Large Network Autoregressions with Unknown Adjacency Matrix

Kenwin Maung[1]

[1]University of Rochester

Motivation

Economic **networks**:

- **cross-sectional dependence** and (spatial) **interactions**
- **between** economic **units** (agents) or **variables**

Motivation

Economic **networks**:

- **cross-sectional dependence** and (spatial) **interactions**
- **between** economic **units** (agents) or **variables**

Significance:

- Manifest when agents/variables interact; **encodes information about economic behavior**
- **Scale** and **type** of interconnectivity will expand: *social media, financial integration, trade liberalization etc.*
- **Applications**: asset pricing (Ahern, 2013), exchange rates (Richmond, 2019), firm growth (Allen et al, 2019), corporate finance (Gofman and Wu, 2022) etc.

This paper

Estimate networks in a **large N large T** framework using a **Network Autoregression** (Zhu et al. 2018, Chen, 2020):

$$Y_t = \beta_0 W_0 Y_t + \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t$$

β_0: contemporaneous network effect;

β_j: dynamic network effect;

W: $N \times N$ adjacency matrix;

W_0: W with 0s on diagonal;

This paper

Estimate networks in a **large N large T** framework using a **Network Autoregression** (Zhu et al. 2018, Chen, 2020):

$$Y_t = \beta_0 W_0 Y_t + \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t$$

β_0: contemporaneous network effect;

β_j: dynamic network effect;

W: $N \times N$ adjacency matrix;

W_0: W with 0s on diagonal;

Key contribution: Estimation without knowing W beforehand

Solutions in the literature

Solution 1: Sparsity

- **Sparsity**: Assume many coefficients actually 0.
- Estimate with shrinkage (e.g. Lasso, SCAD).

Limitation

- Strong assumption in many applications. E.g. international trade networks are dense (De Benedictis and Tajoli, 2011).

Solution 2: (Weighted) Aggregation

Reduce dimension using spatial weights:

Dynamic spatial Durbin model; **Global VAR** (Pesaran et al, 2004); **Infinite-dimensional VAR** (Chudik and Pesaran, 2011); **Stochastic block VAR** (Guðmundsson and Brownlees, 2021); **Network autoregression** (Zhu et al, 2018)

Solutions in the literature

Solution 2: (Weighted) Aggregation

Reduce dimension using spatial weights:

Dynamic spatial Durbin model; **Global VAR** (Pesaran et al, 2004); **Infinite-dimensional VAR** (Chudik and Pesaran, 2011); **Stochastic block VAR** (Guðmundsson and Brownlees, 2021); **Network autoregression** (Zhu et al, 2018)

Complications:

- Do all applications have natural spatial weights?
- How to specify weights? Ad-hoc?

Proposed solution

Treat adjacency matrix, W **, as unknown and estimate it**. But W is $N \times N$.

Focus on **centralities** (low-rank representations):

- **Position** and **influence** of nodes (units) in the network
- **Hub** and **Authority** centralities

Proposed solution

Hub and **Authority** centralities:

A node is an authority if it is linked to by hubs; it is a hub if it links to authorities.

Webpage network:

- Hub score of a node depends on the total authority score of web pages it links to;
- Authority score of a webpage depends on total hub score of the webpages it receives links from;

Proposed solution

Given adjacency matrix W, hub and authority centralities are extracted as left and right singular vectors of W.

Cai et al. (2021) model this as:

$$W = dab^\top$$

d: scalar; a, b: $N \times 1$ vector.

Implication: N^2 coefficients in $W \Rightarrow N + N$ coefficients.

