Quantification of Spillovers and Recovery of the Social Network Structure: Relaxing the Assumption of Persistence

Steven Vethman, 383672

March 20, 2019

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics, Department of Econometrics
Msc. Econometrics & Management Science: Econometrics
Supervisor: Wendun Wang
Second assessor: Yutao Sun

Abstract

Quantification of spillovers in social network analysis requires information on the social ties between agents. As information on the social structure is rarely available, recent research such as De Paula et al. (2018) and Manresa (2016) have proposed methods that can quantify spillovers and recover the social network structure. Identification depends on the assumptions of a sparse and persistent network structure. This paper relaxes the assumption of persistence by means of structural breaks. In addition, the importance of different scenarios of structural breaks is inspected with focus on whether the break affects the spillovers. Next to that, a flexible algorithm is proposed to detect time-invariant social interactions. A simulation study has suggested that additional information on which parameters pertain the break date is valuable. Relative to a baseline model that assumes all parameters alter at the break, the flexible algorithm has the advantage in break date estimation, parameter estimation and network recovery for settings with at least partial persistence of social interactions. The detection of persistent spillovers is demonstrated to benefit predominantly the identification of key players in the network. The proposed algorithms are illustrated with an empirical application to R&D spillovers in the Electronics industry. Future research is necessary for informative inference and is recommended to explore more directions of relieving the assumption of a persistent network structure.

Key words: social network analysis, sparsity, persistence, structural breaks, change point, adaptive weights, R&D spillovers.

Contents

1 Introduction

Social interactions have become recognized to influence the behavior and outcomes of agents in many contexts. For this reason, social interactions are increasingly embedded in fields of research such as crime, where identifying the well-connected individuals can help reduce crime rates (Liu et al., 2012), education, where students' performance is found to be interdependent (Calvó-Armengol et al., 2009), and innovation, where R&D spillovers between firms are substantial (Bloom et al., 2013). Social ties between agents may bring about externalities and quantifying these spillovers are vital for understanding economic phenomena and their policy implications.

Omission or disregard of these social interactions result to inconsistent results due to misspecification. To incorporate the externalities, a common assumption is that social ties between agents are known and fixed over time. Complete information is however rarely available and research has approximated ties based on measures such as distance. Recent papers such as Manresa (2016) and De Paula et al. (2018) have shown advancements by proposing methods for panel data without information on social ties that can estimate direct and spillover effects as well as retrieve the network structure. Their main assumption for identification of the social structure is that social ties are sparse and persistent over time. Sparsity imposes that few of all potential pairs between agents are connected. Persistence requires that connections between agents are time-invariant and based on long-term characteristics. The assumption of a persistent networks is often out of necessity to decrease the complexity of estimation rather than a representation of reality. Networks in many fields of research are dynamic in practice, for example in politics with elections and senators voting strategies, in biology with genetic regulatory networks, and in finance with relational patterns of stocks (Kolar et al., 2010).

In this paper, I build on recent advancements in literature and investigate relaxing the assumption of the persistence of social ties. This is examined by allowing time-variance in terms of structural breaks in the parameters in the model. Both Manresa and De Paula et al. already hint at this possible extension of their approach and their recommendation is extended by exploring the different origins of structural breaks in the panel model. This paper investigates the scenarios where changes stem from the social structure and/or the other model parameters. The investigation is twofold: First, I establish the importance of distinguishing between these scenarios. Second, I propose a model that is flexible towards these scenarios such that efficiency may be gained by identifying which components of the model are time-invariant.

This paper uses the model of Manresa (2016) as a starting point given its flexibility in network recovery and its adaptability with respect to structural breaks. Her linear panel data model is applicable for economic contexts where characteristics of agents affect their own outcome but may also affect the outcome of other agents in the sample. The model is first extended to enable simultaneous estimation of parameter coefficients and break detection. Thereafter, the model is adapted to implore whether knowledge of the scenario is valuable. Finally, by help of the group fused lasso (Tibshirani et al., 2005) and the adaptive group fused lasso (Qian & Su, 2016) a model is proposed flexible to the scenarios. A simulation study has demonstrated the relative performance of the algorithms in terms of break detection, estimation accuracy and network recovery. Results suggest that information concerning the scenario is valuable and the proposed flexible algorithm has potential for settings where (partial) time-invariance of the network structure is present. An empirical application to R&D spillovers in the United States Electronics industry il-

lustrates the proposed algorithms. The remainder of the paper is structured as follows. Section 2 contains the literature review which posits this paper in its field of research. Section 3 consists of the methodology discussing the settings, the algorithms and their building blocks. Section 4 and 5 concern the simulation study and the interpretation of the results. Section 6 consists of an empirical illustration of the algorithms related to R&D spillovers. Section 7 concludes with a conclusion and discussion.

2 Literature review

The role of social networks is increasingly studied by the econometric literature where Jackson et al. (2017) and De Paula (2017) provide a recent overview. Social interactions are often represented by use of graph theory where agents are represented as nodes and their connections as edges. The social interactions are deemed to transpire between pairs of agents (nodes) that have a connection (edge). An adjacency matrix of size N x N is often used to summarize all pairs of nodes as entries (i, j). Every element of the matrix can either have a nonzero value indicating an interaction from j to i or have a zero entry indicating no edge.

The econometric literature on social networks is divided into two main fields. This paper is located in the first that concerns economic processes on which networks play a role, i.e quantification of the spillover- or peer effects. The second concerns why and how agents link, i.e the study of network formation (for an overview see e.g. A. Chandrasekhar, 2016). This division is natural due to the high complexity when both the drivers for the links as well as their effects are to be estimated. Both fields originally treated the links as given, although information on social ties between agents is often unavailable in practice. Therefore, the literature has posed to solve this dearth via postulated ties where agents are assumed linked based on common observables. Geographical proximity is one option which is closely related to the spatial econometrics literature (see e.g. LeSage and Pace (2009) and Elhorst (2014) for an introduction and overview to spatial econometrics). Another option is that social ties relate to homophily, i.e. agents connect based on the similarity of their characteristics (e.g. Graham, 2017 or Graham, 2016). More recently, the literature has posited the use of self-reported ties for which the data set of Banerjee et al. (2013) concerning the diffusion of micro finance of 75 rural villages in India is widely applied (e.g. A. G. Chandrasekhar and Jackson (2014) and Leung (2015)).

Nonetheless, the literature has recognized that the methods implementing postulated or self-reported ties do not fully solve the problem of missing information on social networks. Postulated ties are criticized for not capturing the complete nor relevant social network structure given that links can form on a multitude of different dimensions. Even concerning a narrow topic such as micro-finance, the study of Banerjee et al. (2013) indicated that 13 different questions regarding the exchange of favors resulted in 13 substantially different social structures. Furthermore, De Paula (2017) argues that for the cost and difficulty to acquire the self reported ties the benefit in form of information is limited due to its incompleteness and censored nature. Partial sampling of all nodes (agents) may lead to measurement error and biased estimation(A. Chandrasekhar & Lewis, 2016).

To overcome these limitations recent papers have proposed models that recover the social structure from the observational panel data without network data. Manresa (2016) uses penalized linear regression on panel data where characteristics of agents affect their own outcome but may also affect the outcome of other agents in the sample. De Paula et al. (2018) employs the Adaptive Elastic Net GMM method of (Caner & Zhang, 2014)

where both the characteristics and the outcome of an agent may affect other agents. That is, in terms of Manski (1993), both exogenous and endogenous social effects are considered. Another paper that recovers the network structure from panel data is Rose (2016). His paper applies an extension of the Self-Tuning Instrumental Variable (STIV) estimator of (Gautier & Tsybakov, 2014). Similar to De Paula et al. (2018), Rose (2016) considers both exogenous and endogenous social effects. Rose (2016) states that his method is applicable for settings with a large cross-section (N) relative to the time span (T), while De Paula et al. (2018) concerns settings where $N/T \to \infty$ or $N, T \to \infty$. On the other hand, De Paula et al. (2018) argues that the auxiliary conditions of Rose (2016) are complex and hard to verify in practice. In addition, Lam and Souza (2014) propose a method for panel data that identifies exogenous and endogenous social effects by use of the adaptive lasso of (Zou, 2006). The endogenous social effects require additional requirements related to the magnitude and the authors argue that for large cross-section relative to the time dimension the performance deteriorates fast. The authors also expanded their technique for inclusion of instrumental variables in the manuscript Lam and Souza (2015).

To quantify the spillover effects while also identifying the social structure the authors rely on penalization of parameters related to the social effects. Main conditions for the identification by use of penalization are the assumptions of sparsity and persistence. To elaborate, sparsity in social network econometrics translates to every agent having a small number of influencers relative to the cross-section as well as the time dimension of the panel. The model proposed by De Paula et al. (2018) has even more stringent assumption on the influencers of each agent concerning the magnitudes of each spillover effect which are necessary to incorporate endogenous effects. The model by Manresa (2016) only considers settings for exogenous social effects and therefore has no restriction on the magnitude of the spillover effects. This paper positions itself regarding the assumption of persistence, i.e. it aims to alleviate this restriction of a time-invariant social structure. The papers of Manresa (2016) and De Paula et al. (2018) hint towards the potential of relaxing the assumption by means of sample splits, but leave the matter of break detection and the execution for further research.

In this field of econometric research, literature on time-variant network structure is still scarce. Taking a wider scope concerning network research, specifically computer science, dynamic networks have shown an increase in implementation. Kendrick et al. (2018) provide an overview of recent applications of methods for change point detection in social networks. They showcase methods that employ Bayesian change point detection analysis for detection of structural breaks in networks. Those methods detect changes by utilizing differences in metrics such as size, density or the number of communities. They find that the number of nodes and the number of links are overall the recommended choice for detecting structural changes. This notion has shaped the design of the simulation study and inclusion of these metrics may support further extension of this research. Next to that, Kolar et al. (2010) show that capturing dynamic networks has potential with machine learning techniques. Their methods build on a temporally smoothed l_1-regularized logistic regression formalism which consider either smooth or structural changes. Their application on senate voting record data has shown that realistic dynamic social interactions were estimated on top of information that static models had captured. Furthermore, Antoch et al. (2018) investigate change point detection for panel data models with a large cross-section relative to time span. In relation to this research, their approach of investigating the accuracy of break detection concerning subsets of the parameters or the cross-section is similar. To conclude, dynamic networks and their recovery have a foot-

ing in the field of computer science. This research attempts to gain ground in the same direction for the econometric field of social network analysis by exploring a relaxation of persistence in network recovery and quantification of spillovers.

3 Methodology

This section demonstrates the setting and the methods applied for this study. I start with a clarification of the general notation. Thereafter, the setting and the method by Manresa (2016) are discussed. This is followed by an introduction of the break detection technique and an elaboration of the three scenarios with the accompanying first three algorithms. After these are established, the proposed algorithm is presented with focus on how a time-variant network structure is allowed for and to what degree it is flexible to the proposed scenarios. Finally, additional features are described.

3.1 Notation

Throughout this study the following notation is applied. Bold lowercase letters denote vectors, $\boldsymbol{x}$; bold uppercase letters denote matrices, $\boldsymbol{X}$; the jth row of $\boldsymbol{X}$ is denoted as $\boldsymbol{x_j}$. The transpose of a matrix is denoted by $\boldsymbol{X'}$. Unless stated otherwise, N and T are the dimension of the cross-section and time-span considered in the panel data. I denote $i = 1, ..., N$ and $t = 1, ..., T$. I_p stands for a identity matrix of size (p,p), $O_{(q,r)}$ is a matrix of zeros of size (q,r) and $1_{(n,m)}$ is a matrix of ones of size (n,m). Moreover, $\sum_{j \neq i}$ is denoted as the short version for $\sum_{j=1, j \neq i}^{N}$. The l_1-norm is denoted as $|.|$, which for $|x|$ denotes the absolute value of x and for $|\boldsymbol{x}|$ denotes the sum of all elements of $\boldsymbol{x}$ in absolute values.

3.2 General Setting

The general setting of this paper is a linear panel model with time-variant and heterogeneous parameters. This includes the social network parameters which are pair-specific and time-dependent. This model is illustrated in (1),

$$y_{it} = \alpha_{it} + \beta_{it} x_{it} + \sum_{j \neq i} \gamma_{ij,t} x_{jt} + \boldsymbol{\delta}_t' \boldsymbol{w}_{it} + u_{it}, \tag{1}$$

where y_{it} denotes the individual outcome at time t; x_{it} denotes an individual characteristic that may generate spillovers; $\boldsymbol{w}_{it}$ is a vector of control variables and u_{it} is the disturbance term. Here, $i = 1, ..., N$ and $t = 1, ..., T$ mark the position in the cross-section and time-dimension, respectively. The fixed effects, α_{it}, are time-dependent and individual-specific; β_{it} is the time-variant coefficient capturing the heterogeneous effect of own characteristics and $\gamma_{ij,t}$ are time-variant pair-specific parameters that depict the effect of the characteristic of individual j on the outcome of individual i at time t. The parameter $\boldsymbol{\delta}_t$ captures the time-dependent effects of common controls. This setting allows for heterogeneity and time-variance in the model which creates additional challenges for estimation. To enable identification the heterogeneity and the time-variance of the parameters are restricted. In the subsequent sections the building blocks for the proposed algorithms are introduced. Their accompanying restrictions lead to the concrete setting of this paper.

3.3 Manresa (2016)

The main building block is the method for network recovery of Manresa (2016). Her method is applicable for panel data where parameters are persistent over time such that the setting for now simplifies to equation (2),

$$y_{it} = \alpha_i + \beta_i x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + \boldsymbol{\delta}' \boldsymbol{w}_{it} + u_{it}. \tag{2}$$

This is the baseline setting for the model of Manresa (2016) [1]. The social interaction structure is captured by including the characteristics of others as independent variables. The pair-specific parameters, γ_{ij}, identify the structure of the social interactions as well as quantify the spillover effects. The matrix $\boldsymbol{\Gamma}$ is defined as the matrix containing all γ_{ij} at their positions (i, j) and zeros on the diagonal and is denoted as the social interaction matrix for further reference. This is where the model of Manresa (2016) deviates from research where the social network structure is treated as given. In mathematical terms the inclusion of social effects is often stated as $\rho_0 \sum_{j \neq i} a_{ij} x_{jt}$ instead of $\sum_{j \neq i} \gamma_{ij} x_{jt}$, which in matrix notation is $\rho_0 \boldsymbol{A}\boldsymbol{X}$ instead of $\boldsymbol{\Gamma}\boldsymbol{X}$. Here, $\boldsymbol{A}$ denotes the known adjacency matrix that consists of zeros and nonzero values indicating the structure of the social interactions, while ρ_0 is estimated to capture the magnitude of spillover-effects. For Manresa (2016), a spillover effect equal to zero, $\gamma_{ij} = 0$, is also interpreted as the absence of a social tie. A spillover effect not equal to zero, $\gamma_{ij} \neq 0$, indicates both the existence of a social interaction from j to i as well as the magnitude of its social effect. The fact that connections are unspecified ex-ante circumvents misspecification of the social structure. On the other hand, the number of regressors increase linearly with the cross-section which may entail that they outnumber the periods of observation. As a consequence, the author restricts the model to settings where the number of nonzero spillovers is assumed smaller than the number of time periods in the sample. In terms of model (2) this assumption of sparsity is written as:

$$\sum_{j \neq i} \mathbb{I}\{\gamma_{ij} \neq 0\} = s_i << T \text{ for all } i = 1, ..., N. \tag{3}$$

The sparsity assumption only limits the number of sources from which an individual can receive spillovers but still allows an individual to impact any number of others in the sample. In terms of the social interaction matrix, this only constrains the row-sums while leaving the columns unrestricted. Next to that, the impact of each spillover is left unrestricted, allowing ample flexibility for $\boldsymbol{\Gamma}$ to identify the key senders and receivers by inspection of the magnitudes of the spillovers. Furthermore, the model allows the social network structure to be asymmetric, i.e. a spillover from individual j to i does not require the reverse spillover to be equal in magnitude nor that it is reciprocated, $\gamma_{ij} \neq \gamma_{ji}$. This setting of an asymmetric sparse social interaction matrix with no restriction on the magnitude of spillovers is tailored for network recovery with minimal restrictions. This level of flexibility is not attained by similar methods from De Paula et al. (2018) and Lam and Souza (2014), for their methods include endogenous effects which Manresa (2016) assumes to be absent.

Additional to the sparsity of the network, consistency of the model relies on conditions on the variation of the regressors and the error term. In this paper, I highlight the

[1] The full model of Manresa (2016) adds time-fixed effects to this setting, however, to focus on the aim of this research, these effects are omitted. Next to that, only for the simplified model without time-fixed effects she has studied the rate of convergence of the parameters.

assumptions made, however for full proofs of the consistency of the network recovery I refer to Manresa (2016).

Box 1: Assumption M1 Condition Sparse Eigenvalues

Let $\boldsymbol{M}$ *be the Gram matrix,* $\boldsymbol{M} = \frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$*, where* $\boldsymbol{X}$ *with dimensions* N x T *consists of the regressor related to the spillovers with each* $\tilde{\boldsymbol{x}}_i = (\tilde{\boldsymbol{x}}_{i1}, \ldots, \tilde{\boldsymbol{x}}_{iT})$ *in deviation from the individual-specific mean over time. The minimal and maximal m-sparse eigenvalues of* M *are defined as follows:*

$$\xi_{min}(c_s)(\boldsymbol{M}) = \min_{\boldsymbol{\delta}\in\Delta(m)} \boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta}, \quad \xi_{max}(c_s)(\boldsymbol{M}) = \max_{\boldsymbol{\delta}\in\Delta(m)} \boldsymbol{\delta}'\boldsymbol{M}\boldsymbol{\delta},$$

where

$$\Delta(\boldsymbol{M}) = \left\{\boldsymbol{\delta} \in \mathbb{R} : \sum_{j=1}^{N} \mathbb{I}\{\delta_j \neq 0\} < m, ||\boldsymbol{\delta}|| = 1\right\}.$$

For any $c > 0$*, there exist constants* $0 < \kappa_1 < \kappa_2 < \infty$ *which are independent of* N *and* T *but may depend on* c*, such that, with probability approaching* 1*, as* $N, T, s \to \infty$*, where* $s/T \to 0$*,*

$$\kappa_1 \leq \xi_{min}(c_s)(\boldsymbol{M}) \leq \xi_{max}(c_s)(\boldsymbol{M}) \leq \kappa_2.$$

The assumption in Box 1 limits the degree of cross-sectional correlation between the characteristics of different individuals. In other words, it ensures that the variation of e.g. sources $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ can be distinguished as a potential source of spillovers[2]. Further assumptions on the regressors and error term are closely related to the use of LASSO for the penalization of specific parameters in the linear panel model, see Box 2.

