Spillover of risks and connectedness of an economic system

Fabio Canova
BI Norwegian Business School, CAMP, and CEPR
April 2023

Outline

- Economic measures of systemic risk.

- Networks. Spillovers,connectedness and systemicness.

- VAR-based spillovers/connectedness measures.

- Partial correlation-based spillovers/connectedness measures.

- Estimation of spillovers/connectedness in large systems (penalized, machine learning, Bayesian approaches).

- Stress testing.The Coincident indicator of systemic stress (CISS).

References

Acharya, V. Pedersen, L., Philippon T. and P. Richardson (2017). Measuring Systemic Risk, Review of Financial Studies, 30, 2-47.

Adrian, T. and M. Brunnemeier, (2017). CoVaR, American Economic Review, 106, 1705-1741.

Banburra, M., Giannone, D., and L. Reichlin (2010). Large Bayesian VAR, Journal of Applied Econometrics, 25, 71-92.

Barigozzi, M. and C. Brownlees (2018), NETS: Network Estimation for Time Series," Journal of Applied Econometrics, 34, 347-364.

Belmonte,M. Koop, G. and D. Korobilis (2014). Hierarchical Shrinkage in time varying parameters, Journal of Forecasting, 33, 80-94 .

Billio, M., Getmansky, M., Lo, A. and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in finance and insurance sectors, Journal of Financial Economics, 104, 535-559.

Bonaldi, P., A. Hortacsu, and J. Kastl (2013). An Empirical Analysis of Systemic Risk in the EURO-zone," Manuscript, University of Chicago.

Born, B., Mueller, G., Shularick, M and P. Sedlacek (2019). The costs of economic nationalism: Evidence from the brexit experiment. Economic Journal, 132, 2722-2744.

Brownless, C., and R. Engle (2017), SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, Review of Financial Studies, 30,48-79.

Canova, F. (2004). Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach, International Economic Review, 45, 2004, 49-77.

Canova, F. and M. Ciccarelli (2004). Forecasting and Turning points predictions in Bayesian panel VAR model, Journal of Econometrics, 120, 327-359.

Canova, F. and M. Ciccarelli (2009). Estimating multi-country VAR models, International Economic Review, 50, 929-961.

Canova, F. and M. Ciccarelli (2013). Panel Vector Autoregressive Models: A survey in T. B. Fomby, L. Kilian, A. Murphy (ed.) VAR Models in Macroeconomics — New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Volume 32), Emerald Group Publishing Limited, 2013, 205-246.

Dees, S., Di Mauro,F., Pesaran, H., and L.V. Smith (2007).Exploring the international linkages of the euro area: a global VAR analysis, Journal of Applied Econometrics, 22, 1-38.

Demirer, M., Diebold, F.. Liu, L. and K. Yilmaz, (2018). Estimating global bank network connectedness, Journal of Applied Econometrcis, 33, 1-15.

Dempster, A. P. (1972). Covariance selection. Biometrics, 28, 157-175.

De Paula, A. (2017). Econometrics of Network Models. in Advances in Economics and Econometrics: Theory and Applications, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press.

De Paula, A., Rasul, I. and S. Pedro (2018) Recovering social networks from panel data. Identification, simulation and one application, CEPR working paper 12792.

De Santis, R. and S. Zimic (2018). Spillovers among sovereign debt markets: identification by absolute magnitude restrictions, Journal of Applied Econometrics, 33, 727-747.

Diebold, F. and K. Yielmaz (2009). Measuring financial asset returns and volatility spillovers, with application to global equity markets, Economic Journal, 119, 158-171.

Diebold, F. and K. Yielmaz (2014). On the network topology of variance decomposition; measuring the connectedness of financial firms, Journal of Econometrics, 182, 119-134.

Faust, J. and E. Leeper (1997). Do Long Run Restrictions Really Identify Anything?, Journal of Business and Economic Statistics, 15, 345-353.

Giglio, S.W., B.T. Kelly, and S. Pruitt (2016). Systemic Risk and the Macroeconomy: An Empirical Evaluation, Journal of Financial Economics, 119, 457-471.

Gudmundson, G and C. Brownlees (2021). Detecting groups in large VARs, forthcoming, Journal of Econometrics

Hollo, D, Kremer, M. and M. Lo Duca (2012). CISS : a composite indicator of systemic stress in the financial system, ECB working paper 1426.

Kastl, J. (2017). Recent advances in empirical analysis of financial markets: industrial organization meets finance, In B. HonorÃ©, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), Advances in Economics and Econometrics: Eleventh World Congress (Econometric Society Monographs), 231-270. Cambridge: Cambridge University Press.

Kinn, D. (2017). Synthetic control methods and big data, BI manuscript.

Kocherols, T. (2018).Contagion stress testing, Norges Bank manuscript.

Korobilis, D. and K. Yilmaz (2018) Measuring Dynamic Connectedness with large Bayesian VAR models, Essex Finance Center, wp 27:01-2018.

Koop, G. and D. Korobilis (2013) large Time varying parameters VAR models, Journal of Econometrics, 177, 185-198.

Idier,J., Lamé G. and J.S. Mésonnier (2013). How useful is the Marginal Expected Shortfall for the measurement of systemic exposure? A practical assessment. ECB working paper 1546.

Moratis, G. and P. Sakellaris (2021). Measuring the systemic importance of banks. Journal of Financial Stability, 54, 100878.

Meinshausen, N. and P. Buhlmann (2006). High dimensional graphs and variable selection with the lasso. The Annals of Statistics, 34, 1436-1462.

Park, T. and R. Casella (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103, 681-688.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series B (Methodological), 267-288.

Zou, H. and H. Zhang (2009), On the Adaptive Elastic Net with a Diverging Number of Parameters, Annals of Statistics, 37, 1733.

1 Introduction

• Systemic risk management crucial after 2008. Need to monitor financial and non-bank institutions.

• Many measures of systemic risk: some based on economic concepts, others on statistical considerations.

• Some measures are ex-ante (leading), some ex-post (coincident). Some static, some dynamic. Some use balance sheet; other stock returns (levels or volatility) information.

• Useful indicators. None is perfect. Monitoring gains simultaneously using several indicators preferrable.

2 Economic measures of network risk

• Acharya et al. (2017), Adrian and Brunnemeier (2016), Brownlees and Engle (2017): measures of systemicness constructed estimating the impact of risk changes of one individual unit on the system.

• Adrian and Brunnemeier CoVaR: the change in the value at risk (VaR) of the financial institution, conditional on another institution being in distress, relative to its median state.

- CoVAR different from VaR (conditional vs. unconditional fall).

- Leverage, size, maturity mismatch, and asset price booms predict CoVaR.

- VaR value X_i is implicitly defined by the quantile

$$P(X_i \leq VaR_q^i) = q \tag{1}$$

- CoVaR value X_i is implicitly defined by the quantile

$$P(X_i \leq CoVaR_q^{i|j} | X_j \leq VaR_q^j) = q \tag{2}$$

- Excess risk measure:

$$\Delta CoVar_q^{i|j} = CoVaR_q^{i|j} - VAR_q^i \tag{3}$$

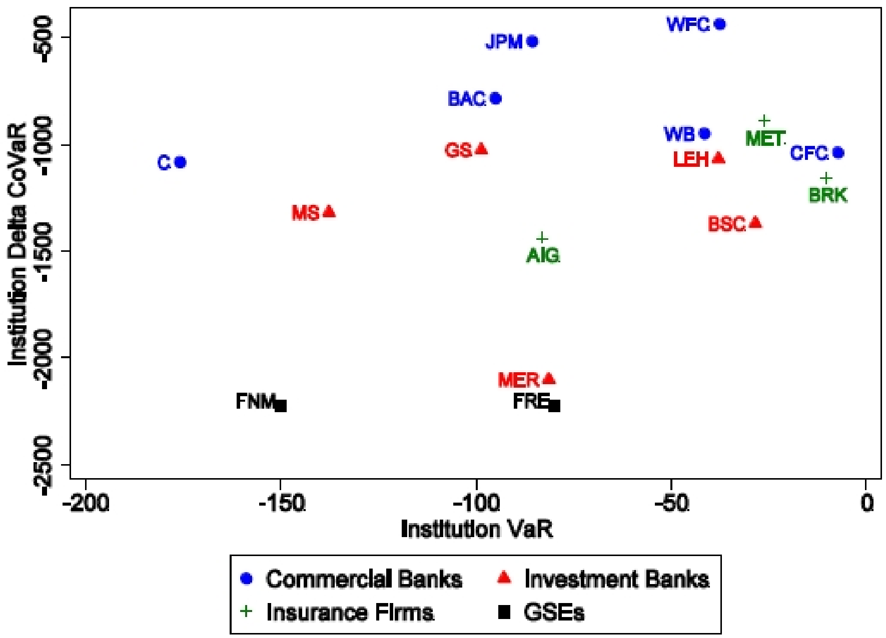
• Various conditioning possibilities:

- *Contribution*:$\Delta CoVaR$: Which institution j contributes most to $VAR_q^{system}|$ when j in distress?

- *Exposure*: $\Delta CoVaR$: which institutions most exposed to systemic stress: i.e. VaR_q^i given that system is in distress?

- *Network*: $\Delta CoVaR$: $\sum_i VAR_q^i$ conditional on j.

ΔCoVaR vs. VaR



- VaR and Δ CoVaR relationship is very weak
- Data up to 12/06

• Acarya et al: Marginal Expected shortfall (MES). The MES is the short-run expected equity loss, conditional on the market taking a loss greater than its Value-at-Risk at α%. If r_{it} (r_{mt}) is the daily (log) stock return of firm i (the market return the firm belongs to), MES is:

$$MES_{it} = E_t(R_{it+1}|R_{mt+1} \leq q_{\alpha,t}(R_{mt+1})) \equiv E_t(R_{it+1}|R_{mt+1} \leq C) \tag{4}$$

where C is a constant defining the "tail risk" of the market.

- The Expected **market** shortfall (ES) is the expected loss in the market index, conditional on a market loss larger than C (ω_i weight).

$$ES_t = E_t(R_{mt+1}|R_{mt+1} \leq C) = \sum_i \omega_i E_t(R_{it+1}|R_{mt+1} \leq C) \tag{5}$$

- If i belongs to the market, MES_{it} is the derivative of ES with respect to the firm market share (or capitalization) and reflects firm i participation in overall systemic risk.

- If a firm is not in the market index, MES measures the sensitivity (resilience) of this firm's stock price to exceptionally bad market events.

Bank / Variable	Loss	MES	CARTIER1	NPL	WFUND	HOL	CIL	LIQ	SIZE
Citigroup Inc.	1	6*	7*	6*	6*	58	49	59	1*
Central Pacific Fin. Corp.	2	26	53	59	43	2*	54	9	47
Regions Fin. Corp.	3	13	8*	17	34	18	35	19	7*
Marsall & Ilsley Corp.	4	14	1*	11	5*	23	10	17	17
Popular, Inc.	5	36	41	1*	2*	30	37	33	20
Zions Bancorp.	6	48	5*	43	33	12	12	38	19
Keycorp	7	2*	11	20	17	45	7*	11	13
Fifth Third Bancorp.	8	24	10*	12	44	37	14	13	12
Huntington Bancshares Inc.	9	38	27	7*	54	27	30	40	22
Suntrust Banks, Inc.	10	5*	3*	14	35	25	28	18	6*
Success ratio		30%	50%	30%	30%	10%	10%	10%	30%

Table 6: Rankings of the worst 10 stock return performers during the crisis according to various pre-crisis indicators as measured in 2007 Q2. The (dynamic BE) MES are estimated using information up to 2007Q2 only (ex ante view). * denotes a bank correctly identified ex-ante as incurring one of the top-10 losses.

• Brownlees and Engle: SRISK. An estimate of the amount of capital that a financial institution needs to raise if a financial crisis occurs.

- Computed recursively with market data on equities and balance sheet data on liabilities as

$$\begin{aligned} SRISK_{it} &= E_{t-1}(Capitalshortfall_{it}|Crisis_t) \\ &= E_{t-1}(K(Debt_{it} + Equity_{it}) - Equity_{it}|Crisis_t) \\ &= KDebt_{it} - (1-K)(1-LRMES_{it})Equity_{it} \end{aligned} \tag{6}$$

where K is the prudential level of equity to assets ($K \approx 0.08$); $LRMES_{it}$ is the expected decline in equity values following a financial crisis.

- $SRISK_{it}$ is reduced by reducing size, leverage or risk (measured by the comovements of firm equity with broad equity measures, i.e. firm's β).

• Why is $SRISK_{it}$ a measure of systemic risk?

- If we have a financial crisis, all firms with positive $SRISK_{it}$ will try simultaneously to raise capital and the only source of these funds is likely to be taxpayers.

- The larger is $SRISK_{it}$, the more serious the threat to financial stability.

- $SRISK_{it}$ is estimated conditional on an endogenous variable - it is a stress test does not measure causality.

• Macro-finance spiral:

- Firms with high SRISK recognize their vulnerability and will begin to deleverage and derisk, and this will impact on the real economy. If only a few firms have high SRISK, the remaining may take up the slack.

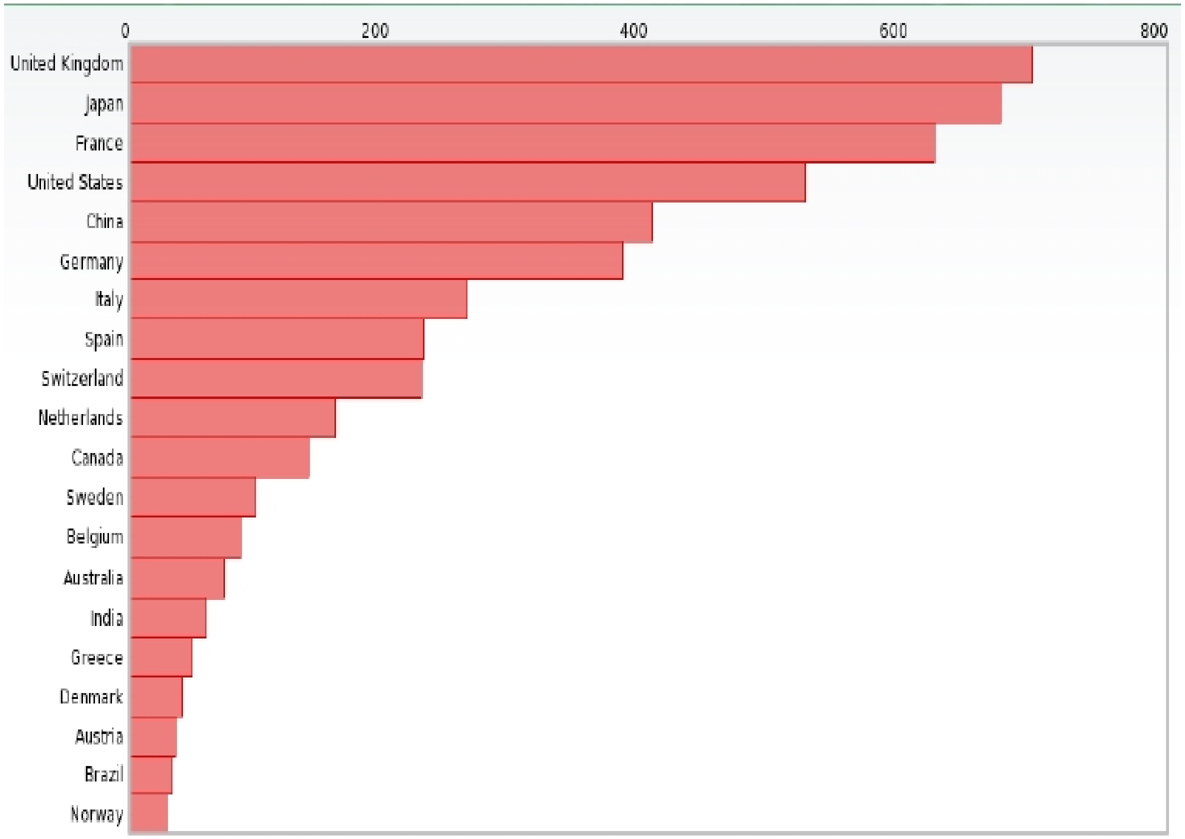
- As the macro economy slows down, stock prices fall, volatility rises, and SRISK increases.

- Investors recognize financial institution weakness and lower valuations further increasing SRISK. Forward looking investors could make this happen in one step.

- Without taxpayer support, bankruptcies and other failures will occur until the return to capital is high enough to bring new capital to the industry.

- If taxpayers step in, the spiral can be stopped before the bottom. However, this erodes market discipline and may impose huge regulatory costs on the financial sector going forward.

- Advance regulation is needed. Regulation should be countercyclical.



0
200
400
600
800
United Kingdom
Japan
France
United States
China
Germany
Italy
Spain
Switzerland
Netherlands
Canada
Sweden
Belgium
Australia
India
Greece
Denmark
Austria
Brazil
Norway

Systemic Risk Rankings for 2012-08-31 (MES is equity loss for a 2% daily market decline)

Institution	SRISK%	RNK	SRISK ($ m)	MES	Beta	Cor	Vol	Lvg	MV
Deutsche Bank AG	7.35	1	141,203	5.86	2.27	0.68	40	85.17	32,840.0
Barclays PLC	6.44	2	123,811	6.01	2.33	0.51	42.5	70.10	35,576.6
Credit Agricole SA	6.17	3	118,623	8.00	3.09	0.58	53.6	153.16	14,564.5
BNP Paribas	5.90	4	113,425	6.03	2.34	0.62	38.3	44.67	54,460.8
Royal Bank of Scotland Group PLC	5.25	5	100,830	5.30	2.06	0.51	40.3	52.88	40,506.7
Societe Generale	4.06	6	77,997	7.49	2.90	0.62	48.2	74.15	20,647.5
ING Groep NV	3.79	7	72,865	5.98	2.32	0.57	33.9	52.08	29,294.9
Lloyds Banking Group PLC	3.25	8	62,515	4.24	1.65	0.52	30.6	39.55	37,214.0
UBS AG-REG	3.19	9	61,386	4.83	1.87	0.57	32	34.25	42,853.3
Banco Santander SA	3.10	10	59,507	5.50	2.13	0.59	44.7	22.74	70,453.8
UniCredit SpA	2.88	11	55,396	6.64	2.57	0.55	47.8	50.13	22,930.0
HSBC Holdings PLC	2.60	12	49,874	3.07	1.19	0.59	23	16.60	158,925.5
Credit Suisse Group AG	2.48	13	47,578	4.65	1.80	0.50	32.1	42.39	25,481.4
Commerzbank AG	2.18	14	41,954	5.99	2.32	0.59	43.5	89.72	9,204.3
AXA SA	1.94	15	37,230	5.57	2.15	0.65	33.3	27.15	34,108.8
Intesa Sanpaolo SpA	1.90	16	36,574	7.37	2.86	0.55	53.4	31.60	25,543.5
Natixis	1.67	17	32,044	6.15	2.35	0.57	48.3	75.79	8,430.5
Dexia SA	1.52	18	29,294	6.24	2.43	0.37	78.1	1,093.52	490.0
Nordea Bank AB	1.51	19	29,008	3.62	1.40	0.58	22.9	24.04	37,452.1
Banco Bilbao Vizcaya Argentari	1.41	20	27,175	6.02	2.34	0.60	43.3	18.84	41,089.0

- Why does an institution have positive SRISK?

- Externalities. If only one firm has high SRISK, there is no spiral.

- Regulatory incentives. Implicit and explicit government guarantees.

- Risk based capital measures ignore correlation and hence leads to nondiversified asset mix. Risk weights may be poor proxy for true risk.

- Miscalculation 1: use short run (rather than long run) risk measures to choose leverage.

- Miscalculation 2: value exotic securities without recognizing their risks.

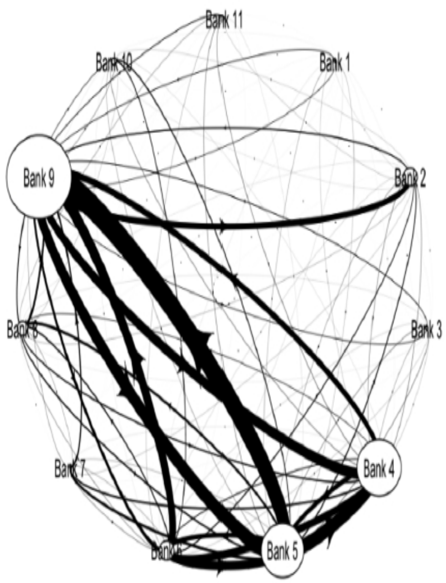
- Miscalculation 3: housing prices can go down.

3 Statistical measures of network risk

• Construct a network of financial linkages and evaluate the extent to which shocks from one unit spill to another or to all others units. No direct measure of the costs of systemicness is obtained.

• If the network is observable, can use standard network measures. If the network is not observable, can use auction data, bank funding costs, stock prices, or CDSs to estimate the strength of the links across institutions.

• Estimating network linkages has **computation and conceptual costs** if there are a large number of units-links. Need methods which enforce **sparcity** and **deal with large dimensionality** in estimating network spillovers.



Source: Denbee, Julliard, Li and Yuan (2014)

Source: Paul Butler

• A **network** $\mathcal{N}$ is composed of N units (nodes) and L links between nodes.

• A network is characterized by an **adjacency** matrix: A_{ij} with $A_{ij} = 1$ if i and j are linked and zero otherwise and $A_{ii} = 0, \ \forall i$.

• The **distance** between nodes i and j, denoted by s_{ij}, is the smallest number of links to be traversed between i and j.

• A **node degree** is the number of links from i to other nodes: $\delta_i = \sum_{j=1}^{N} A_{ij}$; the node degree is symmetric.

Example 3.1 *Suppose N=3, and* $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. *Then* $s_{13} = 2$ *and* $\delta_1 = 1, \delta_2 = 1, \delta_3 = 0$

• The degree distribution $p(\delta)$ is the probability distribution of node degrees. It is discrete, with support in [0, N-1].

• The **mean** of the degree distribution $E(\delta)$ is a canonical measure of connectedness. The larger is $E(\delta)$, the higher is the network connectedness.

• The **diameter** of the network is the maximum distance between two nodes $s_{\mathsf{max}} = \mathsf{max}_{i,j}\, s_{i,j} \approx \frac{\log N}{\ln E(\delta)}$ for N large.

• There is a relationship between the diameter of the network, s_{max}, the size of the network, N, and the mean of the degree distribution, $E(\delta)$. The larger is the mean degree, the smaller is the diameter of the network.

• Even if i and j are not connected contemporaneously, $A_{ij} = 0$, it may be connected via k in two steps. Thus, the above concepts can be extended at different horizon h, e.g. $A_{ij}^{h}, h = 0, 1, 2, \ldots$.

Standard measures of network connectedness/spillovers

- **Katz centrality** (measures the importance of a node within a network).

$$k_i(\lambda) = \sum_{h=1}^{\infty} \sum_{j=1}^{N} \lambda^h A_{i,j}^h \tag{7}$$

where $A_{i,j}^h$ measures the link from i to j at horizon h, and λ^h is a weight.

- If the network is stationary $k_i(\lambda^h) = 1$, and for $h \to \infty$ $k_i(\lambda) = ((I - A')^{-1} - I)$, and a k_{ij} is the response of unit i to a 100bp permanent change in unit j.

Katz centrality concept can be used to define the following two relevant notions:

• **Systemicness**: $k^s_j(\lambda) = \frac{1}{N}\sum_{h=1}^{\infty}\sum_{i=1}^{N}\lambda^h A^h_{i,j}$: the average effect of an exogenous shock to unit j variable on all other units (spillovers from j).

• **Vulnerability**: $k^v_i(\lambda) = \frac{1}{N}\sum_{h=1}^{\infty}\sum_{j=1}^{N}\lambda^h A^h_{i,j}$: the average effect of an exogenous shock to all unit variables on unit i (spillovers to i).

