0099) | (0.0103) |
| true non-zeros | 0.6903 | 0.7963 | 0.9480 | 0.6823 | 0.8400 | 0.9013 | 0.7877 | 0.9530 | 0.9917 |
| | (0.1713) | (0.1102) | (0.0557) | (0.1262) | (0.1297) | (0.1221) | (0.0846) | (0.0464) | (0.0167) |
| true strong links[b] | 18.9000 | 23.6300 | 28.0000 | 20.1100 | 24.7800 | 26.8500 | 21.6200 | 28.0500 | 29.6400 |
| | (5.4910) | (3.3474) | (2.0938) | (3.6622) | (4.0666) | (3.8987) | (2.9671) | (1.5594) | (0.5777) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

[a] Fraction of true zeros and of true non-zeros recovered correctly.

[b] The links with weights bigger than 0.3 are classified as strong.

Table 3: *Simulation results for Political Party network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| MAD($\hat{W}$) | 0.0971 | 0.0728 | 0.0558 | 0.0985 | 0.0636 | 0.0538 | 0.1110 | 0.0357 | 0.0246 |
| | (0.0696) | (0.0803) | (0.0524) | (0.0506) | (0.0405) | (0.0668) | (0.0656) | (0.0064) | (0.0041) |
| true zeros[a] | 0.9493 | 0.9622 | 0.9700 | 0.9489 | 0.9671 | 0.9698 | 0.9128 | 0.9422 | 0.9549 |
| | (0.0405) | (0.0419) | (0.0306) | (0.0271) | (0.0230) | (0.0361) | (0.0346) | (0.0116) | (0.0107) |
| true non-zeros | 0.3982 | 0.5093 | 0.5624 | 0.3951 | 0.5067 | 0.5458 | 0.4602 | 0.6467 | 0.7333 |
| | (0.0817) | (0.0881) | (0.0794) | (0.0786) | (0.0707) | (0.0931) | (0.0758) | (0.0613) | (0.0521) |
| true strong links[b] | 15.3700 | 20.2300 | 22.2400 | 15.2700 | 20.2000 | 22.2100 | 15.5400 | 23.0600 | 26.2000 |
| | (3.2213) | (3.6896) | (3.3246) | (3.3540) | (3.1042) | (3.9982) | (3.1056) | (2.1687) | (2.0101) |
| true weak links | 0.0300 | 0.1600 | 0.2600 | 0.1100 | 0.0900 | 0.1900 | 0.7600 | 1.8100 | 2.5300 |
| | (0.1714) | (0.4197) | (0.5966) | (0.3145) | (0.3208) | (0.5064) | (0.9003) | (1.4120) | (1.6172) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

[a] Fraction of true zeros and of true non-zeros recovered correctly.

[b] The links with weights bigger than 0.3 are classified as strong.

Table 4: *Simulation results for Village network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| MAD($\hat{W}$) | 0.0568 | 0.0289 | 0.0203 | 0.0485 | 0.0271 | 0.0189 | 0.0437 | 0.0205 | 0.0143 |
| | (0.0206) | (0.0077) | (0.0078) | (0.0102) | (0.0053) | (0.0078) | (0.0466) | (0.0022) | (0.0014) |
| true zeros[a] | 0.9701 | 0.9894 | 0.9937 | 0.9784 | 0.9901 | 0.9944 | 0.9559 | 0.9749 | 0.9823 |
| | (0.0136) | (0.0042) | (0.0045) | (0.0066) | (0.0030) | (0.0050) | (0.0468) | (0.0042) | (0.0038) |
| true non-zeros | 0.1828 | 0.2262 | 0.2603 | 0.1786 | 0.2332 | 0.2618 | 0.2471 | 0.3241 | 0.3507 |
| | (0.0237) | (0.0243) | (0.0222) | (0.0258) | (0.0249) | (0.0192) | (0.0306) | (0.0288) | (0.0236) |
| true strong links[b] | 27.0200 | 44.9200 | 54.1700 | 28.4600 | 46.1800 | 54.6900 | 31.1000 | 45.9200 | 57.4300 |
| | (3.8898) | (4.8151) | (4.5706) | (4.2603) | (4.3886) | (4.7048) | (5.4393) | (3.9227) | (3.0920) |
| true weak links | 1.3100 | 0.2600 | 0.7200 | 0.2800 | 0.3900 | 0.6800 | 9.2700 | 13.4100 | 13.9300 |
| | (1.5288) | (0.6908) | (1.5247) | (0.8771) | (1.5301) | (1.3772) | (4.8966) | (4.5083) | (4.2409) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

[a] Fraction of true zeros and of true non-zeros recovered correctly.

[b] The links with weights bigger than 0.3 are classified as strong.

Table 5: *Simulation results for Watts Strogatz network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| MAD($\hat{W}$) | 0.0542 | 0.0359 | 0.0294 | 0.0517 | 0.0378 | 0.0359 | 0.0320 | 0.0227 | 0.0184 |
| | (0.0089) | (0.0130) | (0.0143) | (0.0077) | (0.0164) | (0.0359) | (0.0030) | (0.0017) | (0.0012) |
| true zeros[a] | 0.9799 | 0.9892 | 0.9925 | 0.9809 | 0.9883 | 0.9878 | 0.9604 | 0.9724 | 0.9799 |
| | (0.0056) | (0.0076) | (0.0088) | (0.0051) | (0.0089) | (0.0205) | (0.0052) | (0.0041) | (0.0043) |
| true non-zeros | 0.1339 | 0.1586 | 0.1701 | 0.1354 | 0.1596 | 0.1778 | 0.2477 | 0.2760 | 0.2705 |
| | (0.0182) | (0.0157) | (0.0163) | (0.0201) | (0.0174) | (0.0211) | (0.0214) | (0.0207) | (0.0217) |
| true strong links[b] | 21.1200 | 32.3900 | 36.6900 | 21.9000 | 31.8800 | 36.8600 | 26.9200 | 38.3900 | 42.9900 |
| | (3.5370) | (4.0847) | (3.8995) | (3.0732) | (4.0509) | (5.6712) | (3.5124) | (2.3394) | (1.4805) |
| true weak links | 0.4200 | 0.7400 | 1.0200 | 0.5500 | 0.7900 | 2.2200 | 17.6900 | 21.3400 | 20.9600 |
| | (1.2963) | (1.2841) | (1.6936) | (1.8056) | (1.4020) | (2.8018) | (4.5186) | (4.6887) | (5.2875) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

[a] Fraction of true zeros and of true non-zeros recovered correctly.

[b] The links with weights bigger than 0.3 are classified as strong.

## 5.1 General results

A positive observation shared by all methods and networks is that the performance drastically improves with higher sample size and the variation of each statistic decreases. This convergence seems to occur at the same rate for all three methods. Especially the precision of correctly discovering non-zero edges improves. In particular, the number of true zeros estimated as non-zeros decreases while the total number of edges recovered increases. For each method, the edge recovery rate is highest for Erdos-Renyi network which can be attributed to the fact that all its edges have a value of one and thus are strong. The results for the other three networks show that strong links can be more easily identified than weak ones. This can be explained by the fact that as the number of weak links present in the network increases, their value decreases (due to the row-sum normalisation). Hence, these edges have less influence on the GMM objective function and it is harder to distinguish them from zero.

Moreover, there is a lot of variability across the simulation runs which can be attributed to the fact that only 100 (as opposed to 1000 in De Paula et al. (2020)) runs have been performed for each combination of network, method and sample size. This, unfortunately, makes the results less reliable as they depend on the generated data sets. Nonetheless, since the same random seed has been used for each simulation, the data sets are identical across various methods and consequently the comparison is valid. Because the relation between the outcomes of the three methods is mostly constant, I only analyse the results for $T = 150$.

## 5.2 Comparison of the methods

Even though the three techniques have very similar performances, each method has its strengths and weaknesses on which I elaborate here. SCAD mostly outperforms the other two methods according to the mean absolute deviation of the estimated social interactions matrix, meaning that on average it comes closest to the true values. Moreover, it always has the highest fraction of true non-zeros. Although it discovers only a couple more strong links, it is significantly better at recovering weaker ones. However, SCAD always falsely finds many more weak edges than the number present in the network. For example, for Village network, SCAD defines on average 650 links as weak which is more than 17 times as many as other methods find. Out of those links only around 2.15% are truly non-zero. This explains why the percentage of correctly classified zeros is slightly lower for SCAD. Nonetheless, the hit rate for weak edges is even higher for SCAD than for the other two methods.

On the other hand, Adaptive LASSO and Adaptive Elastic Net often estimate too few weak links which is especially visible for the Watts Strogatz network. For this network, the number of edges categorised as weak corresponds to only about 11% of the true number for Adaptive LASSO and to 15% for Adaptive Elastic Net. Next to that, usually none or one of these are correct. A similar tendency is observed for the Village network. From this, it can be concluded that the entries of social interaction matrix are excessively penalised by Adaptive LASSO and Adaptive Elastic Net and, consequently, the weak links are pushed towards zero.

With regards to discovering the nodes with the highest out-degree, the given methods are very similar. For the Erdos-Renyi and Village network, they all identify two out of three nodes correctly. For Political Party Network, Adaptive Elastic Net and SCAD are slightly better since they discover all nodes while Adaptive LASSO finds two. Lastly, for the Watts Strogatz network, none of the nodes with the highest out-degree are found. This is reasonable since by construction the nodes in this network have initially the same popularity which is slightly adjusted through random rewiring.

Next, I discuss the quality of the *"post-"* estimates of the parameters $\beta$, $\rho$ and $\gamma$. For Adaptive LASSO and Adaptive Elastic Net, $\hat{\beta}$ always approaches the true value. The estimated $\hat{\rho}$ is relatively accurate for Erdos-Renyi network but varies a lot for other ones. Further, the $\hat{\gamma}$ is always biased downward. These outcomes can be explained by the fact that $\rho$ and $\gamma$ are more dependent on the estimated $\hat{W}$ than $\beta$ as the latter is the coefficient of the individual's own effects. In general, the bias decreases for all three parameters with a larger $T$ which is caused by the improved estimate of $W_0$. The results for SCAD are rather surprising. The parameters which are based on $\hat{W}_{scad}$ are around ten times smaller than their true values. A possible explanation for this could be a large number of edges estimated as non-zero as in this way the effects are spread out across the individuals and, consequently, the coefficients are pushed towards zero. Also, there is no clear evidence that those estimates converge to the true values with larger $T$.