Finally, in the appendix of Manresa (2016) the consistency of the method's model selection is demonstrated. The efficacy of the method relies on the probability of estimating the wrong source of spillover approaching zero when spillovers are of substantial size and N and T increase with a restriction on their proportion. To illustrate, consider this reduced model where each individual has only one source of spillovers, denoted as $j(i)$:

$$y_{it} = \gamma_{ij(i)} x_{j(i)t} + u_{it}.$$

The accuracy of the estimated source of spillover relies on the fact that the characteristic of that source is more correlated to the outcome of the individual than the characteristic of any other, i.e.

$$\widehat{j(i)} = \underset{k\neq i}{\operatorname{argmax}} |\widehat{Cov}(y_{it}, x_{kt})|.$$

As a result, the probability that the estimated source of spillover, $\widehat{j(i)}$, is not the right source, $j(i)$, is equal to the probability that the maximum sample covariance between the outcome and each of the other individuals' characteristics is larger in absolute value than the covariance between y_{it} and $x_{j(i)t}$.

$$\mathbb{P}\left(\widehat{j(i)} = j(i)\right) = \mathbb{P}\left(|\widehat{Cov}(y_{it}, x_{j(i)t}| < \sup_{k\neq j(i)} |\widehat{Cov}(y_{it}, x_{kt}|\right) \tag{4}$$

[2] Manresa (2016) refers to Bickel, Ritov, Tsybakov, et al. (2009) for more details on the Restricted Eigenvalue Conditions

The full proof in the appendix of Manresa (2016) shows from there on that this probability approaches zero under the conditions that (i) there is limited multicollinearity of $x_{1t}, \ldots, x_{Nt}$, (ii) the estimated coefficients $\gamma_{ij(i)}$ are bounded away from zero, and (iii) $\frac{\log(N)}{T} \to 0$, i.e. N may only be larger than T to a certain extent.

Box 2: Assumptions M2 Time Series LASSO

Let $k > 0$ be a constant independent of N, T, and s.

2a. $\forall i, j : \gamma_{ij} \in \mathbf{\Gamma}$ *where* $\mathbf{\Gamma} \subset \mathbb{R}$, *is a compact set.*

The parameter space is required to be compact.

2b. $\mathbb{E}[x_{jt}^2] \leq k$, *and* $\mathbb{E}[(\frac{1}{T}\sum_{t=1}^{T} x_{jt})^2] \leq k$, $\forall j = 1, \ldots, N$.

This condition rules out non-stationary regressors and limits their time-series dependence.

2c. $\mathbb{E}[u_{it}|x_{j1}, \ldots, x_{jt}] = 0 \ \forall i, j, t$.

Requires the exogenous regressors as well as idiosyncratic shocks to be independent of past values of covariates.

2d. *There are constants* $a > 0$, $d_1 > 0$ *and a sequence* $\omega[t] \leq e^{-at^{d_1}}$ *s.t.,* $\forall i, j \in \{1, \ldots, N\}$, $\{u_{it}x_{jt}\}_t$ *and* $\{u_{it}\}$ *are strongly mixing processes with mixing coefficients* $\omega[t]$.

This condition on time-series dependence of $\{u_{it}x_{jt}\}_t$ and is met for example when $\{u_{it}\}_t$ and $\{x_{jt}\}_t$ are both mixing and independent. This assumption bounds the noise in estimation for an appropriate choice of penalty tuning parameter λ_i, as N and T approach infinity(Manresa, 2016). The use of this assumption excludes $x_{it} = y_{it-1}$, since auto-regressive models are not necessarily mixing. In the supplementary appendix, however, Manresa provides an alternative set of assumptions to allow for $x_{it} = y_{it-1}$.

2e. *There are constants* $b > 0$ *and* $d_2 > 0$ *such that* $\mathbb{P}(|u_{it}| > m) \leq e^{1-\frac{m}{b}^{d_2}}$ *for all* i, j, t, *and* $m > 0$.

The tail of u_{it} is required to decay exponentially fast.

Now that the general setting for the model of Manresa (2016) is introduced, the estimation process is demonstrated. The criterion to minimize is the following:

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{\Gamma},\boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \alpha_i - \beta_i x_{it} - \sum_{j\neq i}\gamma_{ij}x_{jt} - \boldsymbol{\delta}'\boldsymbol{w}_{it})^2 + \lambda\sum_{j\neq i}|\gamma_{ij}|\phi_{ij}, \tag{5}$$

where λ is the tuning parameter for the penalty level and ϕ_{ij} denotes the pair-specific weights. The objective function has two parts; the first involves the fit of the model in terms of the sum of squared residuals (SSR) and the second concerns penalization of the social interaction matrix, $\mathbf{\Gamma}$. In equations to follow, the sum of squared residuals per individual concerning the period $[1, T]$ is denoted as $Q^i_{[1,T]}(\alpha_i, \beta_i, \boldsymbol{\gamma_i}, \boldsymbol{\delta})$. Likewise, the SSR for the whole sample is denoted $Q_{[1,T]}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{\Gamma}, \boldsymbol{\delta})$. The penalization term is the force behind the assumption of sparsity. As the SSR reduces with the number of parameters,

γ_{ij}, nonzero, the penalty term increases with the absolute value of the nonzero elements in $\mathbf{\Gamma}$. The balancing of these opposing forces is implemented by the tuning parameter, λ, and the pair-specific adaptive weights, ϕ_{ij}. Elaboration of the tuning parameters and weights is postponed to a separate discussion in section 3.7.

For execution of the method, Manresa (2016) relies on an iterative algorithm based on the Pooled Lasso estimator, see Box 3.

Box 3: Social Network Recovery and Estimation with Manresa (2016)

1. Choose the initial values: $\boldsymbol{\delta}^0$ and $\boldsymbol{\beta}^0$. Set $m = 1$.
2. Obtain $\boldsymbol{\alpha}^{(m)}$ and $\mathbf{\Gamma}^{(m)}$ by solving the Lasso estimator for each i:

$$(\alpha_i^{(m)}, \boldsymbol{\gamma}_i^{(m)}) = \underset{\alpha_i, \boldsymbol{\gamma}_i}{\operatorname{argmin}} Q_{[1,T]}^i(\alpha_i, \beta_i^{(m-1)}, \boldsymbol{\gamma}_i, \boldsymbol{\delta}^{(m-1)}) + \lambda_i \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} \tag{6}$$

3. Update the values of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ by OLS on the whole sample.

$$(\boldsymbol{\beta}^{(m)}, \boldsymbol{\delta}^{(m)}) = \underset{\boldsymbol{\beta}, \boldsymbol{\delta}}{\operatorname{argmin}} Q_{[1,T]}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}, \mathbf{\Gamma}^{(m)}, \boldsymbol{\delta}) \tag{7}$$

4. Set m = m + 1. Go to Step 2 until convergence.

Thereafter, Manresa (2016) utilizes Post-Lasso by use of Ordinary Least Squares (OLS) on all the identified parameters different from zero. This eliminates the shrinkage bias and is superior to estimation by the Pooled Lasso Estimator under the condition that the method found the correct model specification.

3.4 Break detection

The method of Manresa (2016) is adapted for break detection based on the least-squares method, which dates back to the renowned work of Bai (1997). His work established break detection in time-series based on the fit of the multiple regression model, allowing for the break to occur in all or in some of the parameters. This paper follows more directly along the recent panel data version of Bai (1997), Baltagi et al. (2016). In equation (8), the setting for the least squares method for break detection is illustrated for panel settings with structural breaks:

$$y_{it} = \begin{cases} \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + e_{it}, & t = 1, \ldots, \tau^*, \\ \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{z}_{it}'\boldsymbol{\theta}_i + e_{it}, & t = \tau^* + 1, \ldots, T, \end{cases} \tag{8}$$

where $i = 1, \ldots, N$, $\boldsymbol{x}_{it}$ a (p x 1) vector of all independent variables and $\boldsymbol{z}_{it} = \boldsymbol{R}^T \boldsymbol{x}_{it}$ denotes a (q x 1) subvector of $\boldsymbol{x}_{it}$ containing the variables considered for structural breaks. Here, $\boldsymbol{R} = (O_{q\text{x}(p-q)}, I_q)$ with $q \leq p$. The execution is described in Box 4.

Box 4: Break detection with Baltagi et al. (2016)

Let $\tau \in (1, T)$ be the candidate break dates. For every τ define matrix $\boldsymbol{X}_i^*(\tau) = (0, \ldots, 0, \boldsymbol{x}'_{i,\tau+1}, \ldots, \mathbf{x}'_{iT})'$ and rewrite model (8) in a compact matrix form:

$$\forall i \in \{1, \ldots, N\} : \ \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{X}_i^*(\tau)\boldsymbol{\theta}_i + \boldsymbol{e}_i = \mathbb{X}_i(\tau)\mathbf{b}_i + \boldsymbol{e}_i \tag{9}$$

Compute for $\forall i \in \{1, \ldots, N\}$:

$$\hat{\mathbf{b}}_i(\tau) = \begin{pmatrix} \hat{\boldsymbol{\beta}}_i(\tau) \\ \hat{\boldsymbol{\theta}}_i(\tau) \end{pmatrix} = [\mathbb{X}_i(\tau)'\mathbb{X}_i(\tau)]^{-1}\mathbb{X}_i(\tau)'\mathbf{y}_i \tag{10}$$

$$SSR_i(\tau) = [\mathbf{y}_i - \mathbb{X}_i(\tau)\hat{\mathbf{b}}_i(\tau)]'[\mathbf{y}_i - \mathbb{X}_i(\tau)\hat{\mathbf{b}}_i(\tau)] \tag{11}$$

$$\hat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}} \sum_{i=1}^{N} SSR_i(\tau). \tag{12}$$

In words, the technique detects the break date by inspecting the SSR of all estimated models concerning the potential break dates and selecting the break date corresponding to the model with minimal SSR. The models are estimated with a transformation on the independent variables such that $\boldsymbol{\theta}$ in equation (8) measures the change of the slope coefficients after the break. While Baltagi et al. (2016) concerns multiple breaks, their method is adapted and demonstrated for a single structural break, given that multiple breaks is outside the scope of this paper. Assumptions additional to those of Manresa (2016) related to the location and size of the break are necessary and presented in Box 5.

Box 5: Assumptions B Structural Break and Distubance Term

(1) *Location:* $\tau^* \in (1, T)$

The break date is an integer bounded away from the start and end of the sample-period such that there are sufficient observations to consistently estimate separate coefficients before and after the break. This condition is less restricting for settings where $\frac{N}{T} \to \infty$, however, for use of the method of Manresa (2016) this condition is indispensable.

(2) *Magnitude: Let us define the magnitude of the collective change in parameters in terms of equation (8) by:* $\chi_N = \sum_{i=1}^{N} \boldsymbol{\theta}_i'\boldsymbol{\theta}_i$. *Then,* $\chi_N \to \infty$ *and (i)* $\frac{\chi_N}{N}$ *is bounded as* $N \to \infty$*; (ii)* $\chi_N \frac{T}{N} \to \infty$ *and* $\chi_N \sqrt{\frac{T}{N}} \to \infty$ *as* $(N, T) \to \infty$.

For the setting of this paper, these conditions assure that the case of no structural break is excluded while 2(i) and 2(ii) impose an upper- and lower bound on $\chi_N \frac{T}{N}$, respectively. Hence, the change in parameters needs to be bounded and of significant size.

For this research, I undertake the strictest version of the assumptions of the disturbance term of Baltagi et al. (2016). That is, (i) the disturbances u_{it}, $i = 1, \ldots, N$, are cross-sectionally independent; (ii) u_{it} is independent of x_{it} for all i and t; (iii) u_{it} are serially uncorrelated. This paper focuses on inspecting whether relaxing the assumption of persistence has potential and leaves robust analysis concerning the disturbance term for future research. Moreover, this paper investigates the structural breaks in subsets of

all parameters. With Assumption B2 of Box 5 in mind, the authors assert that a decrease in cross-sectional observations with a common break deteriorates the accuracy of break detection. Likewise, precision of break date estimation is expected to worsen when a smaller subset of parameters are time-variant. For more specifics on the method, I refer to Baltagi et al. (2016).

3.5 Scenarios

In this section, the method of Manresa (2016) is expanded by incorporating the technique of Baltagi et al. (2016) regarding different origins of structural breaks. Without information on the origin of the break the safe option is to assume that every parameter exhibits a change. This ensures consistency of the model with the cost of potential over-parameterization. Break detection and estimation may benefit substantially from the time-invariant parameters being estimated over the full time dimension. For panel models including a social network, this rises a natural distinction whether the break pertains to the social effects. I consider three different scenarios: (I) all parameters are involved in the structural break; (II) the social interaction matrix is persistent, while the other parameters change; (III) the structural break solely concerns the social interactions.

I hypothesize that identifying these scenarios is important in terms of efficiency in estimation. Recall that the social effects in Manresa (2016) are inferred by N x $(N-1)$ candidate parameters to be penalized. Information whether social effects are time-variant is worth identifying given that it may reduce the candidate parameters from 2 x N x $(N-1)$ to N x $(N-1)$. To investigate the importance of considering the scenarios, I put forward a respective algorithm for each scenario that estimates the parameters as if the scenario is known. The algorithm belonging to Scenario (I) functions as the baseline model, which is the method applied in case of no prior knowledge with respect to the scenarios. Analysis with these algorithms demonstrates the first hypothesis of this paper related to the importance of identifying the scenarios:

1. Information regarding the scenario gives substantial advantage over the baseline model in terms of break detection, accuracy of parameter estimation and network recovery.

The consistency of the following algorithms rely on the assumptions of the building blocks. The methods accompanying the scenarios solely split the full time-dimension into sample splits such that the assumptions need be adapted in terms of the sub-periods. Recall that by use of the break detection technique of Baltagi et al. (2016) the fit of the model is utilized to observe where the model is changing. Hence, for break detection, the method of Manresa (2016) is applied in subsequent sub-periods inspecting all potential breakpoints τ in a trimmed subset of the full time dimension: $[> 1, < T]$. The trimmed subset is applied such that the number of observations in each sub-period is sufficient for consistent estimation of Manresa (2016). Using the trimmed subset for τ, let the smallest sub-period considered be denoted by $\tilde{T}$. As a result, the assumptions of sparsity is altered to:

$$\sum_{j \neq i} \mathbb{I}\{\gamma_{ij} \neq 0\} = s_i << \tilde{T} \text{ for all } i = 1, ..., N, \tag{13}$$

and as such the consistency of network recovery requires adaptation of the third condition to: $\frac{\log(N)}{\tilde{T}} \to 0$.

From this point forward, I adapt the heterogeneous $\boldsymbol{\beta}$ to a homogeneous β such that the cross-section can be utilized for their estimation in the update step displayed in equation (7). Heterogeneity in direct effects is still considered as an extension. The baseline setting of this paper then boils down to equation (14).

$$y_{it} = \begin{cases} \alpha_i + \beta x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + \boldsymbol{\delta}' \boldsymbol{w}_{it} + u_{it}, & t = 1, \dots, \tau^*, \\ a_i + b x_{it} + \sum_{j \neq i} g_{ij} x_{jt} + \boldsymbol{d}' \boldsymbol{w}_{it} + u_{it}, & t = \tau^* + 1, \dots, T. \end{cases} \tag{14}$$

where β denotes the direct effects for the sub-period $[1,\ \tau]$ and b for $[\tau + 1,\ T]$, likewise for individual fixed effects $\boldsymbol{\alpha}$ and $\boldsymbol{a}$, the social interaction matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{G}$ as well as the common controls $\boldsymbol{\delta}$ and $\boldsymbol{d}$.

The objective function to minimize comprises of an addition of two criteria as described in equation (5), one related to each sub-period of the sample split, see equation (15). The weights for the sub-period after the break are denoted by f_{ij}. Box 6 presents the adapted method of Manresa (2016) to incorporate break detection, further denoted as Algorithm 1 or the baseline model. The outer framework resembles that of Box 4, where for each potential break date the SSR is collected from estimation on both sample splits and the break date associated with the minimal SSR is chosen. The inner framework resembles the block coordinate descent of Manresa (2016) with iterations between the network steps (16), (17) and the update steps (18), (19).

$$\underset{\boldsymbol{\alpha},\beta,\boldsymbol{\Gamma},\boldsymbol{\delta},\boldsymbol{a},b,\boldsymbol{G},\boldsymbol{d}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{\tau^*} (y_{it} - \alpha_i - \beta x_{it} - \sum_{j \neq i} \gamma_{ij} x_{jt} - \boldsymbol{\delta}' \boldsymbol{w}_{it})^2 + \sum_{t=\tau^*+1}^{T} (y_{it} - a_i - b x_{it} - \sum_{j \neq i} g_{ij} x_{jt} - \boldsymbol{d}' \boldsymbol{w}_{it})^2$$
$$+ \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} + \lambda_{2i} \sum_{j \neq i} |g_{ij}| f_{ij}. \tag{15}$$

Conditional on the fact that the break is found accurately the algorithms apply the method of Manresa to the sample splits, where the block coordinate descent on the same convex criterion ensures convergence in each step. The same statement applies to the algorithms which estimate as if the scenario is known, i.e. estimation with knowledge of the persistent parameters. Algorithm 2 corresponds to Scenario (II) which is the setting described in equation (14) with the restriction $\boldsymbol{\Gamma} = \boldsymbol{G}$. This reduces the parameters to be penalized by N x $(N-1)$. As a result, performance related network recovery is expected to be superior relative to the baseline model. Moreover, this advantage is anticipated to indirectly enhance break detection and accuracy of the other parameters. For Algorithm 3 corresponding to Scenario (III), the setting is as in equation (14) with the added restrictions $\boldsymbol{\alpha} = \boldsymbol{a}$, $\beta = b$ and $\boldsymbol{\delta} = \boldsymbol{d}$. The advantage of Algorithm 3 is mainly expected in terms of parameter estimation, given that the complexity of time-varying spillovers remains. In Scenario (III) the break only pertains to the spillovers, which I expect is harder to detect due to imperfect recovery and the shrinkage bias of LASSO. For the other scenarios, the addition of cross-sectional variation for the direct effects and common controls can be used to estimate the difference in magnitude between breaks.

I posit the execution of Algorithm 2 and 3 in Appendix B to avoid their duplicity with respect to Algorithm 1. Both algorithms are similar to Algorithm 1, except for their ex-ante knowledge of the parameters that are persistent across sub-periods. The objective criteria including the restrictions are represented in the equations (21) and (22) below.

Box 6: Algorithm 1 for Scenario (I)

For each candidate break date $\tau \in (1, T)$ perform (1) - (5).

(1) Choose an initial $\beta^{(0)}$, $b^{(0)}$ $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{d}^{(0)}$ and set m = 1.