4 SVAR-based measures of spillovers and connectedness.

• Standard network measures based on Katz' centrality common (see e.g. Bonaldi et al, 2013; De Paula, 2016). For macro-finance problems, structural vector autoregressions (SVAR) measures more common.

• A panel VAR model (a VAR with variables across units) is network and the variance decomposition defines an adjacency matrix.

• Variance decomposition: how much of forecast error variation in one location (market, asset) is due to imported shocks (vulnerability) or how important one location is in spreading shocks elsewhere (systemicness).

• Three major differences with standard adjacency matrices:

1) The variance decomposition entries are not zero or one, but weights.

2) A variance decomposition defines directional links (potentially asymmetric).

3) There are rows/columns constraints to the weights (they need to sum to 1) and the diagonal elements are not zero.

• A **connectedness index** measures of the strength of directional spillovers (onto one unit, or from one unit to others). A **spillover index** is a synthetic measure of global interdependences (similar to Katz centrality).

- VAR and MAR:

$$
\begin{aligned}
y_t &= \psi(\ell) y_{t-1} + \epsilon_t, \quad \epsilon_t \sim (0, \Sigma) && (8) \\
&= \phi(\ell)\epsilon_t && (9)
\end{aligned}
$$

if $I - \psi(\ell)\ell$ is invertible, where y_t is a vector, dim (ϵ_t)=dim (y_t).

- Structural MAR (SMAR):

$$
y_t = A(\ell) A_0 u_t, \quad u_t \sim (0, I), \tag{10}
$$

where $\epsilon_t = QBu_t \equiv A_0 u_t$, $B = \text{diag}\{\sigma_i\}, A(\ell) = \phi(\ell)Q^{-1}$, $A(\ell) = A_0 + A_1\ell + \ldots$; $QBBQ^{-1} = \Sigma$; $\text{diag}\{Q_{ii}\} = 1$

- Connectedness to i at horizon H:

$$C_i(H) = 100 * \sum_{h=0}^{H-1} (\sum_{j=1}^{N})_{i \neq j} A_{h,ij}^2 \tag{11}$$

- Connectedness to j at horizon H

$$C_j(H) = 100 * \sum_{h=0}^{H-1} (\sum_{i=1}^{N})_{i \neq j} A_{h,ij}^2 \tag{12}$$

- Spillover index at horizon H:

$$S(H) = 100 * \frac{\sum_{h=0}^{H-1} (\sum_{i,j=1}^{N})_{i \neq j} A_{h,ij}^2}{\sum_{h=0}^{H-1} trace(A_h A_h')} \tag{13}$$

Example 4.1 *Let N=3 and* $y_t = \psi(L)y_{t-1} + \epsilon_t$. *Let B=I so that* $Q = A_0$. *Let* $H = 0$ *(instantaneous spillovers) and*

$$E(\epsilon_t\epsilon_t') = E(A_0 u_t u_t' A_0') = A_0 A_0' \equiv \begin{bmatrix} A^2_{0,11} & A^2_{0,12} & A^2_{0,13} \\ A^2_{0,21} & A^2_{0,22} & A^2_{0,23} \\ A^2_{0,31} & A^2_{0,32} & A^2_{0,33} \end{bmatrix}.$$

- *Connectedness to* Y_{2t} $S_{2,.}(0) = A^2_{0,21} + A^2_{0,23}$.

- *Connectedness from* Y_{2t} $S_{.,2}(0) = A^2_{0,12} + A^2_{0,32}$.

- *System spillovers* $S_{.,.}(0) = A^2_{0,12} + A^2_{0,13} + A^2_{0,21} + A^2_{0,23} + A^2_{0,31} + A^2_{0,32}$.

- *Spillover index = system spillovers/ total forecast error variations*

$$S(0) = 100 * \frac{S_{.,.}(0)}{S_{.,.}(0)+A^2_{0,11}+A^2_{0,22}+A^2_{0,33}}.$$

- *For H=1, the spillover index is:*

$$S(1) = 100 * \frac{S_{.,.}(0)+S_{.,.}(1)}{S_{.,.}(0)+A^2_{0,11}+A^2_{0,22}+A^2_{0,33}+S_{.,.}(1)+A^2_{1,11}+A^2_{1,22}+A^2_{1,33}}.$$

- *For H=2, the spillover index is:*

$$S(2) = 100 * \frac{S_{.,.}(0)+S_{.,.}(1)+S_{.,.}(2)}{S_{.,.}(0)+A^2_{0,11}+A^2_{0,22}+A^2_{0,33}+S_{.,.}(1)+A^2_{1,11}+A^2_{1,22}+A^2_{1,33}+S_{.,.}(2)+A^2_{2,11}+A^2_{2,22}+A^2_{2,33}}.$$

• Common to summarize VAR interdependences with the **bordered spillover matrix**.

Example 4.2 *Suppose the variance decomposition at horizon H is* $var_H = \begin{bmatrix} 40 & 40 & 20 \\ 30 & 50 & 20 \\ 30 & 30 & 40 \end{bmatrix}$. *The bordered spillover matrix is* $V_H = \begin{bmatrix} 40 & 40 & 20 & 60 \\ 30 & 50 & 20 & 50 \\ 30 & 30 & 40 & 60 \\ 60 & 70 & 40 & 340 \end{bmatrix}$.
The last column represents "connectedness to" and the last row "connectedness from". 340 is the total spillovers. Then $S(H) = 340/470 =$ 72.34% *is the spillover index and average spillover is 72.32/6=12.02%.*

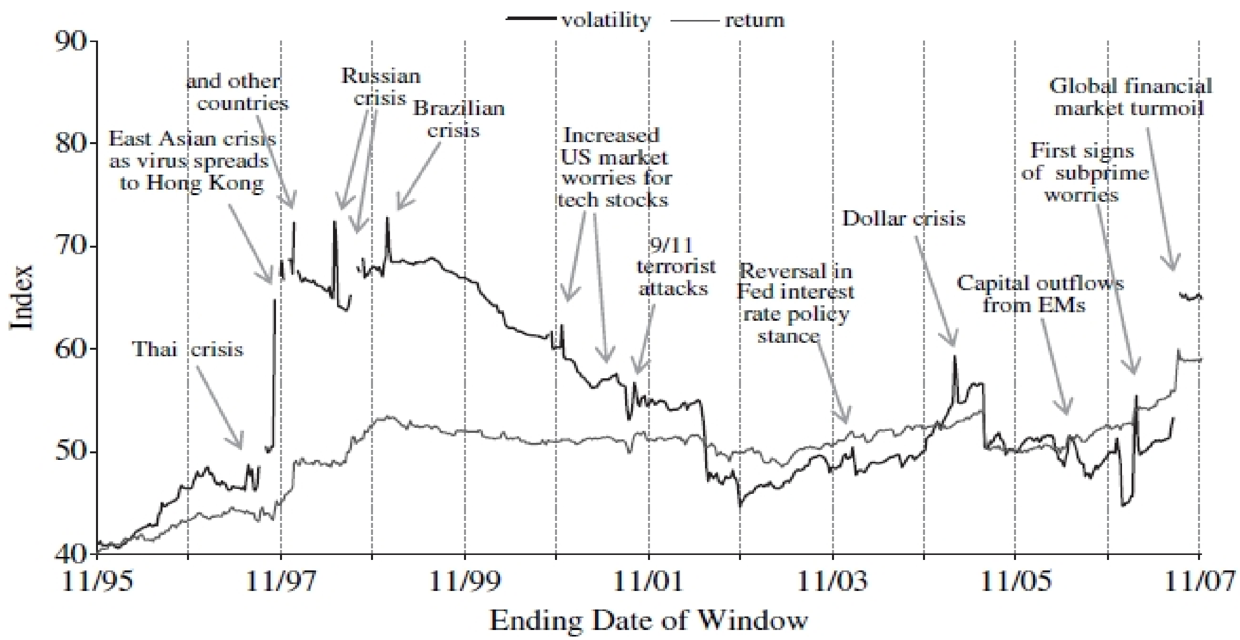
• Analogously with traditional measures we can define systemicness and vulnerability for variance decomposition.

Example 4.3 *Continuing with example 4.2, using the last row and the last column of V_H, we have that the vector of (average) systemicness is [30 25 30] and the vector of (average) vulnerability is [30, 35, 20]. Here unit 2 is less systemic and more vulnerable than the other two.*

• **What is** y_t? The level of a variable (e.g. stock returns) or its volatility. Volatilities typically preferred because they move together in crisis (returns move together in crises and in business cycle swings). Other level variables (e.g. bank funding costs) could be used. Ideally, use high frequency balance sheet or publically available data (bonds yields).

• If y_{it} is a vector, each i, could study cross-market, cross-country spillovers.

• Rolling estimates of connectedness and spillover indices describe time evolution. Can check whether time variations are linked to interesting events.



Source: Diebold and Yielmaz (2009).

Problems

- Computing the variance decomposition requires:

i) an horizon h (what is the relevant h?)

ii) the variables and the number of lags to be included in the VAR.

iii) an identification scheme to obtain structural shocks (VAR residuals are correlated).

- Are the conclusions robust to i)-ii)-iii)?

• The number and the type of variables and the number of lags are crucial to determine the quality of the estimate of $A(\ell)$.

• Unless Σ is almost diagonal, spillover estimates likely to depend on how shocks are identified.

• Rolling window estimates easy to compute but

i) How long should the window be?

ii) How much overlap across windows should we allow for?

• **The information set is crucial for the results**: If a variable is erroneously omitted, included shocks pick up spillovers generated by the omitted one

• Consider a bivariate VAR(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

where Y_{1t} is $m_1 \times 1$ vector and Y_{2t} is a $m_2 \times 1$ vector. Suppose Y_{2t} are omitted. Solving for Y_{2t} from the second set of equations we get

$$\begin{aligned} Y_{1t} &= A_{11}Y_{1t-1} + A_{12}A_{21}\sum_{j=0}^{\infty} A_{22}^{j} Y_{1t-2-j} + A_{12}\sum_{j=0}^{\infty} A_{22}^{j} e_{2t-1-j} + e_{1t} \\ &= C(\ell)Y_{1t-1} + D(\ell)e_{2t-1} + e_{1t} \qquad (14) \end{aligned}$$

• Since $v_t = D(\ell)e_{2t-1} + e_{1t}$, the representation for Y_{1t} is a $VARMA(\infty, \infty)$.

• v_t is combination of e_{1t} and e_{2t}; the timing of v_t is incorrect.

Example 4.4 *Suppose $m = 4$. If one uses bivariate VARs, there are three possible models. For example, the system with variables 1 and 3 has shocks*

$$\begin{bmatrix} \upsilon_{1t} \\ \upsilon_{2t} \end{bmatrix} \equiv \begin{bmatrix} e_{1t} \\ e_{3t} \end{bmatrix} - \Xi(\ell)\Phi(\ell) \begin{bmatrix} e_{2t} \\ e_{4t} \end{bmatrix} \text{ where } \Xi(\ell) = \begin{bmatrix} A_{12}(\ell) & A_{14}(\ell) \\ A_{32}(\ell) & A_{34}(\ell) \end{bmatrix}$$

$$\Phi(\ell) = \begin{bmatrix} A_{22}(\ell) & A_{24}(\ell) \\ A_{42}(\ell) & A_{44}(\ell) \end{bmatrix}^{-1}.$$ *Easy to verify that:*

- *Even if e_t's are contemporaneously and serially uncorrelated, υ_t's are contemporaneously and serially correlated, unless $\Xi(\ell)$ is block diagonal.*

- *Two small scale VAR systems, both with $m_1 = 2 < m = 4$ variables, have different innovations and different variance decompositions. For example, the system with variables 1 and 2 has shocks*

$$\begin{bmatrix} \upsilon_{1t} \\ \upsilon_{2t} \end{bmatrix} \equiv \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} - \tilde{\Xi}(\ell)\tilde{\Phi}(\ell) \begin{bmatrix} e_{3t} \\ e_{4t} \end{bmatrix} \textit{ where } \tilde{\Xi}(\ell) = \begin{bmatrix} A_{13}(\ell) & A_{14}(\ell) \\ A_{23}(\ell) & A_{24}(\ell) \end{bmatrix}$$

$$\tilde{\Phi}(\ell) = \begin{bmatrix} A_{33}(\ell) & A_{34}(\ell) \\ A_{43}(\ell) & A_{44}(\ell) \end{bmatrix}^{-1}.$$

•υ_t *is a linear combination of* **current and past** e_t*'s. The timing of innovations is preserved if and only if the* m_1 *included variables are Granger causally priori to the* $m - m_1$ *omitted ones.*

Example 4.5 *Use a VAR(2) with daily 2009d1-2012d1 data. The variables are 2 years bond returns for DE, FR, IT and ES with US bond returns as exogenous. Identify shocks with Choleski (Ok since covariance matrix of VAR residuals is almost diagonal). Use a Minnesota prior with overall tightness equal to 0.2. The bordered spillover matrix at H=15 (days) is*

$$\begin{array}{c} \\ DE \\ FR \\ IT \\ ES \\ \text{From} \end{array} \begin{array}{c} \begin{array}{ccccc} DE & FR & IT & ES & \text{To} \end{array} \\ \left[\begin{array}{ccccc} 0.92 & 0.01 & 0.01 & 0.03 & 0.05 \\ 0.03 & 0.95 & 0.00 & 0.01 & 0.04 \\ 0.00 & 0.00 & 0.96 & 0.03 & 0.03 \\ 0.00 & 0.00 & 0.01 & 0.98 & 0.01 \\ 0.03 & 0.01 & 0.02 & 0.07 & 0.26 \end{array} \right] \end{array}$$

Here $S(15) = 26/407 = 6.38\%$; *the average spillover* $6.38/8 = 0.79\%$.

• *If we use a VAR with bonds markets of Italy and Germany only, the bordered spillover matrix at H=15 days is*

$$\begin{array}{cc} & \begin{array}{ccc} DE & IT & \text{To} \end{array} \\ \begin{array}{c} DE \\ IT \\ \text{From} \end{array} & \left[\begin{array}{ccc} 0.995 & 0.005 & 0.005 \\ 0.003 & 0.997 & 0.003 \\ 0.003 & 0.005 & 0.016 \end{array}\right] \end{array}$$

Here $S(15) = 16/2008 = 0.79\%$; *the average spillover* $0.79/4 = 0.20\%$.

• **The lag length of the estimated VAR matters**.

Example 4.6 *Continuining with example 4.5, suppose you estimate the same system with 4 lags. The bordered spillover matrix at H=15 (days) is*

$$\begin{array}{c} \\ DE \\ FR \\ IT \\ ES \\ \text{From} \end{array} \begin{array}{c} \begin{array}{ccccc} DE & FR & IT & ES & \text{To} \end{array} \\ \left[\begin{array}{ccccc} 0.95 & 0.02 & 0.01 & 0.01 & 0.04 \\ 0.47 & 0.50 & 0.02 & 0.01 & 0.50 \\ 0.04 & 0.02 & 0.91 & 0.03 & 0.09 \\ 0.01 & 0.06 & 0.37 & 0.56 & 0.44 \\ 0.52 & 0.10 & 0.40 & 0.05 & 2.14 \end{array}\right] \end{array}$$

Here $S(15) = 214/506 = 42.29\%$*; the average spillover* $42.29/8 = 5.29\%$*.*

- **The horizon H also matters.**

Example 4.7 *Continuining with example 4.5, suppose you estimate the same system with 2 lags. The bordered spillover matrix at H=45 (days) is*

$$
\begin{array}{c c}
 & \begin{array}{ccccc} DE & FR & IT & ES & \text{To} \end{array} \\
\begin{array}{c} DE \\ FR \\ IT \\ ES \\ \text{From} \end{array} &
\left[\begin{array}{ccccc}
0.99 & 0.01 & 0.00 & 0.00 & 0.01 \\
0.67 & 0.32 & 0.01 & 0.00 & 0.68 \\
0.01 & 0.00 & 0.98 & 0.01 & 0.02 \\
0.01 & 0.02 & 0.50 & 0.47 & 0.53 \\
0.69 & 0.03 & 0.51 & 0.01 & 2.48
\end{array} \right]
\end{array}
$$

Here $S(45) = 248/524 = 47.32\%$*; the average spillover* $47.32/8 = 5.91\%$.

- **Shock identification matters unless Σ is diagonal.**

Example 4.8 *Suppose* $Q = \begin{bmatrix} 1 & b \\ c & 1 \end{bmatrix}, B = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$. *Since* $\boldsymbol{\Sigma}_\epsilon$ *is symmetric,* $\boldsymbol{\Sigma} = QBB'Q'$ *implies three restrictions in four unknown, i.e.*

$$\begin{aligned} \boldsymbol{\Sigma}_{11} &= \sigma_1^2 + b^2\sigma_2^2 && (15) \\ \boldsymbol{\Sigma}_{12} &= c\sigma_1^2 + b\sigma_2^2 && (16) \\ \boldsymbol{\Sigma}_{22} &= c^2\sigma_1^2 + \sigma_2^2 && (17) \end{aligned}$$

- *Identification problem! Typical choices:*

• *Zero restriction:. The system is assumed to be lower (upper) triangular on impact, i.e.* $\hat{A}_0 = \begin{bmatrix} (\sigma_1^2 + b^2\sigma_2^2)^{0.5} & 0 \\ \frac{c\sigma_1^2 + b\sigma_2^2}{[(\sigma_1^2 + b\sigma_2^2)^{0.5}]} & [(c^2\sigma_1^2 + \sigma_2^2) - \frac{(c\sigma_1^2 + b^2\sigma_2^2)}{(\sigma_1^2 + b^2\sigma_2^2)]^{0.5}} \end{bmatrix}$

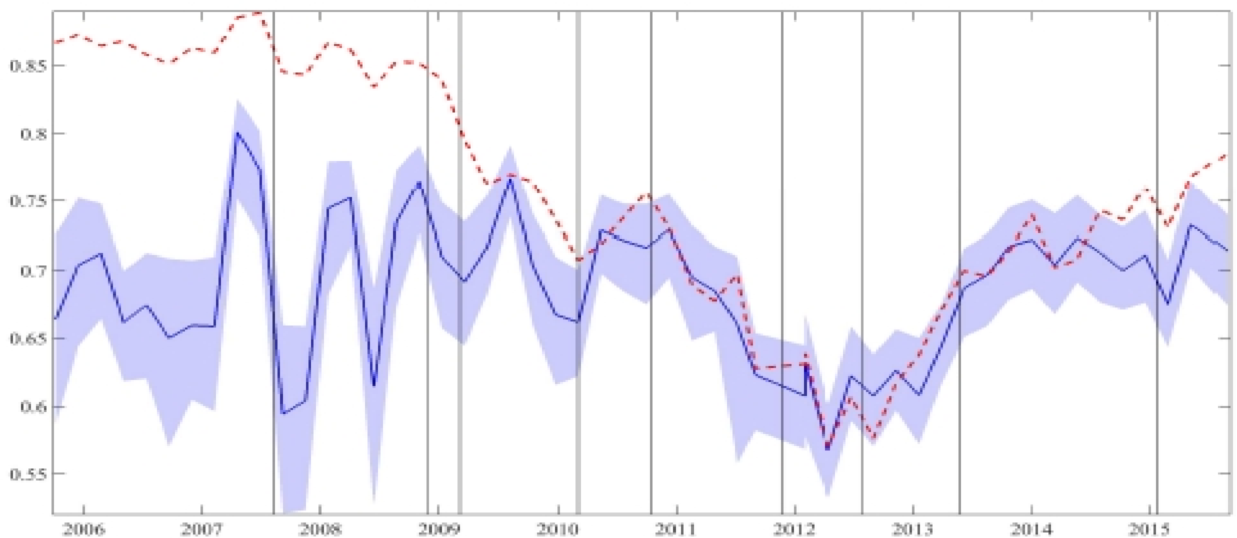
• *Pesaran - Shin GIR:* $\hat{A}_0 = \begin{bmatrix} \frac{\sigma_1^2 + b^2\sigma_2^2}{[(\sigma_1^2 + b^2\sigma_2^2)^{0.5}]} & \frac{c\sigma_1^2 + b\sigma_2^2}{[(c^2\sigma_1^2 + \sigma_2^2)^{0.5}]} \\ \frac{c\sigma_1^2 + b\sigma_2^2}{[(\sigma_1^2 + b\sigma_2^2)^{0.5}]} & \frac{c^2\sigma_1^2 + b\sigma_2^2}{[(c^2\sigma_1^2 + \sigma_2^2)^{0.5}]} \end{bmatrix}$.

• *Sign restrictions: e.g.* $\hat{A}_{0,12} > 0$ *or* $\hat{A}_{0,21} > 0$. *Problem: how do you distinguish two shocks with such a restriction? Restrictions are not mutually exclusive.*

• *Quantity restrictions (De Santis and Zimic, 2018):* $|\hat{A}_{0,12}|/|\hat{A}_{0,11}| < 1$, *and* $|\hat{A}_{0,21}|/|\hat{A}_{0,22}| < 1$, *i.e. instantaneous response of "other" units smaller than the instantaneous response of "own" unit.*

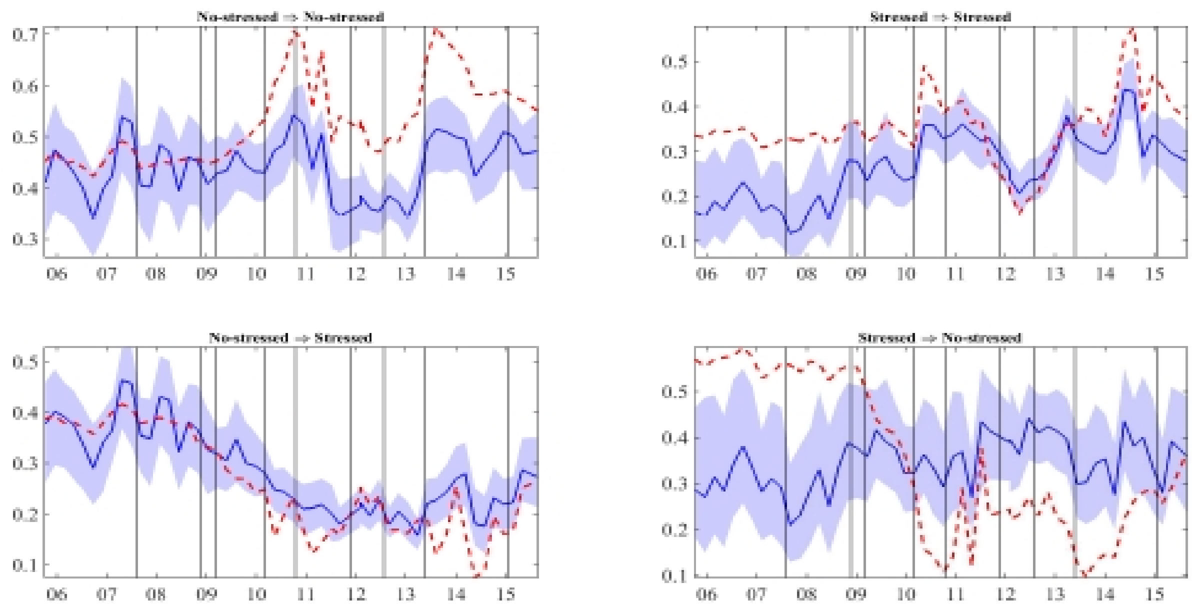
De Santis and Zimic, 2018: Spillovers in bond markets

Figure 8: Total connectedness among sovereign yields excluding monetary policy shocks



Note: This figure shows total connectedness among sovereign yields excluding monetary policy shocks ranging between 0 (no connectedness) and 1 (full connectedness). The blue line and the shaded area provide respectively the median estimate and the 68% error bands obtained using the magnitude restriction method. The dotted red line is the median estimate obtained using GIRFs. The vertical bars denote 9 August 2007 (the interbank credit crisis), 25 November 2008 (US LSAP), 5 March 2009 (BoE APF), 5 March 2010 (the Greek revised budget deficit), 18 October 2010 (the Deauville agreement upon Private Sector Involvement), 24 November 2011 (Fitch downgrade of Portugal's sovereign debt), 26 July 2012 (Draghi's speech), 22 May 2013 (FED tapering announcement), 22 January 2015 (ECB PSPP). Sample period: 3 January 2005 - 24 August 2015.

