INFERENCE ON SOCIAL EFFECTS WHEN THE NETWORK IS UNKNOWN USING CONVEX OR LINEAR PROGRAMMING

ERIC GAUTIER

Abstract. The paper considers panel data models with social interactions when the network is unobserved and there are endogenous and exogenous social effects. The system of equations is estimated jointly by a convex program. The method does not require the knowledge of the variances of the errors nor a subgaussian assumption. It is possible to incorporate linear and sign restrictions and to impose that the matrices of endogenous and exogenous social effects have same nonzeros. The confidence sets are obtained by solving convex programs. Because the network is unknown, it is not known which exogenous variable does not have a direct effect and can be a valid excluded instrument for the endogenous variables having a direct effect. Therefore we extend the framework of Gautier and Tsybakov (2011, 2014) which handles unknown exclusion restrictions to systems of simultaneous equations. The confidence sets are robust to identification and can be infinite when there is not enough sparsity/exclusion restrictions, when, for some included endogenous regressor, all instruments are too weak or when the number of time periods is too small. The second half of the paper presents an alternative approach that only relies on linear programming. This is numerically very attractive for the study of large networks.

1. Introduction

Consider a population $i = 1, \dots, N$ of individuals, firms, countries, etc., that are observed over time $t = 1, \dots, T$. The model is a system of simultaneous equations where the outcome $y_{i,t}$ of agent i at time t is determined simultaneously with the outcome of the other agents $j \neq i$ at time t according to the following model

$$y_{i,t} = \alpha_i + \sum_{j\neq i} \beta_{j,i} y_{j,t} + \sum_{j\neq i} \gamma_{j,i} z_{j,t} + \theta^T x_{i,t} + \delta^T v_t + \epsilon_{i,t}, \tag{1.1}$$

$$(z_{j,s})_{j=1}^N, \widetilde{z}_{i,s},\ x_{i,s} \perp \epsilon_{i,t} \quad \forall s = 1, \dots, T, \tag{1.2}$$

Date: This version: April 2015.

Keywords: Instrumental variables, endogeneity, panels, simultaneous equations, heterogeneity, peer effects, networks, group sparsity, *STIV* estimator, sensitivity characteristics, high-dimensional regression, convex programming, confidence sets, variable selection, unknown variance, sign consistency.

Eric Gautier acknowledges financial support from the grants ANR-13-BSH1-0004 and ERC POEMH.

$$\epsilon_{i,t} \text{ are i.i.d. with } \mathbb{E}[\epsilon_{i,t}] = 0,\ \mathbb{E}[\epsilon_{i,t}^4] < \infty \text{ for } t = 1, \ldots, T,\ i = 1, \ldots, N \tag{1.3}$$

where $\perp$ denotes independence. The fixed effect α_i accounts for the unobserved heterogeneity of agents that is fixed over time. It could be thought as a random variable which could be arbitrarily dependent with $\epsilon_{i,t}$ and the right-hand side variables appearing in (1.1). Writing $\alpha_i = \alpha + \eta_i$ where $\mathbb{E}[\eta_i] = 0$, η_i and η_j for $i \neq j$ could be arbitrarily correlated so that (1.3) allows for arbitrary correlation accross i between the composite errors $\eta_i + \epsilon_{i,t}$ of the system of simultaneous equations. Also $\eta_i + \epsilon_{i,t}$ can be heteroskedastic accross i. The fixed effect also allows for a situation where agents are influenced by the random shocks $\epsilon_{j,1}$ of other agents at time 1 but not later in time. The assumption that $\epsilon_{i,t}$ are i.i.d. and independent of the exogenous variables is maintained to use a bootstrap method and obtain confidence sets which are not too conservative as well as for the simplicity of the arguments to handle the serial dependence of the instruments. It is not essential for identification.

The term $\theta^T x_{i,t}$ is of the form $\eta z_{i,t} + \tilde{\theta}^T \widetilde{x}_{i,t}$ where $\eta z_{i,t}$ accounts for the direct effect of the exogenous variable $z_{i,t}$ of agent i on the outcome $y_{i,t}$ and $\tilde{x}_{i,t}$ is a vector of control variables to justify the exogeneity of $(z_{j,s})_{j=1}^N$. The vector $x_{i,t}$ is of size p. The vector v_t is a vector of functions of t of size q. The vector $\widetilde{z}_{i,t}$ of dimension $l_i \times 1$ is a vector of variables which are independent of $\epsilon_{i,t}$ and serve to instrument the endogenous variables $y_{i,t}$. It can contain for example: lagged, present and even leads of these exogenous variables, lagged values of the variables $y_{i,t}$ (if we further assume that α_i is independent from of $\epsilon_{i,t}$) or lagged first differences $y_{i,s-1} - y_{i,s-2}$ for $s \leq t$, as well as transformations of them.

The coefficients $\beta_{i,j}$ (resp. $\gamma_{i,j}$) account for the individual endogenous (resp. exogenous) effects of each other agent j in the population on the outcome for agent i. Some of these coefficients could be zero in the absence of direct effect. This means that the true model is smaller

$$y_{i,t} = \alpha_i + \sum_{j \in \mathcal{P}_i} \beta_{i,j} y_{j,t} + \sum_{j \in \mathcal{P}_i} \gamma_{i,j} z_{j,t} + \theta^T x_{i,t} + \delta^T v_t + \epsilon_{i,t} \tag{1.4}$$

where $\mathcal{P}_i$ is the set of agents directly affecting the outcome of agent i. Because the network is not known by the econometrician, one cannot estimate the smaller model directly. For this reason we consider the high-dimensional model (1.1) which contains the true parsimonious model as a submodel. When one knows the groups $\mathcal{P}_i$ then one has valid instruments for the endogenous variables y_j for j in $\mathcal{P}_i$. One simply uses the exogenous variables of those who are not in $\mathcal{P}_i$. The strength of these instrumental variables depends on their correlation with the endogenous variables that have a nonzero coefficient. Because we consider a setup where the groups $\mathcal{P}_i$ are unknown, one does not know which exogenous variable could be used to instrument the outcomes of those in $\mathcal{P}_i$. As we shall see, the moment condition $\mathbb{E}[w\epsilon_{i,t}] = 0$ where w is a vector of all exogenous variables, gives rise to overdetermined systems for every sufficiently sparse model of the form (1.4) so that the model could be identified without relying on exogenous variables outside those of the model. The confidence sets of this paper, like those in Gautier and Tsybakov (2011, 2014), are robust to identification. The confidence sets can also be infinite when the instruments are too weak or when the true model is not sparse enough to yield overidentification. They are solutions of convex (or linear) programs and we do not rely on a pretest nor test inversion at every possible value of the parameters.

Because we assume independence between the exogenous variables and the errors, Gautier and Tsybakov (2011, 2014) provides a solution to estimate a model where the direct effect η of the own exogenous regressor $z_{i,t}$, the correlated effect $\widetilde{\theta}$ and the time effect δ are heterogeneous (*i.e.*, can be different accross agents and thus are indexed by i). One simply estimates each equation separately using the *STIV* estimator. Heterogeneity of the coefficients implies a very high-dimensional model. Also, we assume that the network is fixed over time. This is a strong assumption which is more likely to hold over short periods. It is also relatively rare to encounter data sets where T is large. For all these reasons our baseline model (1.1) with the constraints(1.5), (1.6), (1.7) below is relatively homogeneous. A section of this paper explains how to modify the proposed procedure to handle other specifications among which heterogeneous coefficients η_i, $\widetilde{\theta}_i$ and/or δ_i, models where we replace the scalar variables $z_{i,j}$ by vectors, models with high-dimensional vectors of controls or time effects or autoregressive models.

We first impose a structure that is present in (1.4). We impose that

$$\text{The matrices } (\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}} \text{ and } (\gamma_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}} \text{ have the same zeros} \tag{1.5}$$

where by convention $\beta_{i,i} = \gamma_{i,i} = 0$ for $i = 1, \dots, N$. This is called *group sparsity.*

Also, it is usual in the peer effects literature (see, *e.g.*, Manski (1997), Lee (2007), Bramoullé et al (2009) and Davezies et al (2009)) to consider more homogeneous specifications. For example in Bramoulé (2009) one has that $\beta_{i,j} = \overline{\beta}\mathbb{1}_{j\in\mathcal{P}_i}/s_i$ and $\gamma_{i,j} = \overline{\gamma}\mathbb{1}_{j\in\mathcal{P}_i}/s_i$ where s_i is the cardinality of the set $\mathcal{P}_i$. Because the procedure in this paper is based on linear programming ideas we impose the following constraints which are in the spirit of the model considered thus far in the peer effects literature

$$\forall i \neq k,\ \sum_{j\neq i}\beta_{i,j} = \sum_{j\neq k}\beta_{k,j} \text{ and } \sum_{j\neq i}\gamma_{i,j} = \sum_{j\neq k}\gamma_{k,j}, \tag{1.6}$$

$$\exists (\xi_1, \xi_2) \in \{-1, 1\}^2 : \ \forall i, j = 1, \dots, N, \ \xi_1 \beta_{i,j} \geq 0 \text{ and } \xi_2 \gamma_{i,j} \geq 0 \ . \tag{1.7}$$

Under these constraints the matrices $(\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ and $(\gamma_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ are proportional to weighting matrices where the weights are nonnegative and the row sum of the weight matrices is 1. The proportionality constants $\overline{\beta} = \sum_{j \neq i} \beta_{i,j}$ and $\overline{\gamma} = \sum_{j \neq i} \gamma_{i,j}$ for arbitrary $i \in \{1, \dots, N\}$ are the endogenous and exogenous social effects. In Bramoullé et al. (2009) all nonzeros entries on a row of the matrices are equal. This structure is not imposed in this paper to allow estimation using a simple linear or convex program.

Considering that the fixed effect α_i is a nuisance parameter, the dimensionality of this problem can be slightly reduced by taking first differences. This yields the following equations for $t = 2, \dots, T$

$$\Delta(y_i)_t = \Delta \left(\sum_{j \neq i} \beta_{i,j} y_j + \sum_{j \neq i} \gamma_{i,j} z_j + \theta^T x_i + \delta^T v + \epsilon_i \right)_t , \tag{1.8}$$

where we denote for a vector valued sequence $(u_t)_{t=1,\dots,T}$ and $t = 2, \dots, T$, $\Delta(u)_t = u_t - u_{t-1}$. Taking into account the homogeneity constraints (1.6), this equation has $2(N^2 - 2N) + p + q - 1$ unknowns. But the number of nonzero coefficients $2 \sum_{i=1}^N (s_i - 1) - 1 + p + q$ can be much smaller. In particular it could be linear in N if most agents have few connections. The method proposed in this paper has rates of convergence which, up to a logarithmic factor in the number of instruments (possibly invalid) and a constant, are the same as if one knew the identity of the nonzero coefficients and were estimating the low dimensional model. We also provide one type of confidence set that also shares this "adaptivity" property.

Related literature includes the peer effects literature already mentionned as well as the literature on spatial models (see in particular Lee and Yu (2010) which considers a spatial autoregressive panel data models with fixed effects).

Estimation of sparse networks is an active field of research in statistics. The more closely related model is the Gaussian Structural Equation Model (SEM). There one observes an i.i.d. sample of vectors $(y_{i,t})_{i=1,\dots,N}$ from the model

$$y_{i,t} = \sum_{j \neq i} \beta_{i,j} y_{j,t} + \epsilon_{i,t}, \ i = 1, \dots, N \ .$$

Identification relies strongly on the fact that the errors are iid normal and the directed acyclic graph (DAG) structure (see Peters and Bühlmann (2014)). The DAG structure means that $\boldsymbol{\beta} = P \mathcal{T} P^T$ where $\boldsymbol{\beta}$ is the transpose of the matrix $(\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ with $\beta_{i,i} = 0$, $\mathcal{T}$ is strictly lower triangular and P is a permutation matrix. This precludes for example that y_i has direct effect on y_j and that y_j has direct effect on y_i which typically happens in Economics in a system of simultaneous equations. The estimation procedures are currently very demanding: van de Geer and Bühlmann (2013) considers a ℓ_0 penalized log likelihood estimator and Champion (2015) a ℓ_1 penalized log likelihood estimator where the estimation is constrained to matrices $\boldsymbol{\beta}$ that are compatible with the direct acyclic graph structure (this is a nonconvex set).

A few articles in econometrics are closely related. Manresa (2013) studies the case when there is no endogeneous effects. Lam and Souza (2014) allows for endogenous effects but does not use instrumental variables. It requires much more assumptions on the data generating process. Among others it is assumed that the variance of the errors $\epsilon_{i,t}$ decays to zero uniformly in i when t goes to infinity, that $\max_i \max\{\sum_{j\neq i}|\beta_{i,j}|, \sum_{j\neq i}|\gamma_{i,j}|\} < 1$. Both Manresa (2013) and Lam and Souza (2014) rely on an iterative procedure for which, up to our knowledge, the properties are not well understood. Rose (2015) considers a fully heterogeneous model that can be estimated equation by equation using the *STIV* estimator of Gautier and Tsybakov (2014). This paper also contributes to the recent literature on high-dimensional panel data models (*c.f.* Belloni et al. (2014) and the references therein) by allowing for a high-dimensional vector of endogenous regressors. The first approach of this paper is based on convex programs. There are widely available solvers as well as simple interfaces (for example CVX, *c.f.* Grant and Boyd (2013)) to solve convex programs. The linear programming solution is an important extension to handle many moments and to calculate the confidence sets. It builds on Gautier and Tsybakov (2013). Linear programming is much faster than Lasso and in contrast with Gautier and Tsybakov (2013) we propose a method that also allows to obtain robust confidence sets using linear programming.

2. Identification under sparsity

Consider the model for the outcome of agent i. Condition (1.2) leads to the linear system

$$\mathbb{E}\left[\Delta\left(w\right)_t \Delta\left(\sum_{j\neq i}\beta_{i,j}y_j + \sum_{j\neq i}\gamma_{i,j}z_j + \theta^T x_i + \delta^T v\right)_t\right] = \mathbb{E}\left[\Delta\left(w\right)_t \Delta(y_i)_t\right] \tag{2.1}$$

where we denote for $t = 1,\dots,T$ by w_t the vector where we stack all $(z_{i,t})_{i=1}^N$, $x_{i,t}$ and v_t. Due to the independence assumption made in (1.2) one could add an infinite number of moment conditions

$$\mathbb{E}\left[\phi(w_1,\dots,w_T)\Delta\left(\sum_{j\neq i}\beta_{i,j}y_j + \sum_{j\neq i}\gamma_{i,j}z_j + \theta^T x_i + \delta^T v\right)_t\right] = \mathbb{E}\left[\phi(w_1,\dots,w_T)\Delta(y_i)_t\right] \tag{2.2}$$

where ϕ is an arbitrary function. When α_i is independent of $\epsilon_{i,t}$ then $y_{j,1}, \ldots, y_{j,t-1}$ for $j = 1, \ldots, N$ are independent of $\epsilon_{i,t}$ and can be included as arguments of the function ϕ. When α_i is not independent of $\epsilon_{i,t}$ then $y_{j,2} - y_{j,1}, \ldots, y_{j,t-1} - y_{j,t-2}$ for $j = 1, \ldots, N$ are independent of $\epsilon_{i,t}$ and can instead be included as arguments of the function ϕ.

In the analysis that follows we do not include the extra moment conditions (2.2). The aim is to show that the strong exogeneity assumption (1.2) is not essential for identification. We do not account for the restrictions (1.6) and (1.7) which could have identification power. The analysis follows Gautier and Tsybakov (2011, 2014) and shows that sparsity can yield identification.

Equation (2.1) corresponds to $N + p + q$ equations and $2N - 2 + p + q$ unknowns. Imposing $N - 2$ independent linear restrictions on the parameters can yield identification. Imposing more than $N - 1$ independent linear restrictions on the parameters leads potentially to overidentification. An example of such restrictions consists in knowing a priori that some coefficients among $\beta_{i,j}$ and $\gamma_{i,j}$ are zero. Imposing that some of the $\gamma_{j,i}$ are zero means that there are (unknown) exclusion restrictions. When $\gamma_{k,i} = 0$, $\Delta(z_k)_t$ can be used to instrument the right-hand side endogenous outcome of another agent, say $\Delta(y_j)_t$ for $j \neq k, i$. For it to be a relevant instrument one needs the partial correlation between $\Delta(z_k)_t$ and $\Delta(y_j)_t$ to be nonzero. This is similar in spirit to what is called an intransitive triad in Bramoullé (2009). Assume now that $(\beta_{i,j})_{j \neq i}$ and $(\gamma_{i,j})_{j \neq i}$ both have s_i nonzeros, that it corresponds to the same coordinates, and that it is known which coefficients are zero. This yields a system of $N + p + q$ equations with $2s_i + p + q$ unknowns so that one could have overidentification, knowing the restrictions, when $2s_i < N$.

If one does not know the identity of the nonzero coordinates but imposes an upper bound s_i on their number per vector where $s_i < N - 1$, then there are $\binom{s_i}{N-1}$ systems with $2s_i + p + q$ unknowns and $N + p + q$ equations of the form

$$\mathbb{E}\left[\Delta\left(w\right)_t \Delta\left(\sum_{j \in J} \beta_{i,j} y_j + \sum_{j \in J} \gamma_{i,j} z_j + \theta^T x_i + \delta^T v\right)_t\right] = \mathbb{E}\left[\Delta\left(w\right)_t \Delta(y_i)_t\right] \tag{2.3}$$

where J is a subset of $\{j \in \{1, \ldots, N\} : j \neq i\}$ of size s_i. The image of the matrices

$$\mathbb{E}\left[\Delta\left(w\right)_t \left(\Delta(y)_{J,t}^T, \Delta(z)_{J,t}^T, \Delta(x_i)_t^T, \Delta(v)_t^T\right)\right],$$

where $\Delta(u)_{J,t} = (\Delta(u)_{j,t})_{j \in J}$, is a sub vector space of $\mathbb{R}^N$ of dimension at most $2s_i + p + q$. Its Lebesgue measure is 0. Identification is obtained when the vector $\mathbb{E}\left[\Delta\left(w\right)_t \Delta(y_i)_t\right]$ lies only in the space corresponding to the set J of the true model and for that J the matrix has full rank (*i.e.* $2s_i + p + q$). Sparsity can thus yield identification in a case where exclusion restrictions are not known in advance and one allows all exogenous variables to have a direct effect on the outcome assumption. The condition for identification could in principle be tested but considering all possible submatrices is NP-hard. Rather we propose an inference method that is robust to identification. This yields infinite confidence sets when one is at the verge of identification due to weak instruments or does not impose enough sparsity or T is too small.