Model framework

Network autoregression with P lags:

$$Y_t = \beta_0 W_0 Y_t + \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t$$

Model framework

Network autoregression with P lags:

$$
\begin{aligned}
Y_t &= \beta_0 W_0 Y_t + \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t \\
(I - \beta_0 W_0) Y_t &= \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t \\
&\equiv W \mathcal{Y}_{t-P}^{t-1} \beta + \tilde{u}_t \\
&= \tilde{a} \tilde{b}^\top \mathcal{Y}_{t-P}^{t-1} \beta + \tilde{u}_t
\end{aligned}
$$

where

$$\mathcal{Y}_{t-P}^{t-1} = [Y_{t-1}, \ldots, Y_{t-P}]$$

is a $N \times P$ matrix.

Network autoregression with P lags:

$$
\begin{aligned}
Y_t &= \beta_0 W_0 Y_t + \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t \\
(I - \beta_0 W_0) Y_t &= \beta_1 W Y_{t-1} + \ldots + \beta_P W Y_{t-P} + \tilde{u}_t \\
&\equiv W \mathcal{Y}_{t-P}^{t-1} \beta + \tilde{u}_t \\
&= \tilde{a} \tilde{b}^\top \mathcal{Y}_{t-P}^{t-1} \beta + \tilde{u}_t
\end{aligned}
$$

where

$$
\mathcal{Y}_{t-P}^{t-1} = [Y_{t-1}, \ldots, Y_{t-P}]
$$

is a $N \times P$ matrix.

Furthermore,

$$
\tilde{u}_t = \tilde{\Lambda} f_t + \tilde{\varepsilon}_t,
$$

f_t: r common shocks;

$\tilde{\varepsilon}_t$: N idiosyncratic shocks;

Two-step Estimation

Reduced form:

$$\begin{aligned} Y_t &= (I - \beta_0 W_0)^{-1} W \mathcal{Y}_{t-P}^{t-1} \beta + (I - \beta_0 W_0)^{-1} \tilde{\Lambda} f_t + (I - \beta_0 W_0)^{-1} \tilde{\varepsilon}_t \\ &\equiv A \mathcal{Y}_{t-P}^{t-1} \beta + \Lambda f_t + \varepsilon_t \\ &\equiv a b^\top \mathcal{Y}_{t-P}^{t-1} \beta + u_t \end{aligned}$$

*$rank(A) = rank(W)$ since $(I - \beta_0 W_0)^{-1}$ is fullrank.,

Identification: Normalize $\|A\|_F = \|a\|_2 \|b\|_2 = 1$.

Two-step Estimation

$$Y_t = ab^\top \mathcal{Y}_{t-P}^{t-1} \beta + u_t$$

*u_t is uncorrelated over time

Iterative least squares: Step 1

1. Guess $\hat{\beta}^{(m)}$, $\hat{b}^{(m)}$;
2. Estimate $\hat{a}^{(m+1)} = (\sum_t Y_t \hat{\beta}^{(m)\top} \mathcal{Y}_{t-P}^{t-1\top} \hat{b}^{(m)})(\sum_t \hat{b}^{(m)\top} \mathcal{Y}_{t-P}^{t-1} \hat{\beta}^{(m)} \hat{\beta}^{(m)\top} \mathcal{Y}_{t-P}^{t-1\top} \hat{b}^{(m)})^{-1}$;
3. Update $\hat{b}^{(m+1)} = (\sum_t \hat{a}^{(m+1)\top} Y_t \hat{\beta}^{(m)\top} \mathcal{Y}_{t-P}^{t-1\top})(\sum_t \mathcal{Y}_{t-P}^{t-1} \hat{\beta}^{(m)} \hat{\beta}^{(m)\top} \mathcal{Y}_{t-P}^{t-1\top})^{-1}$;
4. Update $\hat{\beta}^{(m+1)} = (\sum_t \mathcal{Y}_{t-P}^{t-1\top} \hat{b}^{(m+1)} \hat{a}^{(m+1)\top} \hat{a}^{(m+1)} \hat{b}^{(m+1)\top} \mathcal{Y}_{t-P}^{t-1})^{-1} (\sum_t \mathcal{Y}_{t-P}^{t-1\top} \hat{b}^{(m+1)} \hat{a}^{(m+1)\top} Y_t)$;
5. Repeat until convergence.