Figure 9: Directional connectedness among sovereign yields of stressed and non-stressed countries



Note: This figure shows directional connectedness among sovereign yields of stressed and non-stressed countries due to shocks to sovereign yields ranging between 0 (no connectedness) and 1 (full connectedness). The non-stressed country group is composed of Austria, Belgium, France, Germany, Netherlands, UK and US. The stressed country group is composed of Greece, Ireland, Italy, Portugal and Spain. The blue line and the shaded area provide respectively the median estimate and the 68% error bands obtained using the magnitude restriction method. The dotted red line is the median estimate obtained using GIRFs. The vertical bars denote 9 August 2007 (the interbank credit crisis), 25 November 2008 (US LSAP), 5 March 2009 (BoE APF), 5 March 2010 (the Greek revised budget deficit), 18 October 2010 (the Deauville agreement upon Private Sector Involvement), 24 November 2011 (Fitch downgrade of Portugal's sovereign debt), 26 July 2012 (Draghi's speech), 22 May 2013 (FED tapering announcement), 22 January 2015 (ECB PSPP). Sample period: 3 January 2005 - 24 August 2015.

• Rolling windows estimates generate artificial persistence (depending on the window size). Thus connectedness and systemicness measures are not very responsive to changes in the underlying fundamentals. Hard to use for real time monitoring.

• Alterantive: specify a time varying coefficient (TVC) VAR.

$$\begin{aligned} Y_t &= A_t Y_{t-1} + E_t \\ A_t &= A_{t-1} + U_t \end{aligned} \tag{18}$$

$E_t \sim N(0, \Sigma_t)$ and $U_t \sim N(0, \Omega_t)$.

• Σ_t and Ω_t could a stochastic volatility format $\Sigma_t = \Sigma_{t-1} + V_t$ and $\Omega_t = \Omega_{t-1} + \chi_t$,

• They could be specified using "forgetting factors" if $\dim(y_t)$ is large (see Koop and Kourobilis, 2013): e.g. $\Omega_t = \rho\Omega_{t-1}$ where $\rho \in [0.98, 0.99]$. Advantage: can use Kalman filter to estimate, no MCMC needed.

Kourobilis and Yielmaz (2018)

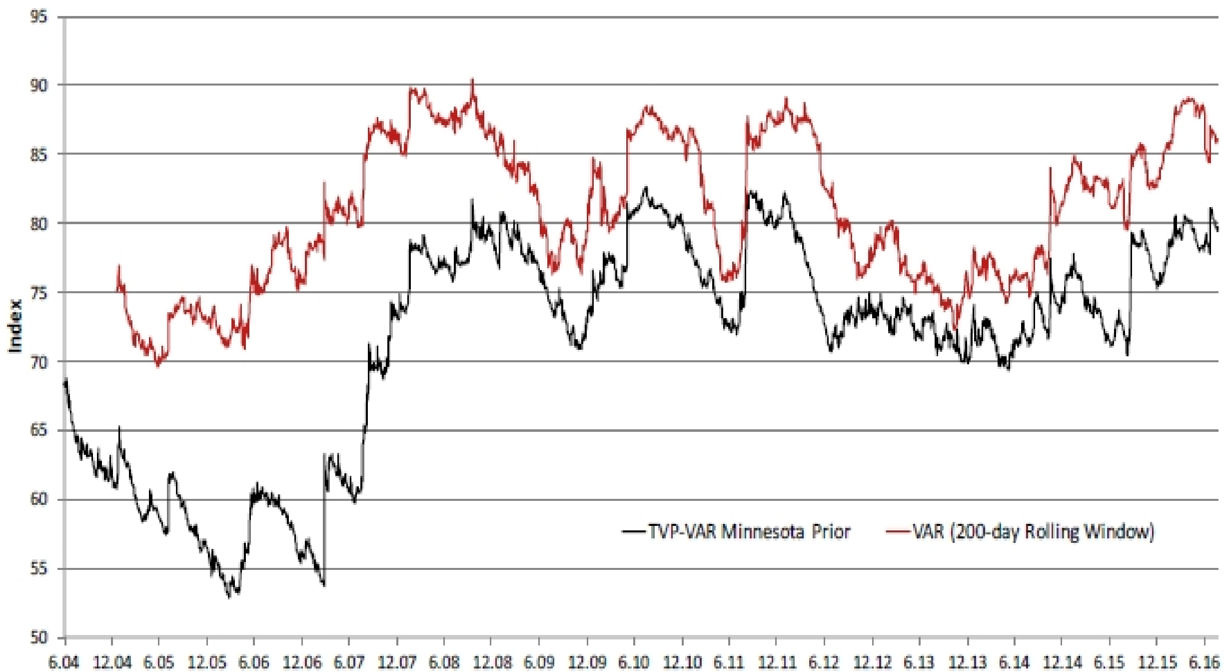


Figure 3: Total Connectedness: TVP-VAR vs VAR with 200-day Rolling Window

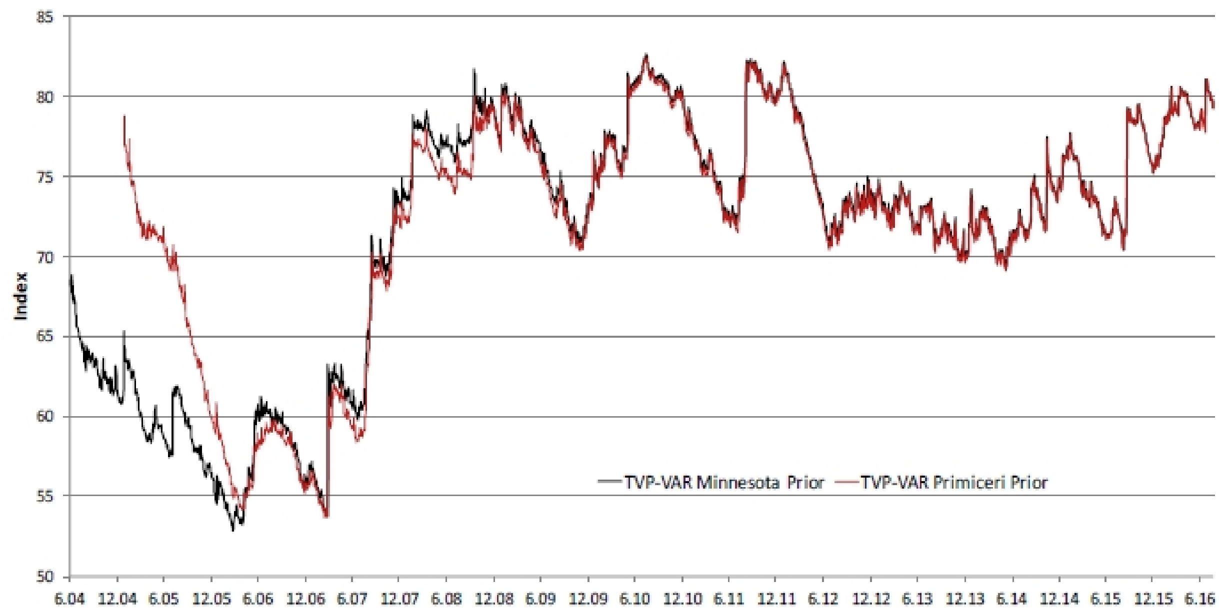


Figure 4: Total Connectedness: TVP-VAR Model with Primiceri Prior vs Minnesota Prior

The TVC-VAR has better predictive performance

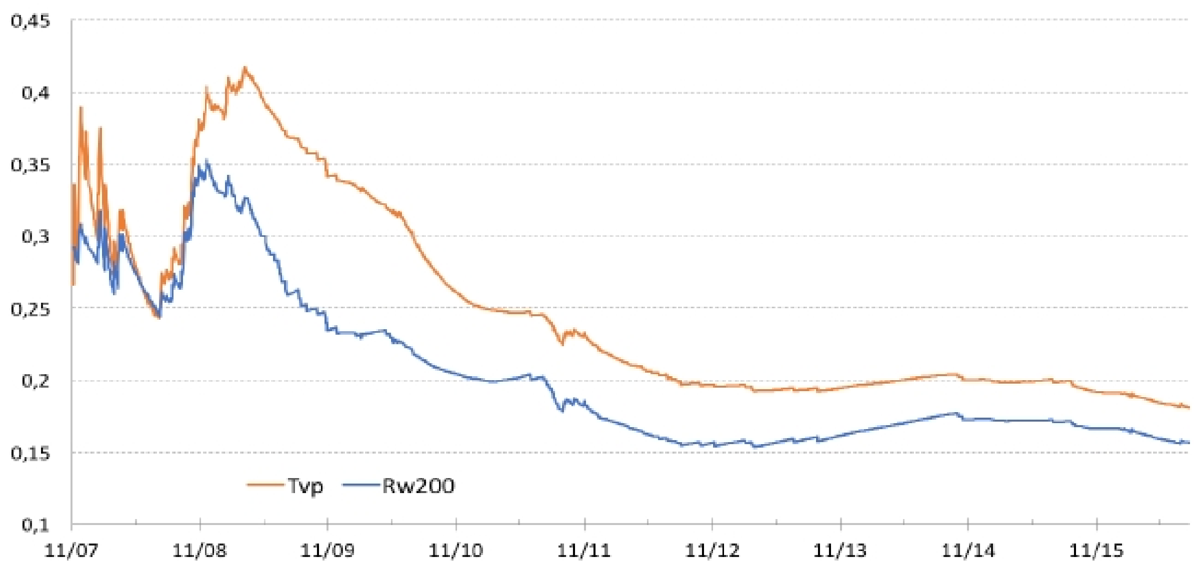


Figure 5: Extreme Event Predictive Performance (Logit; One-day Ahead): TVP vs RW

• A Bayesian VAR can be used. However, with daily data T is typically large, so BVAR estimates $\approx$ VAR estimates.

• Separating samples (and downweighting the first) is possible if we want to give more weght to financial crisis period, to unconventional monetary policy period, etc.

• Split $y_t = [y_{1t}, y_{2t}]$. Estimate A_j using

$$\hat{A}_j = (Y_{2t-j}^T Y_{2t-j} + \kappa_1 Y_{1t-j}^T Y_{1t-j})^{-1}(Y_{2t-j}^T Y_{2t} + \kappa_2 Y_{1t-j}^T Y_{1t}) \quad (19)$$

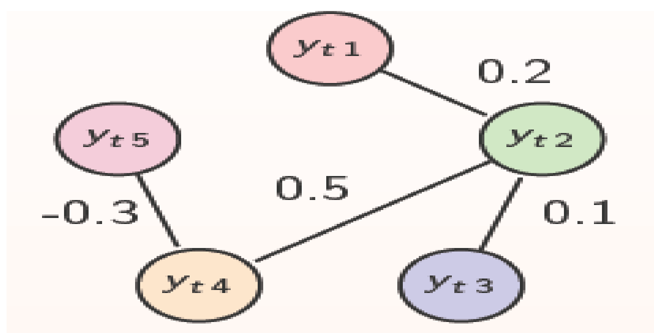
where $\kappa_j < 1, j = 1, 2$ and $\kappa_1 \neq \kappa_2$ (see Hamilton and Baumeister, 2015).

• Alternative to variance decomposition is a historical decomposition: how important is the imported component in explaining deviation from predicted path at t. Useful for real time monitoring.

5 Partial correlation analysis

• An alternative to the variance decomposition to estimate the network connectedness is the partial correlation approach of Dempster (1972), Meinshausen and Buhlmann (2006).

• Suppose we have 5 unit with iid observations y_{it}. The network is an " undirected " graph where (i) the units represent *vertices* , (ii) a link between units (i,j) denotes (significant) partial correlation and (iii) the value reported measures the partial correlation, for example:



- Assume that the network is large but the number of links is limited.

- Find non-zero links and estimate them.

- Linear partial correlation: $\rho^{ij} = corr(y_{it}|y_{jt}, y_{kt}, i \neq j, k)$. Typically measured via the linear regression

$$y_{1t} = c + \beta_{12}y_{2t} + \beta_{13}y_{3t} + \beta_{14}y_{4t} + \ldots + u_t \tag{20}$$

- If $\beta_{13} \neq 0, y_{1t}, y_{3t}$ are partially correlated.

- If units i, j are partially correlated, they are (unconditionally) correlated.

• Limitations of traditional partial correlation analysis:

(i) It needs iid data.

(ii) It does not allow to measure serial dependence (e.g. y_{1t} may be related to $y_{3t-p}, p > 0$ even if it is unrelated to y_{3t}).

• Barigozzi and Brownlees (2018): long run partial correlation. Advantages:

(i) It accounts for non-zero partial correlations at leads and lags.

(ii) It relatively easy to estimate.

• Define $\Sigma_L = \lim_{T\to\infty} \frac{1}{T} cov(\sum_{t=1}^{T} y_t)$ to be the long run covariance matrix of y_t. Let $K = \Sigma_L^{-1}$ be the concentration matrix with typical element $\{k_{ij}\}$. The long run partial correlation is

$$\rho_L^{ij} = -\frac{k_{ij}}{(k_{ii}k_{jj})^{0.5}} \tag{21}$$

• If $\rho_L^{ij} \neq 0$ unit , $i.j$ are linked in the long run.

• K fully characterizes the network properties. Estimation of ρ_L^{ij} is equivalent to estimation of the elements of the K matrix.

• h-period partial correlation. Let $\Sigma_L(h) = \lim_{T\to\infty} \frac{1}{T}$ $cov(\sum_{t=1}^{T} \omega_t y_t)$, where ω_t is a set of (e.g. box, tent, spline) weights different from zero only on a h-sample.

- Estimation:

1) Specify a VAR(q)

$$y_t = \sum_{j=1}^{q} A_j y_{t-j} + e_t, \quad e_t \sim iid(0, \Sigma) \tag{22}$$

2) Compute K_L as

$$K_L = (1 - \sum_{j=1}^{q} A_j)' \Sigma^{-1} (1 - \sum_{j=1}^{q} A_j) \equiv (1 - G)' C (1 - G) \tag{23}$$

- G measures long run predictive system relationships.

- C measures contemporaneous network of the innovations.

- K_L is a combination of long run and contemporaneous networks.

- If dimension of y_t is small, estimation of K_L straightforward.

• If y_t is large, and expect not all elements of A_j to be nonzero, how do estimate G, C? Use a two-step adaptive LASSO approach:

1) Find some preliminary estimate $\tilde{A}_j$ and $\tilde{\Sigma}$ (say, by OLS)

2) Given $\tilde{A}_j$, estimate G via $\hat{G} = \sum_{j=1}^{p} \hat{A}_j$, where $\hat{A}_j$ minimizes $\sum_{t=1}^{T}(y_t - \sum_{j=1}^{q} A_q y_{t-j})^2 + \lambda^G \sum_{j=1}^{q} \frac{|A_j|}{|\tilde{A}_j|}$ and $|A_j|$ is the sum of the absolute value of the components of A_j, and λ^G the Lagrangian of the problem.

3) Construct $\hat{e}_t = y_t - \sum_j \hat{A}_j y_{t-j}$. An estimator $\hat{C}$ of the matrix C is obtained minimizing $\sum_{t=1}^{T} \sum_{i=1}^{N} (\hat{e}_{ti} - \sum_{j \neq i}^{N} \rho_{ij} (\frac{\tilde{c}_{ii}}{\tilde{c}_{jj}})^{0.5} \hat{e}_{tj})^2 + \lambda^C \sum_{i=2}^{N} \sum_{j=1}^{i-1} |\rho^{ij}|$ where $\tilde{c}_{ii} =$diag$\{\tilde{\Sigma}_{ii}^{-1}\}$ is and λ^C is a Lagreangean of the problem.

• Tuning parameters λ^G, λ^C chosen by BIC (via grid search) on a training sample or by leave-one-out techniques.

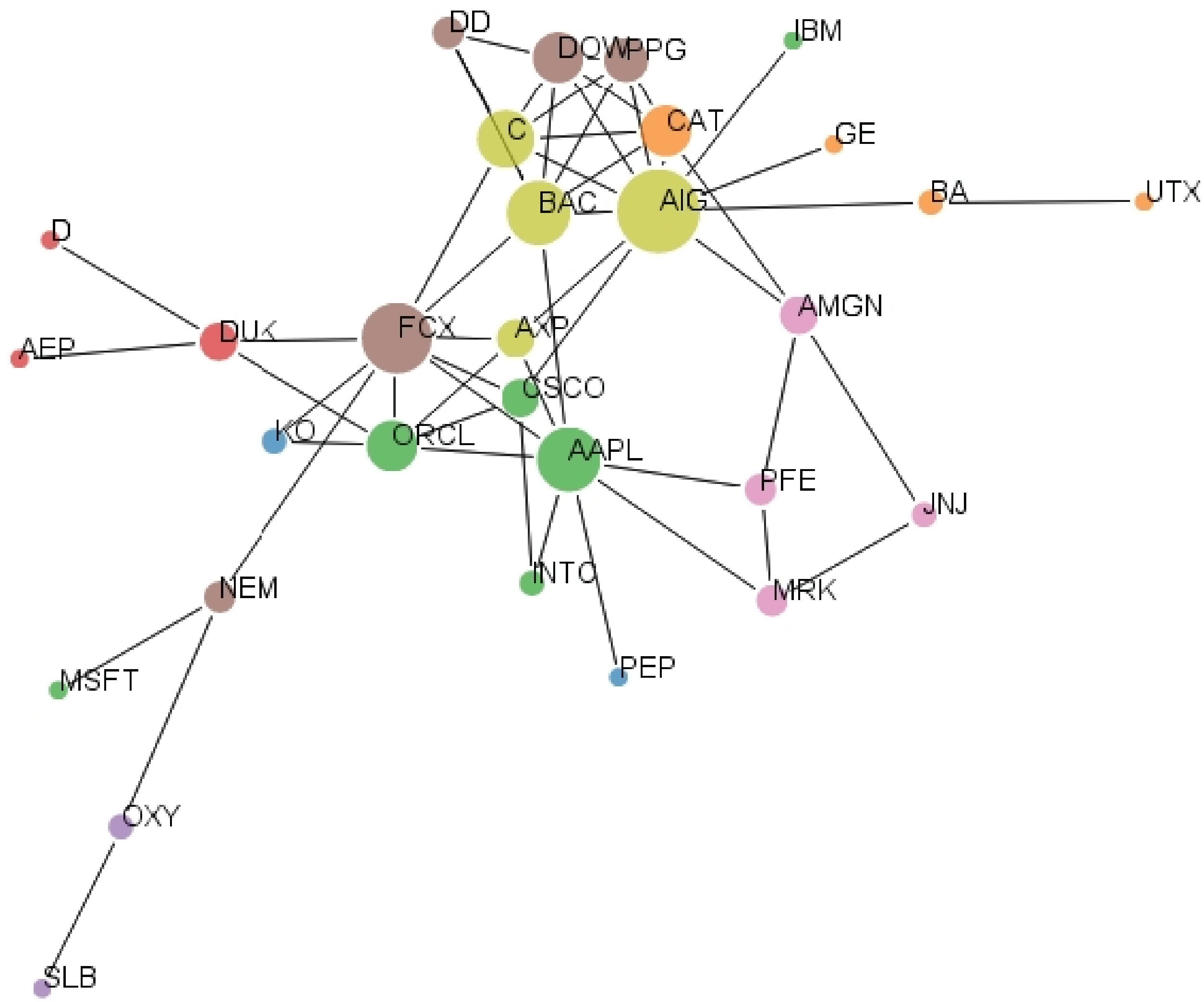
Example 5.1 *US log stock returns for 41 Blue chip industries, 1990:1 to 2010:12. Preliminary run a factor model to eliminate common variations*

$$r_{it} = \beta_i r_{mt} + u_{it} \tag{24}$$

Use $\hat{u}_{1t} = r_{it} - \hat{\beta}_i r_{mt}$ *to construct a network system. Results:*

• *57 out of 820 long run links are significant. 12% are due to dynamic partial correlation, 88% to contemporaneous partial correlation* $\rightarrow$ *lots of contemporaneous spillovers.*

• *The market portfolio explains 25% of the variability of* r_{it}*; network spillovers about 15%.*


DD
DOW
PPG
IBM
C
CAT
GE
BAC
AIG
BA
UTX
D
DUK
FCX
AXP
AMGN
AEP
CSCO
KO
ORCL
AAPL
PFE
JNJ
INTC
MRK
NEM
PEP
MSFT
OXY
SLB

- Partial correlation analysis faces the same problems as VAR analysis as far as lag length, choice of variables, identification and short data sets.

- Long run partial correlation analysis does not take into account (conditional) predictable variations (long run variance decomposition does).

6 Estimating the connectedness of high dimensional SVAR networks.

• When the dimension of the network is large, estimation of an unrestricted SVAR/ partial correlation problematic because the degrees of freedom are small, even if one uses just one lag.

• If N=100, there are 100^2 AR1 and $(100 * 99)/2$ covariance parameters to be estimated. Needs a lot of data!!

• Can use lag selection criteria (AIC,HQ, or BIC) to choose the VAR lag length but, if N is large, this may still be insufficient to estimate the network links (the adjacency matrix).

• How do you deal with dimensionality? Many approaches available.

6.1 Machine learning approaches

• Machine learning is a generic name given to methods which collapse large scale information into simple indicators.

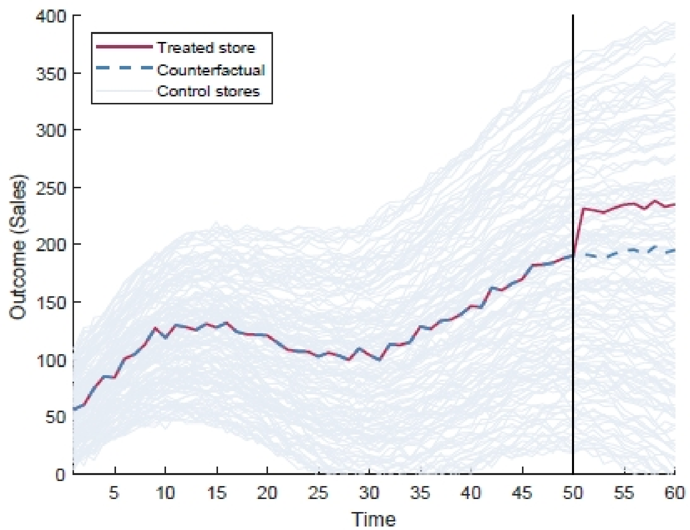
• Typical (prediction) problem: there are many possible predictors, hard to choose one (little guidance from theory); OLS does not work when $J > T$ (collinearity, large standard errors, etc.)

• Idea: use a **weighted** combination of available unit information.

• Can be used micro problems (policy intervention exercises) , in macro setups (see Born et al., 2019) and in finance (measuring network links).

• How do we find weights?

A counterfactual based on Big Data



- Synthetic Control
  - Weighted combination of control units
- Big Data
  - Volume ($J \gg T_0$)
  - Variety
  - Velocity

• Construct a synthetic counterfactual which mimics the relevant data before the break (in-sample). Measure the difference with actual path after the break (out-of- sample).

