Cross-Sectional Dynamics Under Network Structure: Theory and Macroeconomic Applications

Marko Mlikota

University of Pennsylvania

This Version: June 19, 2023

[latest version: click here]

Abstract

Many environments in economics feature a cross-section of units linked by bilateral ties. I develop a framework for studying dynamics of cross-sectional variables exploiting this network structure. It is a vector autoregression in which innovations transmit cross-sectionally only via bilateral links and which can accommodate rich patterns of how network effects of higher order accumulate over time. The model can be used to estimate dynamic network effects, with the network given or inferred from dynamic cross-correlations in the data. It also offers a dimensionality-reduction technique for modeling (cross-sectional) processes, owing to networks' ability to summarize complex relations among units by relatively few non-zero bilateral links. In a first application, I estimate how sectoral productivity shocks transmit along supply chain linkages and affect dynamics of sectoral prices in the US economy. The analysis suggests that network positions can rationalize not only the strength of a sector's impact on aggregates, but also its timing. In a second application, I model industrial production growth across 44 countries by assuming global business cycles are driven by bilateral links which I estimate. This reduces out-of-sample mean squared errors by up to 23% relative to a principal components factor model.

JEL codes: C32, C38, D57, E37.

Key words: Vector Autoregression, Spatial Autoregression, Dynamic Peer Effects, High-Dimensional Time Series, Sparse Factors, Input-Output Economy, Price Dynamics, Global Business Cycles.

---

* Correspondence: Department of Economics, University of Pennsylvania, 133 South 36th Street, Philadelphia, PA 19104-6297. Email: mlikota@sas.upenn.edu. This paper constitutes part of my doctoral dissertation at the University of Pennsylvania. I am highly indebted to my advisors, Frank Diebold and Frank Schorfheide, for their invaluable support and guidance. For helpful comments and discussions I am also grateful to Xu Cheng, Marco Del Negro, Thorsten Drautzburg, Simon Freyaldenhoven, Wayne Gao, Bo Honoré, Leon Huetsch, Daniel Lewis, Aaron Mora, Serena Ng, Michael Pollmann, Giorgio Primiceri, Joao Ritto, Elisa Rubbo, Adrien Wicht as well as seminar and conference participants at Penn, 7th Lindau Meeting on Economic Sciences, Geneva Graduate Institute, SNDE Symposium 2023, and Philadelphia Fed.

1 Introduction

Numerous economic environments feature a cross-section of units connected by a network of bilateral ties. For example, countries are connected via flows of trade and capital, industries are linked through supply chains, and individuals in a society form a network by virtue of being acquainted to one another. As demonstrated theoretically and documented empirically,[1] networks can rationalize comovement in variables measured at the cross-sectional level; GDP across countries varies depending on demand and supply by trade partners, firms adjust their prices in response to price increases by suppliers, individuals receive information and form opinions by interacting with their social network.

What is less well understood, however, is how network-induced comovements play out over time. With regard to the timing of network effects, the literature considers two restrictive cases. The first assumes that innovations transmit via bilateral links contemporaneously (e.g. Acemoglu et al. (2012, 2016); Elliott et al. (2014)). This leads to a static framework and implies that connections of all order play out (simultaneously). For example, an individual talks to all their friends, who in turn talk to all their respective friends, etc., so that at each point in time everyone's opinion incorporates those of all members of society and within the same period fully adjusts to any new information gathered by even its most distant member. The second case posits that network effects materialize exactly one link per period (e.g. Long and Plosser (1983), Golub and Jackson (2010)). This assumption is tenable in theoretical contributions, but in empirical studies a period is defined by data and it remains an empirical question how far a shock travels through the network in one observational period. Only for studies limited to steady state comparisons – i.e. long-term effects of permanent shocks – the timing of network effects is irrelevant.[2]

I build an econometric framework to study the dynamics of cross-sectional variables when units are connected through a network. It is a vector autoregression (VAR) parameterized based on the assumption that innovations transmit cross-sectionally only via bilateral links. Transmission is assumed to be uni-directional[3] and links fixed over time. The framework can accommodate general patterns on how innovations travel through the network as time progresses, i.e. which connection-orders matter for transmission at which horizons. Observational dependence arises as the interplay between temporal distance and cross-sectional distance encoded by the network. Correspondingly, stationarity can be characterized in terms of eigenvalues of the network adjacency matrix and roots of an AR process defined by the timing of network effects. I show that this timing is fundamentally related to the frequency of network interactions relative to the frequency of observation.

[1]See references in the following paragraph and subsequent literature review.

[2]These long-term effects turn out to be the same as the effects in the static framework of contemporaneous linkages. See Section 2.2.2 and Appendix A.4.

[3]i.e. either downstream or upstream. The distinction is only relevant for directed networks.

The Network-VAR (NVAR) is useful in two rather distinct lines of empirical work with cross-sectional time series. On the one hand, it can be used to estimate dynamic peer effects, i.e. to quantify how innovations transmit along bilateral links over time and come to shape cross-sectional dynamics. Thereby, the network can be taken as given or inferred from dynamic (cross-)correlations in the data, possibly aided by shrinking towards observed links. With both the network and effect-timing estimated, the NVAR is also applicable as a dimensionality-reduction technique for modeling (cross-sectional) processes. It assumes that dynamics are generated by innovation transmission along a (small) set of bilateral links. Given the network, inference on the timing of network effects boils down to a linear regression with covariates that summarize lagged observations using bilateral links. Joint inference is implemented easily by iteration on analytically available conditional estimators, with a frequentist as well as Bayesian interpretation. I illustrate each of these two model uses with a respective application.

In the first application, I estimate how sectoral productivity shocks propagate through the supply chain network and shape the monthly dynamics of Producer Price Indices (PPI) in the US economy. I show that the NVAR approximates the process of sectoral prices in a Real Business Cycle (RBC) input-output economy with time lags between the production of goods and their subsequent use as intermediaries in producing other goods. Long-term effects of permanent price increases in this framework are equivalent to the effects of price increases in static models with contemporaneous input-output conversion. By estimating the timing of input-output conversion, the NVAR decomposes these overall, long-term effects over time and estimates transition dynamics induced by sectoral price shocks.

The results suggest that network positions have implications not only for the strength of sectoral shocks' effects on aggregates – as documented in existing literature – but also for its timing, with no clear relationship between the two. How quickly a shock in a sector affects aggregate PPI is determined by the sector's importance as an immediate – as opposed to further upstream – supplier to relevant sectors in the economy. Owing to their position at the top of supply chains, the response to price increases in energy-related sectors is estimated as particularly slow to unravel.

In the second application, I model industrial production growth across 44 countries by estimating an underlying network as relevant for dynamics. This provides a novel perspective on global business cycles by assuming that the dynamic comovement in economic activity across countries is the result of bilateral connections. The NVAR yields a sparse, yet flexible way of approximating cross-sectional processes even in high dimensions. Sparsity is obtained because dynamics are driven by bilateral links and because units can be connected even in absence of a direct link between them. As a result, the dynamic comovement of the whole, potentially high-dimensional cross-section can be modeled with relatively few non-zero bilateral links. This is reminiscent of the assumption that longer-term dynamics are driven by a set of shorter-term dynamics, which is upheld by the general class of VARMA(p, q)

models. Flexibility is owed to the fact that the network is estimated and that the model can accommodate general patterns of how network effects of higher order accrue over time.

An equivalence result suggests that the NVAR is preferred to a factor model whenever cross-sectional dynamics are composed of many micro links rather than driven by a few influential units. This corresponds to the case of numerous sparse factors with differing sets of non-zero loadings across units, or, equivalently, a sparse, yet high-rank network adjacency matrix. In my application, the NVAR – with the network estimated by selecting links to zero – leads to reductions in out-of-sample mean squared errors of up to 23% relative to a principal components factor model, in particular for horizons up to six months.

Related Literature This paper adds to the growing literature on networks in econometrics.[4] In particular, there is a large literature on spatial autoregressive models (SAR). It is mostly concerned with identifying network effects (and effects of other covariates) in a static framework of contemporaneous dependencies (Manski, 1993; Lee, 2007; Bramoullé et al., 2009). In contrast, I analyze lagged network effects. As opposed to other studies in this category (Knight et al., 2016; Zhu et al., 2017; Yang and Lee, 2019), I cast them in an explicit time series model and study dynamic, contagion-like innovation transmission through the network. For this purpose, I generalize the time profile of network effects, I explicitly relate the model's time series properties to the network and timing of network effects, and I show how to conduct inference on the latter two.

A few weakly connected strands of literature incorporate networks into time series analyses. At a fundamental level, my work relates to Diebold and Yilmaz (2009, 2014), who map variance decompositions of VARs into networks with the goal of understanding dynamic connectedness.[5] In contrast, I map networks into VARs. In particular, I use a network to model the conditional mean function, restricting innovations to transmit via bilateral links. This leads to rich patterns of multi-step causality, connecting my work to Dufour and Renault (1998). Relative to other studies restricting time series models using networks (e.g. Pesaran et al. (2004), Chudik and Pesaran (2011), Barigozzi et al. (2022), Mehl et al. (2022), Caporin et al. (2023)), I focus on a simple case of one variable per cross-sectional unit and one type of connection among units. This leads to a clear relation between the model's time series properties on the one hand and the network and timing of network effects on the other. It also generates analytical and intuitive expressions for the estimators and allows me to examine the relation to factor models. Approaches for achieving shock identification using networks are discussed in Hipp (2020) and Dahlhaus et al. (2021). Bykhovskaya (2021) builds a time series model for the evolution of the network itself.

With my first application, I address the macroeconomic literature on production networks (Long and Plosser (1983), Acemoglu et al. (2012), Horvath (2000), Foerster et al. (2011),

[4] See Bramoullé et al. (2016) and Graham (2020) for general references on networks in economics and econometrics.

[5] Another way to represent dynamics by graphs is offered in Barigozzi and Brownlees (2018).

Bouakez et al. (2014)). In various environments, it has been shown that networks amplify idiosyncratic shocks and generate cross-sectional comovement (Acemoglu et al. (2012), Giroud and Mueller (2019), Giovanni et al. (2018)). Thereby, contemporaneous network interactions are standardly assumed; e.g. firms produce output by using contemporaneously produced input goods. This framework is silent on how networks drive aggregate and cross-sectional dynamics.[6] An exception is Long and Plosser (1983), who study a Real Business Cycle (RBC) economy with a one period delay in input-output conversion and show that it leads to endogenous business cycles (persistence in aggregate TFP). Building on their work, I assume general time lags in input-output conversion and potentially differing frequencies of network interaction and data observation. This implies that observed sectoral prices (and output) follow an NVAR process, which allows me to estimate how the overall, long-term amplification of sectoral shocks reported in the literature unfolds over time.

With the second application of the NVAR, this paper addresses the vast literature on dimensionality-reduction techniques for time series modeling. The NVAR combines insights from two commonly used approaches. Compared to reduced rank regression and factor models (Velu et al., 1986; Stock and Watson, 2002), it offers a particular way of finding the linear combination that effectively summarizes the information in the lagged values of the process, namely by bilateral links among cross-sectional units (or variables, in case of non-cross-sectional time series). Compared to variable selection methods such as Lasso (Tibshirani, 1996),[7] it places exclusion restrictions on network-links, which in turn summarize the information in predictors, rather than on predictors themselves. This leads to additional sparsity as the same links are used to summarize information at all lags, although different connection-orders are used at different lags.[8] As discussed in Boivin and Ng (2006), factor models' forcasting performance deteriorates under high dispersion in factor loadings across series. This notably includes case of sparse factors and units differing in the set of factors they load on.[9] The NVAR improves upon factor models in this case by naturally incorporating sparse factors as locally important nodes in the network.

The remainder of this paper is structured as follows. The model and its properties are discussed in Section 2. Section 3 treats inference. In Section 4, I study how input-output connections shape the dynamics of sectoral prices in the US economy, taking the network as given. In Section 5, I illustrate the merits of the NVAR as a dimensionality-reduction technique for modeling cross-sectional processes, and I apply it to forecast cross-country industrial production. Section 6 concludes.

[6]Under contemporaneous interactions, network effects of all order play out simultaneously. In other words, network effects themselves are static; networks can only amplify existing dynamics – obtained thanks to agents' intertemporal optimization problems in a structural model or due to persistence in shocks – but not drive dynamics themselves. See discussion in Section 2.2.2.

[7]See Hsu et al. (2008) and Camehl (2022) for applications of Lasso in the context of VARs.

[8]Other approaches bridging sparse and factor models are usually interested in capturing the cross-sectional correlation in the errors left after factor extraction. See e.g. Fan et al. (2021).

[9]For analyses of sparse factors, see Onatski (2012) and Freyaldenhoven (2022).

2 Lagged Network Effects & Cross-Sectional Dynamics

After providing some basic background on networks in Section 2.1, I present the NVAR in Section 2.2, building on a simple example. Thereby, I explicitly examine the relation between the frequencies of network interaction and observation. Also, I discuss stationarity and the relation to contemporaneous network interactions. Estimation is deferred to Section 3.

2.1 Bilateral Connections in Networks

A network is represented by an $n \times n$ adjacency matrix A with elements a_{ij}. I consider a directed and weighted network, which means that $a_{ij} \in [0,1]$ shows the strength of the link from cross-sectional unit i to unit j, with $a_{ij} \neq a_{ji}$ possibly. If $a_{ij} = 0$, I say unit i is not connected to unit j. The set of bilateral links $\{a_{ij}\}_{i,j=1:n}$ give rise to a plethora of higher-order connections among units, referred to as walks.[10]

Definition 1 (Walk)**.** *A walk from i to j is the product of a sequence of links $a_{i_k,i_{k+1}}$ between units $i_1, i_2, ..., i_K$ such that $a_{i_k,i_{k+1}} \neq 0 \;\forall\; k$, $i_1 = i$, $i_K = j$. For example,*

$$a_{i,i_2,...,i_{K-1},j} \equiv a_{i,i_2} \left[\prod_{k=2}^{K-2} a_{i_k,i_{k+1}} \right] a_{i_{K-1},j} \; ,$$

is a walk from unit i to unit j of length K.

A walk is the product of bilateral links a_{ij} that lead from unit i to unit j over some intermediary units, all of which are sequentially connected. Just as element (i,j) in the matrix A shows the walk from i to j of length one (direct link), simple matrix algebra reveals that $(A^K)_{ij}$ contains the sum of walks from i to j of length K.[11] I refer to this quantity as the Kth-order connection from i to j.

Consider the following example:

$$A = \begin{bmatrix} 0 & 0 & .8 \\ .7 & 0 & .6 \\ 0 & .8 & 0 \end{bmatrix} \; , \quad A^2 = \begin{bmatrix} 0 & .64 & 0 \\ 0 & .48 & .56 \\ .56 & 0 & .48 \end{bmatrix} \; , \quad A^3 = \begin{bmatrix} .448 & 0 & .384 \\ .336 & .448 & .288 \\ 0 & .384 & .448 \end{bmatrix} \; .$$

Even though unit 3 is not directly connected to unit 1 ($a_{31} = 0$), there exists a second-order connection via unit 2 ($a_{32}a_{21} \neq 0$). For example, in a production network, unit 1 could be a supplier to unit 2, who in turn is a supplier to unit 3.

[10] Whenever convenient to simplify notation, I write $a : b$ for the set of integers $\{a, a+1, ..., b\}$, $a \leq b$.

[11] In the case of an unweighted network, $a_{ij} \in \{0,1\}$ and so any walk $a_{i,i_2,...,i_{K-1},j} \in \{0,1\}$, which means that $(A^K)_{ij}$ contains the number of walks from i to j.

2.2 Lagged Innovation Transmission via Bilateral Links

Underlying the proposed NVAR is the core assumption that innovations u_{it} to a process y_{it} transmit cross-sectionally only via bilateral links. This transmission is assumed to operate only in one direction through the network. Specifically, the direct link from i to j, a_{ij}, is a vehicle for innovation transmission from j to i. Innovations can be correlated cross-sectionally (but not over time). For expositional simplicity, I assume $\mathbb{E}[y_{it}] = 0 \ \forall \ i, t$.

2.2.1 Simple Example: NVAR(1,1)

Let $y_t = [y_{1t}, ..., y_{nt}]'$ and $u_t = [u_{1t}, ..., u_{nt}]'$, and consider the following VAR(1):

$$y_t = \Phi y_{t-1} + u_t \ , \quad \mathbb{E}_{t-1}[u_t] = 0 \ ,^{12} \qquad \text{with} \quad \Phi = \alpha A \ , \quad \alpha \in \mathbb{R} \ . \tag{1}$$

Taking Φ to be proportional to the adjacency matrix A of a network that connects the cross-sectional units, one obtains a process that relates the dynamics of the cross-sectional time series y_t to the bilateral links among cross-sectional units.[13] Long and Plosser (1983) derive such a process for sectoral output and prices in an RBC production economy with a one period delay in converting inputs into output[14]. Golub and Jackson (2010) use it in their study of societal opinion formation through friendship ties. Under this process, the one period-ahead expected value of y_{it} is proportional to a weighted sum of one period-lagged values of y_{jt} for all units j to which i is directly linked, with weights given by the strength of direct links a_{ij}: $\mathbb{E}_{t-1}[y_{it}] = \alpha \sum_{j=1}^{n} a_{ij} y_{j,t-1}$. In the example from Long and Plosser (1983), the price charged tomorrow by firms in sector i is expected to be a weighted average of prices charged by their suppliers today. Dynamics of y_t, as summarized by Granger-causality at different horizons $h = 1, 2, ...$, are shaped by hth order connections encoded in A:

$$GC_{ij}^h \equiv \frac{\partial y_{i,t+h}}{\partial y_{jt}} \big| \mathcal{F}_t^y = \frac{\partial y_{i,t+h}}{\partial u_{jt}} \big| \mathcal{F}_t^y = (\alpha^h A^h)_{ij} \ .$$

As a result, given all other variables y_k , $k \neq j$, y_j is useful in forecasting y_i at horizon h iff there is an hth order connection from i to j. The strength of this relationship is determined by the strength of this connection, i.e. by the sum of all walks from i to j of length h. Note that GC_{ij}^h is also referred to as the Generalized Impulse Response Function (GIRF). It is generalized because it is not concerned with identification, but the derivative is taken with respect to (potentially correlated) reduced form errors in u_t.

Fig. 1 provides an example. It depicts the Granger-causality pattern for the process in Eq. (1) and the network from Section 2.1. Each panel (i, j) shows the network connection

[12] $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_{t-1}^y]$, where $\mathcal{F}_{t-1}^y = \{y_{t-j}\}_{j=1}^{\infty}$ is the information set at $t-1$.

[13] As explained in Section 2.2.2, I call this process NVAR(1,1).

[14] See Appendix C.1 and the discussion in Section 4.

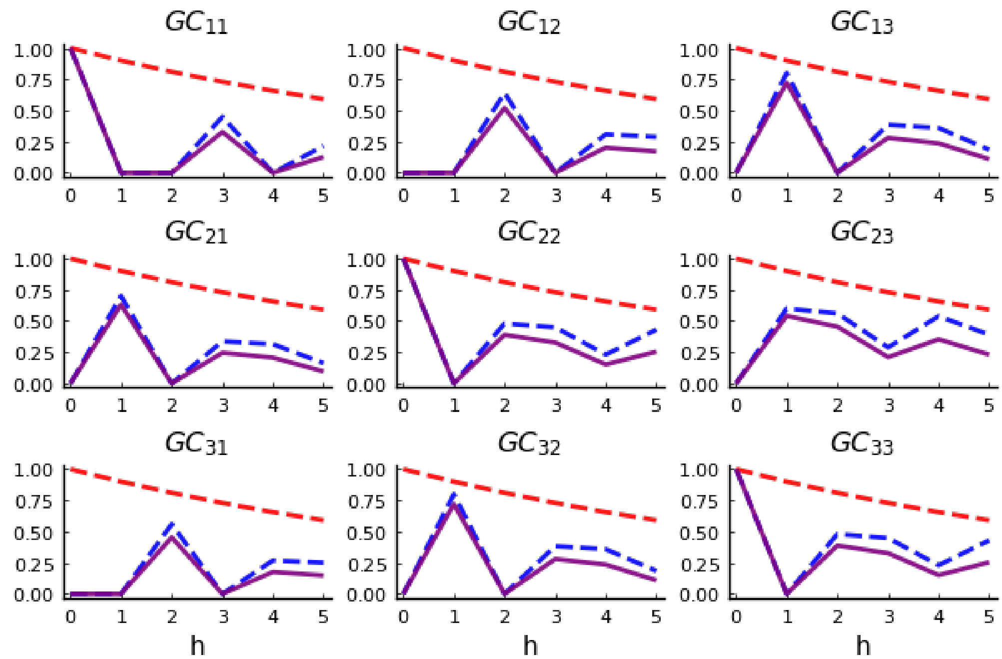


Figure 1: Example Generalized Impulse Responses: NVAR(1, 1)

Notes: Panel (i, j) shows $(A^h)_{ij}$ in blue, α^h in red and $GC^h_{ij} = (\alpha^h A^h)_{ij}$ in purple.

from i to j at different orders h, $(A^h)_{i,j}$, in blue, the decaying series α^h for $\alpha = 0.9$ in red, and their product, the GIRF, in purple.[15] By definition, the contemporaneous responses to all but a series' own innovation are zero. From horizon $h = 1$ onwards, the GIRF for every pair (i, j) is proportional to the network connections from i to j of relevant order.

Two points are worth highlighting. First, the relation between the shocked and responding unit in the network shapes not only the strength of the impulse-response, but also its timing. For example, while unit 2 is directly linked to unit 1 and therefore experiences the latter's innovation with a lag of one period, unit 3 only has an indirect, second-order connection to unit 1 and is therefore impacted by its innovation only after two periods. Second, lagged network interactions can themselves be a source of persistence. Even with all diagonal elements in A equal to zero – i.e. no unit is connected directly to itself – and ruling out autocorrelation in disturbances u_{it}, y_{it} still persistently reacts to changes in u_{it} because of spillback effects. For example, unit 3 is linked to 2, which itself is linked to unit 3. Therefore, after an initial adjustment to its own disturbance, unit 3 experiences further rounds of adjustments because its initial response led to a response of unit 2.[16]

These results relate to the discussion in Dufour and Renault (1998), who point out that

[15] I take $\alpha \in [0, 1]$ purely for illustration purposes. As discussed in Section 2.2.2, it is not required for stationarity.

[16] For larger networks, such second-round responses can surpass the initial response and lead to a hump-shaped response. Generally, if unit i has weak lower-order connections to j but strong higher-order connections, we can have $\alpha^h(A^h)_{ij} > \alpha^{h+1}(A^{h+1})_{ij}$ but $\alpha^h(A^h)_{ij} < \alpha^{\tilde{h}}(A^{\tilde{h}})_{ij}$ for some h and $\tilde{h} > h + 1$.

Granger-causality can take the form of chains. Specifically, even if a series X does not Granger-cause a series Y at horizon 1, under the presence of a third series Z, X might Granger-cause Y at higher horizons as the causality could run from X to Z to Y. They examine conditions under which noncausality at a given horizon implies noncausality at higher horizons. If innovations transmit only via bilateral links, these generally non-trivial conditions boil down to the existence or non-existence of network connections of relevant order between the concerned variables (units under cross-sectional time series).

While useful for theoretical work such as Long and Plosser (1983) and Golub and Jackson (2010), the process in Eq. (1) is of limited use for empirical studies as it entertains a very restrictive mapping between network connections and observed dynamics in y_t. First, it presumes that the frequencies of network interactions and observation are equal, i.e. that innovations travel through the network at the speed of one link per period. This would imply, for instance, that firms in sector i do not adjust their price in response to price increases by suppliers situated two positions upstream of i (suppliers of suppliers) earlier than with a lag of two periods.[17] Second, it assumes complete transmission at a single lag. For example, given a price increase by a supplier-sector j, firms in sector i fully adjust their own price after one period. Further price adjustments in the next periods are possible only to the extent that j is also a supplier to other suppliers of i, i.e. only if there is a second-order connection. In the following, I extend the simple process above along both of these dimensions.

2.2.2 General Model: NVAR(p, q)

The NVAR embodies two ideas. First, innovation transmission along bilateral links is not instantaneous, but happens with a lag and is possibly spread out over several periods. Second, the frequency of such network interactions can differ from the frequency of observation. Combining these ingredients yields a model that can accommodate general patterns on how innovations transmit through the network over time.

Let the cross-sectional time series x_τ evolve according to

$$x_\tau = \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ , \quad \alpha = [\alpha_1, ..., \alpha_p]' \in \mathbb{R}^p \ , \quad \mathbb{E}_{\tau-1}[v_\tau] = 0 \ . \tag{2}$$

Compared to the process in Eq. (1), the process in Eq. (2) allows connections of order lower than h to affect dynamics at horizon h:

Proposition 1 (Granger-Causality in NVAR$(p, 1)$)**.**
Let x_τ evolve as in Eq. (2). Assuming $\alpha_l \neq 0 \; \forall \; l$, x_j Granger-causes x_i at horizon h iff there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}{+}1, ..., h\}$, where $\underline{k} = ceil(h/p)$.[18]

[17] It also implies that their initial response cannot occur later than with a lag of two periods.
[18] $ceil(x)$ rounds $x \in \mathbb{Q}$ up to the next integer.

See Proposition 4 and its proof in Appendix A. The proof establishes that the GIRF is of the form

$$\frac{\partial x_{i,\tau+h}}{\partial v_{j,\tau}} \big| \mathcal{F}_\tau^x = c_{\underline{k}}^h(\alpha) \left[A^{\underline{k}}\right]_{ij} + ... + c_h^h(\alpha) \left[A^h\right]_{ij} \ . \tag{3}$$

The coefficients $\{c_k^h(\alpha)\}_{k=\underline{k}:h}$ are polynomials of $\{\alpha_l\}_{l=1:p}$ and show the importance of different connection-orders for the impulse response at a given horizon h. Eq. (2) specifies that x_τ is shaped by lagged network interactions, whereby innovation transmission along a bilateral link takes p periods to fully materialize.[19] As elaborated on below, $\sum_{l=1}^{p} \alpha_l$ can be thought of as the overall strength of innovation transmission, while the individual elements in α show the time profile of this transmission.[20] Note that it is assumed to be the same for all unit pairs (i, j).

The process x_τ evolves at frequency τ, which I shall call the network interaction frequency.[21] It might not coincide with the frequency of observation. In particular, if data is observed at a lower frequency than network interactions occur, dynamics at horizon h can be driven by connections of order higher than h, as several rounds of transmission can happen in one period of observation. In addition, this leads to network-induced cross-sectional correlation in innovations to the observed process even in absence of correlation in v_τ. Let the observed data be $\{y_t\}_{t=1:T}$. Also, for the follwing, assume v_τ follows a Normal distribution.