Since none of the three methods can be globally preferred based on its accuracy, it is interesting to compare their computational intensity. Although for all methods the same maximum estimation time is employed for the particle swarm algorithm, a different number of optimisations is necessary per 100 simulations. The number of optimisations is 280 for Adaptive LASSO, 460 for Adaptive Elastic Net and 420 for SCAD. Hence, despite on average slightly superior outcomes of Adaptive Elastic Net in contrast to Adaptive LASSO, the latter might be preferred for its speed. I note that those numbers result from my choices concerning the fitting of tuning parameters and the number of iterative evaluations for SCAD. A concise summary of the comparison is presented in Table 6.

Table 6: *Overview of comparison of the methods*

| Method | Adaptive LASSO | Adaptive Elastic Net | SCAD |
|---|---|---|---|
| MAD($\hat{W}$) | middle | middle | lowest |
| true zeros recovered | middle | highest | lowest |
| true non-zeros recovered | middle | lowest | highest |
| Hit rate strong links | middle | middle | middle |
| Hit rate weak links | lowest | middle | highest |
| Computation intensity | lowest | highest | middle |

[a] Averaged over networks and sample sizes.

## 5.3 Comparison to De Paula et al. (2020)

Since De Paula et al. (2020) do not report the performance metrics for the Village network, I focus on the comparison of the Erdos-Renyi and Political Party networks, for the sample size $T = 150$. For the Erdos-Renyi network, the results are very similar. However, De Paula et al. (2020) always recover significantly more non-zeros. For the Political Party network the estimates are quite close to De Paula et al. (2020) in terms of finding zero links. Also, similarly to them, both methods are less good at finding weak links as compared to strong links. Still, in this paper substantially fewer non-zeros are recovered.

Moreover, the authors always find a visible difference between Adaptive LASSO and Adaptive Elastic Net while in this implementation the two methods have approximately the same performance. Surprisingly, the results for Adaptive LASSO are superior to those found by De Paula et al. (2020). Concerning the *"post-"* estimates of $\beta$, $\gamma$ and $\rho$, for De Paula et al. (2020) the parameters converge to the true values for Adaptive Elastic Net, but are strongly biased for Adaptive LASSO. In this study, on the other hand, estimates of both techniques are similar and, apart from $\hat{\beta}$, slightly biased.

Since the values used for the tuning parameters are larger than in De Paula et al. (2020), all estimates are pushed more towards zero. This partly explains why fewer non-zeros are found. Additionally, the simplifications made increase the chance of the optimisation algorithm ending in a local minimum. Lastly, the apparent differences for Adaptive LASSO could be caused by a slightly different implementation of the algorithm by De Paula et al. (2020) as a detailed presentation is not provided.

Table 7: *Simulation results for Erdos-Renyi network, comparison to De Paula et al. (2020)*

| Method | Adaptive LASSO | | Adaptive Elastic Net | |
|---|---|---|---|---|
| results by | De Paula et al. (2020) | this paper | De Paula et al. (2020) | this paper |
| MAD($\hat{W}$) | 0.033 | 0.023 | 0.001 | 0.026 |
| true zeros[a] | 0.878 | 0.958 | 0.997 | 0.979 |
| true non-zeros | 1.000 | 0.948 | 0.962 | 0.901 |
| $\hat{\beta}$ | 0.254 | 0.392 | 0.400 | 0.387 |
| $\hat{\gamma}$ | 0.999 | 0.394 | 0.498 | 0.301 |
| $\hat{\rho}$ | 0.965 | 0.418 | 0.283 | 0.352 |

[a] Fraction of true zeros and of true non-zeros recovered correctly.

Table 8: *Simulation results for Political Party network, comparison to De Paula et al. (2020)*

| Method | Adaptive LASSO | | Adaptive Elastic Net | |
|---|---|---|---|---|
| results by | De Paula et al. (2020) | this paper | De Paula et al. (2020) | this paper |
| MAD($\hat{W}$) | 0.032 | 0.056 | 0.010 | 0.054 |
| true zeros[a] | 0.880 | 0.9700 | 0.989 | 0.970 |
| true non-zeros | 1.000 | 0.562 | 0.960 | 0.546 |
| $\hat{\beta}$ | 0.259 | 0.384 | 0.398 | 0.393 |
| $\hat{\gamma}$ | 0.999 | 0.136 | 0.463 | 0.114 |
| $\hat{\rho}$ | 0.979 | 0.379 | 0.223 | 0.329 |

[a] Fraction of true zeros and of true non-zeros recovered correctly.

# 6 Case study: tax interdependence in Western Europe

## 6.1 Theoretical framework

After discovering the theoretical properties of the different estimation methods, I now apply these techniques in a real-world setting. Similarly to De Paula et al. (2020) I focus on the topic of tax competition, however, on an international level. The interconnections between the tax rates of different political institutions are usually explained by two well-known theories, the theory of factor mobility and the theory of yardstick competition. The former one was introduced by Tiebout (1956) and states that, given a certain level of public goods provision, if the mobility of capital and labour is enabled, the individuals and/or firms will migrate to the region with more favourable tax rates. The theory of yardstick competition, on the other hand, says that during the elections voters compare the taxes with the neighbouring jurisdictions which affects the tax setting by politicians (Shleifer, 1985).

Initially, those theories mainly referred to competition among the regions of the same country since the costs of obtaining information about and moving to other regions are relatively low. However, the recent political and economic integration enables to extend the notion of tax competition to an international level. The most trivial example of a high level of integration is the European Union. Its history began with the establishment of European Coal and Steel Community (ECSC) in 1951. Since then, the goal of creating a single market was pursued through various treaties such as the Treaty of Rome in 1957, which established the European Economic Community, and the Maastricht Treaty in 1992. Moreover, in 1985 a Schengen Agreement was signed by several European countries which created a border-free area thereby largely facilitating the labour mobility.

In the case study I conduct, I investigate how the statutory corporate income tax rates are interrelated among fifteen countries in Western European. The reasons to focus on the statutory corporate income tax are the following. First of all, literature shows that tax interdependence is stronger for mobile factors (Tiebout, 1956; Besley, Griffith, & Klemm, 2001). Secondly, in the past decades, the corporate income tax rates have been decreasing for all countries which can be seen from Figure 3 in Appendix C. Also, the ratio of the corporate income tax rate to effective labour tax rate declined over time for the OECD member countries which includes the countries in this study (Bretschger & Hettich, 2002). This suggests that taxation of corporations has been affected the most by globalisation. Next to that, the corporate income tax rates are relevant for the profit shifting by the firms and for attracting new business by the governments and it is easily observed by both.

Multiple studies find evidence that European countries influence each others' corporate income tax rates (Redoano, 2014; Davies & Voget, 2008). Moreover, Cassette and Paty (2008) show that the effects differ for Western Europe as compared to Central and Eastern Europe. In particular, they find the most profound dependence for Western Europe which is the reason why I limit my analysis to this group. The existing research uses postulated interaction matrix $W_0$ by deriving it from geographic and/or economic proximity. This case study differs in that I estimate the links directly from the data and then investigate their possible origin. In this way, biases caused by invalid interaction matrices are avoided.

## 6.2 Methods

### 6.2.1 Estimation

To estimate the interaction matrix as precisely as possible, I extend the model presented in section 2 by including multiple characteristics of the countries. I assume that the $W_0$ does not differ for various characteristics, but that parameters $\beta$ and $\gamma$ do. Consequently, the adjusted model extends the equation 4 in the following way:

$$(I - H)y_t^* = (I - H)(I - \rho_0 W_0)^{-1} \left( \sum_{k=1}^{K} (\beta_{0,k} I + \gamma_{0,k} W_0) x_{t,k}^* \right) + \nu_t, \tag{14}$$

with $K$ being the number of covariates. For this application, $y_t^*$ and $x_{t,k}^*$ are again obtained by subtracting the mean of the corresponding variable (outcome or characteristic) over time to account for individual fixed effects. Furthermore, the $g_{NT}(\theta)$ in GMM objective function (see equation 5) becomes a $N^2 \times K$ dimensional vector defined as $g_{NT}(\theta) = (I - H) \sum_{t=1}^{T} [x_{1t,1}^* \nu_t(\theta)' ... x_{Nt,K}^* \nu_t(\theta)']'$. The residuals $\nu_t(\theta)$ are defined similarly as before but using the equation 14.

To improve the model fit, I enlarge the set of tuning parameters for the case study. I use the set $\{0, 0.025, 0.05, 0.1, 0.15, 0.2, 0.225\}$ to construct the grid for Adaptive Elastic Net and I do not estimate Adaptive LASSO because zero is included in this set. For SCAD, the parameter $\lambda$ for SCAD is chosen from $\{3.25, 3, 2.75, 2.5, 2.25, 2, 1.75, 1.5, 1.25, 1, 0.75, 0.5\}$. Furthermore, I set the number of initial conditions to 10, adjust the maximum time for the Particle Swarm Optimisation to 300 seconds and perform five instead of three iterations of optimisation for SCAD. Lastly, I note no *"post-"* estimation of the parameters is used due to additional complexity caused by multiple covariates.

For the interpretation of the outcomes, I consider a link as non-zero if it is larger or equal to 0.1 and as strong if it is larger or equal to 0.7. It may occur that the links used for the row-sum normalisation are estimated as negative. In this case, I set them to zero to avoid falsely classifying relations that are not present.