Two-step Estimation

$$Y_t = ab^\top \mathcal{Y}_{t-P}^{t-1} \beta + u_t$$

*u_t is uncorrelated over time

Principal components analysis: Step 2

1. Define $\hat{Y}_t = Y_t - \hat{a}\hat{b}^\top \mathcal{Y}_{t-P}^{t-1} \hat{\beta}$ and we have a pure factor model:

$$\begin{aligned} \hat{Y}_t &= u_t \\ &= \Lambda f_t + \varepsilon_t \end{aligned}$$

2. Let $\hat{Y} = [\hat{Y}_{P+1}, \ldots, \hat{Y}_T]^\top$,

$$\text{estimator of } F \left(= [f_{P+1}^\top, \ldots, f_T^\top] \right) = r \text{ eigenvectors of } \frac{\hat{Y}^\top \hat{Y}}{(T-P)};$$

3. Estimate

$$\hat{\Lambda} = \frac{\hat{Y}^\top \hat{F}}{T-P}.$$

Asymptotic theory: Assumptions (1)

Assumption 1 (Composite errors) (i) f_t is a zero mean stationary white noise processes; (ii) $\|\varepsilon_t\| = O(\sqrt{N})$ almost surely; (iii) $f_t \perp \varepsilon_t$.

Assumption 2 (Factor structure) Normalize $F^\top F/T = I_r$ where $F = [f_1, \ldots, f_T]^\top$ and let $\Lambda^\top\Lambda$ be a diagonal matrix. Assume

(i) $E\|f_t\|^4 \leq C_f < \infty$, $\frac{1}{T}\sum_{t=1}^T f_t f_t^\top \to^p \Sigma_f > 0$ as $T \to \infty$ for some $r \times r$ matrix Σ_f,

(ii) $\|\lambda_i\| \leq C_\lambda < \infty$ where λ_i is the i^{th} row of Λ, $\frac{1}{N}\Lambda^\top\Lambda \to \Sigma_\lambda > 0$ as $N \to \infty$ for some $r \times r$ matrix Σ_λ.

Asymptotic theory: Assumptions (2)

Assumption 3 (Stationarity) Normalize $\|A\|_F = 1$ and define β_j to the be the j-th element of β. Assume the following stability condition holds for $P \geq 1$: The roots of the P-th order polynomial equation

$$z^P - |\beta_1| z^{P-1} - |\beta_2| z^{P-2} - \ldots - |\beta_{P-1}| z^1 - |\beta_p| = 0$$

lie inside the unit circle.

Assumption 4 (Rates) Let $N, T, P \to \infty$ and assume that

$$\frac{N \log N}{T - P} \to 0.$$

Asymptotic theory: Stationarity

Proposition

Under Assumptions 1-3, the (reduced form) model is stable.

Asymptotic theory: Consistency

Theorem

Under Assumptions 1-3, and with $A = ab^\top$, we have:

$$\left\| \hat{\beta}^\top \otimes \hat{a}\hat{b}^\top - \beta^\top \otimes ab^\top \right\|_S \to^p 0;$$

Alternatively, without assuming $A = ab^\top$ instead, we get:

$$\left\| \hat{\beta}^\top \otimes \hat{A} - \beta^\top \otimes A \right\|_S \to^p 0,$$

where $\|\cdot\|_S$ refers to the spectral norm of a matrix.

Asymptotic theory: Normality

Theorem

Let $L_{N,P}$ be a $q \times (2N+P)$ non-stochastic matrix such that $L_{N,P}L_{N,P}^{\top} \to L$ where L is some $q \times q$ non-negative symmetric matrix, for a finite q. Define $\hat{\theta} = (\hat{a}, \hat{b}, \hat{\beta}^{\top})^{\top}$. Given Assumptions 1-4 and $A = ab^{\top}$, we have:

$$L_{N,P}\Omega^{-1/2}\sqrt{(T-P)/N}(\hat{\theta}-\theta) \to^{d} N(0, L),$$

where $\Omega = H^{-1}E[G_t \Sigma G_t^{\top}]H^{-1\top}$, $\Sigma = Var(u_t)$.