If y_t is a stock variable, we can write $\{y_t\}_{t=1:T} = \{x_{tq}\}_{t=1:T}$ for some $q \in \mathbb{Q}_{++}$ and represent the dynamics of y_t as a state space system:

$$\begin{aligned} x_\tau &= \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ , \\ y_{\tau/q} &= x_\tau \quad \text{if } \tau/q \in \mathbb{N} \ , \end{aligned} \tag{4}$$

whereby $\tau = 1 : T_\tau$.[22] I will dub this process NVAR(p, q) (for stock variables).

If $q^{-1} \in \mathbb{N}$, observational frequency either coincides with network interaction frequency ($q = 1$) or is an integer-multiple thereof. In either case, all x_τ are observed. Under Normality

[19] By this I mean the transmission from one unit to another disregarding the responses of other units. This transmission constitutes the whole impulse-response if two units do not share any higher-order connections: if $a_{ij} \neq 0$, but $(A^h)_{ij} = 0 \ \forall \ h > 1$, then $\frac{\partial x_{i,\tau+h}}{\partial v_{j,\tau}} \big| \mathcal{F}_\tau^x = \alpha_h$ for $h = 1 : p$ and zero otherwise.

[20] They are allowed to be negative. For example, under $p = 2$, $(\alpha_1 > 0, \alpha_2 < 0)$ with $\alpha_1 + \alpha_2 > 0$ signifies an initial overreaction and subsequent correction of unit i's series after an innovation at units j to which i is connected.

[21] As Eq. (2) and Eq. (3) make clear, τ denotes a frequency at which innovation transmission occurs over a set of time intervals, all of which are of integer length. If transmission happens at regular intervals, τ is simply the frequency at which it takes one period of time for an innovation to transmit (partially) along a direct link from one cross-sectional unit to another. However, as the subsequent discussion shows, under Normality of v_τ, τ is not unique, but one can write the process at an integer-multiple frequency of τ without changing its distributional properties.

[22] T_τ is such that the number of elements in the set $1 : T_\tau$ that are integer-multiples of q is equal to T.

of v_τ, dynamics of y_t can be represented by an NVAR$(p/q, 1)$:

$$y_t = \gamma_1 A y_{t-1} + ... + \gamma_{p^*} A y_{t-p^*} + u_t \ , \quad \gamma_l = \begin{cases} \alpha_{lq} & \text{if } l \text{ is multiple of } q^{-1} \\ 0 & \text{otherwise} \end{cases} \ , \quad u_t \sim v_\tau \ ,^{23}$$

with $p^* = p/q \in \mathbb{N}$. For example, if network interactions occur at quarterly frequency, but observations are monthly, we have $q = 1/3$, and the observed monthly series depends on its value three months ago, six months ago, etc., up to $3p$ months ago.

If $q \in \mathbb{N}\backslash\{1\}$, network interaction frequency is an integer-multiple of observational frequency such that we observe x_τ every q periods. If x_τ is stationary,[24] dynamics of y_t can be approximated arbitrarily well by the following restricted VARMA process:

$$y_t \approx \sum_{l=1}^{p^*} \Phi_l y_{t-l} + u_t \ , \quad u_t = \sum_{l=0}^{p^*-1} \Theta_l \eta_{t-l} \ , \quad \Phi_l = \sum_{g=1}^{q^*} \gamma_{lg} A^g \ , \tag{5}$$

for p^* large and $q^* = p^* q$. The coefficients γ_{lg} are polynomials of $\{\alpha_l\}_{l=1:p}$ (or zero), and $\eta_t = [v'_\tau, ..., v'_{\tau-q+1}]'$ stacks all high-frequency innovations that occurred in-between the periods of observation $t-1$ and t. The $n \times nq$ matrices $\{\Theta_l\}_{l=0}^{p^*-1}$ are made up of $n \ \times n$ blocks, all of which are polynomials in A of the form of that for Φ_l. Appendix A.2 illustrates for the case of $p = 3$ and $q = 2$.

For other cases with $q \in \mathbb{Q}_{++}$, the same result obtains as under $q \in \mathbb{N}\backslash\{1\}$. We can write $q = q_1 q_2$ with $q^{-1} \in \mathbb{N}$ and $q_2 \in \mathbb{N}$ and deduce the process for y_t by defining the auxiliary process z_{τ^*}. It is obtained by writing x_τ at the higher frequency τ^* given by $\tau = \tau^* q_1$ as an NVAR$(p/q_1, 1)$, just as illustrated above. Then, $\{y_t\}_{t=1:T}$ contains a snapshot of z_{τ^*} every q_2 periods. This takes care of cases such as tri-weekly network interactions and monthly observations, or vice versa.

If y_t is a flow variable, we can write $y_t = x_\tau + ... + x_{\tau-q+1}$ for $\tau = tq$ provided that $q \in \mathbb{N}$, i.e. the network interaction frequency either coincides with the observational frequency ($q = 1$) or is an integer-multiple thereof.[25] In that case, dynamics of y_t are given by the following state space system:

$$\begin{aligned} x_\tau &= \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ , \\ y_{\tau/q} &= x_\tau + ... + x_{\tau-q+1} \quad \text{if } \tau/q \in \mathbb{N} \ . \end{aligned} \tag{6}$$

This is the NVAR(p, q) for flow variables. For $q \in \mathbb{N}\backslash\{1\}$, analogous conclusions about the

[23] As can be easily verified, under distributional equivalence of u_t and v_τ, $\mathbb{E}[y_t y_{t-h}] = \mathbb{E}[x_\tau x_{\tau-hq}] \ \forall \ h$, and under Normality, the first two moments fully characterize the process. Trivially, for $q = 1$ Normality is not required.

[24] See below and Appendix A.3 for a discussion on stationarity.

[25] Analogous calculations apply if $y_t = (x_\tau + ... + x_{\tau-q+1})/q$.

dynamics of y_t can be drawn as in the case of stock variables.[26] For $q \notin \mathbb{N}$, no state space representation can be found for y_t without assumptions that allow us to break down a flow variable from lower to higher frequency.

This time-aggregation of lagged transmission patterns has interesting implications for the mapping from network connectedness to dynamics. If $x_\tau \sim \text{NVAR}(p, 1)$, we know by Proposition 1 that $\partial x_{\tau+h}/\partial x_\tau \big| \mathcal{F}_\tau^x = \partial x_{\tau+h}/\partial v_\tau \big| \mathcal{F}_\tau^x$ for $h = 1, 2, ...$ is composed of network-connections of order $k \in \{ceil(h/p), ..., h\}$. For $h = 0$, we have $\partial x_\tau/\partial v_\tau \big| \mathcal{F}_\tau^x = I$. As Eq. (5) suggests, if x_τ is observed infrequently, several rounds of innovation transmission can occur in one observational period. This implies that not only connections of order lower than h, but also those of order higher than h can matter for dynamic relationships at horizon h. Similar intuition applies for the case of flow variables.

Formally, if y_t is a stock variable and $q \in \mathbb{N}$, we get that

$$\frac{\partial y_{t+h}}{\partial y_t} \big| \mathcal{F}_t^y = \frac{\partial x_{(t+h)q}}{\partial v_{tq}} \big| \mathcal{F}_{tq}^x = \frac{\partial x_{\tau+hq}}{\partial v_\tau} \big| \mathcal{F}_\tau^x$$

for $h = 1, 2, ...$ is composed of network-connections of order $k \in \{ceil(hq/p), ..., hq\}$. Because x_τ is observed every q periods, up to q rounds of transmission can occur in one period of observation, and connections of order up to hq can matter at horizon h. Regarding the contemporaneous innovation to the observed process, $\eta_t = [v'_{tq}, ..., v'_{tq-q+1}]'$, we obtain that $\partial y_t/\partial \eta_t \big| \mathcal{F}_t^y$ can be composed of connections of order $k \leq q - 1$, depending on which of the terms $v_{tq}, ..., v_{tq-q+1}$ is behind the change in η_t.[27] Even without cross-sectional correlation in high-frequency innovations v_τ, there will be cross-sectional correlation in the innovations η_t to the observed process, induced by network interactions materializing at a higher frequency than data is observed.

If y_t is a flow variable and $q \in \mathbb{N}$, we get that $\partial y_{t+h}/\partial y_t \big| \mathcal{F}_t^y$ can be composed of network-connections of order $k \in \{ceil((q(h-1)+1)/p), ..., hq\}$, depending on which of the terms in $y_t = x_{tq} + ... + x_{q(t-1)+1}$ is responsible for the change in y_t.[28] Again, for $\partial y_t/\partial \eta_t \big| \mathcal{F}_t^y$, connections of order $k \leq q - 1$ can matter, depending on which of the terms $v_{tq}, ..., v_{tq-q+1}$ is behind the change in η_t.[29]

---

[26] See Appendix A.2.

[27] If it is the first term, then network-connections do not matter: $k \in \emptyset$. If it is the second term, then only first-order connections matter: $k = \{ceil(1/p), ..., 1\} = \{1\}$. If it is the last term, then connections of order $k \in \{ceil((q-1)/p), ..., q-1\}$ matter.

[28] If it is the first term, we get

$$\frac{\partial y_{t+h}}{\partial x_{tq}} \big| \mathcal{F}_t^y = \frac{\partial (x_{(t+h)q} + ... + x_{(t+h)q-q+1})}{\partial x_{tq}} \big| \mathcal{F}_t^x q = \frac{\partial x_{\tau+hq}}{\partial x_\tau} \big| \mathcal{F}_\tau^x + ... + \frac{\partial x_{\tau+hq-q+1}}{\partial x_\tau} \big| \mathcal{F}_\tau^x$$

and so connection-orders $k \in \{ceil((q(h-1)+1)/p), ..., hq\}$ matter. Analogous calculations show that if it is the last term, connection-orders $k \in \{ceil(hq/p), ..., hq+q-1\}$ matter.

[29] With a subtle change to before. If it is the first term, then network-connections do not matter: $k \in \emptyset$. If it is the second term, then only first-order connections matter: $k \in \{ceil(1/p), ..., 1\} \cup \emptyset = \{1\}$. If it is the

Stationarity As the following propositions show, stationarity of the NVAR can be chracterized in terms of eigenvalues of the network adjacency matrix A and the time profile of network effects α. These results simplify checking for stationarity, in particular if A is high-dimensional.

Proposition 2 (Stationarity of NVAR$(p, 1)$)**.**
Let x_τ follow an NVAR$(p, 1)$:

$$x_\tau = \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ ,$$

with $v_\tau \sim WN$, and assume $\alpha_l \neq 0$ for at least one l. Define $a = \sum_{l=1}^{p} |\alpha_l|$.
Then, x_τ is weakly stationary iff the univariate AR(p) process

$$\check{x}_\tau = \lambda_i \alpha_1 \check{x}_{\tau-1} + ... + \lambda_i \alpha_p \check{x}_{\tau-p} + \check{v}_\tau$$

is weakly stationary for all eigenvalues λ_i of A.
Furthermore, a sufficient condition for weak stationarity of x_τ is is that for all eigenvalues λ_i of A it holds that $|\lambda_i| < 1/a$. If in addition $\alpha_l \geq 0 \ \forall \ l$, then this condition is both necessary and sufficient.

See Corollary 1 and Proposition 8 and their proofs in Appendix A. Intuitively, these conditions ensure that $lim_{k\to\infty} \ c(\alpha, k)A^k = 0$, for any polynomial in α of order k, $c(\alpha, k)$.[30] This in turn ensures that long-term effects of shocks go to zero as the horizon increases. As expected, stationarity of the underlying high-frequency NVAR$(p, 1)$ implies stationarity of the observed NVAR(p, q) process:

Proposition 3 (Stationarity of NVAR(p, q), $q > 1$)**.**
Let x_τ follow an NVAR$(p, 1)$

$$x_\tau = \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ ,$$

with $v_\tau \sim WN$. Let $q \in \mathbb{N}\backslash\{1\}$ and consider the time series y_t defined by $\{y_t\}_{t=1}^{T} = \{x_{tq}\}_{t=1}^{T}$ and z_t defined by $\{z_t\}_{t=1}^{T} = \{x_{tq} + ... + x_{(t-1)q+1}\}_{t=1}^{T}$. Then if x_τ is weakly stationary, so are y_t and z_t.

See Proposition 9 and its proof in Appendix A.[31]

third term, then connections of order $k \in \{ceil(2/p), ..., 2\} \cup \{ceil(1/p), ..., 1\} \cup \emptyset = \{1, 2\}$ matter. If it is the last term, then connections of order $k \in \{1, ..., q-1\}$ matter.

[30] As the proof of Proposition 4 illustrates, dynamics at any horizon h are determined as a linear combination of walks of different orders, with order k multiplied by products of α_ls such that the sum of their exponents is k.

[31] Note that by the discussion above, for stock variables, this proposition applies more generally for $q \in \mathbb{Q}_{++}$ if v_τ is Gaussian.

Contemporaneous Innovation Transmission Via Bilateral Links The NVAR abstracts from contemporaneous network interactions, which feature prominently in the econometric literature on Spatial Autoregressive (SAR) models and the macroeconomic literature on production networks. In that case, the implicit assumption is that connections of all order materialize in any given period of observation:

$$\tilde{y} = A\tilde{y} + \varepsilon = (I - A)^{-1}\varepsilon = (A + A^2 + A^3 + ...)\varepsilon \ .$$

Contemporaneous interactions rationalize the cross-sectional comovement among $\{\tilde{y}_i\}_{i=1}^n$ as the network-induced amplification of cross-sectionally uncorrelated, idiosyncratic shocks $\{\varepsilon_i\}_{i=1}^n$. Ultimately, contemporaneous interactions concern shock identification, which is not the focus of the present analysis. Instead, the interest lies in how networks shape innovation transmission over time, regardless of the origin of these innovations.

Contemporaneous links are useful for quantifying overall connectedness via networks, but they are silent on how networks drive dynamics.[32] Nevertheless, models with contemporaneous and lagged network interactions are related. By Proposition 10 in Appendix A.4, the (contemporaneous) response of $\tilde{y}_i$ to a (transitory or persistent) innovation to $\tilde{y}_j$ under contemporaneous interactions is equal to the long-run response of y_{it} to a persistent innovation to y_{jt} under lagged interactions in a corresponding NVAR$(p, 1)$. Specifically, for

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \qquad \text{and} \qquad \tilde{y} = aA\tilde{y} + \varepsilon \ , \ \ a = \sum_{l=1}^{p} \alpha_l \ ,$$

we have

$$lim_{h\to\infty} \left[\frac{\partial y_{t+h}}{\partial u_t} + \frac{\partial y_{t+h}}{\partial u_{t+1}} + ... + \frac{\partial y_{t+h}}{\partial u_{t+h}} \right] = \frac{\partial \tilde{y}}{\partial \varepsilon} = (I - aA)^{-1} \ ,$$

provided the processes are stationary. Both responses are given by element (i, j) of the Leontief inverse $(I - aA)^{-1}$, which is a sufficient statistic for the (long-term or overall, static) cross-sectional comovement of interest. The difference between the two specifications is that, by taking a stance on the time profile of network interactions, y_t contains information on how any such long-term effect materializes over time. In contrast, the timing of interactions is irrelevant if the interest lies only in steady state comparisons rather than full transition dynamics. As shown in Appendix A.4, the same result applies even if only a snapshot or an average of realizations from such an NVAR$(p, 1)$ is observed every q periods, $q \in \mathbb{N}\backslash\{1\}$. If sums are observed, the long-term response is scaled up by q.

[32] At least in absence of further structure, such as provided by a dynamic macroeconomic model with intertemporally linked optimization problems of agents who are impacted by disturbances to $\tilde{y}$. In this case, even though within the same period idiosyncratic shocks travel through the whole network and effects of all order play out, agents can smooth adjustment to these (amplified) shocks over several periods. Even then, networks only amplify dynamics but are not capable of causing dynamics themselves.

Note that the timing of the long-term response to a permanent shock provides evidence on the timing of this impulse-response more generally, regardless of the nature of the shock. This is because for any VAR, the response to a permanent shock is equal to the cumulative response to a temporary shock (to the same variable). Therefore, the fraction of the long-term response which materialized until horizon h is equal to the area under the IRF to a temporary shock until horizon h as a fraction of the total area. As a result, a slow long-term response to a permanent shock implies a persistent response to a temporary shock.

3 Applicability & Inference

The NVAR is applicable in two rather distinct lines of empirical work with cross-sectional time series. On the one hand, it can be used to estimate dynamic peer effects. These could be of interest themselves or as a way of circumventing the requirements on A for identification in SAR models with contemporaneous interactions. The NVAR can also be used for modeling cross-sectional processes more generally, say if interest lies in forecasting. In either case, one might have data on the network A and be willing to condition on it, or one might prefer to infer it from the data.

In Section 3.1, I discuss the estimation of the time profile of network effects α, conditioning on A. In Section 3.2, I discuss joint inference on (α, A), with A identified from dynamic cross-sectional correlations in the data, possibly aided by shrinking towards observed network links.

3.1 Timing of Network Effects

The first part of this section is devoted to the estimation of α in the NVAR$(p, 1)$ from Eq. (2). The second part deals with the case when data is observed at a higher frequency than network interactions take place, i.e. stock variables in the NVAR(p, q) with $q > 1, q \in \mathbb{Q}$ and flow variables in the NVAR(p, q) with $q > 1, q \in \mathbb{N}$.[33] Details are in Appendix B.1.

NVAR$(p, 1)$ The NVAR$(p, 1)$ from Eq. (2),

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ , \quad \alpha = [\alpha_1, ..., \alpha_p]' \in \mathbb{R}^p \ ,$$

[33] The case of stock variables in the NVAR(p, q) with $q^{-1} \in \mathbb{N}\backslash\{1\}$ is subsumed in the first part (with straightforward adjustments), as all relevant realizations of the underlying NVAR$(p, 1)$ are observed, at least if the process is Gaussian. Also, under Gaussianity, the case of stock variables with $q \in (0, 1] \cap \mathbb{Q}, q^{-1} \notin \mathbb{N}$ can be written to fit in the second part. See Section 2.2.2.

can be written as the linear regression

$$y_t = X_t\alpha + u_t \ , \quad \text{or} \quad y_{it} = x_{it}'\alpha + u_{it} \ , \tag{7}$$

where the $n \times p$ matrix X_t summarizes the information in lags 1 to p of y_t using first-order network connections:

$$X_t = \begin{bmatrix} x_{1t}' \\ ... \\ x_{nt}' \end{bmatrix} = \left[Ay_{t-1}, Ay_{t-2}, ..., Ay_{t-p}\right] \ . \tag{8}$$

Because the network A is taken as given, the dependence of X_t on it is suppressed.

As a result, given A, α can be estimated by Least Squares (LS). This yields the following optimization problem:

$$\underset{\alpha}{min} \ \frac{1}{nT}\sum_{t=1}^{T}(y_t - X_t\alpha)'\,\Sigma^{-1}\,(y_t - X_t\alpha) \ , \tag{9}$$

where $\Sigma = \mathbb{V}[u_t]$. This leads to

$$\hat{\alpha}_{LS}|\Sigma = \left[\sum_{t=1}^{T} X_t'\Sigma^{-1}X_t\right]^{-1}\left[\sum_{t=1}^{T} X_t'\Sigma^{-1}y_t\right] \ . \tag{10}$$

Under $\Sigma = I$, we obtain the Ordinary LS (OLS) estimator, which takes the form of a pooled OLS estimator:

$$\hat{\alpha}_{LS}|(\Sigma = I) = \left[\sum_{t=1}^{T} X_t'X_t\right]^{-1}\left[\sum_{t=1}^{T} X_t'y_t\right] = \left[\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}\right] \ .$$

A (feasible) Generalized LS (GLS) estimator can be obtained either as a two-step procedure or by iterating on $\hat{\alpha}_{LS}|\Sigma$ and $\hat{\Sigma}|\alpha = \frac{1}{T}\sum_{t=1}^{T} u_t(\alpha, A)u_t(\alpha, A)'$ until convergence (see Meng and Rubin (1993)).

Assuming i) the model is specified correctly, i.e. $y_{it} = x_{it}'\alpha + u_{it}$, ii) $\mathbb{E}_{t-1}[u_t] = 0$, and iii) the observed network adjacency matrix A_n converges to some limit A in a way so that for all t and for $l, k = 1 : p$, as $n \longrightarrow \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(A_{n,i\cdot}y_{t-l})'\,(A_{n,i\cdot}y_{t-k}) \longrightarrow \mathbb{E}\left[(A_{i\cdot}y_{t-l})'\,(A_{i\cdot}y_{t-k})\right] \ ,$$

$$\text{and} \quad \frac{1}{n}\sum_{i=1}^{n}(A_{n,i\cdot}y_{t-l})'\,u_{it} \longrightarrow \mathbb{E}\left[(A_{i\cdot}y_{t-l})'\,u_{it}\right] \ ,$$

we get that $\hat{\alpha}_{OLS}$ is consistent for α as $n \longrightarrow \infty$. If in addition i) $\mathbb{E}_{t-1}[u_{it}u_{is}] = \sigma^2$ if $t = s$ and zero otherwise, and ii) $\forall\ t$ and $l, k = 1 : p$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(A_{n,i\cdot}y_{t-l}\right)' u_{it} \Rightarrow N\left(\mathbb{E}\left[\left(A_{i\cdot}y_{t-l}\right)' u_{it}\right], \mathbb{V}\left[\left(A_{i\cdot}y_{t-l}\right)' u_{it}\right]\right)$ as $n \longrightarrow \infty$, then

$$\sqrt{n}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \frac{\sigma^2}{T}\mathbb{E}[x_{it}x_{it}']^{-1}\right) \quad \text{as } n \longrightarrow \infty \ .$$

Under $T \longrightarrow \infty$, $\hat{\alpha}_{OLS}$ is consistent if i) the model is specified correctly, i.e. $y_t = X_t\alpha + u_t$, ii) $\mathbb{E}_{t-1}[u_t] = 0$, and iii) y_t is ergodic and strictly stationary. Assuming in addition $\mathbb{E}_{t-1}[u_t u_t'] = \Sigma$ yields

$$\sqrt{T}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \mathbb{E}[X_t'X_t]^{-1}\mathbb{E}[X_t'\Sigma X_t]\mathbb{E}[X_t'X_t]^{-1'}\right) \quad \text{as } T \longrightarrow \infty \ .$$

The asymptotic variance simplifies to $\sigma^2\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}']^{-1}$ if $\Sigma = \sigma^2 I$. It further simplifies to $\frac{\sigma^2}{n}\mathbb{E}[x_{it}x_{it}']^{-1}$ if $\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}'] = n\mathbb{E}[x_{it}x_{it}']$. Assuming either the conditions under $T \longrightarrow \infty$ asymptotics, including these latter two, or the conditions under $n \longrightarrow \infty$ asymptotics, we get

$$\sqrt{nT}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \sigma^2\mathbb{E}[x_{it}x_{it}']^{-1}\right) \quad \text{as } n, T \longrightarrow \infty \ .$$

NVAR(p, q), with $q > 1$ As discussed in Section 2.2.2, if x_τ follows an NVAR$(p, 1)$ and a snapshot of x_τ is observed every q periods, i.e. $\{y_t\}_{t=1}^{T} = \{x_{tq}\}_{t=1}^{T}$, with $q > 1, q \in \mathbb{Q}$, the dynamics of y_t are represented by the state space system in Eq. (4):

$$\begin{aligned} x_\tau &= \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ , \quad \tau = 1 : T_\tau \ , \\ y_{\tau/q} &= x_\tau \quad \text{if } \tau/q \in \mathbb{N} \ , \end{aligned}$$

and similarly if x_τ and y_t are flow variables and $\{y_t\}_{t=1}^{T} = \{x_{tq} + ... + x_{(t-1)q+1}\}_{t=1}^{T}$ for $q > 1, q \in \mathbb{N}$ is observed (see Eq. (6)).

In principle, an estimator for α could be obtained via the Expectation-Maximization (EM) algorithm (see Appendix B.1). However, point identification of α is not guaranteed. For example, under $q = 2$ and $p = 1$, the observed process follows

$$y_t = \alpha_1^2 A^2 y_{t-1} + \eta_t \ , \quad \text{with } \eta_t = v_{2t} + \alpha_1 A v_{2t-1} \ ,$$

which suggests that α_1 is identified only up to sign. Calculations in Appendix B.1 suggest that for general p, under $q = 2$, the vector $(\alpha_1, \alpha_3, ...)$ is identified only up to sign. For other q, characterization of the identified set remains elusive. As discussed in Section 2.2.2, the observed process can be approximated arbitrarily well as a VARMA with coefficient-matrices equal to polynomials of $\{\alpha_l A\}_{l=1}^{p}$. This suggests that the mapping between parameters in the process for $\{y_t\}_{t=1}^{T}$ and α (the parameters in the process for $\{x_\tau\}_{\tau=1}^{T_\tau}$) is not bijective, just as in the problem of estimating continuous time models using discrete time data (see e.g. Phillips (1973)).

Even if several α rationalize the data equally well, nevertheless useful inference could be drawn, in particular if interest lies only in properties of the process at observational frequency. Furthermore, the whole set of identified α can be obtained using Bayesian methods, implemented with the Gibbs sampler of Carter and Kohn (1994). With appropriate assumptions on the distribution of v_τ and priors for α and the unobserved states in $x_{1:T_\tau}$, the marginal posterior $p(\alpha|Y_{1:T})$ equals the objective function from the EM algorithm. As a result, the set of frequentist point estimates from the EM algorithm can be obtained as the set of modes of $p(\alpha|Y_{1:T})$. For example, under $v_\tau \sim N(0, I)$, $p(\alpha)p(x_{1:T_\tau}) \propto c$, modes of $p(\alpha|Y_{1:T})$ are OLS estimates. Appendix B.2 contains more details.

3.2 Joint Estimation: Network & Effect-Timing

No matter whether one is interested in estimating dynamic peer effects or approximating dynamics of cross-sectional processes, in many cases network data might be missing or it appears restrictive to condition on the available data. This section discusses joint estimation of (α, A). Again the first part deals with the estimation of an NVAR$(p, 1)$, while the second part discusses the case of an NVAR(p, q) with $q > 1$.