(2) Obtain $\boldsymbol{\alpha}^{(m)}$, $\boldsymbol{a}^{(m)}$ and $\boldsymbol{\Gamma}^{(m)}$, $\boldsymbol{G}^{(m)}$ by solving Lasso for each i for both sub-periods.

$$(\alpha_i^{(m)}, \boldsymbol{\gamma}_i^{(m)}) = \underset{\alpha_i, \boldsymbol{\gamma}_i}{\operatorname{argmin}} Q^i_{[1,\tau]}(\alpha_i, \beta^{(m-1)}, \boldsymbol{\gamma}_i, \boldsymbol{\delta}^{(m-1)}) + \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} \tag{16}$$

$$(a_i^{(m)}, \boldsymbol{g}_i^{(m)}) = \underset{a_i, \boldsymbol{g}_i}{\operatorname{argmin}} Q^i_{[\tau+1,T]}(a_i, b^{(m-1)}, \boldsymbol{g}_i, \boldsymbol{d}^{(m-1)}) + \lambda_{2i} \sum_{j \neq i} |g_{ij}| f_{ij} \tag{17}$$

(3) Update β, $\boldsymbol{\delta}$, b and $\boldsymbol{d}$ by OLS splitting the full data matrix in sub-periods.

$$(\beta^{(m)}, \boldsymbol{\delta}^{(m)}) = \underset{\beta, \boldsymbol{\delta}}{\operatorname{argmin}} Q_{[1,\tau]}(\boldsymbol{\alpha}^{(m)}, \beta, \Gamma^{(m)}, \boldsymbol{\delta}) \tag{18}$$

$$(b^{(m)}, \boldsymbol{d}^{(m)}) = \underset{b, \boldsymbol{d}}{\operatorname{argmin}} Q_{[\tau+1,T]}(\boldsymbol{a}^{(m)}, b, \boldsymbol{G}^{(m)}, \boldsymbol{d})] \tag{19}$$

(4) Set m = m + 1 and return to (2) until convergence.

(5) Store the SSR related to this τ using the converged coefficients.

$$\begin{aligned} SSR(\tau) &= Q_{[1,\tau]}(\hat{\boldsymbol{\alpha}}(\tau), \hat{\beta}(\tau), \hat{\boldsymbol{\Gamma}}(\tau), \hat{\boldsymbol{\delta}}(\tau)) + Q_{[\tau+1,T]}(\hat{\boldsymbol{a}}(\tau), \hat{b}(\tau), \hat{\boldsymbol{G}}(\tau), \hat{\boldsymbol{d}}(\tau)) \\ &= SSR(\tau)_{[1,\tau]} + SSR(\tau)_{[\tau+1,T]} \end{aligned}$$

(6) The break date is estimated where SSR is minimal and parameters are estimated with Post-Lasso using OLS on all regressors corresponding to nonzero parameters.

$$\hat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}}\ SSR(\tau). \tag{20}$$

$$\underset{\boldsymbol{\alpha}, \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\Gamma}, \boldsymbol{\delta}, \boldsymbol{d}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \mathbb{I}_{\{t<\tau\}}[\alpha_i + \beta x_{it} + \boldsymbol{w}'_{it}\boldsymbol{\delta}] - \mathbb{I}_{\{t>\tau\}}[a_i + b x_{it} + \boldsymbol{w}'_{it}\boldsymbol{d}] - \sum_{j \neq i} \gamma_{ij} x_{jt})^2$$

$$+ \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij}, \tag{21}$$

$$\underset{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{G}, \boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \alpha_i - \beta x_{it} - \mathbb{I}_{\{t<\tau\}} \sum_{j \neq i} \gamma_{ij} x_{jt} - \mathbb{I}_{\{t>\tau\}} \sum_{j \neq i} g_{ij} x_{jt} - \boldsymbol{w}'_{it}\boldsymbol{\delta})^2$$

$$+ \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} + \lambda_{2i} \sum_{j \neq i} |g_{ij}| f_{ij}, \tag{22}$$

For details on execution, weights and penalty tuning parameters I refer to section 3.7.

3.6 Flexible Algorithm

The fourth algorithm proposed in this paper is flexible to the three distinguished scenarios. The flexibility is attained by using a technique similar to the Group Fused Lasso (GFL) and the adaptive group fused lasso (AGFL) (Qian & Su, 2016). The GFL was designed for problems where parameters are ordered and are penalized by l_1-norms of both the coefficients and their successive differences. The AGFL is an adaptation of the Group Fused Lasso of Tibshirani et al. (2005) with respect to temporal order designed to detect multiple common structural breaks of an unknown number. I elaborate on the AGFL, given that this paper also uses an temporal extension of the GFL. The technique starts with a setting where the parameters of interest are fully time-variant, i.e. each $t \in [1, T]$ has its time-dependent parameters:

$$y_{it} = \mu_i + \boldsymbol{\beta}_t' \boldsymbol{x}_{it} + u_{it}, \qquad i = 1, \dots, N, \ t = 1, \dots, T, \tag{23}$$

where $\boldsymbol{x}_{it}$ and $\boldsymbol{\beta}_t$ are p x 1 vectors of exogenous regressors and their corresponding coefficients, u_{it} is the error term and μ_i denote the individual fixed effects. Consequently the authors propose estimating $\boldsymbol{B} = (\boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_T')'$ with penalized least squares estimation by minimizing the following convex objective function:

$$(\boldsymbol{B}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta y_{it} - \boldsymbol{\beta}_t' \boldsymbol{x}_{it} + \boldsymbol{\beta}_{t-1}' \boldsymbol{x}_{i,t-1} \right)^2 + \lambda_1 \sum_{t=2}^{T} \dot{\boldsymbol{w}}_t ||\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}||, \tag{24}$$

where $||.||$ denotes the Frobenius norm and λ_1 denotes the tuning parameter that depends on N and T. The data-driven weights are denoted by $\dot{\boldsymbol{w}}_t$ and defined as

$$\dot{\boldsymbol{w}}_t = ||\dot{\boldsymbol{\beta}}_t - \dot{\boldsymbol{\beta}}_{t-1}||^{-2}, \tag{25}$$

where $\dot{\boldsymbol{\beta}}_{\boldsymbol{t}}$ denote preliminary estimates of $\boldsymbol{\beta}_t$. Penalization of the AGFL is applied to differences $\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}$ based on their numerical proximity to examine where the fit of the data suggest that $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1}$. Common breaks are detected where successive differences are not penalized to zero such that $\boldsymbol{\beta}_t \neq \boldsymbol{\beta}_{t-1}$. Incorporating this notion to the network parameters of the baseline model is the idea on which the proposed fourth algorithm rests. For clarity, I refer to the final criterion for Algorithm 4 in equation (26) on which is elaborated hereafter. The fourth algorithm differs from the AGFL for it does not use penalization to search for structural breaks. Instead of penalizing coefficients of subsequent time-periods, the proposed algorithm penalizes parameters concerning subsequent sub-periods. The adaptive weights and penalty are employed to detect which parameters do not change at the break. Weights are larger for parameters of successive sub-periods that have numerical proximity, given that weights are inversely related to the differences. In case the penalty pushes the coefficient for the difference in parameters to be zero, e.g. $g_{ij} - \gamma_{ij} = 0$, the estimated coefficients are equal in both sub-periods and Post-Lasso can utilize both sub-periods to estimate the spillover from j to i. This increase in variation over time can in turn lead to more accurate estimation of the persistent spillover. Note that similar to the GFL the l_1-norm is applied for penalization, given that it adapts more naturally to the method of Manresa.

The fourth algorithm focuses on relaxing the restriction of persistence while ensuring a convex objective function that converges. Initially, this paper aimed to penalize the successive differences of all parameters. In that case, the steps in block coordinate descent may minimize different criteria at each step. For this reason, it is chosen to only penalize

the parameters in the network steps. Hence, the fourth algorithms aims to detect spillovers and individual fixed effects that are persistent.

$$\underset{\boldsymbol{\alpha},\beta,\boldsymbol{\Gamma},\boldsymbol{\delta},\boldsymbol{a},b,\boldsymbol{G},\boldsymbol{d}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{\tau^*}(y_{it}-\alpha_i-\beta x_{it}-\sum_{j\neq i}\gamma_{ij}x_{jt}-\boldsymbol{\delta}'\boldsymbol{w}_{it})^2+\sum_{t=\tau^*+1}^{T}(y_{it}-a_i-bx_{it}-\sum_{j\neq i}g_{ij}x_{jt}-\boldsymbol{\delta}'\boldsymbol{w}_{it})^2$$

$$+\lambda_{1i}\sum_{j\neq i}|\gamma_{ij}|\phi_{ij}+\lambda_{2i}\sum_{j\neq i}|g_{ij}-\gamma_{ij}|\psi_{ij}+\lambda_{3i}|a_i-\alpha_i|\zeta_i, \tag{26}$$

I introduce notation for the difference in parameters of the subsequent sub-periods, similar to the break detection technique concerning θ in equation (8): $\breve{\gamma}_{ij}=g_{ij}-\gamma_{ij}$ and $\breve{\alpha}_i=a_i-\alpha_i$. Moreover, let us denote $\boldsymbol{X}'_{(i)}$ the T x $(N-1)$ matrix, which represent the transpose of $\boldsymbol{X}$ excluding the column of exogoneous regressors for agent i. $\boldsymbol{W}_{i,[1,\tau]}$ denotes the p x τ matrix of controls concerning the period $[1,\tau]$. To avoid confusion, I present the objective function again with notation in terms of $\breve{\alpha}_i$ and $\breve{\boldsymbol{\gamma}}_i$:

$$\sum_{i=1}^{N}\Bigg[\boldsymbol{y}_i-\begin{pmatrix}1_{(\tau,1)} & O_{(\tau,1)}\\ 1_{(T-\tau,1)} & 1_{(T-\tau,1)}\end{pmatrix}\begin{pmatrix}\alpha_i\\ \breve{\alpha}_i\end{pmatrix}-\begin{pmatrix}\boldsymbol{x}'_{i,[1,\tau]} & O_{(\tau,1)}\\ O_{(T-\tau,1)} & \boldsymbol{x}'_{i,[\tau+1,T]}\end{pmatrix}\begin{pmatrix}\beta\\ b\end{pmatrix}$$

$$-\begin{pmatrix}\boldsymbol{X}'_{(i)[1,\tau]} & O_{(\tau,1)}\\ \boldsymbol{X}'_{(i),[\tau+1,T]} & \boldsymbol{X}'_{(i),[\tau+1,T]}\end{pmatrix}\begin{pmatrix}\boldsymbol{\gamma}'_i\\ \breve{\boldsymbol{\gamma}}'_i\end{pmatrix}-\begin{pmatrix}\boldsymbol{W}'_{i,[1,\tau]} & O_{(\tau,p)}\\ O_{(T-\tau,p)} & \boldsymbol{W}'_{i,[\tau+1,T]}\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}\\ \boldsymbol{d}\end{pmatrix}\Bigg]^2$$

$$+\lambda_{1i}\sum_{j\neq i}|\gamma_{ij}|\phi_{ij}+\lambda_{2i}\sum_{j\neq i}|\breve{\gamma}_{ij}|\psi_{ij}+\lambda_{3i}|\breve{\alpha}_i|\zeta_i. \tag{27}$$

Here, ψ_{ij} and ζ_i denote the adaptive weights of $\breve{\gamma}_{ij}$ and $\breve{\alpha}_i$, respectively. Please see section 3.7 for details on the adaptive weights. Let the first part of this objective function concerning the fit of the model, $[\ .\ ]^2$, be denoted by: $\breve{Q}^i(\alpha_i,\breve{\alpha}_i,\beta,b,\boldsymbol{\gamma}_i,\breve{\boldsymbol{\gamma}}_i,\boldsymbol{\delta},\boldsymbol{d})$. The fourth algorithm is presented in Box 7.

The flexible algorithm can be interpreted as that it starts from the perspective of the baseline model. For this reason, Algorithm 4 and Algorithm 1 both commence at a consistent model specification when the origin of the structural break is unknown. Consequentially, Algorithm 4 aims to detect the time-invariant parameters by using the numerical proximity of their coefficients across sub-periods. Note that the efficacy of the method depends on time-varying and persistent parameters to be estimated with large and minimal differences in coefficients, respectively. Hence, Algorithm 4 is even more dependent on Assumption B2 in Box 5 related to substantial changes in magnitude at the break. Otherwise, the risk of estimating time-varying parameters as persistent deteriorates estimation. This proposed algorithm is analyzed in its capability to identify the persistent parameters and as such obtain information on the scenario, which relates to the second hypothesis:

Detection of persistent parameters advantages the flexible algorithm in terms of break detection, parameter estimation and network recovery relative to the baseline model.

The advantage of the flexible algorithm is also compared to the advantage of the algorithm corresponding to the scenario such that one can judge whether the algorithm

has "identified the scenario". Moreover, the flexible algorithm is able to be more flexible than the scenarios described in this thesis. The algorithm is flexible in terms of which parameters are time-invariant: e.g. it can detect that only the individual fixed effects across time are time-invariant and leave the rest flexible. In addition, the algorithm is flexible in terms of heterogeneous structural breaks, i.e. it can let individual j have time-invariant indirect effects, while those pertaining to individual h alter at the structure break. This reflects a statement of Bardwell et al. (2018) indicating that while not all time-series may see a change point at the exact same break date, we search for a method that encourages, but does not force, common break dates. Algorithm 4 takes a step into this direction with its detection of persistence per individual entry of $\boldsymbol{\Gamma}$.

Box 7: Algorithm 4 Flexible to the Scenarios

For each candidate break date $\tau \in (1, T)$ perform (1) - (5).

(1) Choose an initial $\beta^{(0)}$ and $b^{(0)}$ $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{d}^{(0)}$ and set m = 1.

(2) Obtain $\boldsymbol{\alpha}^{(m)}$, $\boldsymbol{a}^{(m)}$ and $\boldsymbol{\Gamma}^{(m)}$ $\boldsymbol{G}^{(m)}$ by solving Lasso for each i for the full period:

$$(\alpha_i^{(m)}, \breve{\alpha}_i, \boldsymbol{\gamma}_i^{(m)}, \breve{\boldsymbol{\gamma}}_i) = \underset{\alpha_i, \breve{\alpha}_i, \boldsymbol{\gamma}_i, \breve{\boldsymbol{\gamma}}_i}{\operatorname{argmin}} \breve{Q}_i(\alpha_i, \breve{\alpha}_i, \beta^{(m-1)}, b^{(m-1)}, \boldsymbol{\gamma}_i, \breve{\boldsymbol{\gamma}}_i, \boldsymbol{\delta}^{(m-1)}, \boldsymbol{d}^{(m-1)}) + \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} + \lambda_{2i} \sum_{j \neq i} |\breve{\gamma}_{ij}| \psi_{ij} + \lambda_{3i} |\breve{\alpha}_i| \zeta_i. \tag{28}$$

Thereafter, $\boldsymbol{a} = \boldsymbol{\alpha} + \breve{\boldsymbol{\alpha}}$ and $\boldsymbol{G} = \boldsymbol{\Gamma} + \breve{\boldsymbol{\Gamma}}$

(3) Update β, $\boldsymbol{\delta}$, b and $\boldsymbol{d}$ by OLS splitting the full data matrix into sub-periods.

$$(\beta^{(m)}, \boldsymbol{\delta}^{(m)}) = \underset{\beta, \boldsymbol{\delta}}{\operatorname{argmin}} Q_{[1,\tau]}(\boldsymbol{\alpha}^{(m)}, \beta, \boldsymbol{\Gamma}^{(m)}, \boldsymbol{\delta}) \tag{29}$$

$$(b^{(m)}, \boldsymbol{d}^{(m)}) = \underset{b, \boldsymbol{d}}{\operatorname{argmin}} Q_{[\tau+1,T]}(\boldsymbol{a}^{(m)}, b, \boldsymbol{G}^{(m)}, \boldsymbol{d}) \tag{30}$$

(4) Set m = m + 1 and return to (2) until convergence.

(5) Store the SSR related to this τ using the converged coefficients.

$$SSR(\tau) = Q_{[1,\tau]}(\boldsymbol{\alpha}(\tau), \beta(\tau), \boldsymbol{\Gamma}(\tau), \boldsymbol{\delta}(\tau)) + Q_{[\tau+1,T]}(\boldsymbol{a}(\tau), b(\tau), \boldsymbol{G}(\tau), \boldsymbol{d}(\tau))$$

(6) The break date is estimated where SSR is minimal and parameters are estimated by Post-Lasso by OLS including all regressors corresponding to nonzero parameters.

$$\widehat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}} \, SSR(\tau).$$

3.7 Tuning parameter and Adaptive Weights

Two types of parameters are utilized in the algorithms proposed to ensure that the penalty for sparsity and detecting time-invariant parameters is appropriately allocated.

1. The penalty tuning paramater: λ, which in Algorithm 3 and Algorithm 4 consists of multiple parameters: λ_1, up to, λ_3. This tuning parameter determines the level of penalization and in case of multiple λ's it can differentiate this level across different

types of parameters. In this paper, the estimation method enabled individual specific tuning parameter λ_i such that the flexibility for each agent to have a distinct number of nonzero spillovers is maintained. For computational efficiency, I limit the flexibility of the distribution of the penalty by setting all λ_i in a single estimation step equal to each other. That is, in Algorithm 3: $\lambda_{1i} = \lambda_{2i}$ and in Algorithm 4: $\lambda_{1i} = \lambda_{2i} = \lambda_{3i}$. As a result, the performance of Algorithm 3 and Algorithm 4 may be slightly limited. Related to the performance of Algorithm 4 relative to Algorithm 1, the conclusions are at least not biased in favor of the proposed method of this paper. Furthermore, the R package "glmnet" is applied for the individual LASSO estimations, which is computationally efficient for sparse problems. I utilize cross-validation of a grid of 100 lambdas in the range $[10^{-1}, 10^{6}]$, where the chosen λ_{min} minimizes the error based on the cross-validation[3]. See the vignette for more details on the execution of the network steps with "glmnet" (Friedman, Hastie, & Tibshirani, 2010).

2. The weights: $\phi_{ij}, f_{ij}, \psi_{ij}, \zeta_i$. These weights are adaptive, i.e. adjust based on coefficients estimated at the previous iteration. The weights allocate the penalty such that it is focused on the non-links and on merging the time-invariant parameters. The choice of adaptive weights also follows Manresa (2016) her intuition for network recovery, given that the sources with a larger covariance also constitute a larger coefficient in the absolute sense, guaranteed in the standardized settings utilized. The adaptive weights are also observed in De Paula et al. (2018) and fits naturally to the AGFL of Qian and Su (2016). Let m denote the current iteration in the execution such that the weights are defined as follows:

$$\phi_{ij}^{(m)} = (\gamma_{ij}^{(m-1)})^{-2} \; ; \; f_{ij}^{(m)} = (g_{ij}^{(m-1)})^{-2} \tag{31}$$

$$\psi_{ij}^{(m)} = (g_{ij}^{(m-1)} - \gamma_{ij}^{(m-1)})^{-2} = (\breve{\gamma}_{ij}^{(m-1)})^{-2} \; ; \; \zeta_i^{(m)} = (a_i^{(m-1)} - \alpha_i^{(m-1)})^{-2} = (\breve{\alpha}_i^{(m-1)})^{-2},$$

where the weights are set to a large number in the case that they concern previously estimated non-links, which would result to the undefined $\frac{1}{0}$. For the first iteration, a two stage estimation is performed to compensate for the absence of a previous iteration. The first stage executes the network retrieval step (equations (16), (17) and (28)) with all weights equal to 1. The second stage utilizes the estimated coefficients of the first stage as coefficients in place of those of a previous estimation.