General problem

- A synthetic control weight is defined by

$$\hat{\omega} = argmin_{\omega \in \Omega}(\frac{1}{T}\sum_{t=1}^{T} \ell(y_t^*, y_t)) \quad (25)$$

where ℓ is a loss function, y_t is a vector and y_t^* is the outcome variable.

- Methods differ in the selection of ℓ and in the choice of Ω.

- OLS: $\ell(.) = (y_t^* - \beta_0 - \beta_1 y_t)^2$; Ω unrestricted. Then: $\hat{\omega} = \beta = (\beta_0, \beta_1)$.

- Prediction: $\hat{y}_{t+h}^* = \ell_{\hat{\omega}}(y_{t+h})$. Counterfactual difference: $y_{t+h}^* - \hat{y}_{t+h}^*$.

Constrained square loss specification

- $\beta_0 = 0$, $0 \leq \omega_i \leq 1, \sum \omega_i = 1$.
- Solves

$$\hat{\omega} = argmin_{\omega \in \Omega}\{\frac{1}{T}(\sum_t (y_t - \sum_{j=1}^{J} \omega_j y_{jt}))^2\} \tag{26}$$

- Replicate y_t using a convex combination of regressors y_{jt}.
- Imposes restrictions on Ω: some ω_j may be set to zero.

Difference -in-difference

- Square loss function; $\omega = 1/J, \forall j$.

- Solves

$$\hat{\beta}_0 = argmin_{\beta_0}\{\frac{1}{T}(\sum_t (y_t - \beta_0 - \frac{1}{J}\sum_{j=1}^{J} y_{jt}))^2\} \tag{27}$$

- For $T = 2$, specification implies parallel trends.

- Compare to GVAR approach (see later).

6.2 Penalized approach (regularization)

• Regularized regression introduced to reduce the consequences of multi-collinearity.

• When variables are highly correlated, a large (positive) coefficient on one variable can be compensated by a large (negative) coefficient in another (think of the first and second lags in an AR(2) model).

• Regularization imposes an upper bound on coefficient values, therefore producing a more parsimonious solution and a set of coefficients with smaller variance.

• There are a number of regularization techniques. The most common are: Ridge and Lasso.

6.2.1 Ridge shrinkage

• A ridge estimator solves:

$$\begin{aligned}\beta^{ridge} &= argmin_{\beta \in R^q}\{\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{j=1}^{q}\beta_j^2\} \\ &= argmin_{\beta \in R^q}(y - X^T\beta)^T(y - X\beta) + \lambda\beta^T\beta \qquad (28)\end{aligned}$$

where $\lambda > 0$ is a Lagrange multiplier of the constraint that β has to be small, $X = X^* - \bar{X}$ is the centered matrix with the stacked lags of Y.

• Solution:

$$\beta^{ridge} = (X^TX + \lambda I)^{-1}X^Ty \qquad (29)$$

• Ridge estimator increases the smallest eigenvalues of the data matrix (X^TX) by a factor λ. Useful because when N is large, (X^TX) is ill-conditioned (close to singular) and can't be inverted.

• Note

$$E(\beta^{ridge}|X) = \beta - \lambda(X^T X + \lambda I)^{-1}\beta \tag{30}$$

$$tr(VAR(\beta^{ridge}|X)) = \sigma^2 \sum_{j=1}^{q} \frac{d_j^2}{d_j^2 + \lambda} \leq \sigma^2 \sum_{j=1}^{q} \frac{1}{d_j^2} = tr(VAR(\beta^{ols}|X)) \tag{31}$$

• β^{ridge} is biased; the bias is proportional to λ; but it has smaller variance than β^{ols}. The MSE of β^{ridge} typically smaller than the MSE of β^{ols}.

• It easy to verify that:

$$\beta^{ridge} = (I_N + \lambda(X'X)^{-1})^{-1}\beta^{ols} \tag{32}$$

When X is orthonormal, (32) becomes

$$\beta^{ridge} = (I_N + \lambda(X'X)^{-1})^{-1}\beta^{ols} = ((1+\lambda)I)^{-1}X^T y = \frac{1}{1+\lambda}\beta^{OLS} \tag{33}$$

• OLS regression coefficients shrunk toward zero as $\lambda \to \infty$ and equal OLS estimates if $\lambda = 0$.

• Letting $X = UDV^T$ (the singular value decomposition), the effective number of estimated parameters is equal to

$$tr(X(X^T X + \lambda I)^{-1} X^T)) = UD(D^2 + \lambda I)^{-1} DU^T = \sum_{j=1}^{q} \frac{d_j^2}{d_j^2 + \lambda} \tag{34}$$

where d_j is the j-th diagonal element of D.

• When $\lambda \to 0$ the number of effective parameters is q; when $\lambda \to \infty$ the effective number of parameters is zero. **Regularization reduces the number of effective number of parameters to be estimated**.

• Choice of λ arbitrary. Typical $\lambda = 0.01 - 0.02$. λ can also be selected to minimize some criterion function (e.g. AIC or an other information criteria) or by cross validation (leave-one-out) techniques.

6.2.2 Lasso shrinkage

• A Lasso estimator solves:

$$\begin{aligned}\beta^{lasso} &= argmin_{\beta \in R^q}\{\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda \sum_{j=1}^{q} |\beta_j|\} \\ &= argmin_{\beta \in R^q}(y - X\beta)^T(y - X\beta) + \lambda|\beta| \qquad (35)\end{aligned}$$

(see Tibshirani, 1996).

• No closed form solution for the estimator. (35) defines a non-linear optimization problem. Solution obtained by numerical methods.

• Ridge imposes a "soft" threshold on the coefficients. Lasso imposes a "harder" threshold, in the sense that more coefficients are set to zero.

6.2.3 General Shrinkage estimators

- General optimization problem:

$$\begin{aligned}
&\min_\beta \sum_t (y_t - \sum_i x_{it}^T \beta_i)^2, \quad i = 1, ..K \\
&\text{subject to} \sum_i |\beta_i|^m \leq c
\end{aligned} \tag{36}$$

where c is an (arbitrary) constant, x_{it} includes the lags of the endogenous variables y_{it}. Letting λ is a Lagrange multiplier, the problem becomes:

$$\min_\beta (\sum_t (y_t - \sum_i \beta_i x_{it})^2 + \lambda \sum_i |\beta_i|^m) \tag{37}$$

- Lasso is the solution to (37) for m=1 (absolute penalty); Ridge is the solution when m=2 (quadratic penalty).

- Elastic-net extension

$$\min_{\beta}\Big(\sum_t \big(y_t - \sum_i x_{it}^T \beta_i\big)^2 + \lambda \sum_i \big(0.5|\beta_i| + 0.5\beta_i^2\big)\Big) \tag{38}$$

where λ is selected equation by equation (Zou and Zhang, 2009). Elastic-net estimator combines a Lasso and a Ridge penalty. Useful because it has the property that the selected $\beta_i \neq 0$ makes the model closest in a Kullback-Leibler sense to the true DGP.

- Adaptive Lasso and Adaptive elastic-net.

$$\begin{aligned} |\beta_i| &\rightarrow \eta_i|\beta_i|,\ \eta_i = (\beta_i^{lasso})^{-1} \\ 0.5|\beta_i| + 0.5\beta_i^2 &\rightarrow \eta_i^*(0.5|\beta_i| + 0.5\beta_i^2),\ \eta_i^* = (\beta_i^{ols})^{-1} \end{aligned} \tag{39}$$

Example 6.1 *Bonaldi et al. (2013) use an adaptive elastic-net approach to estimate the funding costs network of banks in the euro area.*

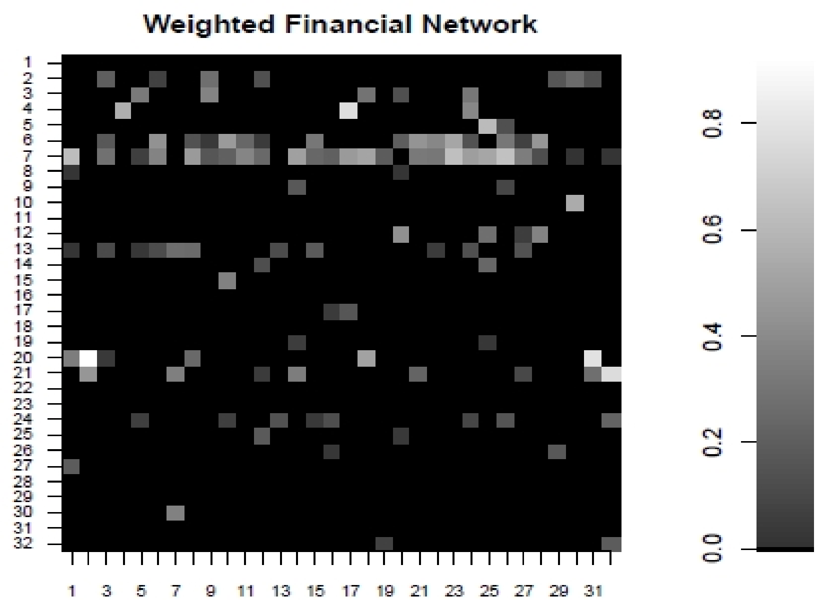


Figure 3: Elastic Net Coefficients: (1st order) Effect of a Shock to Row Bank's (lagged) Funding Costs on Column Bank's Funding Costs.

Table 5: Total effect of a 100bp shock to each of the top 5 systemic banks on all banks (Country means reported)

	AT	BE	DE	DK	ES	FR	GR	IE	IT	JP	NL	UK
ES1	0.8937	1.3736	0.9709	0.5674	0.8817	0.9537	1.5968	1.1110	0.4809	0.8688	0.8830	1.1407
ES2	1.1032	1.0339	0.8771	0.3494	1.0523	0.8919	1.0288	1.1240	0.5164	0.8336	0.7824	1.4665
UK	0.5315	0.7345	0.6390	0.2723	0.6793	0.6047	0.9229	0.8541	0.4016	0.6100	0.4762	0.7018
FR	0.4207	0.5801	0.4669	0.2368	0.6058	0.4688	0.6491	0.6228	0.2421	0.4080	0.4028	0.5633
FR	0.2874	0.4417	0.3122	0.1825	0.4443	0.3067	0.5135	0.3573	0.1547	0.2794	0.2840	0.3668

[a] Averages across banks within a country reported.

[b] The total effect is computed using the Katz centrality measure with $\alpha = 0.9$.

Note: first column has banks in Spain (ES1, ES2), UK, and France

• Regularization approaches are flexible (the prediction function non-linear). Same class as HP or exponential smoothing filters.

• They handle dimensionality problems via choice of λ.

• β^m may be sparse but the covariance matrix of the VAR shocks is still of large dimension. **Shrinkage estimators are not enough when computing adjacency matrix of a large scale SVAR.**

6.2.4 Structural time series approaches

- β_0 is allowed to be time varying; Ω unrestricted.

- Model

$$\begin{aligned} y_t^* &= \beta_{0t} + \omega^T y_t + e_t^* \\ \beta_{0t} &= \beta_{0t-1} + \eta_t + u_{1t} \\ \eta_t &= \eta_{t-1} + u_{2t} \end{aligned} \tag{40}$$

- Local linear trend model (Harvey, 1989); long run risk model (Bansal and Yaron, 2004).

- Flexible setup. Dimensionality is handled with restrictions on covariances of u_{1t}, u_{2t}.

- Classical or Bayesian estimation is possible. Recursive predictions possible, see e.g. Kinn (2017).

6.3 A Bayesian approach

• Bayesian approach provides a coherent framework to think about penalized regressions/general shrinkage setups/structural time series approaches.

• The terms appearing in the estimators arising in addition of those present in OLS can be obtained with a particular set of prior restrictions.

• Can handle dimensionality and time variations.

• Can produce classical estimators under certain conditions.

- Assume a prior is of the form: $\beta = \bar{\beta} + v,\ v \sim N(0, \Sigma_b)$, where $\bar{\beta}, \Sigma_b$ are known (or estimable in a training sample). Then:

$$g(\beta) \propto |\Sigma_b|^{-0.5} exp[-0.5(\beta - \bar{\beta})'\Sigma_b^{-1}(\beta - \bar{\beta})] \tag{41}$$

- Let the VAR be written in SES form as (see appendix)

$$y_t = (I_m \otimes x_t)\alpha + e_t \equiv X_t\alpha + e_t \tag{42}$$

where $e_t \sim (0, \Sigma_e \otimes I)$ and suppose Σ_e is fixed.

- Combining (41) with the likelihood of (42) we obtain:

$$\begin{aligned} g(\beta|y) &\propto g(\beta)L(\beta|y) = \exp\{-0.5(\beta - \tilde{\beta})'Z'Z(\beta - \tilde{\beta})\} \\ &= \exp\{-0.5(\beta - \tilde{\beta})'\tilde{\Sigma}_b^{-1}(\beta - \tilde{\beta})\} \end{aligned} \tag{43}$$

where $z = [\Sigma_b^{-0.5}\bar{\beta}, (\Sigma_e^{-0.5} \otimes I_T)y]'$; $Z = [\Sigma_b^{-0.5}, (\Sigma_e^{-0.5} \otimes X)]'$ and

$$\tilde{\beta} = (Z'Z)^{-1}(Z'z) = [\Sigma_b^{-1} + (\Sigma_e^{-1} \otimes X'X)]^{-1}[\Sigma_b^{-1}\bar{\beta} + (\Sigma_e^{-1} \otimes X)'y]$$

Conclusion: $g(\beta|y)$ (the posterior) is $N(\tilde{\beta}, \tilde{\mathbf{\Sigma}}_b)$ where

- $\tilde{\beta} = [\mathbf{\Sigma}_b^{-1} + (\mathbf{\Sigma}_e^{-1} \otimes X'X)]^{-1}[\mathbf{\Sigma}_b^{-1}\bar{\beta} + (\mathbf{\Sigma}_e^{-1} \otimes X)'y]$

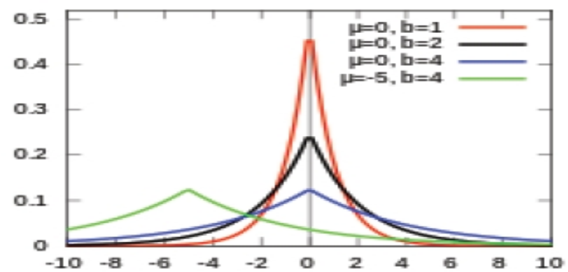
- $\tilde{\mathbf{\Sigma}}_b = [\mathbf{\Sigma}_b^{-1} + (\mathbf{\Sigma}_e^{-1} \otimes X'X)]^{-1}$.

- Linear combination of prior and sample information. Weight given by relation precision of information (recall conjugate prior analysis).

- Construction of $\tilde{\beta}, \tilde{\mathbf{\Sigma}}_b$ requires knowledge of $\mathbf{\Sigma}_e$. If $\mathbf{\Sigma}_e$ is unknown, use $\hat{\mathbf{\Sigma}}_e = \frac{1}{T-1}\hat{e}'\hat{e}$ in formulas, where $\hat{e}_t = y_t - (I \otimes X)\beta_{ols}$.

- Ridge estimator obtained setting $\bar{\beta} = 0, \mathbf{\Sigma}_b = \lambda * \mathbf{\Sigma}_e = \lambda * \text{diag}\{\sigma_i^2\}$.

• Lasso estimator obtained using the same set up but assuming a Laplace prior for β. Laplace distribution strongly peaked at zero (i.e. Lasso sets more coeffficients to zero than Ridge).



Laplace distributions are sharply peaked at their mean with more probability density concentrated there compared to a normal distribution.

• Without further restrictions the dimensionality of $(\beta, \Sigma_b, \Sigma_e)$ may still a problem in large networks. Need additional restrictions.

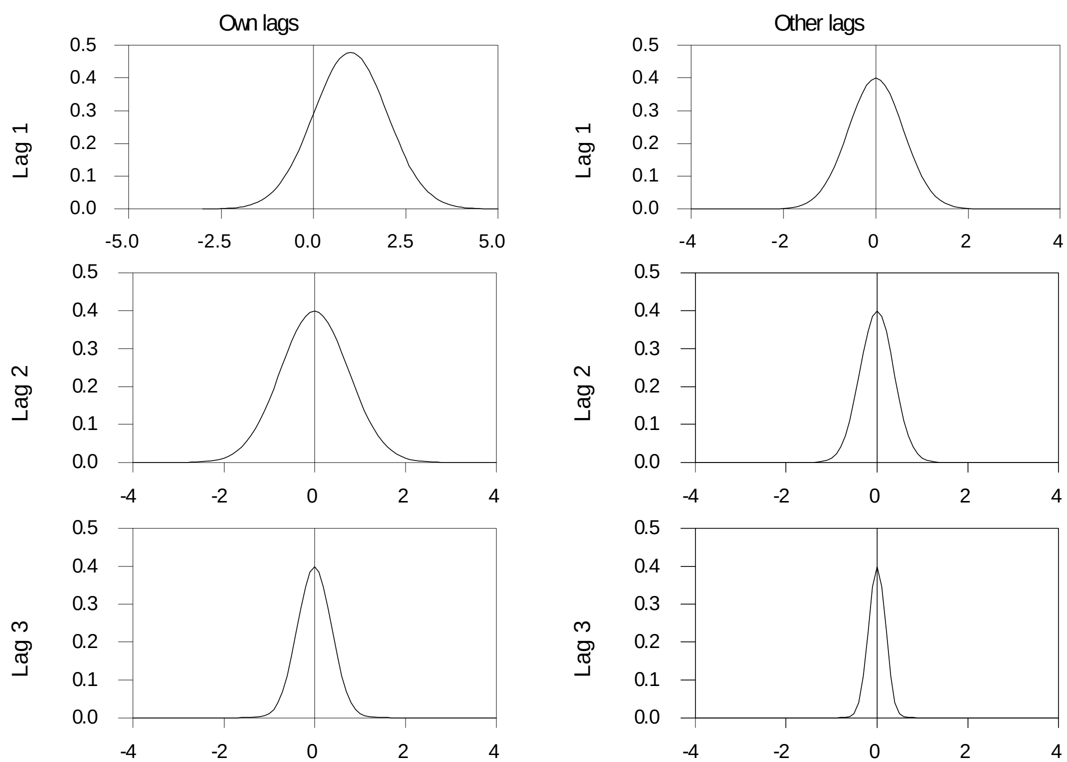
6.3.1 Litterman (Minnesota) structure

Here $\bar{\beta}, \Sigma_b$ have special structure:

- $\bar{\beta} = 0$ except $\bar{\beta}_{(j=i,\ell=1)} = 1$;
- $\Sigma_b = diag(\sigma(\phi))$ where

$$
\begin{aligned}
\sigma_{ij,\ell} &= \frac{\phi_1}{h(\ell)} \ \text{ if } \ i = j \\
&= \phi_1 \frac{\phi_2}{h(\ell)} * (\frac{\sigma_j}{\sigma_i})^2 \ \text{ otherwise} \\
&= \phi_1 * \phi_4 * (\frac{\sigma_j}{\sigma_i})^2 \ \text{ for exogenous variables}
\end{aligned} \tag{44}
$$

$\phi_1, \phi_2 \phi_4$ are parameter; $h(\ell)$ decay function (e.g. harmonic decay $h(\ell) = \ell^{(2*\phi_3)}$; geometric decay $h(\ell) = \phi_3^{-\ell+1}$; linear decay $h(\ell) = \ell$); $(\frac{\sigma_j}{\sigma_i})^2$scaling factor; i= variable, j=equation, ℓ = lag.


Own lags
Other lags
Lag 1
Lag 2
Lag 3
0.5
0.4
0.3
0.2
0.1
0.0
-5.0
-2.5
0.0
2.5
5.0
-4
-2
0
2
4

Logic for this (shrinkage) prior:

• Mean chosen so that the VAR is N a-priori random walks.

• Decrease dimensionality of covariance matrix by setting $\Sigma_b = \Sigma_b(\phi)$ where $\Sigma_b(\phi)$ is diagonal (no relationship among equations and coefficients);

- ϕ_1 measures the relative importance of the prior to the data.

- The variance of lags of LHS variables shrinks to zero as lags increase ($\phi_3 < 1$). Variance of lags of other RHS variables shrinks to zero at a different rate ($\phi_2 \leq 1$ is relative importance of other variables).

- Variance of the exogenous variables is regulated by ϕ_4. If ϕ_4 is large, prior information on the exogenous variables diffuse.

• **Litterman prior imposes no dynamic network links and restricted contemporanous links!**.

• How do we choose $\phi = (\phi_1,\ \phi_2, \ldots)$.

1) Rules of thumb. Typical default values: $\phi_1 = 0.2\hat{}2,\ \phi_2 = 0.5\hat{}2, \phi_4 = 10^5$, an harmonic specification for $h(\ell)$ with $\phi_3 = 1$ or 2 and uninformative prior for the exogenous variables.

2) Estimate ϕ using ML-II approach. That is, maximize $\mathcal{L}(\phi|y) = \int f(\beta|y, \phi)g(\beta|\phi)$ on training sample. where $f(.)$ is the likelihood of the model.

• The scaling factors $(\frac{\sigma_i}{\sigma_j})^2$ are typically estimated from the data (in a training sample).

Example 6.2 *(ML-II approach) Suppose* $y_t = Bx_t + e_t$, $e_t \sim N(0, \sigma_e^2)$, σ_e^2 *known and let* $B = \bar{B} + \upsilon$, *where* $\upsilon \sim N(0, \Sigma_\upsilon^2)$, $\bar{B}$ *fixed and* $\Sigma_\upsilon^2 = f(\phi)^2$, *where* ϕ *is a set of hyperparameters.*

Then $y_t = \bar{B}x_t + \epsilon_t$ *where* $\epsilon_t = e_t + \upsilon x_t$ *and posterior kernel is:*

$$\grave{g}(\phi|y) = \frac{1}{(2\pi f(\phi)^2 tr|X'X| + \sigma_e^2)^{0.5}} \exp\{-0.5 \frac{(y - \bar{B}x)^2}{\sigma_e^2 + f(\phi)^2 tr|X'X|}\} \tag{45}$$

• *Could maximize (45) directly or compute prediction error decomposition of* $\grave{g}(\phi|y)$ *with the Kalman filter and then find modal estimates of* ϕ*.*

- A few applications of ML-II approach: Giannone, Primiceri, Lenza (2015); Carriero, Kapetanios, Marcellino (2014).

• Large scale (Litterman) prior: Banburra et al. (2010). The degree of shrinkage (ϕ_1) is function of the number of variables in the system. Thus, the larger is the system, the smaller is ϕ_1. Good for forecasting.

- With Litterman prior the size of Σ_b is reduced. There are other tricks to reduce the dimensionality of Σ_b, e.g. Kronecker structure (see Canova and Ciccarelli, 2009) or a shrinkage prior based on distance of units, etc.

6.3.2 Hierarchical priors

• Up to now the hyperparameters ϕ are treated as fixed. When ϕ are treated as random, computations become more involved.

• No closed form solution for the posterior shape is available; no closed form solution for the moments of the posterior exists.

• Need to use MCMC (Gibbs, Metropolis-Gibbs) samplers to draw sequences from the posterior.

• Are there significant gains from using random prior parameters (relative to empirical-based or rules of thumb choices)? Not much is known, see Carriero et al., (2014), Giannone et al., (2015).

A Hierarchical Bayesian VARs

$$\begin{aligned} y_t &= X_t\beta + e_t \quad e_t \sim N(0,\Sigma) & (46)\\ \beta &= M_0\theta + v \quad v \sim N(0, D_0) & (47)\\ \theta &= M_1\mu + \zeta \quad \zeta \sim N(0, D_1) & (48) \end{aligned}$$

Assume M_0, M_1, D_1 known; $X_t = (I \otimes x_t)$.