NVAR$(p,1)$ The NVAR$(p, 1)$ from Eq. (2),

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ , \quad \alpha = [\alpha_1, ..., \alpha_p]' \in \mathbb{R}^p \ ,$$

can be written as the linear regression

$$y_t = A z_t + u_t \ , \quad \text{or} \ \ Y = Z A' + U \ , \tag{11}$$

whereby $z_t = \sum_{l=1}^p \alpha_l y_{t-l} = \tilde{X}_t \alpha$ with $\tilde{X}_t = \left[y_{t-1}, y_{t-2}, ..., y_{t-p} \right]$, and the $T \times n$ matrices Y, Z and U stack y_t, z_t and u_t along rows, respectively. Note that $X_t = A\tilde{X}_t$, as defined in Eq. (8). To simplifty notation, I suppress the dependence of X_t on A and that of z_t on α.

To render (α, A) jointly identified despite their multiplicative interaction, we can normalize $||\alpha||_1 = 1$. Alternatively, fix one α_l and drop it from α, with appropriate redefinitions of y_t and $\tilde{X}_t$ (and z_t and X_t). The latter normalization facilitates asymptotic analysis, but requires $\alpha_l \neq 0$ in the true data generating process.[34] In practice, the former is easier to implement (see below).

[34] Also, if $a_{ij} \geq 0$ is imposed, it requires knowledge on the sign of α_l.

Least Squares (LS) estimation of (α, A) with a Ridge-penalty to shrink A to some matrix B yields the following optimization problem:

$$\underset{\alpha,A}{min}\ \frac{1}{nT}\sum_{t=1}^{T}\left(y_t - A\tilde{X}_t\alpha\right)'\Sigma^{-1}\left(y_t - A\tilde{X}_t\alpha\right) + \tilde{\lambda}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\ , \tag{12}$$

where $\Sigma = \mathbb{V}[u_t]$ and b_{ij} is element (i,j) in B. This leads to $\hat{\alpha}_{LS}|(A,\Sigma)$ as in Eq. (10). Also,

$$\hat{A}_{LS}|(\alpha,\Sigma) = [Y'Z + \lambda\Sigma B]\,[Z'Z + \lambda\Sigma]^{-1}\ , \tag{13}$$

with $\lambda = nT\tilde{\lambda}$, which is equal to the mode of the conditional posterior $p(A|Y,\alpha,\Sigma)$ under the priors $a_{ij} \sim N(b_{ij}, \lambda^{-1})$ and assuming $u_t \overset{iid}{\sim} N(0,\Sigma)$ (see Appendix B.2). As $\lambda \longrightarrow \infty$, $\hat{A}_{LS}|(\alpha,\Sigma) \longrightarrow B$ and we condition the analysis on the network adjacency matrix B, just as in Section 3.1. As $\lambda \longrightarrow 0$, $\hat{A}_{LS}|(\alpha,\Sigma)$ is inferred from the data alone. No domain restrictions on A are imposed because any solution $(\hat{\alpha}, \hat{A})$ can be rescaled to yield $\hat{a}_{ij} \in [-1,1]\ \forall\, i,j$, so that A can be interpreted as a network. To enforce $\hat{a}_{ij} \in [0,1]$ even under low λ, $a_{ij} \geq 0$ must be imposed.

Under OLS (i.e. setting $\Sigma = I$ in the above expressions), the unconditional estimator $(\hat{\alpha}_{OLS}, \hat{A}_{OLS})$ is obtained by iterating on $\hat{\alpha}_{OLS}|A$ and $\hat{A}_{OLS}|\alpha$ until convergence, as outlined in Meng and Rubin (1993). To impose the normalization $||\alpha||_1 = 1$, rescale $\hat{\alpha}_{OLS}|A$ appropriately before proceeding to $\hat{A}_{OLS}|\alpha$ in the iteration. For GLS, a third iteration step is added for $\hat{\Sigma}|(\alpha,A) = \frac{1}{T}\sum_{t=1}^{T} u_t(\alpha,A)u_t(\alpha,A)'$.[35] From a forecasting point of view, it is optimal to select λ by adding an iteration step and setting

$$\lambda|A = \left[\frac{1}{n}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\right]\ .$$

This is the mode of $p(\lambda|Y,\alpha,\Sigma)$ under a hierarchical Bayes model with a Uniform prior for λ, treated as a hyperparameter (see Appendix B.2). As discussed in Giannone et al. (2015), the shape of the posterior of λ coincides with that of the marginal likelihood, a measure of out-of-sample forecasting performance.

A Ridge- rather than Lasso-penalty is chosen only for analytical convenience. Under a Lasso-penalty for $A - \tilde{\lambda}\sum_{i,j=1}^{n}|a_{ij} - b_{ij}|$ – an analytical expression can only be obtained under OLS, imposing $a_{ij} \geq 0$ and selecting links to zero; $b_{ij} = 0$. And even then, the expression only shows the conditional estimator for a column in A, with elements $\hat{a}_{ij,OLS}|(\alpha, A_{i,-j})$.[36]

[35] Analogously, Bayesian inference could be implemented by Gibbs sampling, i.e. iteratively drawing from the conditional posteriors for the three quantities α, A and Σ to obtain their joint posterior. See Appendix B.2. However, for large n, this could be prohibitively costly.

[36] $A_{i,-j}$ denotes all elements in row i of A except that in column j. Under $a_{ij} \geq 0$ and $b_{ij} = 0$ but with GLS, $\hat{A}_{LS}|(\alpha,\Sigma)$ is the mode of a truncated Normal distribution, which is costly to evaluate even for moderate n. See Appendix B.2.

Let $\theta = (\alpha, A)$. Assuming i) y_t is ergodic and strictly stationary, ii) $\mathbb{E}[X_t' X_t]$ and $\mathbb{E}[z_t z_t']$ are of full rank (at θ_0), and iii) $\tilde{\lambda}_{n,T} = o(1)$, we get that, as $T \longrightarrow \infty$, $\hat{\theta}_{OLS}$ is consistent for θ_0 defined by

$$\alpha_0 | A_0 = \mathbb{E}[X_t' X_t]^{-1} \mathbb{E}[X_t' y_t] \ , \quad A_0 | \alpha_0 = \mathbb{E}[y_t z_t'] \mathbb{E}[z_t z_t']^{-1} \ .$$

Asymptotic Normality is obtained if in addition i) the model is specified correctly, i.e. $y_t = A\tilde{X}_t \alpha + u_t$, ii) $\mathbb{E}_{t-1}[u_t] = 0$, iii) $\mathbb{E}_{t-1}[u_t u_t'] = \Sigma$ and iv) $\tilde{\lambda}_{n,T} = o(T^{-\frac{1}{2}})$.[37]

NVAR(p,q) with $q > 1$: As in the estimation of $\alpha | A$, an estimator for (α, A) under an NVAR(p, q) with $q > 1$ can be obtained using the EM algorithm. This amounts to adding one data-augmentation step to the above iterations, in which $\hat{x}_{1:T_\tau} \equiv \mathbb{E}[x_{1:T_\tau} | y_{1:T}, \alpha, A]$ is obtained from the Kalman Smoother. The joint estimation of (α, A) possibly exacerbates the identification problem discussed in Section 3.1. However, as before, its extent can be assessed by using the proportionality of the objective function in Eq. (12) and posterior density $p(\alpha, A|Y)$ and inspecting (the set of) modes of the latter.[38] Furthermore, in some cases, only properties of the process at observational frequency are of interest (e.g. for forecasting y_t at observational frequency).

4 Input-Output Links & Sectoral Price Dynamics

How do price innovations propagate across sectors in an economy? Given an observed price increase in, say, energy-related sectors, what is the expected path of aggregate prices? How do we expect prices in another sector to react? With sectors linked through an input-output network, the answers depend on the positions of the shocked (and responding) sector in the network as well as on the velocity at which a shock travels through the network.

The literature so far has used the assumption of contemporaneous transmission of idiosyncratic shocks to document that input-output linkages can rationalize the sectoral comovement of prices at a given point in time. In the following, I use the NVAR to analyze the dynamic aspects of this comovement. While existing literature shows that a sector's position in the supply chain network determines the strength of its effects on prices in other sectors and aggregates, the present analysis aims at establishing whether this is true also for the timing of these effects.

Consistent with the literature on granular origins of business cycles, I consider the propagation of relative price changes induced by supply-side TFP shocks, as motivated by an

[37] The same applies analogously under a Lasso-penalty for A, although no analytical expression for the conditional estimator can be found in that case. Under $a_{ij} \geq 0$ (in which case the Lasso-estimator for $A|\alpha, \Sigma$ can be found when selecting to $B = 0$), only consistency goes through as A_0 is likely not interior. See Appendix B.2.

[38] See the discussion in Appendix B.1, which is straightforwardly extended to the joint estimation of (α, A).

input-output economy in the Real Business Cycle (RBC) tradition, which I discuss in Section 4.1. With simple extensions to the baseline model, sectoral prices and output in this economy evolve (at some model-frequency) according to an NVAR$(p, 1)$. The analysis provides some intuition on the types of structural models that map into an NVAR, and it illustrates the determinants of dynamics in the NVAR using an actual application. However, more work needs to be done to establish whether there is indeed a role for lagged input-output linkages in driving dynamics of output and prices.

After theoretically motivating the analysis in Section 4.1, I discuss the data in Section 4.2 and the estimation procedure in Section 4.3. Results are presented in Section 4.4.

4.1 Theory

This section extends a benchmark input-output economy by introducing time lags in input-output conversion. This yields sectoral prices that evolve (approximately) as an NVAR$(p, 1)$. The derivation here is based on Carvalho and Tahbaz-Salehi (2019), who discuss a static input-output economy. Details are provided in Appendix C.1.

Assume there are n sectors, in each of which a representative firm produces a differentiated good i by combining labor services l_{it} and goods produced by other sectors j, $\{x_{ijt}\}_{j=1}^n$, using a Cobb-Douglas production function. Firms maximize profits taking prices as given. The profits of firm i in period t are

$$\Pi_{it} = p_{it} y_{it} - w_t l_{it} - \sum_{j=1}^{n} p_{jt} x_t^{ij} \ ,$$

$$y_{it} = z_{it} l_{it}^{b_i} \prod_{j=1}^{n} x_{ijt}^{a_{ij}} \ , \quad b_i > 0 \ , \quad a_{ij} \geq 0 \ , \quad b_i + \sum_{j=1}^{n} a_{ij} = 1 \ ,$$

where z_{it} denotes TFP in sector i and w_t is the price of labor. x_t^{ij} denotes the amount of good j purchased in period t. As discussed below, it can differ from the amount of good j used in the production at time t, x_{ijt}. Under perfect competition and constant returns to scale (CRS) Cobb-Douglas production functions, prices are entirely determined by supply. Nevertheless, to show that the following results hold in general equilibrium and to obtain results for output dynamics, I assume there is a representative household which supplies one unit of labor inelastically and exhibits log-preferences over the n goods:

$$u(\{c_{it}\}_{i=1}^n) = \sum_{i=1}^{n} \gamma_i \, ln(c_{it}/\gamma_i) \ , \quad \sum_{i=1}^{n} \gamma_i = 1 \ .$$

Different assumptions on the timing of input-output conversion lead to different dynamics of sectoral prices and output in this economy. In the following, I focus on prices and relegate further results, including output dynamics, to Appendix C.1. Let $x^{ij}_{t,t-h}$ denote the use of good j purchased at time $t-h$ in the production of good i at time t.

Most of the literature assumes that inputs are converted into outputs in the same period when they are produced and purchased, i.e. $x_{ijt} = x^{ij}_{t,t} = x^{ij}_{t}$. This leads to a static economy with contemporaneous network effects. We obtain the following equation for sectoral prices $p_t = (p_{1t}, ..., p_{nt})'$ as a function of sectoral productivities $z_t = (z_{1t}, ..., z_{nt})'$ and input-output relations summarized by the input-output matrix A:

$$\tilde{p}_t = k^p + A\tilde{p}_t + \varepsilon_t \ ,$$

where $\tilde{p}_t = ln(p_t/w_t)$, $\varepsilon_t = -ln(z_t)$ and k^p is a vector of constants. This equation fully characterizes prices in this economy, whereby wages are taken as the numéraire.

To analyze the cases of lagged input-output conversion, I additionally assume perfect foresight. If, as in Long and Plosser (1983), it takes one period to convert purchased inputs into output, i.e. $x_{ijt} = x^{ij}_{t,t-1} = x^{ij}_{t-1}$, we obtain that sectoral prices approximately follow an NVAR(1,1):

$$\tilde{p}_t = k^{p1}_t + A\tilde{p}_{t-1} + \varepsilon_t \ ,$$

where $k^{p1}_t = k^{p1} - (\iota - b) ln\left(G^w_t\right)$. Thereby, k^{p1} is a vector of constants, ι is a vector of ones, $b = (b_1, ..., b_n)'$ contains sectoral labor shares and $G^w_t = w_t/w_{t-1}$ is wage growth in period t. This process only deviates from an NVAR$(1,1)$ to the extent that the numéraire w_t changes in value. This result can easily be extended to input-output conversion at single lags of arbitrary length; if it takes p periods to convert inputs into output, $\tilde{p}_t$ approximately follows an NVAR$(p,1)$ where the coefficients in front of all but the pth lag are zero.

As shown in Appendix C.1, this economy leads to almost the same steady state as the above economy with contemporaneous network interactions.[39] However, while the latter is always in steady state, this economy is dynamic and after a disturbance to ε_t only asymptotically converges to the steady state. For empirical analyses one has to take a stance on what a period in this model signifies (relative to an observational period in the data).

An NVAR where the last several lags matter for dynamics is obtained if firms use inputs purchased in the past several periods in their production at time t. To model this case, I assume that x_{ijt} aggregates quantities of input j purchased at different periods in the past using a Constant Elasticity of Substitution (CES) aggregator. To keep the exposition tractable, let x_{ijt} include amounts of good j bought at $t-1$ and $t-2$, $x^{ij}_{t,t-1}$ and $x^{ij}_{t,t-2}$.[40] This means that a good perishes after two periods, at least with regard to its suitability as an input in production. An extension to arbitrary lengths p is straightforward. As in

[39] Differences vanish as the discount factor $\beta \longrightarrow 1$.

[40] Therefore, the amount of good j purchased at time t can be used in production at periods $t+1$ and $t+2$: $x^{ij}_t = x^{ij}_{t+1,t} + x^{ij}_{t+2,t}$.

the Long and Plosser (1983) economy above, the presumption is that storage is done by the buyer. We then have

$$x_{ijt} = \left[\eta_1 (x^{ij}_{t,t-1})^r + \eta_2 (x^{ij}_{t,t-2})^r\right]^{1/r} \quad , \quad \eta_1, \eta_2 \geq 0 \ , \quad \eta_1 + \eta_2 = 1 \ , \quad r > 0 \ .$$

In the Cobb-Douglas case $r \longrightarrow 0$, we obtain that sectoral prices approximately follow an NVAR(2,1):

$$\tilde{p}_t = k^{p2}_t + \eta_1 A \tilde{p}_{t-1} + \eta_2 A \tilde{p}_{t-2} + \varepsilon_t \ ,$$

where $k^{p2}_t = k^{p2} - (\iota - b)\left[\eta_1 ln\left(G^w_t\right) + \eta_2 ln\left(G^w_t G^w_{t-1}\right)\right]$. Again, this relation is only approximate because the numéraire can change in value.[41]

Under a more general elasticity of substitution r, excluding the case of perfect substitutability ($r = 1$), we can derive a similar result by log-linearizing around the steady state. Let a hat denote percentage deviation from steady state, whereby, with slight abuse of notation, $\hat{p}_t$ denotes this deviation for p_t/w_t. We obtain

$$\hat{p}_t = \hat{k}^{p3}_t + \chi_1 \check{A} \hat{p}_{t-1} + \chi_2 \check{A} \hat{p}_{t-2} + \hat{\epsilon}_t \ .$$

In this expression, χ_1, χ_2 are non-negative scalars that sum to one, $\check{A}$ contains scaled bilateral links $a_{ij}/(1 + b_i(1 - r))$ and $\hat{\epsilon}_t$ contains scaled TFP deviations $\hat{\epsilon}_{it} = -\frac{2-r}{1+b_i(1-r)}\hat{z}_{it}$. These scalings vanish as we move towards the case of perfect substitutability, $r \longrightarrow 1$. The vector $\hat{k}^{p3}_t$ is composed of elements $\hat{k}^{p3}_{it} = \frac{1-\phi_i}{\phi_i}\hat{y}_{it} + (1 - b_i/\phi_i)\left[\chi_1 \hat{G}^w_t + \chi_2(\hat{G}^w_t + \hat{G}^w_{t-1})\right]$, with $\phi_i = (1 + b_i(1 - r))/(2 - r)$. Hence, for general elasticities of substitution r, the process of sectoral prices differs from an NVAR(2,1) not only by the extent that the numéraire changes, but also as sectoral output changes. Note that the output-term vanishes as $r \longrightarrow 1$.

To sum up, under general lags in input-output conversion, the log of sectoral prices, now denoted by x_τ, evolves at some model-frequency (approximately) as an NVAR$(p, 1)$,

$$x_\tau \approx \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + \upsilon_\tau \ ,$$

with $\alpha_l \geq 0 \ \forall \ l$ and $\sum_{l=1}^p \alpha_l = 1$. Also, $\sum_j a_{ij} < 1 \ \forall \ i$. These restrictions imply that the process is stationary.[42] In the empirical analysis that follows, I allow the model-frequency to differ from the observational frequency, as discussed in Section 2, and I infer their relation from the data by model selection criteria.

A difference to the (unrestricted) NVAR$(p, 1)$ from Section 2 stands out: the domain restrictions $\alpha_1, ..., \alpha_p \geq 0$ imply that the impulse response to a shock in sector j has the

[41] Note that $G^w_t G^w_{t-1} = \frac{w_t}{w_{t-2}}$ is the wage growth from $t - 2$ to t.

[42] Berman and Plemmons (1979, p. 37) show that for an element-wise nonnegative matrix with row sums strictly smaller than 1, the absolute value of the largest Eigenvalue is strictly less than 1. Stationarity then follows by Proposition 2.

same sign for all units i. There are two reasons for this. First, perfect competition implies that prices equal marginal costs and prevents strategic price setting by firms. Second, Cobb-Douglas production functions imply constant input shares regardless of prices and prevent upstream propagation of price shocks. Note that the model can nevertheless rationalize price movements in opposite directions because in the same period some sectors might experience positive, others negative shocks, while the remaining sectors differ in the extent to which they are impacted by the two owing to different positions in the network.

4.2 Data

To construct the network of sectoral links, I use annual data on input-output matrices provided by the Bureau of Economic Analysis (BEA). Following the theory in Section 4.1 and most of the literature, I simplify the analysis by assuming constant network connections over time. I take the input-output data for 2010, roughly the midpoint of the sample of sectoral Producer Price Indices (PPI) (see below). Due to availability of the latter, I consider the level of 64 mostly three- and four-digit sectors rather than the finer level of around 400 six-digit commodities (NAICS classification). The analysis is restricted to non-farm and non-governmental sectors. Following Acemoglu et al. (2016), links a_{ij} are defined as

$$a_{ij} \equiv \frac{sales_{j \longrightarrow i}}{sales_i} \ ,$$

where $sales_{j \longrightarrow i}$ is the total value of goods and services purchased by sector i from sector j as determined by the corresponding entry in the BEA's "use" table. The value of a_{ij} shows how many dollars worth of output of sector j sector i needs to purchase in order to produce one dollar's worth of its own output.[43]

The corresponding time series of sector-level PPI is obtained from the Bureau of Labor Statistics (BLS). Data availability narrows the analysis to 51 sectors and the time frame January 2005 - August 2022. This includes the Great Recession as well as the COVID-19 recession. More details on the matching of PPI and input-output data are provided in Appendix C.2.

Most of the raw log PPI series show a clear upward trend. To render the series stationary,

[43] As discussed in Appendix C.1, the expression for a_{ij} in steady state changes slightly in economies with different lags of input-output conversion. For example, in the Long and Plosser (1983) economy, the above a_{ij} would need to be multiplied by β^{-1}, the inverse of the discount factor. Under Cobb-Douglas aggregation of inputs purchased in the past two periods, one would need to multiply by $(\alpha_1\beta + \alpha_2\beta^2)^{-1}$. For general CES aggregation, this constant is also a function of the elasticity r. However, these differences in the proper calibration of a_{ij} vanish as $\beta \longrightarrow 1$.

(a) Weighted In-Degrees (b) Weighted Out-Degrees

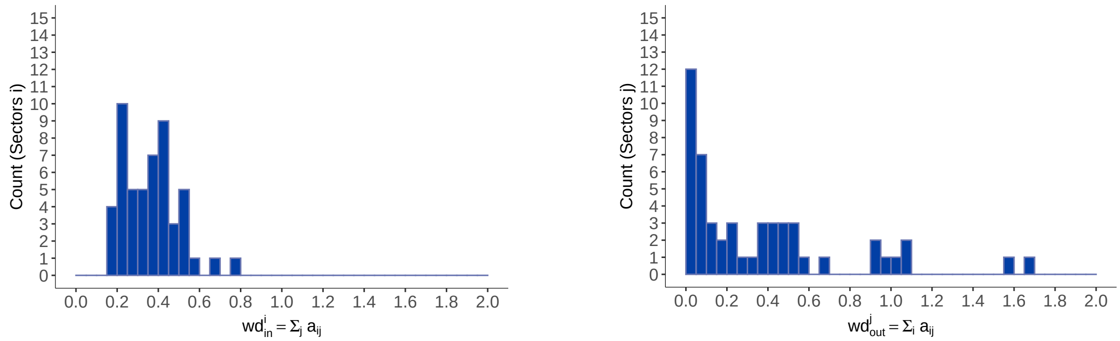


Figure 2: Weighted In-Degrees & Out-Degrees

Notes: The left panel plots weighted in-degrees, equal to the column-wise sums of A, which show the differing reliance on intermediate inputs across sectors. The right panel plots weighted out-degrees, equal to the row-wise sums of A, which show the differing importance of a sector as a supplier to other sectors in the economy.

I estimate and subtract a linear trend and a seasonality component.[44] In the theoretical model, any time trends in sectoral prices are given by idiosyncratic trends in sectoral TFP levels amplified by the network. However, for these trends the exact timing of network effects is irrelevant, just as it is irrelevant for the steady state. Therefore, given the goal of the present analysis, no information is lost by subtracting time trends.

The majority of links in the network are weak. Even though the fraction of non-zero links is 73.55%, only 16.88% are above 0.01. Nevertheless, and as expected at this level of aggregation, this network density is much higher than the 3% reported for the finer level of 417 sectors in Carvalho (2014). As illustrated in the left panel of Fig. 2, weighted in-degrees, $wd_i^{in} \equiv \sum_j a_{ij}$, lie below 1 for all sectors, as posited by theory. The heterogeneity in this statistic across sectors shows that they rely to different extent on intermediary inputs in production. The right panel shows the weighted out-degrees, $wd_j^{out} \equiv \sum_i a_{ij}$, and illustrates that most sectors are specialized input-suppliers, while there are also a few general-purpose suppliers. Even though only 16.88% of links are non-zero, the average distance in the network is 2.41. This means that each sector is on average 1.4 in-between suppliers away from other sectors. The longest distance, or diameter of the network, is 7, which means that it takes at most 6 in-between suppliers for a sector to reach another sector. Appendix C.2 contains more details on input-output data.

The left panel in Fig. 3 depicts the raw PPI series for a few sectors. It provides evidence of considerable heterogeneity in price dynamics across sectors, even disregarding the highly

[44] Given the raw series of the natural logarithm of PPI in sector i, p_{it}, I estimate

$$p_{it} = \beta_i t + \sum_{m=1}^{12} \gamma_{im} \mathbf{1}\{\text{observation } t \text{ is in month } m\} + e_{it} \ ,$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. In turn, I set $y_{it} = \hat{e}_{it}$ and base the subsequent analysis on y_t.

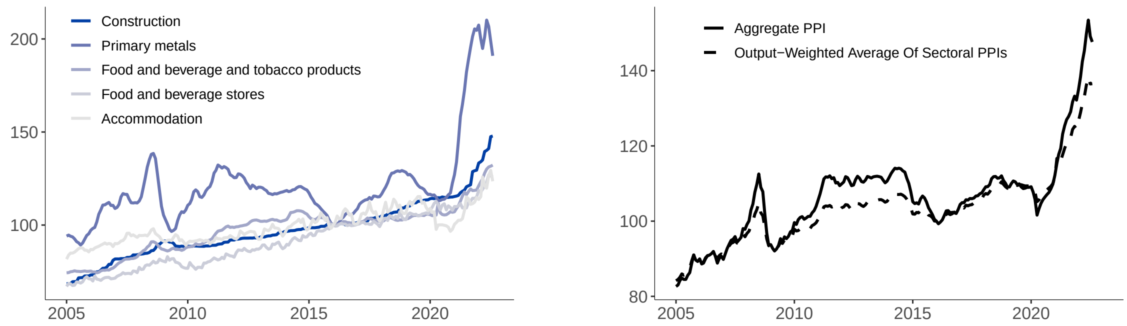


Figure 3: Sectoral and Aggregate PPI

Notes: The left panel shows the raw PPI series for a few selected sectors. The right panel compares the aggregate PPI from the FRED Database and the output-weighted average of PPIs of sectors included in the analysis.

volatile energy-sectors. The mean, standard deviation and range of sectoral PPI changes can be found in Table A-1. Oftentimes, studies on production networks are interested in implications for an aggregate variable, given by some weighted sum of the same variable measured at the cross-sectional level. The right panel of Fig. 3 shows that an output-weighted average of sectoral PPIs included in the analysis replicates the actual aggregate PPI fairly well, despite the fact that some sectors are excluded due to data limitations.[45] Aggregate PPI shows a clear upward trend, with a smaller spike around the Great Recession as well as a very pronounced spike in the aftermath of the COVID-19 recession.[46]

Suggestive evidence that network proximity does not only have implications for the contemporaneous, cross-sectional correlation of inflation across sectors, but also for dynamics is provided in Fig. 4. The lightest-blue line plots the contemporaneous correlation of prices in two sectors against their distance(s). It reproduces for prices the finding in Carvalho (2014) that sectoral comovement decreases with the distance between sectors, although this relationship is much less pronounced at the higher level of disaggregation analyzed here. However, it is not only the contemporaneous comovement between sectors that decreases with distance, but also the comovement of sector i's inflation with lagged values of sector j's inflation is declining with the distance from sector i to sector j. This is illustrated by the remaining lines in Fig. 4, which show this correlation for lags ranging from one to twelve months in darker shades of blue. In fact, the downward slope is more pronounced for higher lags.[47]

[45] The aggregate PPI is obtained from the FRED database of the Federal Reserve Bank of St. Louis. Weights are constructed using sectoral output in 2010.