3. Notations

Denote by I_d the identity of size d, by $\mathbf{1}$ a vector of ones and by $\mathbf{0}$ a vector of zeros. For a sequence $(t_{k,t})_{\substack{k=1,\dots,K\\t=1,\dots,T}}$ we denote by $\boldsymbol{T}$ the transpose of the $T\times K$ matrix with corresponding elements, for example we write

$$\boldsymbol{Y}=\begin{pmatrix} y_{1,1} & \cdots & y_{N,1}\\ \vdots & & \vdots\\ y_{1,T} & \cdots & y_{N,T}\end{pmatrix}.$$

For such a sequence $(t_{k,t})_{\substack{k=1,\dots,K\\t=1,\dots,T}}$ we denote by $\boldsymbol{T_{-k}}$ the $T\times(K-1)$ the matrix where one suppresses the k^{th} column and by $\boldsymbol{T_k}$ the $T\times 1$ matrix where one only keeps the k^{th} column. As a result $\boldsymbol{T_{-k,l}}$ is the l^{th} column of $\boldsymbol{T_{-k}}$. The case of $x_{i,t}$ and $\widetilde{z}_{i,t}$ is specific since they are vectors, and we use the notation

$$\boldsymbol{X_i}=\begin{pmatrix} (x_{i,1})_1 & \dots & (x_{i,1})_p\\ \vdots & & \vdots\\ (x_{i,T})_1 & \dots & (x_{i,T})_1\end{pmatrix},\quad \widetilde{\boldsymbol{Z}}_{\boldsymbol{i}}=\begin{pmatrix} (\widetilde{z}_{i,1})_1 & \dots & (\widetilde{z}_{i,1})_{l_i}\\ \vdots & & \vdots\\ (\widetilde{z}_{i,T})_1 & \dots & (\widetilde{z}_{i,T})_{l_i}\end{pmatrix}.$$

We also denote by $\boldsymbol{\beta}$ (resp. $\boldsymbol{\gamma}$) the transpose of the matrix $(\beta_{i,j})_{\substack{i=1,\dots,N\\j=1,\dots,N}}$ (resp. $(\gamma_{i,j})_{\substack{i=1,\dots,N\\j=1,\dots,N}}$) with $\beta_{i,i}=0$ (resp. $\gamma_{i,i}=0$) for $i=1,\dots,N$ and by $\boldsymbol{\beta_{-i,i}}$ (resp. $\boldsymbol{\gamma_{-i,i}}$) the i^{th} column of $\boldsymbol{\beta}$ (resp. $\boldsymbol{\gamma}$) where one removes the i^{th} row. We also denote by $\Theta=(B^T,G^T,\theta^T,\delta^T)^T$ the vector of unknown parameters where $B=(\boldsymbol{\beta_{-1,1}}^T,\dots,\boldsymbol{\beta_{-N,N}}^T)^T$ and $G=(\boldsymbol{\gamma_{-1,1}}^T,\dots,\boldsymbol{\gamma_{-N,N}}^T)^T$.

We denote by $|\cdot|_p$ the usual ℓ_p norms for $1\le p\le\infty$ and define the $|\cdot|_{2,1}$-norm of the matrix (B,G) by

$$|(B,G)|_{2,1}=\sum_{k=1}^{N(N-1)}\sqrt{B_k^2+G_k^2}\,.$$

This mixed norm is a convex criterion that is used to ensure group sparsity in the model that we study. In contrast the mixed norm $|\cdot|_{2,0}$-norm of the vector (B,G) which is defined as the number of nonzeros of the vector $\left(\sqrt{B_k^2+G_k^2}\right)_{k=1}^{N(N-1)}$ is not convex. In Section 5 the $|\cdot|_{\infty,1}$-norm is considered instead. It is defined for a matrix (B,G) by

$$|(B,G)|_{\infty,1} = \sum_{k=1}^{N(N-1)} \max\left\{|B_k|,|G_k|\right\} \ .$$

The Helmert's transformation (see, *e.g.*, Arellano and Bover (1995)) is defined as

$$\widetilde{D} = (DD^T)^{-1/2}D$$

where

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}$$

of dimension $(T-1)\times T$ and the square-root is the upper triangular Cholesky factorization. Recall that ordinary least squares when $\widetilde{D}$ is applied to the data corresponds to generalized least squares when the transformation D is used. Recall that $\widetilde{D}\mathbf{1}=0$, $\widetilde{D}^T\widetilde{D} = I_T - \frac{1}{T}\mathbf{1}\mathbf{1}^T = D^T(DD^T)^{-1}D$ is the within-group operator which is a projection matrix, that $\widetilde{D}\widetilde{D}^T = I_{T-1}$ and that for $t=1,\dots,T-1$,

$$\left(\widetilde{D}\boldsymbol{\epsilon_i}\right)_t = \sqrt{\frac{T-t}{T-t+1}}\left(\epsilon_{i,t} - \frac{\sum_{s=t+1}^T \epsilon_{i,s}}{T-t}\right) \ .$$

Because of this expression we can see that lagged values of the outcome variables $y_{i,t}$ can be used as supplemental instruments when α_i is assumed independent of $\epsilon_{i,t}$. Indeed

$$\forall s<t,\ \mathbb{E}\left[y_{i,s}\left(\widetilde{D}\boldsymbol{\epsilon_i}\right)_t\right] = \mathbb{E}\left[y_{i,s}\left(\widetilde{D}(\boldsymbol{\epsilon_i}+\alpha_i\mathbf{1})\right)_t\right] = 0 \ .$$

For example one could take $\widetilde{z}_{i,1}=0$ and for $t=2,\dots,T-1$, $\widetilde{z}_{i,t}=y_{i,t-1}$; $\widetilde{z}_{i,1}=0$, $\widetilde{z}_{i,2}=0$ and $t=3,\dots,T-1$, $\widetilde{z}_{i,t}=y_{i,t-2}$; $\widetilde{z}_{i,1}=0$ and for $t=2,\dots,T-1$, $\widetilde{z}_{i,t}=y_{j,t-1}$ for $j\neq i$, etc. One has

$$\mathbb{E}\left[\boldsymbol{\widetilde{z}_i}^T\widetilde{D}\boldsymbol{\epsilon_i}\right] = \mathbb{E}\left[\boldsymbol{\widetilde{z}_i}^T\widetilde{D}\left(\boldsymbol{\epsilon_i}+\alpha_i\mathbf{1}\right)\right] = \mathbb{E}\left[\left(\widetilde{D}^T\boldsymbol{\widetilde{z}_i}\right)^T\left(\boldsymbol{\epsilon_i}+\alpha_i\mathbf{1}\right)\right] = 0 \ .$$

When α_i is not independent of $\epsilon_{i,t}$ then one could use as instruments $y_{i,s}-y_{i,s-1}$ for $2\le s\le t-1$.

We use the notation diag$\left\{(a_k)_{k=1,\dots,d}\right\}$ to denote the block diagonal matrix with block diagonal elements given by the matrices of the sequence $(a_k)_{k=1,\dots,d}$. Let us introduce the following normalization matrices. They are introduced to obtain a procedure that is invariant to the scale of the regressors $\boldsymbol{D^R} = \mathrm{diag}\left\{\left(\boldsymbol{D^B},\boldsymbol{D^G},\boldsymbol{D^X},\boldsymbol{D^V}\right)\right\}$ and instruments $\boldsymbol{D^I} = \mathrm{diag}\left\{\left(\boldsymbol{D^G},\boldsymbol{D^{\widetilde{Z}}},\boldsymbol{D^X},\boldsymbol{D^V}\right)\right\}$, where

$$\boldsymbol{D^B} = \operatorname{diag}\left\{\left(\boldsymbol{D^{Y_{-i}}}\right)_{i=1,\dots,N}\right\},\ \boldsymbol{D^G} = \operatorname{diag}\left\{\left(\boldsymbol{D^{Z_{-i}}}\right)_{i=1,\dots,N}\right\},\ \boldsymbol{D^{\widetilde{Z}}} = \operatorname{diag}\left\{\left(\boldsymbol{D^{\widetilde{Z}_i}}\right)_{i=1,\dots,N}\right\},$$

$$\boldsymbol{D^{Y_{-i}}} = \operatorname{diag}\left\{\left(\left(\frac{1}{T}\boldsymbol{Y_{-i,k}}^T \widetilde{D}^T \widetilde{D} \boldsymbol{Y_{-i,k}}\right)^{-1/2}\right)_{k=1,\dots,N-1}\right\},$$

$$\boldsymbol{D^{Z_{-i}}} = \operatorname{diag}\left\{\left(\left(\frac{1}{T}\boldsymbol{Z_{-i,k}}^T \widetilde{D}^T \widetilde{D} \boldsymbol{Z_{-i,k}}\right)^{-1/2}\right)_{k=1,\dots,N-1}\right\},$$

$$\boldsymbol{D^{\widetilde{Z}_i}} = \operatorname{diag}\left\{\left(\left(\frac{1}{T-1}\boldsymbol{\widetilde{Z}_{i,k}}^T \boldsymbol{\widetilde{Z}_{i,k}}\right)^{-1/2}\right)_{k=1,\dots,l_i}\right\},$$

and

$$\boldsymbol{D^X} = \operatorname{diag}\left\{\left(\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X_{i,k}}^T \widetilde{D}^T \widetilde{D} \boldsymbol{X_{i,k}}\right)^{-1/2}\right)_{k=1,\dots,p}\right\},$$

$$\boldsymbol{D^V} = \operatorname{diag}\left\{\left(\left(\frac{1}{T}\boldsymbol{V_k}^T \widetilde{D}^T \widetilde{D} \boldsymbol{V_k}\right)^{-1/2}\right)_{k=1,\dots,q}\right\}.$$

Define the $T \times (N-1)$ matrix $\boldsymbol{\overline{Y}_{-i}} = \boldsymbol{D^{Y_{-i}}} \widetilde{D}^T \widetilde{D} \boldsymbol{Y_{-i}}$ (resp. $\boldsymbol{\overline{Z}_{-i}} = \boldsymbol{D^{Z_{-i}}} \widetilde{D}^T \widetilde{D} \boldsymbol{Z_{-i}}$) and denote its elements by $\overline{y}_{-i,k,t} = \left(\boldsymbol{\overline{Y}_{-i,k}}\right)_t$ (resp. $\overline{z}_{-i,k,t} = \left(\boldsymbol{\overline{Z}_{-i,k}}\right)_t$). Define the $T \times l_i$ matrix $\boldsymbol{\overline{\widetilde{Z}}_i} = \boldsymbol{D^{\widetilde{Z}_i}} \widetilde{D}^T \widetilde{D} \boldsymbol{\widetilde{Z}_i}$ and denote its elements by $\overline{\widetilde{z}}_{i,t,k} = \left(\boldsymbol{\overline{\widetilde{Z}}_i}\right)_{t,k}$. Define the $T \times p$ matrix $\boldsymbol{\overline{X}_i} = \boldsymbol{D^X} \widetilde{D}^T \widetilde{D} \boldsymbol{X_i}$ and denote its elements by $\overline{x}_{i,t,k} = \left(\boldsymbol{\overline{X}_i}\right)_{t,k}$. Finally define the $T \times q$ matrix $\boldsymbol{\overline{V}} = \boldsymbol{D^V} \widetilde{D}^T \widetilde{D} \boldsymbol{V}$ and denote its elements by $\overline{v}_{t,k} = \left(\boldsymbol{\overline{V}}\right)_{t,k}$.

The following matrix

$$\Psi = \frac{1}{T-1}\begin{pmatrix} \boldsymbol{\overline{Z}_{-1}}^T\left(\boldsymbol{\overline{Y}_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Z}_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{X}_1}, \boldsymbol{\overline{V}}\right) \\ \vdots \\ \boldsymbol{\overline{Z}_{-N}}^T\left(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Y}_{-N}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Z}_{-N}}, \boldsymbol{\overline{X}_N}, \boldsymbol{V}\right) \\ \boldsymbol{\overline{\widetilde{Z}}_1}^T\left(\boldsymbol{\overline{Y}_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Z}_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{X}_1}, \boldsymbol{\overline{V}}\right) \\ \vdots \\ \boldsymbol{\overline{\widetilde{Z}}_N}^T\left(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Y}_{-N}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\overline{Z}_{-N}}, \boldsymbol{\overline{X}_N}, \boldsymbol{\overline{V}}\right) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\overline{X}_i}^T\left(\boldsymbol{\overline{Y}_{-1}}, \dots, \boldsymbol{\overline{Y}_{-N}}, \boldsymbol{\overline{Z}_{-1}}, \dots, \boldsymbol{\overline{Z}_{-N}}, \boldsymbol{\overline{X}_i}, \boldsymbol{\overline{V}}\right) \\ \frac{1}{\sqrt{N}}\boldsymbol{\overline{V}}^T\left(\boldsymbol{\overline{Y}_{-1}}, \dots, \boldsymbol{\overline{Y}_{-N}}, \boldsymbol{\overline{Z}_{-1}}, \dots, \boldsymbol{\overline{Z}_{-N}}, \sum_{i=1}^N \boldsymbol{\overline{X}_i}, \boldsymbol{\overline{V}}\right) \end{pmatrix}.$$

plays an important role in the analysis. Its dimension is $L \times (2N(N-1)+p+q)$ where $L = \sum_{i=1}^N (N-1+l_i)+p+q$.

For a vector $V \in \mathbb{R}^K$, let $J(V) = \{k \in \{1,\dots,K\}: \ V_k \neq 0\}$ be its support. Denote by $|J|$ the cardinality of a set $J \subseteq \{1,\dots,K\}$ and by J^c its complement: $J^c = \{1,\dots,K\} \setminus J$. Denote by $J_{\rm end} = \{1,\dots,N(N-1)\}$ and by $J_{\rm ex} = \{N(N-1)+1,\dots,2N(N-1)+p+q\}$ the indices of the coefficients of the endogenous variables, respectively exogenous variables in Θ. Given a subset J of $J_{\rm end}$, denote $J_{\rm ex}^{\cap J} = (N(N-1)+J) \cup \{2N(N-1)+1,\dots,2N(N-1)+p+q\}$ where $N(N-1)+J$ is the set of indices in J to which we add $N(N-1)$. For $V \in \mathbb{R}^K$ and a set of indices $J \subseteq \{1,\dots,K\}$, define $V_J = (V_1 1\!\!1_{\{1\in J\}},\dots,V_K 1\!\!1_{\{K\in J\}})^T$, where $1\!\!1_{\{\cdot\}}$ is the indicator function. For a vector $V \in \mathbb{R}^K$, set $\overrightarrow{\mathrm{sign}(V)} = (\mathrm{sign}(V_1),\dots,\mathrm{sign}(V_K))$ where

$$\mathrm{sign}(t) = \begin{cases} 1 & \text{if } t>0 \\ 0 & \text{if } t=0 \\ -1 & \text{if } t<0 \end{cases}$$

For $a \in \mathbb{R}$, set $a_+ = \max(0,a)$ and $1/0 = \infty$. The convention $\inf \emptyset = \infty$ is used throughout the text.

Denote the identified set by

$$\mathcal{I}dent = \left\{ \Theta : \begin{array}{l} \mathbb{E}\left[\Delta(w)_t \Delta\left(\sum_{j\neq i}\beta_{i,j}y_j + \sum_{j\neq i}\gamma_{i,j}z_j + \theta^T x_i + \delta^T v\right)_t\right] = \mathbb{E}\left[\Delta(w)_t \Delta(y_i)_t\right] \\ \mathbb{E}\left[\widetilde{z}_{i,t}\Delta\left(\sum_{j\neq i}\beta_{i,j}y_j + \sum_{j\neq i}\gamma_{i,j}z_j + \theta^T x_i + \delta^T v\right)_t\right] = \mathbb{E}\left[\widetilde{z}_{i,t}\Delta(y_i)_t\right] \end{array} \right\}. \tag{3.1}$$

We will sometimes restrict the class of models to sparse models and make inference on the *sparse identifiable parameters*:

$$\mathcal{B}_s = \mathcal{I}dent \bigcap \{\Theta : \ |J(B)| = |J(G)| \le s\}$$

for some upper bound s in $\{1,\dots,N(N-1)\}$ on the sparsity. This is the set of vectors of coefficients compatible with (1) the moment restrictions and (2) a prior upper bound on the number of non-zero coefficients. These sets satisfy

$$\forall s \le s' \le N(N-1), \quad \mathcal{B}_s \subseteq \mathcal{B}_{s'} \subseteq \mathcal{B}_{N(N-1)} = \mathcal{I}dent \ .$$

4. Approach based on convex programming

4.1. **Estimator.** The proposed estimator is an extension of the *STIV* estimator of Gautier and Tsybakov (2011, 2014).

For a constant $r > 0$, a parameter (Θ, σ), where $\Theta = (B^T, G^T, \theta^T, \delta^T)^T$, is said to satisfy the *IV-constraint* if it belongs to the set

$$\widehat{\mathcal{I}}(r\sigma) = \bigcup_{(\eta_1,\eta_2)\in\{-1,1\}^2} \widehat{\mathcal{I}}_{\eta_1,\eta_2}(r\sigma) \tag{4.1}$$

$$\widehat{\mathcal{I}}_{\eta_1,\eta_2}(r\sigma) = \left\{ \begin{array}{l} B, G \in \mathbb{R}^{N(N-1)} : \ \eta_1 B \geq \mathbf{0}, \ \eta_2 G \geq \mathbf{0}, \ \theta \in \mathbb{R}^p, \ \delta \in \mathbb{R}^q, \\ B = (\boldsymbol{\beta_{-1,1}}^T, \ldots, \boldsymbol{\beta_{-N,N}}^T)^T, \ G = (\boldsymbol{\gamma_{-1,1}}^T, \ldots, \boldsymbol{\gamma_{-N,N}}^T)^T, \\ \mathbf{1}^T\boldsymbol{\beta_{-1,1}} = \ldots = \mathbf{1}^T\boldsymbol{\beta_{-N,N}}, \ \mathbf{1}^T\boldsymbol{\gamma_{-1,1}} = \ldots = \mathbf{1}^T\boldsymbol{\gamma_{-N,N}}, \\ \boldsymbol{R_i} = \boldsymbol{Y_i} - \boldsymbol{Y_{-i}}\boldsymbol{\beta_{-i,i}} - \boldsymbol{Z_{-i}}\boldsymbol{\gamma_{-i,i}} - \boldsymbol{X_i}\theta - \boldsymbol{V}\delta, \quad \forall i = 1\ldots, N, \\ \frac{1}{T-1} \max_{i=1\ldots,N} \max\left\{ \left| \overline{\boldsymbol{Z}}_{\boldsymbol{-i}}{}^T \boldsymbol{R_i} \right|_\infty, \left| \overline{\widetilde{\boldsymbol{Z}}}_{\boldsymbol{i}}{}^T \boldsymbol{R_i} \right|_\infty \right\} \leq \sigma r \\ \frac{1}{(T-1)\sqrt{N}} \max\left\{ \left| \sum_{i=1}^N \overline{\boldsymbol{X}}_{\boldsymbol{i}}{}^T \boldsymbol{R_i} \right|_\infty, \left| \overline{\boldsymbol{V}}^T \sum_{i=1}^N \boldsymbol{R_i} \right|_\infty \right\} \leq \sigma r \end{array} \right\} \tag{4.2}$$

where

$$\widehat{Q}(\Theta) = \frac{1}{N(T-1)} \sum_{i=1}^N \left| \widetilde{D} \left(\boldsymbol{Y_i} - \boldsymbol{Y_{-i}}\boldsymbol{\beta_{-i,i}} - \boldsymbol{Z_{-i}}\boldsymbol{\gamma_{-i,i}} - \boldsymbol{X_i}\theta - \boldsymbol{V}\delta \right) \right|_2^2 . \tag{4.3}$$

The first conditions in the definition of $\widehat{\mathcal{I}}_{\eta_1,\eta_2}(r\sigma)$ account for the additional structure (1.6), (1.7) that is imposed. If one does not impose this structure these constraints can be dropped and the set $\widehat{\mathcal{I}}(r\sigma)$ is no longer a union of sets.