### 6.2.2 Interpretation of the links

To determine the main sources of interrelated corporate income taxes among the Western European countries, I make use of the Logit model. For this, I first transform the estimated links into a binary variable and next construct a linear probability model of the form:

$$\hat{W}_{ij} = \lambda_0 + \sum_{l=1}^{L} \lambda_l X_{ij,l} + \sum_{s=L+1}^{S} \lambda_s X_{j,s}, \tag{15}$$

where $X_{ij,l}$ reflects a certain proximity $l$ of the pair of states $i, j$ and $X_{j,s}$ corresponds to some characteristic of state $j$. Since the sample consists of 15 states, $15 \cdot (15 - 1) = 210$ observations are used for the Logit model. The estimated parameters are interpreted based on the Log odds ratio which can be written as follows:

$$log\left(\frac{Pr(W_{ij} = 1|X_{ij}, X_j)}{Pr(W_{ij} = 0|X_{ij}, X_j)}\right) = \lambda_0 + \sum_{l=1}^{L} \lambda_l X_{ij,l} + \sum_{s=L+1}^{S} \lambda_s X_{j,s}. \tag{16}$$

Here, $Pr(W_{ij} = 1|X_{ij}, X_j)$ is the probability of a link being present and $Pr(W_{ij} = 0|X_{ij}, X_j)$ the probability of a link being absent.

## 6.3 Data

The analysis is performed using annual data for the fifteen countries in Western Europe for 38 years, from 1981 up to and including 2018. The sources and brief description of the data are summarised in Table 21 in Appendix C. As mentioned earlier, the outcomes $y_{it}$ for the estimation are the corporate income tax rates in each country in a given year. The characteristics $x_{it,k}$, which I include for estimation of $W_0$, are the fraction of the population in non-working age (the dependency ratio), the total tax revenues as a percentage of GDP, the unemployment rate, the population density and the net inflows of Foreign Direct Investment (FDI) as a percentage of GDP. The variables are chosen in such a way that the social-economic characteristics are representative for public expenditures and economic characteristics reflect the attractiveness of the country for businesses. I note that for Luxembourg the data on FDI is not available for first the 21 years which I solve by replacing it with the mean across other countries for the respective years. Since in this earlier period there are little differences in FDI among the countries, this should not affect the validity of outcomes.

Table 22 in Appendix C depicts the mean and standard deviation of the variables used. It can be observed that there is a lot of variation. Concerning the corporate income tax rate, the countries for which its average is the highest are Belgium, France and Italy. It is usually around ten percentage points lower for Finland, Ireland and Luxembourg. Next to that, one can see on GDP per capita and unemployment that there is a discrepancy between northern and southern European countries.

To interpret the estimated links, I construct three types of variables. Firstly, the variables reflecting the geographic closeness. These are a dummy for geographical neighbours and the distance between the capitals of the corresponding pair of countries.[8] Next, variables representing economical and social proximity. I derive these from the absolute values of the differences in GDP per capita and in proportion of the population in non-working age, both calculated using the country's average over time. Thirdly, I use a dummy variable for the countries being unofficially considered as tax havens.[9] Within the given group of countries, these are Ireland, Luxembourg and The Netherlands (Sawulski, 2020; Cassette & Paty, 2008). The choice is confirmed by the fact that for these countries the FDI corresponds to a substantially larger fraction of GDP than for others (see Table 22 in Appendix C). Moreover, Ireland, Luxembourg and The Netherlands have favourable tax regulations for

[8] I include Belgium, France, Ireland and The Netherlands as neighbours of Great Britain.

[9] For $\hat{W}_{ij}$, $x_{j,s}$ is set to one if country $j$ is known as tax haven.

certain types of enterprises (Genschel, 2002). While in Ireland and Luxembourg corporate tax rates are the lowest, The Netherlands gives companies access to non-European tax havens such as Aruba and The Netherlands Antilles (Cassette & Paty, 2008).

### 6.4 Results

The networks estimated by both Adaptive Elastic Net and SCAD have a density of around 20% and find approximately the same number of links (see Table 9). As expected, SCAD recovers slightly more weak links than Adaptive Elastic Net. The same table also shows the estimated parameters $\rho$, $\beta$ and $\gamma$. Surprisingly, both methods find that $\rho$ is equal to zero meaning that the corporate tax rates of the countries do not influence each other directly. For countries own characteristics, the corporate income tax rate seems to increase with a higher fraction of population in non-working age and higher population density. The interpretation of the parameters corresponding to the impact of the characteristics of other countries is rather ambiguous as it differs for two methods. I note that as no *"post-"* estimation is performed, these results should be interpreted with caution.

To get further insights, the plots of the estimated networks are presented in Figure 1. The networks found by both methods do not contain isolated nodes and the relations go beyond the neighbouring countries. This suggests a high level of connectivity among the countries in Western Europe. However, when looking at the links more closely, it can be observed that the two methods lead to different outcomes. This is confirmed by Tables 10 and 11 which summarise the number of incoming and outgoing connections for each country. The only country for which the results of Adaptive Elastic Net and SCAD are very similar is The Netherlands. With the out-degree of 6 and 7 for the two methods respectively, it seems to have the highest influence on the other countries. This is in line with the fact that The Netherlands has more favourable corporate taxation policy relative to other member states. Contradicting results are obtained for Greece, Portugal and Spain. According to Adaptive Elastic Net, those countries have very little influence on the corporate income tax rates in the other states. SCAD, on the other hand, finds a high number of outgoing links for those countries.

Table 9: *Characteristics of the estimated networks of Western European countries*

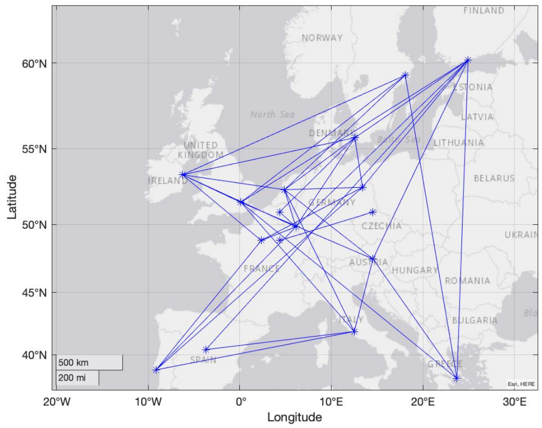
| | Adaptive Elastic Net | SCAD |
|---|---|---|
| All links | 40 | 44 |
| Strong links | 23 | 20 |
| Weak links | 17 | 24 |
| Reciprocal links | 2 | 4 |
| Density | 0.1905 | 0.2095 |
| Number of components | 2 | 4 |
| Size largest component | 14 | 12 |
| Clustering coefficient | 0.3125 | 0.2857 |
| St. Dev. diagonal of $\hat{W}^2$ | 0.2222 | 0.8338 |
| In-degree distribution | 2.6667 (0.8997) | 2.9333 (1.2799) |
| Out-degree distribution | 2.6667 (1.5887) | 2.9333 (2.4044) |
| $\hat{\rho}$ | 0.0000 | 0.0000 |
| $\hat{\beta}_1$ (POPNONWORK) | 1.0000 | 1.0000 |
| $\hat{\beta}_2$ (TAXES) | 0.0000 | 0.7863 |
| $\hat{\beta}_3$ (UNEMPL) | 0.0000 | 0.0000 |
| $\hat{\beta}_4$ (POPDENS) | 0.0737 | 0.0850 |
| $\hat{\beta}_5$ (FDI) | 0.0000 | 0.0000 |
| $\hat{\gamma}_1$ (POPNONWORK) | 0.0000 | 0.0000 |
| $\hat{\gamma}_2$ (TAXES) | 1.0000 | 0.0000 |
| $\hat{\gamma}_3$ (UNEMPL) | 0.0000 | 1.0000 |
| $\hat{\gamma}_4$ (POPDENS) | 0.0661 | 0.0000 |
| $\hat{\gamma}_5$ (FDI) | 0.0000 | 0.0000 |

Table 10: *Summary statistics for estimation results per country, Adaptive Elastic Net*

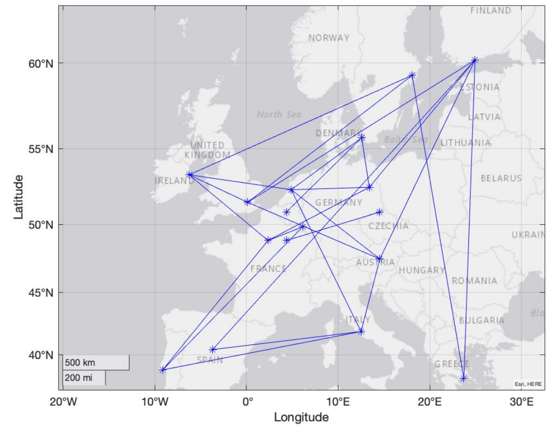
| | AUT | BEL | DNK | FIN | FRA | DEU | GRC | IRL | ITA | LUX | NLD | PRT | ESP | SWE | GBR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| In-degree | 2 | 2 | 2 | 4 | 2 | 4 | 3 | 4 | 3 | 3 | 2 | 3 | 2 | 1 | 3 |
| Out-degree | 4 | 1 | 4 | 3 | 3 | 1 | 2 | 2 | 2 | 4 | 6 | 1 | 0 | 3 | 4 |

Table 11: *Summary statistics for estimation results per country, SCAD*

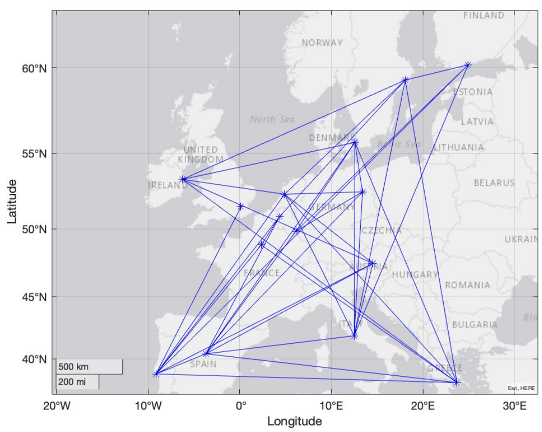
| | AUT | BEL | DNK | FIN | FRA | DEU | GRC | IRL | ITA | LUX | NLD | PRT | ESP | SWE | GBR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| In-degree | 4 | 4 | 3 | 3 | 2 | 3 | 2 | 2 | 6 | 3 | 3 | 3 | 1 | 4 | 1 |
| Out-degree | 1 | 0 | 3 | 2 | 1 | 3 | 6 | 4 | 1 | 1 | 7 | 5 | 7 | 3 | 0 |