[46] The latter is included in the analysis because it contains potentially valuable information on how price shocks transmit through the input-output network.

[47] Note that Fig. 4 plots mean correlations by distance and masks plenty of heterogeneity across sector-pairs.

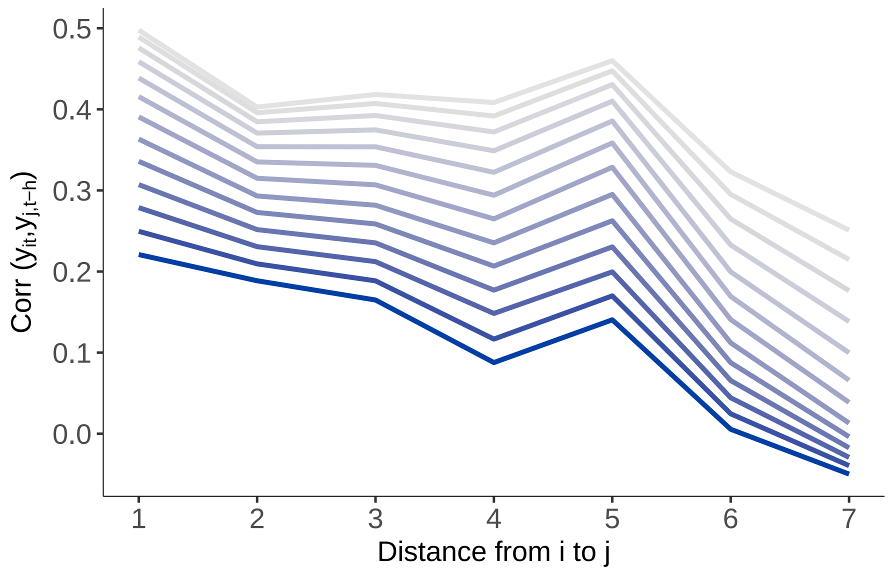


Figure 4: Network Distance and the Correlation of Sectoral Inflation

Notes: The figure plots the average correlation of sectoral prices for different distances between them. The lightest blue line refers to contemporaneous correlations. Darker lines show the average correlation of a sector i with lagged values of a sector j as a function of the distance from i to j. Lags range from 0 to 12 months. The series refer to de-trended and de-seasonalized log PPIs.

4.3 Estimation

The theoretical analysis in Section 4.1 suggests that sectoral prices, at some model-frequency, evolve (approximately) as an NVAR$(p, 1)$,

$$x_\tau = \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + v_\tau \ .$$

Relative to the unconstrained estimation of $\alpha|A$ treated in Section 3.1, the model from Section 4.1 features $\alpha_l \geq 0 \ \forall \ l$ and $\sum_{l=1}^{p} \alpha_l = 1$, which requires some adjustments to the estimation procedure.

Because of these restrictions, I drop α_p from α and impose the domain restrictions $\alpha_l \in [0, 1]$ for $l = 1 : p-1$ and $\sum_{l=1}^{p-1} \alpha_l \leq 1$. In turn, $\alpha_p = 1 - \sum_{l=1}^{p-1} \alpha_l$ with $\alpha_p \in [0, 1]$. Consistent with the literature on granular origins of business cycles, I assume that $v_{i\tau}$ is uncorrelated across i and τ with $\mathbb{E}[v_\tau] = 0$ and $\mathbb{V}[v_\tau] = \Sigma$, $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$. Furthermore, I assume Normality of v_τ and consider Maximum Likelihood (ML) estimation of $\theta = (\alpha, \sigma)$, where $\alpha = (\alpha_1, .., \alpha_{p-1})$ and $\sigma = (\sigma_1, ..., \sigma_n)$.

I allow the frequency of network interactions to differ from the frequency of observation and infer their relation from the data by model selection criteria. As in Section 2.2.2, let $\{y_t\}_{t=1}^T = \{x_{qt}\}_{t=1}^T$ denote the observed series. I consider $q = \frac{1}{3}, \frac{1}{2}, 1, 2, 4,$ which under monthly observations corresponds to quarterly, bi-monthly, monthly, bi-weekly and weekly network interactions, respectively.

Regardless of the value for q, the ML estimator (MLE) $\hat{\theta}_{ML}$ cannot be obtained analytically because of the non-trivial domain restrictions for α. Additionally, for $q > 1$, the likelihood is a nonlinear function of θ. As a result, obtaining $\hat{\theta}_{ML}$ using nonlinear optimization would be inefficient and would suffer from local optima issues. Instead, I implement the MLE using Bayesian methods. Under a prior distribution $p(\theta)$ proportional to a constant, the posterior $p(\theta|Y)$ is proportional to the likelihood $p(Y|\theta)$:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \propto p(Y|\theta)p(\theta) \propto p(Y|\theta) \ .$$

Therefore, $\hat{\theta}_{ML}$ is equal to the posterior mode. The posterior can be obtained efficiently using the Sequential Monte Carlo (SMC) algorithm.[48] I use independent, Uniform priors for $\{\alpha_l\}_{l=1:p-1} \in [0,1]^{p-1}$, truncated to satisfy the additional domain restriction $\sum_{l=1}^{p-1} \alpha_l \leq 1$. The resulting distribution is derived in Appendix C.3. The priors for σ_i are also independent Uniform distributions, ranging from zero to upper bounds large enough to ensure that the domain encompasses $\hat{\sigma}_{i,ML}$.

4.4 Results

Table 1 reports the Marginal Data Density (MDD) for different specifications of the NVAR. The values for q along rows refer to quarterly, bi-monthly, monthly, bi-weekly and weekly network interaction frequencies, respectively. The values for p in the columns indicate how many of up to six past months matter for dynamics. The most preferred specification features monthly network interactions and lags up to six months. Model selection according to the Bayesian or Akaike Information Criteria lead to the same conclusion (see Table A-2). The following analysis is based on this preferred NVAR$(6,1)$.

Table 2 reports the estimation results for α. The first column shows the MLE, approximated by the Maximum A-Posteriori (MAP) estimator, i.e. the posterior draw (particle in the SMC algorithm) with the highest likelihood. It is very close to the posterior mean, reported in the second column. The 95% Bayesian Highest Posterior Density (HPD) sets together with the peaked marginal posteriors shown in Fig. A-4 illustrate that α is estimated very precisely. This is not surprising as there are $nT = 51 \cdot 206 = 10,506$ observations and only $n + p - 1 = 56$ parameters.

[48] See Herbst and Schorfheide (2015) for a general discussion of the SMC algorithm and Appendix C.3 for more details on its implementation for this application. I choose it over alternative posterior sampling techniques because it is parallelizable, allows for an effective tuning of the sampling accuracy and recent advances show how to speed up its computations by using a well-designed proposal density.

Table 1: Model Selection: Log MDD

		p					
		$1q$	$2q$	$3q$	$4q$	$5q$	$6q$
q	$1/3$			19079			19044
	$1/2$		19384		18768		18690
	1	20153	20056	19675	19879	18899	20218
	2	17546	19570	19248	20142	18662	19636
	4	18517	19808	19754	19655	18904	19301

Notes: The table shows values for the natural logarithm of the Marginal Data Density (MDD) across model specifications. The values for q (from top to bottom) refer to quarterly, bi-monthly, monthly, bi-weekly and weekly network interactions, respectively, while $p = mq$ implies that the last m months matter for dynamics.

Table 2: Estimation Results: α

	MLE	Mean	Low	High
α_1	0.1550	0.1557	0.1370	0.1745
α_2	0.3460	0.3382	0.3168	0.3605
α_3	0.2816	0.2865	0.2644	0.3129
α_4	0.0915	0.0991	0.0785	0.1174
α_5	0.1045	0.0975	0.0837	0.1135

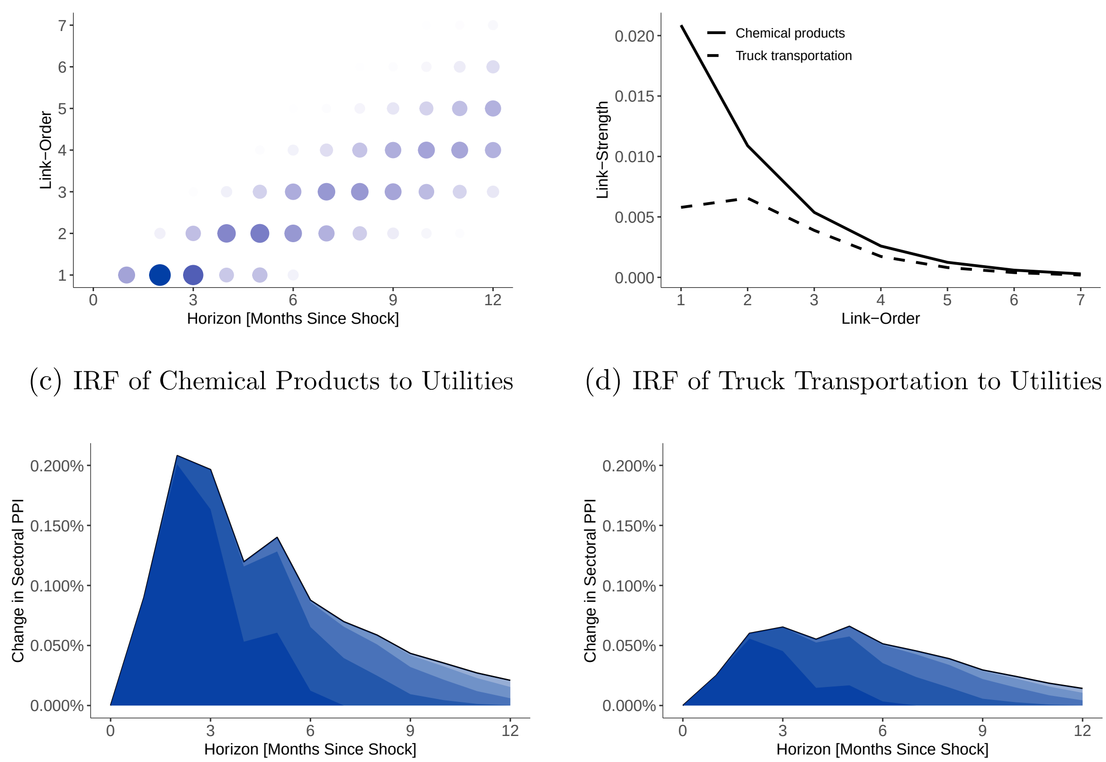
Notes: The first column shows the Maximum Likelihood or Maximum A-Posteriori (MAP) Estimator, the second refers to the posterior mean, and Low and High report the bounds of the 95% Bayesian HPD credible sets.

The dynamics of y_t can be summarized by impulse response functions (IRF). By Section 2.2, under an NVAR$(6,1)$, the impulse response of y_t at horizon h comprises supply chain connections of order $k \in \underline{k} : h$, with $\underline{k} = ceil(h/6)$:

$$\frac{\partial y_{i,t+h}}{\partial u_{j,t}} = \left[\frac{\partial y_{t+h}}{\partial u_t}\right]_{ij} = c^h_{\underline{k}}(\alpha)\left[A^{\underline{k}}\right]_{ij} + ... + c^h_h(\alpha)\left[A^h\right]_{ij} \ . \tag{14}$$

The coefficients $\{c^h_k(\alpha)\}_{k=\underline{k}:h}$ are functions of α and show the importance of (upstream) supply chain connections of different order for the response of sectoral prices at any one horizon h. As the present analysis abstracts from hetereogeneity in α, these coefficients are constant across time and sector-pairs.

Fig. 5 illustrates this composition of impulse responses. The dots in the top left panel



(a) Relevance of Link-Orders Across Horizons (b) Input-Output Links to Utilities Sector (c) IRF of Chemical Products to Utilities (d) IRF of Truck Transportation to Utilities

Figure 5: Impulse Responses: Transmission of Price Shocks via Supply-Chain Links

Notes: The top left panel shows the importance of different connection-orders for shock transmission as a function of the time elapsed since a shock took place. The top right panel shows the supply chain connections of different order from the sectors "Chemical Products" and "Truck Transportation" to the utilities sector, and the bottom panels show their resulting IRFs to an increase in the price of utilities by one standard deviation.

depict the coefficients $\{c_k^h\}_{k=\underline{k}:h}$ with horizons h on the x-axis and connection-orders k on the y-axis. Larger values are represented by larger and darker dots. As stated above, under $\alpha_l > 0$ for $l = 1:6$, orders $\underline{k} : h$ matter for propagation at horizon h. Hence, there are $h - \underline{k} + 1$ dots aligned vertically at horizon h. As time passes, a shock spreads through the network and reaches more distant nodes. However, it is the exact values of $\{\alpha_l\}_{l=1:6}$ that determine the exact width and speed of this propagation. This is illustrated by the differing sizes and colors of the dots.

The top right panel in Fig. 5 shows the strength of network connections of different order from the sectors "Chemical Products" and "Truck Transportation" to the sector "Utilities", respectively. The former sector is more dependent on utilities as a supplier than the latter, as evidenced by stronger network connections, in particular of first and second order. As Eq. (14) above makes clear, such network-connections are the second building block of impulse responses in the NVAR.

The lower panels of Fig. 5 illustrate the resulting impulse responses. The different shades of blue depict the individual terms $c_k^h(\alpha)\left[A^k\right]_{ij}$, which show the contribution of network-

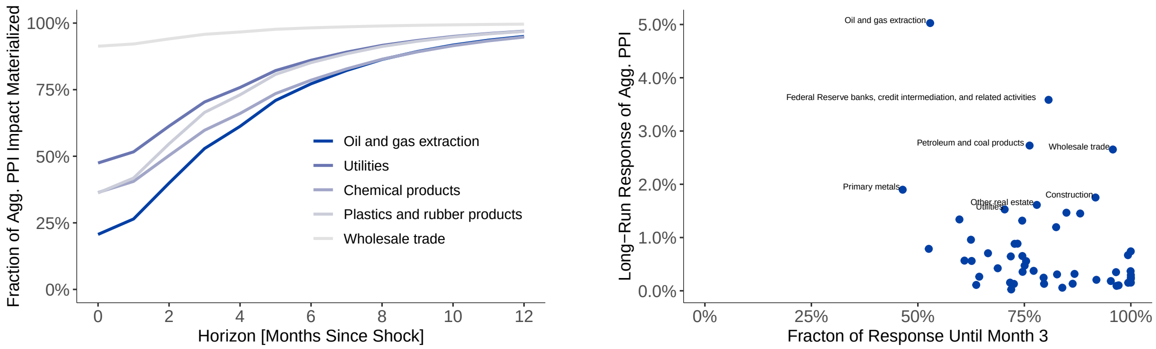


Figure 6: Size and Timing of Aggregate PPI Response to Sectoral Shocks

Notes: The left panel shows the time profile of the effect of sectoral price disturbances on aggregate PPI for a few selected sectors. The right panel relates the strength of the effect on aggregate PPI to its timing. The shock sizes are equal to one standard deviation of the respective sectoral disturbance.

connections of order k to the impulse-response of i to j at horizon h. Darker shades refer to network connections of lower order. As a result of its stronger network-connections to the utilities sector, the price of chemical products reacts more strongly to a one-standard deviation increase in the price of utilities than does the price of truck transportation. The price of chemicals rises quickly and peaks after two months. In contrast, the price of truck transportation increases slowly and remains slightly elevated without a noticeable peak. It is in particular the direct and second-order supply-chain connections that make up the difference between the two responses, in line with the top right panel. Longer-term responses are driven by higher-order connections and after nine months they are of similar size for the two sectors as they share similarly strong higher-order connections to the utilities sector.

The literature on granular origins of business cycles shows that the effects of sectoral price shocks on aggregate prices are stronger for sectors with more central positions in the supply chain network. The present NVAR$(6,1)$ leads to the same long-term responses of prices to permanent shocks as in the static framework of contemporaneous network interactions used in that literature (see Section 2.2s). The comparative advantage of the present framework is that it permits studying how the effects of a shock unfold over time. In the following, I focus on the responses of aggregate PPI, in line with the literature, but the same analysis could also be conducted for sector-pairs (i,j).

The left panel of Fig. 6 shows the time profiles of aggregate PPI responses to price shocks in different sectors. It suggests that sectors differ in the speed at which they impact aggregate PPI. For example, the response of aggregate PPI to a shock to wholesale trade prices materializes rather quickly, while its response to an increase in the price of oil and gas extraction takes time. As revealed by the IRF discussion above, this is because wholesale trade connects to other sectors mostly as a direct or lower-order supplier, while the oil and gas extraction sector sits further upstream in its supply-chain relationships. In case of the aggregate PPI, the relevant counterpart is a weighted average of customer-sectors, with weights given by their contribution to aggregate output.

The right panel of Fig. 6 plots the strength of aggregate PPI responses against the fractions which materialize during the first quarter after the shock. Although stronger effects tend to take more time to realize, there is no clear relationship between the strength and timing of responses. For example, the construction and primary metals sectors have similar overall effects on aggregate prices. This means that other sectors (or the output-weighted average of them) have similar overall connections to both, as judged by the sum of connections of all order in the Leontief inverse.[49] Yet the impact of price increases in the construction sector materializes much more quickly since this sector is more relevant as an immediate supplier to relevant sectors in the economy compared to the primary metals sector.

In sum, the present analysis confirms the result from existing literature that the stronger the connections from sector i to sector j, the more pronounced is the response of sector i's PPI to a price shock in sector j. In addition, the analysis suggests that how fast sector i responds depends on the importance of sector j as a more immediate – rather than further upstream – supplier to sector i. The exact mapping from network-connections to impulse responses is determined by the extent to which connections of different order matter at different horizons. This is true not only for prices in a sector i, but also for a weighted average of sectoral prices, such as the aggregate PPI.

5 Forecasting Global Industrial Production Growth

How does economic activity co-move across countries? Given an expansion in one country, how do we expect economic activity in other countries to react? The international economics literature has long been interested in spillover and spillback effects across countries and the transmission of US shocks in particular. In this section, I shed light on global business cycles from a novel perspective, by assuming that the dynamic comovement in economic activity across countries is the result of bilateral connections, which I estimate. This is in starkest contrast with factor models, which in this context posit that it is the result of exposure to a few influential economies.

The previous section examined a case where an observed cross-sectional time series is arguably driven by one particular network of bilateral links for which data is available, and the interest lies in quantifying how network effects materialize over time. In this section, I consider the case where no network data is available, yet the assumption of an underlying network structure that shapes (cross-sectional) dynamics appears reasonable.[50] I first provide some intuition on the merits of the NVAR as a tool for approximating cross-sectional time

[49]Put simply, when summing up the connections of all order from sectors i to the construction sector and taking a weighted average, one gets a similar number as for the primary metals sector. See the expression for the Leontief inverse in Section 2.2.

[50]Note that in principle the NVAR could be applied for general, not necessarily cross-sectional time series.

series in Section 5.1 and discuss the relation to alternative methods. The application to cross-country industrial production growth is set up in Section 5.2, with results presented in Section 5.3.

5.1 Modeling Cross-Sectional Processes by Sparse Networks

Consider the problem of approximating dynamics of a cross-sectional time series y_t. Even for intermediate sizes of the cross-section, an unrestricted VAR(p) is not feasible. Modeling the series as an NVAR(p, q) process and estimating (α, A) gives a sparse, yet flexible and interpretable alternative.

Sparsity is obtained by the assumption that innovations transmit cross-sectionally only via bilateral links. As a result, the information in the high-dimensional vector of potential covariates – the lagged values of y_t – is compressed into a low-dimensional vector of regressors X_t that summarizes this information using network connections:

$$y_t = X_t\alpha + u_t \ , \quad X_t = [Ay_{t-1}, ..., Ay_{t-l}] \ .$$

Furthermore, because two units can be connected even in absence of a direct link between them, the dynamic, cross-sectional comovement may potentially be captured by relatively few non-zero bilateral links. In other words, A can be sparse, leading to additional parsimony.[51] Assuming that dynamic relations across all unit-pairs (i, j) are driven by a relatively small set of bilateral links is akin to the assumption that longer-term dynamics are driven by a set of shorter-term dynamics, which is upheld by the general class of VARMA(p, q) models.

Flexibility is obtained because the connections in A are estimated and because the NVAR(p, q) can capture rather general patterns of which connection-orders matter at which horizons, in particular for $q > 1$. The VARMA process in Eq. (5), which approximates the dynamics of y_t under an NVAR(p, q) with $q > 1$, brings to mind functional approximation of the linear projection of y_t on the information set at $t - 1$ using a polynomial expansion in A. Thereby, adding a term $\gamma_{lk}A^k$ to the equation satisfies the two main requirements on basis functions, orthogonality and locality: the term i) adds new, orthogonal information to that captured by lower powers of A, ii) adds different information across node-pairs (i, j), and iii) adds this information only at lag l.

Comparison to Alternatives There is a vast literature on modeling high-dimensional time series. The methods by which parsimony is induced can be roughly split into three categories.[52] Variable selection methods like Lasso or boosting aim at finding the most important predictors by excluding less relevant ones. Instead of imposing outright exclusion

[51] However, if $a_{ij} = 0$, y_j can Granger-cause y_i one period ahead only under an NVAR(p, q) with $q > 1$. See Proposition 1.

[52] See Carriero et al. (2011) for an extensive discussion.

restrictions, shrinkage methods such as Ridge regression or Minnesota-type priors do that by downweighting less relevant ones. Finally, factor models and reduced rank regression models reduce dimensionality by summarizing a large set of predictors by a few linear combinations of them.

The NVAR combines insights from factor models and variable selection. Compared to factor models, it offers a particular way of finding the linear combination that effectively summarizes the information in the high-dimensional set of predictors $y'_{t-1}, y'_{t-2}, ...$, namely by the set of bilateral links among cross-sectional units. Compared to variable selection methods, the NVAR places exclusion restrictions on bilateral links a_{ij}, which in turn summarize the information in the predictors, rather than on the predictors themselves. As a result, it entertains the additional sparsity assumption that, for any variable i, the same linear combinations of predictor-variables j matter at all lags $l = 1 : p$. Nevertheless, dynamics at different horizons h are driven by different linear combinations, i.e. different sets of connection-orders, even more so if the network interaction frequency is allowed to be higher than the frequency of observation (see Proposition 1 and discussion in Section 2.2.2). This additional restriction can become important in higher dimensions.

The rank of the network adjacency matrix A in the NVAR$(p, 1)$ is related to the number of factors in a factor model, arguably the most popular method for modeling high dimensional time series in macroeconomics. It is easy to see that an NVAR$(p, 1)$ permits a factor structure, with the number of factors given by the number of non-redundant columns in A. The equivalence result in Appendix D.1 establishes in addition that, for large n, a factor model can be cast as an NVAR$(p, 1)$ – with the number of factors again equal to the rank of A – provided that the factors themselves evolve according to an NVAR$(p, 1)$. Note that for $p = 1$, the latter condition just requires the factors to evolve according to a VAR(1), while in case of a single factor, it requires the factor to follow an AR(p).

Based on these insights, the NVAR is expected to better capture cross-sectional dynamics when these are composed of many (seemingly neglibile) links rather than driven by a few influential units. And indeed, in many cases, we expect A to be sparse, yet of close-to-full rank. For example, most countries trade only with a subset of other countries, but act as a significant trading partner to at least one other country. Similarly, most sectors supply only a small subset of other sectors in the economy, yet for most sectors we can find at least one other sector whose output or price-setting behavior critically depends on that of the sector in question. In principle, the same can apply not only for units in cross-sectional time series, but for variables in multivariate time series more generally.

Regardless of the rank of A, the NVAR is expected to better capture the dynamics of y_{it} for units i with a dependence structure in $A_{i\cdot}$ (or factor loadings) that differs considerably from that of other units. As pointed out in Boivin and Ng (2006), the more dispersion there is in the factor loadings across series, the worse will be the forecasting performance of a factor model.[53] This dispersion notably includes the case of weak factors, as captured by a sparse loading matrix (or a sparse adjacency matrix A in the case of an NVAR). The NVAR naturally incorporates weak factors as locally important units in the network. Moreover, under the NVAR, sparsity of A leads not only to cross-sectional differences in the strength of exposure to some given unit, but also to differences in the timing of this exposure[54]. Therefore, the NVAR is further preferred to factor models whenever some notion of cross-sectional distance is expected to be relevant for the timing of impulse responses.

Even in case the NVAR offers no advantage to factor models in terms of modeling and forecasting cross-sectional dynamics, it may be preferred for other reasons. First, it estimates a network as relevant for dynamics and, relatedly, offers an interpretable way of approximating the dynamics in y_t. Second, it facilitates the analysis of spillover and spillback effects as it estimates the whole set of bilateral Granger-causality patterns. Third, the estimated network offers a possible method for shock identification in high dimensions, the assumption being that the same bilateral links that rationalize lagged innovation transmission are also behind contemporaneous shock transmission.

5.2 Forecasting Setup

I use the NVAR to model the dynamics of monthly industrial production (IP) growth across countries. IP data is obtained from the IMF and OECD databases. Based on the raw data, I compute growth rates relative to the same month a year ago. Data availability narrows the sample to 44 countries and the time frame January 2001 to July 2022. In all of the following, I limit the analysis to pre-COVID-19. The data is summarized in Table A-3 in Appendix D.2.

To assess forecasting performance, I first estimate an NVAR$(p, 1)$ as well as a factor model based on data from January 2001 to December 2017 and consider out-of-sample forecasting performance for horizons up to 24 months ahead. The sample is iteratively increased by one month and the analysis is repeated until the sample end date reaches December 2019. Forecasts for periods after January 2020 are excluded from the assessment.

I estimate the NVAR$(p, 1)$ enforcing $a_{ij} \geq 0$ and using a Lasso-penalty to select links in A to zero (see Section 3.2). I select the optimal degree of sparsity in A based on BIC by counting the number of non-zero elements in $\hat{A}(\lambda)$, as suggested in Zou et al. (2007).[55] Once the NVAR is estimated, forecasts are obtained in the same way as for any VAR(p) model. The factor model is estimated using principal components. I select the number of factors

[53] Even if the number of factors is selected separately for each series, the forecasts for series that depend on less dominant factors are nevertheless more noisy than forecasts for series that depend on the most dominant factors. This is because including more estimated factors induces more sampling variability into the forecasts.

[54] See causality chain discussion in Section 2.2 and Proposition 1.

[55] Also, I normalize $||\alpha||_1 = 1$, though this is without influence on the results.