Simulation set-up

Experiment 1

$$A = \left[\frac{1}{N}\right]_{N\times N}$$

and

$$a = (1/\sqrt{N}, \ldots, 1/\sqrt{N})^\top, b = (1/\sqrt{N}, \ldots, 1/\sqrt{N})^\top.$$

Experiment 2 (Two-neighbor model; Chudik and Pesaran, 2011)

$$A = \begin{bmatrix} 0.8 & -0.2 & 0 & \ldots & \ldots & 0 \\ -0.2 & 0.8 & -0.2 & 0 & \ldots & 0 \\ 0 & -0.2 & 0.8 & -0.2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & -0.2 & 0.8 & -0.2 \\ 0 & \ldots & \ldots & \ldots & -0.2 & 0.8 \end{bmatrix}.$$

Simulation set-up

- $P = 4$
- $\beta = (0.8, -0.4, 0.2, 0.1)$
- Number of common shocks = 1, i.i.d standard normal
- $\Lambda = (\underbrace{1, ..., 1}_{N/2}, \underbrace{0.5, ..., 0.5}_{N/2})$
- ε_t follows $N(0, \Sigma)$, $\Sigma_{(i,j)} = 0.5^{|i-j|}$

Simulation result

Table 1: Simulation results from experiment 1 and 2

Experiment 1	N		
	Reduced-rank (algorithm 1)		
	10	50	100
T			
300	1.22 (0.4)	3.36 (0.58)	6.41 (1.03)
400	1.02 (0.31)	2.78 (0.46)	4.80 (0.71)
500	0.85 (0.25)	2.31 (0.34)	3.88 (0.44)

Experiment 2	N			
	Reduced-rank (algorithm 1)		Arbitrary structure (algorithm 2)	
	10	20	10	20
T				
500	1.11 (0.085)	1.24 (0.109)	0.66 (0.072)	1.32 (0.083)
600	1.08 (0.076)	1.21 (0.087)	0.59 (0.066)	1.20 (0.075)
700	1.08 (0.07)	1.17 (0.076)	0.55 (0.06)	1.10 (0.069)

Recovering structural form

Structural VAR:

$$Y_t = (I - \beta_0 W_0)^{-1} W \mathcal{Y}_{t-P}^{t-1} \beta + (I - \beta_0 W_0)^{-1} \tilde{\Lambda} f_t + (I - \beta_0 W_0)^{-1} \tilde{\varepsilon}_t$$

Reduced form:

$$Y_t = A W \mathcal{Y}_{t-P}^{t-1} \beta + \Lambda f_t + \varepsilon_t$$

Estimate (over-identified) nonlinear system:

$$\hat{A} = \hat{a} \hat{b}^\top = (I - \beta_0 W_0)^{-1} W$$

$$Var(\hat{\varepsilon}_t) = (I - \beta_0 W_0)^{-1} (I - \beta_0 W_0)^{-1\top}$$

subject to $W = \tilde{a} \tilde{b}^\top$; $\|\tilde{a}\|_2 = \|\tilde{b}\|_2 = 1$.

Empirical application: Linkages in the US stock market

- Study a network of large-cap (monthly) stock returns;
- Greater degree of co-movement of asset prices may pose systemic stability issues;
- Adverse idiosyncratic shocks may more readily spillover to other stocks;
- Portfolios, ETFs, indices that include these stocks take on large correlated exposures;
- Increases the likelihood of simultaneous losses for investors.