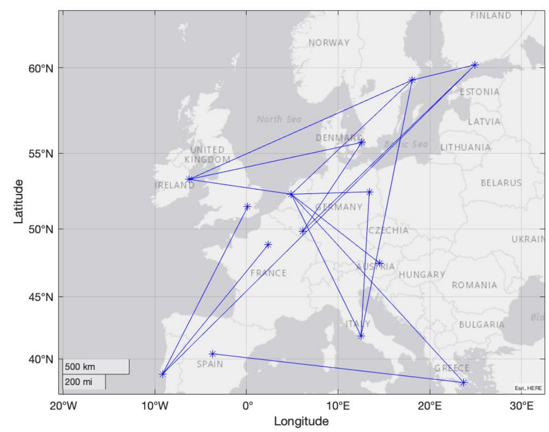

(a) Adaptive Elastic Net, all links




(b) Adaptive Elastic Net, strong links




(c) SCAD, all links




(d) SCAD, strong links

Figure 1: *The graphs of the estimated networks*

I now elaborate on the results of the Logit regressions which are displayed in Table 12. The outcomes of regression which uses the $\hat{W}$ from Adaptive Elastic Net are all insignificant even at a 10% level. When using the $\hat{W}$ from SCAD, however, all variables except for GDP homophily seem to be relevant. Hence, I will focus on discussing those results now. The negative intercept means that if all other regressors are zero, a relation of the two countries is unlikely to be present. Furthermore, on average being neighbours increases the probability of interconnected corporate income tax rates. However, the probability is also increasing with distance. This unexpected result could be caused by the fact that the distances are generally not large for the countries analysed. Therefore, the costs of moving location of a corporation within the European Union are more likely to depend on the language differences and the relative facilities for the business rather than on the geographic distance between countries. The negative coefficient for demographic homophily confirms that the corporate income tax rates might be more correlated for the countries with a similar dependency ratio. Also, the hypothesis that the states with favourable corporate tax regulations (tax havens) are more likely to influence other countries is supported.

Table 12: *Results of Logit regression with links between countries as dependent variable*

| | Adaptive Elastic Net | SCAD |
|---|---|---|
| Constant | − 0.9529 | − 2.2862*** |
| | (0.6055) | (0.6032) |
| Neighbours | − 0.1803 | 0.8685* |
| | (0.5279) | (0.5211) |
| Distance | − 0.0003 | 0.0008*** |
| | (0.0003) | (0.0003) |
| Demographic homophily | − 10.6149 | − 31.0112* |
| | (17.1690) | (16.0573) |
| GDP homophily | − 0.0000 | − 0.0000 |
| | (0.0000) | (0.0000) |
| Tax haven | 0.7309 | 1.1864** |
| | (0.4468) | (0.4651) |

Notes: *** denotes significance at 1%, ** at 5% and * at 10%.
Standard errors in parentheses.

# 7 Conclusion and discussion

## 7.1 Main simulation study

The objective of the main part of this research was to determine what are the relative strengths and weaknesses of Adaptive LASSO, Adaptive Elastic Net and SCAD in the context of estimation of network structure. In general, all methods showed similar performance and neither of them can be said to be universally the best. Firstly, each method is accurate for shrinking the estimates of redundant parameters to zero although Adaptive Elastic Net is usually slightly better than others. Furthermore, SCAD is superior in correctly finding the non-zero edges, both in absolute number and in the hit rate of recovery. However, this method also estimates substantially more edges falsely as non-zero which in a certain situation could be a strong drawback. An advantage that Adaptive LASSO has over the other two methods is the shorter estimation time.

Based on these observations, I conclude that the choice among the three techniques depends on the researcher's priority. For his/her goal could be either maximising the number of links found or minimising the amount of falsely discovered links. For example, for the networks similar to Erdos-Renyi it is reasonable to prefer Adaptive LASSO or Adaptive Elastic Net as the intention is to correctly discover the strong links. However, if, like in Watts Strogatz network, the total number of edges is dominated by the weak links, SCAD might be a better choice.

Nonetheless, it is important to be aware of the limitations of the analysis when assessing the quality of the results. First of all, it should be noted that the implementation of the methods is simplified in a number of ways. The small number of simulation runs, few initial conditions and restricted calculation time doubtlessly increase the variability and thereby deteriorate the reliability of results. Therefore, given the availability of computation facilities, it would be valuable to conduct the study with a larger scope. Another suggestion for further research is to see whether more time-efficient optimisation algorithms could be implemented. Since the non-linearity of the GMM objective function restrains from using linear programming or algorithms which use derivatives, one should look into a different formulation of the problem. For example, the parameters of the reduced form could be estimated from which then the original parameters would be obtained.

Secondly, the performance of SCAD is strongly affected by a large number of weak edges. In particular, as the downward biased estimates of $\beta$, $\gamma$ and $\rho$are used for the BIC criterion, they influence the selection of tuning parameter $\lambda$. Therefore, a suboptimal $\lambda$ might be chosen which in turn might worsen the quality of the results. A possible solution could be setting a higher threshold for classifying a link as non-zero. Also, larger $\lambda$ could be used to increase the weight of the penalty.

A lot of improvement could be achieved by advancing the procedure for selecting the tuning penalty parameters. The most straightforward extension would be to increase the grid of parameter values and evaluate those for a larger fraction of total simulation runs. Even better results could be obtained by making use of the various cross-validation approaches, such as widely used five-fold cross-validation or generalised cross-validation. These techniques have not been applied in this paper due to time constraints.

## 7.2 Case study

In the case study, I aimed to identify how the statutory corporate income tax rates are interrelated among fifteen Western European countries. The two techniques employed, Adaptive Elastic Net and SCAD, recover quite different networks. Nonetheless, both estimate relatively dense networks with links spread evenly across Western Europe. From this, I conclude that tax competition is indeed present and its dynamics is governed by a complex combination of various factors. To make reliable conclusions, additional research is necessary.

The conducted case study also posses several limitations. First of all, as 38 years is a quite long period, the relations among countries might have changed in this time frame. Hence, it would be interesting to further investigate how the adjacency matrix evolves throughout time. Secondly, due to the restricted set of analysed countries, the possible influence of other countries is not captured by the model. An interesting extension would be to apply the study to a larger geographic area. Furthermore, it is important to acknowledge that the corporate tax rates alone do not contain complete information on how the countries compete for enterprises. Incorporating the insights about possible influential taxation policies could significantly improve the quality of results.

# References


Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013a). The diffusion of microfinance. *Science*, *341*(6144).

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013b). *The Diffusion of Microfinance.* Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/U3BIHX doi: 10.7910/DVN/U3BIHX

Besley, T., Griffith, R., & Klemm, A. (2001). Empirical evidence on fiscal interdependence in OECD countries.

Blume, L. E., Brock, W. A., Durlauf, S. N., & Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, *123*(2), 444–496.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, *150*(1), 41–55.

Bretschger, L., & Hettich, F. (2002). Globalisation, capital mobility and tax competition: theory and evidence for OECD countries. *European journal of political economy*, *18*(4), 695–716.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., & Pan, M. (2017). Using aggregated relational data to feasibly identify network structure without network data.

Caner, M., & Zhang, H. H. (2014). Adaptive Elastic Net for Generalized Methods of Moments. *Journal of Business & Economic Statistics*, *32*(1), 30–47.

Carvalho, V. M., Nirei, M., Saito, Y., & Tahbaz-Salehi, A. (2016). Supply chain disruptions: Evidence from the Great East Japan Earthquake. *Columbia Business School Research Paper*(17-5).

Cassette, A., & Paty, S. (2008). Tax competition among Eastern and Western European countries: With whom do countries compete? *Economic Systems*, *32*(4), 307–325.

Davies, R., & Voget, J. (2008). Tax competition in an expanding European Union.

De Paula, Á., Rasul, I., & Souza, P. (2020). Identifying network ties from panel data: theory and an application to tax competition. *arXiv preprint arXiv:1910.07452*.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, *5*(1), 17–60.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Genschel, P. (2002). Globalization, tax competition, and the welfare state. *Politics & society*, *30*(2), 245–275.

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347–367). Elsevier.

Jackson, M. O., Rogers, B. W., & Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, *55*(1), 49–95.

Kim, Y., Choi, H., & Oh, H.-S. (2008). Smoothly Clipped Absolute Deviation on high dimensions. *Journal of the American Statistical Association*, *103*(484), 1665–1673.

Kranton, R. E., & Minehart, D. F. (2001). A theory of buyer-seller networks. *American economic review*, *91*(3), 485–508.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, *60*(3), 531–542.

Newman, M. E., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, *99*(suppl 1), 2566–2572.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, 758–765.

OECD.Stat. (2019). *OECD statistics.* Retrieved from https://stats.oecd.org

Redoano, M. (2014). Tax competition among European countries. Does the EU matter? *European Journal of Political Economy*, *34*, 353–371.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, *116*(2), 681–704.

Sawulski, J. (2020). Tax unfairness in the european union: Towards greater solidarity in fighting tax evasion. *Sociologický časopis/Czech Sociological Review*.

Shleifer, A. (1985). A theory of yardstick competition. *The RAND journal of Economics*, 319–327.

The MathWorks, I. (2015). *Build watts-strogatz small world graph model.* Retrieved from https://nl.mathworks.com/help/matlab/math/build-watts-strogatz-small-world-graph-model.html

TheWorldBank. (2019). *World bank open data.* Retrieved from https://data.worldbank.org

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, *64*(5), 416–424.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies*, *68*(2), 261–295.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.

Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 801–817.

Zou, H. (2006). The Adaptive LASSO and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Zou, H., & Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of statistics*, *36*(4), 1509.