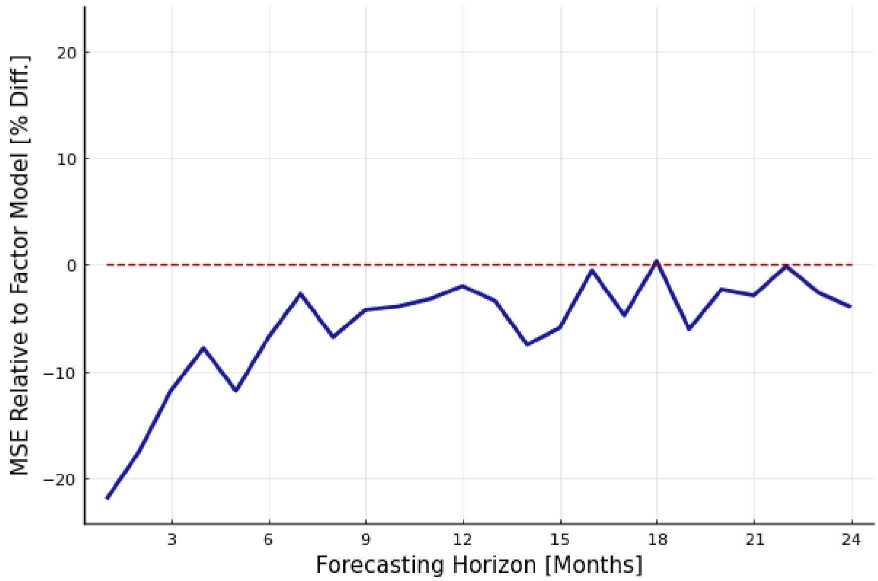


Figure 7: Out-of-Sample Forecasting Performance: NVAR(4, 1) vs. Factor Model

Notes: The plot depicts the percentage difference between the out-of-sample Mean Squared Errors generated by the NVAR(1, 1) to those generated by the Principal Components Factor Model.

based on the information criterion developed in Bai and Ng (2002). To construct forecasts, I fit a VAR(p) for the factors.

5.3 Results

The results of the forecasting exercise are shown in Fig. 7. It reports the average out-of-sample Mean Squared Error (MSE) across countries under the estimated NVAR(4, 1) relative to those obtained under the factor model. The NVAR(4, 1) yields a substantial reduction in MSE compared to the factor model. This holds in particular for forecasts up to six months ahead. The results for alternative choices of p are similar (see Fig. A-6).

On top of the good forecasting performance, the NVAR estimates the network as relevant for IP dynamics and the full set of Granger-causality patterns among units. The following discussion is based on the estimated network using data from January 2001 to December 2017 and setting $\lambda = 0$. Note that even in this case, the estimated A is sparse as $a_{ij} \in \mathbb{R}_+$ is enforced. In the network, 22% of the links are non-zero. Excluding links below 0.05, the network density drops to 11%. Nevertheless, $\hat{A}$ has rank 38 and is therefore close to full-rank. For higher values of λ, selected as optimal by BIC and used to construct the forecasts in Fig. 7, this number is lower, but always stays at levels that exceed the factor dimensions commonly selected in applications.

Fig. 8 shows weighted out-degrees, $\hat{wd}^j_{out} = \sum_i \hat{a}_{ij}$, a measure of country j's influence on IP dynamics of other countries in the sample. Without any information beyond the IP series across countries, the NVAR estimates the most influential country to be the United

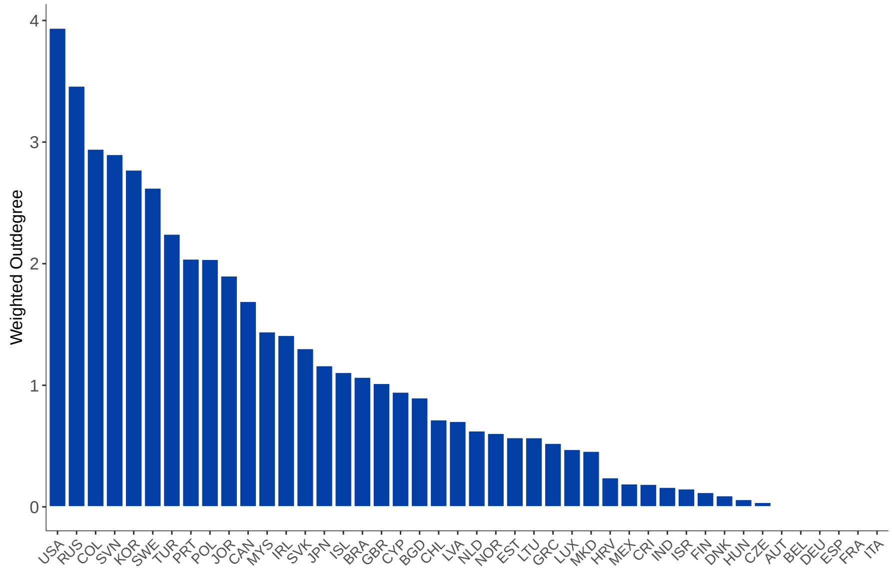


Figure 8: Weighted Out-degrees in the Estimated Network

Notes: The plot depicts the weighted outdegrees in the estimated network as relevant for monthly industrial production dynamics across countries.

States, in line with expectations. The second most influential country is Russia, another large economy and major energy-exporter. The ordering of countries according to this measure does also show some surprises. In particular, Germany, France and Italy are estimated to not influence any other country in the sample. This is presumably due to high (contemporaneous) correlation of economic activity among countries in the Euro Area and EU. As a result, the model likely attributes innovations coming from these three major European economies to Slovenia, Sweden, Portugal and Poland, all of which are estimated to be among the most influential countries. Such results can be avoided by including prior information. For example, one could shrink links to some measure of bilateral connection from the data, such as capital or trade flows.

In the top left panel of Fig. 9, I illustrate the propagation pattern of innovations to IP growth through the network as captured by $\hat{\alpha}$. As in Fig. 5 in Section 4, the dots show which connection orders matter for innovation propagation at which horizon, with the strength reflected by the dots' size and shading. In contrast to Fig. 5, however, the coefficients in α are not restricted to be positive in this application. To distinguish positive from negative transmission, I show the former in blue and the latter in red. The plot suggests that following an incrase in a country's IP growth, other countries' response features an initial overshooting and subsequent correction. The exact magnitudes of these forces depend on network connections of different order between any given pair (i, j). The top right panel of Fig. 9 reports these connections from Germany and Finland, respectively, to the United States. While Germany is estimated to have a strong direct link to the US, the dependence of

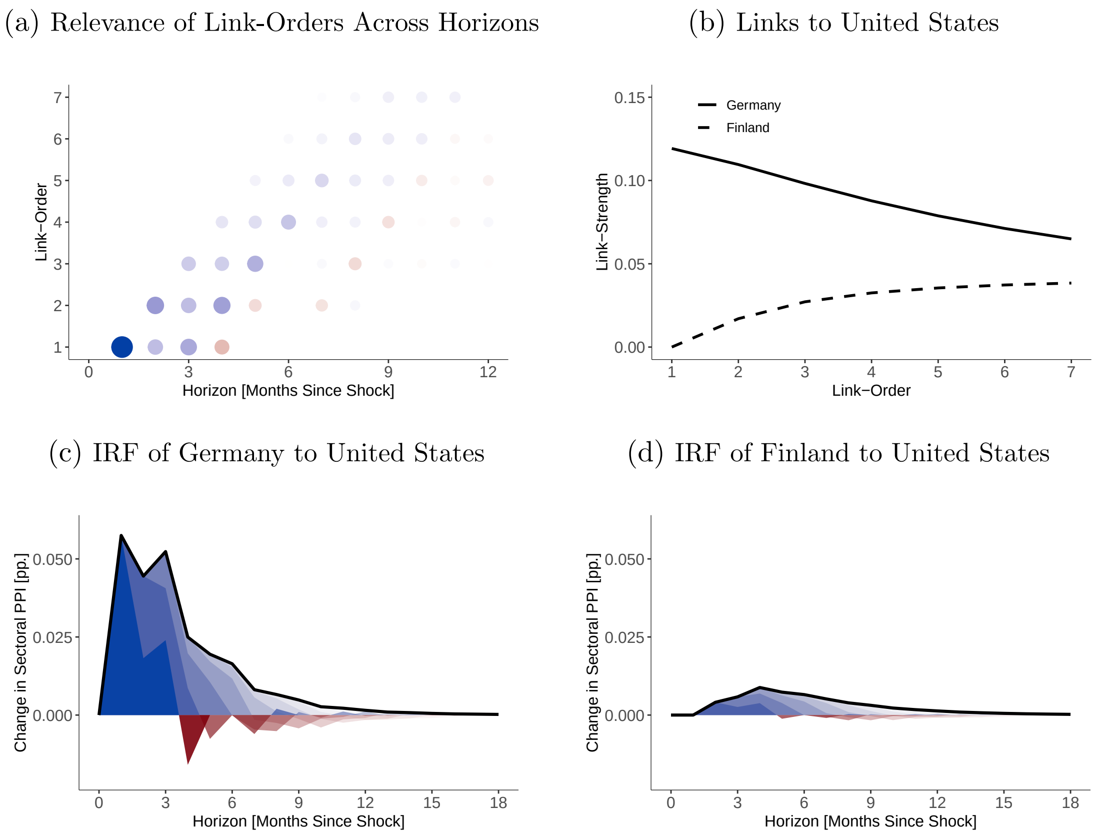


Figure 9: Impulse Responses: Network-Induced Transmission of Economic Activity

Notes: The top left panel shows the importance of different connection-orders for transmission as a function of the time elapsed since an innovation took place. The top right panel shows the connections of all order from Germany and Finland to the United States, and the bottom panels show their resulting IRFs to a one standard deviation increase in US industrial production.

Finnish IP on that of the US comes only from higher-order connections and is weaker overall. As a result, the lower panels of Fig. 9 show that IP in Germany responds much faster and stronger to an increase in US IP growth. In contrast, it takes time for this increase to travel through the network and affect economic activity in Finland. Both impulse responses show a sinusoidal pattern of innovation transmission via any one given order of network connections, reflecting the initial overshooting and subsequent correction. This is depicted by different shading for different link-orders.[56]

6 Conclusion

In this paper, I develop an econometric framework that explicitly relates the dynamics of cross-sectional variables to the bilateral links among cross-sectional units. In a first applica-

[56]Note that the interpretative decomposition of impulse responses in Fig. 9 is impacted by the normalization applied. To generate the figure, I re-scale the estimated network such that $lim_{k\to\infty}\hat{A}^k = 0$ by dividing the estimated adjacency matrix by its largest Eigenvalue (in absolute value).

tion, I use it to estimate how supply chain linkages affect the dynamics of sectoral prices in the US economy. In a second application, I apply it as a dimensionality-reduction technique for modeling cross-country industrial production growth.

As discussed in the introduction, most existing studies in macroeconomics establish relations between network statistics and cross-sectional properties in static environments. The proposed NVAR can be used to examine the relation between network properties and cross-sectional dynamics, avoiding the simplifying assumption of contemporaneous network interactions. Thereby, the model could be augmented to accommodate heterogeneous propagation patterns across units or over time. By adding covariates, it could assess to what extent dynamics are driven by a particular network as opposed to other forces.

The NVAR presented in this chapter assumes that innovation transmission occurs along bilateral links which are fixed over time. An important methodological step forward would be to develop a tractable framework that links such network effects to network formation. In many cases, units can adjust their network position in reaction of innovations transmitted through the network.

A further promising direction for future research is the use of networks - estimated or not - for shock identification. This possibly offers a solution to the challenge of finding a convincing identification strategy for cross-sectional time series, in particular under high dimensions.

References

Acemoglu, D., U. Akcigit, and W. Kerr (2016): "Networks and the macroeconomy: an empirical exploration," *NBER Macroeconomics Annual*, 30, 273–335.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Bai, J. and S. Ng (2002): "Determining The Number Of Factors In Approximate Factor Models," *Econometrica*, 70, 191–221.

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

Barigozzi, M., G. Cavaliere, and G. Moramarco (2022): "Factor Network Autoregressions," *Manuscript, University of Bologna.*

Berman, A. and R. J. Plemmons (1979): *Nonnegative Matrices in the Mathematical Sciences.*

Boivin, J. and S. Ng (2006): "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194.

Bouakez, H., E. Cardia, and F. Ruge-Murcia (2014): "Sectoral price rigidity and aggregate dynamics," *European Economic Review*, 65, 1–22.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

Bramoullé, Y., A. Galeotti, and B. W. Rogers (2016): *The Oxford Handbook of the Economics of Networks.*

Bykhovskaya, A. (2021): "Time Series Approach to the Evolution of Networks: Prediction and Estimation," *Journal of Business & Economic Statistics, forthcoming.*

Cai, M., M. D. Negro, E. Herbst, E. Matlin, R. Sarfati, and F. Schorfheide (2021): "Online estimation of DSGE models," *Econometrics Journal*, 24, C33–C58.

Camehl, A. (2022): "Penalized estimation of panel vector autoregressive models: A panel LASSO approach," *International Journal of Forecasting, forthcoming.*

Caporin, M., D. Erdemlioglu, and S. Nasini (2023): "Estimating Financial Networks by Realized Interdependencies: A Restricted Vector Autoregressive Approach," *Manuscript, University of Padova.*

Carriero, A., G. Kapetanios, and M. Marcellino (2011): "Forecasting large datasets with Bayesian reduced rank multivariate models," *Journal of Applied Econometrics*, 26, 735–761.

Carter, C. K. and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Carvalho, V. M. (2014): "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, 28, 23–48.

Carvalho, V. M. and A. Tahbaz-Salehi (2019): "Production Networks: A Primer," *Annual Review of Economics*, 635–663.

Cesa-Bianchi, A. and A. Ferrero (2021): "The Transmission of Keynesian Supply Shocks," *Manuscript, Bank of England.*

Chudik, A. and M. H. Pesaran (2011): "Infinite-dimensional VARs and factor models," *Journal of Econometrics*, 163, 4–22.

Dahlhaus, T., J. Schaumburg, and T. Sekhposyan (2021): "Networking the yield curve: implications for monetary policy," *ECB Working Paper Series.*

Diebold, F. X. and K. Yilmaz (2009): "Measuring Financial Asset Return and Volatility Spillovers, With Application to Global Equity Markets," *The Economic Journal*, 119, 158–171.

——— (2014): "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, 182, 119–134.

Dufour, J.-M. and E. Renault (1998): "Short Run and Long Run Causality in Time Series: Theory," *Econometrica*, 66, 1099–1125.

Elliott, M., B. Golub, and M. O. Jackson (2014): "Financial networks and contagion," *American Economic Review*, 104, 3115–3153.

Fan, J., R. Masini, and M. C. Medeiros (2021): "Bridging factor and sparse models," *Manuscript, Princeton University.*

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119.

Freyaldenhoven, S. (2022): "Identification Through Sparsity in Factor Models: the L1-rotation criterion," *Manuscript, Federal Reserve Bank of Philadelphia.*

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97, 436–451.

Giovanni, J. D., A. A. Levchenko, and I. Mejean (2018): "The micro origins of international business-cycle comovement," *American Economic Review*, 108, 82–108.

Giroud, X. and H. M. Mueller (2019): "Firms' internal networks and local economic shocks," *American Economic Review*, 109, 3617–3649.

Golub, B. and M. O. Jackson (2010): "Naïve Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149.

Graham, B. S. (2020): "Network data," *Handbook of Econometrics.*

Herbst, E. P. and F. Schorfheide (2015): *Bayesian Estimation of DSGE Models*, Princeton University Press.

Hipp, R. (2020): "On Causal Networks of Financial Firms: Structural Identification via Non-Parametric Heteroskedasticity," *Staff Working Paper, Bank of Canada.*

Horvath, M. (2000): "Sectoral Shocks and Aggregate Fluctuations," *Journal of Monetary Economics*, 45, 69–106.

Hsu, N. J., H. L. Hung, and Y. M. Chang (2008): "Subset selection for vector autoregressive processes using Lasso," *Computational Statistics and Data Analysis*, 52, 3645–3657.

Knight, M. I., M. A. Nunes, and G. P. Nason (2016): "Modelling, Detrending and Decorrelation of Network Time Series," *Manuscript, University of Bristol.*

Lee, L. F. (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140, 333–374.

Long, J. B. J. and C. I. Plosser (1983): "Real Business Cycles," *Journal of Political Economy*, 93, 36–69.

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

Mehl, A., M. Mlikota, and I. V. Robays (2022): "Why Does a Dominant Currency Replace Another?" *Manuscript, European Central Bank.*

Meng, X.-L. and D. B. Rubin (1993): "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika*, 80, 267–278.

Mlikota, M. and F. Schorfheide (2022): "Sequential Monte Carlo With Model Tempering," *Manuscript, University of Pennsylvania.*

Onatski, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168, 244–258.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004): "Modeling Regional Interdependences Using a Global Error-Correcting Macroeconometric Model," *Journal of Business and Economic Statistics*, 22, 129–162.

Phillips, P. C. B. (1973): "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351–362.

Stock, J. H. and M. W. Watson (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.

Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986): "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105–118.

Yang, K. and L. F. Lee (2019): "Identification and estimation of spatial dynamic panel simultaneous equations models," *Regional Science and Urban Economics*, 76, 32–46.

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017): "Network vector autoregression," *Annals of Statistics*, 45, 1096–1123.

Zou, H., T. Hastie, and R. Tibshirani (2007): "On the "degrees of freedom" of the lasso," *Annals of Statistics*, 35, 2173–2192.

Appendix

Cross-Sectional Dynamics Under Network Structure: Theory and Macroeconomic Applications

Marko Mlikota
University of Pennsylvania

A NVAR Model

A.1 Granger-Causality

Proposition 4 (Granger-Causality in NVAR$(p,1)$).
Let y_t follow an NVAR(p,1):

$$y_t = \sum_{l=1}^{p} \Phi_l y_{t-l} + u_t \ , \quad \Phi_l = \alpha_l A \ , \quad \alpha_l \in \mathbb{R} \ ,$$

and assume $\alpha_l \neq 0$ for $l = 1:p$,. Then $\frac{\partial y_{i,t+h}}{\partial y_{j,t}} | \mathcal{F}_t > 0 \Leftrightarrow (A^k)_{ij} > 0$ for at least one $k \in \underline{k} : h$, $\underline{k} = ceil(h/p)$, i.e. y_j Granger-causes y_i at horizon h if and only if there exists a walk from i to j of at least one length $k \in \underline{k} : h$.

Proof: Using the companion form of this process, we have

$$\frac{\partial y_{t+h}}{\partial y_t} = [I_n, 0_{n\times n(p-1)}] F^h [I_n, 0_{n\times n(p-1)}]' = (F^h)_{11} \ ,$$

$$F = \begin{bmatrix} \Phi_1 & \Phi_2 & ... & \Phi_{p-1} & \Phi_p \\ I_n & 0_n & ... & 0_n & 0_n \\ 0_n & I_n & ... & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & ... & I_n & 0_n \end{bmatrix} \ , \quad \Phi_l = \alpha_l A \ .$$

I will prove the following claim by induction: $(F^h)_{1l}$, the $n \times n$ matrix in position $(1,l)$ of the $np \times np$ matrix F, has powers of A in the set $ceil(\frac{h+l-1}{p}) : h$. Note that the claim is true for $h = 1$. Assume it is true for h. For $h+1$ we have

$$F^{h+1} = \begin{bmatrix} (F^h)_{11} & (F^h)_{12} & \dots & (F^h)_{1p} \\ (F^h)_{21} & (F^h)_{22} & \dots & (F^h)_{2p} \\ \vdots & & \ddots & \vdots \\ (F^h)_{p1} & (F^h)_{p2} & \dots & (F^h)_{pp} \end{bmatrix} \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & I_n & \dots & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & \dots & I_n & 0_n \end{bmatrix}$$
$$= \begin{bmatrix} (F^h)_{11}\Phi_1 + (F^h)_{12} & (F^h)_{11}\Phi_2 + (F^h)_{13} & \dots & (F^h)_{11}\Phi_{p-1} + (F^h)_{1p} & (F^h)_{11}\Phi_p \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{bmatrix} ,$$

where I only show the first row of blocks in F^{h+1} as only they are relevant to the argument.

Let $h + m$ for some $m \in 1 : p-1$ be a multiple of p such that $ceil(\frac{h+l-1}{p}) = \frac{h+m}{p}$ for $l = 1 : m+1$, while $ceil(\frac{h+l-1}{p}) = \frac{h+m}{p} + 1$ for $l = m+2 : p$. This means that $(F^h)_{1l}$ for $l = 1 : m+1$ have powers of A in $\frac{h+m}{p} : h$, while $(F^h)_{1l}$ for $l = m+2 : p$ have powers in $\frac{h+m}{p} + 1 : h$. Then, using the equation above, $(F^{h+1})_{1l}$ for $l = 1 : m$ have powers of A in $\frac{h+m}{p} : h+1 = ceil(\frac{h+1+l-1}{p}) : h+1$, while $(F^{h+1})_{1l}$ for $l = 1 : m$ have powers in $\frac{h+m}{p} + 1 : h+1 = ceil(\frac{h+1+l-1}{p}) : h+1$. Note that these sets are independent of m and therefore the claim holds for $h+1$ in all possible cases. ■

A.2 Time-Aggregation of Lagged Transmission Patterns

In this section, I show that the state space representations in Eq. (4) and Eq. (6) for $q \in \mathbb{N}\backslash\{1\}$ lead to a VARMA process for the observed series $\{y_t\}_{t=1}^T$ of the type of Eq. (5). For illustration purposes, let x_τ follow an NVAR$(3, 1)$.

Suppose we observe x_τ every $q = 2$ periods. Under monthly observations, this would mean that network interactions happen bi-weekly. Suppose the particular realization at period τ, x_τ, is observed. Inserting sequentially for the non-observed $x_{\tau-1}$ and $x_{\tau-3}$, we get

$$\begin{aligned} x_\tau = &\left[\alpha_2 A + \alpha_1^2 A^2\right] x_{\tau-2} + \left[(\alpha_1\alpha_2 + 2\alpha_1\alpha_3)A^2\right] x_{\tau-4} \\ &+ v_\tau + \alpha_1 A v_{\tau-1} + (\alpha_3 A + \alpha_1\alpha_2 A^2) v_{\tau-3} + \text{terms in } x_{\tau-6}, x_{\tau-7} . \end{aligned}$$

This can be written (for a generic observational period t) as

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \text{terms in } x_{\tau-6}, x_{\tau-7} ,$$

where $y_{t-l} = x_{\tau-2l}$ for $l = 0, 1, 2$, $\eta_t = [v'_\tau, v'_{\tau-1}]'$, $\eta_{t-1} = [v'_{\tau-2}, v'_{\tau-3}]'$, and

$$\Phi_1 = \alpha_2 A + \alpha_1^2 A^2 , \quad \Phi_2 = (\alpha_1\alpha_2 + 2\alpha_1\alpha_3)A^2 , \quad \Theta_0 = [I_n, \alpha_1 A] , \quad \Theta_1 = \left[0_n, \alpha_3 A + \alpha_1\alpha_2 A^2\right] .$$

The exact process $\{y_t\}_{t=1}^T = \{x_{qt}\}_{t=1}^T$ contains infinitely many lags with ever higher powers of A at higher lags. However, if x_τ is stationary, then so is y_t and the latter can be approximated with some finite number of lags p^*.[57]

Suppose instead that we observe $y_{\tau/q} = x_\tau + x_{\tau-1}$ for $\tau/2 \in \mathbb{N}$. Starting from $y_t = x_\tau + x_{\tau-1}$, iteratively inserting for x_τ and $x_{\tau-1}$ and adding and subtracting an $x_{\tau-3}$-term so as to align the coefficients in front of $x_{\tau-2}$ and $x_{\tau-3}$, we obtain

$$\begin{aligned} y_t &= x_\tau + x_{\tau-1} \\ &= [(\alpha_1+\alpha_2)A + \alpha_1^2 A^2](x_{\tau-2}+x_{\tau-3}) + v_\tau + [I+\alpha_1 A]v_{\tau-1} \\ &\quad + [(\alpha_3-\alpha_1)A + \alpha_1(\alpha_2-\alpha_1)A^2]x_{\tau-3} + [\alpha_3 A + \alpha_1\alpha_3 A^2]x_{\tau-4} \ . \end{aligned}$$

Inserting for the terms in the second row and adding and subtracting an $x_{\tau-5}$-term, we get

$$\begin{aligned} y_t &= x_\tau + x_{\tau-1} \\ &= [(\alpha_1+\alpha_2)A + \alpha_1^2 A^2](x_{\tau-2}+x_{\tau-3}) + v_\tau + [I+\alpha_1 A]v_{\tau-1} + [(\alpha_3-\alpha_1)A + \alpha_1(\alpha_2-\alpha_1)A^2]v_{\tau-3} \\ &\quad + [\alpha_3 A + \alpha_1(2\alpha_3-\alpha_1)A^2 + \alpha_1^2(\alpha_2-\alpha_1)A^3](x_{\tau-4}+x_{\tau-5}) + \text{terms in } x_{\tau-5}, x_{\tau-6} \ , \end{aligned}$$

which can be written as

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \text{terms in } x_{\tau-5}, x_{\tau-6} \ ,$$

where $y_{t-l} = x_{\tau-2l} + x_{\tau-2l-1}$ for $l = 0, 1, 2$, $\eta_t = [v_\tau', v_{\tau-1}']'$, $\eta_{t-1} = [v_{\tau-2}', v_{\tau-3}']'$, and

$$\begin{aligned} &\Phi_1 = (\alpha_1+\alpha_2)A + \alpha_1^2 A^2 \ , \quad \Phi_2 = \alpha_3 A + \alpha_1(2\alpha_3-\alpha_1)A^2 + \alpha_1^2(\alpha_2-\alpha_1)A^3 \ , \\ &\Theta_0 = [I_n, I+\alpha_1 A] \ , \quad \Theta_1 = \left[0_n, (\alpha_3-\alpha_1)A + \alpha_1(\alpha_2-\alpha_1)A^2\right] \ . \end{aligned}$$

As in the case of stock variables, the exact process $\{y_t\}_{t=1}^T$ contains infinitely many lags with ever higher powers of A at higher lags. However, if x_τ is stationary, then so is y_t and the latter can be approximated well with a finite number of lags.

Both for stock and flow variables, the procedure of finding the coefficient matrices $\{\Phi_l\}_{l=1}^p$ and $\{\Theta_l\}_{l=0}^{p-1}$ in the approximating VARMA-representation can be formalized as an algorithm for general p and q. However, this does not offer any benefits in terms of estimation or model analysis relative to the state space representations in Eq. (4) and Eq. (6).

