Identification and Estimation in Linear Network Models with Unobserved Network Ties

K. Chountas

New York University Abu Dhabi

CRETE 2025

The Linear Network Model

The researcher observes $(y_{it}, x_{it} : t = 1, ..., n)$ for each unit $i = 1, ..., M$. Outcomes are determined according to:

$$y_{it} = \gamma_0 \sum_{j \neq i} w_{0,ij} y_{it} + x_{it}^T \beta_0 + \sum_{j \neq i} w_{0,ij} x_{jt}^T \delta_0 + e_{it},$$

1. **Model's Unknown Parameters:**
   - $\theta_0 = (\beta_0, \delta_0, \gamma_0)$, where:
     - γ_0, δ_0 represent the *endogenous* and *contextual* effects, respectively.
   - Network (weighted) Adjacency Matrix W_0:
     - The (i, j)-th entry, $w_{0,ij}$, reflects the strength of the link between units i and j.
2. **Covariates x_{it}:**
   - $x_{it} \in \mathbb{R}^k$ is assumed to be exogenous. If a subset is endogenous, it is assumed that valid instruments are available.
   - Covariates may include network-level regressors.
3. **Error Term e_{it}:**
   - e_{it} may include additively separable unobservables specific to each unit (i) and network-wide shocks at time t.

The Linear Network Model: Matrix Form

Stacking observations across units:

$$Y_t = X_t\beta_0 + W_0X_t\delta_0 + \gamma_0 W_0 Y_t + e_t,$$

for each $t = 1, ..., n$, where

1. Y_t is an $M-$vector of dependent variables,
2. X_t is a $M \times k$ matrix of explanatory variables,
3. e_t is an $M-$vector of errors,

and,

$$W_0 = \begin{bmatrix} 0 & w_{0,12} & \cdots & w_{0,1M} \\ w_{0,21} & 0 & \cdots & w_{0,2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{0,M1} & w_{0,M2} & \cdots & 0 \end{bmatrix}.$$

Literature: Identification with Known W_0

The identification analysis focuses on $\theta_0 = (\beta_0, \delta_0, \gamma_0)$, treating W_0 as known.

1. **Linear-in-Means Model:** $W_0 = (\iota\iota' - I) / (M - 1)$
   - Suffers from the *reflection problem*: difficult to distinguish between endogenous and contextual effects.
   - Identification is possible if multiple group sizes are observed (Lee, 2007; Davezies et al., 2009).
   - Alternative strategies exploit higher-order moments for identification (Glaeser et al., 1996; C. D. Rose, 2017).
2. **Network Model:**
   - Social interactions are structured through a network rather than the linear-in-means form.
   - θ_0 is identified if I, W_0, and W_0^2 are linearly independent (Bramoullé et al., 2009).

Literature: Identification with Known W_0

The adjacency matrix W_0 is rarely observed. Researchers often rely on:

- Postulating network ties or approximating the underlying network structure with a simplified network (e.g., linear-in-means).
- Treating social ties as known functions of certain proximity measures (Pinkse et al., 2002).
- Assuming a network formation model and estimating it jointly with the structural parameters (Goldsmith-Pinkham and Imbens, 2013; Hsieh et al., 2020; Johnsson and Moon, 2021).

However,

- These approaches are imperfect solutions to the fundamental problem of missing data.
- Misspecifying the network may invalidate the identification and estimation of social effects that rely on the network structure (Bramoullé et al., 2020; Griffith, 2022).

Literature: Identification with Unknown W_0

Recent literature explores identification with W_0 as an unknown parameter:

1. **C. Rose (2018)**
   - Proposes an identification strategy relying on network sparsity, analogous to exclusion restrictions in simultaneous equations.
   - Inspired by Gautier and Tsybakov (2014) on identification in models with many endogenous regressors and instruments with unknown exclusion restrictions.
2. **De Paula et al. (2024)**
   - Provide sufficient conditions for set identification of (θ_0, W_0).
   - Point identification is achieved with additional structure on W_0 or sign restrictions on θ_0.