Zou, H., & Zhang, H. H. (2009). On the Adaptive Elastic Net with a diverging number of parameters. *Annals of statistics*, *37*(4), 1733.

# A Appendix: Estimation

In this Appendix, I describe the implementation of instrumental variable regression using peers-of-peers covariates. The procedure is similar to Bramoullé, Djebbari, and Fortin (2009) with some adjustments.

First of all, I define $y$ and $x$ as an $NT \times 1$ vectors with the outcomes and characteristics of $N$ individuals for $T$ instances respectively. Also, I combine the parameters of interest into a vector $\psi = [\rho, \gamma, \beta]'$. Then, I construct the following four $NT \times NT$ block-diagonal matrices with blocks of dimension $N \times N$:

– $B$ with blocks given by $(I - H)$;

– $W_b$ with blocks given by $(I - H)\hat{W}$;

– $W_b^2$ with blocks given by $(I - H)\hat{W}^2$;

– $K(\psi)$ with blocks given by $\hat{W}(I - \rho - \hat{W})^{-1}[(I - H)\beta I + \gamma\hat{W})$.

In the next step, I create four matrices:

– $\tilde{X} = [W_b y \;\; Bx \;\; W_b x]$, the matrix of explanatory variables;

– $S = [Bx \;\; W_b x \;\; W_b^2 x]$, the matrix of instruments for the first step;

– $P = S(S'S)^{-1}S$, the weighting matrix for the first step;

– $Z(\psi) = [K(\psi)x \;\; Bx \;\; W_b x]$, the matrix of instruments for the second step.

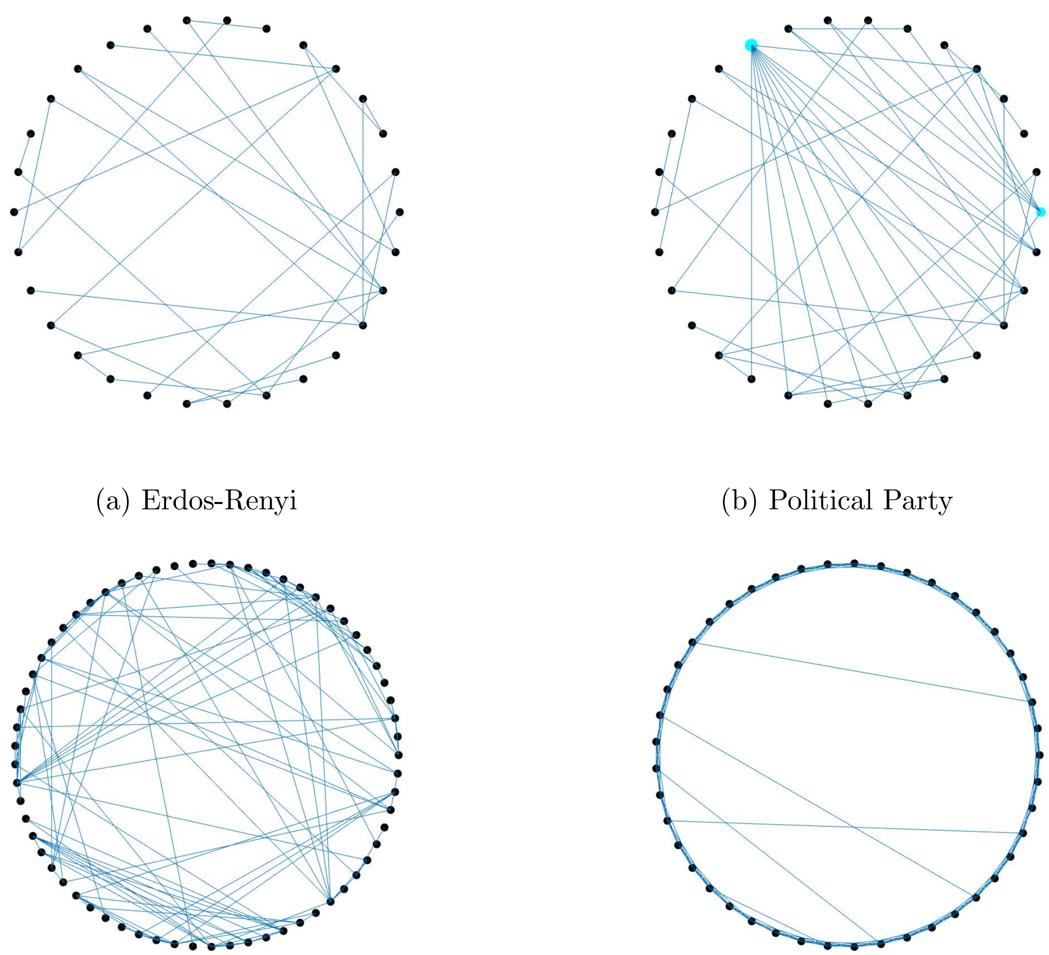
Using these, the estimate of the first step is obtained as:

$$\hat{\psi}_1 = (\tilde{X}'P\tilde{X})^{-1}\tilde{X}Py. \tag{17}$$

Furthermore, the estimate of the second and final step can be calculated as:

$$\hat{\psi}_2 = (Z(\hat{\psi}_1)'\tilde{X})^{-1}Z(\hat{\psi}_1)'y. \tag{18}$$

# B Appendix: Simulation Study



(a) Erdos-Renyi (b) Political Party

(c) Village (d) Watts Strogatz

Figure 2: *The graphs of the selected networks*

Table 13: *Results of Peers-of-Peers regression for Erdos-Renyi network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| $\hat{\beta}$ | 0.6630 | 0.3803 | 0.3924 | 0.3820 | 0.3978 | 0.3867 | 0.0381 | 0.0388 | 0.0391 |
| | (2.8668) | (0.0945) | (0.1675) | (0.0465) | (0.0752) | (0.0293) | (0.1153) | (0.1174) | (0.1369) |
| $\hat{\gamma}$ | 0.4860 | 0.1620 | 0.3938 | 0.1508 | 0.2474 | 0.3010 | 0.0244 | 0.0352 | 0.0467 |
| | (12.9025) | (0.3134) | (0.1344) | (0.1256) | (0.1855) | (0.1644) | (0.0801) | (0.1108) | (0.1417) |
| $\hat{\rho}$ | − 0.1077 | 0.3352 | 0.4180 | 0.2083 | 0.2066 | 0.3523 | 0.0498 | 0.0601 | 0.0443 |
| | (30.8550) | (0.6767) | (0.1675) | (0.2416) | (0.5117) | (0.1687) | (0.1612) | (0.1893) | (0.1369) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

Table 14: *Results of Peers-of-Peers regression for Political Party network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| $\hat{\beta}$ | 0.3682 | 0.3783 | 0.3835 | 0.4165 | 0.3860 | 0.3934 | 0.0347 | 0.0389 | 0.0399 |
| | (0.0979) | (0.0277) | (0.0220) | (0.3966) | (0.0238) | (0.1038) | (0.1058) | (0.1175) | (0.1206) |
| $\hat{\gamma}$ | 0.0917 | 0.0956 | 0.1359 | 0.1799 | 0.1140 | 0.1724 | 0.0035 | 0.0217 | 0.0278 |
| | (0.3841) | (0.1703) | (0.0962) | (1.0397) | (0.0976) | (0.2598) | (0.0271) | (0.0717) | (0.0885) |
| $\hat{\rho}$ | 0.2054 | 0.3795 | 0.3788 | − 0.0115 | 0.3285 | 0.3855 | 0.0655 | 0.0785 | 0.0767 |
| | (1.0503) | (0.3716) | (0.2611) | (2.3777) | (0.2386) | (0.6621) | (0.1058) | (0.2510) | (0.2465) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

Table 15: *Results of Peers-of-Peers regression for Village network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| $\hat{\beta}$ | 0.4134 | 0.4198 | 0.4115 | 0.4239 | 0.4167 | 0.4090 | 0.0445 | 0.0381 | 0.0385 |
| | (0.1212) | (0.0323) | (0.0172) | (0.0368) | (0.0202) | (0.0185) | (0.1226) | (0.1152) | (0.1163) |
| $\hat{\gamma}$ | 0.0969 | 0.1270 | 0.1772 | 0.0626 | 0.1302 | 0.1890 | 0.0091 | 0.0207 | 0.0280 |
| | (0.3045) | (0.1432) | (0.0716) | (0.1851) | (0.0596) | (0.0672) | (0.0378) | (0.0643) | (0.0861) |
| $\hat{\rho}$ | − 0.0527 | 0.1284 | 0.2113 | 0.0822 | 0.1609 | 0.2177 | 0.0445 | 0.0394 | 0.0413 |
| | (0.7198) | (0.3737) | (0.0956) | (0.4481) | (0.1262) | (0.1052) | (0.1503) | (0.1200) | (0.1267) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

Table 16: *Results of Peers-of-Peers regression for Watts Strogatz network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| $\hat{\beta}$ | 0.4059 | 0.4074 | 0.4005 | 0.4023 | 0.4058 | 0.4012 | 0.0359 | 0.0372 | 0.0379 |
| | (0.0444) | (0.0278) | (0.0193) | (0.0343) | (0.1016) | (0.0830) | (0.1084) | (0.1124) | (0.1143) |
| $\hat{\gamma}$ | 0.0714 | 0.1381 | 0.1845 | 0.0663 | 0.1635 | 0.2352 | 0.0195 | 0.0252 | 0.0294 |
| | (0.1907) | (0.1152) | (0.0895) | (0.1288) | (0.2606) | (0.4009) | (0.0663) | (0.0792) | (0.0894) |
| $\hat{\rho}$ | 0.1712 | 0.2362 | 0.2224 | 0.2111 | 0.1467 | 0.1209 | 0.0693 | 0.0573 | 0.0479 |
| | (0.5112) | (0.2018) | (0.2470) | (0.2827) | (0.7040) | (1.0542) | (0.2208) | (0.1763) | (0.1457) |

The reported results are averaged over 100 simulation runs. Standard deviations in parentheses.