The constant r is chosen to obtain a set of admissible parameters $\widehat{\mathcal{I}}(r\sigma)$ large enough so that it contains the true vector with probability $1-\alpha$ where α is a confidence level. The choice of r depends on the assumption made on the data generating process. A typical ("reference") behavior using a simple union bound is

$$r \sim \sqrt{\frac{\log(L/\alpha)}{T}}. \tag{4.4}$$

For simplicity it is assumed (1.3) and Assumption (4.1) so that one chooses r to be the $1-\alpha$ quantile of the following statistic

$$\begin{aligned} \mathcal{S} = \frac{1}{T-1} \max \Bigg\{ & \max_{\substack{i=1,\ldots,N \\ k=1,\ldots,N-1}} \left| \sum_{t=1}^T \overline{z}_{-i,t,k} e_{i,t} \right|, \max_{\substack{i=1,\ldots,N \\ k=1,\ldots,l_i}} \left| \sum_{t=1}^T \overline{\widetilde{z}}_{i,t,k} e_{i,t} \right|, \\ & \max_{k=1,\ldots,p} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T \overline{x}_{i,t,k} e_{i,t} \right|, \max_{k=1,\ldots,q} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T \overline{v}_{t,k} e_{i,t} \right| \Bigg\} \end{aligned} \tag{4.5}$$

where $e_{i,t}$ are i.i.d. standard normal random variables. It can be easily computed by Monte-Carlo techniques. This is the multiplier bootstrap of Chernozhukov, Chetverikov and Kato (2013).

Definition 4.1. *The Network Self-Tuned IV (*N-STIV*) estimator is any solution* $(\widehat{\Theta}, \widehat{\sigma})$ *of the following minimization problem:*

$$\min_{\substack{\sigma>0,\ \Theta=(B^T,G^T,\theta^T,\delta^T)^T \\ \Theta\in\widehat{\mathcal{I}}(r\sigma),\ \widehat{Q}(\Theta)\le\sigma^2}} \left(\left|\left((\boldsymbol{D^B})^{-1}B, r(\boldsymbol{D^G})^{-1}G\right)\right|_{2,1} + c\sigma \right), \tag{4.6}$$

where $0 < c < N/\sqrt{2}$.

For an estimator $\widehat{\Theta}$ there corresponds implicitly estimators $\widehat{B}$, $\widehat{\boldsymbol{\beta_{-i,i}}}$, $\widehat{\beta_{i,j}}$, $\widehat{G}$, $\widehat{\theta}$, $\widehat{\delta}$, etc. Each set $\widehat{\mathcal{I}}_{\eta_1,\eta_2}(r\sigma) \cap \left\{\Theta :\ \widehat{Q}(\Theta) \le \sigma^2\right\}$ is convex and $(\widehat{\Theta}, \widehat{\sigma})$ is the minimum of the 4 minima obtained by minimizing the objective function 4.6 on the 4 sets $\widehat{\mathcal{I}}_{\eta_1,\eta_2}(r\sigma)$ for $(\eta_1, \eta_2) \in \{-1, 1\}^2$.

The idea behind the estimator is to select among all admissible vectors that lie in the set $\widehat{\mathcal{I}}(r\sigma)$ one which is the most parsimonious as defined by the convex objective function. The natural objective function would involve a mixed $|\cdot|_{2,0}$-norm. This delivers a *NP*-hard problem where one has to search among all possible submodels. The mixed $|\cdot|_{2,1}$-norm is a convex relaxation of the $|\cdot|_{2,0}$-norm. A mixed norm is used to ensure group sparsity. Lounici et al (2011) also considered the used of a mixed-norm without endogenity and without the sensitivity analysis that is used in this paper. The estimator (4.6) is an extension of the *STIV* estimator of Gautier and Tsybakov (2011, 2014) which can be viewed as adding the *IV*-constraint to the square-root Lasso of Belloni, Chernozhukov and Wang (2011) or a pivotal Dantzig selector of Candès and Tao (2007) based on the *IV*-constraint. Gautier and Tsybakov (2011, 2014) also study the case where there is no ℓ_1, mixed norm in this paper, in (4.6). In this case one minimizes the least squares criterion subject to the constraint that Θ belongs to $\widehat{\mathcal{I}}(r\sigma)$. This was a new method to do robust inference in the presence of weak instruments which does not involve a pretest or test inversion for all possible values of the parameters.

4.2. **Sensitivity characteristics.** Due to sparsity, one is interested in $|\Psi\Delta|_\infty$ for vectors Δ in the cone

$$C_J = \bigcup_{(\eta_1,\eta_2)\in\{-1,1\}^2} C_{J,\eta_1,\eta_2} \tag{4.7}$$

where

(4.8)

$$C_{J,\eta_1,\eta_2} = \left\{ \begin{array}{ll} \Delta \in C_{\eta_1,\eta_2} :\ \left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{2,1} \le & \left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1} \\ & +c\left(\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right)\right) \end{array} \right\}$$

and

$$(4.9)\qquad C_{\eta_1,\eta_2} = \left\{ \begin{array}{l} \Delta^B, \Delta^G \in \mathbb{R}^{N(N-1)},\ \Delta^\theta \in \mathbb{R}^p,\ \Delta^\delta \in \mathbb{R}^q : \\ \Delta^B = (\boldsymbol{D^B})^{-1}\left(B^{(1)} - B^{(2)}\right),\ \Delta^G = (\boldsymbol{D^G})^{-1}\left(G^{(1)} - G^{(2)}\right)\ , \\ \text{where for } j = 1, 2,\ \eta_1 B^{(j)} \geq \mathbf{0},\ \eta_2 G^{(j)} \geq \mathbf{0}\ , \\ B^{(j)} = (\boldsymbol{\beta^{(j)}_{-1,1}}^T, \ldots, \boldsymbol{\beta^{(j)}_{-N,N}}^T)^T,\ G^{(j)} = (\boldsymbol{\gamma^{(j)}_{-1,1}}^T, \ldots, \boldsymbol{\gamma^{(j)}_{-N,N}}^T)^T\ , \\ \mathbf{1}^T \boldsymbol{\beta^{(j)}_{-1,1}} = \ldots = \mathbf{1}^T \boldsymbol{\beta^{(j)}_{-N,N}},\ \mathbf{1}^T \boldsymbol{\gamma^{(j)}_{-1,1}} = \ldots = \mathbf{1}^T \boldsymbol{\gamma^{(j)}_{-N,N}} \end{array} \right\}$$

The set J that will be used later is the set $J(B)$ which is assumed to be equal to $J(G)$. The last inequality comes from the fact that the estimator minimizes the criterion function (4.6) and that B and G are sparse. Using the Hölder inequality, the inequality that is explicited in (4.8) implies that

$$(4.10)\qquad (1 - c\sqrt{2}/N)\left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{2,1} \leq (1 + c\sqrt{2}/N)\left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1} + cr\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right)/\sqrt{N}$$

so that (Δ^B_J, Δ^G_J) contains most of the mass of the vector (Δ^B, Δ^G). The conditions in the definition of C_{η_1,η_2} account for the additional structure (1.6), (1.7) that is imposed. If one does not impose this structure these constraints can be dropped and the set C_J is no longer a union of sets.

Let us now introduce various sensitivity characteristics which are key quantities to measure the accuracy of the estimator. They extend the restricted eigenvalues of Bickel, Ritov and Tsybakov (2009) and provide sharper bounds even without endogeneity (*c.f.*, Gautier and Tsybakov (2011, 2014)). The following sensitivities are introduced to bound the estimation of the variance of the errors

$$\kappa^\sigma_J = \inf_{\substack{\Delta \in C_J \\ r|\Delta^G|_1 + |\Delta^B|_1 + r\sqrt{N}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) = N}} |\Psi\Delta|_\infty$$

$$\kappa_{2,1,J} = \inf_{\substack{\Delta \in C_J \\ \left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1} = 1}} |\Psi\Delta|_\infty\ .$$

The *coordinate-wise sensitivities* for $k \in \{1, \ldots, 2N(N-1) + p + q\}$ are the quantities

$$\kappa^*_{k,J} = \inf_{\substack{\Delta \in C_J \\ \Delta_k = 1}} |\Psi\Delta|_\infty\ .$$

Finally, the following sensitivites are used for inference on the social effects

$$\kappa^{\text{end}}_J = \inf_{\substack{\Delta \in C_J \\ \forall i = 1,\ldots,N,\ \mathbf{1}^T\left(\boldsymbol{\beta}_{(1),-i,i} - \boldsymbol{\beta}_{(2),-i,i}\right) = 1}} |\Psi\Delta|_\infty$$

$$\kappa^{\text{ex}}_J = \inf_{\substack{\Delta \in C_J \\ \forall i = 1,\ldots,N,\ \mathbf{1}^T\left(\boldsymbol{\gamma}_{(1),-i,i} - \boldsymbol{\gamma}_{(2),-i,i}\right) = 1}} |\Psi\Delta|_\infty\ .$$

They are coordinate wise sensitivities for a linear combination of coefficients, with the extra restrictions coming from the fact that it is imposed that these effects are homogeneous. Note that, by similar arguments as in the proof of Proposition 4.1 (see as well Gautier and Tsybakov (2011, 2014)), one can show that

$$\kappa_J^{\sigma} \geq \frac{N - c\sqrt{2}}{4|J| + (p+q)\sqrt{N}} \kappa_{\infty,J}^{r} \tag{4.11}$$

$$\kappa_{2,1,J} \geq \frac{1}{\sqrt{2}|J|} \kappa_{\infty,J}^{r} \tag{4.12}$$

where

$$\kappa_{\infty,J}^{r} = \min \left\{ \inf_{\substack{\Delta \in C_J \\ \left|\Delta_{J_{\mathrm{end}} \cap J}\right|_\infty = 1 \\ r\left|\Delta_{J_{\mathrm{ex}}^{\complement} \cap J}\right|_\infty \leq 1}} |\Psi \Delta|_\infty , \inf_{\substack{\Delta \in C_J \\ \left|\Delta_{J_{\mathrm{ex}}^{\complement} \cap J}\right|_\infty = \frac{1}{r} \\ \left|\Delta_{J_{\mathrm{end}} \cap J}\right|_\infty \leq 1}} |\Psi \Delta|_\infty \right\} .$$

All sensitivities depend on c but this dependence is not made explicit for the simplicity of notations. For the sake of simplicity the paper focuses on estimation of the individual coordinates. This is essential for a fine study of the variable selection (here the recovery of the adjacency matrix) and is in the spirit of Lounici (2008) but with an analysis based on the sensitivities. Estimation of vectors with arbitrary ℓ_p norm could also be done like in Gautier and Tsybakov (2011, 2014).

4.3. **Choice of r.** For a fixed $\Theta \in \mathcal{I}dent$ (or $\Theta \in \mathcal{B}_s$) one denotes by $\boldsymbol{R_i} = \boldsymbol{Y_i} - \boldsymbol{Y_{-i}}\boldsymbol{\beta_{-i,i}} - \boldsymbol{Z_{-i}}\boldsymbol{\gamma_{-i,i}} - \boldsymbol{X_i}\theta - \boldsymbol{V}\delta$ and by $\boldsymbol{\epsilon_i} = \boldsymbol{R_i} - \boldsymbol{\alpha_i}$. One has $\widetilde{D}\boldsymbol{\epsilon_i} = \widetilde{D}\boldsymbol{R_i}$ so that

$$\widehat{Q}(\Theta) = \frac{1}{N(T-1)} \sum_{i=1}^{N} \left| \widetilde{D}\boldsymbol{\epsilon_i} \right|_2^2 .$$

Denoting by $\mu_2 = \mathbb{E}[\epsilon_{i,t}^2]$ and $\mu_4 = \mathbb{E}[\epsilon_{i,t}^4]$, one can easily check using Chebyshev's inequality that the following result holds.

Lemma 4.1. *For every* $\Theta \in \mathcal{I}dent$, $\mathbb{E}\left[\widehat{Q}(\Theta)\right] = \mu_2$ *and*

$$\mathbb{P}\left(\left| \widehat{Q}(\Theta) - \mu_2 \right| \geq \epsilon \right) \leq \frac{\mu_4 - \mu_2^2}{\epsilon^2 NT} .$$

Denote by $\mathbb{P}(\Theta)$ the distribution of $(z_{i,t}^T, \widetilde{z}_{i,t}^T, x_{i,t}^T)_{t=1}^T$ and by $\mathcal{P}$ the class of distributions $\mathbb{P}(\Theta)$ that satisfy the following assumption.

Assumption 4.1. $\mathbb{P}(\Theta)$ *is such that conditions* (1.2) *and* (1.3) *hold and there exists constants* $\overline{c}$, $\overline{C}$, B_T *such that:*

(*i*) $\forall i = 1, \ldots, N,\ k = 1, \ldots, N-1,\ |\overline{z}_{-i,t,k}| \le B_T,$

$\forall i = 1, \ldots, N,\ k = 1, \ldots, l_i,\ |\overline{\overline{z}}_{i,t,k}| \le B_T,$

$\forall i = 1, \ldots, N,\ k = 1, \ldots, p,\ |\overline{x}_{i,t,k}| \le B_T,$

$\forall k = 1, \ldots, q,\ |\overline{v}_{t,k}| \le B_T;$

(*ii*) $\mu_4 \le \overline{C};$

(*iii*) $B_T^4(\log(LT))^7/T \le \overline{C}T^{-\overline{c}}.$

For every $\Theta \in \mathcal{I}dent$ and $\mathbb{P}(\Theta) \in \mathcal{P}$, $\widehat{Q}(\Theta)$ is a consitent estimator of μ_2 when $NT \to \infty$.

Our main analysis is carried out, by working on the event

$$\mathcal{G}(\Theta) = \left\{ \begin{array}{l} \max_{\substack{i=1,\ldots,N \\ k=1,\ldots,N-1}} \left| \frac{1}{T-1} \sum_{t=1}^T \overline{z}_{-i,t,k}\epsilon_{i,t} \right| \le r\sqrt{\widehat{Q}(\Theta)} \\ \max_{\substack{i=1,\ldots,N \\ k=1,\ldots,l_i}} \left| \frac{1}{T-1} \sum_{t=1}^T \overline{\overline{z}}_{i,t,k}\epsilon_{i,t} \right| \le r\sqrt{\widehat{Q}(\Theta)} \\ \max_{k=1,\ldots,p} \left| \frac{1}{(T-1)\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T \overline{x}_{i,t,k}\epsilon_{i,t} \right| \le r\sqrt{\widehat{Q}(\Theta)} \\ \max_{k=1,\ldots,q} \left| \frac{1}{(T-1)\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T \overline{v}_{t,k}\epsilon_{i,t} \right| \le r\sqrt{\widehat{Q}(\Theta)} \end{array} \right\} \tag{4.13}$$

Recall that the constant r is adjusted to be the $1-\alpha$ quantile of $\mathcal{S}$ in (4.5). This is because $\mathbb{P}(\mathcal{G}(\Theta)) = \mathbb{E}[\mathbb{P}(\mathcal{G}(\Theta)|w_s\ \forall 1 \le s \le T)]$ so that from Corollary 2.1 in Chernozhukov, Chetverikov and Kato (2013), $\mathbb{P}(\mathcal{G}(\Theta)) \ge 1-\alpha$ asymptotically.

For such a choice of r,

$$\underline{\lim}_{\substack{T\to\infty, \\ B_T^4(\log(LT))^7/T \le \overline{C}T^{-\overline{c}}}} \inf_{c,\Theta,\mathbb{P}:\ c\in(0,N/\sqrt{2}),\ \Theta\in\mathcal{I}dent,\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}(\mathcal{G}(\Theta)) \ge 1-\alpha\ .$$

Taking the infimum over the parameter c is innocuous as neither $\mathcal{I}dent$, $\mathcal{G}(\Theta)$ nor $\mathcal{P}$ depend on c.

4.4. **Rates of convergence.** Let us assume a prior upper bound $|J(B)| \le s$. Reducing $\mathcal{P}$ if necessary one can assume the following.

Assumption 4.2. *For every $\widetilde{\alpha} \in (0,1)$, $\Theta \in \mathcal{B}_s$ and $\mathbb{P}(\Theta) \in \mathcal{P}$, there exists σ_*, $\tau_*(c,r) < \infty$, $\kappa_*^{\rm end} > 0$, $\kappa_*^{\rm ex} > 0$, $v_k > 0$ and $\kappa_{*k} > 0$ for $k = 1, \ldots, 2N(N-1)+p+q$ and an event $\widetilde{\mathcal{G}}(\Theta)$ such that*

$$\underline{\lim}_{\substack{T\to\infty, \\ B_T^4(\log(LT))^7/T \le \overline{C}T^{-\overline{c}}}} \inf_{\Theta,\mathbb{P}:\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}\left(\widetilde{\mathcal{G}}(\Theta)\right) \ge 1-\widetilde{\alpha}$$

and on $\widetilde{\mathcal{G}}(\Theta)$

$$\widehat{Q}(\Theta) \le \sigma_*^2\ , \tag{4.14}$$

$$\left(1 - \frac{r}{c\kappa_{2,1,J(B)}}\right)_+^{-1} \left(1 + \frac{r}{c\kappa_{2,1,J(B)}}\right) \le \tau_*(c,r)\ , \tag{4.15}$$

$$\kappa^{\text{end}}_{J(B)} \geq \kappa^{\text{end}}_{*} \ , \tag{4.16}$$

$$\kappa^{\text{ex}}_{J(B)} \geq \kappa^{\text{ex}}_{*} \ , \tag{4.17}$$

and for every $k = 1, \dots, 2N(N-1) + p + q$*,*

$$\kappa^{*}_{k,J(B)} \geq \kappa_{*k} \ , \tag{4.18}$$

$$(\boldsymbol{D}^{\boldsymbol{R}})^{-1}_{k,k} \geq v_k \ . \tag{4.19}$$

An upper bound of the type (4.14) is easily obtained from Lemma 4.1. When T exceeds the number of instruments which itself exceeds the number of right-hand side variables in (1.8) Ψ is likely to have full column rank so that $\kappa^{\text{end}}_{*} > 0$, $\kappa^{\text{ex}}_{*} > 0$ and $\kappa_{*k} > 0$ for $k = 1, \dots, 2N(N-1)+p+q$ hold. In the high-dimensional case, the restriction $\Delta \in C_{J(B)}$ (which includes the restrictions in C_{η_1,η_2} as well as the one explicited in (4.8)) when $|J(B)|$ is not too large makes (4.18) likely to hold. Condition (4.19) is very mild. It is a lower bound on the square-root of the empirical second moments of the transformed regressors. The constants in Assumption 4.2 can depend on Θ.

The following theorem gives rates of convergence for the *N-STIV* estimator.