In Eq. (5), q^* is claimed to be equal to p^*q. This is because the largest power of A in any Φ_l cannot exceed p^*q (and the coefficients γ_{lg} can be zero). To see this, note that starting from the equation for x_τ and iteratively inserting for all $x_{\tau-j}$, $j = 1:h-1$, one would obtain $\partial x_{\tau+h}/\partial x_\tau \big| \mathcal{F}_\tau^x$ as the coefficient-matrix in front of $x_{\tau-h}$, which we know consists of powers of A in the set $k \leq h$. However, when performing the time-aggregation for stock variables, every qth x_τ is not inserted for, which means that powers accumulate slower. As a result,

[57] See Appendix A.3.

the coefficient-matrix in front of $y_{t-l} = x_{(t-l)q}$ contains powers not higher than lq. This is maximized for $l = p^*$. The same result applies also to flow variables.

A.3 Stationarity

Proposition 5 (Stationarity of NVAR$(1,1)$)**.**
Let y_t follow an NVAR(1,1):

$$y_t = aAy_{t-1} + u_t \ , \quad a \in \mathbb{R} \ ,$$

with $u_t \sim WN$ and assume $a \neq 0$. Then y_t is weakly stationary iff for all eigenvalues λ_i of A it holds that $|\lambda_i| < 1/|a|$.

Proof: This follows directly from the fact that λ_i is an eigenvalue of A iff $a\lambda_i$ is an eigenvalue of aA. Formally, let

$$\begin{aligned} \mathcal{L} &= \{\lambda_i : |\lambda_i I - A| = 0\} \ , \\ \tilde{\mathcal{L}} &= \left\{\tilde{\lambda}_i : |\tilde{\lambda}_i I - aA| = 0\right\} \ . \end{aligned}$$

We have

$$\begin{aligned} y_t \text{ stationary} \quad &\Leftrightarrow \quad \forall\ \tilde{\lambda}_i \in \tilde{\mathcal{L}} \ , \quad |\tilde{\lambda}_i| < 1 \\ &\Leftrightarrow \quad \forall\ \tilde{\lambda}_i \in \tilde{\mathcal{L}} \ , \quad |\tilde{\lambda}_i/a| = |\tilde{\lambda}_i|/|a| < 1/|a| \\ &\Leftrightarrow \quad \forall\ \lambda_i \in \mathcal{L} \ , \quad |\lambda_i| < 1/|a| \ , \end{aligned}$$

where the last equivalence follows from

$$|\tilde{\lambda}_i I - aA| = |a(\tilde{\lambda}_i/aI - A)| = a^n|\tilde{\lambda}_i/aI - A| = 0 \Leftrightarrow |\tilde{\lambda}_i/aI - A| = 0 \ . \quad \blacksquare$$

Proposition 6.
Let y_t follow an NVAR$(p,1)$:

$$y_t = \alpha_1 Ay_{t-1} + ... + \alpha_p Ay_{t-p} + u_t \ ,$$

with $u_t \sim WN$, and assume $\alpha_l \neq 0$ for at least one l. Define $a = \sum_{l=1}^{p} |\alpha_l|$, and let

$$\tilde{y}_t = aA\tilde{y}_{t-1} + \tilde{u}_t \ .$$

Then, y_t is weakly stationary if $\tilde{y}_t$ is weakly stationary.

Proof: Let

$$\begin{aligned}
\mathcal{Z} &= \{z_i : |I - \alpha_1 A z_i - ... - \alpha_p A z_i^p| = |I - (\alpha_1 z_i + ... + \alpha_p z_i^p)A| = 0\} \ , \\
\tilde{\mathcal{Z}} &= \{\tilde{z}_i : |I - \tilde{z}_i a A| = 0\} \ .
\end{aligned}$$

The proof shall show

$$\forall\ \tilde{z}_i \in \tilde{\mathcal{Z}}, \quad |\tilde{z}_i| > 1 \quad \Rightarrow \quad \forall\ z_i \in \mathcal{Z}, \quad |z_i| > 1 \ .$$

We have

$$\begin{aligned}
& \forall\ \tilde{z}_i \in \tilde{\mathcal{Z}}\ , \quad |\tilde{z}_i| > 1 \\
\Leftrightarrow\ & \forall\ \tilde{z}_i \in \tilde{\mathcal{Z}}, \quad |a\tilde{z}_i| = a|\tilde{z}_i| > a \\
\Leftrightarrow\ & \forall\ z_i \in \mathcal{Z}\ , \quad |\alpha_1 z_i + ... + \alpha_p z_i^p| > a \\
\Rightarrow\ & \forall\ z_i \in \mathcal{Z}\ , \quad |z_i| > 1 \ .
\end{aligned}$$

To show the last implication, suppose first that the statement on the second-last line is true, but the statement on the last line is not. Then $\exists z_i \in \mathcal{Z}$ s.t. $|z_i| \leq 1$. In turn,

$$|\alpha_1 z_i + ... + \alpha_p z_i^p| \leq |\alpha_1 z_i| + ... + |\alpha_p z_i^p| \leq |\alpha_1 z_i| + ... + |\alpha_p z_i| \leq (|\alpha_1| + ... + |\alpha_p|)|z_i| = a|z_i| \leq a \ ,$$

a contradiction. ■

Proposition 7.
Let y_t follow an NVAR$(p,1)$:

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ ,$$

with $u_t \sim WN$, and assume $\alpha_l \geq 0$ for $l = 1:p$ and $\alpha_l > 0$ for at least one l. Define $a = \sum_{l=1}^p \alpha_l$, and let

$$\tilde{y}_t = aA\tilde{y}_{t-1} + \tilde{u}_t \ .$$

Then, y_t is weakly stationary iff $\tilde{y}_t$ is weakly stationary.

Proof: The proof is equivalent to that of Proposition 6, except that if $\alpha_l \geq 0\ \forall\ l$, the last implication is both-sided:

$$\begin{aligned}
& \forall\ z_i \in \mathcal{Z}\ , \quad |z_i| > 1 \\
\Rightarrow\ & \forall\ z_i \in \mathcal{Z}\ , \quad |\alpha_1 z_i + ... + \alpha_p z_i^p| > |(\alpha_1 + ... + \alpha_p) z_i| = |a z_i| = a|z_i| > a \ . \quad ■
\end{aligned}$$

Corollary 1 (Stationarity of NVAR$(p,1)$ I)**.**
Let y_t follow an NVAR$(p,1)$:

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ ,$$

with $u_t \sim WN$, and assume $\alpha_l \neq 0$ for at least one l. Define $a = \sum_{l=1}^{p} |\alpha_l|$.
Then y_t is weakly stationary if for all eigenvalues λ_i of A it holds that $|\lambda_i| < 1/a$.
If in addition $\alpha_l \geq 0 \ \forall \ l$, then this condition is both necessary and sufficient.

Proposition 8 (Stationarity of NVAR$(p,1)$ II)**.**
Let y_t follow an NVAR$(p,1)$:

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ ,$$

with $u_t \sim WN$ and $\alpha_l \neq 0$ for at least one l.
Then, y_t is weakly stationary iff the univariate AR(p) process

$$\check{y}_t = \lambda_i \alpha_1 \check{y}_{t-1} + ... + \lambda_i \alpha_p \check{y}_{t-p} + \check{u}_t$$

is weakly stationary for all eigenvalues λ_i of A.

Proof: Stationarity of y_t is equivalent to the statement that for all eigenvalues l_i of

$$F = \begin{bmatrix} \alpha_1 A & \alpha_2 A & ... & \alpha_{p-1} A & \alpha_p A \\ I_n & 0_n & ... & 0_n & 0_n \\ 0_n & I_n & ... & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & ... & I_n & 0_n \end{bmatrix}$$

it holds that $|l_i| < 1$. We have

$$\begin{aligned} & |l_i I - F| = 0 \\ \Leftrightarrow \quad & \left| l_i^p I - l_i^{p-1} \alpha_1 A - ... - l_i \alpha_{p-1} A - \alpha_p A \right| = 0 \\ \Leftrightarrow \quad & l_i^{n(p-1)} \left| l_i I - \left(\alpha_1 + \alpha_2 / l_i + ... + \alpha_p / l_i^{p-1} \right) A \right| = 0 \\ \Leftrightarrow \quad & \left(l_i^{p-1} \left(\alpha_1 + \alpha_2 / l_i + ... + \alpha_p / l_i^{p-1} \right) \right)^n \left| \frac{l_i}{\alpha_1 + \alpha_2 / l_i + ... + \alpha_p / l_i^{p-1}} I - A \right| = 0 \\ \Leftrightarrow \quad & \left| \frac{l_i}{\alpha_1 + \alpha_2 / l_i + ... + \alpha_p / l_i^{p-1}} I - A \right| = 0 \ . \end{aligned}$$

This establishes a relation between the eigenvalues l_i of F and the eigenvalues λ_i of A. Given an eigenvalue l_i of F, we know $l_i / \left(\alpha_1 + \alpha_2/l_i + ... + \alpha_p/l_i^{p-1}\right)$ is an eigenvalue of A. Conversely, given an eigenvalue λ_i of A, all eigenvalues l_i that solve

$$l_i^p - l_i^{p-1}\lambda_i\alpha_1 - ... - l_i\lambda_i\alpha_{p-1} - \lambda_i\alpha_p = 0$$

are eigenvalues of F. This equation is the characteristic polynomial for stationarity of the AR(p) process $\breve{y}_t$ defined above. ■

Proposition 9 (Stationarity Preservation Under Time-Aggregation).
Let x_τ follow an NVAR($p, 1$)

$$x_\tau = \alpha_1 A x_{\tau-1} + ... + \alpha_p A x_{\tau-p} + \upsilon_\tau \ ,$$

with $\upsilon_\tau \sim WN$. Let $q \in \mathbb{N}\backslash\{1\}$ and consider the time series y_t defined by $\{y_t\}_{t=1}^T = \{x_{tq}\}_{t=1}^T$ and z_t defined by $\{z_t\}_{t=1}^T = \{x_{tq} + ... + x_{(t-1)q+1}\}_{t=1}^T$. Then if x_τ is weakly stationary, so are y_t and z_t.

Proof: Weak stationarity of x_τ is defined by the two conditions

1. $\mathbb{E}[x_\tau] = \mathbb{E}[x_{\tau-l}] \ \forall \ l$
2. $Cov(x_\tau, x_{\tau-h}) = Cov(x_{\tau-l}, x_{\tau-l-h}) \ \forall \ l, h$

They imply that

1. $\mathbb{E}[y_t] = \mathbb{E}[x_{tq}] = \mathbb{E}[x_{(t-l)q}] = \mathbb{E}[y_{t-l}] \ \forall \ l$
2. $Cov(y_t, y_{t-h}) = Cov(x_{tq}, x_{(t-h)q}) = Cov(x_{(t-l)q}, x_{(t-l-h)q}) = Cov(y_{t-l}, y_{t-l-h}) \ \forall \ l, h$,

which in turn is the definition of stationarity for y_t. Similarly, they imply that

1. $\mathbb{E}[z_t] = \mathbb{E}[x_{tq} + ... + x_{(t-1)q+1}] = \mathbb{E}[x_{(t-l)q} + ... + x_{(t-l-1)q+1}] = \mathbb{E}[y_{t-l}] \ \forall \ l$
2. $$\begin{aligned} Cov(z_t, z_{t-h}) &= Cov(x_{tq} + ... + x_{(t-1)q+1} \ , \ x_{(t-h)q} + ... + x_{(t-h-1)q+1}) \\ &= Cov(x_{(t-l)q} + ... + x_{(t-l-1)q+1} \ , \ x_{(t-l-h)q} + ... + x_{(t-l-h-1)q+1}) \\ &= Cov(z_{t-l}, z_{t-l-h}) \ \forall \ l, h \ , \end{aligned}$$

which is the definition of stationarity for z_t. ■

A.4 Impulse-Responses

Proposition 10 (Long-Term Response in NVAR$(p,1)$)**.**
Let y_t follow an NVAR$(p,1)$:

$$y_t = \alpha_1 A x_{t-1} + ... + \alpha_p A x_{t-p} + u_t \ .$$

Define $a = \sum_{l=1}^{p} \alpha_l$, and let

$$y = aAy + \varepsilon \ .$$

Assume y_t is stationary. Then, the long-term response of y_t to a permanent increase in u_t is equivalent to the (contemporaneous) response of y to a disturbance in ε, $\partial y/\partial \varepsilon$, i.e.

$$R \equiv lim_{h\to\infty}\left[\frac{\partial y_{t+h}}{\partial u_t} + \frac{\partial y_{t+h}}{\partial u_{t+1}} + ... + \frac{\partial y_{t+h}}{\partial u_{t+h}}\right] = \frac{\partial y}{\partial \varepsilon} \ .$$

Proof: First, note that

$$y = (I - aA)^{-1}\varepsilon \ ,$$

and therefore $\partial y/\partial \varepsilon = (I - aA)^{-1}$.

Turning to y_t, note that under stationarity

$$R = \lim_{h\to\infty}\sum_{j=0}^{h+1}\frac{\partial y_{t+h}}{\partial u_{t+h-j}} = \lim_{h\to\infty}\sum_{j=0}^{h+1}\frac{\partial y_{t+j}}{\partial u_t} = \sum_{j=0}^{\infty}\frac{\partial y_{t+j}}{\partial u_t} \ .$$

To get the impulse response function for x_t, write it in companion form as

$$z_t = F z_{t-1} + e_t \ ,$$

where $z_t = [y_t', y_{t-1}', ..., y_{t-p+1}']'$ and $e_t = [u_t', 0', ..., 0']'$ are np-dimensional vectors, and the $n \times n$ matrix F is defined as

$$F = \begin{bmatrix} \alpha_1 A & \alpha_2 A & ... & \alpha_{p-1} A & \alpha_p A \\ I_n & 0_n & ... & 0_n & 0_n \\ 0_n & I_n & ... & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & ... & I_n & 0_n \end{bmatrix} .$$

The impulse response of x_t to a disturbance in v_t is then given by $n \times n$ upper left block in

F^h, denoted by $(F^h)_{11}$:

$$\frac{\partial y_{t+h}}{\partial u_t} = \frac{\partial y_{t+h}}{\partial z_{t+h}} \frac{\partial z_{t+h}}{\partial e_t} \frac{\partial e_t}{\partial u_t} = [I_n, 0_n, ..., 0_n] \frac{\partial z_{t+h}}{\partial e_t} [I_n, 0_n, ..., 0_n]' = (F^h)_{11} \ .$$

Note that

$$\sum_{j=0}^{\infty} \frac{\partial z_{t+j}}{\partial e_t} = \sum_{j=0}^{\infty} F^j = (I - F)^{-1} \ .$$

Therefore,

$$R = \sum_{j=0}^{\infty} \frac{\partial y_{t+j}}{\partial u_t} = \sum_{j=0}^{\infty} [I_n, 0_n, ..., 0_n](I - F)^{-1} [I_n, 0_n, ..., 0_n]' = \left((I - F)^{-1}\right)_{11} \ .$$

Let M be the inverse of $(I - F)$ and partition it into p^2 blocks of dimension $n \times n$, denoted by $\{M_{lm}\}_{l,m=1:p}$. We have

$$I = M(I - F)$$

$$= \begin{bmatrix} M_{11} & M_{12} & M_{13} & ... & M_{1,p-1} & M_{1p} \\ M_{21} & M_{22} & M_{23} & ... & M_{2,p-1} & M_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{p1} & M_{p2} & M_{p3} & ... & M_{p,p-1} & M_{pp} \end{bmatrix} \begin{bmatrix} I - \alpha_1 A & -\alpha_2 A & -\alpha_3 A & ... & -\alpha_{p-1} A & -\alpha_p A \\ -I_n & I_n & 0_n & ... & 0_n & 0_n \\ 0_n & -I_n & I_n & ... & 0_n & 0_n \\ \vdots & & \ddots & \ddots & & \vdots \\ 0_n & 0_n & ... & -I_n & I_n & 0_n \\ 0_n & 0_n & ... & 0_n & -I_n & I_n \end{bmatrix}$$

As it turns out, the first row of this product is sufficient to solve for the object of interest, $M_{11} = ((I - F)^{-1})_{11}$. Comparing the left- and right-hand sides for the last element, block $(1, p)$, we get

$$0_n = -M_{11} \alpha_p A + M_{1p} \ ,$$

which implies $M_{1p} = M_{11} \alpha_p A$. For elements $l = 2, ..., p - 1$ we get

$$0_n = -M_{11} \alpha_l A + M_{1l} - M_{1,l+1} \ ,$$

which implies

$$M_{12} = M_{11} \alpha_2 A + M_{13} = M_{11} \alpha_2 A + M_{11} \alpha_3 A + M_{14} = ... = M_{11}(\alpha_2 + ... + \alpha_p) A \ .$$

The first element gives

$$I_n = M_{11}(I - \alpha_1 A) - M_{12} = M_{11}\left(I - (\alpha_1 + \alpha_2 + ... + \alpha_p)A\right) = M_{11}(I - aA) \ ,$$

which implies $M_{11} = \left((I-F)^{-1}\right)_{11} = (I - aA)^{-1}$. ■

Note that a process not being observed in every period does not change its long-term response to a permanent increase in the underlying high-frequency innovation. Specifically, if x_τ follows an NVAR$(p,1)$ and $\{y_t\}_{t=1}^T = \{x_{tq}\}_{t=1}^T$ for some $q \in \mathbb{N}\backslash\{1\}$, we have

$$\begin{aligned} lim_{h\to\infty}\left[\frac{\partial y_{t+h}}{\partial v_{tq}} + \frac{\partial y_{t+h}}{\partial v_{tq+1}} + ... + \frac{\partial y_{t+h}}{\partial v_{(t+h)q}}\right] &= lim_{h\to\infty}\sum_{j=0}^{hq}\frac{\partial y_{t+h}}{\partial v_{tq+j}} \\ &= lim_{h\to\infty}\sum_{j=0}^{hq}\frac{\partial x_{(t+h)q}}{\partial v_{tq+j}} \\ &= lim_{h\to\infty}\sum_{j=0}^{hq}\frac{\partial x_{\tau+hq}}{\partial v_{\tau+j}} \\ &= lim_{h\to\infty}\sum_{j=0}^{h}\frac{\partial x_{\tau+h}}{\partial v_{\tau+j}} \\ &= (I - aA)^{-1} \ . \end{aligned}$$

Instead, if $\{y_t\}_{t=1}^T = \{x_{tq} + ... + x_{(t-1)q+1}\}_{t=1}^T$ for some $q \in \mathbb{N}\backslash\{1\}$, we have

$$\begin{aligned} lim_{h\to\infty}\sum_{j=0}^{hq}\frac{\partial y_{t+h}}{\partial v_{tq+j}} &= lim_{h\to\infty}\sum_{j=0}^{hq}\sum_{k=0}^{q-1}\frac{\partial x_{(t+h)q-k}}{\partial v_{tq+j}}\mathbf{1}\left\{j \le hq-k\right\} \\ &= lim_{h\to\infty}\sum_{k=0}^{q-1}\sum_{j=0}^{hq-k}\frac{\partial x_{(t+h)q-k}}{\partial v_{tq+j}} \\ &= q\ lim_{h\to\infty}\sum_{j=0}^{hq-k}\frac{\partial x_{(t+h)q-k}}{\partial v_{tq+j}} \\ &= q\ lim_{h\to\infty}\sum_{j=0}^{h}\frac{\partial x_{\tau+h}}{\partial v_{\tau+j}} \\ &= q(I - aA)^{-1} \ . \end{aligned}$$

Note that if y_t is obtained by averaging instead of summing up, the q in this expression would cancel and we would obtain the same long-run response to a permanent disturbance as that of the underlying high-frequency process x_τ.

B Estimation

B.1 Timing of Network Effects $\alpha|A$

NVAR($p,1$): Asymptotic Properties of $\hat{\alpha}_{OLS}$

The OLS estimator for α from Section 3.1 is given by

$$\hat{\alpha}_{OLS} = \left[\sum_{t=1}^{T} X_t' X_t\right]^{-1} \left[\sum_{t=1}^{T} X_t' y_t\right] = \left[\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right]^{-1} \left[\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}\right] .$$

Large n asymptotics To establish consistency, assume

1. Model is specified correctly: $y_{it} = x_{it}'\alpha + u_{it}$.
2. $\mathbb{E}_{t-1}[u_t] = 0$.
3. The observed network adjacency matrix A_n converges to some limit A in a way so that for all t and for $l,k = 1:p$, as $n \longrightarrow \infty$,
   (a) $\frac{1}{n}\sum_{i=1}^{n} (A_{n,i\cdot}y_{t-l})' (A_{n,i\cdot}y_{t-k}) \longrightarrow \mathbb{E}\left[(A_{i\cdot}y_{t-l})' (A_{i\cdot}y_{t-k})\right]$; and
   (b) $\frac{1}{n}\sum_{i=1}^{n} (A_{n,i\cdot}y_{t-l})' u_{it} \longrightarrow \mathbb{E}\left[(A_{i\cdot}y_{t-l})' u_{it}\right]$.

Then $\hat{\alpha}_{OLS}$ is consistent, i.e. $\hat{\alpha}_{OLS} \longrightarrow \alpha$ as $n \longrightarrow \infty$. By condition 1,

$$\hat{\alpha}_{OLS} = \left[\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right]^{-1} \left[\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\alpha + \frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}u_{it}\right] .$$

Condition 3 ensures that $\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}' \longrightarrow \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_{it}x_{it}']$ and $\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}u_{it} \longrightarrow \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_{it}u_{it}]$ are defined. By condition 2 and the Law of Iterated Expectations (LIE), $\mathbb{E}[x_{it}u_{it}] = \mathbb{E}[x_{it}\mathbb{E}_{t-1}[u_{it}]] = 0$. As usual, assembling these pieces by applying Slutsky's theorem yields consistency.

To establish asymptotic Normality, assume in addition

1. $\mathbb{E}_{t-1}[u_{it}u_{is}] = \sigma^2$ if $t = s$ and zero otherwise.
2. $\forall\, t$ and $l,k = 1:p$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (A_{n,i\cdot}y_{t-l})' u_{it} \Rightarrow N\left(\mathbb{E}\left[(A_{i\cdot}y_{t-l})' u_{it}\right], \mathbb{V}\left[(A_{i\cdot}y_{t-l})' u_{it}\right]\right)$ as $n \longrightarrow \infty$.

Then $\sqrt{n}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \frac{\sigma^2}{T}\mathbb{E}[x_{it}x_{it}']^{-1}\right)$. We have

$$\sqrt{n}(\hat{\alpha}_{OLS} - \alpha) = \left[\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right]^{-1}\left[\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}u_{it}\right] .$$

Condition 2 and Slutsky's theorem ensure that $\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}u_{it} \Rightarrow N\left(\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}\right], \mathbb{V}\left[\frac{1}{T}\sum_{t=1}^{T} x\right.\right.$ As before, $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}\right] = 0$ and so $\mathbb{V}\left[\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}\right] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T} x_{is}u_{is}\right)'\right] = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[x_{it}x_{is}'u_{it}u_{is}] = \frac{\sigma^2}{T}\mathbb{E}[x_{it}x_{it}']$ by condition 1 and LIE.

Large T asymptotics To establish consistency, assume

1. Model is specified correctly: $y_t = X_t\alpha + u_t$.
2. $\mathbb{E}_{t-1}[u_t] = 0$.
3. y_t is ergodic and strictly stationary.

Then $\hat{\alpha}_{OLS}$ is consistent, i.e. $\hat{\alpha}_{OLS} \longrightarrow \alpha$ as $T \longrightarrow \infty$. By condition 1,

$$\hat{\alpha}_{OLS} = \left[\frac{1}{T}\sum_{t=1}^{T} X_t'X_t\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} X_t'X_t\alpha + \frac{1}{T}\sum_{t=1}^{T} X_t'u_t\right] .$$

By the Weak Law of Large Numbers (WLLN) for ergodic and strictly stationary time series (owing to condition 3) and the Continuous Mapping Theorem (CMT), $\frac{1}{T}\sum_{t=1}^{T}(A_n y_{t-l})'(A_n y_{t-k}) \longrightarrow \mathbb{E}\left[(A_n y_{t-l})'(A_n y_{t-k})\right]$ so that $\frac{1}{T}\sum_{t=1}^{T} X_t'X_t \longrightarrow \mathbb{E}[X_t'X_t]$. By condition 2 and LIE, $X_t'u_t$ is a Martingale Difference Sequence (MDS), so that by the WLLN for MDS, $\frac{1}{T}\sum_{t=1}^{T} X_t'u_t \longrightarrow 0$.

To establish asymptotic Normality, assume in addition

1. $\mathbb{E}_{t-1}[u_t u_t'] = \Sigma$.

Then $\sqrt{T}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \mathbb{E}[X_t'X_t]^{-1}\mathbb{E}[X_t'\Sigma X_t]\mathbb{E}[X_t'X_t]^{-1\prime}\right)$. We have

$$\sqrt{T}(\hat{\alpha}_{OLS} - \alpha) = \left[\frac{1}{T}\sum_{t=1}^{T} X_t'X_t\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'u_t\right] .$$

By the Central Limit Theorem (CLT) for ergodic and strictly stationary time series and CMT, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'u_t \Rightarrow N\left(\mathbb{E}[X_t'u_t], \mathbb{V}[X_t'u_t]\right)$. Because $\mathbb{E}[X_t'u_t] = \mathbb{E}[X_t'\mathbb{E}_{t-1}[u_t]] = 0$, $\mathbb{V}[X_t'u_t] = \mathbb{E}[X_t'u_t u_t'X_t] = \mathbb{E}[X_t'\Sigma X_t]$ by condition 1 and LIE.

Note that if $\Sigma = \sigma^2 I$, this boils down to $\sqrt{T}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \sigma^2\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}']^{-1}\right)$. If in addition, $\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}']$ can be written as $n\mathbb{E}[x_{it}x_{it}']$, the asymptotic variance becomes $\frac{\sigma^2}{n}\mathbb{E}[x_{it}x_{it}']^{-1}$.

Large (n, T) asymptotics To establish consistency, assume either the conditions for consistency under large n asymptotics or the conditions for consistency under large T asymptotics. Then $\hat{\alpha}_{OLS} \longrightarrow \alpha$ as $n, T \longrightarrow \infty$. To establish asymptotic Normality, assume either the conditions under large n asymptotics or i) the conditions under large T asymptotics, ii) $\Sigma = \sigma^2 I$, and iii) $\mathbb{E}[\sum_{i=1}^n x_{n,it} x'_{n,it}] = n\mathbb{E}[x_{it} x'_{it}]$. Then, $\sqrt{nT}(\hat{\alpha}_{OLS} - \alpha) \Rightarrow N\left(0, \sigma^2 \mathbb{E}[x_{it} x'_{it}]^{-1}\right)$ as $n, T \longrightarrow \infty$.