Table 17: *Characteristics of simulated networks for Erdos-Renyi network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| Total edges[a] | 184.7000 | 72.9000 | 150.1200 | 80.4400 | 82.3800 | 67.7600 | 206.3600 | 200.9800 | 162.2100 |
| | (170.5470) | (30.6246) | (23.1945) | (39.0853) | (56.3794) | (25.9701) | (55.9790) | (51.5971) | (46.4330) |
| Strong edges[b] | 65.6300 | 46.0900 | 36.4800 | 52.6200 | 44.9400 | 38.3500 | 52.5900 | 37.2100 | 33.9100 |
| | (47.2948) | (18.1052) | (6.3524) | (27.6123) | (17.5843) | (10.2832) | (21.6865) | (2.8965) | (1.9853) |
| Weak edges | 119.0700 | 26.8100 | 113.6400 | 27.8200 | 37.4400 | 29.4100 | 153.7700 | 163.7700 | 128.3000 |
| | (130.0441) | (21.2002) | (23.3535) | (23.0245) | (42.3259) | (21.5744) | (57.1853) | (51.8706) | (46.2701) |
| Reciprocal edges | 39.0200 | 6.1600 | 16.7700 | 6.7200 | 8.6300 | 7.0300 | 30.0900 | 30.6600 | 22.5300 |
| | (71.4382) | (5.5700) | (5.2510) | (7.8626) | (15.5139) | (5.9143) | (20.7623) | (14.6558) | (11.6363) |
| Number of components | 3.2900 | 7.5400 | 1.0700 | 4.9600 | 7.5500 | 9.0300 | 1.0000 | 1.0500 | 1.1200 |
| | (5.0478) | (6.6034) | (0.3555) | (4.3622) | (7.1015) | (7.2425) | (0.0000) | (0.2190) | (0.3562) |
| Size largest component | 27.4800 | 22.5000 | 29.9300 | 25.3500 | 22.3900 | 20.6500 | 30.0000 | 29.9500 | 29.8800 |
| | (5.5368) | (7.8012) | (0.3555) | (5.3170) | (8.2498) | (8.4464) | (0.0000) | (0.2190) | (0.3562) |
| St. Dev. diagonal of $\hat{W}^2$ | 1.2274 | 0.4965 | 0.2942 | 0.8150 | 0.5377 | 0.4357 | 0.4901 | 0.2137 | 0.2340 |
| | (2.8619) | (0.6133) | (0.1031) | (2.2250) | (0.8446) | (0.4218) | (1.2301) | (0.0534) | (0.0400) |
| In-degree distribution[c] | 6.1567 | 2.4300 | 5.0040 | 2.6813 | 2.7460 | 2.2587 | 6.8787 | 6.6993 | 5.4070 |
| | (1.6998) | (1.2181) | (1.5956) | (1.3242) | (1.2215) | (1.1450) | (1.9477) | (1.9241) | (1.7179) |
| Out-degree distribution | 6.1567 | 2.4300 | 5.0040 | 2.6813 | 2.7460 | 2.2587 | 6.8787 | 6.6993 | 5.4070 |
| | (1.8053) | (1.4945) | (2.0379) | (1.4393) | (1.5642) | (1.5330) | (2.1437) | (2.3638) | (2.3014) |
| Highest out-degree[d] | {29,20,28} | {29,28,25} | {29,20,28} | {29,28,25} | {29,28,20} | {29,28,20} | {29,20,28} | {29,20,28} | {29,20,28} |

Standard deviations in parentheses. [a] All edges estimated as $> 10^{-6}$ are considered as non-zero.

[b] An edge is define as strong if its weight is >0.3. [c] For in- and out-degree distribution the mean and standard deviation in parenthesis are reported.

[d] Three nodes with highest out-degree.

Table 18: *Characteristics of simulated networks for Political Party network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| Total edges | 90.3200 | 81.8200 | 76.2000 | 91.9700 | 77.3400 | 73.1900 | 207.1000 | 206.4100 | 174.9600 |
| | (51.2693) | (52.1157) | (44.4008) | (39.0921) | (36.7616) | (45.6700) | (75.7403) | (50.6986) | (45.7536) |
| Strong edges[b] | 58.1100 | 51.6300 | 46.3300 | 58.5200 | 47.6600 | 45.2100 | 70.2300 | 41.6700 | 36.9100 |
| | (31.5646) | (33.8454) | (22.8040) | (21.9469) | (16.4981) | (27.4614) | (27.0749) | (4.2783) | (2.8073) |
| Weak edges | 32.2100 | 30.1900 | 29.8700 | 33.4500 | 29.6800 | 27.9800 | 136.8700 | 164.7400 | 138.0500 |
| | (28.0912) | (32.5824) | (27.1958) | (25.7279) | (25.5927) | (29.0315) | (73.0140) | (49.9538) | (45.8747) |
| Reciprocal edges | 7.4600 | 6.3200 | 5.6300 | 6.8500 | 4.5800 | 5.6000 | 30.1200 | 29.0500 | 21.4700 |
| | (13.1966) | (11.5521) | (10.0309) | (6.8451) | (5.9936) | (9.1398) | (34.0666) | (14.8721) | (11.4684) |
| Number of components | 4.4000 | 7.7000 | 8.6900 | 3.8900 | 7.5800 | 10.7600 | 1.0000 | 1.0000 | 1.0200 |
| | (4.9298) | (7.1088) | (7.0878) | (4.2685) | (7.0327) | (8.0405) | (0.0000) | (0.0000) | (0.1407) |
| Size largest component | 26.3400 | 22.5500 | 21.8800 | 26.7200 | 22.5400 | 19.6700 | 30.0000 | 30.0000 | 29.9800 |
| | (5.3091) | (7.8320) | (7.6571) | (5.1091) | (7.9155) | (8.5269) | (0.0000) | (0.0000) | (0.1407) |
| Clustering coefficient | − 0.0656 | 16.2453 | − 0.2961 | − 0.2485 | 0.5353 | 0.8971 | − 2.8993 | 0.0610 | 0.0647 |
| | (2.0806) | (161.4271) | (2.5633) | (1.8335) | (5.5742) | (8.3358) | (32.3519) | (0.0773) | (0.0551) |
| St. Dev. diagonal of $\hat{W}^2$ | 0.8363 | 0.7975 | 0.5748 | 0.7842 | 0.4402 | 0.5017 | 0.6922 | 0.1126 | 0.0756 |
| | (1.5832) | (2.6414) | (1.3946) | (1.0133) | (0.7542) | (1.7419) | (0.8372) | ( 0.0517) | (0.0394) |
| In-degree distribution[c] | 3.0107 | 2.7273 | 2.5400 | 3.0657 | 2.5780 | 2.4397 | 6.9033 | 6.8803 | 5.8320 |
| | (1.3539) | (1.2699) | (1.2036) | (1.3943) | (1.2447) | (1.1664) | (1.9094) | (1.8986) | (1.7411) |
| Out-degree distribution | 3.0107 | 2.7273 | 2.5400 | 3.0657 | 2.5780 | 2.4397 | 6.9033 | 6.8803 | 5.8320 |
| | (1.5536) | (1.7195) | (1.8126) | (1.5552) | (1.6600) | (1.7463) | (2.1588) | (2.5118) | (2.6354) |
| Highest out-degree[d] | {11,29,28} | {11,1,28} | {11,1,29} | {29,28,11} | {11,29,5} | {11,1,5} | {11,29,28} | {11,1,5} | {11,1,5} |

Standard deviations in parentheses. [a] All edges estimated as $> 10^{-6}$ are considered as non-zero.

[b] An edge is define as strong if its weight is $> 0.3$. [c] For in- and out-degree distribution the mean and standard deviation in parenthesis are reported.

[d] Three nodes with highest out-degree.

Table 19: *Characteristics of simulated networks for Village network*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| Total edges | 362.9000 | 132.4800 | 120.0100 | 204.0600 | 138.1100 | 120.1300 | 1277.9800 | 1053.3000 | 734.7700 |
| | (172.2815) | (33.0456) | (38.3831) | (108.1123) | (56.9199) | (56.6859) | (397.3944) | (242.1023) | (193.2740) |
| Strong edges[b] | 144.9100 | 94.9900 | 84.4000 | 124.8000 | 92.6200 | 81.7500 | 266.7700 | 92.4600 | 85.0600 |
| | (43.2869) | (14.2343) | (14.8092) | (21.2351) | (9.9258) | (15.4693) | (105.6971) | (5.2904) | (3.8868) |
| Weak edges | 217.9900 | 37.4900 | 35.6100 | 79.2600 | 45.4900 | 38.3800 | 1142.7100 | 960.8400 | 649.7100 |
| | (150.6328) | (24.9073) | (27.7769) | (100.3145) | (54.3757) | (46.5550) | (346.6761) | (242.3843) | (193.6178) |
| Reciprocal edges | 36.0800 | 10.1600 | 11.6000 | 17.6400 | 11.9700 | 11.7600 | 266.7700 | 196.3600 | 108.7200 |
| | (22.2582) | (4.7348) | (4.0651) | (15.2259) | (8.4069) | (5.5960) | (173.2975) | (70.2250) | (43.4721) |
| Number of components | 1.6900 | 17.8400 | 26.1000 | 6.5700 | 19.2200 | 29.4700 | 1.0000 | 1.0000 | 1.0000 |
| | (2.0035) | (11.1905) | (14.1611) | (5.4740) | (12.4345) | (15.8653) | (0.0000) | (0.0000) | (0.0000) |
| Size largest component | 64.1600 | 45.0900 | 33.1700 | 58.3100 | 42.7300 | 29.6000 | 65.0000 | 65.0000 | 65.0000 |
| | (2.4770) | (13.9480) | (18.5009) | (6.5885) | (15.6288) | (19.8194) | (0.0000) | (0.0000) | (0.0000) |
| Clustering coefficient | − 0.3688 | 0.0026 | 0.0581 | − 0.0876 | 0.0076 | 0.5113 | 0.0507 | 0.0418 | 0.0724 |
| | (3.7345) | (0.3370) | (0.2170) | (5.4740) | (0.3191) | (0.2566) | (3.8016) | (0.0827) | (0.0565) |
| St. Dev. diagonal of $\hat{W}^2$ | 0.9230 | 0.5296 | 0.4845 | 0.7762 | 0.5275 | 0.4820 | 0.7439 | 0.2823 | 0.2663 |
| | (0.6953) | (0.1281) | (0.1329) | (0.3003) | (0.1247) | (0.1459) | (1.6722) | (0.0735) | (0.0456) |
| In-degree distribution[c] | 5.5831 | 2.0382 | 1.8463 | 3.1394 | 2.1248 | 1.8482 | 19.6612 | 16.2046 | 11.3042 |
| | (1.9014) | (1.1689) | (1.0392) | (1.5152) | (1.1817) | (1.0005) | (3.4744) | (3.2628) | (2.8226) |
| Out-degree distribution | 5.5831 | 2.0382 | 1.8463 | 3.1394 | 2.1248 | 1.8482 | 19.6612 | 16.2046 | 11.3042 |
| | (2.0084) | (1.3290) | (1.3157) | (1.5218) | (1.3696) | (1.3215) | (3.4856) | (3.4226) | (3.0122) |
| Highest out-degree[d] | {35,54,57} | {35,40,54} | {35,57,54} | {35,54,55} | {35,57,40} | {35,57,54} | {35,64,54} | {35,57,55} | {35,57,54} |

Standard deviations in parentheses. [a] All edges estimated as $> 10^{-6}$ are considered as non-zero.