Priors: $p(\Sigma) \sim iW(\bar{S}, s)$; $p(D_0) \sim iW(\bar{D}_0, \rho)$; $p(\mu) \propto 1$.

Conditional Posteriors:

1) $(\beta|\psi_{-\beta}, Y, X) \sim N(\tilde{\beta}, \tilde{\Omega})$.

2) $(\Sigma|\psi_{-\Sigma}, Y, X) \sim iW(\tilde{\Sigma}, s+T)$

3) $(\theta|\psi_{-\theta}, Y, X) \sim N(\tilde{D}_1(D_1^{-1}M_1\mu + M_0'D_0^{-1}\beta), \tilde{D}_1)$

4) $(D_0|\psi_{-D_0}, Y, X) \sim iW(\tilde{D}_0, \rho + 1)$
5)$(\mu|\psi_{-\mu}, Y, X) \sim N(\hat{\mu}, \mathbf{\Sigma}_\mu)$

where

$\tilde{\Omega} = (D_0^{-1} + \sum_t X_t' \mathbf{\Sigma}^{-1} X_t)^{-1}$;
$\tilde{\beta} = \tilde{\Omega}(D_0^{-1} M_0 \theta + \sum_t X_t' \mathbf{\Sigma}^{-1} y_t)$;
$\tilde{\mathbf{\Sigma}}^{-1} = \bar{S} + \sum_t (Y_t - X_t\beta)(y_t - X_t\beta)'$;
$\tilde{D}_1 = (D_1^{-1} + M_0' D_0^{-1} M_0)^{-1}$;
$\tilde{D_0}^{-1} = D_0^{-1} + \sum_{g=1}^{M} (\beta_g - \theta)(\beta_g - \theta)'$
$\hat{\mu} = (M_1' M_1)^{-1} (M_1 \theta)$
$\mathbf{\Sigma}_\mu = (\theta - \hat{\mu} M_1)'(\theta - \hat{\mu} M_1)$

Use these five conditional posteriors in a Gibbs sampler.

Another useful hierarchical VAR:

$$\begin{aligned} y_t &= (I \otimes x_t)\beta + e_t \quad e_t \sim N(0, \mathbf{\Sigma}) &(49)\\ \beta &= \bar{\beta} + v \quad v \sim N(0, \mathbf{\Sigma} \otimes \Omega * \zeta) &(50)\\ \zeta &= \bar{\zeta} + \epsilon \quad \epsilon \sim N(0, \eta) &(51) \end{aligned}$$

where $(\bar{\beta},\ \Omega, \bar{\zeta}, \eta)$ are known (or estimable).

• Compute the joint posterior of $(\beta, \zeta, \mathbf{\Sigma})$.

• Interest is in $g(\zeta|y, X, y_p) = \int g(\zeta, \beta, \mathbf{\Sigma}|y, X, y_p)d\beta d\mathbf{\Sigma}$, where y_p are the initial conditions.

• For one example where $g(\zeta|y, X, y_p)$ is analytically available see Canova (2007, chapter 9).

Example 6.3 *Moratis and Sakellaris (2021): Use 77 banks from 26 countries issuing CDS. Sample: 2008:1-2017:6. Estimate the VAR*

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + BX_t + e_t \tag{52}$$

where Y_t is the log differences of daily CDS spreads, X_t are exogenous factors affecting all spreads (VIX, 3mo US Tbills, etc.) plus country specific and bank specific controls.

- *Use Pesaran-Shin approach to identify shocks.*

- *Rank banks based on their connectedness: sum of systemicness score (effect on the others of a bank specific shock) and vulnerability score (effect of shocks of all other banks on that bank).*

Table 7: Individual systemic importance

Panel A – Ranked by total score

Rank by score	Rank by bank assets	Bank Name	Home-Country	Region	Assets (billion US $)	Score	From others (Aggr.)	To others (Aggr.)
1	27	Intesa Sanpaolo	Italy	Europe	766	1.056	0.505	0.551
2	16	Banco Santander	Spain	Europe	1416	1.038	0.464	0.574
3	26	BBVA	Spain	Europe	773	0.987	0.454	0.533
4	5	BNP Paribas	France	Europe	2196	0.982	0.445	0.537
5	21	Unicredit S.p.A.	Italy	Europe	908	0.975	0.489	0.485
6	14	Barclays PLC	UK	Europe	1490	0.961	0.454	0.507
7	12	Deutsche Bank	Germany	Europe	1682	0.954	0.428	0.526
8	9	Credit Agricole Group	France	Europe	1821	0.942	0.424	0.518
9	15	Societe Generale	France	Europe	1461	0.940	0.427	0.512
10	72	Credit Lyonnais	France	Europe	120	0.938	0.445	0.493
11	17	Lloyds Banking Group	UK	Europe	1004	0.925	0.438	0.487
12	25	Credit Suisse Group	Switz.	Europe	806	0.907	0.376	0.531
13	35	Commerzbank	Germany	Europe	549	0.904	0.410	0.494
14	20	UBS Group AG	Switz.	Europe	919	0.890	0.387	0.503
15	32	Standard Chartered Plc	UK	Europe	646	0.870	0.419	0.451
16	68	Banca Monte dei Paschi	Italy	Europe	161	0.861	0.429	0.433
17	19	Royal Bank of Scotland Group	UK	Europe	981	0.845	0.423	0.423
18	29	Rabobank Group	Netherl.	Europe	700	0.841	0.376	0.465
19	22	ING Groep NV	Netherl.	Europe	893	0.823	0.374	0.449
20	76	Mediobanca	Italy	Europe	95	0.821	0.389	0.433

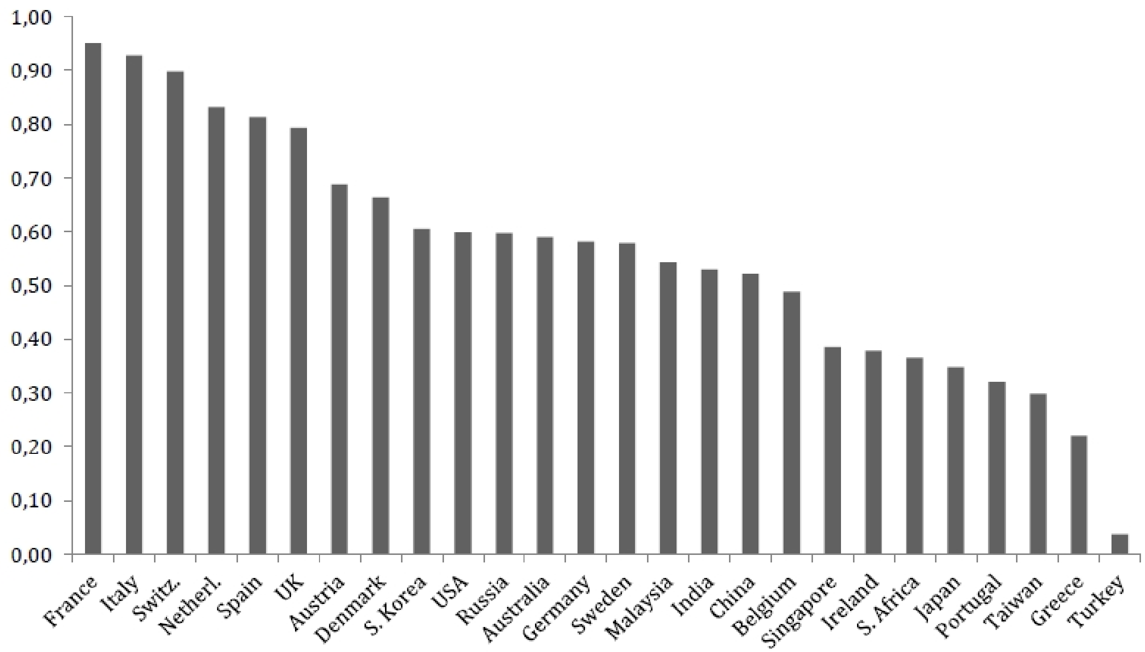
Table 10 shows that banks with higher systemic importance tend to have higher externalities ratios. In particular, banks in the first quartile have average individual directionality of 54%, whereas banks in the last quartile average 30%.

Table 10: Directionality of Bank Systemic Importance

Quartile	*Directionality*
1st	54%
2nd	53%
3rd	47%
4th	30%

Note: Table presents the average directionality for banks in each quartile, when ranked by systemic importance. Directionality is the ratio between the *individual externalities* and the *total individual contagion.*

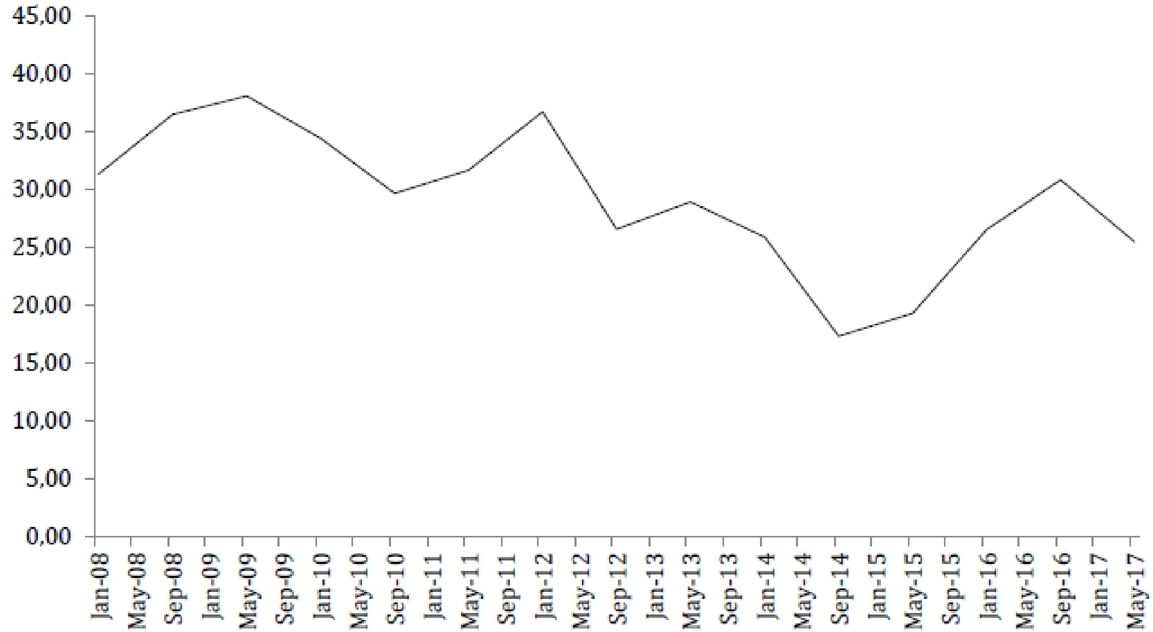
Figure 2: Average systemic risk per bank – Own shocks are excluded



Note: Results cover the period January 2008-June 2017.

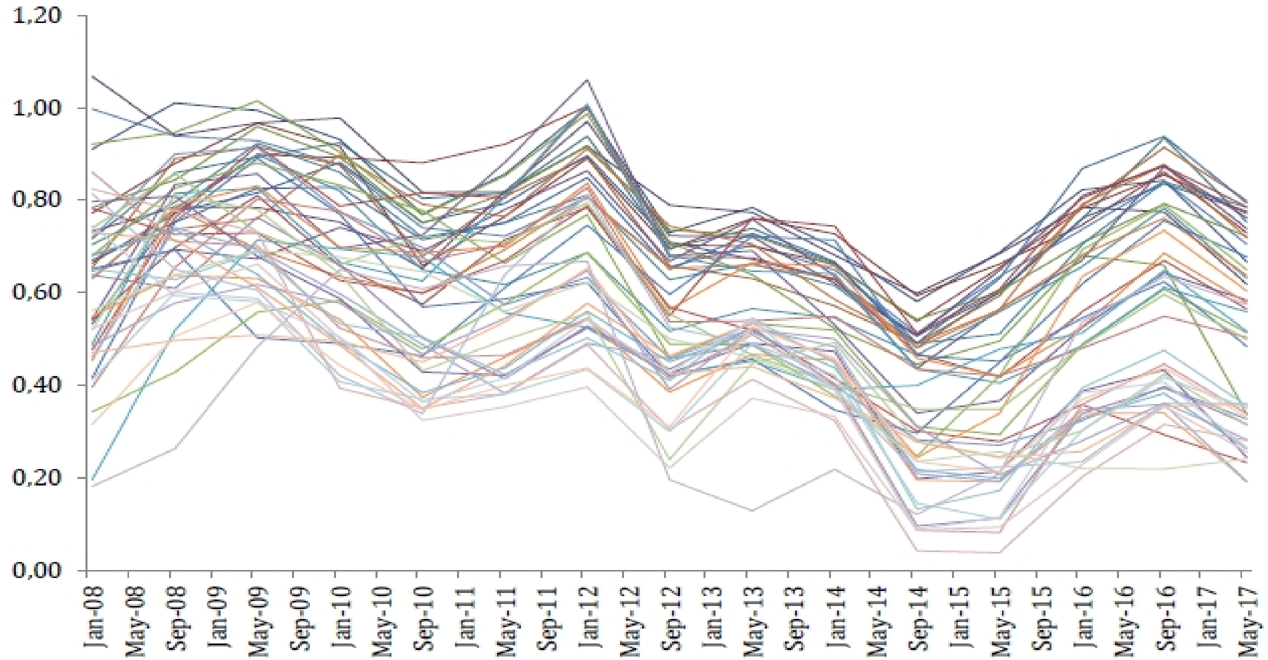
Systemic Risk(i)=$\sum_{j=1}^{N}(vulnerability_{j,i} + systemicness_{j,i})/N$

Figure 5: Total Systemic Risk Index (TSRI)



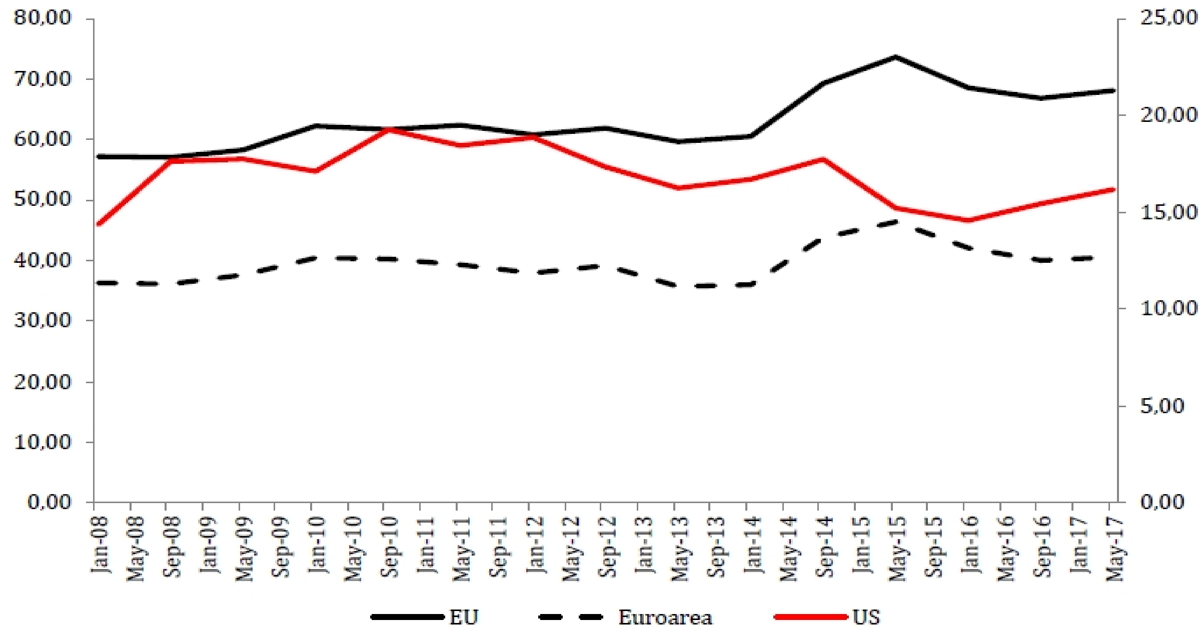
Note: TSRI is defined as the average response per bank in the connectedness matrix. The length of the window is 340 days and the step is 150 days.

Figure 3: Evolution of Bank Systemic Importance



Note: Each line depicts the total systemic risk *(externalities + vulnerability)* for each bank in the sample. The length of the rolling window is 340 days and the step is 150 days.

Figure 8: European, US and Euro-area banks' contribution to total systemic risk



Note: Figure presents the ratio of aggregated systemic importance scores of banks in Europe, Eurozone and the U.S. over those in the total sample. US banks' contribution is measured on the right axis.

Example 6.4 *Use a BVAR with daily data on 2 and 10 years bond yields for DE, FR, IT, ES for 2009d1-2012d1 and US 2 and 10 years yields as exogenous. Use 2 lags and a Minnesota prior with overall tightness at 0.1. Cholesky shocks. The bordered spillover matrix at 15 days horizon is*

$$\begin{array}{cc}
 & \begin{array}{ccccccccc} DE2 & DE10 & FR2 & FR10 & IT2 & IT10 & ES2 & ES10 & \text{To} \end{array} \\
\begin{array}{c} DE2 \\ DE10 \\ FR2 \\ FR10 \\ IT2 \\ IT10 \\ ES2 \\ ES10 \\ \text{From} \end{array} &
\left[\begin{array}{ccccccccc}
86 & 04 & 03 & 00 & 01 & 00 & 01 & 01 & 14 \\
00 & 82 & 06 & 00 & 00 & 01 & 04 & 01 & 18 \\
05 & 04 & 84 & 00 & 00 & 00 & 00 & 01 & 16 \\
02 & 10 & 00 & 81 & 01 & 01 & 01 & 01 & 19 \\
00 & 02 & 00 & 00 & 90 & 00 & 01 & 01 & 10 \\
00 & 01 & 00 & 00 & 00 & 94 & 01 & 01 & 6 \\
00 & 02 & 00 & 00 & 02 & 02 & 83 & 05 & 17 \\
00 & 04 & 00 & 00 & 01 & 00 & 02 & 88 & 12 \\
07 & 23 & 09 & 00 & 05 & 04 & 10 & 11 & 181
\end{array}\right]
\end{array}$$

$S(15) = 181/381 = 20.82\%$; *average spillovers* $20.82/16 = 1.30\%$.

6.3.3 Bayesian Lasso

• Instead of a Litterman prior one could cut down dimensionality of parameter space with Bayesian Lasso plus restrictions on Σ_b.

• The prior on (β, σ) is (see Park and Casella, 2008)

$$\begin{aligned} g(\beta|\sigma) &= \prod_{j=1}^{p} \frac{\lambda}{2\sigma} \exp(\frac{-\lambda|\beta_j|}{\sigma}) \\ g(\sigma) &= \frac{1}{\sigma} \end{aligned} \tag{53}$$

• A Laplace distribution for β approximated with a mixture of normal with exponential mixing.

$$\frac{a}{2}\exp(-a|z|) = [\int_0^\infty \frac{1}{(2\pi s)^{0.5}}\exp(\frac{-z^2}{2s})][\frac{a^2}{2}\exp(-0.5a^2 s)]ds, \quad a > 0 \tag{54}$$

• A SVAR with Laplace prior on β can be estimated using the following hierarchical model ($\tilde{Y} = Y - \mu 1_N, \Sigma_e = \sigma^2 * I, \mu$ fixed and known).

$$\begin{aligned}
(\tilde{Y}|X,\beta,\sigma^2) &\sim N(X\beta,\sigma^2 I_N) \\
(\beta|\sigma^2,\tau_1^2,\ldots,\tau_p^2) &\sim N(0,\sigma^2 diag(\tau_1^2,\ldots,\tau_p^2)) \\
(\sigma^2,\tau_1^2,\ldots,\tau_p^2) &\sim \pi(\sigma^2)d\sigma^2\frac{1}{\sigma^2}\prod_{j=1}^{p}0.5\lambda^2\exp(\frac{-\lambda^2\tau_j^2}{2})d\tau_j^2
\end{aligned} \tag{55}$$

where $\sigma^2, \tau_1^2, \ldots, \tau_p^2 > 0$.

• λ is typically treated as fixed and selected to maximize the marginal likelihood of the model (in a training sample). Alternatively, if the prior is diffuse, the posterior of β obtained by integrating λ from the joint of (λ, β).

• General penalized Bayesian analysis (m fixed). The prior is:

$$\begin{aligned} g(\beta, \Sigma, \lambda) &= g(\beta|\sigma, \lambda) g(\Sigma|\lambda) g(\lambda) \\ &\propto \prod_{j=1}^{p} \exp(-\lambda(\frac{|\beta_j|^m}{\sigma})) \frac{1}{\sigma} g(\lambda) \end{aligned} \tag{56}$$

• Model with Laplace priors needs to be estimated with a Metropolis step within Gibbs sampler.

• Estimators obtained with Bayesian Lasso are smoother than classical Lasso estimators and tougher in setting coefficient to zero than classical ridge estimators (see Park and Casella, 2008)

6.4 Data grouping

• Another alternative to reduce the dimensionality of the VAR is to group units first and estimate the links across (homogeneous) groups.

• If there are N units, we estimate the links across M "homogenous" groups ($M < N$) pooling the data within each group. i.e. we run a VAR(q) with M units.

• How do you find "homogenous" groups? A-priori: group units in the same industry, in the same country, of the same size, etc. Endogeneously, see Canova (2004).

Example 6.5 *You have 5 units and you need to find two groups. Suppose that you know that units 1-2 are in group 1 and units 4-5 are in group 2.*

1) How would you check if unit 3 goes to group 1 or 2?

- *Null: group 1=[1,2,3] and group 2=[4,5].*

- *Alternative: group 1=[1,2] and group 2=[3,4,5].*

- *Estimate the model under the null and the alternative. Choose the (grouping) with the highest (marginal) likelihood.*

2) How do you assign 5 units to two groups without any prior knowledge?

- *Potential groupings are [1]-[2,3,4,5]; [1,2,3,4]-[5],[1,2,3,5]-[4];[1,2,4,5]-[3];[1,3,4,5]-[2]. [1,2]-[3,4,5]; [1,3]-[2,4,5];[1,4]-[2,3,5]; [1,5]-[2,3,4];[2,4]-[1,3,5];[2,3]-[1,4,5];[2,5]-[1,3,4]; [1,2,4]-[3,5]; [1,2,5]-[3,4]; [1,2,3]-[4,5];*

- *Groupings invariant to relabelling of groups or reordering of units within groups.*

- *Compare 15 (marginal) likelihood. Pick the one with the highest value.*

3) What if you do not know the ordering (label) of the units?