4 Simulation Study

The performance of the proposed algorithms with respect to the scenarios is investigated by a simulation study. In all settings there is one break that occurs at exactly the half of the time-span. The trimmed-subset of potential break dates, $\tau \in (1, T)$, is adjusted per setting based on three factors: (1) There is a necessity to trim the set of potential break-dates given that break detection is inconsistent for too small sub-periods. (2) Minimizing the influence of the chosen subset of break-dates on break detection accuracy; this factor makes the number of potential break dates larger. (3) Computational speed of the algorithms; this factor makes the number of potential break dates smaller. These factors have led to the choice of 11 potential break dates with the true break date at its center,$[\tau^* - 5, \tau * +5]$. Settings with $T = 20$ are considered for an exception where the

[3]Take note that the random folds used for cross-validation in "glmnet" introduce a slight randomness such that convergence is not completely strict. This influence is minimized by successively fixing the chosen tuning parameters after a considerable amount of iterations.

trimmed subset contains 9 potential break dates, $[6, 14]$.

All settings utilize the following standard DGP for $i = 1, \ldots, N$:

$$y_{it} = \begin{cases} \alpha_i + \beta x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + \boldsymbol{\delta}' \boldsymbol{w}_{it} + u_{it}, & t = 1, \ldots, \tau^* \\ a_i + b x_{it} + \sum_{j \neq i} g_{ij} x_{jt} + \boldsymbol{d}' \boldsymbol{w}_{it} + u_{it}, & t = \tau^* + 1, \ldots, T \end{cases} \tag{32}$$

where for Scenario (I) all parameters are time-variant: $(\boldsymbol{\alpha}, \boldsymbol{a}) = (1, 2)$, $(\beta, b) = (1, 2)$, $(\boldsymbol{\delta}, \boldsymbol{d}) = (1, 2)$ and $\boldsymbol{\Gamma} \neq \boldsymbol{G}$. For Scenario (II), the set-up is identical except for the time-invariant network structure $\boldsymbol{\Gamma} = \boldsymbol{G}$. Scenario (III) has persistent parameters $(\alpha, a) = (1, 1)$ $(\beta, b) = (1, 1)$, $(\boldsymbol{\delta}, \boldsymbol{d}) = (1, 1)$ with solely the network structure changing at τ^*, $\boldsymbol{\Gamma} \neq \boldsymbol{G}$. The independent variables, x_{it}, w_{it} and the disturbance term u_{it} are i.i.d. $\sim N(0, 1)$ with the number of controls, $p = 1$. Next to differentiation for the scenarios, four main settings are applied which differ in N, T and the social interaction matrix.

Setting (1) Degree: $N \in (15, 30)$, $T \in (20, 50, 100)$.
Per row of the social interaction matrix ($\boldsymbol{\gamma_i}$): randomly the 'degree' number of elements (γ_{ij}) are set to 1, with the remaining elements of $\boldsymbol{\gamma_i}$ set to 0. The degrees can be high (6), medium (3) or low (1) and may differ between sub-periods. This setting is mainly utilized to investigate the relative performance for different degrees. Distinction is made between settings where the level of degree remains unchanged and those where the degree differs before and after the break. Moreover, the setting allows for analysis in case the sparsity assumption is close to violation, when $s_i = 6$ and $[1, \tau]$ consists of 6 to 14 time periods.

Setting (2) Political Party: $N \in (15, 30)$, $T \in (20, 50, 100, 200, 400)$.
The social interaction matrix concerns two communities representing two political parties of size $N/3$ and $2N/3$. This setting is similar to that of De Paula et al. (2018), where the authors define strong links and weak links as $|\gamma_{ij}| > 0.3$ and $|\gamma_{ij}| < 0.3$, respectively. Strong links (0.6 or 0.8) are established from the party leader to each member. In addition, each individual also receives a small spillover in terms of a weak link (0.2) from a random other politician. Structural breaks in the network structure changes one leader and a switch of $N/3$ politicians to the other party. Setting (2) focuses on the network recovery with regard to strong and weak links. Moreover, the performance is inspected for a broad range of N & T, with particular focus on how fast network structure recovery goes to 100% as $N/T \to 0$.

Setting (3) Village: $N = 65$, $T = 100$.
Each row there is a random strong link of magnitude 0.7. On top of that, 175 weak links are randomly distributed throughout the social interaction network. I follow De Paula et al. (2018) relating the magnitude of the weak links in row j to the number of allocated weak links in row j, i.e. $\frac{0.3}{\textit{No. of weak links in row } j}$. So if there are two weak links allocated in row j they both have a magnitude of 0.15. The setting mimics the situation of Banerjee et al. (2013) concerning a rural village in India. With corresponding levels of N and T to the study of De Paula et al. (2018), the performance of the algorithms are inspected in comparison to their findings. Note that the comparison is imperfect given that this paper additionally concerns structural breaks and their paper incorporates endogenous social effects.

Setting (4) Hierarchy: $N \in (15, 30),\ T \in (20, 50, 100)$.
Each individual has four sources of spillovers, i.e. $\boldsymbol{\gamma}_i$ contains $(\gamma_{ij}, \gamma_{ik}, \gamma_{il}, \gamma_{im}) = (2, 1.5, 1, 0.5)$, where $j, k, l, m \neq i$ and the other $N - 4$ elements of $\boldsymbol{\gamma}_i$ are 0. This setting is utilized to investigate the performance of the algorithm when it is only the change of hierarchy in influencers that causes the break while the sources of spillovers remain the same.

4.1 Performance Measures

The algorithms are evaluated in three aspects: accuracy of coefficient estimates, precision of break date estimation and accuracy of network recovery . Parameter estimation is inspected in terms of Mean Absolute Deviation (MAD). In terms of the parameters of interest, i.e.

$$\begin{aligned}
\text{A:}\ \ \text{MAD}(\boldsymbol{\alpha}, \boldsymbol{a}) &= \frac{1}{2N}\Big(|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}| + |\hat{\boldsymbol{a}} - \boldsymbol{a}|\Big), \\
\text{B:}\ \ \text{MAD}(\beta, b) &= \frac{1}{2}\Big(|\hat{\beta} - \beta| + |\hat{b} - b|\Big), \\
\text{G:}\ \ \text{MAD}(\boldsymbol{\Gamma}, \boldsymbol{G}) &= \frac{1}{2N(N-1)} \sum_{i}^{N} |\hat{\boldsymbol{\gamma}}_{\boldsymbol{i}} - \boldsymbol{\gamma}_{\boldsymbol{i}}| + |\ \hat{\boldsymbol{g}}_{\boldsymbol{i}} - \boldsymbol{g}_{\boldsymbol{i}}|, \\
\text{D:}\ \ \text{MAD}(\boldsymbol{\delta}, \boldsymbol{d}) &= \frac{1}{2p}\Big(|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}| + |\hat{\boldsymbol{d}} - \boldsymbol{d}|\Big),
\end{aligned}$$

where A,B,G and D are used for short notation in the presentation of the results and $|.|$ denotes the l_1-norm. Moreover, the break date estimates, $\hat{\tau}^*$, are assessed by their absolute distance to the true break date τ^*. That is,

$$\text{AD} = |\hat{\tau}^* - \tau^*|. \tag{33}$$

For evaluation of the recovery of the network structure, De Paula et al. (2018) suggests the percentage of true non-zeros and zeros found. Moreover, measures related to their distinction of strong and weak links provide additional information. Finally, an inspection of information from correctly established links relative to misinformation in the forms of misleading links and missing links is incorporated. In summary, the following six measures for network recovery are applied:

1. Nonlinks: The percentage of accurate recovery of zeros,
2. Links: The percentage of accurate recovery of non-zeros,
3. Overall: The overall percentage of accurate recovery of links and non-links,
4. Ratio: The ratio of accurate links over (over-identified links + missed links),
5. Strong: The accurate detection of strong links (including sign),
6. Weak: The accurate detection of weak links (including sign).

Furthermore, two explorations in the form of extensions are included in the simulation study. First, the algorithms are extended from homogeneous to heterogeneous direct

effects. The differences between the main results and this extension are examined as well as whether the relative performance of Algorithm 4 to the baseline model alters. Second, the performance of the algorithms is investigated in case the assumption of a single structural break is violated, i.e. the setting is persistent. The flexible algorithm is expected to be more resilient relative to Algorithm 1, because it may merge all individual specific parameters. Preliminary testing for a break or the number of unknown breaks is left for future research.

5 Results

Full disclosure of the simulation study results are found in Appendix A. This section commences with examinations in terms of the performance categories and is followed by discussions of additional features. Emphasis is placed upon the two hypotheses: (i) the advantage of information on the origin of the structural break (ii) the performance of Algorithm 4 relative to baseline model and the Algorithm corresponding to the scenario.

Table 1: Break Detection

Setting 1		Scenario (I)		Scenario (II)			Scenario (III)		
N	T	A1	A4	A1	A2	A4	A1	A3	A4
Low Degree									
15	20	0.29	0.45	0.52	0.28	0.45	0.93	1.08	1.43
15	50	0.09	0.12	0.14	0.03	0.15	0.77	0.76	0.76
15	100	0.03	0.04	0.07	0.01	0.06	0.45	0.50	0.57
High Degree									
15	20	0.78	0.98	1.64	0.62	1.08	1.33	1.32	1.44
15	50	0.07	0.11	0.33	0.04	0.21	0.19	0.16	0.28
15	100	0.00	0.01	0.08	0.01	0.08	0.05	0.04	0.06
Medium Degree									
15	20			0.95	0.40	0.65			
15	50			0.26	0.05	0.18			
15	100			0.08	0.02	0.07			
Setting 2		Scenario (I)		Scenario (II)			Scenario (III)		
N	T	A1	A4	A1	A2	A4	A1	A3	A4
30	20	0.39	0.47	0.44	0.27	0.38	1.60	1.60	1.89
30	50	0.21	0.24	0.25	0.05	0.21	1.61	1.69	1.69
15	20	0.43	0.48	0.53	0.25	0.48	1.53	1.70	1.83
15	50	0.20	0.22	0.19	0.05	0.17	1.92	1.79	1.79
15	100	0.12	0.15	0.08	0.02	0.09	1.55	1.51	1.53
15	200	0.06	0.08	0.06	0.02	0.04	1.23	1.28	1.43
15	400	0.05	0.05	0.03	0.02	0.04	1.15	1.03	1.16

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "A1" is short for Algorithm 1, likewise for "A2", "A3" and "A4" . Performance measures are as denoted in section 4.1

5.1 Break Detection

Table 1 contains a selected summary of the results for break detection which are here discussed per Scenario. In Scenario (I), Algorithm 4 has higher levels of AD than Algorithm 1 for settings with $T = 20, 50$, especially in Setting 1. Performance for settings with larger T is similar, notably in Setting 2 where one third of the social interaction matrix is persistent. For settings in Scenario (II), Algorithm 2 is shown to substantially have the lowest levels of AD. Overall, Algorithm 4 has higher precision of break detection than Algorithm 1 and this advantage stands out in settings with $T = 20, 50$ and medium to high degree. Table 10 in Appendix A confirms this tendency for $N = 30$. The Monte Carlo simulations suggest that the information of Scenario (III) is not valuable given that Algorithm 1 is either similar or slightly superior to Algorithm 3. Algorithm 4 performs similar to Algorithm 1 in most settings, except for settings with low degree and $T = 20$ where Algorithm 4 consistently has higher levels of AD. Furthermore, the results suggest that break detection with the same number of links in both periods is not sub-optimal with respect to break detection in settings where the number of links alter. In conclusion, information of a time-invariant network structure is valuable for break detection. Algorithm 4 is able to detect the time-invariant network structure and gains advantages for break detection but does not attain estimation as if the scenario is known. On the other hand, break detection of Algorithm 4 is sub-optimal with respect to Algorithm 1 for settings in Scenario (I) and short panels with low degree in Scenario (III).

5.2 Network Recovery

The tables in Appendix A show a comprehensive overview of the performance related to network recovery, whereas Table 2 concerning strong link detection illustrates the main tendencies in a concise format.

The performance measure of misinformation, Ratio, suggests that Algorithm 1, 2 and 3, on average, have the advantage in their corresponding scenario. Strong link detection, however, suggests that only Algorithm 2 has a great advantage over Algorithm 1 in its scenario. This observation is confirmed by the substantially lower levels of MAD($\boldsymbol{\Gamma}$,$\boldsymbol{G}$) relative to Algorithm 1, where this difference is greater than that between Algorithm 3 and Algorithm 1. Information that $\boldsymbol{\alpha}$, β and $\boldsymbol{\delta}$ are time-invariant allows Algorithm 3 to capture a more clean picture of the network, as is seen by the higher levels of Overall and Ratio. In contrast, Algorithm 1 appears just as able to detect the key roles of the social structure. Weak link detection is shown to be difficult for settings to be found in practice.

Furthermore, Table 2 suggest that in Scenario (I) and (III) the fourth algorithm detects key players as well as Algorithm 1 in most settings. However, the measures Ratio and Overall signal that in those scenarios network retrieval of Algorithm 1 is cleaner. In detail, results for Setting 1 indicate that the baseline model is superior in strong link detection for small T and low degree. However, Setting 2 demonstrates that with the addition of one weak link and partial persistence of the network this advantage is not robust. For settings with small T and high degree Algorithm 4 is even superior in network recovery in terms of MAD($\boldsymbol{\Gamma}$,$\boldsymbol{G}$), the ratio related to misinformation and identification of the strong links. I expect this result to stem from random overlap of spillovers between sub-periods due to the small cross-section. With $N = 15$, each $\boldsymbol{\gamma}_i$ and $\boldsymbol{g}_i$ contain 6 out of $(N-1) = 14$ sources of spillovers at random. The odds are that some indirect effects pertain the same pairs in both sub-periods, which the fourth algorithm uses for its advantage. Recall that the ability of Algorithm 4 to detect time-invariance in the network structure is per

pair-specific entry, γ_{ij}. This suggest that partial persistence of spillovers is sufficient to give Algorithm 4 an advantage. Moreover, in Scenario (II) the ability to detect the time-invariant social interactions grants Algorithm 4 network recovery exceeding that of Algorithm 1. For example, see Table 10 where for medium and high degree Algorithm 4 demonstrates strong link identification, ratio of information and accuracy of link detection of (almost) double size. The levels of MAD($\boldsymbol{\Gamma}$,$\boldsymbol{G}$) and Ratio point out that Algorithm 4 does not reach estimation as if the scenario is known. In terms of strong link detection, Algorithm 4 performs close to Algorithm 2 for all levels of T, where only for $T = 20$ Algorithm 2 has a substantial advantage.

Table 2: Strong link identification in (%)

Setting 1		Scenario (I)		Scenario (II)			Scenario (III)		
N	T	A1	A4	A1	A2	A4	A1	A3	A4
Low Degree									
15	20	62.4	52.0	62.2	89.9	83.3	61.0	62.4	54.4
15	50	96.8	92.9	96.9	99.9	99.4	96.5	97.5	92.9
15	100	99.9	99.8	99.9	100.0	100.0	99.9	100.0	99.7
High Degree									
15	20	33.2	41.5	34.0	76.7	62.9	32.9	34.0	43.6
15	50	92.9	92.5	92.8	99.8	98.9	92.9	94.7	92.9
15	100	99.9	99.9	99.9	100.0	100.0	99.9	99.9	99.9
Low to High Degree									
15	20	38.1	40.6				37.9	39.6	43.5
15	50	93.6	92.9				93.3	93.9	93.5
15	100	99.9	99.9				99.9	99.9	99.9
Setting 2		Scenario (I)		Scenario (II)			Scenario (III)		
N	T	A1	A4	A1	A2	A4	A1	A3	A4
30	20	30.8	31.4	34.1	65.4	55.9	29.9	30.8	34.7
30	50	76.2	75.1	79.7	96.9	94.3	75.9	76.8	75.7
15	20	41.1	41.7	44.1	72.7	63.8	39.6	38.6	43.1
15	50	81.6	81.5	85.1	97.8	95.3	80.3	81.9	81.0
15	100	97.2	97.1	98.0	99.9	99.3	97.0	97.3	96.8
15	200	99.8	99.8	99.9	100.0	100.0	99.9	99.8	99.8
15	400	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "A1" is short for Algorithm 1, likewise for "A2","A3" and "A4" . Performance measures are as denoted in section 4.1

All in all, information on the scenario allows a cleaner estimation of the social network structure, while for detection of key players mainly information that spillovers are persistent is valuable. Algorithm 4 is suggested to already benefit from a partially persistent network structure. In those settings, Algorithm 4 is superior to Algorithm 1 in network recovery but only attains estimation close to Algorithm 2 in terms of strong link identification for T bigger than 20.

5.3 Accuracy of Parameter Estimation

The tables in Appendix A show that information on persistent parameters lead to more accurate estimation of individual fixed effects. Moreover, Algorithm 1 relative to Algorithm 4 has lower levels of MAD($\boldsymbol{\alpha}, \boldsymbol{a}$) in Scenario (I) and (II) for settings with $T = 20$ and low degree. Algorithm 4 is slightly superior for settings with medium and high degree in Scenario (II). Remaining settings in Scenario (I) and (II) show similar levels of accuracy. In Scenario (III), detection of time-invariant fixed effects of Algorithm 4 leads estimation to be superior to that of Algorithm 1. Although Algorithm 4 has levels of MAD($\boldsymbol{\alpha}, \boldsymbol{a}$) that approach those of Algorithm 3, accuracy is not as if the scenario is known.

Furthermore, Algorithm 2 and 3 estimate direct effects more accurate in their corresponding scenario. As T increases, results of Setting 2 suggest that estimation of the baseline model attains accuracy as if the scenario is known faster in Scenario (II) than in (III). Algorithm 1 and 4 have similar accuracy for direct effects in almost all settings. This is conform expectations given that both treat direct effects to be time-variant. There is indication in Table 9 and 10 that detecting the persistence of spillovers slightly improves estimation of direct effects, most notably in settings of high degree. Estimation of coefficients for common controls follows the tendencies described regarding the direct effects.

In summary, precision of estimating individual fixed effects and direct effects follow the statement that information on the scenario improves accuracy. In addition, the fourth algorithm benefits from the ability to detect time-invariant individual fixed effects. Finally, detection of a persistent network structure improves estimating direct effects for settings with high degree.