Theorem 4.1. *Under assumptions 4.1 and 4.2, for every* Θ *in* $\mathcal{B}_s$*, on the event* $\mathcal{G}(\Theta) \cap \widetilde{\mathcal{G}}(\Theta)$ *and for the constants of Assumption 4.2, for any solution* $(\widehat{\Theta}, \widehat{\sigma})$ *of the minimzation problem* (4.6)

(i) *for all* $k = 1, \dots, 2N(N-1) + p + q$*, the following inequalities hold*

$$\left|\widehat{\Theta}_k - \Theta_k\right| \leq \frac{\sigma_*(1 + \tau_*(c,r))\ r}{v_k\ \kappa_{*k}} \ , \tag{4.20}$$

$$\left|\widehat{\overline{\beta}}_k - \overline{\beta}_k\right| \leq \frac{\sigma_*(1 + \tau_*(c,r))\ r}{\kappa^{\text{end}}_{*}} \ , \tag{4.21}$$

$$\left|\widehat{\overline{\gamma}} - \overline{\gamma}_k\right| \leq \frac{\sigma_*(1 + \tau_*(c,r))\ r}{\kappa^{\text{ex}}_{*}} \ , \tag{4.22}$$

(ii) *if*

$$\min_{k \in J(B)} v_k \kappa_{*k} |\Theta_k| > \sigma_*(1 + \tau_*(c,r))\ r \ , \tag{4.23}$$

then $J(B) \subseteq J(\widehat{B})$ *and if*

$$\min_{k \in J(B)+N(N-1)} v_k \kappa_{*k} |\Theta_k| > \sigma_*(1 + \tau_*(c,r))\ r \ , \tag{4.24}$$

then $J(G) \subseteq J(\widehat{G})$*.*

For reasonably large sample size (typically $T \gg \log(L)$), the value r is small, and $\tau_*(c,r)$ is approaching 1 as $r \to 0$.

In the model considered in Bramoullé et al (2009) the entries of the matrix $(\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ are such that either $\beta_{i,j} = 0$ or $\beta_{i,j} = \frac{\overline{\beta}}{s_i}$ where s_i denotes the size of the group of peers directly influencing agent i. Assume that for every $i = 1,\dots,N$, $s_i \le \overline{s}$. Therefore, condition (4.23) is satisfied when

$$|\overline{\beta}| > \sigma_* \overline{s}(1+\tau_*(c,r))\ r \frac{1}{\min_{k\in J(B)} v_k \kappa_{*k}}\ .$$

This can typically be achieved when $\sqrt{T}$ is sufficiently large relative to $\sqrt{\log(L/\alpha)}\left(\frac{\overline{s}}{\min_{k\in J(B)} v_k\kappa_{*k}}\right)$. Recovering a direct effect of agent j on agent i is dificult when i has many peers and when the best instrument for $y_{j,t}$ is weak. A similar analysis hold for the matrix $(\gamma_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ of exogenous social effects but there κ_{*k} are usually larger because of exogeneity. This makes (4.24) much likely to hold than (4.23) when both $|\overline{\beta}|$ and $|\overline{\gamma}|$ are nonzero. When T is large enough and one has sufficiently strong instruments for the endogenous effects $J(\widehat{B})$ and $J(\widehat{G})$ can both yield upper estimates of the set of existing connections and can then be interstected.

4.5. **Computable lower bounds on the senstivities.** The set J that matters to evaluate the estimation accuracy is the set $J(B)$. To obtain confidence sets let us rely on lower bounds that does not depend on the perfect knowledge of $J(B)$. As in Gautier and Tsybakov (2011, 2014) two methods could be used: (1) either one has at his disposal a set $\widetilde{J} \supseteq J(B)$ or (2) the sparsity certificate approach where one specifies an upper bound on the number of nonzeros. Let us now provide lower bounds that can be easily computed by convex programs.

Proposition 4.1. *When $0 < c < N/\sqrt{2}$ and $\widetilde{J} \supseteq J$ the following lower bounds hold*

(*i*) $\kappa_J^\sigma \ge \kappa^\sigma\left(\widetilde{J}\right)$ *where*

$$\kappa^\sigma\left(\widetilde{J}\right) = \frac{N - c\sqrt{2}}{4\left|\widetilde{J}\right| + (p+q)\sqrt{N}} \min_{(\eta_1,\eta_2)\in\{-1,1\}^2} \min\left\{\frac{1}{r}\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm ex}}\left(\widetilde{J}\right), \kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm end}}\left(\widetilde{J}\right)\right\}$$

$$\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm ex}}\left(\widetilde{J}\right) = \min_{j\in J_{\rm ex}^{\cap}\widetilde{J}} \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le r\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\le r\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)\left|\widetilde{J}\right| + \frac{c}{\sqrt{N}}(p+q)\right)}} |\Psi\Delta|_\infty$$

$$\kappa^{\sigma,J_{\rm end}}_{\eta_1,\eta_2}\left(\widetilde{J}\right)=\min_{j\in J_{\rm end}\cap\widetilde{J}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le 1\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\le\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)|\widetilde{J}|+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

(ii) *For all* $k=1,\dots,2N(N-1)+p+q$, $\kappa^*_{k,J}\ge\kappa^*_k\left(\widetilde{J}\right)$ *where*

$$\kappa^*_k\left(\widetilde{J}\right)=\min_{(\eta_1,\eta_2,\eta_3)\in\{-1,1\}^3}\min\left\{\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right),\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right)\right\}$$

$$\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right)=\min_{j\in J_{\rm ex}\cap\widetilde{J}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge0\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r|\Delta^\theta|_\infty,r|\Delta^\delta|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le r\eta_3\Delta_j\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\le r\eta_3\Delta_j\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)|\widetilde{J}|+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

$$\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right)=\min_{j\in J_{\rm end}\cap\widetilde{J}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge0\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le\eta_3\Delta_j\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\le\eta_3\Delta_j\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)|\widetilde{J}|+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

(iii) $\kappa^{\rm end}_J\ge\kappa^{\rm end}\left(\widetilde{J}\right)$ *where* $\kappa^{\rm end}\left(\widetilde{J}\right)$ *is obtained like* $\kappa^*_k\left(\widetilde{J}\right)$ *replacing the constraint* $\Delta_k=1$ *by* $\forall i=1,\dots,N,\ \mathbf{1}^T\left(\boldsymbol{\beta}_{(1),-i,i}-\boldsymbol{\beta}_{(2),-i,i}\right)=1$.

(iv) $\kappa^{\rm ex}_J\ge\kappa^{\rm ex}\left(\widetilde{J}\right)$ *where* $\kappa^{\rm ex}\left(\widetilde{J}\right)$ *is obtained like* $\kappa^*_k\left(\widetilde{J}\right)$ *replacing the constraint* $\Delta_k=1$ *by* $\forall i=1,\dots,N,\ \mathbf{1}^T\left(\boldsymbol{\gamma}_{(1),-i,i}-\boldsymbol{\gamma}_{(2),-i,i}\right)=1$.

Proposition 4.2. *When* $0<c<N/\sqrt{2}$ *and* $|J|\le s$ *the following lower bounds hold*

(i) $\kappa^\sigma_J\ge\kappa^\sigma(s)$ *where*

$$\kappa^\sigma(s)=\frac{N-c\sqrt{2}}{4s+(p+q)\sqrt{N}}\min_{(\eta_1,\eta_2)\in\{-1,1\}^2}\min\left\{\frac{1}{r}\kappa^{\sigma,J_{\rm ex}}_{\eta_1,\eta_2}(s),\kappa^{\sigma,J_{\rm end}}_{\eta_1,\eta_2}(s)\right\}$$

$$\kappa^{\sigma,J_{\rm ex}}_{\eta_1,\eta_2}(s)=\min_{j\in J_{\rm ex}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r|\Delta^\delta|_\infty,\left|\Delta^B\right|_\infty\right\}\le r\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}\le r\left(2\sqrt{2}s+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

$$\kappa^{\sigma,J_{\rm end}}_{\eta_1,\eta_2}(s)=\min_{j\in J_{\rm end}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B\right|_\infty\right\}\le1\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}\le\left(2\sqrt{2}s+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

(ii) *For all* $k=1,\dots,2N(N-1)+p+q$, $\kappa^*_{k,J}\geq\kappa^*_k(s)$ *where*

$$\kappa^*_k(s)=\min_{(\eta_1,\eta_2,\eta_3)\in\{-1,1\}^3}\min\left\{\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}(s),\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}(s)\right\}$$

$$\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}(s)=\min_{j\in J_{\rm ex}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\geq 0\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r|\Delta^\delta|_\infty,\left|\Delta^B\right|_\infty\right\}\leq r\eta_3\Delta_j\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}\leq r\eta_3\Delta_j\left(2\sqrt{2}s+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

$$\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}(s)=\min_{j\in J_{\rm end}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\geq 0\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B\right|_\infty\right\}\leq \eta_3\Delta_j\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}\leq \eta_3\Delta_j\left(2\sqrt{2}s+\frac{c}{\sqrt{N}}(p+q)\right)}}|\Psi\Delta|_\infty$$

(iii) $\kappa^{\rm end}_J\geq\kappa^{\rm end}(s)$ *where* $\kappa^{\rm end}(s)$ *is obtained like* $\kappa^*_k(s)$ *replacing the constraint* $\Delta_k=1$ *by* $\forall i=1,\dots,N$, $\mathbf{1}^T\left(\boldsymbol{\beta}_{(1),-i,i}-\boldsymbol{\beta}_{(2),-i,i}\right)=1$.

(iv) $\kappa^{\rm ex}_J\geq\kappa^{\rm ex}(s)$ *where* $\kappa^{\rm ex}(s)$ *is obtained like* $\kappa^*_k(s)$ *replacing the constraint* $\Delta_k=1$ *by* $\forall i=1,\dots,N$, $\mathbf{1}^T\left(\boldsymbol{\gamma}_{(1),-i,i}-\boldsymbol{\gamma}_{(2),-i,i}\right)=1$.

4.6. **Exact recovery of the adjacency matrix.** Theorem 4.1 (iii) provides an upper estimate on the set of nonzero components of the matrices $(\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ and $(\gamma_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$. Exact recovery of the set of nonzeros, and therefore of the adjacency matrix, can be performed as well. For this purpose, a thresholded *N-STIV* estimator $(\widetilde{B},\widetilde{G})$ is used. Its coordinates are defined by

$$\widetilde{B}_k=\begin{cases}\widehat{B}_k & \text{if } |\widehat{B}_k|>\omega_k(s),\\ 0 & \text{otherwise,}\end{cases}\tag{4.25}$$

and

$$\widetilde{G}_k=\begin{cases}\widehat{G}_k & \text{if } |\widehat{G}_k|>\omega_{k+N(N+1)}(s),\\ 0 & \text{otherwise,}\end{cases}\tag{4.26}$$

where

$$\omega_k(s)=\frac{2\widehat{\sigma}\left(\boldsymbol{D^R}\right)_{k,k}\ r}{\kappa^*_k(s)}\left(1-\frac{r}{\kappa^\sigma(s)}\right)^{-1}_+ .$$

To provide garantees for the thresholding rule, westrengthen Assumption 4.2 as follows. Reducing $\mathcal{P}$ if necessary one can assume the following.

Reducing $\mathcal{P}$ if necessary one can assume the following.

Assumption 4.3. *For s in $\{1,\dots,N(N-1)\}$, for every $\widetilde{\alpha}\in(0,1)$ there exist $\sigma_*(s)$, $\tau_*(c,r,s)<\infty$, $\kappa_*^{\rm end}(s)>0$, $\kappa_*^{\rm ex}(s)>0$, $v_k>0$ and $\kappa_{*k}(s)>0$ for $k=1,\dots,2N(N-1)$ such that, for every $\Theta\in\mathcal{B}_s$ there exists an event $\widetilde{\mathcal{G}}(\Theta)$ such that*

$$\underline{\lim}_{\substack{T\to\infty,\\ B_T^4(\log(LT))^7/T\le\overline{C}T^{-\bar{c}}}}\ \inf_{\Theta,\mathbb{P}:\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}}\mathbb{P}\left(\widetilde{\mathcal{G}}(\Theta)\right)\ge 1-\widetilde{\alpha}$$

and on $\widetilde{\mathcal{G}}(\Theta)$

$$\widehat{Q}(\Theta)\le\sigma_*^2(s)\ , \tag{4.27}$$

$$\left(1-\frac{r}{\kappa^\sigma_{J(B)}}\right)_+^{-1}\left(1-\frac{r}{c\kappa_{2,1,J(B)}}\right)_+^{-1}\left(1+\frac{r}{c\kappa_{2,1,J(B)}}\right)\le\tau_*(c,r,s)\ , \tag{4.28}$$

$$\kappa^{\rm end}_{J(B)}\ge\kappa_{*\rm end}(s)\ , \tag{4.29}$$

$$\kappa^{\rm ex}_{J(B)}\ge\kappa_{*\rm ex}(s)\ , \tag{4.30}$$

and for every $k=1,\dots,2N(N-1)$,

$$\kappa^*_{k,J(B)}\ge\kappa_{*k}(s)\ , \tag{4.31}$$

$$(\boldsymbol{D}^{\boldsymbol{R}})^{-1}_{k,k}\ge v_k\ . \tag{4.32}$$

Based on Assumption 4.3, let us consider the following subset $\mathcal{B}_s$ where one removes from the s sparse identifiable vectors those which could be more sparse and have some coordinates which are too small to detect.

$$\mathcal{B}_s(r)=\{\Theta\in\mathcal{B}_s:\ \forall k\in J(B)\cup(J(B)+N(N+1)),\ v_k\kappa_{*k}(s)|\Theta_k|>4\sigma_*(s)\tau_*(c,r,s)\ r\}\ . \tag{4.33}$$

The following theorem shows that, based on thresholding of the *N-STIV* estimator, it is possible to recover the set of non-zero coefficients $J(B)$ and $J(G)$ with probability close to 1and to achieve sign consistency (*i.e.,* to recover the vector of signs of the coefficients of B (resp. G) with probability close to 1).

Reducing $\mathcal{P}$ if necessary one can assume the following.

Theorem 4.2. *Under assumptions 4.1 and 4.3 for s in $\{1,\dots,N-1\}$ and $\widetilde{\alpha}$ in $(0,1)$, for every Θ in $\mathcal{B}_s(r)$, on the event $\mathcal{G}(\Theta)\cap\widetilde{\mathcal{G}}(\Theta)$ one has*

$$\overrightarrow{\mathrm{sign}((\widetilde{B}^T,\widetilde{G}^T)^T)}=\overrightarrow{\mathrm{sign}((B^T,G^T)^T)} \tag{4.34}$$

and thus $J(\widetilde{B})=J(\widetilde{G})=J(B)$, Conditions (4.23), (4.24) and the condition in the definition of $\mathcal{B}_s(r)$ are referred to as *beta-min* assumptions.

4.7. **Confidence sets.**

4.7.1. *Confidence sets based on an estimated support.*

Theorem 4.3. *Let $0 < c < N/\sqrt{2}$, and let the assumptions of Theorem 4.2 hold for s in $\{1,\dots,N-1\}$ and $\widetilde{\alpha}$ in $(0,1)$. Set $\widehat{J} = J(\widetilde{B})$ or $\widehat{J} = J(\widetilde{G})$ where $\widetilde{B}$ and $\widetilde{G}$ defined in (4.25). For every Θ in $\mathcal{B}_s(r)$ on the event $\mathcal{G}(\Theta) \cap \widetilde{\mathcal{G}}(\Theta)$, for any solution $(\widehat{\Theta}, \widehat{\sigma})$ of the minimization problem (4.6) the following inequalities hold:*
for every $k = 1,\dots,2N(N-1)+p+q$,

$$\left|\widehat{\Theta}_k - \Theta_k\right| \le \frac{2\widehat{\sigma} r \left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k}}{\kappa_k^*\left(\widehat{J}\right)}\left(1 - \frac{r}{\kappa^\sigma\left(\widehat{J}\right)}\right)_+^{-1}, \tag{4.35}$$

$$\left|\widehat{\overline{\beta}} - \overline{\beta}\right| \le \frac{2\widehat{\sigma} r}{\kappa^{\mathrm{end}}\left(\widehat{J}\right)}\left(1 - \frac{r}{\kappa^\sigma\left(\widehat{J}\right)}\right)_+^{-1}, \tag{4.36}$$

$$\left|\widehat{\overline{\gamma}} - \overline{\gamma}\right| \le \frac{2\widehat{\sigma} r}{\kappa^{\mathrm{ex}}\left(\widehat{J}\right)}\left(1 - \frac{r}{\kappa^\sigma\left(\widehat{J}\right)}\right)_+^{-1}. \tag{4.37}$$

The sets defined in Theorem 4.3 can be computed from the data. They do not require to know the exact sparsity of the vector Θ however they require an upper bound s and to restrict the parameter space and consider $\mathcal{B}_s(r)$. Because their width depends on $\widehat{J}$ which is the true set $J(\Theta)$, which can be much smaller than s, they are called *adaptive confidence sets* (*c.f.*, Gautier and Tsybakov (2011, 2014) and Nickl and van de Geer (2013)). The confidence sets have coverage $1-\alpha-\widetilde{\alpha}$ because all inequalities in Theorem 4.3 hold on the event $\mathcal{G}(\Theta) \cap \widetilde{\mathcal{G}}(\Theta)$ where

$$\underline{\lim}_{\substack{T\to\infty,\\ B_T^4(\log(LT))^7/T \le \overline{C}T^{-\overline{c}}}} \ \inf_{\Theta,\mathbb{P}:\ \Theta\in\mathcal{B}_s(r),\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}\left(\mathcal{G}(\Theta)\cap\widetilde{\mathcal{G}}(\Theta)\right) \ge 1-\alpha-\widetilde{\alpha}\,.$$

Remark 4.1. *One does not need to introduce the event $\widetilde{\mathcal{G}}(\Theta)$ when the variance of the errors is known and the matrix Ψ is considered as fixed. This does not make sense in the presence of endogenous variables.*

Note that for all $\widetilde{\alpha}$, the sets $\mathcal{B}_s(r)$ increase when T increases and $\mathcal{B}_s = \bigcup_{n\in\mathbb{N}} \mathcal{B}_s(\frac{1}{n})$.

Under Assumption 4.3 one can use either $\widehat{J} = J(\widetilde{B})$ or $\widehat{J} = J(\widetilde{G})$. However the set $\mathcal{B}_s(r)$ might be too small for small T or when instruments are all too weak. It is possible to work with larger sets, for example

$$\mathcal{B}_s^B(r) = \{\Theta \in \mathcal{B}_s :\ \forall k \in J(B),\ v_k\kappa_{*k}(s)|\Theta_k| > 4\sigma_*(s)\tau_*(c,r,s)\ r\}$$

or

$$\mathcal{B}_s^G(r) = \{\Theta \in \mathcal{B}_s :\ \forall k \in J(B) + N(N+1),\ v_k\kappa_{*k}(s)|\Theta_k| > 4\sigma_*(s)\tau_*(c,r,s)\ r\}\ .$$

Because the coordinate-wise sensitivities are usually smaller for the exogeneous regressors, the set $\mathcal{B}_s^G(r)$ is usually larger than $\mathcal{B}_s^B(r)$ so that it is prefereable to use $\widehat{J} = J(\widetilde{G})$.

If instead of working with the assumptions of Theorem 4.3, one relies on the weaker Assumption 4.2, the same plug-in strategy as above can be used. Because of Theorem 4.1 (ii) this yields confidence sets which are more conservative.

One can also use the confidence sets of Theorem 4.3 constructing the thresholded estimator based on $s = |J(\widetilde{G})|$.

4.7.2. *Confidence Sets Under a Sparsity Certificate.* We argued that for the problem at hand and when the matrices $(\beta_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ and $(\gamma_{i,j})_{\substack{i=1,\dots,N\\ j=1,\dots,N}}$ are of the form used in Bramoullé (2009), a *beta-min* assumption is valid when T is not too small and there exists at least one strong instrument per endogenous regressor. Lagged endogenous regressors or their first differences are strongly correlated with the endogenous regressors.