[b] An edge is define as strong if its weight is $> 0.3$. [c] For in- and out-degree distribution the mean and standard deviation in parenthesis are reported.

[d] Three nodes with highest out-degree.

Table 20: *Characteristics of simulated networks for Watts Strogatz*

| Method | Adaptive LASSO | | | Adaptive Elastic Net | | | SCAD | | |
|---|---|---|---|---|---|---|---|---|---|
| Size (T) | 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| Total edges | 103.5700 | 84.3400 | 81.7200 | 101.5700 | 87.4300 | 98.5100 | 595.3700 | 433.7400 | 316.6600 |
| | (40.9057) | (25.4805) | (32.8151) | (44.3904) | (26.4316) | (63.4940) | (82.1484) | (78.4493) | (71.6928) |
| Strong edges[b] | 69.7300 | 59.6800 | 56.5000 | 67.9300 | 61.5000 | 63.9300 | 595.3700 | 59.8100 | 56.3100 |
| | (7.8507) | (11.8235) | (13.3360) | (7.1326) | (14.8966) | (33.3810) | (4.5609) | (3.2866) | (3.3746) |
| Weak edges | 33.8400 | 24.6600 | 25.2200 | 33.6400 | 25.9300 | 34.5800 | 531.6800 | 373.9300 | 260.3500 |
| | (39.1626) | (17.7725) | (23.1868) | (42.6313) | (15.4725) | (40.5049) | (82.6403) | (78.7713) | (71.6733) |
| Reciprocal edges | 8.8000 | 6.9300 | 7.6300 | 8.9700 | 7.5100 | 10.2800 | 121.5600 | 77.3300 | 48.0800 |
| | (6.4714) | (3.4356) | (4.0218) | (8.2284) | (3.6501) | (9.6808) | (27.4463) | (23.7415) | (15.9650) |
| Number of components | 10.2400 | 16.7000 | 17.8500 | 11.0700 | 15.5800 | 15.7800 | 1.0000 | 1.0000 | 1.0100 |
| | (7.3843) | (9.9316) | (9.2030) | (7.9978) | (10.5430) | (10.6483) | (0.0000) | (0.0000) | (0.1000) |
| Size largest component | 32.7000 | 23.9200 | 21.1600 | 32.9300 | 25.3000 | 23.9400 | 45.0000 | 45.0000 | 44.9900 |
| | (10.8567) | (13.6105) | (13.0436) | (10.3653) | (14.0946) | (15.1116) | (0.0000) | (0.0000) | (0.1000) |
| Clustering coefficient | 0.0875 | 0.1220 | 0.1735 | 0.0228 | 0.0781 | 0.3275 | 0.1180 | 0.1892 | 0.2581 |
| | (0.7196) | (0.4489) | (0.3983) | (0.6880) | (0.5373) | (0.6309) | (0.0948) | (0.0860) | (0.0920) |
| St. Dev. diagonal of $\hat{W}^2$ | 0.6130 | 0.5011 | 0.4719 | 0.5739 | 0.4957 | 0.6097 | 0.1993 | 0.2109 | 0.2226 |
| | (0.2111) | (0.2137) | (0.2094) | (0.2148) | (0.2175) | (0.8314) | (0.0873) | (0.0485) | (0.0337) |
| In-degree distribution[c] | 2.3016 | 1.8742 | 1.8160 | 2.2571 | 1.9429 | 2.1891 | 13.2304 | 9.6387 | 7.0369 |
| | (1.2731) | (1.0578) | (1.0006) | (1.2615) | (1.0820) | (1.0797) | (2.8501) | (2.4586) | (2.0824) |
| Out-degree distribution | 2.3016 | 1.8742 | 1.8160 | 2.2571 | 1.9429 | 2.1891 | 13.2304 | 9.6387 | 7.0369 |
| | (1.2977) | (1.1816) | (1.1553) | (1.2943) | (1.1935) | (1.2199) | (2.8844) | (2.5202) | (2.2726) |
| Highest out-degree[d] | {31,44,27} | {31,18,38} | {31,38,18} | {31,12,27} | {31,38,7} | {31,38,18} | {14,21,31} | {31,7,12} | {31,38,37} |

Standard deviations in parentheses. [a] All edges estimated as $> 10^{-6}$ are considered as non-zero.

[b] An edge is define as strong if its weight is $> 0.3$. [c] For in- and out-degree distribution the mean and standard deviation in parenthesis are reported.

[d] Three nodes with highest out-degree.

# C Appendix: Case Study

Table 21: *Description of data on Western European countries, 1981-2018*

| Variable name | Description | Source(s) |
|---|---|---|
| CIT | Statutory corporate income tax rate | OECD.Stat |
| POPNONWORK | Fraction of non-working population (below 20 and above 65) | OECD.Stat |
| TAXES | Total tax revenues as % of GDP | OECD.Stat |
| POPDENS | Population density in people per squared kilometre | OECD.Stat |
| FDI | FDI net inflows as % of GDP | TheWorldBank |
| GDP | GDP per capita in constant US dollars | TheWorldBank |
| UNEMPL | Unemployment as % of total labour force | OECD.Stat for 1981-1990 |
| | | TheWorldBank for 1991-2018 |

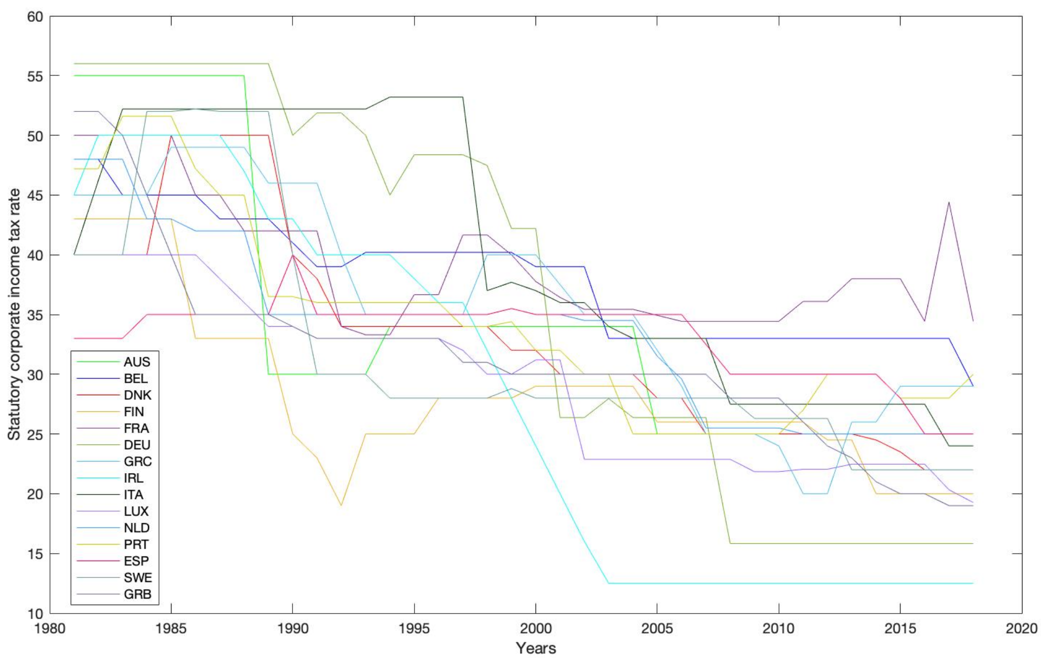



Figure 3: *The Statutory Corporate Income Tax Rate for selected countries, 1981-2018*