Empirical application: Linkages in the US stock market

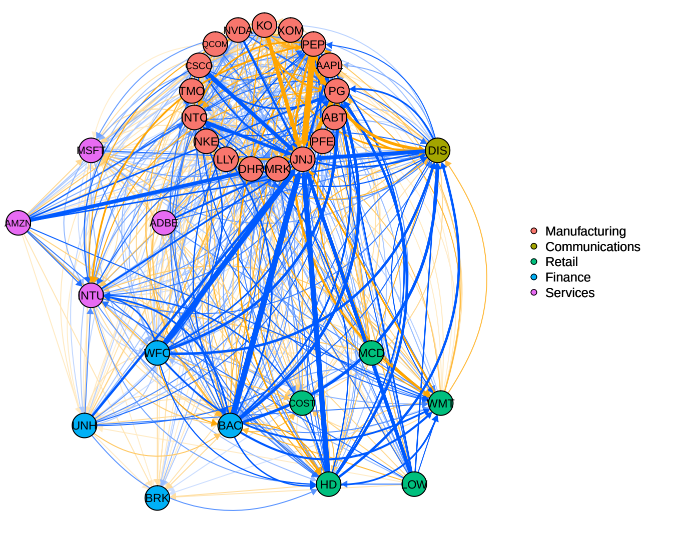
Data: Large-cap firms (top 31 holdings in Vanguard Large-cap ETF, VV)

Sample period: Jan 2010 - Dec 2019;

N = 31, T = 108 $\Rightarrow$ Unrestricted estimation of W **not feasible**;

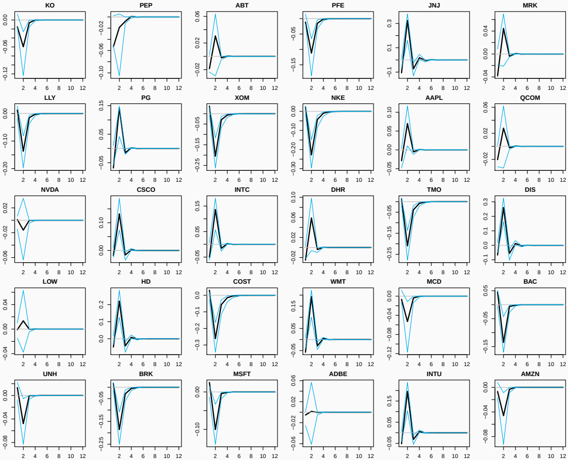
Empirical application: Linkages in the US stock market

Figure 1: Estimated adjacency matrix, W_0.



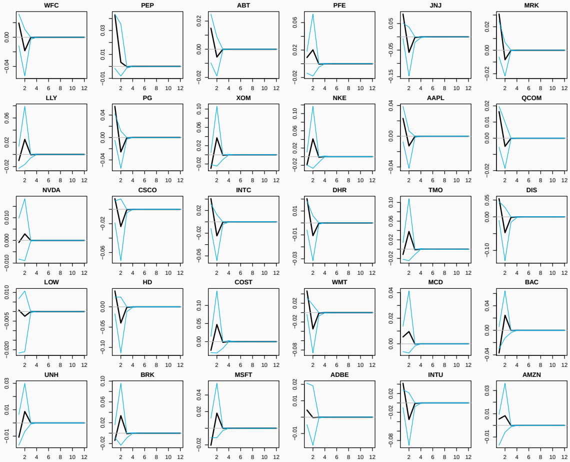
Empirical application: Linkages in the US stock market

Figure 2: Impulse response to a 1 s.d. negative shock to WFC.



Empirical application: Linkages in the US stock market

Figure 3: Impulse response to a 1 s.d. negative shock to KO.



Concluding remarks

- Proposed a novel SVAR methodology to handle large networks observed over time with **unknown adjacency matrices**;
- Reduced-form estimation is **consistent** and **asymptotically normal**;
- Numerical experiments show that the proposed iterated least squares algorithms can handle large network autoregressions well;
- Empirical application highlights the ability of the SVAR to isolate idiosyncratic impacts from common shocks;
- Provided **a tool to identify systemically important actors** in a network;