- *Repeat 2) for all possible reordering of units (5!=120)*

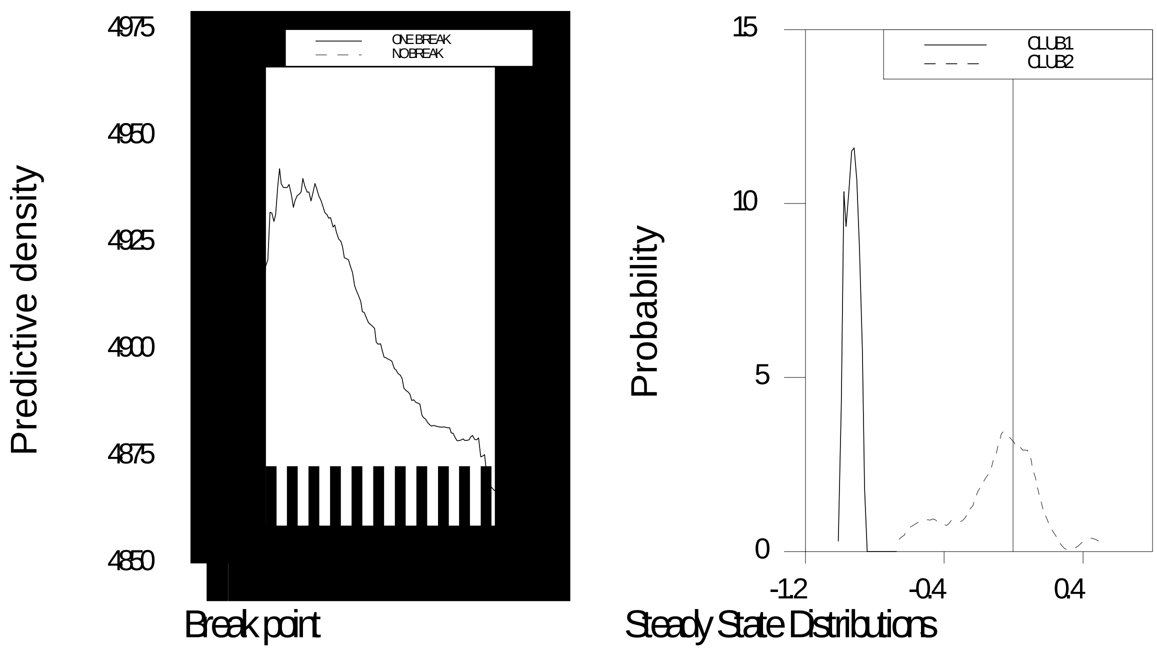
4) What if you do not know how many groups exist?

- *Assume a maximum number of groups.*

- *Repeat 2)-3) for 2,3,4,...G groups. Pick the grouping with the highest (marginal) likelihood.*

- Approach straightforward. Computationally burdensome.

- Can avoid step 4) with some economic argument. Can avoid step 3) if some ordering are more natural than others (e.g. based on asset classes, capital ratios, etc.)



• Alternatives: Bonhomme and Manresa (2015); Brownlee and Gudmundsson (2021)

6.5 Shrinkage Panel VAR

• If you have more than one variable in each unit, the VAR becomes a panel VAR.

• Panel VARs with interdependencies across units and heterogenous dynamics can not be estimated unrestrictedly because the number of parameters is greater than T. What do you do?

• Impose some a-priori restriction (e.g. Global VARs).

• Change setup (e.g. use a factor model).

• Use partial pooling techniques (Canova and Ciccarelli, 2004, 2009, 2013).

- Model

$$y_{it} = D_i(\ell)Y_{t-1} + F_i(\ell)\, W_{t-1} + e_{it} \tag{57}$$

$i = 1, ..., N$, y_{it} is $G \times 1$ vector each i, $Y_t = \left(y'_{1t}, \ldots y'_{Nt}\right)'$.

- W_t are exogenous variables.

- $D_i(\ell)$ has p lags; $F_i(\ell)$ has q lags.

- $D_i(\ell)$ and $F_i(\ell)$ specific to each cross sectional unit.

- Allow for general form of lagged and contemporaneous interdependencies.

• Each equation has $k = NGp + Mq$ coefficients, and $r = NG$ equations: T is always smaller than $k \times r$.

• Parsimonious representation:

$$
\begin{aligned}
Y_t &= X_t\delta + E_t \qquad E_t \sim N(0, \Omega) && (58) \\
\delta &= \Xi\theta + u \qquad u \sim N(0, \Omega \otimes V) && (59)
\end{aligned}
$$

where $X_t = [Y_{t-1}, W_{t-1}]$; $\delta = vec(D_i(L), F_i(L)); \theta = [\theta_1, \theta_2, \theta_3, \ldots]'$.

• **Idea: Factorize the coefficient vector δ into components**: θ is $s \times 1$ vector, Ξ_j are matrices of zeros and ones, $s << k * r$.

• (59) interpreted as a cross-sectional prior or as a part of the model.

Example:

- θ_1 could capture comovements in δ common to all units and variables (a 1×1 vector): can be interpreted as a global factor in the coefficients.

- θ_2 could capture comovements in δ common to all the variables of a unit (a $N \times 1$ vector): can be interpreted as a unit specific factor.

- θ_3 could capture comovements in δ in a variables across all units ($G \times 1$ vector): can be interpreted as a variable specific factor.

- θ_4 captures movements in δ specific to the exogenous variables (1×1).

- u captures unmodelled features of the coefficients vector.

Observable factor model

If (59) is part of the model, using (59) into (58) we have

$$Y_t = \mathcal{Z}_{1t}\theta_1 + \mathcal{Z}_{2t}\theta_2 + \mathcal{Z}_{3t}\theta_3 + \mathcal{Z}_{4t}\theta_4 + v_t = \mathcal{Z}_t\theta + v_t \tag{60}$$

where $\mathcal{Z}_{it} = X_t\Xi_i$, and $v_t = E_t + X_t u$.

• Since Ξ_i are matrices of zeros and ones, regressors of (60) are averages of lags of the VAR variables. Observable factor model dynamically spaning lagged interdependencies between variables and countries. $\theta_i, i = 1, 2, \ldots$ are loadings.

• $\mathcal{Z}_{it}$ are easy to construct (they are observable).

• Analysis feasible with small T and small N and when degrees of freedom in Panel VAR small. Estimate $\theta's$ not VAR coefficients $\delta's$.

Example 6.6 *$G = 2$ variables, $N = 2$ units, 1 lag, no exogenous variables. δ_t is a vector 16×1 . Let*

$$\delta = \Xi_1\theta_1 + \Xi_2\theta_2 + \Xi_3\theta_3 + u$$

θ_1 is scalar, θ_2 is 2×1, θ_3 is 2×1, and the VAR can be rewritten as

$$\begin{bmatrix} y_t^1 \\ x_t^1 \\ y_t^2 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \end{bmatrix} \theta_1 + \begin{bmatrix} \mathcal{Z}_{2,1,t} & 0 \\ \mathcal{Z}_{2,1,t} & 0 \\ 0 & \mathcal{Z}_{2,2,t} \\ 0 & \mathcal{Z}_{2,2,t} \end{bmatrix} \theta_2 + \begin{bmatrix} \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \\ \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \end{bmatrix} \theta_3 + v_t$$

where $\mathcal{Z}_{1t} = y_{t-1}^1 + x_{t-1}^1 + y_{t-1}^2 + x_{t-1}^2$ captures global information, $\mathcal{Z}_{2,1,t} = y_{t-1}^1 + x_{t-1}^1$ captures unit 1 information (across variables), $\mathcal{Z}_{3,1,t} = y_{t-1}^1 + y_{t-1}^2$ captures variable y information (across countries).

How do you estimate the $\theta's$?

- Form $\mathcal{Z}_{it}$.

- If $u = 0$, use OLS to estimate the $\theta's$ in (60).

- If $u \neq 0$, do heteroschedasticity correction (v_t depends on X_t).

- If (59) is treated as a prior, use hierarchical Bayesian methods.

- Easy to use also when the panel VAR has time varying coefficients. If $var(u) \propto var(e_t)$, as in (59), can use Gibbs sampler.

• Hierarchical TVC panel VAR specification:

$$\begin{aligned} Y_t &= X_t\delta_t + E_t & E_t &\sim N(0,\Omega) & (61)\\ \delta_t &= \Xi\theta_t + u_t & u_t &\sim N(0,\Omega\otimes V) & (62)\\ \theta_t &= \theta_{t-1} + \eta_t & \eta_t &\sim N(0,B_t) & (63)\end{aligned}$$

• E_t, u_t, η_t uncorrelated, $V = \sigma^2 I_k$, B_t could be time-varying, e.g. $B_t = \gamma_1 B_{t-1} + \gamma_2 B_0$, with $B_0 = diag\,(B_{01}, B_{02}, B_{03}, B_{04})$.

• Need prior densities for $(\sigma^2, \Omega, B_0, \theta_0)$, see Canova and Ciccarelli (2009).

• Get posterior distribution of (Ω, $\{\theta_t\}_{t=1}^T, \sigma^2$) and of transformations of interest ($Y_{t+\tau|t}$, variance decomposition, etc.).

• The BEAR toolbox estimates this kind of models. It imposes however some restrictions (see manual).

Example 6.7 *Use a TVC panel BVAR with daily data on 2 and 10 years bond returns for DE, FR, IT, ES for 2009d1-2012d1. Use 2 lags and the BEAR options "static structural factor". The bordered spillover matrix at H=15 (days) is*

$$\begin{array}{cc}
 & \begin{array}{ccccccccc} DE2 & DE10 & FR2 & FR10 & IT2 & IT10 & ES2 & ES10 & \text{To} \end{array} \\
\begin{array}{c} DE2 \\ DE10 \\ FR2 \\ FR10 \\ IT2 \\ IT10 \\ ES2 \\ ES10 \\ \text{From} \end{array} &
\left[\begin{array}{ccccccccc}
49 & 34 & 00 & 02 & 09 & 00 & 04 & 05 & 51 \\
28 & 63 & 00 & 01 & 02 & 03 & 00 & 01 & 37 \\
09 & 17 & 34 & 26 & 08 & 00 & 03 & 00 & 66 \\
06 & 27 & 37 & 37 & 00 & 04 & 00 & 03 & 63 \\
06 & 07 & 04 & 12 & 60 & 10 & 01 & 00 & 40 \\
10 & 06 & 04 & 09 & 32 & 33 & 00 & 01 & 66 \\
04 & 04 & 00 & 13 & 29 & 00 & 33 & 05 & 66 \\
07 & 05 & 01 & 12 & 17 & 15 & 14 & 26 & 74 \\
70 & 100 & 46 & 75 & 97 & 32 & 22 & 15 & 920
\end{array}\right]
\end{array}$$

$S(15) = 920/1255 = 73.3\%$; *the average spillover is* $73.3/16 = 4.58\%$.

Relationship Panel VARs and GVARs

• Can we approximate a heterogenous interdependent panel VAR with a GVAR (Dees et al., 2007)?

• What is a GVAR?

$$y_{it} = A_i(\ell)y_{it-1} + B_i(\ell)\bar{y}_t + e_t \tag{64}$$

where $\bar{y}_t = \sum_i w_i y_{it}$ and w_i are trade weights, financial weights, regional weights, etc.

• In general, the approximation is poor.

Example 6.8 *Consider a two unit panel VAR:*

$$y_{1t} = A_{11}(\ell)y_{1t-1} + A_{12}(\ell)y_{2t-1} + e_{1t} \quad (65)$$

$$y_{2t} = A_{21}(\ell)y_{1t-1} + A_{22}(\ell)y_{2t-1} + e_{2t} \quad (66)$$

Let $\bar{y} = 0.5(y_{1t} + y_{2t})$. *Then a GVAR for the first unit is:*

$$\begin{aligned} y_{1t} &= B_{11}(\ell)y_{1t-1} + B_{12}(\ell)\bar{y}_t + u_{1t} \\ y_{1t} &= B_{11}(\ell)y_{1t-1} + 0.5B_{12}(\ell)(y_{1t} + y_{2t}) + u_{1t} \\ (1 - 0.5B_{12}(\ell))y_{1t} &= B_{11}(\ell)y_{1t-1} + 0.5B_{12}(\ell)y_{2t} + u_{1t} \\ y_{1t} &= C_{11}(\ell)y_{1t-1} + C_{12}(\ell)y_{2t} + C_{13}(\ell)u_{1t} \quad (67) \end{aligned}$$

where $C_{13}(\ell) \equiv (1 - 0.5B_{12}(\ell))^{-1}; C_{11}(\ell) \equiv C_{13}(\ell)B_{11}(\ell)$; $C_{12}(\ell) \equiv C_{13}(\ell)B_{12}(\ell)$.

- *Timing of y_{2t} incorrect.*

- *Error term is serially correlated.*

What if we use $\bar{y}_{t-1}$ in the GVAR?

Suppose the model for unit 1 is $y_{1t} = A_{11}(\ell)y_{1t-1} + A_{12}(\ell)y_{2t-1} + A_{13}(\ell)y_{3t-1} + e_{1t}$ *and let* $\bar{y}_t = \frac{1}{3}(y_{1t} + y_{2t} + y_{3t})$. *Then*

$$\begin{aligned} y_{1t} &= B_{11}(\ell)y_{1t-1} + B_{12}(\ell)\bar{y}_{t-1} + u_{1t} \\ &= B_{11}(\ell)y_{1t-1} + \frac{B_{12}(\ell)}{3}(y_{1t-1} + y_{2t-1} + y_{3t-1}) + u_{1t} \\ &= A_{11}(\ell)y_{1t-1} + A_{12}(\ell)y_{2t-1} + A_{13}(\ell)y_{3t-1} + e_{1t} \end{aligned} \tag{68}$$

where $A_{11}(\ell) = (B_{11}(\ell) + \frac{B_{12}(\ell)}{3})$; $A_{12}(\ell) = \frac{B_{12}(\ell)}{3} = A_{13}(\ell)$. *Restrictions on the dynamic effects of lags of units 2 and 3!!*

Relationship Panel VARs and Spatial VARs

- What are spatial VARs? Restricted Panel VARs.

- Restrictions come from spatial structure of the data. For example, the interdependences of the variables of two banks serving the same market are larger than those serving two different markets.

- Spatial structure is typically calibrated with distance measures.

Example 6.9 *N=3, Panel-VAR(1),* $9 \times k$ *regression parameters to estimate,* $k = dim(y_t)$.

$$
\begin{aligned}
y_{1t} &= \rho_{11} y_{1t-1} + \rho_{12} y_{2t-1} + \rho_{13} y_{3t-1} + e_{1t} \\
y_{2t} &= \rho_{21} y_{1t-1} + \rho_{22} y_{2t-1} + \rho_{23} y_{3t-1} + e_{2t} \\
y_{3t} &= \rho_{31} y_{1t-1} + \rho_{32} y_{2t-1} + \rho_{33} y_{3t-1} + e_{3t}
\end{aligned} \tag{69}
$$

Spatial structure: unit 1 and 2 are close, unit 3 far away: $\delta < 1$.

$$\begin{aligned} y_{1t} &= \rho_{11} y_{1t-1} + \delta \rho_{11} y_{2t-1} + \delta^2 \rho_{11} y_{3t-1} + e_{1t} \\ y_{2t} &= \delta \rho_{22} y_{1t-1} + \rho_{22} y_{2t-1} + \delta^2 \rho_{22} y_{3t-1} + e_{2t} \\ y_{3t} &= \delta^2 \rho_{33} y_{1t-1} + \delta^2 \rho_{33} y_{2t-1} + \rho_{33} y_{3t-1} + e_{3t} \end{aligned} \tag{70}$$

• *Only* $3 \times k$ *regression parameters to estimate, for a given* δ*! Note:*

$$\begin{aligned} y_{1t} &= \rho_{11}(y_{1t-1} + \delta y_{2t-1} + \delta^2 y_{3t-1}) + e_{1t} \equiv \rho_{11} W_{1t-1} + e_{1t} \\ y_{2t} &= \rho_{22}(\delta y_{1t-1} + y_{2t-1} + \delta^2 y_{3t-1}) + e_{2t} \equiv \rho_{22} W_{2t-1} + e_{2t} \\ y_{3t} &= \rho_{33}(\delta^2 y_{1t-1} + \delta^2 y_{2t-1} + y_{3t-1}) + e_{3t} \equiv \rho_{33} W_{3t-1} + e_{3t} \end{aligned} \tag{71}$$

• *Observable factor structure. Factors and loadings are unit specific. Factors are weighted versions of the Canova and Ciccarelli (2009) world factor.*

• *The same distance structure applies to all the variables in* y_{it}*. Is it a reasonable restriction?*

Warnings

• Lots of information in a network. Typical to compute systemicness and vulnerability but much more in it.

- Check for similarities (nodes which are similar are likely to be linked, e.g. firms in the same industry).

- Check for centrality (which firms/sectors which contribute most to the spreading of the network).

- Check for clustering (is there any pattern in the links according to some firm/sector characteristics?)

For further ideas about estimation and interpretation of networks see:

- De Paula (2017), De Paula et. al (2018)

- Kastl (2016).

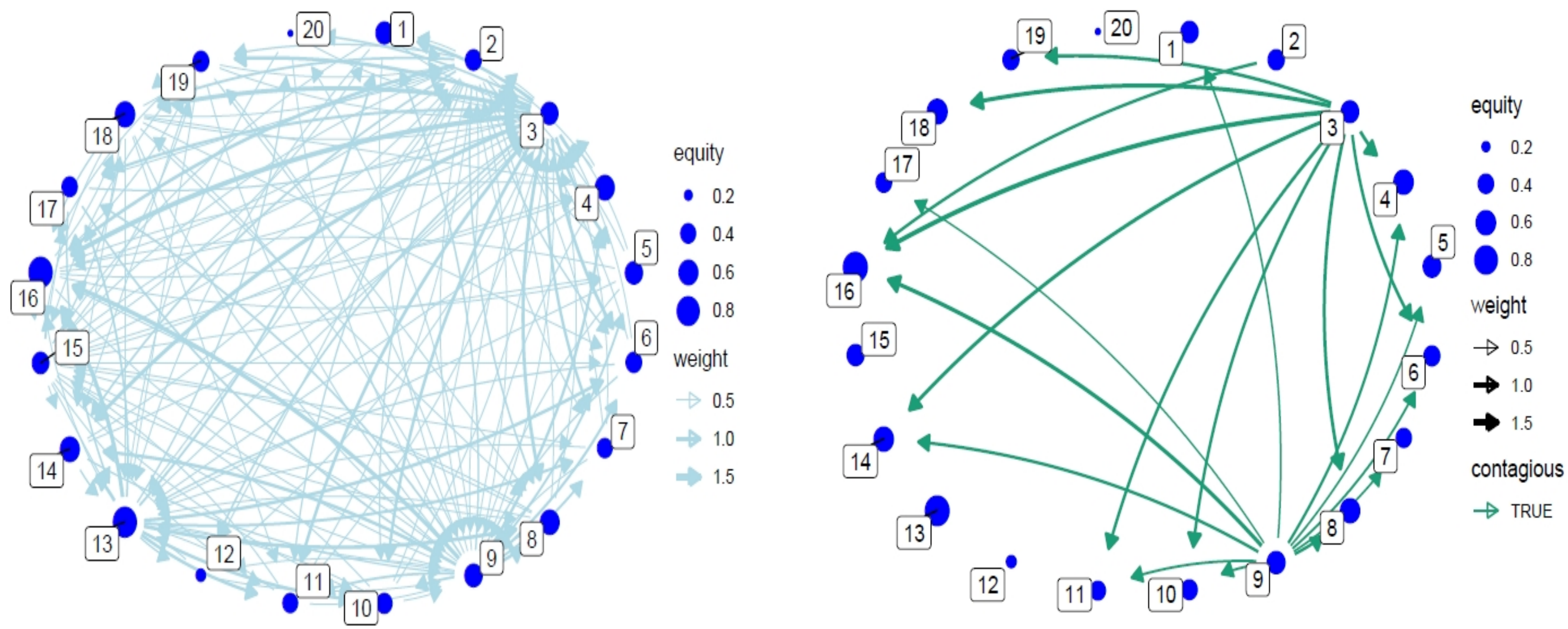
7 Contagion stress testing

• Connectedness/spillover measures examine the linkages of the networks in "normal" time. They do not tell us much about contagion or spillovers during a crisis.

• They are unconditional measures or conditional measures but when the shock initiating the process is "typical" (not unusual, not extreme,..).

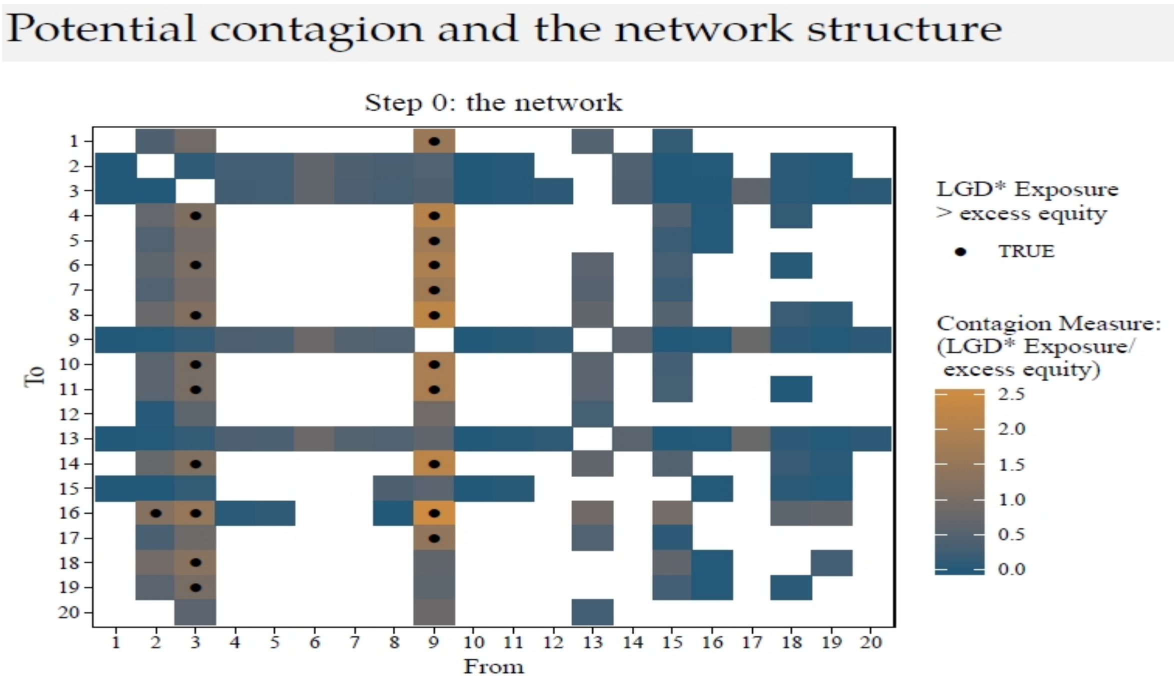
• Often interested in extreme events (e.g. bank failure) and its repercussion on the system, when e.g. the total capital of the economy endogenously adjusts to the initial shock and "multiplies" the intial effect.

• Can how can we measure this?



- Interest is in the systemic effect of disruptions in units 3 and 9.