Table 22: *Mean and St. Dev. (in parenthesis) of the countries' characteristics, 1981-2018*

| | CIT | POPNONWORK | TAXES | POPDENS | FDI | GDP | UNEMPL |
|---|---|---|---|---|---|---|---|
| AUT | 34.5789 | 0.3916 | 41.2863 | 97.6074 | 1.8812 | $3.9873 \cdot 10^4$ | 4.8104 |
| | (11.3057) | (0.0135) | (1.4390) | (4.6811) | (5.3078) | ($0.7333 \cdot 10^4$) | (0.7728) |
| BEL | 38.0632 | 0.4034 | 43.1001 | 341.6317 | 9.1904 | $3.7766 \cdot 10^4$ | 8.4341 |
| | (5.1326) | (0.0055) | (1.1170) | (18.8490) | (13.4378) | ($0.6598 \cdot 10^4$) | (1.4739) |
| DNK | 33.0263 | 0.4036 | 45.3734 | 126.5105 | 1.7603 | $5.2075 \cdot 10^4$ | 6.5542 |
| | (8.6735) | (0.0136) | (1.9239) | (5.3927) | (4.1779) | ($0.8103 \cdot 10^4$) | (1.8157) |
| FIN | 28.3421 | 0.3995 | 42.2299 | 16.9851 | 2.3541 | $3.8396 \cdot 10^4$ | 8.7141 |
| | (6.8577) | (0.0114) | (2.5095) | (0.7072) | (3.4832) | ($0.7958 \cdot 10^4$) | (3.8518) |
| FRA | 39.2703 | 0.4190 | 42.8481 | 111.9015 | 1.5925 | $3.6194 \cdot 10^4$ | 9.1856 |
| | (5.4147) | (0.0086) | (1.5694) | (6.6940) | (1.0285) | ($0.5252 \cdot 10^4$) | (1.8082) |
| DEU | 36.5798 | 0.3849 | 35.7661 | 231.6742 | 1.5318 | $3.6825 \cdot 10^4$ | 7.0770 |
| | (16.7090) | (0.0133) | (0.9464) | (5.0472) | (2.2600) | ($0.6273 \cdot 10^4$) | (2.0307) |
| GRC | 35.7763 | 0.3978 | 29.9739 | 82.1526 | 0.8385 | $2.2471 \cdot 10^4$ | 11.6825 |
| | (8.7801) | (0.0133) | (4.8410) | (3.4039) | (0.4933) | ($0.3631 \cdot 10^4$) | (6.4257) |
| IRL | 28.1053 | 0.4311 | 30.1519 | 57.8641 | 11.7379 | $3.9492 \cdot 10^4$ | 10.5925 |
| | (15.5020) | (0.0445) | (3.5167) | (7.0772) | (15.1660) | ($1.6710 \cdot 10^4$) | (4.3935) |
| ITA | 39.9579 | 0.3936 | 39.0114 | 196.6265 | 0.7526 | $3.2983 \cdot 10^4$ | 9.6034 |
| | (11.1235) | (0.0145) | (3.7152) | (5.0391) | (0.7514) | ($0.4048 \cdot 10^4$) | (1.9572) |
| LUX | 29.2582 | 0.3765 | 36.4101 | 179.5793 | 11.5827 | $8.3571 \cdot 10^4$ | 3.5052 |
| | (7.0175) | (0.0102) | (1.8076) | (34.2980) | (23.7264) | ($2.3118 \cdot 10^4$) | (1.5707) |
| NLD | 33.6342 | 0.3906 | 37.8313 | 467.4527 | 13.0408 | $4.2913 \cdot 10^4$ | 6.0532 |
| | (7.1933) | (0.0129) | (2.2936) | (27.3770) | (19.2472) | ($0.8406 \cdot 10^4$) | (2.2581) |
| PRT | 34.4158 | 0.4103 | 29.5812 | 111.7744 | 2.8178 | $1.9088 \cdot 10^4$ | 7.4189 |
| | (8.3051) | (0.0213) | (3.3913) | (2.5761) | (2.2705) | ($0.3738 \cdot 10^4$) | (3.2321) |
| ESP | 33.0263 | 0.3960 | 31.5074 | 84.0353 | 2.4341 | $2.6094 \cdot 10^4$ | 16.9234 |
| | (3.3671) | (0.0279) | (3.0893) | (6.7846) | (1.3742) | ($0.5108 \cdot 10^4$) | (4.9001) |
| SWE | 32.1105 | 0.4202 | 45.5900 | 21.8981 | 3.2849 | $4.4062 \cdot 10^4$ | 5.9930 |
| | (9.9183) | (0.0058) | (2.1420) | (1.3161) | (4.4362) | ($0.8694 \cdot 10^4$) | (2.6275) |
| GBR | 31.5526 | 0.4146 | 32.5662 | 247.3884 | 3.5321 | $3.3878 \cdot 10^4$ | 7.2073 |
| | (8.0696) | (0.0086) | (1.6752) | (13.0252) | (2.9817) | ($0.6986 \cdot 10^4$) | (2.0608) |

# D Appendix: Codes

In this appendix, I provide the description of the MATLAB functions and scripts used to obtain the presented results. The codes are divided into two folders, "Codes Main" and "Codes Case Study". I first briefly review all the codes used for the simulation study.

***Codes Main: Functions used to create networks***

– *erdos_renyi_func.m*: creates the $W_0$ for Erdos-Renyi network for given $N$.

– *political_party_func.m*: creates the $W_0$ for Political Party network for given $N$.

– *village_func.m*: creates the $W_0$ for the Village network using the original data set.

– *WattsStrogatz.m*: creates a Watts Strogatz graph for given $N$, $k$ and $\beta$ (The MathWorks, 2015).

– *watts_strogatz_func.m*: creates the $W_0$ for Watts Strogatz network using its graph representation.

– *data_simulation_func.m*: simulates the $x^*$ and $y^*$ given $W_0$ and $T$.

***Codes Main: Functions used for estimation and evaluation***

– *theta_to_par_func.m*: transforms the parameter vector $\theta$ into $W$, $\rho$, $\gamma$ and $\beta$.

– *LASSO_initial_fun.m*: computes the starting value of Particles 1 and 2.

– *particle_initial_func.m*: computes the starting value of Particle 3.

– *gmm_func.m*: computes the GMM objective function as a function of $\theta$.

– *gmm_BIC_func.m*: computes the GMM objective function as a function of $\rho$, $\gamma$ and $\beta$.

– *gmm_l_en_func.m*: computes the GMM with LASSO or Elastic Net penalty dependent on value of $p_2$.

– *gmm_al_aen_func.m*: computes the GMM with Adaptive LASSO or Adaptive Elastic Net penalty dependent on value of $p_2$.

– *gmm_scad_func.m*: computes the GMM with SCAD penalty.

– *peers_of_peers_func.m*: computes the *"post-"* estimates of $\rho$, $\gamma$ and $\beta$.

– *BIC_func.m*: computes the BIC given the results of estimation and certain $p$ or $\lambda$.

– *density_network_func.m*: computes the density of a network.

– *network_stats_func.m*: computes a number of network characteristics (also used for the Case Study).

– *network_performance_func.m*: computes a number of performance metrics for $\hat{W}$.

– *MAD_W_func.m*: computes the mean absolute deviation of estimated W.

***Codes Main: Scripts for creating graphs***

– *graph_W0.m*: makes a graph of a given network using $W$.

– *graph_W0_pp.m*: makes a graph of Political Party network using $W$ (it marks the party leaders).

***Codes Main: Scripts for implementation***

To implement those scripts, the Parallel Computing Toolbox should be installed. Moreover, at the beginning of the scripts the sample size $T$ can be predetermined. Then, each of the scripts below implements the simulation and computes the relevant performance metrics and network characteristics.

– *al_er.m*: GMM with Adaptive LASSO penalty for Erdos-Renyi network.

– *al_pp.m*: GMM with Adaptive LASSO penalty for Political Party network.

– *al_vil.m*: GMM with Adaptive LASSO penalty for Village network.

– *al_ws.m*: GMM with Adaptive LASSO penalty for Watts Strogatz network.

– *aen_er.m*: GMM with Adaptive Elastic Net penalty for Erdos-Renyi network.

– *aen_pp.m*: GMM with Adaptive Elastic Net penalty for Political Party network.

– *aen_vil.m*: GMM with Adaptive Elastic Net penalty for Village network.

– *aen_ws.m*: GMM with Adaptive Elastic Net penalty for Watts Strogatz network.

– *scad_er.m*: GMM with SCAD penalty for Erdos-Renyi network.

– *scad_pp.m*: GMM with SCAD penalty for Political Party network.

– *scad_vil.m*: GMM with SCAD penalty for Village network.

– *scad_ws.m*: GMM with SCAD penalty for Watts Strogatz network.

***Codes Main: External data set***

Finally, for the Village network, the original data set needs to be placed in the same folder as the codes. As described in section 4, the data can be retrieved from Harvard Dataverse (Banerjee et al., 2013b). The relevant file can be then found under "Data / 1. Network Data / Adjacency Matrices" and has the following name:

– *adj_rel_HH_vilno_*10*.csv*.

Below, all codes used for the case study are reviewed. Before running them, it is important to open the file *workspace_initial.mat* as the current workspace.

***Case Study: Scripts used for data preparation***

– *data_transformation.m*: performs necessary data transformations and subtracts the mean over time for each variable.

***Case Study: Functions used for estimation and evaluation***

– *EU_gmm_func.m*: computes the GMM objective function as a function of $\theta$.

– *EU_gmm_l_en_func.m*: computes the GMM with LASSO or Elastic Net penalty dependent on value of $p_2$.

– *EU_gmm_al_aen_func.m*: computes the GMM with Adaptive LASSO or Adaptive Elastic Net penalty dependent on value of $p_2$.

– *EU_gmm_scad_func.m*: computes the GMM with SCAD penalty.

– *EU_theta_to_par_func.m*: transforms the parameter vector $\theta$ into $W$, $\rho$, $\gamma$ and $\beta$.

– *EU_LASSO_initial_func.m*: computes the starting value of Particles 1 and 2 using five characteristics.

– *EU_gmm_BIC_func.m*: computes the GMM objective function as a function of $\rho$, $\gamma$ and $\beta$.

– *EU_BIC_func.m*: computes the BIC given the results of estimation and certain $p$ or $\lambda$.

***Case Study: Scripts for implementation***

– *aen_EU.m*: GMM with Adaptive Elastic Net penalty.

– *scad_EU.m*: GMM with SCAD penalty.

– *W_hat_interpretation.m*: estimates Logit model with entries of $\hat{W}$ as dependent variable.

***Case Study: Scripts for creating graphs***

– *corporate_tax_rate_plot.m*: creates a plot of statutory corporate tax rate of each variable over time.

– *aen_all_plot.m*: creates a plot with all links estimated using Adaptive Elastic Net.

– *aen_strong_plot.m*: creates a plot with strong links estimated using Adaptive Elastic Net.

– *scad_all_plot.m*: creates a plot with all links estimated using SCAD.

– *scad_strong_plot.m*: creates a plot with strong links estimated using SCAD.