Table 3: Setting 3 Rural Village: Comparison and Network Recovery

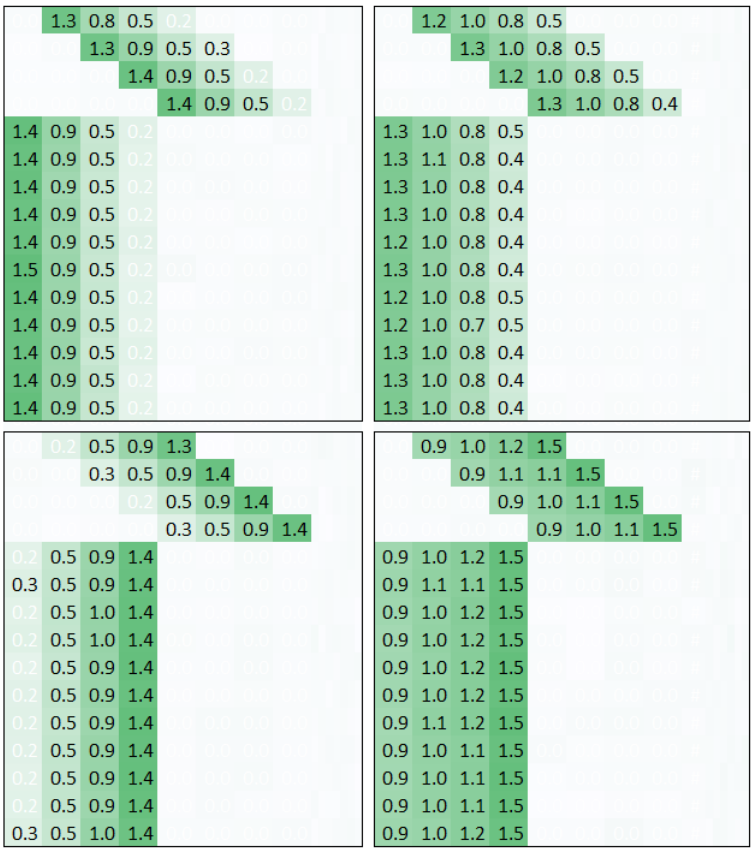
Panel (A)	Algorithm 1		Algorithm 4		True Network	
Scenario	Strong	Links	Strong	Links	Strong	Links
1	65	187	65	171	65	240
2	65	188	65	236	65	240
3	65	183	65	134	65	240
Panel (B)	Algorithm 1			Algorithm 4		
Scenario	Ratio	Strong	Weak	Ratio	Strong	Weak
1	0.23	96.2%	2.9%	0.19	85.0%	4.4%
2	0.23	96.1%	3.0%	0.29	99.5%	6.5%
3	0.23	95.7%	3.0%	0.20	84.2%	4.0%

300 Monte Carlo Simulations are performed for this setting with $N = 65$. Panel (A) contains measures "Strong" and "Links" defined as the number of (strong) links on the average of the estimated social interaction matrices. Links of absolute size bigger than 0.3 are regarded as strong and those smaller than 0.05 are regarded as non-links, similar to De Paula et al. (2018). Panel (B) demonstrates performance measures as denoted in section 4.1

5.4 Additional Settings

Table 3 and Table 15 illustrate results for the setting which mimics the simulation of De Paula et al. (2018) regarding the experiment in rural villages of Banerjee et al. (2013).

Figure 1: Network Recovery of Alg. 1 and Alg. 4 for Setting 4 Hierarchy



Note: 1000 Monte Carlo simulations are used to construct the averages of the social interaction matrices of Algorithm 1 (left) and Algorithm 4 (right). The spillovers before the break are presented above the spillovers after the break. Weak links defined as estimated spillovers with an absolute value smaller than 0.3 are presented as white.

The setting follows the insights discussed before and highlights the finding that weak links ($|\gamma_{ij}| < 0.3$) are not picked up by the methods in realistic settings. De Paula et al. (2018) scrutinize their method by observing whether the average estimated social interaction matrix contains the same amount of links and strong links as the true network contains. This is in contrast to the performance measure I employ that observes whether the method detects the strong spillovers for the correct pairs (i, j) for each simulation. Note that I average the performance measures over the simulations and not calculate the performance measures on an average of the estimated social interaction matrices. To compare, however, I have additionally extracted this information. The method of De Paula et al. (2018) retrieves 194 edges of which 68 strong, while the true network has 240 edges of which 65 strong. Panel A in Table 3 shows that Algorithm 1 retrieves similar numbers as De Paula et al. (2018) while Algorithm 4 detecting the time-invariant network in Scenario (II) provides measures close the numbers of the true network. Please recall the imperfect comparison given that Algorithm 1 and 4 consider a structural break, while De Paula et al. (2018) also incorporate the endogenous social effects. Besides that, the performance measures in panel B suggest that for larger T and more weak links, the benefit of Algorithm 4 is still observed for settings with persistent spillovers. For

Scenario (I) and (III), however, strong link detection of Algorithm 4 deteriorates, whereas Algorithm 1 has solid strong link detection in all scenarios.

Table 4: Comparison Scenario (II) with Heterogeneous and Homogeneous Direct Effects

N T Deg.	Alg.	*Heterogeneous*				*Homogeneous*			
		B	AD	Ratio	Strong	B	AD	Ratio	Strong
15 20 (1,1)	1	0.58	0.91	0.20	48%	0.12	0.52	0.27	62%
	2	0.44	0.80	0.39	74%	0.09	0.28	0.58	90%
	4	0.54	0.91	0.32	68%	0.11	0.45	0.42	83%
15 20 (6,6)	1	1.06	1.86	0.35	29%	0.26	1.64	0.42	34%
	2	0.72	1.28	0.90	58%	0.15	0.62	1.68	77%
	4	0.87	1.55	0.69	49%	0.21	1.08	1.00	63%
15 50 (1,1)	1	0.21	0.40	0.67	93%	0.05	0.14	0.74	97%
	2	0.18	0.14	0.99	99%	0.05	0.03	1.13	100%
	4	0.20	0.30	0.88	99%	0.05	0.15	0.90	99%
15 50 (6,6)	1	0.31	0.78	2.48	84%	0.06	0.33	3.80	93%
	2	0.20	0.25	7.41	99%	0.05	0.04	10.18	100%
	4	0.24	0.58	4.95	96%	0.05	0.21	6.69	99%
30 20 (1,1)	1	0.60	0.88	0.13	38%	0.10	0.42	0.18	54%
	2	0.50	0.71	0.21	70%	0.08	0.28	0.30	87%
	4	0.58	0.90	0.18	61%	0.09	0.39	0.25	79%
30 20 (3,3)	1	0.78	1.48	0.18	25%	0.15	1.17	0.22	32%
	2	0.66	1.03	0.38	58%	0.10	0.43	0.56	77%
	4	0.74	1.27	0.33	49%	0.12	0.75	0.46	65%
30 20 (6,6)	1	0.99	2.00	0.17	18%	0.21	1.97	0.20	21%
	2	0.84	1.33	0.42	45%	0.14	0.87	0.60	59%
	4	0.91	1.75	0.35	37%	0.17	1.41	0.47	48%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

In Setting 4 the algorithms are investigated in a context where spillovers pertain to the same pairs before and after a break, while the magnitudes alter such that the hierarchy of the sources change. The results in Table 16 suggest that Algorithm 4 is advantageous over Algorithm 1 and Algorithm 3 for $T = 20, 50$ in terms of network recovery. Moreover, for settings of $T = 20$ the information that only the social interaction matrix alters is valuable for Algorithm 3 in terms of break detection. Figure 1 demonstrates the estimated social interaction matrices of Algorithm 1 (left) and Algorithm 4 (right). The illustration suggests that the detection of time-invariant social ties has supported Algorithm 4 in detecting the strong links. In contrary, it also illustrates that Algorithm 4 has shrunk parameters $\breve{\gamma}_{ij}$ to zero too often in this setting of small T and a degree close to violation of the sparsity assumption. As a result, an upward bias is shown in the first column of $\boldsymbol{G}$. In general, a downward bias is observed for not all strong links are estimated in every simulation; only 74% are detected on average for Algorithm 4 and 50% for Algorithm 1.

In summary, Setting 3 has shown with an imperfect comparison that algorithms with a structural break at least have the potential to be considered next to methods such as

that of De Paula et al. (2018). Setting 4 shows that Algorithm 4 can use its detection technique for its advantage to identify key players even if their magnitude alter. This comes, however, with the risk of falsely identifying persistent parameters that result to less reliable magnitudes.

5.5 Extension: Heterogeneous Direct Effects

In this section, the extension of the algorithms to heterogeneous effects are portrayed. Given the similarity for execution of the algorithms, Appendix B contains only Algorithm 1 extended with heterogeneous direct effects. Table 4 shows that not only the estimation of direct effects worsens for making them individual-specific, also the network recovery and break detection is indicated to be less accurate. For $T = 50$, the difference for network retrieval becomes less evident, while the advantage for the homogeneous model stays in terms of break detection and estimation of direct effects. The advantage of Algorithm 4 in settings of (partially) persistent network structures described in sections 5.1, 5.2 and 5.3 are analogous for the model with heterogeneous direct effects.

Table 5: Setting 1 Degree Scenario 0: No break

Homogeneous direct effects							
N T Deg.	Alg	A	B	G	Overall	Ratio	Strong
15 20 (1,1)	1	0.41	0.12	0.13	84%	0.28	63%
	4	0.27	0.09	0.10	87%	0.49	88%
15 20 (6,6)	1	0.41	0.12	0.13	84%	0.28	63%
	4	0.27	0.09	0.10	87%	0.49	88%
15 50 (1,1)	1	0.19	0.05	0.06	90%	0.66	96%
	4	0.14	0.05	0.04	92%	0.92	100%
15 50 (6,6)	1	0.24	0.06	0.16	89%	3.58	92%
	4	0.16	0.05	0.10	93%	6.45	99%
30 20 (1,1)	1	0.41	0.09	0.07	90%	0.18	54%
	4	0.26	0.07	0.06	90%	0.30	86%
30 20 (6,6)	1	0.81	0.18	0.25	77%	0.22	23%
	4	0.46	0.14	0.20	80%	0.57	55%
Heterogeneous direct effects							
N T Deg.	Alg	A	B	G	Overall	Ratio	Strong
30 20 (1,1)	1	0.46	0.53	0.07	90%	0.15	43%
	4	0.29	0.46	0.06	90%	0.25	75%
30 20 (6,6)	1	1.05	1.96	0.35	77%	0.19	20%
	4	0.53	0.83	0.23	78%	0.43	44%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm.

5.6 Scenario: No Structural Break

Table 5 shows the results for Algorithm 1 and Algorithm 4 when the assumption of a structural break in the panel data is invalid. For Algorithm 4, the error in parameter estimation is substantially lower and network recovery is superior in terms of the measures Ratio and Strong. These results suggest that Algorithm 4 is preferred for applications where the assumption of a single break might be violated.

6 Empirical Application

6.1 R&D Spillovers

Research and Development (R&D) is a firm's intention to increase productivity via development of new goods or innovation. Given that most inventions that are productive are at least attempted to be imitated by competitors, firms tend to benefit from R&D investments from other firms. For this reason, the social return of R&D investments may be greater than the private return. Policy makers and researchers are interested in the quantification of R&D spillovers, given that it can suggest whether policies such as tax credits for R&D expenditure to lower the private cost of R&D are empirically founded. However, the receivers and senders of these spillover are generally not known as these interactions are not directly observable. To overcome this lack of information previous research has often chosen to proxy for these social ties, e.g. by use of historical patenting behavior of firms (Bloom et al., 2013). More recently, Manresa (2016) and Rose (2016) have applied their proposed methods for the quantification of spillovers and identification of key senders and receivers. In their applications, they assume that the social ties through which R&D spillovers are present are persistent. However, this assumption that the main spillovers originate from the same firms over decades is not infallible, especially in industries where innovation is of essence. Hence, I extend their analysis of R&D spillovers in absence of the assumption of time-invariant social interactions.

6.2 Model

The set-up of this application is deliberately similar to Rose (2016) and Manresa (2016), allowing comparison of the methods. For the model specification I build on the same baseline as Manresa (2016). Firms are assumed to produce according to a Cobb Douglas function extended with the stock of R&D or knowledge capital:

$$Y_{it} = A_i L_{it}^{\delta_l} C_{it}^{\delta_c} K_{it}^{\beta_i} \Pi_{j \neq i} K_{jt}^{\gamma_{ij}} e^{\epsilon_{it}}. \tag{34}$$

Here, L,C and K denote the stock of labor, capital and knowledge respectively, while $\delta_l, \delta_c, \beta_i, \gamma_{ij}$ are their corresponding coefficients. Note here, that I have chosen to use the extension of the algorithms using, β_i, accommodating to the heterogeneous direct effects found by previous research. A_i represents a firm-specific technology shock and ϵ_{it} is assumed to be symmetrically distributed under the conditions in MA2. Then I adjust this setting to allow for time-variance in the parameters in terms of structural breaks.

$$Y_{it} = \begin{cases} A_i L_{it}^{\delta_l} C_{it}^{\delta_c} K_{it}^{\beta_i} \Pi_{j \neq i} K_{jt}^{\gamma_{ij}} e^{\epsilon_{it}} & t = 1, \ldots, \tau^* \\ \\ \mathbb{A}_i L_{it}^{d_l} C_{it}^{d_c} K_{it}^{b_i} \Pi_{j \neq i} K_{jt}^{g_{ij}} e^{\epsilon_{it}} & t = \tau^* + 1, \ldots, T \end{cases} \tag{35}$$

Please note the the deviating notation here of A_i and $\mathbb{A}_i$ which here denote the firm-specific effects capturing the firm-specific persistent differences within their respective sub-periods. Taking logs of (35) the following regression model follows:

$$y_{it} = \begin{cases} \alpha_i + \beta_i k_{it} + \sum_{j \neq i} \gamma_{ij} k_{jt} + \delta_c c_{it} + \delta_l l_{it} + \epsilon_{it} & t = 1, \ldots, \tau^* \\ a_i + b_i k_{it} + \sum_{j \neq i} g_{ij} k_{jt} + d_c c_{it} + d_l l_{it} + \epsilon_{it} & t = \tau^* + 1, \ldots, T \end{cases} \tag{36}$$

Here, the lowercase letters of the previous described variables in capital letters denote the variables in natural logs. This application showcases Algorithm 1 and Algorithm 4 to inspect their differences. The fourth algorithm is able to determine whether the key players concerning R&D spillovers and firm-specific effects are time-invariant while assuming the firm-specific direct effects and controls to exhibit change at the break date. Furthermore, Bloom et al. (2013) has stated that the externalities of R&D investments are not only positive for other firms. R&D investments can also limit the productivity of those linked, given that innovation combined with patents can lead to stealing business of competitors. Hence, depending on the sign, identified spillovers may signify beneficial copying behavior or a product market rivalry effect.

6.3 Endogeneity

Current choices of the inputs labor, capital and knowledge are likely to cause endogeneity being present in the model (36). E.g. by simultaneity firms can specify their amount of R&D investments based on their sales projections or sales made in the year. Thence, to limit endogeneity I apply both the approach of Manresa (2016) and Rose (2016). The former chooses to preserve the regression framework and for that reason chooses to include the inputs of the productive function as lags, see (37). This still provides consistent OLS estimates for sufficient T under the conditions that lagged R&D investments are not correlated with current shocks in output and that there is no persistence in these shocks. As T with respect the number of regressors is not very large, the results should be interpreted with this bias in mind.

$$y_{it} = \begin{cases} \alpha_i + \beta_i k_{it-1} + \sum_{j \neq i} \gamma_{ij} k_{jt-1} + \delta_c c_{it-1} + \delta_l l_{it-1} + \epsilon_{it} & t = 1, \ldots, \tau^* \\ a_i + b_i k_{it-1} + \sum_{j \neq i} g_{ij} k_{jt-1} + d_c c_{it-1} + d_l l_{it-1} + \epsilon_{it} & t = \tau^* + 1, \ldots, T \end{cases} \tag{37}$$

In case these conditions do not hold instrumental variables are needed. This is the approach Rose (2016) applied, which he had based on the work of Bloom et al. (2013). He uses the United States Research and Experimentation Tax Credit policy to capture exogenous variation of the R&D stock. For this paper, I implement three of the five suggested instrumental variables (IV) for reasons of availability and attainability [4]. The IV included concern the federal component of the tax credit for Research and Experimentation, initiated in 1981. The tax credit is firm-specific and is claimed on the difference between actual R&D expenditure and a firm-specific "base". From 1981 to 1989 the base was the maximum of a rolling average of the previous three years' R&D and 50% of current R&D

[4] Complete replication would require additional investigation of all patents and their inventors of the firms in the sample such that state-specific components of the tax credit may be used.

expenditure. From 1990 until 2001 (excl.1995-1996) the base was fixed to be the average of the firm's R&D to sales ratio between 1984 and 1988, multiplied by current sales, up to a maximum of 16%. This calculation follows that of Bloom et al. (2013), however, a clearer overview of the development of the tax credit is presented by Rao (2016, p. 3). The first stage of the IV-regression is executed per individual i noted in (38),

$$k_{it} = \pi_i^0 + \pi_i^1 k_{it-1} + \pi_i^2 tax_{it} + \pi_i^3 tax_{it-1} + \pi_i^4 c_{it-1} + \pi_i^5 l_{it-1} + u_{it}, \tag{38}$$

where $\boldsymbol{\pi}_i$ denote the first stage coefficients, tax_{it} denotes the federal component of the R&D tax credit and u_{it} is the idiosyncratic error. The second stage regression is identical to that of equation (37) with the exception of substituting the fitted values of the first stage for the lag of the R&D stock for the estimation of direct and indirect effects, i.e. $\hat{k}_{it}$, $\hat{k}_{jt}$ instead of k_{it-1}, k_{jt-1}.

The benefit of including the approach of Rose (2016) is the addition of instrumental variables aiming to capture the exogenous variation related to policy intervention and which are widely used by previous research. The drawbacks are that Manresa (2016) is still developing the methodology to cover instrumental variables and the first stage equation requires the additional assumption that the relationship between the instruments and the R&D stock is persistent over the whole sample period. Therefore, the results of the approach of Rose (2016) need be interpreted with caution.

6.4 Data, Measurement and Sample Selection

For data, measurement and sample selection the set-up of Rose (2016) was most reproducible. The sample set by Rose (2016) consists that of 29 firms which are defined by the two-digit SIC code for Electronics. The firms chosen are those in the industry that have a positive R&D stock over the sample period, from 1980 to 2001, and have issued for patents based on data of the United States Patent and Trademark Office. This paper utilizes 26 of the 29 firms over the same period, with the exclusion of three firms due to their unattainability from the database with the license I have available. The annual firm-level data is obtained from the Compustat Accounting database[5]. Output is measure in real sales, where net sales levels are deflated using the CPI[6] with 1996 as the base year. Following Manresa (2016) capital stock is measured in book values and deflated by the CPI and labor stock is approximated by the number of employees. R&D stock is measured using the perpetual inventory method described in Bloom et al. (2013):

$$K_{it} = (1 - \mu)K_{it-1} + I_{it}, \tag{39}$$

where K_{it} and I_{it} stand for the R&D-stock and R&D investments at time t, respecitively. Depreciation is here denoted by μ and set to 15%, following Rose (2016) and Manresa (2016). The initial stock of R&D-stock is calculated based on a steady state level based on a 5% growth level. The descriptive statistics of the data is displayed in Table 6.

[5]Retrieved at 23/12/2018 from https://wrds-web.wharton.upenn.edu/wrds/
[6]Obtained at 23/12/2018 from https://www.bls.gov/cpi/research-series/

Table 6: Descriptive statistics for the electronic industry from 1980 to 2001

Variable	Symbol	Units	Mean	S.D.	Min	Max
Real Sales	Y	Million 1996$	2712.1	5003.8	32.2	34370.3
Real R&D Stock	K	Million 1996$	773.4	1881.5	2.8	16242.2
State tax credit	tax	Million 1996$	5.4	20.9	0.00	319.9
Real Capital Stock	C	Million 1996$	118.8	433.8	0.1	6015.0
Labour Stock	L	Thousands	18.5	28.7	0.3	150.0

To find out whether the instruments are strong predictors of the log R&D stock, I follow Rose (2016) and compute the F statistic on the individual regressions in equation 38 for the restriction: $\pi_i^1 = \pi_i^2 = \pi_i^3 = 0$. Similarly to Rose, the F statistics estimated are large, with the median and minimum F-statistic being 112.56 and 9.04, respectively. The maximum related p-value is 0.001, from which is concluded that the tax credits and the lag are not too weak predictors.