When T is relatively small, one might prefer a different approach. One can construct confidence sets based on an upper bound s on the sparsity. This is a similar idea as undersmoothing in nonparametric inference. It is also possible to draw nested confidence sets for different values of s which include $s = \left|\widehat{J}\right|$ for $\widehat{J} = J(\widehat{G})$, $\widehat{J} = J(\widetilde{G})$, $\widehat{J} = J(\widehat{B})$ and $\widehat{J} = J(\widetilde{B})$.

Theorem 4.4. *Under Assumpion 4.1, for every Θ in $\mathcal{B}_s$, on the event $\mathcal{G}(\Theta)$, for any c in $(0, N/\sqrt{2})$, for any solution $(\widehat{\Theta}, \widehat{\sigma})$ of the minimization problem (4.6), the following inequalities hold: for every $k = 1, \dots, 2N(N-1) + p + q$,*

$$\left|\widehat{\Theta}_k - \Theta_k\right| \le \frac{2\widehat{\sigma} r \left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k}}{\kappa_k^*(s)} \left(1 - \frac{r}{\kappa^{\sigma}(s)}\right)_+^{-1}\ , \tag{4.38}$$

$$\left|\widehat{\overline{\beta}} - \overline{\beta}\right| \le \frac{2\widehat{\sigma} r}{\kappa^{\mathrm{end}}(s)} \left(1 - \frac{r}{\kappa^{\sigma}(s)}\right)_+^{-1}\ , \tag{4.39}$$

and

$$\left|\widehat{\overline{\gamma}}-\overline{\gamma}\right| \le \frac{2\widehat{\sigma}r}{\kappa^{\rm ex}(s)}\left(1-\frac{r}{\kappa^{\sigma}(s)}\right)_{+}^{-1} . \tag{4.40}$$

Each value of c delivers a random set $\mathcal{C}_c$ that only depends on the data and the sparsity certificate s. However, because the set of inequalities in Theorem 4.4 holds on the event $\mathcal{G}(\Theta)$ for every c, the set $\bigcap_{c\in(0,N/\sqrt{2})\cap D}\mathcal{C}_c$ where D is countable yields a measurable set such that $\mathbb{P}\left(\bigcap_{c\in(0,N/\sqrt{2})\cap D}\mathcal{C}_c\right)\ge \mathbb{P}\left(\mathcal{G}(\Theta)\right)$ and thus for any countable set D

$$\underline{\lim}_{\substack{T\to\infty,\\ B_T^4(\log(LT))^7/T\le \overline{C}T^{-\overline{c}}}} \ \inf_{(\Theta,\mathbb{P}):\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}\left(\bigcap_{c\in(0,N/\sqrt{2})\cap D}\mathcal{C}_c\right)\ge 1-\alpha .$$

The more sets $\mathcal{C}_c$ are interstected the smaller the confidence set. In practice one can make an intersection over a finite grid on $(0,N/\sqrt{2})$.

4.7.3. *Discussion.* The confidence sets can be infinite. Indeed the upper bounds in Theorem 4.3 involve the term $\left(1-\frac{r}{\kappa^{\sigma}(\widehat{J})}\right)_{+}^{-1}$, while the upper bounds in Theorem 4.4 involve the term $\left(1-\frac{r}{\kappa^{\sigma}(s)}\right)_{+}^{-1}$. Possibly infinite confidence sets is the price to pay for robustness to identification as explained in Dufour (1997). The confidence sets can be infinite when r is too small. This can be due to T which is too small or $\log(L)$ which is too large. Note that in the specification discused so far, without relying on auxiliary instruments, $\log(L)$ is of the order of $\log(N)$. The confidence sets can also be infinite when the lower bound on the sensitivity in the denominator is too small. This occurs when there is not a single sufficiently strong instrument for one of the endogenous variable. This is unlikely if one uses lagged endogenous regressors or their first differences as instrumental variables. Recall as well that the cone condition for an estimated $\widehat{J}$ gives more weight to the endogenous variables corresponding to the indices in $\widehat{J}$. When a sparsity certificate is used, the lower bound on the sensitivity can be small when s is too large. This confidence sets are two types of uniform confidence sets for identifiable parameters (see Romano and Shaikh (2008)). But here, like in Gautier and Tsybakov (2011, 2014), we restrict our attention to sparse vectors in the identified set.

4.8. ***N-STIV* estimator and confidence sets for different specifications.** This section explains how to modify the procedure and results when one considers specifications which are different from the reference one that we have focused on. Only a few different models are presented for the sake of conciseness.

4.8.1. *High-dimensional set of controls and time effects.* When one is uncertain about which own exogenous variables or time effects to consider one replaces equation (4.6) by

(4.41)
$$\min_{\substack{\sigma>0,\ \Theta=(B^T,G^T,\theta^T,\delta^T)^T\\ \Theta\in\widehat{\mathcal{I}}(r\sigma),\ \widehat{Q}(\Theta)\le\sigma^2}}\left(\left|\left((\boldsymbol{D^B})^{-1}B, r(\boldsymbol{D^G})^{-1}G\right)\right|_{2,1}+r\sqrt{\frac{N}{2}}\left|(\boldsymbol{D^\theta})^{-1}\theta\right|_1+r\sqrt{\frac{N}{2}}\left|(\boldsymbol{D^\delta})^{-1}\delta\right|_1+c\sigma\right)$$

and (4.8) by

(4.42)
$$C_{J,J_1,J_2,\eta_1,\eta_2}=\left\{\begin{array}{ll}\Delta\in C_{\eta_1,\eta_2}: & \left|\left(\Delta^B_{J^c},r\Delta^G_{J^c}\right)\right|_{2,1}+\left(\sqrt{\frac{N}{2}}-\frac{c}{\sqrt{N}}\right)r\left(\left|\Delta^\theta_{J_1^c}\right|_1+\left|\Delta^\delta_{J_2^c}\right|_1\right)\\ & \le\left|\left(\Delta^B_{J},r\Delta^G_{J}\right)\right|_{2,1}+\left(\sqrt{\frac{N}{2}}+\frac{c}{\sqrt{N}}\right)r\left(\left|\Delta^\theta_{J_1}\right|_1+\left|\Delta^\delta_{J_2}\right|_1\right)+\frac{c}{N}\left(r|\Delta^G|_1+|\Delta^B|_1\right)\end{array}\right\}$$

and work with the sensitivities where one replaces C_J by

$$C_{J,J_1,J_2}=\bigcup_{(\eta_1,\eta_2)\in\{-1,1\}^2}C_{J,J_1,J_2,\eta_1,\eta_2}\ .$$

An analogue of Theorem 4.1 holds with Assumption 4.2 replacing the sensitivities for the cone $C_{J(B)}$ by the sensitivities for the cone $C_{J(B),J(\theta),J(\delta)}$. The lower bounds on the sensitivities have to be modified accordingly and we work with:

$$\kappa^{\sigma,J_{\rm ex}}_{\eta_1,\eta_2}\left(\widetilde{J},\widetilde{J}_1,\widetilde{J}_2\right)=\min_{j\in J_{\rm ex}^{\cap\widetilde{J},\widetilde{J}_1,\widetilde{J}_2}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r\left|\Delta^\theta_{\widetilde{J}_1}\right|_\infty,r\left|\Delta^\delta_{\widetilde{J}_2}\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le r\\ \left(1-c\frac{\sqrt{2}}{N}\right)\left(\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}+r\sqrt{\frac{N}{2}}\left(\left|\Delta^\theta_{\widetilde{J}_1^c}\right|_1+\left|\Delta^\delta_{\widetilde{J}_2^c}\right|_1\right)\right)\le r\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)|\widetilde{J}|+\left(1+c\frac{\sqrt{2}}{N}\right)\sqrt{\frac{N}{2}}\left(|\widetilde{J}_1|+|\widetilde{J}_2|\right)\right)}}|$$

$$\kappa^{\sigma,J_{\rm end}}_{\eta_1,\eta_2}\left(\widetilde{J},\widetilde{J}_1,\widetilde{J}_2\right)=\min_{j\in J_{\rm end}\cap\widetilde{J}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r\left|\Delta^\theta_{\widetilde{J}_1}\right|_\infty,r\left|\Delta^\delta_{\widetilde{J}_2}\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le1\\ (1-c\sqrt{2})\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta_{\widetilde{J}_1^c}\right|_1+\left|\Delta^\delta_{\widetilde{J}_2^c}\right|_1\right)\le(\sqrt{2}+2c)|\widetilde{J}|+c(|\widetilde{J}_1|+|\widetilde{J}_2|)}}|\Psi\Delta|_\infty\ ,$$

$$\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J},\widetilde{J}_1,\widetilde{J}_2\right)=\min_{j\in J_{\rm ex}^{\cap\widetilde{J},\widetilde{J}_1,\widetilde{J}_2}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge0\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r\left|\Delta^\theta_{\widetilde{J}_1}\right|_\infty,r\left|\Delta^\delta_{\widetilde{J}_2}\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le r\eta_3\Delta_j\\ (1-c\sqrt{2})\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta_{\widetilde{J}_1^c}\right|_1+\left|\Delta^\delta_{\widetilde{J}_2^c}\right|_1\right)\le r\eta_3\Delta_j\left((\sqrt{2}+2c)|\widetilde{J}|+c(|\widetilde{J}_1|+|\widetilde{J}_2|)\right)}}|\Psi\Delta|_\infty\ ,$$

$$\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J},\widetilde{J}_1,\widetilde{J}_2\right)=\min_{j\in J_{\rm end}\cap\widetilde{J}}\ \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge 0\\ \max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty,r\left|\Delta^\theta_{\widetilde{J}_1}\right|_\infty,r\left|\Delta^\delta_{\widetilde{J}_2}\right|_\infty,\left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\le\eta_3\Delta_j\\ (1-c\sqrt{2})\left|\left(\Delta^B_{\widetilde{J}^c},r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta_{\widetilde{J}_1^c}\right|_1+\left|\Delta^\delta_{\widetilde{J}_2^c}\right|_1\right)\le\eta_3\Delta_j\left((\sqrt{2}+2c)\left|\widetilde{J}\right|+c\left(\left|\widetilde{J}_1\right|+\left|\widetilde{J}_2\right|\right)\right)}}|\Psi\Delta|_\infty\ ,$$

where $J_{\rm ex}^{\cap,J,J_1,J_2}=(N(N-1)+J)\cup(2N(N-1)+J_1)\cup(2N(N-1)+p+J_2)$, $\widetilde{J}\supseteq J$, $\widetilde{J}_1\supseteq J_1$ and $\widetilde{J}_2\supseteq J_2$, and with

$$\kappa^{\sigma,J_{\rm ex}}_{\eta_1,\eta_2}(s,s_1,s_2)=\min_{j\in J_{\rm ex}}\ \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r|\Delta^\delta|_\infty,\left|\Delta^B\right|_\infty\right\}\le r\\ (1-c\sqrt{2})\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta\right|_1+\left|\Delta^\delta\right|_1\right)\le 2r\left(\sqrt{2}s+s_1+s_2\right)}}|\Psi\Delta|_\infty\ ,$$

$$\kappa^{\sigma,J_{\rm end}}_{\eta_1,\eta_2}(s,s_1,s_2)=\min_{j\in J_{\rm end}}\ \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B\right|_\infty\right\}\le 1\\ (1-c\sqrt{2})\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta\right|_1+\left|\Delta^\delta\right|_1\right)\le\left(2\sqrt{2}s+s_1+s_2\right)}}|\Psi\Delta|_\infty\ ,$$

$$\kappa^{*,J_{\rm ex}}_{k,\eta_1,\eta_2,\eta_3}(s,s_1,s_2)=\min_{j\in J_{\rm ex}}\ \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge 0\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r|\Delta^\delta|_\infty,\left|\Delta^B\right|_\infty\right\}\le r\eta_3\Delta_j\\ (1-c\sqrt{2})\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta\right|_1+\left|\Delta^\delta\right|_1\right)\le r\eta_3\Delta_j(2\sqrt{2}s+s_1+s_2)}}|\Psi\Delta|_\infty\ ,$$

$$\kappa^{*,J_{\rm end}}_{k,\eta_1,\eta_2,\eta_3}(s,s_1,s_2)=\min_{j\in J_{\rm end}}\ \inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_k=1\\ \eta_3\Delta_j\ge 0\\ \max\left\{r\left|\Delta^G\right|_\infty,r|\Delta^\theta|_\infty,r\left|\Delta^\delta\right|_\infty,\left|\Delta^B\right|_\infty\right\}\le\eta_3\Delta_j\\ (1-c\sqrt{2})\left|\left(\Delta^B,r\Delta^G\right)\right|_{2,1}+(1-c)r\left(\left|\Delta^\theta\right|_1+\left|\Delta^\delta\right|_1\right)\le\eta_3\Delta_j(2\sqrt{2}s+s_1+s_2)}}|\Psi\Delta|_\infty$$

for $|J|\le s$, $|J_1|\le s_1$ and $|J_2|\le s_2$.

4.8.2. *More heterogeneous specifications.* Consider the model where $\theta^Tx_{i,t}$ is replaced by $\eta_iz_{i,t}+\widetilde{\theta}^T\widetilde{x}_{i,t}$ and the constraints (1.6) and (1.7) are dropped. This means that the total endogenous (resp. exogenous) effect of the members of the peer group $\sum_{j\neq i}\beta_{i,j}$ (resp. $\sum_{j\neq i}\gamma_{i,j}$) is heterogeneous and that two different members of the peer group can have effects with opposite sign.

The proposed method can be easily adapted to this setting (and/or heterogeneous $\widetilde{\theta}$, $\widetilde{\delta}$) by changing the definition of $\boldsymbol{R}_i$ in $\widehat{Q}(\Theta)$ and $\widehat{\mathcal{I}}(r\sigma)$. The matrix Ψ should also be modified

$$\Psi=\frac{1}{T}\boldsymbol{D^{HI}}\begin{pmatrix} \boldsymbol{Z_{-1}}^TM^TM\left(\boldsymbol{Y_{-1}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Z_{-1}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{X_1},\boldsymbol{0},\dots,\boldsymbol{V}\right)\\ \vdots\\ \boldsymbol{Z_{-N}}^TM^TM\left(\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Y_{-N}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Z_{-N}},,\boldsymbol{0},\dots,\boldsymbol{X_N},\boldsymbol{V}\right)\\ \widetilde{\boldsymbol{Z}}_{\boldsymbol{1}}^TM^TM\left(\boldsymbol{Y_{-1}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Z_{-1}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{X_1},\boldsymbol{0},\dots,\boldsymbol{V}\right)\\ \vdots\\ \widetilde{\boldsymbol{Z}}_{\boldsymbol{N}}^TM^TM\left(\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Y_{-N}},\boldsymbol{0},\dots,\boldsymbol{0},\boldsymbol{Z_{-N}},\boldsymbol{0},\dots,\boldsymbol{X_N},\boldsymbol{V}\right)\\ \boldsymbol{X_1}^TM^TM\left(\boldsymbol{Y_{-1}},\dots,\boldsymbol{Y_{-N}},\boldsymbol{Z_{-1}},\dots,\boldsymbol{Z_{-N}},\boldsymbol{X_i},\boldsymbol{V}\right)\\ \frac{1}{\sqrt{N}}\boldsymbol{V}^TM^TM\sum_{i=1}^N\left(\boldsymbol{Y_{-1}},\dots,\boldsymbol{Y_{-N}},\boldsymbol{Z_{-1}},\dots,\boldsymbol{Z_{-N}},\boldsymbol{X_1},\boldsymbol{0},\dots,\boldsymbol{V}\right)\\ \vdots\\ \boldsymbol{X_N}^TM^TM\left(\boldsymbol{Y_{-1}},\dots,\boldsymbol{Y_{-N}},\boldsymbol{Z_{-1}},\dots,\boldsymbol{Z_{-N}},\boldsymbol{X_i},\boldsymbol{V}\right)\\ \frac{1}{\sqrt{N}}\boldsymbol{V}^TM^TM\sum_{i=1}^N\left(\boldsymbol{Y_{-1}},\dots,\boldsymbol{Y_{-N}},\boldsymbol{Z_{-1}},\dots,\boldsymbol{Z_{-N}},\boldsymbol{0},\dots,\boldsymbol{X_N},\boldsymbol{V}\right) \end{pmatrix}\boldsymbol{D^{HR}}$$

with weighting matrices $\boldsymbol{D^{HI}}$ and $\boldsymbol{D^{HR}}$ slightly modified in an obvous way. The statistic $\mathcal{S}$ should also be modified as follows

$$\mathcal{S}=\frac{1}{T}\max\left\{\max_{\substack{i=1,\dots,N\\k=1,\dots,N-1}}\left|\sum_{t=1}^T\overline{z}_{-i,t,k}e_{i,t}\right|,\ \max_{\substack{i=1,\dots,N\\k=1,\dots,l_i}}\left|\sum_{t=1}^T\overline{\overline{z}}_{i,t,k}e_{i,t}\right|,\right.$$
$$\max_{\substack{i=1,\dots,N\\k=1,\dots,N-1}}\ \max_{k=1,\dots,p}\left|\sum_{t=1}^T\overline{x}^H_{i,t,k}e_{i,t}\right|,$$
$$\left.\max_{k=1,\dots,q}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T\overline{v}_{t,k}e_{i,t}\right|\right\}\tag{4.43}$$

with the new reweigted regressors $\overline{x}^H_{i,t,k}$.

One also need to drop the constraints coming from (1.6) and (1.7) are dropped in the set $\widehat{\mathcal{I}}(r\sigma)$ and in the sensitivites. Since the parameters $\overline{\beta}$ and $\overline{\gamma}$ are now heterogenous we do not impose the equality in the sensitivities. New particular parameters of interest are the average endogenous (resp. exogenous) effect $\overline{\overline{\beta}}=\frac{1}{N}\sum_{i=1}^N\sum_{j\neq i}\beta_{i,j}$ (resp. $\overline{\overline{\gamma}}=\frac{1}{N}\sum_{i=1}^N\sum_{j\neq i}\gamma_{i,j}$) and sensitivities and their lower bounds for inference on these parameters can be obtained in an obvious way.

4.8.3. *Autoregressive models.* Because our method is based on instrumental variables there is no specific difficulty to make inference with a specification with an autoregressive term

$$y_{i,t} = \rho y_{i,t-1} + \alpha_i + \sum_{j\neq i} \beta_{i,j} y_{j,t} + \sum_{j\neq i} \gamma_{i,j} z_{j,t} + \theta^T x_{i,t} + \delta^T v_t + \epsilon_{i,t} \ . \tag{4.44}$$

One can also consider various more or less heterogeneous specifications of equation (4.44) and modify accordingly the convex program.

4.8.4. *Exogenous effects with multiple exogenous variables.* When one wants to allow for exogenous effects with multiple (two in the example below) exogenous variables, one can consider, using obvious notations, a specification of the form

$$y_{i,t} = \alpha_i + \sum_{j\neq i} \beta_{i,j} y_{j,t} + \sum_{j\neq i} \gamma^1_{i,j} z^1_{j,t} + \sum_{j\neq i} \gamma^2_{i,j} z^2_{j,t} + \theta^T x_{i,t} + \delta^T v_t + \epsilon_{i,t} \ . \tag{4.45}$$

We do not impose the natural assumtion that there exists c such that $\forall j, i \ \gamma^1_{i,j} = c\gamma^2_{i,j}$ because this would lead a nonconvex program (see Lam and Souza (2014).