Related Literature:

- Blume et al. (2015): θ_0 is identified if certain pairs of units are known not to be linked.
- Lewbel et al. (2023): Study identification of θ_0 with W_0 modeled as a matrix of random coefficients.

Main Assumption

Make the following assumption:

Assumption 1

The outcomes are determined by

$$Y = X\beta_0 + W_0 X\delta_0 + \gamma_0 W_0 Y + e,$$

The model's parameters (θ_0, W_0) satisfy

i. $|\gamma_0| < 1$.

ii. $\delta_0 + \beta_0\gamma_0 \neq 0$.

iii. For each $i = 1, ..., M$, $\sum_{j \neq i} |w_{0,ij}| \leq 1$ and $w_{0,ii} = 0$. Furthermore, there exists some i such that $\sum_{j \neq i} w_{0,ij} = 1$.

Definition: Global Identification

Let $\Pi : \mathcal{A} \times \Theta \longrightarrow \mathbb{R}^{M \times M}$ be defined by

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

where:

- $\mathcal{A}$ is the set of admissible Ws that satisfy the restrictions in Assumption 1(iii).
- Θ is the set of admissible θs that satisfy the restrictions in Assumption 1(i) and 1(ii).

Definition

(θ_0, W_0) is (globally) identified iff $\Pi(W, \theta) = \Pi_0 \implies (\theta, W) = (\theta_0, W_0)$ for each $(W, \theta) \in \mathcal{A} \times \Theta$, where $\Pi_0 = \Pi(W_0, \theta_0)$.

Identification Strategy (De Paula et al.,2024)

Assumption 1 holds. Additionally, make the following assumption:

Assumption 2 (De Paula et al., 2024)

The diagonal entries of W_0^2 are not proportional to the vector of all ones.

Define two regions of the parameter space:

$$\Theta_+ = \{\theta \in \Theta : \delta + \beta\gamma > 0\}, \quad \Theta_- = \{\theta \in \Theta : \delta + \beta\gamma < 0\}.$$

Let

$$\Pi : \mathcal{A} \times \Theta_\pm \to \mathbb{R}^{M \times M}.$$

Then:

- On Θ_+, (θ_0, W_0) is globally identified.
- Similarly, on Θ_-, (θ_0, W_0) is globally identified.

Identification Strategy: De Paula et al. (2024)

- Point identification requires that the sign of $\delta_0 + \beta_0\gamma_0$ is known. This is often implausible in real-world applications.
- The sign of $\delta_0 + \beta_0\gamma_0$ can be inferred from the reduced-form matrix under the additional assumptions that $\gamma_0 > 0$ and that W_0 is non-negative.
- Alternative conditions (De Paula et al., 2024) are also available, but involve eigenvalue restrictions (real and bounded magnitude) that are less interpretable.

Identification Strategy: This Paper

Consider the collection of functions indexed by θ:

$$\{\Pi_\theta \text{ defined by } W \mapsto \Pi(W, \theta) : \theta \in \Theta\}.$$

Assumption 1(i) and 1(iii) ensure that Π_θ is analytic for each $\theta \in \Theta$, and hence can be written as a polynomial in W. Then

$$W \in \bigcup_{\theta \in \Theta} \Pi_\theta^{-1}(\{\Pi_0\}) \implies W \in \text{Com}[\Pi_0],$$

or, equivalently

$$W \notin \text{Com}[\Pi_0] \implies W \notin \bigcup_{\theta \in \Theta} \Pi_\theta^{-1}(\{\Pi_0\}),$$

where,

$$\text{Com}[\Pi_0] := \left\{ W \in \mathbb{R}^{M \times M} \middle| \, W\Pi_0 = \Pi_0 W \right\}.$$

Identification Strategy: This Paper

- The above can be interpreted as follows: If W maps to the reduced-form matrix Π_0 under the (unknown) mapping Π_θ, then W must commute with Π_0.
- Thus, if the commutator of Π_0, denoted $\text{Com}[\Pi_0]$, is one-dimensional, then W_0 is identified.