6.5 Results

The results are estimated with Algorithm 1 and Algorithm 4. Similar to Rose (2016), the data is initially transformed with forward orthogonal deviations. The trimmed subset of potential break-dates ranges from 1987-1993. For the data-driven approach of the tuning parameter, cross-validation over a logarithmic grid of 300 values is applied ranging from $[10^{-3}, 10^{6}]$. This application focuses on illustrating the usage of the algorithms with the aim to demonstrate the opportunity of detecting persistent spillovers while assuming time-variance. This paper does not state that its proposed methods are proven sufficiently for this application to consider the significance of the inference. Therefore, point estimates should be interpreted with caution. Moreover, the simulation study exposed that the algorithms are only applicable for settings where spillovers are large enough.

Table 7 summarizes the results for both the approaches of (A) Manresa (2016) and (B) Rose (2016). For setting (A) the results suggest that the change occurs at 1991. Private effects of R&D increased after the break according to Algorithm 1, while Algorithm 4 suggest that the average intensity of spillovers became twice more positive. In setting B, Algorithm 1 deviates most from the other estimations with a break at 1987 and suggesting that from 1981 to 1987 externalities of R&D suggest a highly competitive nature. Algorithm 4 in setting B demonstrates that the break at 1991 follows the tendencies of setting A that the private effects of R&D became more intense as well as social effects less positive and/or more negative.

Network recovery is visually portrayed in Appendix C Figures 2 to 5. Overall, the figures suggest that Siliconix Inc. (SILI), Thomas Betts Corp. (TNB), Harris Corp. (HRS) and Cooper Industries Plc. (CBE) are key influencers in the sample of innovating firms in the electronics industry. Comparing Algorithm 1 and 4 in setting A, the intensive and extensive margin of the sub-periods are illustrated as in Table 7. However, by three observations the period after the break appears similarly estimated by the algorithms: (i) SILI and HRS are the main sources of positive spillovers; (ii) both algorithms identify the two key receivers Standard Microsystems Corp. (SMSC) and INTC (Intel Corp.); (iii) almost identical estimation of competitive spillovers sent by Woodhead Industries Inc. (WDHD). The algorithms show more diverging social interaction matrices in setting B relative to setting A. This is likely to originate from Algorithm 1 estimating the break at

Table 7: R&D Spillovers in the Electronic Industry

Description	Definition	(A) Lags		(B) IV	
		Alg. 1	Alg. 4	Alg. 1	Alg. 4
Direct effect of R&D	$\frac{1}{N}\sum_{i=1}^{N}\beta_i$	0.40	0.80	0.99	0.85
	$\frac{1}{N}\sum_{i=1}^{N}b_i$	1.32	0.85	0.85	1.24
Intensive margin of spillovers	$\frac{\sum_{i=1}^{N}\sum_{j\neq i}\gamma_{ij}}{\sum_{i=1}^{N}\sum_{j\neq i}\mathbb{I}\{\gamma_{ij}\neq 0\}}$	0.22	0.11	-0.91	0.09
	$\frac{\sum_{i=1}^{N}\sum_{j\neq i}g_{ij}}{\sum_{i=1}^{N}\sum_{j\neq i}\mathbb{I}\{g_{ij}\neq 0\}}$	0.15	0.20	0.08	-0.10
Extensive margin of spillovers	$\frac{\sum_{i=1}^{N}\sum_{j\neq i}\mathbb{I}\{\gamma_{ij}\neq 0\}}{N(N-1)}$	0.11	0.08	0.07	0.11
	$\frac{\sum_{i=1}^{N}\sum_{j\neq i}\mathbb{I}\{g_{ij}\neq 0\}}{N(N-1)}$	0.08	0.12	0.14	0.16
Break date	τ	1991	1991	1987	1991

1987. The most robust finding is that SILI sends mainly large positive R&D spillovers between 1991 and 2000, except for its suggested rival SMSC.

The figures in Appendix C 2 to 5 are constructed such that the firms are sorted in increasing size along the axis. Here, size is measured in total sales over the full sample period. In contrast to Rose (2016), there is no indication that larger firms tend to send more spillovers. This deviation may have origin in our assumptions regarding persistence. Another disparity is that Rose forced the spillovers from other firm's R&D stock to be positive and spillovers captured as endogenous social effects to be negative. In case negative and positive externalities are both present between pairs of firms, Algorithm 1 and Algorithm 4 cannot distinguish between them. For this reason, the algorithms can only identify social ties where R&D either dominantly related to additional competition between firms or to spreading productivity by reproducible innovation.

7 Conclusion

Recently proposed panel models that quantify spillovers and recover sparse network structures rely heavily on the assumption of persistent social interactions. This paper relaxes the assumption of persistence by means of structural breaks, using Manresa (2016) as a starting point for this exploration. Different scenarios of breaks are considered with focus on whether the social interaction matrix structurally changes. Information on the scenarios is evaluated by comparison of algorithms that estimate as if the scenario is known relative to the baseline model that assumes all parameters alter at the break date. A simulation study has indicated that information of time-invariant parameters improves parameter estimation and enhances the retrieval of information relative to the misinformation from the estimated network. Identification of key players and break detection tend to improve mainly from information that the spillovers are persistent.

Furthermore, this paper has proposed an algorithm that detects persistent social interactions. This flexible algorithm uses numerical proximity and penalization to identify whether coefficients in successive periods need be estimated as time-invariant parameters. The value of this algorithm is measured relative to the baseline model and to what extent performance attains levels as if the scenario is known. Results suggest that performance

of the flexible algorithm is dependent on whether the additional effort to detect persistent spillovers is worthwhile. Worthwhile settings comprise those with at least a subset of the social interactions persistent. On average, a short time-span enhances the relative advantage of detecting persistent spillovers. This suggests that when the time-variation available is small, doubling the number of observations is more valuable for extracting key information. Short panel data are common in econometric research on social network analysis and is more suitable for this single-break investigation. Next to that, requirement of partial persistence is already a relaxation of persistence. Hence, one can conclude that the flexible algorithm has shown potential to relax the assumption of persistence in relevant settings. The potential of the algorithm is demonstrated in terms of break detection, accuracy of parameter estimation and network recovery and predominantly in identification of key players.

Additional inquiries have shown that identification of persistent social ties with varying magnitudes requires the difference in magnitudes across sub-periods to be of significant size. Otherwise, the algorithm deems these spillovers time-invariant and merges their magnitude. The advantage of the flexible algorithm in its worthwhile settings also pertain to the extended algorithm considering heterogeneous direct effects. Moreover, in case the setting is persistent and the algorithms erroneously assume a structural break, the flexible algorithm is less affected by this mistake. Furthermore, an empirical application has shown that the flexible algorithm and the baseline model recover both different and similar patterns of R&D spillovers in the Electronics industry of the United States. However, more research is necessary to conclude more regarding their relative performance in practice.

All in all, this paper concludes that the notions of different scenarios of structural breaks and detection of time-invariant spillovers by means of penalization are valuable to consider in the relaxation of the persistent network structure.

7.1 Discussion

This paper has functioned as an exploration of relaxing the assumption of a persistent social network structure by means of scenarios of structural breaks. I do not claim that the algorithms themselves pertain to optimal methods in this field of research nor that their execution is optimal. With the conclusion of this paper, I suggest future research to review this exploration for multiple methods such as from Lam and Souza (2014), De Paula et al. (2018) and Rose (2016).

Next to that, I recommend an extension to multiple breaks as well as addition of an information criterion to establish the number of unknown breaks. Hui et al. (2015) may provide direction with their extended regularized information criterion that is specific for tuning parameters of the adaptive lasso. Furthermore, the methods in this paper are unable to detect changes at the boundaries of the time-span. Bardwell et al. (2018) shows promising results for change point detection in a multivariate panel model close to the end of the time-period, which could contribute to the field of social network econometrics.

Moreover, in the aspect of execution, Manresa (2016) suggests in her supplementary appendix that a choice of a higher penalty tuning parameter λ can increase the performance of her method. I expect this result to originate from the noise of over-identified links of small magnitude, which a higher penalty may shrink to zero. In turn, Post-Lasso can gain efficiency for having less parameters to estimate. In opposition, this choice is difficult to argue for in combination with using a data-driven approach for the tuning-parameter. Besides that, for the proposed flexible algorithm a higher λ choice may also entail that more parameters are mistakenly estimated to be time-invariant. Another re-

mark related to the execution is that the trimmed subsets of potential break dates are chosen to be small, (9 or 10), for computational efficiency. This has the drawback of suggesting the break date precision too favorable, however, the relative precision across algorithms remained valid and may have supported the comparability across time-spans. The number of Monte Carlo simulations applied in the simulation study are sub-optimal, which is due to the break detection and cross-validation requiring two days for a 1000 simulations for a setting with $N = 15$ and three algorithms.

Furthermore, future research is recommended to investigate relaxation of the assumption of a persistent network structure in different directions. Machine learning techniques may provide the additional tools for estimating time-varying networks from characteristics. This direction is for example taken by Kolar et al. (2010), who propose their temporally smoothed l_1-regularized logistic regression formalism. They also argue that including as much as prior information as possible is pivotal, which implies Bayesian methods to be considered. Future research is encouraged to aim for time-varying networks allowing both smooth as well as structural changes.

On another angle, future research is urged to the opportunity of limiting the heterogeneity of social effects to enable time-variance in the structure. Manresa (2016) implores this suggestion in her supplementary appendix by relating social network analysis to her previous research in Bonhomme and Manresa (2015). She adapts her model to let social effects depend on the group-membership of individuals. Inspiration for limiting heterogeneity is also attainable in the literature of stochastic block models with community detection in sparse networks (e.g. Guédon & Vershynin, 2016).

References

Antoch, J., Hanousek, J., Horváth, L., Hušková, M., & Wang, S. (2018). Structural breaks in panel data: Large number of panels and short length time series. *Econometric Reviews*, 1–24.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, *79*(4), 551–563.

Baltagi, B. H., Feng, Q., & Kao, C. (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, *191*(1), 176–195.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Bardwell, L., Fearnhead, P., Eckley, I. A., Smith, S., & Spott, M. (2018, 2). Most recent changepoint detection in panel data. *Technometrics*. doi: 10.1080/00401706.2018.1438926

Bickel, P. J., Ritov, Y., Tsybakov, A. B., et al. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, *37*(4), 1705–1732.

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347–1393.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*(3), 1147–1184.

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

Caner, M., & Zhang, H. H. (2014). Adaptive elastic net for generalized methods of moments. *Journal of Business & Economic Statistics*, *32*(1), 30–47.

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, 303–357.

Chandrasekhar, A., & Lewis, R. (2016, November). Econometrics of sampled networks. *Unpublished manuscript, Stanford*. doi: Availableathttps://stanford.edu/˜arungc/CL.pdf

Chandrasekhar, A. G., & Jackson, M. O. (2014). Tractable and consistent random graph models. *NBER Working Paper No. w20276. Available at SSRN: https://ssrn.com/abstract=2463391*.

De Paula, A. (2017). Econometrics of network models. In *Advances in economics and econometrics: Theory and applications, eleventh world congress* (pp. 268–323).

De Paula, A., Rasul, I., & Souza, P. (2018). Recovering social networks from panel data: Identification, simulations and an application. *CeMMAP Working Paper version April 2018*.

Elhorst, J. P. (2014). Spatial panel data models. In *Spatial econometrics* (pp. 37–93). Springer, Berlin, Heidelberg.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from http://www.jstatsoft.org/v33/i01/

Gautier, E., & Tsybakov, A. (2014). High-dimensional instrumental variables regression and confidence sets. *arXiv preprint arXiv:1105.2454 (under second revision for Econometrica)*.

Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation* (Tech. Rep.). National Bureau of Economic Research.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, *85*(4), 1033–1063.

Guédon, O., & Vershynin, R. (2016). Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, *165*(3-4), 1025–1049.

Hui, F. K., Warton, D. I., & Foster, S. D. (2015). Tuning parameter selection for the adaptive lasso using eric. *Journal of the American Statistical Association*, *110*(509), 262–269.

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

Kendrick, L., Musial, K., & Gabrys, B. (2018). Change point detection in social networks—critical review with experiments. *Computer Science Review*, *29*, 1–13.

Kolar, M., Song, L., Ahmed, A., Xing, E. P., et al. (2010). Estimating time-varying networks. *The Annals of Applied Statistics*, *4*(1), 94–123.

Lam, C., & Souza, P. C. (2014). Regularization for spatial panel time series using the adaptive lasso. *Unpublished. London School of Economics, Department of Statistics*.

Lam, C., & Souza, P. C. (2015). *One-step regularized spatial weight matrix and fixed effects estimation with instrumental variables* (Tech. Rep.). Manuscript.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, *188*(1), 182–195.

Liu, X., Patacchini, E., Zenou, Y., & Lee, L.-F. (2012). Criminal networks: Who is the key player? *FEEM Working Paper No. 39.2012. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2089267*.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.

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

Qian, J., & Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, *191*(1), 86–109.

Rao, N. (2016). Do tax credits stimulate r&d spending? the effect of the r&d tax credit in its first decade. *Journal of Public Economics*, *140*, 1–12.

Rose, C. (2016). Identification of spillover effects using panel data. *Unpublished Manuscript. Toulouse School of Economics, Toulouse*.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(1), 91–108.

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

Table 8: Setting 1 Degree: Scenario (I)

N T Deg.	Alg.	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong
15 20 (1,1)	1	0.396	0.114	0.129	0.115	0.292	85%	64%	84%	0.27	62%
	4	0.497	0.127	0.149	0.126	0.452	84%	54%	82%	0.20	52%
15 20 (1,6)	1	0.560	0.181	0.301	0.173	0.677	84%	41%	74%	0.37	38%
	4	0.642	0.179	0.307	0.175	0.753	83%	44%	74%	0.40	41%
15 20 (6,6)	1	0.748	0.243	0.474	0.222	0.782	83%	36%	64%	0.41	33%
	4	0.757	0.237	0.469	0.218	0.978	78%	44%	64%	0.50	41%
15 50 (1,1)	1	0.187	0.049	0.058	0.047	0.092	90%	97%	91%	0.75	97%
	4	0.195	0.049	0.069	0.049	0.124	87%	93%	87%	0.51	93%
15 50 (1,6)	1	0.207	0.054	0.106	0.055	0.075	89%	94%	90%	2.41	94%
	4	0.216	0.055	0.113	0.055	0.127	87%	93%	89%	2.00	93%
15 50 (6,6)	1	0.231	0.060	0.157	0.059	0.067	88%	93%	90%	3.81	93%
	4	0.252	0.063	0.187	0.064	0.110	73%	93%	81%	1.96	93%
15 100 (1,1)	1	0.120	0.030	0.031	0.032	0.034	93%	100%	93%	1.05	100%
	4	0.120	0.031	0.035	0.032	0.041	89%	100%	89%	0.66	100%
15 100 (1,6)	1	0.122	0.032	0.051	0.032	0.028	92%	100%	94%	4.29	100%
	4	0.122	0.033	0.051	0.032	0.031	90%	100%	92%	3.30	100%
15 100 (6,6)	1	0.126	0.033	0.071	0.033	0.004	92%	100%	95%	8.86	100%
	4	0.126	0.033	0.080	0.033	0.015	74%	100%	85%	2.61	100%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 9: Setting 1 Degree: Scenario (II) for N = 15

N T Deg.	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong
15 20 (1,1)	1	0.406	0.123	0.130	0.119	0.515	85%	63%	84%	0.27	62%
	2	0.330	0.091	0.084	0.089	0.278	88%	90%	88%	0.58	90%
	4	0.429	0.111	0.110	0.107	0.450	86%	84%	86%	0.42	83%
15 20 (3,3)	1	0.557	0.181	0.263	0.176	0.949	84%	48%	77%	0.42	47%
	2	0.392	0.110	0.148	0.107	0.403	86%	85%	86%	1.34	85%
	4	0.522	0.148	0.198	0.140	0.648	83%	76%	81%	0.85	75%
15 20 (6,6)	1	0.767	0.264	0.476	0.243	1.637	83%	37%	65%	0.42	34%
	2	0.508	0.151	0.277	0.143	0.620	83%	77%	80%	1.68	77%
	4	0.651	0.214	0.360	0.196	1.076	79%	65%	73%	1.00	63%
15 50 (1,1)	1	0.186	0.050	0.059	0.050	0.142	90%	97%	91%	0.74	97%
	2	0.169	0.046	0.032	0.046	0.034	93%	100%	93%	1.13	100%
	4	0.180	0.049	0.044	0.048	0.146	91%	99%	92%	0.90	99%
15 50 (3,3)	1	0.203	0.052	0.093	0.054	0.256	89%	96%	91%	2.14	96%
	2	0.174	0.045	0.048	0.046	0.050	92%	100%	94%	3.69	100%
	4	0.187	0.048	0.066	0.050	0.178	90%	99%	92%	2.77	99%
15 50 (6,6)	1	0.233	0.063	0.158	0.063	0.327	88%	93%	90%	3.80	93%
	2	0.179	0.047	0.072	0.048	0.042	92%	100%	95%	10.18	100%
	4	0.198	0.053	0.098	0.052	0.206	89%	99%	93%	6.69	99%
15 100 (1,1)	1	0.122	0.031	0.032	0.031	0.069	93%	100%	93%	1.04	100%
	2	0.117	0.030	0.020	0.030	0.014	94%	100%	94%	1.36	100%
	4	0.120	0.030	0.026	0.031	0.055	93%	100%	93%	1.18	100%
15 100 (3,3)	1	0.122	0.033	0.047	0.031	0.081	92%	100%	94%	3.44	100%
	2	0.117	0.031	0.031	0.031	0.015	93%	100%	95%	4.51	100%
	4	0.119	0.032	0.038	0.031	0.073	92%	100%	94%	3.77	100%
15 100 (6,6)	1	0.125	0.033	0.071	0.033	0.084	92%	100%	95%	8.72	100%
	2	0.118	0.032	0.046	0.032	0.012	93%	100%	96%	11.58	100%
	4	0.121	0.032	0.055	0.032	0.080	92%	100%	95%	10.02	100%

1000 Monte Carlo Simulations are performed for these settings with $N = 15$. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 10: Setting 1 Scenario (II) for N = 30