5. Large networks via linear programming

5.1. **Preliminaries.** The procedure presented in the preceeding sections involved convex programming. However it is possible to only rely on linear programming. This is in the spririt of the method in Gautier and Tsybakov (2013). We call it *LPN-STIV* estimator for Linear Programming Network Self Tuned IV estimator. The bounds are slightly larger but a linear programming approach but their computation is faster for large values of N.

Definition 5.1. *The first-stage* LPN-STIV $\left(\widehat{\Theta}^{1S}, \widehat{\sigma}^{1S}\right)$ *is any solution of*

$$\min_{\substack{\sigma>0,\ \Theta\in\widehat{\mathcal{I}}(r\sigma)\\ \Theta=(B^T,G^T,\theta^T,\delta^T)^T}} \left(\left| \left((\boldsymbol{D^B})^{-1}B, r(\boldsymbol{D^G})^{-1}G \right) \right|_{\infty,1} + c\sigma \right), \tag{5.1}$$

where $0 < c < 1/2$*.*

This can be rewritten as the following linear program

$$\min_{\substack{\sigma>0,\ \Theta\in\widehat{\mathcal{I}}(r\sigma),\ w\in\mathbb{R}^{N(N-1)}\\ \Theta=(B^T,G^T,\theta^T,\delta^T)^T\\ w\geq\mathbf{0}\\ \forall j=1,\dots,N(N-1),\ \left|\left((\boldsymbol{D^B})^{-1}B\right)_j\right|\leq w_j\\ \forall j=1,\dots,N(N-1),\ r\left|\left((\boldsymbol{D^G})^{-1}G\right)_j\right|\leq w_j}} \left(\sum_{j=1,\dots,N(N-1)} w_j + c\sigma \right) . \tag{5.2}$$

This first stage estimator is used to obtain a data-driven upper bound of $\widehat{Q}(\Theta)$. Two sensitivities are particularly important for this analysis. They are $\kappa_J^{\sigma,\gamma}$ and

$$\kappa_{\infty,1,J}^{\gamma}=\inf_{\substack{\Delta\in C_J^{\gamma}\\ \left|\left(\Delta_J^B,r\Delta_J^G\right)\right|_{\infty,1}=1}}|\Psi\Delta|_{\infty}$$

and are defined in the same way as in the previous sections working on the cone C_J^{γ} for γ positive, defined like C_J (*c.f.* (4.7)), replacing the cones C_{J,η_1,η_2} in (4.7) by

(5.3)

$$C_{J,\eta_1,\eta_2}^{\gamma}=\left\{\begin{array}{ll}\Delta\in C_{\eta_1,\eta_2}: & \left|\left(\Delta_{J^c}^B,r\Delta_{J^c}^G\right)\right|_{\infty,1}\\ & \le(2\gamma+1)\left|\left(\Delta_J^B,r\Delta_J^G\right)\right|_{\infty,1}+\gamma\left(r\left(|\Delta_J^G|_1+|\Delta^{\theta}|_1+|\Delta^{\delta}|_1\right)+|\Delta_J^B|_1\right)\end{array}\right\}.$$

Proposition 5.1. *When $0<c<1/2$ and $|J|\le s$, the following lower bounds hold*

(*i*) $\kappa_J^{\sigma,\gamma}\ge\kappa^{\sigma,\gamma}(s)$ *where*

$$\kappa^{\sigma,\gamma}(s)=\frac{1}{4(2\gamma+1)s+(\gamma+1)(p+q)}\min_{(\eta_1,\eta_2)\in\{-1,1\}^2}\min\left\{\frac{1}{r}\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm ex},\gamma}(s),\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm end},\gamma}(s)\right\}$$

$$\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm ex},\gamma}(s)=\min_{j\in J_{\rm ex}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_{\infty},r|\Delta^{\theta}|_{\infty},r|\Delta^{\delta}|_{\infty},\left|\Delta^B\right|_{\infty}\right\}\le r\\ \forall j,\ \left|\Delta_j^B\right|_{\infty}\le w_j,\ r\left|\Delta_j^G\right|_{\infty}\le w_j\\ \sum_j w_j\le r(2(2\gamma+1)s+\gamma(p+q))}}|\Psi\Delta|_{\infty}$$

$$\kappa_{\eta_1,\eta_2}^{\sigma,J_{\rm end},\gamma}(s)=\min_{j\in J_{\rm end}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_{\infty},r|\Delta^{\theta}|_{\infty},r|\Delta^{\delta}|_{\infty},\left|\Delta^B\right|_{\infty}\right\}\le 1\\ \forall j,\ \left|\Delta_j^B\right|_{\infty}\le w_j,\ r\left|\Delta_j^G\right|_{\infty}\le w_j\\ \sum_j w_j\le(2(2\gamma+1)s+\gamma(p+q))}}|\Psi\Delta|_{\infty}$$

(*ii*) $\kappa_{\infty,1,J}^{\gamma}\ge\kappa_{\infty,1}^{\gamma}(s)$ *where*

$$\kappa_{\infty,1}^{\gamma}(s)=\frac{1}{s}\min_{(\eta_1,\eta_2)\in\{-1,1\}^2}\min\left\{\frac{1}{r}\kappa_{\infty,1,\eta_1,\eta_2}^{J_{\rm ex},\gamma}(s),\kappa_{\infty,1,\eta_1,\eta_2}^{J_{\rm end},\gamma}(s)\right\}$$

$$\kappa_{\infty,1,\eta_1,\eta_2}^{J_{\rm ex},\gamma}(s)=\min_{j\in J_{\rm ex}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_{\infty},r|\Delta^{\theta}|_{\infty},r|\Delta^{\delta}|_{\infty},\left|\Delta^B\right|_{\infty}\right\}\le r\\ \forall j,\ \left|\Delta_j^B\right|_{\infty}\le w_j,\ r\left|\Delta_j^G\right|_{\infty}\le w_j\\ \sum_j w_j\le r(2(2\gamma+1)s+\gamma(p+q))}}|\Psi\Delta|_{\infty}$$

$$\kappa_{\infty,1,\eta_1,\eta_2}^{J_{\rm end},\gamma}(s)=\min_{j\in J_{\rm end}}\inf_{\substack{\Delta\in C_{\eta_1,\eta_2}\\ \Delta_j=1\\ \max\left\{r\left|\Delta^G\right|_{\infty},r|\Delta^{\theta}|_{\infty},r|\Delta^{\delta}|_{\infty},\left|\Delta^B\right|_{\infty}\right\}\le 1\\ \forall j,\ \left|\Delta_j^B\right|_{\infty}\le w_j,\ r\left|\Delta_j^G\right|_{\infty}\le w_j\\ \sum_j w_j\le(2(2\gamma+1)s+\gamma(p+q))}}|\Psi\Delta|_{\infty}.$$

Moreover, setting

$$c = \frac{\lambda\ r}{\kappa^{\gamma}_{\infty,1}(s)} \tag{5.4}$$

with

$$\frac{\lambda(\lambda-1)}{\lambda+1} = \frac{\gamma\kappa^{\gamma}_{\infty,1}(s)}{(\gamma+1)\kappa^{\sigma,\gamma}(s)}$$

in the definition of (5.1) *yields that, for any* Θ *in* $\mathcal{B}_s$*, on the event* $\mathcal{G}(\Theta)$,

$$\sqrt{\widehat{Q}(\Theta)} \le C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})} \tag{5.5}$$

where

$$C(\gamma, r, s) = \left(1 - \frac{2r(\lambda+1)}{(\lambda-1)\kappa^{\sigma,\gamma}(s)}\right)_+^{-1} . \tag{5.6}$$

Definition 5.2. *The second-stage* LPN-STIV $\widehat{\Theta}^{2S}$ *is any solution of*

$$\min_{\substack{\Theta\in\widehat{\mathcal{I}}\left(rC(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\right)\\ \Theta=(B^T,G^T,\theta^T,\delta^T)^T}} \left|\left((\boldsymbol{D}^{\boldsymbol{B}})^{-1}B, r(\boldsymbol{D}^{\boldsymbol{G}})^{-1}G\right)\right|_{\infty,1} . \tag{5.7}$$

The analysis of the second-stage estimator is based on the cone C_J^{2S} defined like C_J (*c.f.* (4.7)), replacing the cones C_{J,η_1,η_2} in (4.7) by

$$C^{2S}_{J,\eta_1,\eta_2} = \left\{\Delta \in C_{\eta_1,\eta_2}:\ \left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{\infty,1} \le \left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{\infty,1}\right\} . \tag{5.8}$$

They are denoted $\kappa^{*2S}_{k,J}$, $\kappa^{\mathrm{end},2S}_J$ and $\kappa^{\mathrm{ex},2S}_J$.

5.2. **Rates of convergence.** Let us introduce the following matrix

$$\overline{\Psi} = \frac{1}{T}\boldsymbol{D}^{\boldsymbol{R}} \begin{pmatrix} \boldsymbol{Y_{-1}}^T M^T M\left(\boldsymbol{Y_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Z_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{X_1}, \boldsymbol{V}\right) \\ \vdots \\ \boldsymbol{Y_{-N}}^T M^T M\left(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Y_{-N}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Z_{-N}}, \boldsymbol{X_N}, \boldsymbol{V}\right) \\ \boldsymbol{Z_{-1}}^T M^T M\left(\boldsymbol{Y_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Z_{-1}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{X_1}, \boldsymbol{V}\right) \\ \vdots \\ \boldsymbol{Z_{-N}}^T M^T M\left(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Y_{-N}}, \boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{Z_{-N}}, \boldsymbol{X_N}, \boldsymbol{V}\right) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{X_i}^T M^T M\left(\boldsymbol{Y_{-1}}, \dots, \boldsymbol{Y_{-N}}, \boldsymbol{Z_{-1}}, \dots, \boldsymbol{Z_{-N}}, \boldsymbol{X_i}, \boldsymbol{V}\right) \\ \frac{1}{\sqrt{N}}\boldsymbol{V}^T M^T M\sum_{i=1}^N\left(\boldsymbol{Y_{-1}}, \dots, \boldsymbol{Y_{-N}}, \boldsymbol{Z_{-1}}, \dots, \boldsymbol{Z_{-N}}, \boldsymbol{X_i}, \boldsymbol{V}\right) \end{pmatrix} \boldsymbol{D}^{\boldsymbol{R}}$$

and

$$\overline{\kappa}_J = \sup_{\substack{\Delta\in C^{\gamma}_J \\ \left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{\infty,1}=1}} \left|\Delta^T\overline{\Psi}\Delta\right| .$$

Let us make the following assumption.

Assumption 5.1. *For every $\widetilde{\alpha}\in(0,1)$ and $\Theta\in\mathcal{B}_s$, there exists σ_*, $C_*(\gamma,r,s)$, $\overline{\tau}_*(\gamma,r,s)$, $\kappa_*^{\mathrm{end},2S}>0$, $\kappa_*^{\mathrm{ex},2S}>0$, $v_k>0$ and $\kappa_{*k}^{2S}>0$ for $k=1,\dots,2N(N-1)+p+q$ and an event $\widetilde{\mathcal{G}}(\Theta)$ such that*

$$\underline{\lim}_{\substack{T\to\infty,\\ B_T^4(\log(LT))^7/T\le\overline{C}T^{-\overline{c}}}}\ \inf_{\Theta,\mathbb{P}:\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}}\mathbb{P}\left(\widetilde{\mathcal{G}}(\Theta)\right)\ge 1-\widetilde{\alpha}$$

and on $\widetilde{\mathcal{G}}(\Theta)$

$$\widehat{Q}(\Theta)\le\sigma_*^2\ , \tag{5.9}$$

$$C(\gamma,r,s)\le C_*(\gamma,r,s)\ , \tag{5.10}$$

$$\frac{2\lambda}{(\lambda-1)_+}\left(\frac{1}{\kappa^{\sigma}(s)}+\frac{2r\lambda\overline{\kappa}\sqrt{\widehat{Q}(\Theta)}}{(\lambda-1)_+(\kappa_{\infty,1}(s))^2}\right)\le\overline{\tau}_*(\gamma,r,s)\ , \tag{5.11}$$

$$\kappa_{J(B)}^{\mathrm{end},2S}\ge\kappa_*^{\mathrm{end},2S}\ , \tag{5.12}$$

$$\kappa_{J(B)}^{\mathrm{ex},2S}\ge\kappa_*^{\mathrm{ex},2S}\ , \tag{5.13}$$

and for every $k=1,\dots,2N(N-1)+p+q$,

$$\kappa_{k,J(B)}^{*2S}\ge\kappa_{*k}^{2S}\ , \tag{5.14}$$

$$(\boldsymbol{D}^{\boldsymbol{R}})_{k,k}^{-1}\ge v_k\ . \tag{5.15}$$

The following theorem gives rates of convergence for the *LPN-STIV* estimator.

Theorem 5.1. *Under assumptions 4.1 and 5.1, for every Θ in $\mathcal{B}_s$, on the event $\mathcal{G}(\Theta)\cap\widetilde{\mathcal{G}}(\Theta)$ and for the constants of Assumption 4.2, for any solution $\widehat{\Theta}^{2S}$ of the minimzation problem* (5.7)

(i) *for all $k=1,\dots,2N(N-1)+p+q$, the following inequalities hold*

$$\left|\widehat{\Theta}_k^{2S}-\Theta_k^{2S}\right|\le\frac{2C_*(\gamma,r,s)\sigma_*(1+r\overline{\tau}_*(\gamma,r,s))\ r}{v_k\ \kappa_{*k}}\ , \tag{5.16}$$

$$\left|\widehat{\overline{\beta}}^{2S}-\overline{\beta}^{2S}\right|\le\frac{2C_*(\gamma,r,s)\sigma_*(1+r\overline{\tau}_*(\gamma,r,s))\ r}{\kappa_*^{\mathrm{end}}}\ , \tag{5.17}$$

$$\left|\widehat{\overline{\gamma}}^{2S}-\overline{\gamma}^{2S}\right|\le\frac{2C_*(\gamma,r,s)\sigma_*(1+r\overline{\tau}_*(\gamma,r,s))\ r}{\kappa_*^{\mathrm{ex}}}\ , \tag{5.18}$$

(ii) *if*

$$\min_{k\in J(B)}v_k\kappa_{*k}|\Theta_k|>2C_*(\gamma,r,s)\sigma_*(1+r\overline{\tau}_*(\gamma,r,s))\ r\ , \tag{5.19}$$

then $J(B) \subseteq J(\widehat{B})$ *and if*

$$\min_{k \in J(B)+N(N-1)} v_k \kappa_{*k} |\Theta_k| > 2C_*(\gamma, r, s)\sigma_*(1 + r\overline{\tau}_*(\gamma, r, s))\ r\ , \tag{5.20}$$

then $J(G) \subseteq J(\widehat{G})$.

5.3. Computable lower bounds on the senstivities.

Proposition 5.2. *When* $0 < c < 1/2$ *and* $\widetilde{J} \supseteq J$*, the following lower bounds hold*

(i) *For all* $k = 1, \ldots, 2N(N-1)+p+q$, $\kappa^{*2S}_{k,J} \geq \kappa^{*2S}_{k}\left(\widetilde{J}\right)$ *where*

$$\kappa^{*2S}_{k}\left(\widetilde{J}\right) = \min_{(\eta_1,\eta_2,\eta_3)\in\{-1,1\}^3} \min\left\{\kappa^{*,J_{\mathrm{ex}},2S}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right), \kappa^{*,J_{\mathrm{end}},2S}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right)\right\}$$

$$\kappa^{*,J_{\mathrm{ex}},2S}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right) = \min_{j \in J_{\mathrm{ex}} \cap \widetilde{J}} \inf_{\substack{\Delta \in C_{\eta_1,\eta_2} \\ \Delta_k = 1 \\ \eta_3 \Delta_j \geq 0 \\ \forall l \in \widetilde{J},\ \left|\Delta^B_l\right|_\infty \leq w_l,\ r\left|\Delta^G_l\right|_\infty \leq w_l,\ w_l \leq \eta_3 \Delta_j \\ \sum_{l \in \widetilde{J}} w_l \leq sr\eta_3\Delta_j}} |\Psi\Delta|_\infty$$

$$\kappa^{*,J_{\mathrm{end}},2S}_{k,\eta_1,\eta_2,\eta_3}\left(\widetilde{J}\right) = \min_{j \in J_{\mathrm{end}} \cap \widetilde{J}} \inf_{\substack{\Delta \in C_{\eta_1,\eta_2} \\ \Delta_k = 1 \\ \eta_3 \Delta_j \geq 0 \\ \forall l \in \widetilde{J},\ \left|\Delta^B_l\right|_\infty \leq w_l,\ r\left|\Delta^G_l\right|_\infty \leq w_l,\ w_l \leq \eta_3 \Delta_j \\ \sum_{l \in \widetilde{J}} w_l \leq s\eta_3\Delta_j}} |\Psi\Delta|_\infty$$

(ii) $\kappa^{\mathrm{end},2S}_{J} \geq \kappa^{\mathrm{end},2S}(s)$ *where* $\kappa^{\mathrm{end},2S}(s)$ *is obtained like* $\kappa^{*2S}_{k}(s)$ *replacing the constraint* $\Delta_k = 1$ *by* $\forall i = 1, \ldots, N,\ \mathbf{1}^T\left(\boldsymbol{\beta}_{(1),-i,i} - \boldsymbol{\beta}_{(2),-i,i}\right) = 1.$

(iii) $\kappa^{\mathrm{ex},2S}_{J} \geq \kappa^{\mathrm{ex},2S}(s)$ *where* $\kappa^{\mathrm{ex},2S}(s)$ *is obtained like* $\kappa^{*2S}_{k}(s)$ *replacing the constraint* $\Delta_k = 1$ *by* $\forall i = 1, \ldots, N,\ \mathbf{1}^T\left(\boldsymbol{\gamma}_{(1),-i,i} - \boldsymbol{\gamma}_{(2),-i,i}\right) = 1.$

Proposition 5.3. *When* $0 < c < 1/2$ *and* $|J| \leq s$*, the following lower bounds hold*

(i) *For all* $k = 1, \ldots, 2N(N-1)+p+q$, $\kappa^{*2S}_{k,J} \geq \kappa^{*2S}_{k}(s)$ *where*

$$\kappa^{*2S}_{k}(s) = \min_{(\eta_1,\eta_2,\eta_3)\in\{-1,1\}^3} \min\left\{\kappa^{*,J_{\mathrm{ex}},2S}_{k,\eta_1,\eta_2,\eta_3}(s), \kappa^{*,J_{\mathrm{end}},2S}_{k,\eta_1,\eta_2,\eta_3}(s)\right\}$$

$$\kappa^{*,J_{\mathrm{ex}},2S}_{k,\eta_1,\eta_2,\eta_3}(s) = \min_{j \in J_{\mathrm{ex}}} \inf_{\substack{\Delta \in C_{\eta_1,\eta_2} \\ \Delta_k = 1 \\ \eta_3 \Delta_j \geq 0 \\ \forall l,\ \left|\Delta^B_l\right|_\infty \leq w_l,\ r\left|\Delta^G_l\right|_\infty \leq w_l,\ w_l \leq \eta_3 \Delta_j \\ \sum_{l} w_l \leq 2sr\eta_3\Delta_j}} |\Psi\Delta|_\infty$$

$$\kappa^{*,J_{\mathrm{end}},2S}_{k,\eta_1,\eta_2,\eta_3}(s) = \min_{j \in J_{\mathrm{end}}} \inf_{\substack{\Delta \in C_{\eta_1,\eta_2} \\ \Delta_k = 1 \\ \eta_3 \Delta_j \geq 0 \\ \forall l,\ \left|\Delta^B_l\right|_\infty \leq w_l,\ r\left|\Delta^G_l\right|_\infty \leq w_l,\ w_l \leq \eta_3 \Delta_j \\ \sum_{l} w_l \leq 2s\eta_3\Delta_j}} |\Psi\Delta|_\infty$$