This leads to the following assumption:

Assumption 2

For $\mathcal{W}$, the class of network matrices with no self-loops, it holds that

$$\mathcal{W} \cap \text{Com}[\Pi_0] = \{cW_0\}, \quad \text{for some } c \in \mathbb{R}.$$

Relation to Existing Results

- Assumption 2 (De Paula et al., 2024) is implied by our Assumption 2, but no additional restrictions are imposed on the lower-dimensional parameters θ_0.
- The problem is identifiable for almost every value of W_0, or for none, depending on the restrictions on the support of W_0 (Lemma 2.1, Barsotti et al., 2014).
- Support restrictions are more interpretable. For example:
  - If the network is known to be bipartite, Assumption 2 generally fails.
  - In other cases, non-identification occurs on a set of measure zero.
- The condition is also sufficient for identifying θ_0, making the result applicable to various estimation strategies.

Related Work Using Assumption 2

- Assumption 2 appears in Barsotti et al. (2014) in the context of recovering the transition matrix of a Markov chain observed at random times.
- The key feature in this application is that the transition matrix of the observed Markov chain can be expressed as a polynomial in the transition matrix of the original Markov chain.
- De Castro et al. (2017) use a similar strategy to reconstruct undirected graphs based on estimates of the graph's eigenspaces.
- Their approach leverages the fact that if the estimated eigenspaces are close to those of the graph's adjacency matrix, the two matrices will nearly commute.

Unit-Specific Fixed Effects and Common Shocks

- Unit-specific fixed effects can be easily incorporated: The reduced-form parameters can be estimated using either a demeaned or first-differenced equation as the estimating equation.
- Existing results do not account for common shocks:
  1. C. Rose (2018) does not consider common shocks at all.
  2. De Paula et al. (2024) address identification with common shocks, but their Proposition 1 contains a technical issue in its proof.
- This paper provides sufficient conditions for identification: The key condition boils down to the existence of a group of units that is isolated from the rest of the network.
- Analogous requirements are common in the peer-effects literature (e.g., Lewbel et al., 2023; Boucher et al., 2024).

Estimation-Fixed Network Size

For $\widehat{\Pi}_k$ being an estimator of Π_{0k} we have, for each $k = 1, ..., K$,

$$\left\| W_0 \widehat{\Pi}_k - \widehat{\Pi}_k W_0 \right\|_F \leq 2M \left\| \widehat{\Pi}_k - \Pi_{0k} \right\|_F \text{ a.s.}$$

Then, W_0 can be estimated using

$$\widehat{W} \in \arg\min_{W \in \mathcal{W}} \left\| \sum_{k=1}^{K} \left(W \widehat{\Pi}_k - \widehat{\Pi}_k W \right) \right\|_F^2 .$$

- The estimator is shown to be asymptotically normal.
- The estimator has a closed form and is easy to implement.

Estimation: High-Dimensional Networks

- **Adaptive Elastic Net GMM (Caner and Zhang, 2014):** Employed in De Paula et al. (2024), this method enforces sparsity on the network matrix itself but requires $n/M \to 0$.
- **Reduced-Form Lasso (Meinshausen & Bühlmann, 2006):** When $M \gg n$, estimate the reduced-form parameters via a Lasso-type estimator.
- **Note:** Sparsity in the adjacency matrix need not imply sparsity in the reduced-form matrix. Notions of sparsity must control the influence of indirect connections.
- **Sparsity Restriction (This Paper):** Assume the underlying graph has L disjoint components, with each component's vertex set $V_{0,\ell}$ satisfying

$$\max_{1 \le \ell \le L} |V_{0,\ell}| \;\le\; s_n \;=\; o(n).$$

Estimation: Group Lasso

- In some applications, it is reasonable to assume symmetry in the connectivity properties of the graph:

  i is connected to j iff j is connected to i (e.g., friendship networks).