N T Deg.	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong
30 20 (1,1)	1	0.404	0.096	0.068	0.091	0.422	91%	56%	90%	0.18	54%
	2	0.350	0.075	0.057	0.074	0.280	90%	88%	90%	0.30	87%
	4	0.475	0.090	0.065	0.087	0.392	89%	81%	89%	0.25	79%
30 20 (3,3)	1	0.588	0.152	0.144	0.146	1.170	90%	34%	85%	0.22	32%
	2	0.436	0.095	0.106	0.102	0.434	87%	78%	86%	0.56	77%
	4	0.590	0.120	0.120	0.122	0.752	87%	67%	85%	0.46	65%
30 20 (6,6)	1	0.816	0.215	0.261	0.209	1.968	90%	23%	77%	0.20	21%
	2	0.619	0.145	0.203	0.148	0.866	84%	60%	80%	0.60	59%
	4	0.736	0.172	0.221	0.170	1.406	86%	50%	79%	0.47	48%
30 50 (1,1)	1	0.205	0.040	0.047	0.041	0.294	91%	96%	91%	0.37	96%
	2	0.178	0.034	0.025	0.032	0.048	94%	100%	94%	0.60	100%
	4	0.196	0.038	0.034	0.038	0.180	92%	100%	93%	0.47	100%
30 50 (3,3)	1	0.232	0.045	0.079	0.047	0.354	88%	93%	88%	0.82	92%
	2	0.183	0.034	0.037	0.034	0.058	93%	100%	93%	1.57	100%
	4	0.207	0.040	0.051	0.042	0.316	91%	99%	91%	1.21	99%
30 50 (6,6)	1	0.300	0.063	0.145	0.063	0.684	84%	84%	84%	1.07	83%
	2	0.191	0.036	0.052	0.035	0.048	92%	100%	93%	3.19	100%
	4	0.229	0.046	0.078	0.046	0.468	88%	99%	90%	2.07	99%

500 Monte Carlo Simulations are performed for these settings with $N = 30$. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 11: Setting 1 Degree Scenario (III)

N T Deg.	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong
15 20 (1,1)	1	0.401	0.113	0.131	0.113	0.934	85%	62%	84%	0.26	61%
	3	0.269	0.072	0.114	0.073	1.082	89%	63%	87%	0.36	62%
	4	0.312	0.108	0.139	0.111	1.429	84%	56%	83%	0.22	54%
15 20 (1,6)	1	0.558	0.173	0.298	0.158	1.093	85%	40%	74%	0.37	38%
	3	0.370	0.106	0.283	0.104	1.049	87%	41%	77%	0.42	40%
	4	0.405	0.158	0.295	0.153	1.196	83%	47%	74%	0.43	44%
15 20 (6,6)	1	0.744	0.234	0.476	0.214	1.328	83%	36%	64%	0.40	33%
	3	0.507	0.154	0.462	0.141	1.320	85%	36%	65%	0.42	34%
	4	0.546	0.203	0.454	0.190	1.438	78%	46%	65%	0.53	44%
15 50 (1,1)	1	0.190	0.051	0.061	0.049	0.768	90%	97%	90%	0.71	97%
	3	0.129	0.034	0.049	0.032	0.758	93%	98%	93%	1.00	98%
	4	0.158	0.052	0.068	0.050	0.756	87%	93%	87%	0.51	93%
15 50 (1,6)	1	0.210	0.055	0.107	0.055	0.211	89%	93%	90%	2.40	93%
	3	0.142	0.037	0.100	0.037	0.221	91%	94%	92%	2.78	94%
	4	0.174	0.056	0.111	0.055	0.250	87%	94%	89%	2.03	94%
15 50 (6,6)	1	0.230	0.060	0.156	0.060	0.185	88%	93%	90%	3.86	93%
	3	0.155	0.041	0.141	0.041	0.163	90%	95%	92%	4.96	95%
	4	0.206	0.065	0.184	0.062	0.279	73%	93%	81%	1.98	93%
15 100 (1,1)	1	0.121	0.032	0.032	0.031	0.454	93%	100%	93%	1.03	100%
	3	0.085	0.021	0.028	0.021	0.504	94%	100%	95%	1.36	100%
	4	0.101	0.032	0.035	0.031	0.569	89%	100%	89%	0.65	100%
15 100 (1,6)	1	0.123	0.033	0.052	0.032	0.067	92%	100%	94%	4.17	100%
	3	0.086	0.022	0.050	0.022	0.072	93%	100%	95%	4.70	100%
	4	0.100	0.032	0.051	0.032	0.095	90%	100%	92%	3.33	100%
15 100 (6,6)	1	0.125	0.033	0.072	0.032	0.045	92%	100%	95%	8.68	100%
	3	0.088	0.023	0.068	0.022	0.041	93%	100%	96%	11.51	100%
	4	0.105	0.033	0.079	0.032	0.064	74%	100%	85%	2.63	100%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 12: Setting 2 Political Parties Scenario (I)

N	T	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong	Weak
15	20	1	0.385	0.115	0.131	0.114	0.432	86%	30%	79%	0.18	41%	2%
		4	0.457	0.116	0.135	0.113	0.478	85%	30%	78%	0.18	42%	2%
15	50	1	0.193	0.051	0.075	0.051	0.199	90%	49%	85%	0.41	82%	2%
		4	0.197	0.052	0.076	0.051	0.215	88%	50%	83%	0.38	82%	4%
15	100	1	0.122	0.032	0.045	0.031	0.116	92%	61%	88%	0.66	97%	9%
		4	0.122	0.032	0.047	0.032	0.150	89%	62%	86%	0.54	97%	15%
15	200	1	0.083	0.022	0.029	0.021	0.064	93%	73%	91%	1.01	100%	34%
		4	0.083	0.022	0.030	0.021	0.078	90%	71%	88%	0.74	100%	34%
15	400	1	0.057	0.015	0.018	0.015	0.050	94%	87%	93%	1.76	100%	68%
		4	0.057	0.015	0.020	0.015	0.051	91%	83%	90%	1.05	100%	60%
30	20	1	0.385	0.092	0.067	0.090	0.386	91%	22%	87%	0.11	31%	2%
		4	0.511	0.094	0.077	0.090	0.466	89%	23%	85%	0.10	31%	2%
30	50	1	0.208	0.041	0.055	0.041	0.206	91%	47%	88%	0.25	76%	3%
		4	0.216	0.041	0.053	0.042	0.244	91%	46%	88%	0.24	75%	3%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 13: Setting 2 Political Parties Scenario (II)

N	T	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong	Weak
15	20	1	0.389	0.116	0.133	0.115	0.532	86%	31%	79%	0.19	44%	2%
		2	0.334	0.088	0.099	0.090	0.248	88%	45%	83%	0.33	73%	1%
		4	0.434	0.108	0.120	0.110	0.479	86%	41%	80%	0.27	64%	2%
15	50	1	0.192	0.052	0.074	0.051	0.192	90%	51%	85%	0.43	85%	2%
		2	0.174	0.046	0.045	0.045	0.050	92%	61%	88%	0.68	98%	8%
		4	0.184	0.050	0.059	0.048	0.165	91%	59%	87%	0.58	95%	6%
15	100	1	0.122	0.032	0.044	0.031	0.083	92%	62%	88%	0.68	98%	9%
		2	0.117	0.031	0.029	0.030	0.024	93%	73%	91%	1.02	100%	33%
		4	0.120	0.032	0.037	0.031	0.093	92%	70%	89%	0.86	99%	25%
15	200	1	0.084	0.021	0.029	0.022	0.061	93%	73%	91%	1.00	100%	34%
		2	0.082	0.020	0.018	0.021	0.019	94%	87%	93%	1.83	100%	67%
		4	0.082	0.021	0.022	0.021	0.037	94%	83%	93%	1.50	100%	57%
15	400	1	0.058	0.014	0.018	0.015	0.028	94%	87%	93%	1.73	100%	68%
		2	0.057	0.014	0.011	0.015	0.022	94%	98%	95%	3.16	100%	93%
		4	0.057	0.014	0.012	0.015	0.036	95%	95%	95%	3.15	100%	86%
30	20	1	0.389	0.093	0.068	0.090	0.438	91%	23%	87%	0.12	34%	2%
		2	0.351	0.077	0.063	0.075	0.274	90%	41%	87%	0.20	65%	2%
		4	0.481	0.090	0.070	0.087	0.376	89%	36%	86%	0.17	56%	3%
30	50	1	0.209	0.043	0.055	0.041	0.245	91%	48%	88%	0.26	80%	3%
		2	0.180	0.035	0.032	0.034	0.048	93%	60%	91%	0.45	97%	6%
		4	0.201	0.040	0.041	0.038	0.208	92%	58%	90%	0.38	94%	5%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 14: Setting 2 Political Parties Scenario (III)

N	T	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong	Weak
15	20	1	0.388	0.112	0.130	0.105	1.526	86%	29%	79%	0.17	40%	2%
		3	0.255	0.069	0.115	0.068	1.695	90%	26%	82%	0.18	39%	1%
		4	0.287	0.098	0.124	0.098	1.826	86%	31%	79%	0.18	43%	2%
15	50	1	0.198	0.052	0.080	0.051	1.917	89%	49%	84%	0.39	80%	2%
		3	0.133	0.035	0.066	0.034	1.790	92%	48%	87%	0.46	82%	1%
		4	0.157	0.050	0.076	0.050	1.794	88%	50%	83%	0.38	81%	4%
15	100	1	0.124	0.033	0.046	0.033	1.551	92%	62%	88%	0.65	97%	9%
		3	0.086	0.023	0.042	0.023	1.513	93%	61%	89%	0.72	97%	7%
		4	0.102	0.032	0.047	0.032	1.532	89%	62%	86%	0.54	97%	14%
15	200	1	0.083	0.022	0.030	0.022	1.225	93%	73%	90%	0.99	100%	33%
		3	0.059	0.015	0.028	0.016	1.280	94%	72%	92%	1.10	100%	31%
		4	0.068	0.022	0.030	0.022	1.433	90%	71%	88%	0.74	100%	33%
15	400	1	0.058	0.015	0.018	0.015	1.151	94%	87%	93%	1.78	100%	68%
		3	0.041	0.011	0.016	0.011	1.031	96%	86%	95%	2.26	100%	65%
		4	0.047	0.015	0.020	0.015	1.163	91%	83%	90%	1.06	100%	61%
30	20	1	0.392	0.081	0.067	0.082	1.598	91%	21%	87%	0.11	30%	2%
		3	0.258	0.048	0.061	0.052	1.600	94%	20%	89%	0.12	31%	1%
		4	0.273	0.075	0.069	0.075	1.888	90%	25%	86%	0.11	35%	2%
30	50	1	0.209	0.039	0.056	0.039	1.605	90%	47%	88%	0.24	76%	3%
		3	0.139	0.027	0.043	0.024	1.688	94%	45%	91%	0.32	77%	1%
		4	0.155	0.038	0.049	0.037	1.688	91%	45%	88%	0.26	76%	3%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 15: Setting 3 Rural Village

Scen	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong	Weak
1	1	0.133	0.017	0.021	0.017	0.080	95.4%	33.7%	92.0%	0.23	96.2%	2.9%
	4	0.138	0.018	0.024	0.018	0.066	94.4%	31.5%	90.9%	0.19	85.0%	4.4%
2	1	0.134	0.018	0.021	0.018	0.160	95.3%	33.7%	91.9%	0.23	96.1%	3.0%
	2	0.122	0.016	0.013	0.015	0.005	96.8%	37.1%	93.5%	0.32	99.9%	7.8%
	4	0.128	0.017	0.016	0.016	0.065	96.3%	36.7%	93.0%	0.29	99.5%	6.5%
3	1	0.134	0.018	0.021	0.017	1.225	95.3%	33.6%	91.9%	0.23	95.7%	3.0%
	3	0.091	0.012	0.017	0.012	1.065	97.0%	32.2%	93.4%	0.27	96.3%	1.8%
	4	0.103	0.018	0.022	0.017	1.000	95.0%	30.8%	91.4%	0.20	84.2%	4.0%

300 Monte Carlo Simulations are performed for settings with $N = 65$. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

Table 16: Setting 4 Hierarchy Scenario (III)

N	T	Alg	A	B	G	D	AD	Nonlinks	Links	Overall	Ratio	Strong
30	20	1	0.737	0.160	0.201	0.155	1.748	89%	38%	82%	0.30	36%
		3	0.469	0.106	0.193	0.099	1.596	90%	40%	84%	0.34	39%
		4	0.469	0.140	0.186	0.134	1.776	86%	66%	83%	0.56	65%
15	20	1	0.651	0.202	0.340	0.192	1.172	83%	52%	74%	0.56	50%
		3	0.432	0.128	0.318	0.118	1.242	85%	54%	77%	0.65	53%
		4	0.454	0.165	0.303	0.161	1.285	81%	76%	79%	1.04	74%
15	50	1	0.211	0.055	0.115	0.056	0.482	89%	88%	88%	2.10	88%
		3	0.146	0.036	0.105	0.039	0.513	91%	89%	90%	2.49	88%
		4	0.172	0.055	0.110	0.055	0.473	88%	95%	90%	2.68	92%
15	100	1	0.125	0.032	0.060	0.032	0.278	91%	97%	93%	3.80	96%
		3	0.088	0.022	0.057	0.022	0.305	92%	97%	94%	4.35	96%
		4	0.105	0.032	0.063	0.032	0.353	91%	98%	93%	3.95	97%
15	200	1	0.084	0.022	0.036	0.022	0.228	93%	100%	95%	5.33	99%
		3	0.060	0.015	0.035	0.016	0.211	94%	100%	96%	6.55	99%
		4	0.070	0.022	0.037	0.022	0.226	93%	100%	95%	5.60	99%
15	400	1	0.057	0.015	0.024	0.014	0.131	94%	100%	96%	6.83	100%
		3	0.041	0.010	0.022	0.010	0.146	96%	100%	97%	10.70	100%
		4	0.046	0.015	0.022	0.014	0.170	95%	100%	96%	9.37	100%

1000 and 500 Monte Carlo Simulations are performed for settings with $N = 15$ and $N = 30$, respectively. "Deg." denotes the level of degree in each sub-period. "Alg." is the abbreviation of Algorithm. Performance measures are as denoted in section 4.1

9 Appendix B: Algorithms

Scenario (II): Algorithm 2

For each candidate break date $\tau \in (1, T)$ perform (1) - (5).

(1) Choose an initial $\beta^{(0)}$ and $b^{(0)}$ $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{d}^{(0)}$ and set m = 1.

(2) Obtain $\boldsymbol{\alpha}^{(m)}$, $\boldsymbol{a}^{(m)}$ and $\boldsymbol{\Gamma}^{(m)}$ by solving Lasso for each i for the full period.

$$(\alpha_i^{(m)}, a_i^{(m)}, \boldsymbol{\gamma}_i^{(m)}) = \underset{\alpha_i, a_i, \boldsymbol{\gamma}_i}{\operatorname{argmin}} [Q^{2i}(\alpha_i, \beta^{(m-1)}, \boldsymbol{\delta}^{(m-1)}, a_i, \beta^{(m-1)}, \boldsymbol{d}^{(m-1)}, \boldsymbol{\gamma}_i) + \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij}]$$

where Q^{2i} is the least-squares objective per individual considering the sample splits such that the subsequent time-periods are related to the time-variant parameters as shown in equation 21.

(3) Update β ,b, $\boldsymbol{\delta}$,$\boldsymbol{d}$ by OLS concerning the full data matrix.

$$(\beta^{(m)}, b^{(m)}, \boldsymbol{\delta}^{(m)}, \boldsymbol{d}^{(m)}) = \underset{\beta, b, \boldsymbol{\delta}, \boldsymbol{d}}{\operatorname{argmin}} [Q^2(\boldsymbol{\alpha}^{(m)}, \boldsymbol{a}^{(m)}, \beta, b, \boldsymbol{\delta}, \boldsymbol{d}, \Gamma^{(m)})] \tag{40}$$

(4) Set m = m + 1 and return to (2) until convergence.

(5) Store the SSR related to this τ using the converged coefficients.

$$\begin{aligned} SSR(\tau) &= Q^2(\hat{\boldsymbol{\alpha}}(\tau), \hat{\beta}(\tau), \hat{\boldsymbol{\Gamma}}(\tau), \hat{\boldsymbol{\delta}}(\tau)) + Q^2(\hat{\boldsymbol{a}}(\tau), \hat{b}(\tau), \hat{\boldsymbol{\Gamma}}(\tau), \hat{\boldsymbol{d}}(\tau)) \\ &= SSR(\tau)_{[1,\tau]} + SSR(\tau)_{[\tau+1,T]} \end{aligned}$$

(6) The break date is estimated where SSR is minimal and parameters are estimated by Post-Lasso

$$\hat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}}\ SSR(\tau). \tag{41}$$

Scenario (III): Algorithm 3

For each candidate break date $\tau \in (1, T)$ perform (1) - (5).

(1) Choose an initial $\beta^{(0)}$ and $\boldsymbol{\delta}^{(0)}$ and set m = 1.

(2) Obtain $\boldsymbol{\alpha}^{(m)}$, $\boldsymbol{\Gamma}^{(m)}$, and $\boldsymbol{G}^{(m)}$ by solving Lasso for each i over $[1, T]$.

$$(\alpha_i^{(m)}, \boldsymbol{\gamma}_i^{(m)}, \boldsymbol{g}_i^{(m)}) = \underset{\alpha_i, \boldsymbol{\gamma}_i^{(m)}, \boldsymbol{g}_i^{(m)}}{\operatorname{argmin}} [Q^{3i}(\alpha_i, \beta^{(m-1)}, \boldsymbol{\delta}^{(m-1)}, \boldsymbol{\gamma}_i, \boldsymbol{g}_i) + \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} + \lambda_{2i} \sum_{j \neq i} |g_{ij}| f_{ij}]$$

where Q^{3i} is the least-squares objective per individual considering the sample splits such that the subsequent time-periods are related to the time-variant parameters as shown in equation 22.

(3) Update β and $\boldsymbol{\delta}$ by OLS concerning the full data matrix.

$$(\beta^{(m)}, \boldsymbol{\delta}^{(m)}) = \underset{\beta, \boldsymbol{\delta}}{\operatorname{argmin}} [Q^3(\boldsymbol{\alpha}^{(m)}, \beta, \boldsymbol{\delta}, \boldsymbol{\Gamma}^{(m)}, \boldsymbol{G}^{(m)})] \tag{42}$$

(4) Set m = m + 1 and return to (2) until convergence.

(5) Store the SSR related to this τ using the converged coefficients.

$$\begin{aligned} SSR(\tau) &= Q^3(\hat{\boldsymbol{\alpha}}(\tau), \hat{\beta}(\tau), \hat{\boldsymbol{\Gamma}}(\tau), \hat{\boldsymbol{\delta}}(\tau)) + Q^3(\hat{\boldsymbol{\alpha}}(\tau), \hat{\beta}(\tau), \hat{\boldsymbol{G}}(\tau), \hat{\boldsymbol{\delta}}(\tau)) \\ &= SSR(\tau)_{[1,\tau]} + SSR(\tau)_{[\tau+1,T]} \end{aligned}$$

(6) The break date is estimated where SSR is minimal and parameters are estimated by Post-Lasso

$$\hat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}}\ SSR(\tau). \tag{43}$$

Algorithm 1: Heterogeneous Direct Effects

For each candidate break date $\tau \in (1, T)$ perform (1) - (5).

(1) Choose an initial $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{b}^{(0)}$ $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{d}^{(0)}$ and set m = 1.