(ii) $\kappa_J^{\mathrm{end},2S} \geq \kappa^{\mathrm{end},2S}(s)$ *where* $\kappa^{\mathrm{end},2S}(s)$ *is obtained like* $\kappa_k^{*2S}(s)$ *replacing the constraint* $\Delta_k = 1$ *by* $\forall i = 1, \dots, N,\ \mathbf{1}^T \left(\boldsymbol{\beta}_{(1),-i,i} - \boldsymbol{\beta}_{(2),-i,i}\right) = 1.$

(iii) $\kappa_J^{\mathrm{ex},2S} \geq \kappa^{\mathrm{ex},2S}(s)$ *where* $\kappa^{\mathrm{ex},2S}(s)$ *is obtained like* $\kappa_k^{*2S}(s)$ *replacing the constraint* $\Delta_k = 1$ *by* $\forall i = 1, \dots, N,\ \mathbf{1}^T \left(\boldsymbol{\gamma}_{(1),-i,i} - \boldsymbol{\gamma}_{(2),-i,i}\right) = 1.$

5.4. **Exact recovery of the adjacency matrix.** Theorem 5.1 (iii) provides an upper estimate on the set of nonzero components of the matrices $(\beta_{i,j})_{\substack{i=1,\dots,N \\ j=1,\dots,N}}$ and $(\gamma_{i,j})_{\substack{i=1,\dots,N \\ j=1,\dots,N}}$. Exact recovery of the set of nonzeros, and therefore of the adjacency matrix, can be performed as well. For this purpose, a thresholded *LPN-STIV* estimator $(\widetilde{B}^{2S}, \widetilde{G}^{2S})$ is used. Its coordinates are defined by

$$\widetilde{B}_k^{2S} = \begin{cases} \widehat{B}_k^{2S} & \text{if } |\widehat{B}_k^{2S}| > \omega_k^{2S}(s) \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{5.21}$$

and

$$\widetilde{G}_k^{2S} = \begin{cases} \widehat{G}_k^{2S} & \text{if } |\widehat{G}_k^{2S}| > \omega_{k+N(N+1)}^{2S}(s) \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{5.22}$$

where

$$\omega_k^{2S}(s) = \frac{2C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k}\ r}{\kappa_k^{*2S}} \ .$$

Assumption 5.2. *For* s *in* $\{1, \dots, N(N-1)\}$*, for every* $\widetilde{\alpha} \in (0,1)$ *there exist* $\sigma_*(s)$, $\overline{\tau}_*(\gamma, r, s) < \infty$, $\kappa_*^{\mathrm{end},2S}(s) > 0$, $\kappa_*^{\mathrm{ex},2S}(s) > 0$, $v_k > 0$ *and* $\kappa_{*k}^{2S}(s) > 0$ *for* $k = 1, \dots, 2N(N-1) + p + q$ *such that, for every* $\Theta \in \mathcal{B}_s$ *there exists an event* $\widetilde{\mathcal{G}}(\Theta)$ *such that*

$$\underline{\lim}_{\substack{T\to\infty, \\ B_T^4(\log(LT))^7/T \leq \overline{C}T^{-\overline{c}}}} \ \inf_{\Theta,\mathbb{P}:\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}\left(\widetilde{\mathcal{G}}(\Theta)\right) \geq 1 - \widetilde{\alpha}$$

and

$$\widehat{Q}(\Theta) \leq \sigma_*^2(s) \ , \tag{5.23}$$

$$C(\gamma, r, s) \leq C_*(\gamma, r, s) \ , \tag{5.24}$$

$$\frac{2\lambda}{(\lambda-1)_+}\left(\frac{1}{\kappa^{\sigma}(s)} + \frac{2r\lambda\overline{\kappa}\sqrt{\widehat{Q}(\Theta)}}{(\lambda-1)_+(\kappa_{\infty,1}(s))^2}\right) \leq \overline{\tau}_*(\gamma, r, s) \ , \tag{5.25}$$

$$\kappa_{J(B)}^{\mathrm{end},2S} \geq \kappa_*^{\mathrm{end},2S}(s) \ , \tag{5.26}$$

$$\kappa_{J(B)}^{\mathrm{ex},2S} \geq \kappa_*^{\mathrm{ex},2S}(s) \ , \tag{5.27}$$

and for every $k = 1, \dots, 2N(N-1) + p + q$,

$$\kappa^{*2S}_{k,J(B)} \geq \kappa^{2S}_{*k}(s) \ , \tag{5.28}$$

$$(\boldsymbol{D}^{\boldsymbol{R}})^{-1}_{k,k} \geq v_k \ . \tag{5.29}$$

Based on Assumption 5.2, let us consider the following subset $\mathcal{B}_s$ where one removes from the s sparse identifiable vectors those which are more sparse but have some coordinates which are too small to detect.

(5.30)

$$\mathcal{B}_s(r) = \{\Theta \in \mathcal{B}_s : \ \forall k \in J(B) \cup (J(B) + N(N+1)), \ v_k \kappa_{*k}(s) |\Theta_k| > 4C_*(\gamma, r, s)\sigma_*(s)(1 + r\overline{\tau}_*(\gamma, r, s)) \ r\} \ .$$

The following theorem shows that, based on thresholding of the *LPN-STIV* estimator, it is possible to recover the set of non-zero coefficients $J(B)$ and $J(G)$ with probability close to 1and to achieve sign consistency (*i.e.,* to recover the vector of signs of the coefficients of B (resp. G) with probability close to 1).

Theorem 5.2. *Under assumptions 4.1 and 5.2 for* s *in* $\{1, \dots, N-1\}$ *and* $\widetilde{\alpha}$ *in* $(0,1)$*, for every* Θ *in* $\mathcal{B}_s(r)$*, on the event* $\mathcal{G}(\Theta) \cap \widetilde{\mathcal{G}}(\Theta)$ *one has*

$$\overrightarrow{\mathrm{sign}((\widetilde{B}^T, \widetilde{G}^T)^T)} = \overrightarrow{\mathrm{sign}((B^T, G^T)^T)} \tag{5.31}$$

and thus $J(\widetilde{B}) = J(\widetilde{G}) = J(B)$,

5.5. Confidence sets.

5.5.1. *Confidence sets based on an estimated support.*

Theorem 5.3. *Let* $0 < c < 1/2$*, and let the assumptions of Theorem 5.2 hold for* s *in* $\{1, \dots, N-1\}$ *and* $\widetilde{\alpha}$ *in* $(0,1)$*. Set* $\widehat{J} = J(\widetilde{B})$ *or* $\widehat{J} = J(\widetilde{G})$ *where* $\widetilde{B}$ *and* $\widetilde{G}$ *defined in* (4.25)*. For every* Θ *in* $\mathcal{B}_s(r)$ *on the event* $\mathcal{G}(\Theta) \cap \widetilde{\mathcal{G}}(\Theta)$*, for any solution* $(\widehat{\Theta}, \widehat{\sigma})$ *of the minimization problem (5.1) the following inequalities hold:*
For every $k = 1, \dots, 2N(N-1) + p + q$,

$$\left|\widehat{\Theta}^{2S}_k - \Theta_k\right| \leq \frac{2\left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k} C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})} \ r}{\kappa^{*2S}_k\left(\widehat{J}\right)} \ , \tag{5.32}$$

$$\left|\widehat{\overline{\beta}}^{2S} - \overline{\beta}\right| \le \frac{2C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa^{\text{end},2S}\left(\widehat{J}\right)} \ , \tag{5.33}$$

$$\left|\widehat{\overline{\gamma}}^{2S} - \overline{\gamma}\right| \le \frac{2C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa^{\text{ex},2S}\left(\widehat{J}\right)} \ . \tag{5.34}$$

If instead of working with the assumptions of Theorem 5.3, one relies on the weaker Assumption 5.1, the same plug-in strategy as above can be used. Because of Theorem 5.1 (ii) this yields confidence sets which are more conservative.

One can also use the confidence sets of Theorem 5.3 constructing the thresholded estimator based on $s = |J(\widetilde{G})|$.

5.5.2. *Confidence Sets Under a Sparsity Certificate.*

Theorem 5.4. *Under Assumpion 4.1, for every Θ in $\mathcal{B}_s$, on the event $\mathcal{G}(\Theta)$, for any solution $(\widehat{\Theta}, \widehat{\sigma})$ of the minimization problem (4.6), the following inequalities hold:*
for every $k = 1, \dots, 2N(N-1) + p + q$,

$$\left|\widehat{\Theta}_k^{2S} - \Theta_k\right| \le \frac{2\left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k} C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa_k^{*2S}\left(s\right)} \ , \tag{5.35}$$

$$\left|\widehat{\overline{\beta}}^{2S} - \overline{\beta}\right| \le \frac{2C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa^{\text{end},2S}\left(s\right)} \ , \tag{5.36}$$

and

$$\left|\widehat{\overline{\gamma}}^{2S} - \overline{\gamma}\right| \le \frac{2C(\gamma, r, s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa^{\text{ex},2S}\left(s\right)} \ . \tag{5.37}$$

Each value of γ delivers a random set $\mathcal{C}_\gamma$ that only depends on the data and the sparsity certificate s. However, because the set of inequalities in Theorem 5.4 holds on the event $\mathcal{G}(\Theta)$ for every γ, the set $\bigcap_{\gamma\in(0,\infty)\cap D}\mathcal{C}_\gamma$ where D is countable yields a measurable set such that $\mathbb{P}\left(\bigcap_{\gamma\in(0,\infty)\cap D}\mathcal{C}_\gamma\right) \ge \mathbb{P}\left(\mathcal{G}(\Theta)\right)$ and thus for any countable set D

$$\underline{\lim}_{\substack{T\to\infty,\\ B_T^4(\log(LT))^7/T\le\overline{C}T^{-\overline{c}}}} \ \inf_{(\Theta,\mathbb{P}):\ \Theta\in\mathcal{B}_s,\ \mathbb{P}(\Theta)\in\mathcal{P}} \mathbb{P}\left(\bigcap_{\gamma\in(0,\infty)\cap D}\mathcal{C}_\gamma\right) \ge 1-\alpha \ .$$

6. Appendix

Theorem 6.1. *Under Assumpion 4.1, for every Θ in $\mathcal{I}dent$, on the event $\mathcal{G}(\Theta)$, for any solution $(\widehat{\Theta}, \widehat{\sigma})$ of the minimization problem (4.6), the following inequalities hold,*

$$\widehat{\sigma} \leq \tau(c,r)\sqrt{\widehat{Q}(\Theta)} \tag{6.1}$$

where

$$\tau(c,r) = \left(1 - \frac{r}{c\kappa_{2,1,J(B)}}\right)_+^{-1} \left(1 + \frac{r}{c\kappa_{2,1,J(B)}}\right) \ , \tag{6.2}$$

for every $k = 1, \ldots, N(N-1) + p + q$,

$$\left|\widehat{\Theta}_k - \Theta_k\right| \leq \frac{\left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k} \sqrt{\widehat{Q}(\Theta)}(1 + \tau(c,r))r}{\kappa^*_{k,J(B)}} \tag{6.3}$$

$$\left|\widehat{\Theta}_k - \Theta_k\right| \leq \frac{\left(\boldsymbol{D}^{\boldsymbol{R}}\right)_{k,k} 2\widehat{\sigma} r}{\kappa^*_{k,J(B)}} \left(1 - \frac{r}{\kappa^{\sigma}_{J(B)}}\right)_+^{-1} \ , \tag{6.4}$$

$$\left|\widehat{\overline{\beta}} - \overline{\beta}\right| \leq \frac{\sqrt{\widehat{Q}(\Theta)}(1 + \tau(c,r))r}{\kappa^{\text{end}}_{J(B)}} \tag{6.5}$$

$$\left|\widehat{\overline{\beta}} - \overline{\beta}\right| \leq \frac{2\widehat{\sigma} r}{\kappa^{\text{end}}_{J(B)}} \left(1 - \frac{r}{\kappa^{\sigma}_{J(B)}}\right)_+^{-1} \ , \tag{6.6}$$

$$\left|\widehat{\overline{\gamma}} - \overline{\gamma}\right| \leq \frac{\sqrt{\widehat{Q}(\Theta)}(1 + \tau(c,r))r}{\kappa^{\text{ex}}_{J(B)}} \tag{6.7}$$

$$\left|\widehat{\overline{\gamma}} - \overline{\gamma}\right| \leq \frac{2\widehat{\sigma} r}{\kappa^{\text{ex}}_{J(B)}} \left(1 - \frac{r}{\kappa^{\sigma}_{J(B)}}\right)_+^{-1} \ . \tag{6.8}$$

Proof of Theorem 6.1. Consider here a fixed $\Theta \in \mathcal{I}dent$. One has that on the event $\mathcal{G}(\Theta)$, Θ belongs to $\widehat{\mathcal{I}}\left(r\sqrt{\widehat{Q}(\Theta)}\right)$.

Set $\Delta^B = \left(\boldsymbol{D}^{\boldsymbol{B}}\right)^{-1}\left(\widehat{B} - B\right)$, $\Delta^G = \left(\boldsymbol{D}^{\boldsymbol{G}}\right)^{-1}\left(\widehat{G} - G\right)$, $\Delta^\theta = \left(\boldsymbol{D}^{\boldsymbol{X}}\right)^{-1}\left(\widehat{\theta} - \theta\right)$, $\Delta^\delta = \left(\boldsymbol{D}^{\boldsymbol{V}}\right)^{-1}\left(\widehat{\delta} - \delta\right)$ and $\Delta = \left((\Delta^B)^T, (\Delta^G)^T, (\Delta^\theta)^T, (\Delta^\delta)^T\right)^T$. Note that

$$\Delta = \left(\boldsymbol{D}^R\right)^{-1}\left((\widehat{B} - B)^T, (\widehat{G} - G)^T, (\widehat{\theta} - \theta)^T, (\widehat{\delta} - \delta)^T\right)^T.$$

The triangle inequality yields that, on the event $\mathcal{G}(\Theta)$,

$$|\Psi\Delta|_\infty \leq r\left(\widehat{\sigma} + \sqrt{\widehat{Q}(\Theta)}\right) \ . \tag{6.9}$$

On the other hand, $(\widehat{\Theta},\widehat{\sigma})$ minimizes the criterion $\left|\left(\left(\boldsymbol{D^B}\right)^{-1}B, r\left(\boldsymbol{D^G}\right)^{-1}G\right)\right|_{\infty,1}+c\sigma$ on the set $\widehat{\mathcal{I}}$. Thus, on the event $\mathcal{G}(\Theta)$,

$$\left|\left(\left(\boldsymbol{D^B}\right)^{-1}\widehat{B}, r\left(\boldsymbol{D^G}\right)^{-1}\widehat{G}\right)\right|_{2,1}+c\widehat{\sigma}\le\left|\left(\left(\boldsymbol{D^B}\right)^{-1}B, r\left(\boldsymbol{D^G}\right)^{-1}G\right)\right|_{2,1}+c\sqrt{\widehat{Q}(\Theta)}. \tag{6.10}$$

This implies by the triangle inequality and the fact that $J(B)=J(G)$, that on the event $\mathcal{G}(\Theta)$,

$$\left|\left(\Delta^B_{J(B)^c}, r\Delta^G_{J(B)^c}\right)\right|_{2,1}\le\left|\left(\Delta^B_{J(B)}, r\Delta^G_{J(B)}\right)\right|_{2,1}+c\left(\sqrt{\widehat{Q}(\Theta)}-\sqrt{\widehat{Q}(\widehat{\Theta})}.\right).$$

The last inequality holds because by construction $\sqrt{\widehat{Q}(\widehat{\Theta})}\le\widehat{\sigma}$.

Because the function $\gamma\to\sqrt{\widehat{Q}(\gamma)}$ is convex, one has

$$\begin{aligned}\sqrt{\widehat{Q}(\Theta)}-\sqrt{\widehat{Q}(\widehat{\Theta})}&\le\langle w,\Theta-\widehat{\Theta}\rangle\\&=\langle\boldsymbol{D^R}w,(\boldsymbol{D^R})^{-1}(\Theta-\widehat{\Theta})\rangle=-\langle\boldsymbol{D^R}w,\Delta\rangle\,.\end{aligned}$$

where $\langle,\rangle$ is the ℓ_2 scalar product, for any w in the subgradient of $\sqrt{\widehat{Q}}$ at Θ. With the same argument as in the proof of Theorem 6.1 in Gautier and Tsybakov (2011, 2014) one obtains

$$\sqrt{\widehat{Q}(\Theta)}-\sqrt{\widehat{Q}(\widehat{\Theta})}\le\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)\ . \tag{6.11}$$

This implies that on the event $\mathcal{G}(\Theta)$,

$$\left|\left(\Delta^B_{J(B)^c}, r\Delta^G_{J(B)^c}\right)\right|_{2,1}\le\left|\left(\Delta^B_{J(B)}, r\Delta^G_{J(B)}\right)\right|_{2,1}+c\left(\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)\right)\ .$$

In other words, the vector Δ belongs to the cone $C_{J(B)}$.

Also, by (6.10) and the definition of $\kappa_{2,1,J}$, one obtains

$$\begin{aligned}c\widehat{\sigma}&\le\left|\left(\Delta^B_{J(B)}, r\Delta^G_{J(B)}\right)\right|_{2,1}+c\sqrt{\widehat{Q}(\Theta)}\\&\le\frac{|\Psi\Delta|_\infty}{\kappa_{2,1,J(B)}}+c\sqrt{\widehat{Q}(\Theta)}\ .\end{aligned}$$

This yields

$$\left(1-\frac{r}{c\kappa_{2,1,J(B)}}\right)\widehat{\sigma}\le\left(1+\frac{r}{c\kappa_{2,1,J(B)}}\right)\sqrt{\widehat{Q}(\Theta)}\ . \tag{6.12}$$

Inequalities (6.3), (6.5) and (6.7) follow from (6.9), (6.12) and the definition of the sensitivities.