NVAR(p, q), $q > 1$: Identification

With A given, the identification problem under $x_\tau \sim$ NVAR$(p, 1)$ and $\{y_t\}_{t=1}^T = \{x_{tq}\}_{t=1}^T$ for $q > 1$ is akin to that for $\tilde{x}_\tau \sim$ AR(p) and $\{\tilde{y}_t\}_{t=1}^T = \{\tilde{x}_{tq}\}_{t=1}^T$. For example, under $p = 1$ and $q = 2$, we get

$$y_t = \alpha_1^2 A^2 y_{t-1} + \eta_t \ , \quad \text{and} \quad \tilde{y}_t = \alpha_1^2 \tilde{y}_{t-1} + \tilde{\eta}_t \ ,$$

respectively, and in both cases α_1 is identified only up to sign. While characterization of the identified set remains elusive for the former case for all but $(p = 1, q = 2)$, in the latter case it can be analyzed for $q = 2$ and general p.

Let $\gamma_h = \mathbb{E}[\tilde{x}_t \tilde{x}_{t-h}] = \gamma_{-h}$, which can be estimated by the analogy principle from $\hat{\gamma}_h = \frac{1}{T-h} \sum_{t=h+1}^T \tilde{x}_t \tilde{x}_{t-h}$. Under $q = 2$, $\hat{\gamma}_h$ is observed only for h even (and zero). The Yule-Walker equations for an AR(p) lead to the system

$$\begin{bmatrix} \gamma_0 - \sigma^2 & \gamma_1 & \ldots & \gamma_m \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{bmatrix} \begin{bmatrix} \gamma_1 & \gamma_0 & \cdots & & & \gamma_{m-1} \\ \gamma_2 & \gamma_1 & \ddots & & & \vdots \\ \vdots & \vdots & \ddots & & & \\ \gamma_p & \gamma_{p-1} & \cdots & \gamma_1 & \gamma_0 & \cdots \end{bmatrix} ,$$

for $m \geq p - 1$. In principle, this system of (nonlinear) equations could be solved for the unknowns $\{\alpha_l\}_{l=1:p}$ and $\{\gamma_h\}_{h=1,3,\ldots}$. However, the following analysis suggests that $\{\alpha_l\}_{l=1,3,\ldots}$ and $\{\gamma_h\}_{h=1,3,\ldots}$ are identified only up to sign.

Let $\underline{m}$ be the largest odd number in $1 : m$ and $\overline{m}$ the largest even one. We can write

$$\begin{bmatrix} \gamma_1 & \gamma_3 & \ldots & \gamma_{\underline{m}} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_2 & \gamma_4 & \cdots & \gamma_{\underline{m}-1} \\ \gamma_1 & \gamma_1 & \gamma_3 & & \\ \vdots & \vdots & \vdots & & \\ \gamma_{p-1} & \gamma_{p-3} & & & \end{bmatrix} ,$$

and therefore

$$\begin{bmatrix} \gamma_1 \\ \gamma_3 \\ \ldots \\ \gamma_{\underline{m}} \end{bmatrix} = \overline{A} \begin{bmatrix} \gamma_1 \\ \gamma_3 \\ \ldots \\ \gamma_{\underline{m}} \end{bmatrix} + \underline{A} \begin{bmatrix} \gamma_0 \\ \gamma_2 \\ \ldots \\ \gamma_{\overline{m}} \end{bmatrix} = (I - \overline{A})^{-1} \underline{A} \begin{bmatrix} \gamma_0 \\ \gamma_2 \\ \ldots \\ \gamma_{\overline{m}} \end{bmatrix} , \tag{A.1}$$

where only α_l for l even appear in $\overline{A}$ (and its elements are linear in α), and only α_l for l odd appear in $\underline{A}$. The remaining equations give

$$\begin{bmatrix} \gamma_0 - \sigma^2 & \gamma_2 & \ldots & \gamma_{\overline{m}} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_p \end{bmatrix} \begin{bmatrix} \gamma_1 & \gamma_1 & \gamma_3 & \cdots & \gamma_{\overline{m}-1} \\ \gamma_2 & \gamma_0 & \gamma_2 & & \\ \vdots & \vdots & \vdots & & \\ \gamma_p & \gamma_{p-2} & & & \end{bmatrix} ,$$

and therefore

$$\begin{bmatrix} \gamma_0 - \sigma^2 \\ \gamma_2 \\ \ldots \\ \gamma_{\overline{m}} \end{bmatrix} = \underline{B} \begin{bmatrix} \gamma_1 \\ \gamma_3 \\ \ldots \\ \gamma_{\underline{m}} \end{bmatrix} + \overline{B} \begin{bmatrix} \gamma_0 \\ \gamma_2 \\ \ldots \\ \gamma_{\overline{m}} \end{bmatrix} , \tag{A.2}$$

where again only α_l for l even appear in $\overline{B}$, and only α_l for l odd appear in $\underline{B}$. Eq. (A.1) and Eq. (A.2) illustrate that multiplying $(\alpha_1, \alpha_3, \ldots)$ as well as $(\gamma_1, \gamma_3, \ldots)$ by (-1) does not change the system of equations.

NVAR(p, q), $q > 1$: Implementation

For simplicity, in the following, assume Σ is known (or OLS estimation is considered). The EM algorithm constructs the incomplete data-objective function by integrating out the unobserved states $x_{1:T_\tau}$:

$$\hat{\alpha} = \arg\max_{\alpha} \int Q(Y_{1:T}, x_{1:T_\tau}; \alpha) dx_{1:T_\tau} \ .$$

For example, $Q(Y_{1:T}, x_{1:T_\tau}; \alpha)$ could be the incomplete data-likelihood $p(Y_{1:T}, x_{1:T_\tau} | \alpha)$ or it could be the (negative of the) OLS objective function (i.e. Eq. (9) with $\Sigma = I$). Practically, the EM algorithm involves iterating on two steps until convergence: i) given $\hat{\alpha}$, run the Kalman Smoother to obtain $\hat{x}_{1:T_\tau} \equiv \mathbb{E}[x_{1:T_\tau} | \hat{\alpha}, y_{1:T}]$, and ii) given $\hat{x}_{1:T_\tau}$, construct $Q(Y_{1:T}, \hat{x}_{1:T_\tau}; \alpha)$ and find $\hat{\alpha} = \arg\max_\alpha Q(Y_{1:T}, \hat{x}_{1:T_\tau}; \alpha)$.

The Gibbs sampler of Carter and Kohn (1994) yields the joint posterior $p(\alpha, x_{1:T_\tau} | Y_{1:T})$, with marginal posterior

$$p(\alpha|Y_{1:T}) = \int p(\alpha, x_{1:T_\tau}|Y_{1:T})dx_{1:T_\tau} = \int p(Y_{1:T}, x_{1:T_\tau}|\alpha)p(\alpha)p(x_{1:T_\tau})dx_{1:T_\tau} \ .$$

With distributional assumptions such that $p(Y_{1:T}, x_{1:T_\tau}|\alpha)p(\alpha)p(x_{1:T_\tau}) \propto Q(Y_{1:T}, x_{1:T_\tau}; \alpha)$, $\hat{\alpha}$ from above is equal to the mode of $p(\alpha|Y_{1:T})$. For example, under $v_\tau \sim N(0, I)$, $p(\alpha)p(x_{1:T_\tau}) \propto c$, we get the OLS estimator. See derivations in Appendix B.2. Practically, draws from $p(\alpha, x_{1:T_\tau}|Y_{1:T})$ are obtained by iteratively drawing from the conditional posteriors $p(\alpha|Y_{1:T}, x_{1:T_\tau})$ and $p(x_{1:T_\tau}|Y_{1:T}, \alpha)$, with the latter obtained from the Kalman Smoother. The marginal posterior $p(\alpha|Y_{1:T})$ is then obtained by simply ignoring the drawn values that pertain to $x_{1:T_\tau}$.

In case Σ is estimated as well, write $\theta = (\alpha, \Sigma)$ instead of α in $Q(Y_{1:T}, x_{1:T_\tau}; \alpha)$, $\mathbb{E}[x_{1:T_\tau}|\alpha, y_{1:T}]$ and all densities above. In the EM algorithm, add an additional iteration step for $\hat{\Sigma}|\hat{\alpha}, \hat{x}_{1:T_\tau}$. In the Gibbs sampler, add a step to draw from $p(\Sigma|Y_{1:T}, \alpha, x_{1:T_\tau})$. As shown in Appendix B.2, under $v_\tau \sim N(0, \Sigma)$, $p(\alpha)p(x_{1:T_\tau})p(\Sigma) \propto c$ we get the LS estimator.

B.2 Joint Inference: Network & Timing (α, A)

NVAR$(p, 1)$: Derivation of $(\hat{\alpha}_{LS}, \hat{A}_{LS})$ as Posterior Modes

Under $u_t \sim N(0, \Sigma)$, the (conditional) likelihood associated with the NVAR$(p, 1)$ is

$$\begin{aligned} p(Y_{1:n,1:T}|\alpha, Y_{1:n,-p+1:0}) &= \prod_{t=1}^{T} p(y_t|\theta, y_{t-p:t-1}) \\ &= \prod_{t=1}^{T} (2\pi)^{-n/2}|\Sigma|^{-1/2} exp\left\{-\frac{1}{2}u_t(\alpha)'\Sigma^{-1}u_t(\alpha)\right\} \\ &= (2\pi)^{-nT/2}|\Sigma|^{-T/2} exp\left\{-\frac{1}{2}\sum_{t=1}^{T} u_t(\alpha)'\Sigma^{-1}u_t(\alpha)\right\} \ , \end{aligned}$$

where $u_t(\alpha) = y_t - \sum_{l=1}^{p} \alpha_l A y_{t-l}$. Under the priors $a_{ij} \sim N(b_{ij}, \lambda^{-1})$, and $p(\alpha, \Sigma) \propto c$, the posterior of (α, A, Σ) is

$$\begin{aligned} p(\alpha, A, \Sigma|Y) &\propto p(Y|\alpha, A, \Sigma)p(\alpha)p(A)p(\Sigma) \\ &\propto |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t\right\} exp\left\{-\frac{1}{2}\lambda\sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2\right\} \ , \end{aligned}$$

with $u_t(\alpha, A) = y_t - \sum_{l=1}^{p} \alpha_l A y_{t-l}$. Under $\lambda = nT\tilde{\lambda}$, the negative of the logarithm of this posterior is proportional to the objective function in the LS minimization problem in Eq. (12). Therefore, with the proper re-scaling of the penalty-parameter, the (joint) minimzer of the objective function in Eq. (12), $(\hat{\alpha}_{LS}, \hat{A}_{LS})|\Sigma$, is equal to the mode of the joint posterior $p(\alpha, A|Y, \Sigma)$, while the mode of the conditional posterior $p(A|Y, \alpha, \Sigma)$ is equal to $\hat{A}_{LS}|\alpha, \Sigma$ and the mode of $p(\alpha|Y, A, \Sigma)$ is equal to $\hat{\alpha}_{LS}|A, \Sigma$. The OLS estimator $(\hat{\alpha}_{OLS}, \hat{A}_{OLS})$ is the mode of $p(\alpha, A|Y)$, where $u_t \sim N(0, \Sigma)$ with $\Sigma = I$ is assumed in the likelihood.

The conditional posterior of $\alpha|\Sigma$ is

$$\begin{aligned} p(\alpha|Y,\Sigma) &\propto p(Y|\alpha,\Sigma)p(\alpha) \\ &\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\alpha)'\,\Sigma^{-1}\,(y_t - X_t\alpha)\right\} \\ &\propto exp\left\{-\frac{1}{2}\left(\sum_{t=1}^{T}(y_t - X_t\alpha)'\,\Sigma^{-1}\,(y_t - X_t\alpha)\right)\right\} \\ &\propto exp\left\{-\frac{1}{2}\left\{\alpha'\left[\sum_{t=1}^{T}X_t'\Sigma^{-1}X_t\right]\alpha - 2\alpha'\left[\sum_{t=1}^{T}X_t'\Sigma^{-1}y_t\right]\right\}\right\} \ , \end{aligned}$$

which shows that

$$\alpha|Y,\Sigma \sim N\left(\bar{\alpha}, \bar{V}_\alpha\right) \ , \quad \text{with} \quad \bar{V}_\alpha = \left[\sum_{t=1}^{T}X_t'\Sigma^{-1}X_t\right]^{-1} \ , \quad \bar{\alpha} = \bar{V}_\alpha\left[\sum_{t=1}^{T}X_t'\Sigma^{-1}y_t\right] \ .$$

The conditional posterior of $A|\alpha, \Sigma$ is

$$\begin{aligned} p(A|Y,\alpha,\Sigma) &\propto p(Y|\alpha,A,\Sigma)p(A) \\ &\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\alpha)'\,\Sigma^{-1}\,(y_t - X_t\alpha)\right\} exp\left\{-\frac{1}{2}\lambda\sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2\right\} \\ &= exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(Y - ZA'\right)'\left(Y - ZA'\right)\right]\right\} exp\left\{-\frac{1}{2}\lambda tr[(A-B)'(A-B)]\right\} \ , \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(-A(Z'Y + \lambda B'\Sigma) - (Y'Z + \lambda\Sigma B)A' + A(Z'Z + \lambda\Sigma)A'\right)\right]\right\} \ ,^{58} \end{aligned}$$

which lets us deduce that

$$A'|Y,\alpha,\Sigma \sim MN\left(\bar{A}, \bar{U}_A, \bar{V}_A\right) \ , \quad \text{with} \ \bar{U}_A = \left[Z'Z + \lambda\Sigma\right]^{-1} \ , \ \bar{A} = \bar{U}_A\left[Z'Y + \lambda B'\Sigma\right] \ , \ \bar{V}_A = \Sigma \ ,$$

and therefore

$$A|Y,\alpha,\Sigma \sim MN\left(\bar{A}', \bar{V}_A, \bar{U}_A\right) \ .$$

[58] Note that $\sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2 = vec(A-B)'vec(A-B) = tr[(A-B)'(A-B)]$. Also, it holds that

Neither $(\hat{\alpha}_{OLS}, \hat{A}_{OLS})$ nor $p(\alpha, A|Y)$ are available analytically. The former can be obtained by iterating on the conditional estimators $\hat{\alpha}_{LS}|A$ and $\hat{A}_{LS}|\alpha$ until convergence (see Meng and Rubin (1993)). The latter can be obtained analogously by Gibbs sampling, i.e. iteratively drawing from the conditional posteriors $p(\alpha|Y, A)$ and $p(A|Y, \alpha)$.[59]

Note that the mode of $p(\Sigma|Y, \alpha)$ is equal to $\hat{\Sigma}|\alpha = \frac{1}{T}\sum_{t=1}^{T} u_t(\alpha)u_t(\alpha)'$:

$$\begin{aligned} p(\Sigma|Y, \alpha) &\propto p(Y|\alpha, \Sigma) \\ &\propto |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma^{-1}u_t\right\} \\ &= |\Sigma|^{-T/2} exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}U'U\right]\right\} , \end{aligned}$$

where U is $T \times n$ and stacks u_t' along rows. This shows that

$$\Sigma|Y, \alpha \sim \mathcal{IW}(\bar{S}, \bar{v}) , \quad \bar{S} = U'U , \quad \bar{v} = T .$$

As a result, under these prior choices, the mode of $p(\alpha, A, \Sigma|Y)$ is equal to the (GLS) estimator $(\hat{\alpha}, \hat{A}, \hat{\Sigma})$,[60] obtained by iterating on the conditional estimators $\hat{\alpha}_{LS}|\Sigma$ and $\hat{\Sigma}|\alpha$ until convergence (see Meng and Rubin (1993)). Again, Bayesian inference can be implemented analogously by Gibbs sampling, i.e. iteratively drawing from the conditional posteriors $p(\alpha|Y, \Sigma)$ and $p(\Sigma|Y, \alpha)$ to obtain the joint posterior $p(\alpha, \Sigma|Y)$.

Lasso As can be verified, under the priors $a_{ij} \sim$ Exponential(λ) instead of $a_{ij} \sim N(b_{ij}, \lambda^{-1})$, for $\lambda = nT\tilde{\lambda}/2$, $-ln\ p(\alpha, A, \Sigma|Y)$ is proportional to the LS objective function in Eq. (12) with restrictions $a_{ij} \geq 0$ and a Lasso-penalty inducing selection to zero of $\{a_{ij}\}_{i,j=1:n}$, $\tilde{\lambda}\sum_{i,j=1}^{n} |a_{ij}| = \tilde{\lambda}\sum_{i,j=1}^{n} a_{ij}$. As a result, $\hat{A}_{LS}|\alpha, \Sigma$ from this problem is equal to the mode of $p(A|Y, \alpha, \Sigma)$. We get

$$\begin{aligned} p(A|Y, \alpha, \Sigma) &\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - Az_t)'\Sigma^{-1}(y_t - Az_t)\right\} exp\left\{-\lambda\iota' A\iota\right\} \\ &= exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}(Y - ZA')'(Y - ZA')\right]\right\} exp\left\{-\lambda\iota' A\iota\right\} \\ &\propto exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}\left[AZ'ZA' - 2A\left(Z'Y - \lambda\iota\iota'\Sigma\right)\right]\right]\right\} , \end{aligned}$$

$tr[AB] = tr[BA]$, $tr[A] = tr[A']$ and $c\ tr[A] = tr[cA]$.

[59]Also, as in Appendix B.1, under $u_t \sim N(0, \Sigma)$ and $p(\Sigma) \propto c$, the mode of $p(\alpha, A, \Sigma|Y)$ is equal to the GLS estimator $(\hat{\alpha}, \hat{A}, \hat{\Sigma})$, obtained by iterating on $\hat{\alpha}_{LS}|A, \Sigma$, $\hat{A}_{LS}|\alpha, \Sigma$ and $\hat{\Sigma}|\alpha, A = \frac{1}{T}\sum_{t=1}^{T} u_t(\alpha)u_t(\alpha)'$ until convergence.

[60]This is the joint minimizer of the objective function in Eq. (9).

where ι is an n-dimensional vector of ones.[61] This leads to

$$A'|Y,\alpha,\Sigma \sim N\left(\bar{A},\Sigma\otimes\bar{P}^{-1}\right) \ , \quad \text{truncated to } \mathbb{R}_+^{n^2} \ , \quad \text{with } \bar{P}=Z'Z \ , \quad \bar{A}=\bar{P}^{-1}\left[Z'Y-\lambda\iota\iota'\Sigma\right] \ .$$

Under OLS, the estimators $\hat{a}_{ij,LS}|Y,\alpha$ are independent across i. We get

$$\hat{a}_{ij,OLS}|(A_{i,-j},\alpha)=max\{0,\breve{a}_{ij}\} \ , \quad \breve{a}_{ij}=\frac{\sum_{t=1}^{T}(y_{it}-A_{i,-j}z_{-j,t})z_{jt}-\lambda}{\sum_{t=1}^{T}z_{jt}^2} \ .$$

Hyperparameter Selection If λ is treated as an additional parameter to estimate, in spirit of hierarchical modeling, with a flat prior $p(\lambda)\propto c$,[62] its (marginal) posterior coincides in shape with the marginal likelihood, a measure of forecasting accuracy, so that their posterior mode maximizes forecasting performance (see Giannone et al. (2015)). We get

$$p(\lambda|Y,\alpha,A,\Sigma)\propto\lambda^{\frac{n}{2}}exp\left\{-\frac{1}{2}\lambda\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\right\} \ ,$$

which, by maximization, leads to the mode(s) stated in the main text.

NVAR$(p,1)$: Asymptotic Analysis

Let $\theta=(\alpha,A)$. Under OLS, the optimization problem in Eq. (12) is

$$\hat{\theta}_{LS}=\arg\min_{\theta\in\Theta} Q_{n,T}(\theta;Y) \ ,$$
$$\text{with } \ Q(\theta;Y)=\frac{1}{nT}\sum_{t=1}^{T}u_t(\theta)'u_t(\theta)+\tilde{\lambda}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2 \ ,$$

and $u_t(\theta)=y_t-A\tilde{X}_t\alpha$. To render (α,A) identified, fix α_l for some l and drop it from α, with appropriate redefinitions of y_t and $\tilde{X}_t$ (and z_t and X_t). Under the alternative normalization $||\alpha||_1=1$, the following consistency results would go through, but the interior-requirement for asymptotic Normality would be violated.

For consistency under $T\longrightarrow\infty$, take $\Theta=[-c,c]^{p-1+n^2}$ for $c>0$ large such that $\Theta\subset\mathbb{R}^{p-1+n^2}$ is compact, and assume

1. $\tilde{\lambda}_{n,T}=o(1)$.
2. y_t is ergodic and strictly stationary.

[61] Note that $\sum_{i,j=1}^{n}a_{ij}=\iota'A\iota$. On top of the rules for $tr[\cdot]$ referenced above, here I also used $a'Ba=tr[Baa']$.

[62] For example, $\lambda\sim\mathcal{U}(\underline{\lambda},\overline{\lambda})$ for some bounds wide enough.

3. $\mathbb{E}[X_t'X_t]$ and $\mathbb{E}[z_t z_t']$ are of full rank.

By conditions 1 and 2, $Q_{n,T}(\theta;Y)$ converges uniformly in probability to the limit objective function

$$Q(\theta) = \frac{1}{n}\mathbb{E}\left[u_t(\theta)'u_t(\theta)\right] \ ,$$

which is continuous on Θ:

$$\begin{aligned}
&\underset{\theta\in\Theta}{sup}\,\Big|\frac{1}{nT}\sum_{t=1}^{T}u_t(\theta)'u_t(\theta) - \frac{1}{n}\mathbb{E}\left[u_t(\theta)'u_t(\theta)\right] + \tilde{\lambda}_{n,T}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\Big| \\
\leq &\frac{1}{n}\underset{\theta\in\Theta}{sup}\,\Big|\frac{1}{T}\sum_{t=1}^{T}u_t(\theta)'u_t(\theta) - \mathbb{E}\left[u_t(\theta)'u_t(\theta)\right]\Big| + \underset{\theta\in\Theta}{sup}\,\Big|\tilde{\lambda}_{n,T}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\Big|
\end{aligned}$$

converges in probability to zero because, under condition 1,

$$\underset{\theta\in\Theta}{sup}\,\Big|\tilde{\lambda}_{n,T}\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\Big| = \tilde{\lambda}_{n,T}\sum_{i,j=1}^{n}(c+b_{ij})^2 \leq \tilde{\lambda}_{n,T}\sum_{i,j=1}^{n}\tilde{c} = \tilde{\lambda}_{n,T}n^2\tilde{c} \longrightarrow 0 \ ,$$

where $\tilde{c} = \max_{i,j}(c+b_{ij})^2$, while under condition 2, $\frac{1}{T}\sum_{t=1}^{T}u_t(\theta)'u_t(\theta) \longrightarrow \mathbb{E}\left[u_t(\theta)'u_t(\theta)\right]$ by WLLN for ergodic and strictly stationary time series and CMT. Finally, under condition 3, $Q(\theta)$ is uniquely minimized by $\theta_0 = (\alpha_0, A_0)$ defined by

$$\alpha_0|A_0 = \mathbb{E}[X_t(A_0)'X_t(A_0)]^{-1}\mathbb{E}[X_t(A_0)'y_t] \ , \quad A_0|\alpha_0 = \mathbb{E}[y_t z_t(\alpha_0)']\mathbb{E}[z_t(\alpha_0)z_t(\alpha_0)']^{-1} \ .$$

This can be seen by taking first-order conditions (FOC) and noting that for c large enough, we necessarily get a solution that is interior on Θ. Note that without the imposed normalization, θ_0 would not be unique, as for any (α_0, A_0) that solves the above, $(k\alpha_0, k^{-1}A_0)$ for any $k \in \mathbb{R}$ does, too, because $X_t(k^{-1}A_0) = k^{-1}X_t(A_0)$ and $z_t(k\alpha_0) = kz_t(\alpha_0)$.[63]

To establish asymptotic Normality, assume in addition

1. Model is specified correctly: $y_t = A\tilde{X}_t\alpha + u_t$.
2. $\mathbb{E}_{t-1}[u_t] = 0$.
3. $\mathbb{E}_{t-1}[u_t u_t'] = \Sigma$.
4. $\tilde{\lambda}_{n,T} = o(T^{-\frac{1}{2}})$.

[63] With α_l dropped, both of these statements still hold, but the y_t in the expression for $\alpha_0|A_0$ is in fact $y_t - Ay_{t-l}$, while it is unchanged in the expression for $A_0|\alpha_0$.