Network structure



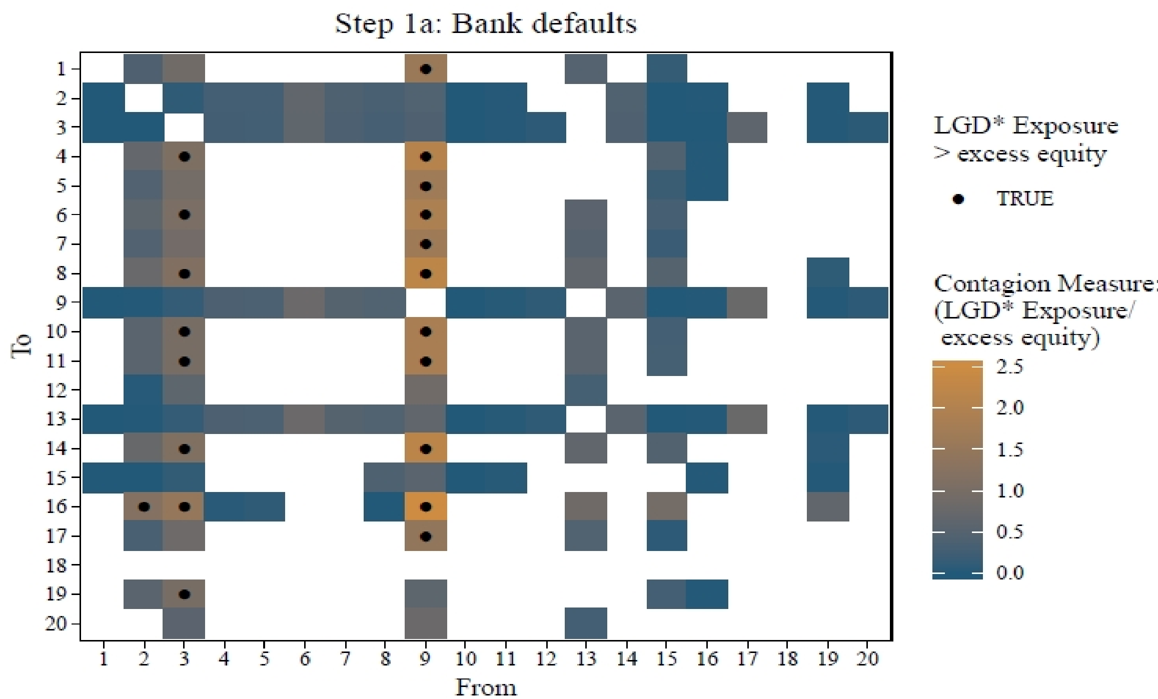
LGD= loss given default (set to 40%)

- Stress testing exercise:

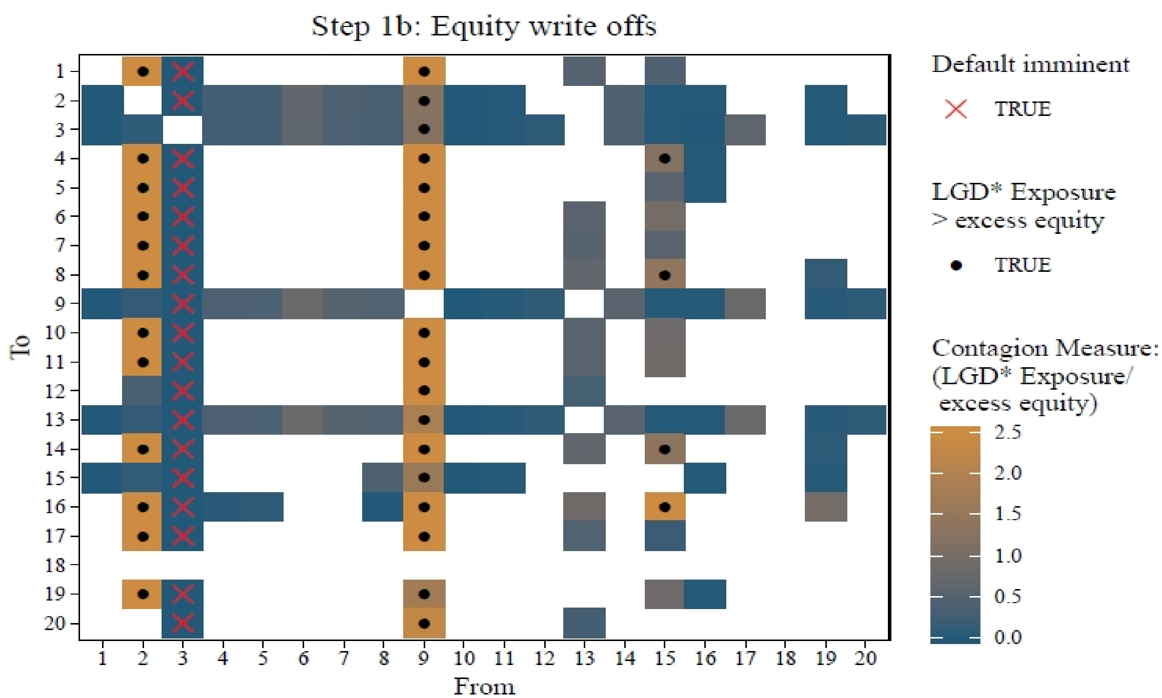
- There is an initial shock. A bank defaults. There is an asset value loss.

- Propagation: Equity are written off. Contagious defaults happen

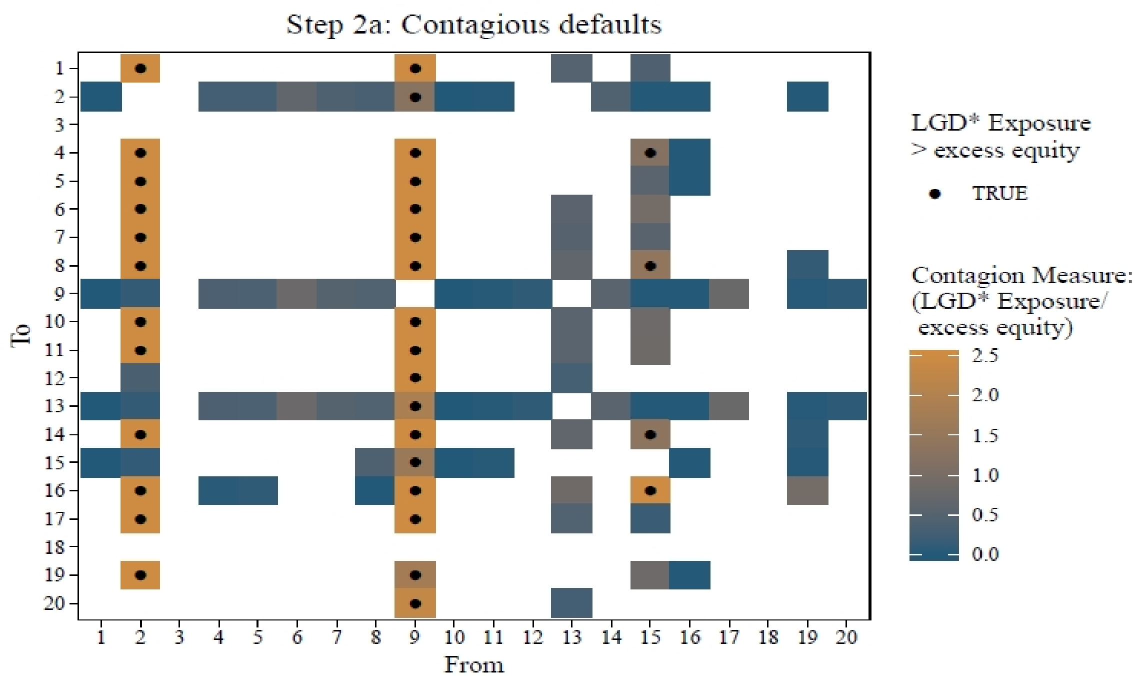
Assumption: Bank 18 defaults



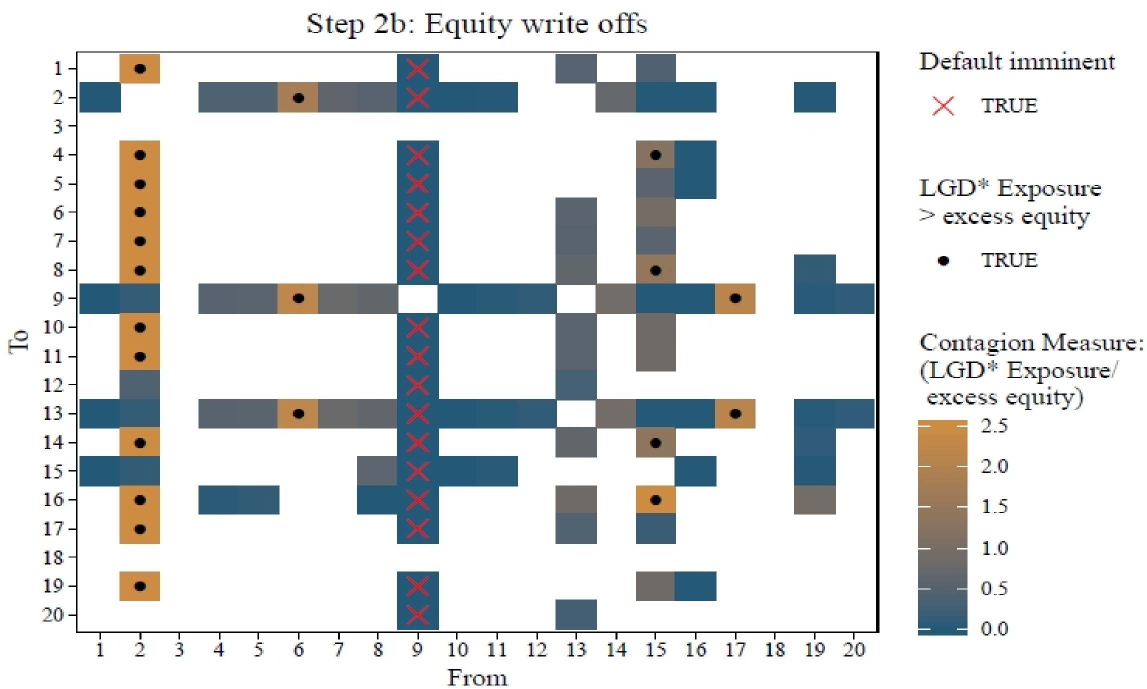
More contagious links and bank 3 will default



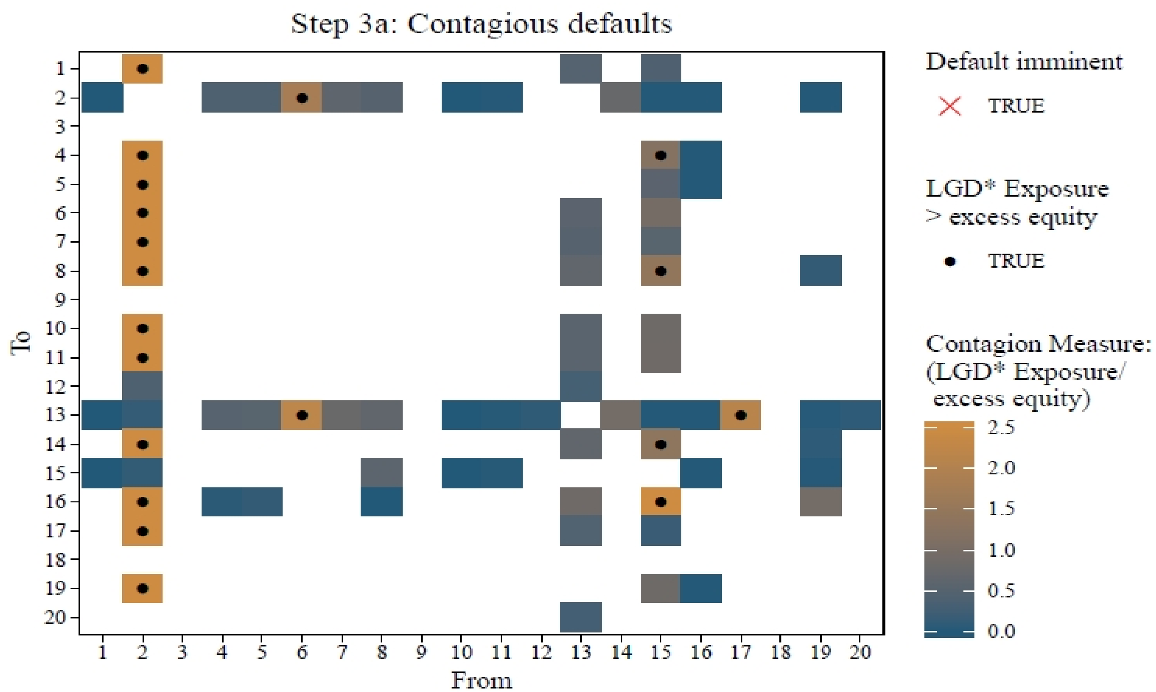
Bank 3 defaults due to contagion



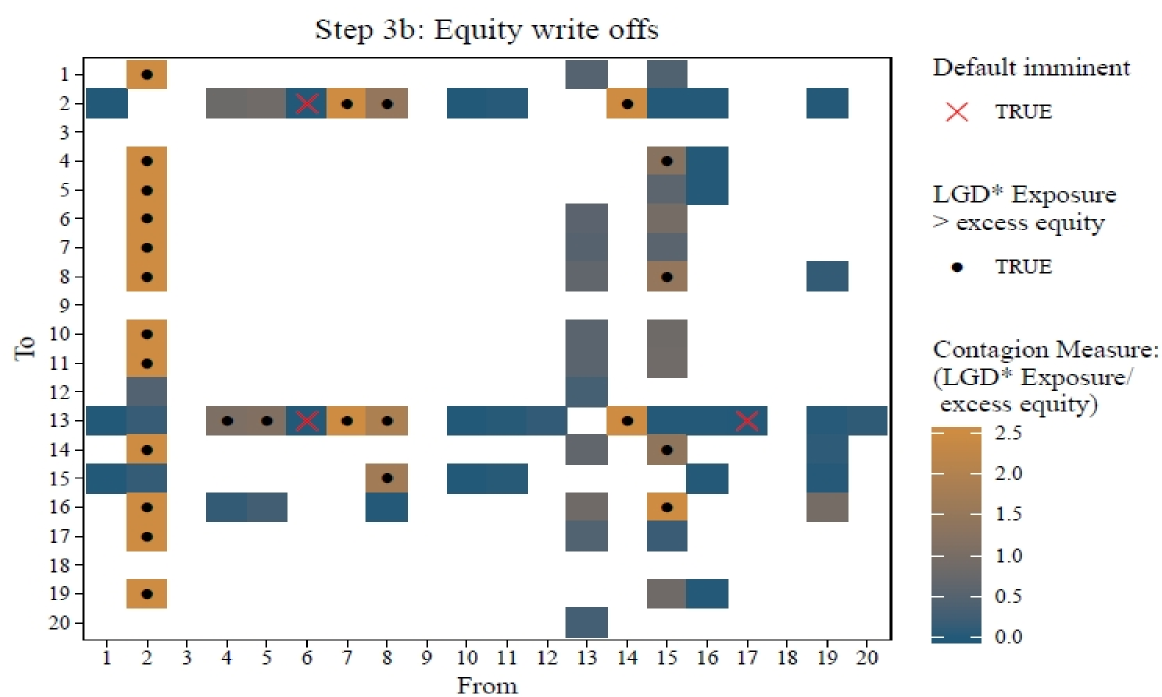
Contagious links appear for 6/17 and 9 defaults



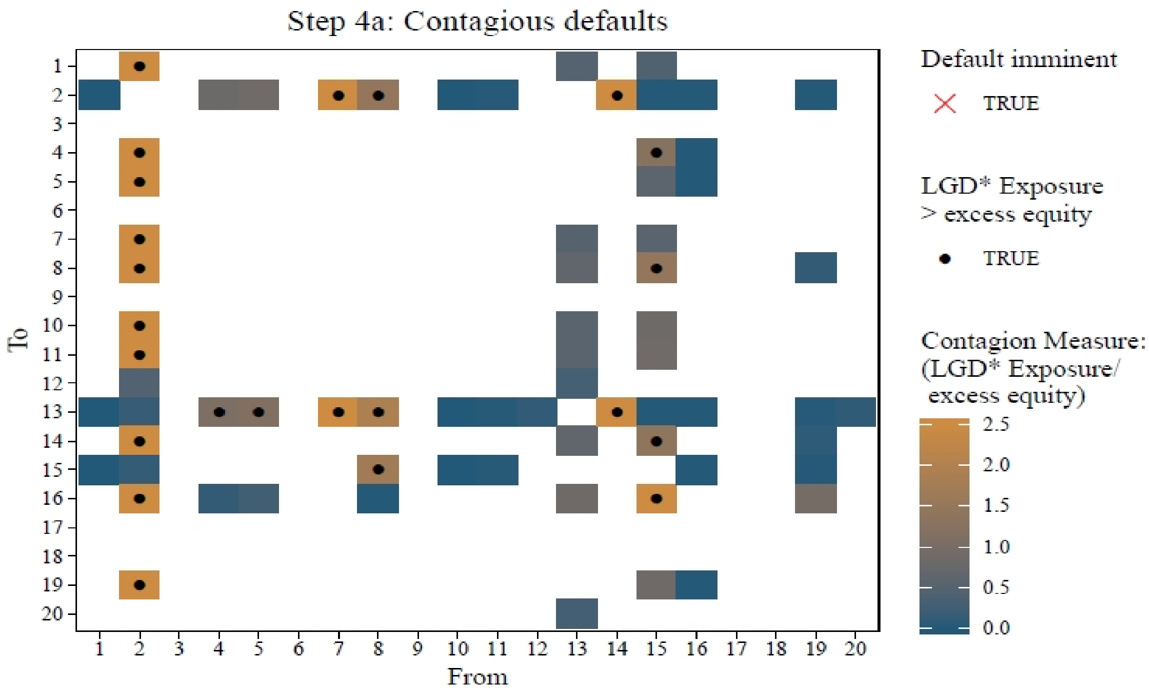
Bank 9 defaults due to contagion



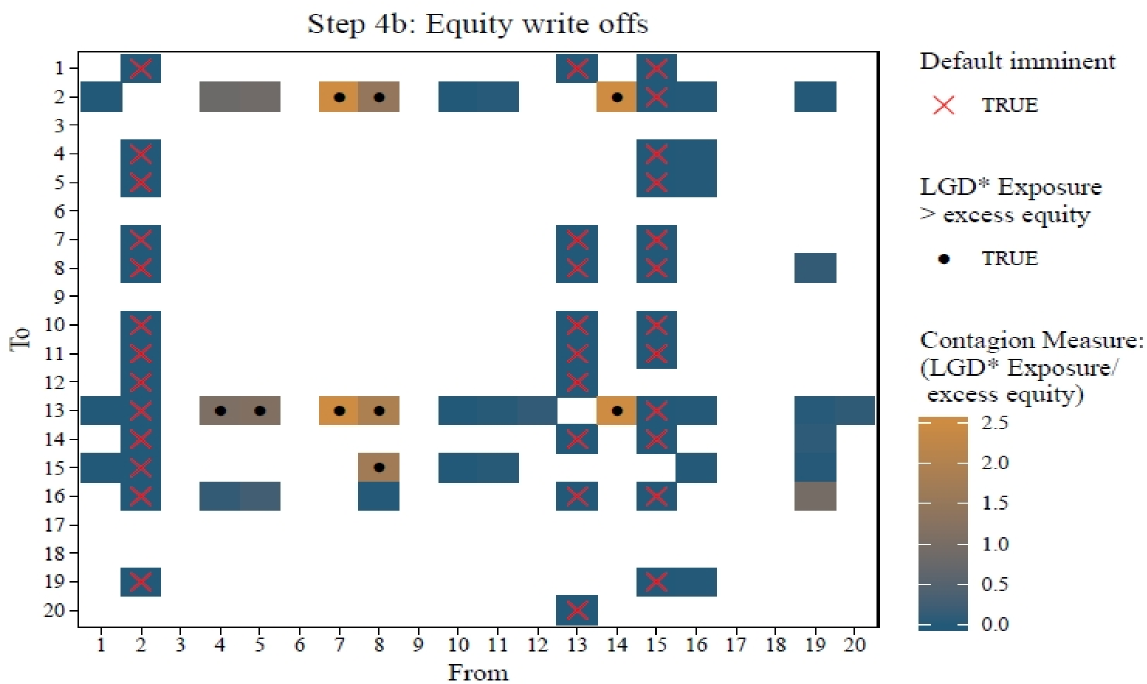
More contagious links and bank 6/17 will default



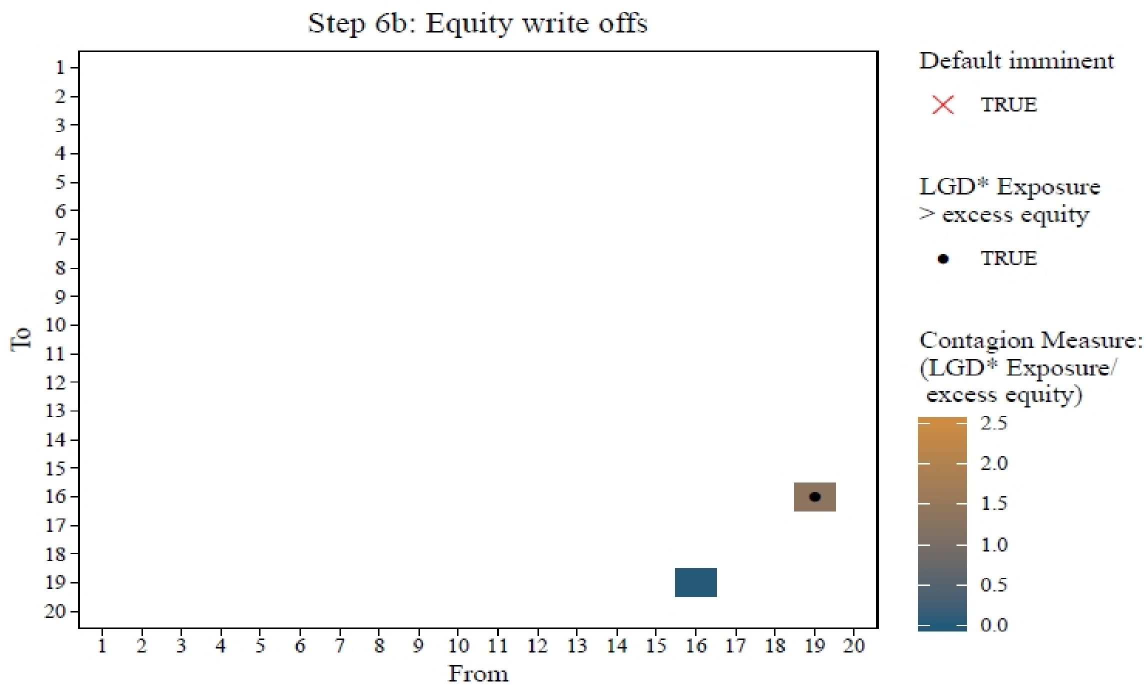
Bank 6/17 defaults due to contagion



Double default triggers unforeseen tripple default



Steady State: 7 banks left. 5 without links



Contagion analysis can reveal vulnerabilities which are hidden to static network analysis

Bank_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Contagion Measures:																				
Number of subsequent defaults if bank_i starts the cascade	13	1	1	12	12	12	12	12	1	13	13	1	1	12	1	15	2	13	13	1
Number of defaults of bank_i in all simulations (n=20)	1	13	13	12	12	12	12	12	14	1	1	1	13	12	13	1	13	2	2	1
Average defaults before bank_i defaults	0.00	3.05	0.75	4.10	3.65	1.70	3.65	3.65	0.75	0.00	0.00	0.00	3.20	3.65	6.40	0.00	1.85	0.50	0.65	0.00
Average defaults after bank_i defaults	0.60	3.25	5.85	1.05	1.30	4.55	1.30	1.30	5.90	0.60	0.60	0.00	3.15	1.30	0.35	0.70	4.45	0.70	0.60	0.00

Policy questions:

- How much capital should there be available to avoid sequenatial defaults?

- Should a regulator prevent the initial default or default at some other level of the cascade?

- Which bank is systemic? The one that triggers the biggest cascade afterwards? Or the bank that is more likely to fail starting from different initial defaults?

- What should a regulator do? Limit the exposure? Build capital buffers? Let the system adjust by itself?

8 Composite Indicator of Systemic Risk (CISS)

• Ex-post measure. Constructed from daily price data. Aggregate various factors of systemic stress. Regularly updated at the ECB site (see http://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=290.CISS.D.U2.Z0Z.4F.EC.SS_CI.IDX)

• Suppose there are N risk factors $z_{it}, i = 1, \ldots, N$. Look for situations where z_i comove and are all high. **Systemic stress is a joint condition of comovement and extremeness**.