Now, using the fact that $\sqrt{\widehat{Q}(\widehat{\Theta})}\le\widehat{\sigma}$, (6.9) yield that

$$|\Psi\Delta|_\infty\le r\left(2\widehat{\sigma}+\sqrt{\widehat{Q}(\Theta)}-\widehat{\sigma}\right) \tag{6.13}$$

$$\le r\left(2\widehat{\sigma}+\sqrt{\widehat{Q}(\Theta)}-\sqrt{\widehat{Q}(\widehat{\Theta})}\right) \tag{6.14}$$

Inequality (6.13) thus yields

$$|\Psi\Delta|_\infty \leq r\left(2\widehat{\sigma} + \frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right)\right) . \tag{6.15}$$

Using the definition of the sensitivities one gets that, on the event $\mathcal{G}_\Theta$,

$$|\Psi\Delta|_\infty \leq r\left(2\widehat{\sigma} + \frac{|\Psi\Delta|_\infty}{\kappa^\sigma_{J(B)}}\right),$$

which implies

$$|\Psi\Delta|_\infty \leq 2\widehat{\sigma} r\left(1 - \frac{r}{\kappa^\sigma_{J(B)}}\right)_+^{-1} . \tag{6.16}$$

The upper bounds (6.4), (6.6) and (6.8) follow from the definition of the sensitivities □

Proof of Theorem 4.1. The rates of convergence are obtained by replacing the random right-hand sides in Theorem 6.1 by the deterministic upper bounds from Assumption 4.1. □

Proof of Proposition 4.1. Let us prove item (i). The other items are obtained similarly.
Any Δ in the cone C_J is also in the cone $C_{\widetilde{J}}$. Now, by the Hölder inequality, any Δ in the cone $C_{\widetilde{J}}$ is such that

$$\left(1 - c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1} \leq \left|\left(\Delta^B_{\widetilde{J}}, r\Delta^G_{\widetilde{J}}\right)\right|_{2,1} + c\left(r\left(\frac{1}{N}\left|\Delta^G_{\widetilde{J}}\right|_1 + \frac{1}{\sqrt{N}}|\Delta^\theta|_1 + \frac{1}{\sqrt{N}}|\Delta^\delta|_1\right) + \frac{1}{N}\left|\Delta^B_{\widetilde{J}}\right|_1\right) .$$

Let us now consider 4 cases.

Case 1: there exists $j \in \widetilde{J}$ such that $\max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_{\widetilde{J}}\right|_\infty\right\} \leq r|\Delta^G_j|$ then one has

$$\left(1 - c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1} \leq r|\Delta^G_j|\left(\sqrt{2}\left(1 + c\frac{\sqrt{2}}{N}\right)\left|\widetilde{J}\right| + \frac{c}{\sqrt{N}}(p+q)\right) \tag{6.17}$$

and

$$\begin{aligned}\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) \leq & \frac{\sqrt{2}}{N}\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1} \\ & + \frac{r}{N}|\Delta^G_{\widetilde{J}}|_1 + \frac{1}{N}|\Delta^B_{\widetilde{J}}|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) ,\end{aligned} \tag{6.18}$$

thus injecting (6.17) into (6.18) yields

$$\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) \leq r|\Delta^G_j|\frac{4\left|\widetilde{J}\right| + (p+q)\sqrt{N}}{N - c\sqrt{2}} . \tag{6.19}$$

This yields

$$\frac{|\Psi\Delta|_\infty}{\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)}\geq\frac{|\Psi\Delta|_\infty}{|\Delta^G_j|}\frac{N-c\sqrt{2}}{r\left(4\left|\widetilde{J}\right|+(p+q)\sqrt{N}\right)}\ . \tag{6.20}$$

Case 2: there exists $j\in J$ such that $\max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\leq|\Delta^B_j|$ then one has

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\leq|\Delta^B_j|\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)\left|\widetilde{J}\right|+\frac{c}{\sqrt{N}}(p+q)\right) \tag{6.21}$$

thus injecting (6.21) into (6.18) yields

$$\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)\leq|\Delta^B_j|\frac{4\left|\widetilde{J}\right|+(p+q)\sqrt{N}}{N-c\sqrt{2}}\ . \tag{6.22}$$

This yields

$$\frac{|\Psi\Delta|_\infty}{r\left(|\Delta^G|_1+|\Delta^\theta|_1+|\Delta^\delta|_1\right)+|\Delta^B|_1}\geq\frac{|\Psi\Delta|_\infty}{|\Delta^B_j|}\frac{N-c\sqrt{2}}{4\left|\widetilde{J}\right|+(p+q)\sqrt{N}}\ . \tag{6.23}$$

Case 3: there exists $j\in\{1,\dots,p\}$ such that $\max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\leq r|\Delta^\theta_j|$ then one has

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{\widetilde{J}^c}, r\Delta^G_{\widetilde{J}^c}\right)\right|_{2,1}\leq r|\Delta^\theta_j|\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)\left|\widetilde{J}\right|+\frac{c}{\sqrt{N}}(p+q)\right) \tag{6.24}$$

thus injecting (6.24) into (6.18) yields

$$\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)\leq r|\Delta^\theta_j|\frac{4\left|\widetilde{J}\right|+(p+q)\sqrt{N}}{N-c\sqrt{2}}\ . \tag{6.25}$$

This yields

$$\frac{|\Psi\Delta|_\infty}{\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)}\geq\frac{|\Psi\Delta|_\infty}{|\Delta^\theta_j|}\frac{N-c\sqrt{2}}{r\left(4\left|\widetilde{J}\right|+(p+q)\sqrt{N}\right)}\ . \tag{6.26}$$

Case 4: there exists $j\in\{1,\dots,q\}$ such that $\max\left\{r\left|\Delta^G_{\widetilde{J}}\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_{\widetilde{J}}\right|_\infty\right\}\leq r|\Delta^\delta_j|$, exactly like case 3 one obtains

$$\frac{|\Psi\Delta|_\infty}{\frac{r}{N}|\Delta^G|_1+\frac{1}{N}|\Delta^B|_1+\frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1+|\Delta^\delta|_1\right)}\geq\frac{|\Psi\Delta|_\infty}{|\Delta^\delta_j|}\frac{N-c\sqrt{2}}{r\left(4\left|\widetilde{J}\right|+(p+q)\sqrt{N}\right)} \tag{6.27}$$

□

Proof of Proposition 4.2. Recall that it is assumed that $|J| \le s$.

Let us prove item (i). The other items are obtained similarly.

Adding $(1-c\sqrt{2}/N)\left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1}$ to both sides of

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{2,1} \le \left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1} + c\left(r\left(\frac{1}{N}\left|\Delta^G_J\right|_1 + \frac{1}{\sqrt{N}}|\Delta^\theta|_1 + \frac{1}{\sqrt{N}}|\Delta^\delta|_1\right) + \frac{1}{N}\left|\Delta^B_J\right|_1\right)$$

one obtains

$$\begin{aligned}\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B, r\Delta^G\right)\right|_{2,1} \le& \left(2-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{2,1} \\ &+ c\left(r\left(\frac{1}{N}\left|\Delta^G_J\right|_1 + \frac{1}{\sqrt{N}}|\Delta^\theta|_1 + \frac{1}{\sqrt{N}}|\Delta^\delta|_1\right) + \frac{1}{N}\left|\Delta^B_J\right|_1\right) .\end{aligned}$$

We have to consider 4 cases like in the proof of Proposition 4.1. Let us only present the case of Case 1, the other ones are treated similarly.

Case 1: for some $j \in \{1, \dots, N(N-1)\}$, $\max\left\{r\left|\Delta^G\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B\right|_\infty\right\} \le r|\Delta^G_j|$, then one has

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B, r\Delta^G\right)\right|_{2,1} \le r|\Delta^G_j|\left(2\sqrt{2}\,|J| + \frac{c}{\sqrt{N}}(p+q)\right) , \tag{6.28}$$

and thus

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B, r\Delta^G\right)\right|_{2,1} \le r|\Delta^G_j|\left(2\sqrt{2}s + \frac{c}{\sqrt{N}}(p+q)\right) . \tag{6.29}$$

Also because

$$\left(1-c\frac{\sqrt{2}}{N}\right)\left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{2,1} \le r|\Delta^G_j|\left(\sqrt{2}\left(1+c\frac{\sqrt{2}}{N}\right)|J| + \frac{c}{\sqrt{N}}(p+q)\right) \tag{6.30}$$

and

$$\begin{aligned}\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) \le& \frac{\sqrt{2}}{N}\left|\left(\Delta^B_{J^c}, r\Delta^G_{J^c}\right)\right|_{2,1} + \frac{1}{N}\left|\Delta^B_J\right|_1 \\ &+ r\left(\frac{1}{N}\left|\Delta^G_J\right|_1 + \frac{1}{\sqrt{N}}|\Delta^\theta|_1 + \frac{1}{\sqrt{N}}|\Delta^\delta|_1\right) ,\end{aligned} \tag{6.31}$$

injecting (6.30) into (6.31) yields

$$\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) \le r|\Delta^G_j|\frac{4\,|J| + (p+q)\sqrt{N}}{N - c\sqrt{2}} . \tag{6.32}$$

This yields

$$\frac{|\Psi\Delta|_\infty}{\frac{r}{N}|\Delta^G|_1 + \frac{1}{N}|\Delta^B|_1 + \frac{r}{\sqrt{N}}\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right)} \ge \frac{|\Psi\Delta|_\infty}{|\Delta^G_j|}\frac{N - c\sqrt{2}}{r\left(4s + (p+q)\sqrt{N}\right)} \tag{6.33}$$

□

Proof of Theorem 4.3 and Theorem 4.4. The only unknown in the bounds of Theorem 6.1 is the set $J(B)$. However, by Theorem 4.2, $J(\widetilde{B})=J(B)$ with probablity close to 1, and thus it is possible to plug in a data-driven $\widehat{J}=J(\widetilde{B})$ instead of $J(B)$. This together with Theorem 6.1 leads to Theorem 4.3. Theorem 4.4 is a simple consequence of Theorem 6.1 and Proposition 4.2. $\square$

Proof of Theorem 4.2. Fix Θ in $\mathcal{B}_s(r)$. From the assumptions one can check that

$$\omega_k(s) \le \frac{2\sigma_*(s)\tau_*(c,r,s)\ r}{v_k\kappa_{*k}(s)}\ .$$

The right-hand side that we denote by ω_k^* is the upper bound that appears in the upper bound in Theorem 4.1 (i) under the stronger Assumption 4.2. By assumption, $|B_k|>2\omega_k^*$ for $k\in J(B)$. Note that the following two cases can occur. First, if $k\in J(B)^c$ (so that $B_k=0$) then, using (6.4), one obtains $|\widehat{B}_k|\le\omega_k(s)$ which implies that $\widetilde{B}_k=0$. Second, if $k\in J(B)$, then using again (6.4) again one gets $||B_k|-|\widehat{B}_k||\le|B_k-\widehat{B}_k|\le\omega_k(s)\le\omega_k^*$. Since $|B_k|>2\omega_k^*$ for $k\in J(B)$, one obtains that $|\widehat{B}_k|>\omega_k^*$ thus $|\widehat{B}_k|>\omega_k(s)$, so that $\widetilde{B}_k=\widehat{B}_k$ and the signs of B_k and $\widehat{B}_k$ coincide. The same is true if one replaces B by G and k by $k+N(N+1)$. This yields the result. $\square$

Theorem 6.2. *Under Assumpion 4.1, for every Θ in $\mathcal{B}_s$, on the event $\mathcal{G}(\Theta)$, for any solution $\widehat{\Theta}^{2S}$ of the minimization problem* (4.6)*, the following inequalities hold, for every $k=1,\dots,N(N-1)+p+q$,*

$$\left|\widehat{\Theta}_k^{2S}-\Theta_k\right| \le \frac{2\left(\boldsymbol{D^R}\right)_{k,k}C(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa_{k,J(B)}^{*2S}}\ , \tag{6.34}$$

$$\left|\widehat{\overline{\beta}}^{2S}-\overline{\beta}\right| \le \frac{2C(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa_{J(B)}^{\mathrm{end},2S}}\ , \tag{6.35}$$

$$\left|\widehat{\overline{\gamma}}^{2S}-\overline{\gamma}\right| \le \frac{2C(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\ r}{\kappa_{J(B)}^{\mathrm{ex},2S}}\ . \tag{6.36}$$

Proof of Proposition 5.1. Let us prove item (i).
Any Δ in the cone $C_{J,\gamma}$ is such that

$$\left|\left(\Delta^B,r\Delta^G\right)\right|_{\infty,1}\le 2(\gamma+1)\left|\left(\Delta^B_J,r\Delta^G_J\right)\right|_{\infty,1}+\gamma\left(r\left(\left|\Delta^G_J\right|_1+|\Delta^\theta|_1+|\Delta^\delta|_1\right)+\left|\Delta^B_J\right|_1\right)\ .$$

Let us show the lower bound considering case 1 out of 4 cases: there one has $\max\left\{r\left|\Delta^G_J\right|_\infty, r|\Delta^\theta|_\infty, r|\Delta^\delta|_\infty, \left|\Delta^B_J\right|_\infty\right\} \le r|\Delta^G_j|$ for some j, which implies

$$\left|\left(\Delta^B, r\Delta^G\right)\right|_{\infty,1} \le r|\Delta^G_j|\left(2(2\gamma+1)s + \gamma(p+q)\right) \ . \tag{6.37}$$

Because

$$r\left(|\Delta^G|_1 + |\Delta^\theta|_1 + |\Delta^\delta|_1\right) + |\Delta^B|_1 \le 2\left|\left(\Delta^B, r\Delta^G\right)\right|_{\infty,1} + r\left(|\Delta^\theta|_1 + |\Delta^\delta|_1\right) \ , \tag{6.38}$$

one obtains that

$$r\left(|\Delta^G|_1 + |\Delta^\theta|_1 + |\Delta^\delta|_1\right) + |\Delta^B|_1 \le r|\Delta^G_j|\left(4(2\gamma+1)s + (\gamma+1)(p+q)\right) \ . \tag{6.39}$$

Item (ii) in case 1 follows from the fact that

$$\left|\left(\Delta^B_J, r\Delta^G_J\right)\right|_{\infty,1} \le rs|\Delta^G_j| \ . \tag{6.40}$$

(EG) □

Proof of Theorem 6.2. Consider $\Theta \in \mathcal{B}_s$ and denote by $\Delta = \boldsymbol{D}^{\boldsymbol{R}}\left(\widehat{\Theta}^{2s} - \widehat{\Theta}\right)$.

Due to Proposition 4.1, one has that on the event $\mathcal{G}(\Theta)$, Θ belongs to $\widehat{\mathcal{I}}\left(rC(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\right)$. The triangle inequality yields that, on the event $\mathcal{G}(\Theta)$,

$$|\Psi\Delta|_\infty \le 2rC(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})} \ . \tag{6.41}$$

On the other hand, $\widehat{\Theta}^{2S}$ minimizes (5.7) on the set $\widehat{\mathcal{I}}\left(rC(\gamma,r,s)\sqrt{\widehat{Q}(\widehat{\Theta}^{1S})}\right)$. Thus, on the event $\mathcal{G}(\Theta)$,

$$\left|\left(\left(\boldsymbol{D}^{\boldsymbol{B}}\right)^{-1}\widehat{B}^{2S}, r\left(\boldsymbol{D}^{\boldsymbol{G}}\right)^{-1}\widehat{G}^{2S}\right)\right|_{\infty,1} \le \left|\left(\left(\boldsymbol{D}^{\boldsymbol{B}}\right)^{-1}B, r\left(\boldsymbol{D}^{\boldsymbol{G}}\right)^{-1}G\right)\right|_{\infty,1} \ . \tag{6.42}$$

This implies by the triangle inequality and the fact that $J(B) = J(G)$, that on the event $\mathcal{G}(\Theta)$,

$$\left|\left(\Delta^B_{J(B)^c}, r\Delta^G_{J(B)^c}\right)\right|_{\infty,1} \le \left|\left(\Delta^B_{J(B)}, r\Delta^G_{J(B)}\right)\right|_{\infty,1} \ .$$

One concludes using the definition of the sensitivities. □

References

[1] Arrelano, M., and 0. Bover (1995): "Another look at the instrumental variable estimation of error-components models ". *Journal of Econometrics,* 68, 29–51.

[2] Belloni, A., V. Chernozhukov, and L. Wang (2011): "Square-Root Lasso: Pivotal Recovery of Sparse Signals Via Conic Programming". *Biometrika,* 98, 791–806.

[3] Belloni, A., V. Chernozhukov, C. Hansen and D. Kozbur (2014): "Inference in High Dimensional Panel Models with an Application to Gun Control ". Preprint arXiv:1411.6507v1.

[4] Bickel, P., J. Y. Ritov, and A. B. Tsybakov (2009): "Simultaneous Analysis of Lasso and Dantzig Selector". *Annals of Statistics,* 37, 1705–1732.

[5] Bramoullé, Y., Djebbari, H., and B. Fortin (2009): "Identification of Peer Effects Through Social Networks ". *Journal of Econometrics,* 150, 41–55.

[6] Candès, E., and T. Tao (2007): "The Dantzig Selector: Statistical Estimation when p is Much Larger Than n". *Annals of Statistics,* 35, 2313–2351.

[7] Champion, M. (2015): "Contribution à la Modélisation et l'Inférence de Réseaux de Régulation de Gènes". Ph.D. manuscript tel-01112126v1.

[8] Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors". *Annals of Statistics,* 41, 2786–2819.

[9] Davezies, L., X. dHaultfoeuille, and D. Fougere (2006): "Identification of peer effects using group size variation textquotedblright. *Econometrics Journal,* 12, 397–413.

[10] Dufour, J.-M. (1997): "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models". *Econometrica,* 65, 1365–1387.

[11] Gautier, E., and A. Tsybakov (2011, 2014): "High-dimensional Instrumental Variables Regression and Confidence Sets". Preprint arXiv:1105.2454. Revision of the preprint posted in (2011).

[12] Gautier, E., and A. Tsybakov (2013): "Pivotal Estimation in High-Dimensional Regression via Linear Programming". in: *Empirical Inference, Festschrift in Honor of Vladimir N. Vapnik*, Springer.

[13] Grant, M., and S. Boyd (2013): "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta". http://cvxr.com/cvx.

[14] Lee, L.-F., and Yu, J. (2010): "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects". *Journal of Econometrics,* 154, 165–185.

[15] Lam, C., and P. CL. Souza (2013): "Regularization for Spatial Panel Time Series Using the Adaptive LASSO". Preprint.

[16] Lee, L.-f. (2007): "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects". *Journal of Econometrics,* 140, 333–374.

[17] Lounici, K. (2008): "Sup-Norm Convergence Rate and Sign Concentration Property of the Lasso and Dantzig Selector". *Electronic Journal of Statistics,* 2, 90–102.

[18] Lounici, K., M. Pontil, S. van de Geer and A. B. Tsybakov (2011): "Oracle Inequalities and Optimal Inference Under Group Sparsity". *Annals of Statistics,* 39, 2164–2204.

[19] Manresa, E. (2013): "Estimating the Structure of Social Interactions Using Panel Data". Working paper.

[20] Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem". *Review of Economic Studies,* 60, 531–542.

[21] Nickl, R., and S. van de Geer (2013): "Confidence Sets in Sparse Regression". *Annals of Statistics,* 41, 2852–2876.

[22] Peters, J., and P. Bühlmann (2014): "Identifiability of Gaussian Structural Equation Models With Equal Error Variance". *Biometrika,* 101, 219–228.

[23] Romano, J. P., and A. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models". *Journal of Statistical Planning and Inference,* 138, 2786–2807.

[24] Rose, C.: "Identification of Networks & Endogenous Spillover Effects using Panel Data".

[25] van de Geer, S., and P. Bühlmann: "ℓ_0-penalized Maximum Likelihood For Sparse Directed Acyclic Graphs". *The Annals of Statistics*, 41, 536–567.

[26] Yu, J., R. de Jong, and L.-f. Lee: "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large". *Journal of Econometrics,* 146, 118–134.

Toulouse School of Economics, 21 allée de Brienne, 31000 Toulouse, France.

E-mail address: eric.gautier@tse-fr.eu