(2) Obtain $\boldsymbol{\alpha}^{(m)}$, $\boldsymbol{a}^{(m)}$ and $\boldsymbol{\Gamma}^{(m)}$, $\boldsymbol{G}^{(m)}$ by solving Lasso for each i for both sub-periods.

$$(\alpha_i^{(m)}, \boldsymbol{\gamma}_i^{(m)}) = \underset{\alpha_i, \boldsymbol{\gamma}_i}{\operatorname{argmin}} Q^i_{[1,\tau]}(\alpha_i, \beta_i^{(m-1)}, \boldsymbol{\gamma}_i, \boldsymbol{\delta}^{(m-1)}) + \lambda_{1i} \sum_{j \neq i} |\gamma_{ij}| \phi_{ij} \tag{44}$$

$$(a_i^{(m)}, \boldsymbol{g}_i^{(m)}) = \underset{a_i, \boldsymbol{g}_i}{\operatorname{argmin}} Q^i_{[\tau+1,T]}(a_i, b_i^{(m-1)}, \boldsymbol{g}_i, \boldsymbol{d}^{(m-1)}) + \lambda_{2i} \sum_{j \neq i} |g_{ij}| \phi_{ij} \tag{45}$$

(3) Update $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{b}$ and $\boldsymbol{d}$ by OLS splitting the full data matrix in sub-periods.

$$(\boldsymbol{\beta}^{(m)}, \boldsymbol{\delta}^{(m)}) = \underset{\boldsymbol{\beta}, \boldsymbol{\delta}}{\operatorname{argmin}} Q_{[1,\tau]}(\boldsymbol{\alpha}^{(m)}, \boldsymbol{\beta}, \Gamma^{(m)}, \boldsymbol{\delta}) \tag{46}$$

$$(\boldsymbol{b}^{(m)}, \boldsymbol{d}^{(m)}) = \underset{\boldsymbol{b}, \boldsymbol{d}}{\operatorname{argmin}} Q_{[\tau+1,T]}(\boldsymbol{a}^{(m)}, \boldsymbol{b}, \boldsymbol{G}^{(m)}, \boldsymbol{d})] \tag{47}$$

(4) Set m = m + 1 and return to (2) until convergence.

(5) Store the SSR related to this τ using the converged coefficients.

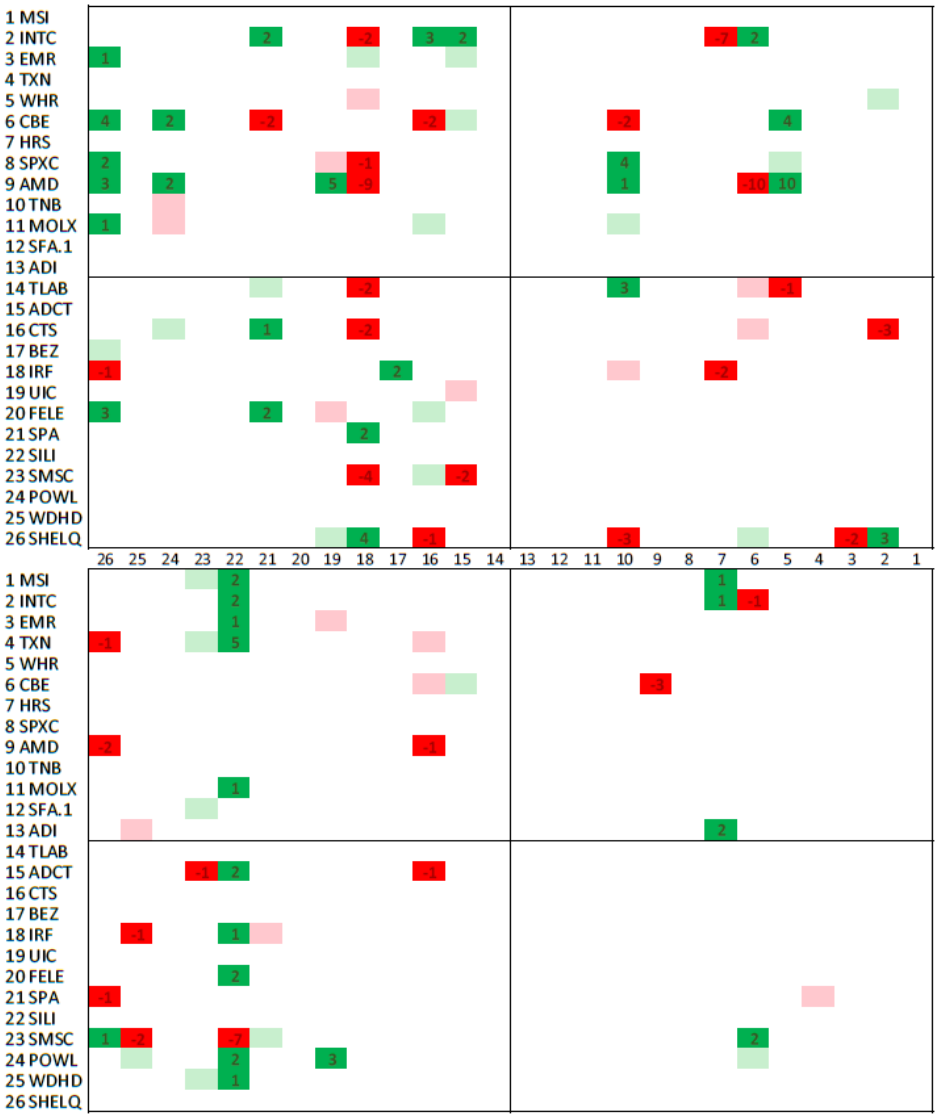
$$\begin{aligned} SSR(\tau) &= Q_{[1,\tau]}(\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau), \hat{\boldsymbol{\Gamma}}(\tau), \hat{\boldsymbol{\delta}}(\tau)) + Q_{[\tau+1,T]}(\hat{\boldsymbol{a}}(\tau), \hat{\boldsymbol{b}}(\tau), \hat{\boldsymbol{G}}(\tau), \hat{\boldsymbol{d}}(\tau)) \\ &= SSR(\tau)_{[1,\tau]} + SSR(\tau)_{[\tau+1,T]} \end{aligned}$$

(6) The break date is estimated where SSR is minimal and parameters are estimated with Post-Lasso using OLS on all regressors corresponding to nonzero parameters.

$$\widehat{\tau}^* = \underset{\tau \in (1,T)}{\operatorname{argmin}}\ SSR(\tau). \tag{48}$$

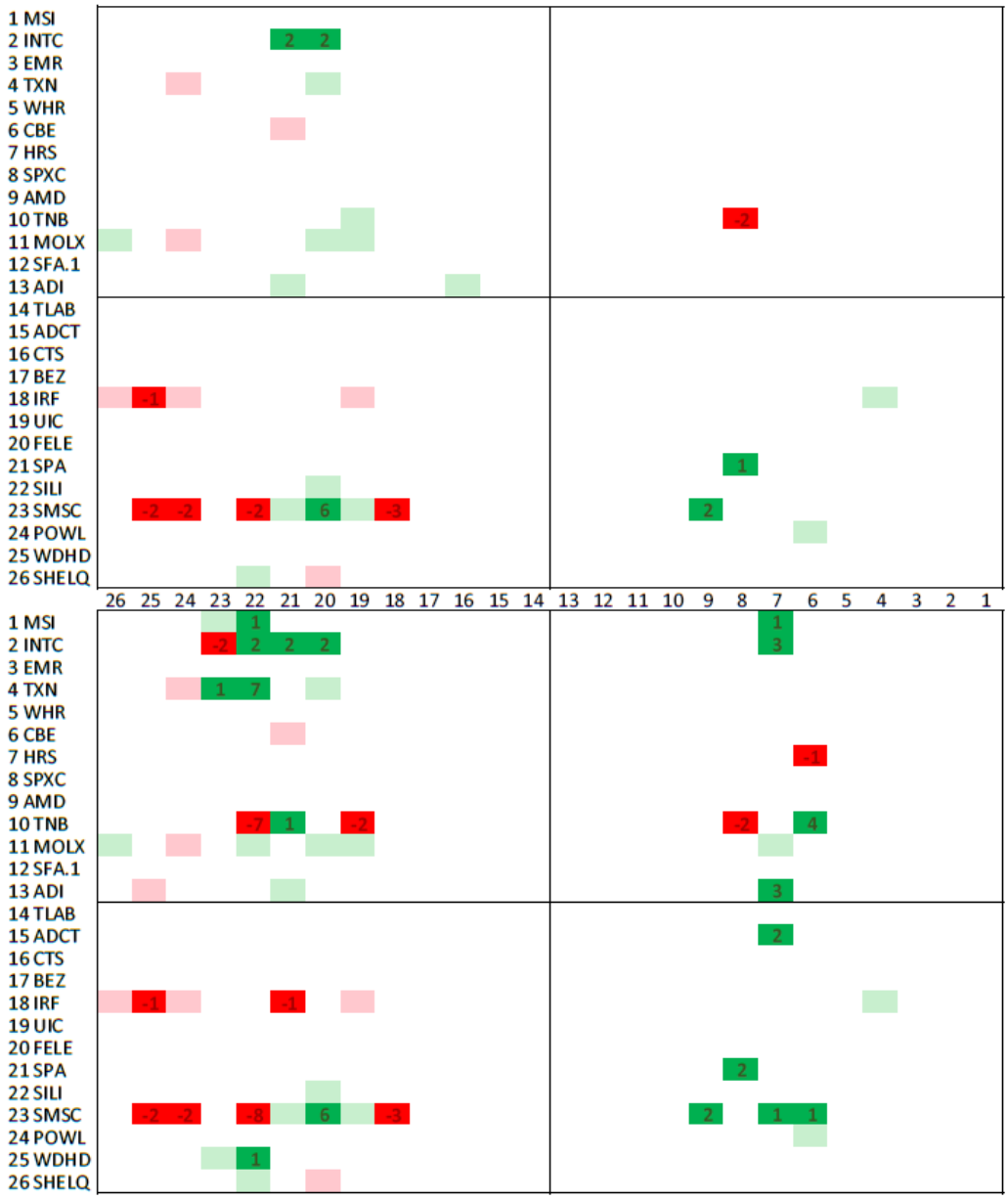
10 Appendix C: Figures

Figure 2: Algorithm 1 Setting (A): R&D Spillovers



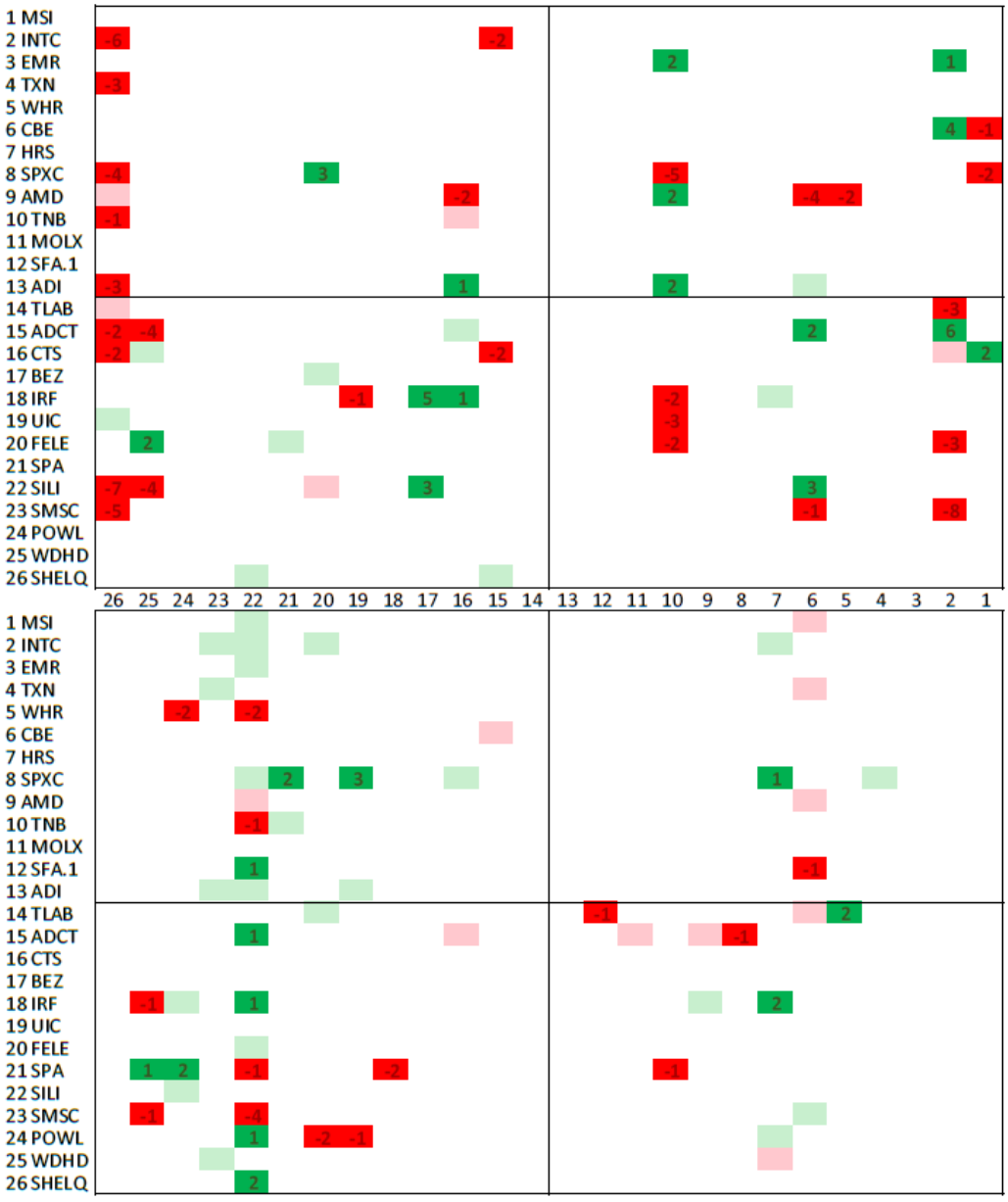
Note: The firms included are shown with their official TIC-code used in the COMPUSTAT database. Cells representing spillovers larger than 1 or smaller than −1 are filled with bright green and red, respectively. Remaining spillovers larger in absolute value than 0.3 are filled with light green or red. The upper and lower illustration represent the social interaction matrix, before and after the break. Lines separating the matrices in quadrants are included to aid inspection.

Figure 3: Algorithm 4 Setting (A): R&D Spillovers



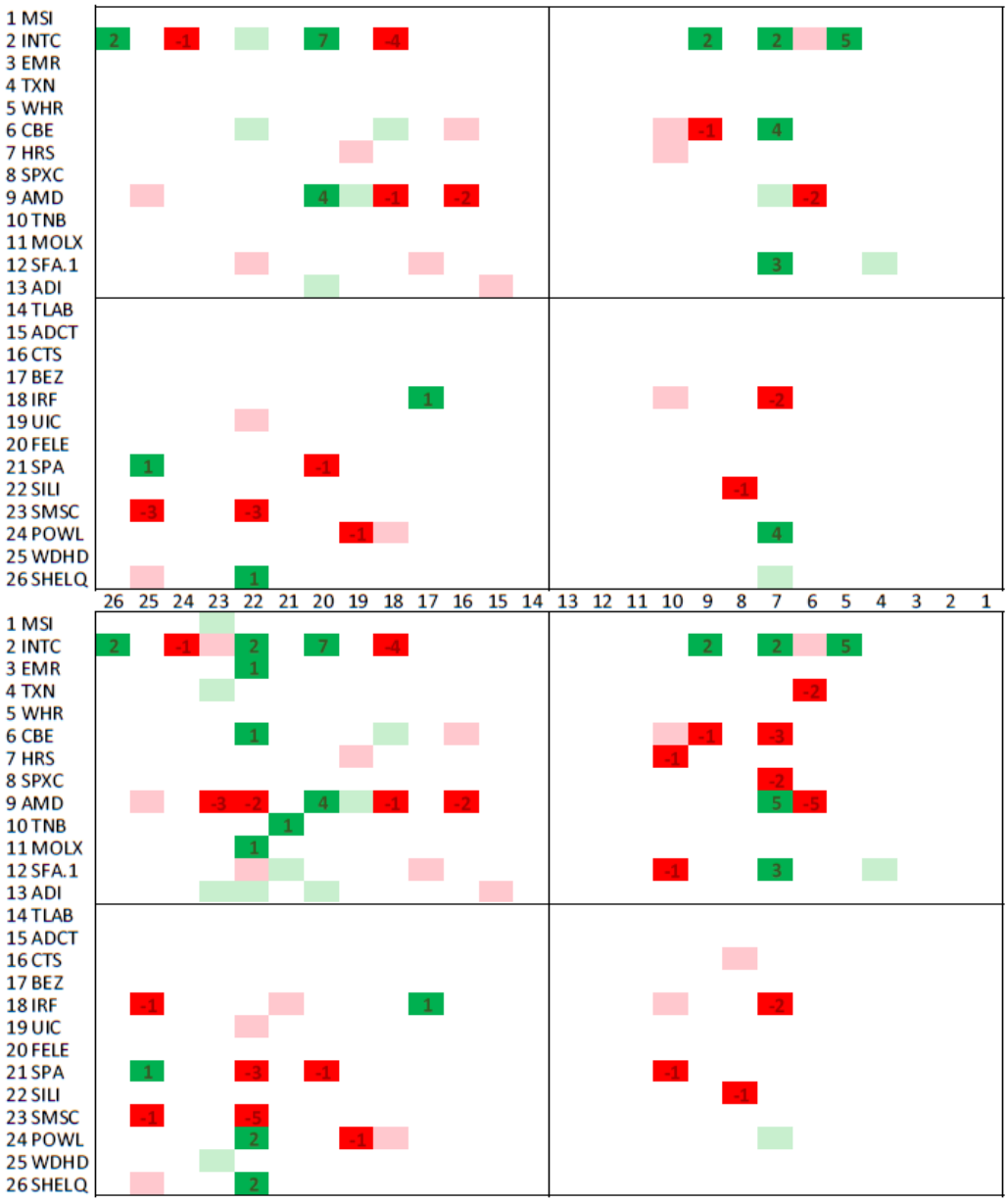
Note: The firms included are shown with their official TIC-code used in the COMPUSTAT database. Cells representing spillovers larger than 1 or smaller than -1 are filled with bright green and red, respectively. Remaining spillovers larger in absolute value than 0.3 are filled with light green or red. The upper and lower illustration represent the social interaction matrix, before and after the break. Lines separating the matrices in quadrants are included to aid inspection.

Figure 4: Algorithm 1 Setting (B): R&D Spillovers



Note: The firms included are shown with their official TIC-code used in the COMPUSTAT database. Cells representing spillovers larger than 1 or smaller than -1 are filled with bright green and red, respectively. Remaining spillovers larger in absolute value than 0.3 are filled with light green or red. The upper and lower illustration represent the social interaction matrix, before and after the break. Lines separating the matrices in quadrants are included to aid inspection.

Figure 5: Algorithm 4 Setting (B): R&D Spillovers



Note: The firms included are shown with their official TIC-code used in the COMPUSTAT database. Cells representing spillovers larger than 1 or smaller than −1 are filled with bright green and red, respectively. Remaining spillovers larger in absolute value than 0.3 are filled with light green or red. The upper and lower illustration represent the social interaction matrix, before and after the break. Lines separating the matrices in quadrants are included to aid inspection.