- I consider a Group-Lasso estimator, solving:

$$\widehat{\Pi} \in \arg \min_{\Pi \in \mathbb{R}^{M \times M}} \left\{ \frac{1}{2n} \| Y - X\Pi^T \|_F^2 + \frac{\lambda}{n} \sum_{i=1}^{M} \sum_{j>i} \| \Pi(i,j) \| \right\},$$

  where,

$$\| \Pi(i,j) \|^2 = \begin{bmatrix} \pi_{ij} & \pi_{ji} \end{bmatrix} \begin{bmatrix} \pi_{ij} \\ \pi_{ji} \end{bmatrix}.$$

  under restricted eigenvalue assumptions (Tsybakov et al., 2009).

- The tuning parameter λ is chosen as in Tsybakov et al. (2009) and Belloni & Chernozhukov (2013) to guarantee the favorable performance of the estimator.

Group Lasso: Properties

Non-asymptotic bound

Under a restricted eigenvalue condition, if

$$P\left(\lambda \geq cn \max_{1\leq i,j\leq M}\left|\frac{2}{n}\sum_{t=1}^{n} U_{it}X_{tj}\right|\right) \geq 1-\alpha,$$

for some constant $c > \sqrt{2}$ and $\alpha \in (0,1)$, then with probability at least $1-\alpha$,

$$\|\widehat{\Pi} - \Pi_0\|_F \leq \left(1+\frac{\sqrt{2}}{c}\right)\frac{\lambda}{n}\frac{\sqrt{sL}}{\kappa_{\overline{c}}\,\kappa(\widehat{m})},$$

where $\overline{c} = (c+\sqrt{2})/(c-\sqrt{2})$.

One can set

$$\lambda := 2c\sigma\sqrt{n}\,\Phi^{-1}\left(1-\frac{\alpha}{2M^2}\right)$$

to ensure the above condition holds.

Empirical Application: Cigarette Demand

- Panel of 46 U.S. states, 1963–1992.
- Baltagi & Levin (1992): minimum real neighbor price as a bootlegging proxy.
- Elhorst (2014): W_0 as row-normalized binary contiguity (adjacency).
- Debarsy et al. (2012),: weights proportional to miles shared on state borders.
- **Limitation:** Pre-specified W_0 ignores other channels (transport links, high population density near state lines).
- **Our Approach:** Treat W_0 as unknown.

Model Specification

We assume the following model:

$$\ln C_{it} = \beta_1 \ln P_{it} + \beta_2 \ln I_{it} + \gamma \sum_{j \neq i} w_{ij} \ln C_{jt} + \alpha_i + e_{it},$$

where e_{it} is assumed to follow a spatial autoregressive process of the form

$$e_{it} = \rho \sum_{j \neq i} w_{ij} e_{jt} + u_{it}, \quad |\rho| < 1.$$

- C_{it}: packs of cigarettes per person i at time t; P_{it}: real price; I_{it}: real income.
- α_i are state fixed effects.

Estimation Results

	Binary contiguity W_{geo}	Estimated $\widehat{W}$
Parameter	(1)	(2)
γ	-0.253^{***} (0.054)	-0.419^{***} (0.061)
ρ	0.444^{***} (0.046)	0.581^{***} (0.043)
$\ln P_{it}$	-0.379^{***} (0.022)	-0.370^{***} (0.022)
$\ln I_{it}$	0.218^{***} (0.036)	0.200^{***} (0.036)
T	29	29
N	46	46

Future Work

- The proposed identification strategy is applicable at each $t = 1, \ldots, n$. Future research will focus on extending the estimation strategy to allow for networks that vary by observables (e.g. time).
- The identifiability condition depends on the rank of an estimable matrix. Testing the rank of this matrix is feasible using existing methods (Kleibergen and Paap, 2006; Donald et al., 2007).
- Examine conditions for identification in models beyond the canonical linear social interactions model. For example, in the framework of Graham (2008), combining conditional variance restrictions along our key identifying assumption (dimension of commutator) ensures identification.