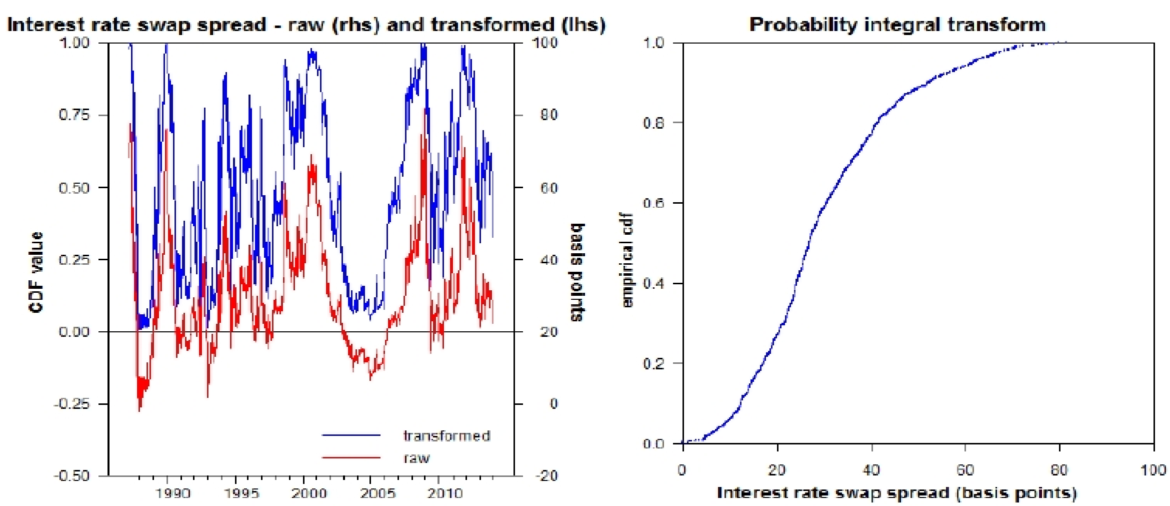
• If G_{1t} defines the comovements and G_{2t} the extremeness of these comovements, a systemic stress index is

$$S_t = \frac{1}{N^2} \sum_i \sum_j G_{1t}(ij) G_{2t}(ij) \tag{72}$$

• Mantel (1967) suggested (72) as scaled matrix association index.

• Each stress factor is computed from row indicators using Probability integral Transform (PIT). This makes z_{it} homogeneous in terms of scale (they are all between 0 and 1) and of distribution (they are all uniform).

Probability integral transform: example case



- Comovements are measured relative to median state

$$Co_t = (z_t - 0.5 i_n)(z_t - 0.5 i_n)^T \tag{73}$$

and made dynamic via (for weekly data)

$$H_t = 0.93 H_{t-1} + 0.07 Co_t \tag{74}$$

The comovement index is constructed as

$$G_{1t}(ij) = \frac{(H_t)_{ij}}{(H_t)_{ii}(H_t)_{jj}} \tag{75}$$

- G_{1t} is a time varying Spearman rank coefficient.

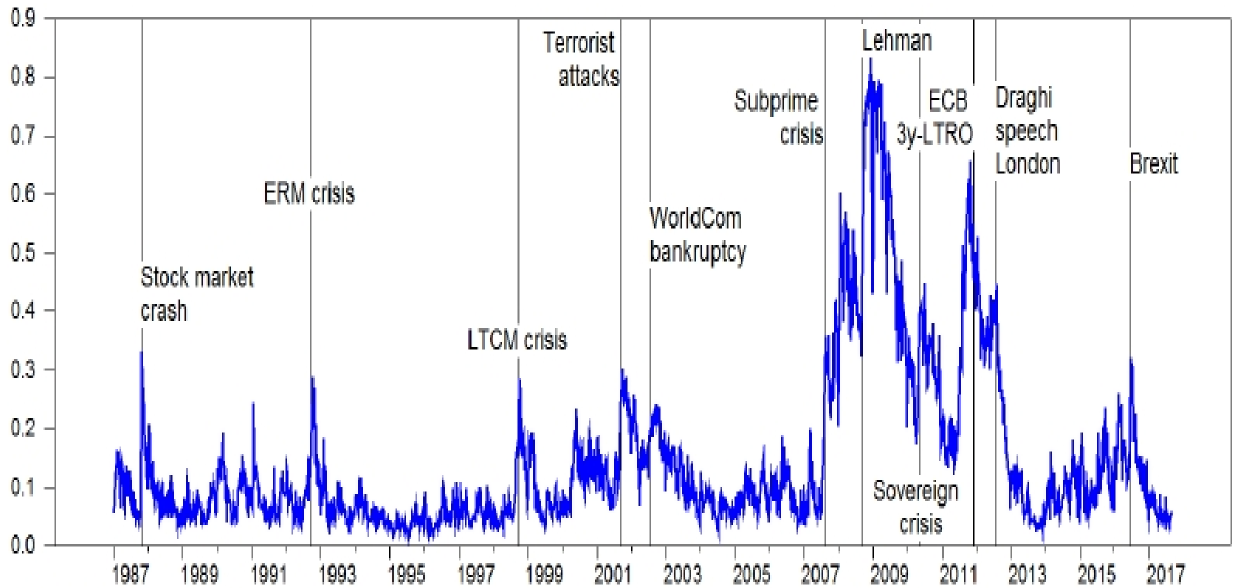
- Extremeness is measured by

$$G_{2t} = 0.93 G_{2t-1} + 0.07 z_t z_t^T \tag{76}$$

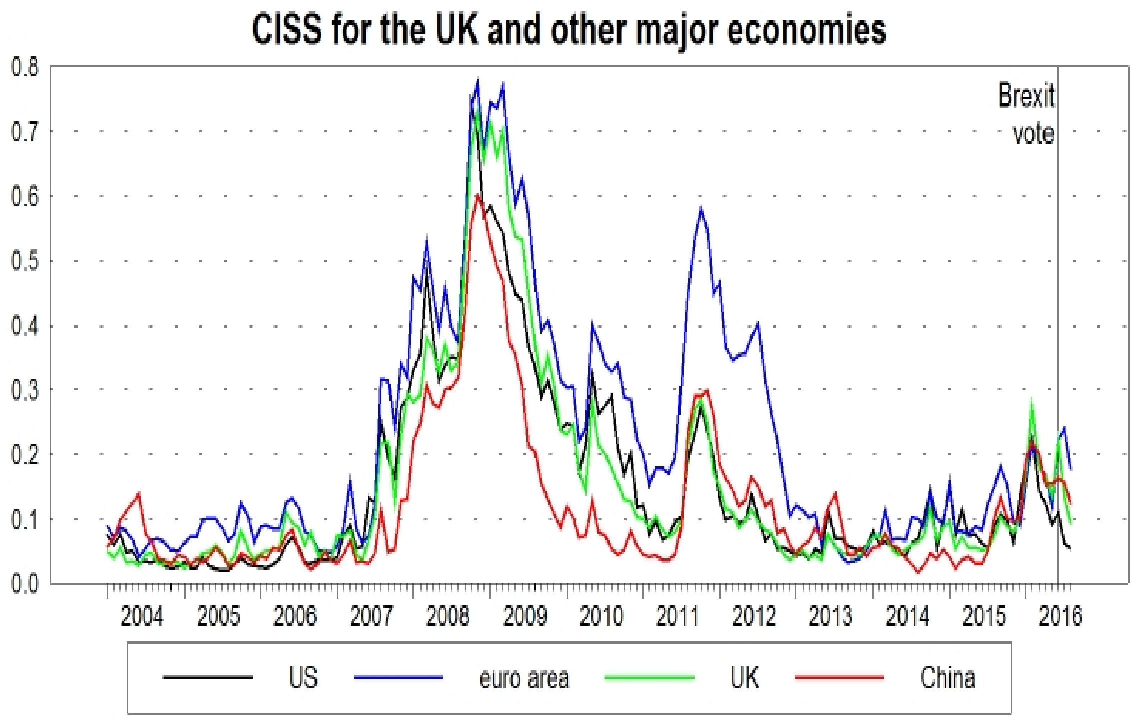
- G_{2t} is at its maximum at $i_N i_N^T$ (when all indicators are high) and the minimum at 0_N (when all indicators are low).

- Raw stress indicators cover money markets; bond and sovereign bond markets; equity markets (non financial firms); financial intermediaries markets, foreign exchange markets.

Euro area CISS and major financial stress events



Notes: weekly data, 2 Jan 1987 to 1 Sept. 2017.



Notes: Monthly data from Jan. 2004 to Oct. 2016

Appendix A:Alternative Representation of VAR(q)

Consider

$$y_t = A(\ell)y_{t-1} + e_t \tag{77}$$

$y_t,\ e_t\ m \times 1$ vectors; A_j is $m \times m$, $e_t \sim (0, \Sigma_e)$.

Different representations are useful for different purposes.

• **Companion form** useful for computing moments, ML estimators. Transform a m-variable VAR(q) into a mq-variable VAR(1).

Example 8.1 *Consider a VAR(3). Let* $\mathbf{Y}_t = [y_t, y_{t-1}, y_{t-2}]'$; $\mathbf{E}_t = [e_t, 0, 0]'$*;*

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & A_3 \\ I_m & 0 & 0 \\ 0 & I_m & 0 \end{bmatrix} \quad \Sigma_E = \begin{bmatrix} \Sigma_e & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Then the VAR(3) can be rewritten as

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{E}_t \quad \mathbf{E}_t \sim \mathbf{N}(0, \Sigma_E) \tag{78}$$

where $\mathbf{Y}_t, \mathbf{E}_t$ *are* $3m \times 1$ *vectors and* $\mathbf{A}$ *is* $3m \times 3m$*.*

• **Simultaneous equation setup** useful for evaluating the likelihood and computing restricted estimates.

There are two alternative representations:

1) Let $x_t = [y_{t-1}, y_{t-2}, \ldots]$; $\mathbf{X} = [x_1, \ldots, x_T]'$ (a $T \times mq$ matrix), $\mathbf{Y} = [y_1, \ldots, y_T]'$ (a $T \times m$ matrix); and if $\mathbf{A} = [A_1', \ldots A_q']'$ is a $mq \times m$ matrix

$$\mathbf{Y} = \mathbf{XA} + \mathbf{E} \tag{79}$$

2) Let i indicate the subscript for the $i - th$ column vector. The equation for variable i is $y_i = x\alpha_i + e_i$. Stacking the columns of y_i, e_i into where $mT \times 1$ vectors we have

$$y = (I_m \otimes x)\alpha + e \equiv X\alpha + e \tag{80}$$

Appendix B: Identification: Obtaining SVARs

VARs are reduced form models. Therefore:

- Shocks are linear combination of meaningful economic disturbances.

- Difficult to relate VAR responses with the responses obtained from theoretical models.

- Can't be used for certain policy analyses (Lucas critique).

What is a SVAR? It is a linear dynamic model of the form:

$$\mathcal{A}_0 y_t = \mathcal{A}_1 y_{t-1} + \ldots + \mathcal{A}_q y_{t-q} + \varepsilon_t \qquad \varepsilon_t \sim (0, \Sigma_\varepsilon) \tag{81}$$

Its reduced form (VAR) is:

$$y_t = A_1 y_{t-1} + \ldots + A_q y_{t-q} + e_t \qquad e_t \sim (0, \mathbf{\Sigma}_e) \tag{82}$$

where $A_j = \mathcal{A}_j \mathcal{A}_0^{-1},\ e_t = \mathcal{A}_0^{-1} \varepsilon_t$.

We want to go from (82) to (81), since (82) is easy to estimate (just use OLS equation by equation). To do this, we need $\mathcal{A}_0$. To estimate $\mathcal{A}_0$, we need restrictions, since $A_j, \mathbf{\Sigma}_e$ have less free parameters than $\mathcal{A}_0,\ \mathcal{A}_j, \mathbf{\Sigma}_\varepsilon$.

- Restrictions should be derived from economic theory !!!

- Distinguish: Stationary vs. nonstationary VARs.

Stationary VARs

$$\text{VAR}: \qquad y_t = A(\ell)y_{t-1} + e_t \qquad e_t \sim (0, \Sigma_e) \tag{83}$$

$$\text{SVAR}: \qquad \mathcal{A}_0 y_t = \mathcal{A}(\ell)y_{t-1} + \epsilon_t \qquad \epsilon_t \sim (0, \Sigma_\epsilon = diag\{\sigma_i\}) \tag{84}$$

(83) and (84) imply

$$\mathcal{A}_0^{-1}\epsilon_t = e_t \tag{85}$$

so that

$$\mathcal{A}_0^{-1}\Sigma_\epsilon \mathcal{A}_0^{\prime -1} = \Sigma_e \tag{86}$$

To recover unknown parameters in $\mathcal{A}_0$ from (86) we need at least as many equations as unknowns.

• Order condition: If the VAR has m variables, need $m(m-1)/2$ restrictions. This is because there are m^2 free parameters on the left hand side of (86) and only $m(m+1)/2$ parameters in Σ_e ($m^2 = m(m+1)/2 + m(m-1)/2$).

• Rank condition: (see Hamilton, 1994, p.332-335).

• Exactly identified vs. overidentified (number of restrictions larger than $m(m-1)/2$).

• Rank and order conditions are valid only for "local identification" (need to be checked at one specific point).

Example 8.2 *i) Cholesky decomposition of Σ_e has exactly $m(m-1)/2$ zeros restrictions. Note $\mathcal{A}_0^{-1}$ is lower triangular and variable i does not affect variable $i-1$ simultaneously, but it affects variable $i+1$.*

ii) $y_t = [GDP_t, P_t, i_t, M_t]$. *Then we need at least 6 restrictions for local identification, e.g.* $\begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_{01} & 1 & 0 & \alpha_{02} \\ 0 & 0 & 1 & \alpha_{03} \\ \alpha_{04} & \alpha_{05} & \alpha_{06} & 1 \end{bmatrix}$.

• Rubio, Waggoner and Zha (2009): sufficient conditions for global identification.

• Valid for exactly or just identified models and for SVAR restrictions which are linear or nonlinear in the parameters.

• When the system is exactly identified a necessary and sufficient condition for global identification is that the rank of Q_j (call it q_j) is equal to $m - j,\ j = 1, 2, \ldots, m$, where Q_j is a matrix composed of zero and ones indicating whether the elements of a column of $[\ \mathcal{A}_0\ A_1]'$ are restricted or not, and A_1 is the first row of the companion form of the system.

Example 8.3 *Let* $\mathcal{A}_0 = \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \end{bmatrix}$. *To check globally identification, rewrite the three zero restrictions into 0 and 1 elements of the matrices* $Q_j, j = 1, 2, 3$, *where* Q_j *is a* 3×3 *matrix of rank* q_j.

In the first column there are two restrictions, in the positions 1 and 2. Therefore, $Q_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ *In the second column there is one restriction, in the position 3. Therefore,* $Q_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ *In the third column there are no restrictions. Therefore,* $Q_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Since $q_1 = 2, q_2 = 1, q_0 = 0$, *the SVAR is globally identified.*

Example 8.4 *(local vs. global identification): Let* $\mathcal{A}_0 = \begin{bmatrix} a_{11} & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}$.

Here $m(m-1)/2 = 3$*: order condition satisfied; rank of Hamilton matrix (at* $a_{11} = a_{22} = a_{33} = 1, a_{13} = a_{21} = a_{32} = 2$*) is 6: rank condition satisfied. SVAR is locally identified. However, since* $q_1 = q_2 = q_3 = 1$*: SVAR is not globally identified. Can also show this in another way. Let* $P = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \end{bmatrix}$. *Note that* $PP = I$. *If we set* $a_{11} = a_{22} = a_{33} = 1, a_{13} = a_{21} = a_{32} = 2$ *then* $\mathcal{A}_0 = \mathcal{A}_0{}^* = \mathcal{A}_0 P$. *Since we have found an observationally equivalent* $\mathcal{A}_0$*, the SVAR is not globally identified.*

- *The SVAR is not globally identified because the restrictions are not "appropriately" placed.*

How do you estimate a stationary SVAR? Use a two-step approach:

- Get (unrestricted) estimates of $A(\ell)$ and Σ_e.

- Use restrictions on $\mathcal{A}_0$ to estimate Σ_ϵ and free parameters of $\mathcal{A}_0$.

- Use $\mathcal{A}(\ell) = \mathcal{A}_0 A(\ell)$ to trace out structural dynamics.

• Unless the system is in Cholesky format, we need ML to estimate $\mathcal{A}_0$ in just identified systems (see appendix).

• For over-identified systems, always need ML to estimate $\mathcal{A}_0$.

• IV approach interpretation.

Example 8.5 *(Blanchard and Perotti, 2002) VAR with* (T_t, g_t, y_t). *Let the SVAR be* $\mathcal{A}_0 y_t = \mathcal{A}(\ell) y_{t-1} + \mathcal{B}\epsilon_t, \epsilon_t \sim (0, \Sigma_\epsilon = diag\{\sigma_i\})$ *Here* $\mathcal{A}_0 e_t = \mathcal{B}\epsilon_t$, *where*

$$\mathcal{A}_0 = \begin{bmatrix} 1 & 0 & a_{01} \\ 0 & 1 & a_{02} \\ a_{03} & a_{04} & 1 \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} 1 & b_1 & 0 \\ b_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Restrictions: no discretionary response of T_t *and* g_t *shocks to* y_t *within the quarter (last column of* $\mathcal{B}$ *has two zeros). Information delays.*

We have a total 6+3 (variance) parameters to estimate. At most 6 parameters in Σ_e. *Need additional restrictions. Get information about* a_{01}, a_{02} *from external sources; further impose either* $b_1 = 0$ *or* $b_2 = 0$

With a_{01}, a_{02} *fixed, two stage approach has a IV interpretation:* e_{1t}, e_{2t} *are predetermined and used as instruments for* e_{3t}.

Nonstationary VARs

Let VAR and SVAR be:

$$\begin{aligned} \Delta y_t &= D(\ell)e_t = D(1)e_t + D^*(\ell)\Delta e_t & (87)\\ \Delta y_t &= \mathcal{D}(\ell)\mathcal{A}_0\epsilon_t = \mathcal{D}(\ell)(1)\mathcal{A}_0\epsilon_t + \mathcal{D}^*(\ell)\mathcal{A}_0\Delta\epsilon_t & (88) \end{aligned}$$

where $D(\ell) = (I - A(\ell)\ell)^{-1},\ \mathcal{D}(\ell) = (1 - \mathcal{A}(\ell)\ell)^{-1},\ D^*(\ell) \equiv \frac{D(\ell)-D(1)}{1-\ell}$, $\mathcal{D}^*(\ell) \equiv \frac{\mathcal{D}(\ell)-\mathcal{D}(1)}{1-\ell}$. Matching coefficients: $\mathcal{D}(\ell)\mathcal{A}_0\epsilon_t = D(\ell)e_t$.

Separating permanent and transitory components and using for the latter only contemporaneous restrictions we have

$$\begin{aligned} \mathcal{D}(1)\mathcal{A}_0\epsilon_t &= D(1)e_t & (89)\\ \mathcal{A}_0\Delta\epsilon_t &= \Delta e_t & (90) \end{aligned}$$

If y_t is stationary, $\mathcal{D}(1) = D(1) = 0$ and (89) is vacuous.

Two types of restrictions to estimate $\mathcal{A}_0$: short and long run.

Example 8.6 *In a VAR(2) imposing (89) requires one restriction. Suppose that $\mathcal{D}(1)^{12} = 0$ (ϵ_{2t} has no long run effect on y_{1t}). If $\Sigma_\epsilon = I$, the three elements of $\mathcal{D}(1)\mathcal{A}_0\Sigma_\epsilon\mathcal{A}_0'\mathcal{D}(1)'$ can be obtained from the Cholesky factor of $D(1)\Sigma_e D(1)'$.*

• Blanchard-Quah: decomposition in permanent-transitory components (use (89)-(90)). If $y_t = [\Delta y_{1t}, y_{2t}]$, $(m \times 1)$; y_{1t} are I(1); y_{2t} are I(0) and $y_t = \bar{y} + D(\ell)\epsilon_t$, where $\epsilon_t \sim iid(0, \Sigma_\epsilon)$

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ 0 \end{pmatrix} + \begin{pmatrix} D_1(1) \\ 0 \end{pmatrix} \epsilon_t + \begin{pmatrix} (1-\ell)D_1^\dagger(\ell) \\ (1-\ell)D_2^\dagger(\ell) \end{pmatrix} \epsilon_t \tag{91}$$

where $D_1(1) = [1, 0]$ and $D^\dagger(\ell) = D(\ell) - D(1)$. Here y_{2t} is any set of stationary variables which is influenced by both shocks.

Identification through heteroskedasticity

Idea: $var(e_t) = \Sigma_1, t = 1, \ldots, T_1, \quad var(e_t) = \Sigma_2, t = T_1 + 1, \ldots, T,$

• Lutkepohl (1996, Chapter 6.1.2): There exists a W and a diagonal Ω with typical element $\omega_i > 0, i = 1, 2, \ldots, m$ such that $\Sigma_1 = WW'$ and $\Sigma_2 = W\Omega W'$.

• W is a full matrix. It is unique up to sign changes if ω_i are distinct. Since it is the same in $\Sigma_1$and Σ_2, the impact effect of (structural) shocks is unchanged.

• Ω incorporates volatility changes (if one $\omega_i \neq 1$, there is a change in volatility) - shocks are normalized to 1 in the first sample and to ω_i in the second.

• If $\mathcal{A}_0^{-1} = W$ all shocks of the system are identified.

• Under the above restrictions, and if the variance of the shocks changes once at know date, shocks can be identified without economic restrictions (!!). If economic restrictions exist, they become overidentifying and can be tested

• If more than two regimes, variance changes provides overidentification restrictions (see Rigobon, 2003).

• Lanne and Lutkepohl (2008): Markov switching structure in the variance of the shocks. Same idea applies.

Example 8.7

$$\begin{aligned} p_t &= \beta y_t + \epsilon_{1t} && (92) \\ y_t &= \alpha p_t + \epsilon_{2t} && (93) \end{aligned}$$

where $E(\epsilon_{1t}\epsilon_{2t}) = 0$. *The covariance matrix of* $[p_t, y_t]'$ *is*

$$V \equiv \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} = \frac{1}{(1-\alpha\beta)^2} \begin{bmatrix} \beta^2\sigma_2^2 + \sigma_1^2 & \beta\sigma_2^2 + \alpha\sigma_1^2 \\ \beta\sigma_2^2 + \alpha\sigma_1^2 & \sigma_2^2 + \alpha^2\sigma_1^2 \end{bmatrix}$$

Thus there are three free elements in V *and 4 structural parameters* $(\alpha, \beta, \sigma_1^2, \sigma_2^2)$ *- the system is underidentified.*

Suppose σ_1^2, σ_2^2 *depend on the state. Let* $s = 1, 2$. *Then*

$$V_1 = \frac{1}{(1-\alpha\beta)^2} \begin{bmatrix} \beta^2\sigma_{21}^2 + \sigma_{11}^2 & \beta\sigma_{21}^2 + \alpha\sigma_{11}^2 \\ \beta\sigma_{21}^2 + \alpha\sigma_{11}^2 & \sigma_{21}^2 + \alpha^2\sigma_{11}^2 \end{bmatrix}$$

$$V_2 = \frac{1}{(1-\alpha\beta)^2} \begin{bmatrix} \beta^2\sigma_{22}^2 + \sigma_{12}^2 & \beta\sigma_{22}^2 + \alpha\sigma_{12}^2 \\ \beta\sigma_{22}^2 + \alpha\sigma_{12}^2 & \sigma_{22}^2 + \alpha^2\sigma_{12}^2 \end{bmatrix}$$

There are six free elements in V_1, V_2 *and six structural parameters* $(\alpha, \beta, \sigma_{11}^2,$ $\sigma_{12}^2\sigma_{21}^2, \sigma_{22}^2)$ *- system just identified by order condition!!*

If we have three variance regimes, we have ($3 \times 3 = 9$) reduced form parameters and 8 structural parameters (3×2 structural variances, α, β). System over-identified.

Important: to identify all the parameters we need:

1) α and β to be unchanged.

2) The variance of both shocks to be changing (if there are two regimes).