As mentioned above, for c large enough, $\theta_0 \in int(\Theta)$. Write $\overrightarrow{A}$ for $vec\,(A)$. By condition 4, CLT for ergodic and strictly stationary time series and CMT,

$$\sqrt{T}Q(1)_{n,T}(\theta_0; Y) = \sqrt{T}\left.\begin{bmatrix}\frac{\partial Q_{n,T}(\theta;Y)}{\partial\alpha}\\ \frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}}\end{bmatrix}\right|_{\theta=\theta_0} = -2\begin{bmatrix}\frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t'(y_t - X_t\alpha_0)\\ \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overrightarrow{[(y_t - A_0z_t)z_t']} - \sqrt{T}\tilde{\lambda}\overrightarrow{[A_0 - B]}\end{bmatrix} \Rightarrow N(\mu, M)\ ,$$

with μ, M given below. Thereby, $\sqrt{T}\tilde{\lambda} \longrightarrow 0$ by condition 4. Using conditions 1 and 2,

$$\mu = -\frac{2}{n}\begin{bmatrix}\mathbb{E}\left[X_t'(y_t - X_t\alpha)\right]\\ \mathbb{E}\overrightarrow{[(y_t - A_0z_t)z_t']}\end{bmatrix} = -\frac{2}{n}\begin{bmatrix}\mathbb{E}\left[X_t'u_t\right]\\ \mathbb{E}\overrightarrow{[u_tz_t']}\end{bmatrix} = 0\ .$$

Using conditions 1, 2 and 3 as well as LIE,

$$M = \frac{4}{n^2}\begin{bmatrix}\mathbb{E}\left[X_t'u_tu_t'X_t\right] & \cdot\\ \mathbb{E}\left[\overrightarrow{[u_tz_t']}u_t'X_t\right] & \mathbb{E}\left[\overrightarrow{[u_tz_t']}\,\overrightarrow{[u_tz_t']}'\right]\end{bmatrix} = \frac{4}{n^2}\begin{bmatrix}\mathbb{E}\left[X_t'\Sigma X_t\right] & \cdot\\ \mathbb{E}\left[z_t\otimes\Sigma X_t\right] & \mathbb{E}\left[z_tz_t'\right]\otimes\Sigma\end{bmatrix}\ .$$

Furthermore, using again the WLLN for ergodic and strictly stationary time series, CMT and conditions 1 and 2,

$$Q^{(2)}_{n,t}(\theta_0; Y) = \left.\begin{bmatrix}\frac{\partial Q_{n,T}(\theta;Y)}{\partial\alpha\partial\alpha'} & \frac{\partial Q_{n,T}(\theta;Y)}{\partial\alpha\partial\overrightarrow{A}'}\\ \frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}\partial\alpha'} & \frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}\partial\overrightarrow{A}'}\end{bmatrix}\right|_{\theta=\theta_0} \longrightarrow \begin{bmatrix}H_{11} & H_{21}'\\ H_{21} & H_{22}\end{bmatrix} \equiv H\ ,$$

with

$$\frac{\partial Q_{n,T}(\theta;Y)}{\partial\alpha\partial\alpha'} = \frac{2}{nT}\sum_{t=1}^{T}X_t'X_t \longrightarrow \frac{2}{n}\mathbb{E}[X_t'X_t] = H_{11}\ ,$$

$$\frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}\partial\alpha'} = -\frac{2}{nT}\sum_{t=1}^{T}\begin{bmatrix}(y_t - X\tilde{X}_t\alpha)\tilde{X}_{t,1\cdot} - z_{1t}A\tilde{X}_t\\ \vdots\\ (y_t - X\tilde{X}_t\alpha)\tilde{X}_{t,n\cdot} - z_{nt}A\tilde{X}_t\end{bmatrix} \longrightarrow \frac{2}{n}\begin{bmatrix}A\mathbb{E}[z_{1t}\tilde{X}_t]\\ \vdots\\ A\mathbb{E}[z_{nt}\tilde{X}_t]\end{bmatrix} = H_{21}\ ,$$

$$\frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}\partial\overrightarrow{A}'} = \frac{2}{nT}\sum_{t=1}^{T}\begin{bmatrix}z_t'\otimes z_{1t}I_n\\ \vdots\\ z_t'\otimes z_{nt}I_n\end{bmatrix} \to \frac{2}{n}\begin{bmatrix}\mathbb{E}[z_t'\otimes z_{1t}I_n]\\ \vdots\\ \mathbb{E}[z_t'\otimes z_{nt}I_n]\end{bmatrix} = H_{22}\ .$$[64]

Overall, we obtain

$$\sqrt{T}(\hat{\theta}_{LS} - \theta_0) \Rightarrow N(0, H^{-1}MH^{-1}) \ .$$

C Input-Output Links & Sectoral Price Dynamics

C.1 Structural Model Details

Contemporaneous Input-Output Conversion

In this case, the amount of good j purchased at t and used in the production at t coincide: $x_{ijt} = x_{t,t}^{ij} = x_t^{ij}$. I will write x^{ij} for this quantity. Because the environment is static, I drop time subscripts for notational simplicity. Firm i solves the problem

$$\max_{l_i, \{x^{ij}\}_{j=1}^n} p_i z_i l_i^{b_i} \prod_{j=1}^n \left(x^{ij}\right)^{a_{ij}} - w l_i - \sum_{j=1}^n p_j x^{ij} \ .$$

The first-order conditions (FOCs) w.r.t. l_i and x^{ij} give

$$l_i = b_i \frac{p_i y_i}{w} \ , \quad x^{ij} = a_{ij} \frac{p_i y_i}{p_j} \ .$$

The latter FOC provides an interpretation of $a_{ij} = (p_j x^{ij})/(p_i y_i)$ as the amount of good j purchased by sector i divided by the total output of sector i. Plugging these expressions into the production function and taking logs yields

$$ln\left(p_i/w\right) = k_i^p + \sum_{j=1}^n a_{ij} ln\left(p_j/w\right) + \varepsilon_i \ ,$$

where $\varepsilon_i = -ln(z_i)$ and the constant $k_i^p = -\left[b_i ln(b_i) + \sum_{j=1}^n a_{ij} ln(a_{ij})\right]$ reflects differences in the reliance on different production factors across sectors i. Stacking this expression for all sectors i yields the equation for sectoral prices in the main text.

---

[64] To see this, note that $[\overrightarrow{(y_t - Az_t)z_t'}]$ consists of n stacked vectors with the one in position l given by $(y_t - Az_t)z_{lt} = (y_t - A\tilde{X}_t\alpha)\tilde{X}_{t,l\cdot}\alpha$, whose derivativate w.r.t α' is $(y_t - A\tilde{X}_t\alpha)\tilde{X}_{t,l\cdot} + z_{1t}A\tilde{X}_t$. Moreover, note that $[\overrightarrow{Az_tz_t'}]$ consists of vectors of the form $Az_tz_{lt} = [A_{1\cdot}z_tz_{lt}, ..., A_{n\cdot}z_tz_{lt}]'$ whose derivative w.r.t. A gives $z_t' \otimes z_{lt}I_n$.

The representative household's problem is

$$\underset{\{c_i\}_{i=1}^n}{max} \sum_{i=1}^n \gamma_i \, ln(c_i/\gamma_i) \; , \quad \text{s.t.} \; \sum_{i=1}^n p_i c_i = w \; .$$

The FOC yields $c_i = \gamma_i \frac{w}{p_i}$. Hence, γ_i is the share of good i in households' expenditures.

The market clearing condition for good j reads $y_j = c_j + \sum_{i=1}^n x^{ij}$. Plugging in the expressions for c_j and x^{ij} and multiplying by p_j/w yields the following expression for the Domar weight of sector j, λ_j:

$$\lambda_j \equiv \frac{y_j p_j}{w} = \gamma_j + \sum_{i=1}^n a_{ij} \lambda_i \; .$$

As a result, the vector of sectoral Domar weights is $\lambda = (I - A')^{-1}\gamma$. The Domar weight of sector i reflects its importance as a supplier to relevant sectors in the economy, with relevance given by households' expenditure share: $\lambda_i = \sum_{j=1}^n \gamma_j l_{ji}$. In this expression, l_{ij} is element (i, j) of the Leontief-inverse $(I - A)^{-1}$. It sums up connections of all order from a sector i to a sector j and therefore shows how important sector j is in i's supply chain. This relation holds regardless of TFP levels in ε. Using the definition of λ_i, we get the following expression for output:

$$ln\,(y) = ln\,(\lambda) - ln\left(\frac{p}{w}\right) = k^y + A ln\,(y) - \varepsilon \; ,$$

with $k^y = (I - A) ln\,(\lambda) - k^p$. The labor market clearing condition reads $\sum_{i=1}^n l_{it} = 1$ and gives $w_t = \sum_{i=1}^n p_{it} y_{it}$, but it can be ignored by Walras' law.

In the unperturbed state $\varepsilon = 0$, we get

$$ln\,(p/w) = (I - A)^{-1} k^p \; ,$$

and $ln\,(y_i) = ln\,(\lambda_i) - ln\,(p_i/w)$.

Single-Lag Input-Output Conversion

Assume good j used in production at time t is purchased at time $t-1$: $x_{ijt} = x^{ij}_{t,t-1} = x^{ij}_{t-1}$. I will write x^{ij}_{t-1} for this quantity. An extension to general lags $p \geq 1$ is straightforward. Firm i's problem is then

$$\underset{\{l_{it}, \{x^{ij}_t, x^{ij}_{t,t-1}\}_{j=1}^n\}_{t=0}^\infty}{max} \sum_{t=0}^\infty \beta^t \left[p_{it} z_{it} l_{it}^{b_i} \prod_{j=1}^n \left(x^{ij}_{t-1}\right)^{a_{ij}} - w l_{it} - \sum_{j=1}^n p_{jt} x^{ij}_{t-1} \right] \; .$$

This leads to the following value function:

$$V_i\left(\left\{x_{t-1}^{ij}\right\}_{j=1}^{n}\right)=\underset{l_{it},\{x_t^{ij}\}_{j=1}^{n}}{max}\; p_{it}z_{it}l_{it}^{b_i}\prod_{j=1}^{n}\left(x_{t-1}^{ij}\right)^{a_{ij}}-w_tl_{it}-\sum_{j=1}^{n}p_{jt}x_{t-1}^{ij}+\beta V_i\left(\left\{x_t^{ij}\right\}_{j=1}^{n}\right)\ .$$

The FOC w.r.t. l_{it} and x_t^{ij} give

$$l_{it}=b_i\frac{p_{it}y_{it}}{w_t}\quad ,\quad x_t^{ij}=\beta a_{ij}\frac{p_{i,t+1}y_{i,t+1}}{p_{jt}}\ .$$

Note that in steady state, the latter expression yields $a_{ij}=\beta^{-1}(p_jx^{ij})/(p_iy_i)$, which means that the meaning of a_{ij} changes slightly compared to the case of contemporaneous input usage before. Plugging these expressions into the production function and taking logs gives

$$ln\left(\frac{p_{it}}{w_t}\right)=k_{it}^{p1}+\sum_{j=1}^{n}a_{ij}ln\left(\frac{p_{j,t-1}}{w_{t-1}}\right)+\varepsilon_{it}\ ,$$

where again $\varepsilon_{it}=-ln(z_{it})$ and $k_{it}^{p1}=-\left[b_iln(b_i)+\sum_{j=1}^{n}a_{ij}ln(\beta a_{ij})+(1-b_i)ln(G_t^w)\right]$ with $G_t^w=w_t/w_{t-1}$. Stacking this expression for all sectors i gives the equation in the main text.

Provided that in every period t households spend all their period t income, w_t, we again get $c_{it}=\gamma_iw_t/p_{it}$. Even if they are endowed with a storage technology, market clearing ensures that the whole output is consumed in period t. For example, with a bond that costs one unit of the numéraire and repays $(1+r_t)$ next period, market clearing implies $r_t=\rho$, where ρ is the households' discount rate.

By market clearing of good j,

$$y_{jt}=c_{jt}+\sum_{i=1}^{n}x_t^{ij}=\gamma_j\frac{w_t}{p_{jt}}+\sum_{i=1}^{n}\beta a_{ij}\frac{p_{i,t+1}y_{i,t+1}}{p_{jt}}\ .$$

Multiplying again by p_{jt} and dividing by w_t gives

$$\lambda_{jt}\equiv\frac{y_{jt}p_{jt}}{w_t}=\gamma_j+\sum_{i=1}^{n}\beta a_{ij}\frac{w_t}{w_{t-1}}\lambda_{i,t+1}\ .$$

Stacking this equation for all i and solving forward shows that, compared to before, Domar weights are adjusted by future changes in the value of the numéraire:

$$\lambda_t=\sum_{h=0}^{\infty}\beta^h\frac{w_{t+h}}{w_t}(A')^h\gamma\ .$$

For output, we obtain

$$ln(y_t) = k_t^{y1} + Aln(y_{t-1}) - \varepsilon_t \ ,$$

where $k_t^{y1} = ln(\lambda_t) - Aln(\lambda_{t-1}) - k_t^{p1}$.

In the steady state with $\varepsilon_t = 0 \ \forall \ t$ we get

$$\lambda = (I - \beta A')^{-1}\gamma \ , \qquad ln\,(p/w) = (I - A)^{-1}k^{p1} \ ,$$

where k^{p1} contains elements $k_i^{p1} = -\left[b_i ln(b_i) + \sum_{j=1}^{n} a_{ij} ln(\beta a_{ij})\right]$. For output we have, as before, $ln\,(y_i) = ln\,(\lambda_i) - ln\,(p_i/w)$. Taking into account the slightly altered meaning of A in this economy, the steady state value for λ is unaltered compared to the above economy with contemporaneous input-output conversion. Specifically, while in the latter $a_{ij} = (p_j x^{ij})/(p_i y_i)$, here $a_{ij} = \beta^{-1}(p_j x^{ij})/(p_i y_i)$. The steady state value for $ln(p/w)$ nevertheless changes sligthly. The difference vanishes as $\beta \longrightarrow 1$.

Multiple-Lags Input-Output Conversion

I start with the general CES case. Firm i's problem is then

$$\max_{\substack{\{l_{it}, \{x_t^{ij}, x_{t,t-1}^{ij},\\ x_{t,t-2}^{ij}\}_{j=1}^{n}\}_{t=0}^{\infty}}} \sum_{t=0}^{\infty} \beta^t \left[p_{it} z_{it} l_{it}^{b_i} \prod_{j=1}^{n} \left[\eta_1 \left(x_{t,t-1}^{ij}\right)^r + \eta_2 \left(x_{t,t-2}^{ij}\right)^r \right]^{\frac{a_{ij}}{r}} - w_t l_{it} - \sum_{j=1}^{n} p_{jt} x_t^{ij} \right]$$

$$\text{s.t.} \quad x_t^{ij} = x_{t,t}^{ij} + x_{t+1,t}^{ij} + x_{t+2,t}^{ij} \ \forall \ t, i, j \ .$$

For each input j, the firm chooses how much to buy in period t, x_t^{ij}, and how to distribute the bought amount for production over periods $t+1, t+2$. Because I abstract from the case of perfect substitutability, I ignore the boundary constraints $l_{it}, x_{t+1,t}^{ij}, x_{t+2,t}^{ij} \geq 0 \ \forall \ t, i, j$.

Let $\breve{x}_{t+h,t}^{ij}$ be the amount of good j purchased at t and not used up in production up to (but not including) period $t+h$. We obtain the following value function:

$$\begin{aligned} V_i\left(\{\breve{x}_{t,t-2}^{ij}\}_j, \{\breve{x}_{t,t-1}^{ij}\}_j\right) = \max_{\substack{l_{it}, \{x_t^{ij},\\ x_{t,t-1}^{ij}, x_{t,t-2}^{ij}\}_j}} & \left[p_{it} z_{it} l_{it}^{b_i} \prod_{j=1}^{n} \left[\eta_1 \left(x_{t,t-1}^{ij}\right)^r + \eta_2 \left(x_{t,t-2}^{ij}\right)^r \right]^{\frac{a_{ij}}{r}} \right. \\ & \left. - w_t l_{it} - \sum_{j=1}^{n} p_{jt} x_t^{ij} \right] + \beta V_i\left(\{\breve{x}_{t+1,t-1}^{ij}\}_j, \{\breve{x}_{t+1,t}^{ij}\}_j\right) \\ \text{s.t.} \quad & \breve{x}_{t+1,t}^{ij} = x_t^{ij} \ , \\ & \breve{x}_{t,t-1}^{ij} = x_{t,t-1}^{ij} + x_{t+1,t-1}^{ij} \ , \\ & \breve{x}_{t,t-2}^{ij} = x_{t,t-2}^{ij} \ . \end{aligned}$$

The problem can be written more compactly as

$$V_i\left(\{x^{ij}_{t,t-2}\}_j, \{\breve{x}^{ij}_{t,t-1}\}_j\right) = \max_{\substack{l_{it},\{x^{ij}_t,\\ x^{ij}_{t+1,t-1}\}_j}} \left[p_{it} z_{it} l_{it}^{b_i} \prod_{j=1}^{n} \left[\eta_1 \left(\breve{x}^{ij}_{t,t-1} - x^{ij}_{t+1,t-1}\right)^r + \eta_2 \left(x^{ij}_{t,t-2}\right)^r\right]^{\frac{a_{ij}}{r}} \right.$$
$$\left. - w l_{it} - \sum_{j=1}^{n} p_{jt} x^{ij}_t \right] + \beta V\left(\{x^{ij}_{t+1,t-1}\}_j, \{x^{ij}_t\}_j\right)$$

This means that in each period t, and for each input j, a firm essentially only chooses how how much to buy for production in $t+1$ and $t+2$ and how much of the leftover amount purchased at $t-1$ to use at t as opposed to leaving it for $t+1$.

Cobb-Douglas Aggregation of Past-Purchased Inputs Under $r \longrightarrow 0$, we have $x_{ijt} = \left(x^{ij}_{t,t-1}\right)^{\eta_1}\left(x^{ij}_{t,t-2}\right)^{\eta_2}$ and the optimality conditions yield

$$l_{it} = b_i \frac{p_{it} y_{it}}{w}\ , \quad x^{ij}_{t,t-1} = \beta \eta_1 a_{ij} \frac{p_{it} y_{it}}{p_{j,t-1}}\ , \quad x^{ij}_{t,t-2} = \beta \eta_2 a_{ij} \frac{p_{it} y_{it}}{p_{j,t-2}}\ .$$

Inserting these expressions into the production function, leads after a little algebra to

$$ln\left(\frac{p_{it}}{w_t}\right) = k^{p2}_t + \sum_{j=1}^{n} a_{ij}\left[\eta_1 ln\left(\frac{p_{j,t-1}}{w_{t-1}}\right) + \eta_2 ln\left(\frac{p_{j,t-2}}{w_{t-2}}\right)\right] + \varepsilon_t\ ,$$

where

$$k^{p2}_{it} = k^{p2}_i - (1-b_i)\left[\eta_1 ln\left(\frac{w_t}{w_{t-1}}\right) + \eta_2 ln\left(\frac{w_t}{w_{t-2}}\right)\right]\ ,$$

and $k^{p2}_i = -b_i ln(b_i) - \sum_{j=1}^{n} a_{ij}\left[\eta_1 ln(\beta a_{ij}) + \eta_2 ln(\beta^2 a_{ij})\right]$. Stacking this equation for all i gives the expression in the main text.

The market clearing condition for good j is now

$$y_{jt} = c_{jt} + \sum_{i=1}^{n} x^{ij}_t = c_{jt} + \sum_{i=1}^{n} x^{ij}_{t+1,t} + x^{ij}_{t+2,t}\ .$$

Plugging in the optimality conditions and multiplying by p_{jt}/w_t to solve for λ_{jt} gives

$$\lambda_{jt} = \gamma_j + \beta\eta_1 \frac{w_t}{w_{t-1}} \sum_{i=1}^{n} a_{ij}\lambda_{i,t+1} + \beta^2\eta_2 \frac{w_t}{w_{t-2}} \sum_{i=1}^{n} a_{ij}\lambda_{i,t+2}\ .$$

When stacked for all i, one could solve forward to obtain λ_t. Its value is independent of TFP levels ε_t. For output we get then

$$ln(y_t) = k_t^{y2} + \eta_1 A ln(y_{t-1}) + \eta_2 A ln(y_{t-2}) - \varepsilon_t \ ,$$

where $k_t^{y2} = ln(\lambda_t) - \eta_1 A ln(\lambda_{t-1}) - \eta_2 A ln(\lambda_{t-2}) - k_t^{p2}$.

In the steady state with $\varepsilon_t = 0 \ \forall \ t$ we get

$$\lambda = (I - (\beta\eta_1 + \beta^2\eta_2)A')^{-1}\gamma \ , \qquad ln\,(p/w) = (I - A)^{-1} k^{p2} \ .$$

For output we have, as before, $ln\,(y_i) = ln\,(\lambda_i) - ln\,(p_i/w)$. Taking into account the slightly altered meaning of A in this economy, the steady state value for λ is again unaltered compared to the above two economies. In this economy, we have

$$a_{ij} = \left[\beta\eta_1 + \beta^2\eta_2\right]^{-1} (p_j x^{ij})/(p_i y_i)$$

in steady state. The steady state value for $ln(p/w)$ nevertheless changes sligthly. Again the difference vanishes as $\beta \longrightarrow 1$.

General CES-Aggregation of Past-Purchased Inputs For general r, the optimality conditions yield

$$l_{it} = b_i \frac{y_{it} p_{it}}{w_t} \ , \quad x_{t,t-1}^{ij} = \left[a_{ij}\eta_1\beta \frac{y_{it}p_{it}/x_{ijt}}{p_{jt-1}}\right]^{\frac{1}{1-r}} \ , \quad x_{t,t-2}^{ij} = \left[a_{ij}\eta_2\beta^2 \frac{y_{it}p_{it}/x_{ijt}}{p_{jt-2}}\right]^{\frac{1}{1-r}} \ ,$$

Inserting the resulting expressions into the equation for x_{ijt} gives

$$x_{ijt} = (p_{it}y_{it})^{\frac{1}{2-r}} \Lambda_{ijt}^{\frac{1-r}{2-r}} \ , \quad \Lambda_{ijt} = \left[\eta_1(\eta_1 a_{ij}\beta)^{\frac{r}{1-r}} (p_{j,t-1})^{-\frac{r}{1-r}} + \eta_2(\eta_2 a_{ij}\beta^2)^{\frac{r}{1-r}} (p_{j,t-2})^{-\frac{r}{1-r}}\right]^{1/r} \ .$$

In turn, inserting this equation for x_{ijt} into the production function and linearizing around a steady state yields

$$\hat{p}_{it} = \hat{k}_{it}^{p3} + \sum_{j=1}^{n} \frac{1}{\phi_i}\frac{a_{ij}}{2-r}\left[\chi_1\hat{p}_{j,t-1} + \chi_2\hat{p}_{j,t-2}\right] + \frac{1}{\phi_i}\hat{\epsilon}_{it} \ ,$$

where

$$\hat{k}_{it}^{p3} = \frac{1-\phi_i}{\phi_i}\hat{y}_{it} - (1 - b_i/\phi_i)\left[\chi_1(\hat{w}_t - \hat{w}_{t-1}) + \chi_2(\hat{w}_t - \hat{w}_{t-2})\right] \ ,$$

and

$$\chi_1 = \frac{(\eta_1\beta^r)^{\frac{1}{1-r}}}{(\eta_1\beta^r)^{\frac{1}{1-r}} + (\eta_2\beta^2 r)^{\frac{1}{1-r}}} \ , \quad \chi_2 = 1 - \chi_1 \ .$$

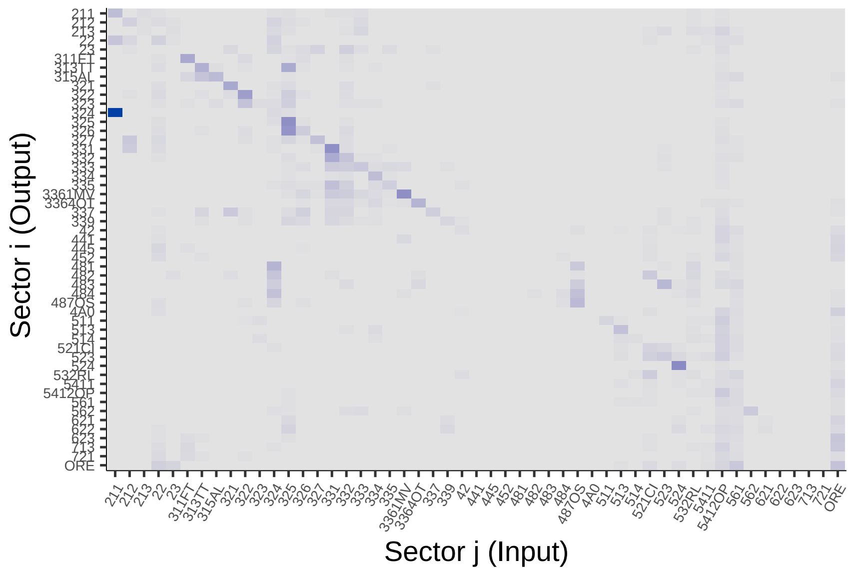


Figure A-1: Tabular Representation of the Input-Output Matrix

Notes: The figure shows the input-output matrix A, with darker shades of blue indicating stronger links a_{ij}.

C.2 Data

The sectors in the PPI and input-output data were matched as follows. Excluding governmental and farming sectors, the BEA input-output data contains 64 sectors. For each of these, I find the corresponding PPI sector. For 13 BEA-sectors, no PPI data is available. Out of the remaining 51, 39 can be matched perfectly, although sometimes the BEA data uses other codes than those of the NAICS classification, which are used in the PPI data. For 12 BEA-sectors, PPI data for only a subset of subsectors which make up these sectors is available. In case data for only one subsector is available, I take this series as an approximation of the sectoral PPI. If multiple subsectors are available, I take an output-weighted average of these subsectors to construct the sectoral PPI. In some cases, some subsectors are excluded because there is no output data available or because the PPI series for this subsector starts late in the sample. Data on sectoral outputs at the fine level of 405 sectors is obtained from the BEA's detailed input-output table for 2010 (also available in 2007). Many of the relevant sectoral and subsectoral PPI series start in December 2003, so that no earlier starting date is possible. I move the starting date of the sample a bit further to January 2005 because this adds two more sectors to the analysis.

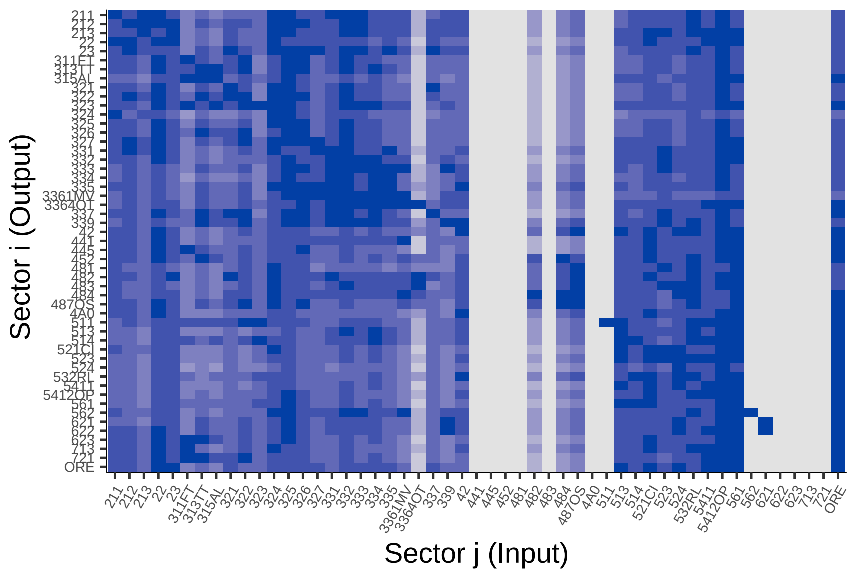
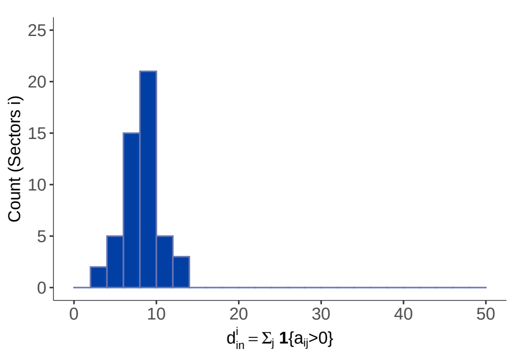


Figure A-2: Tabular Representation of Shortest Paths

Notes: The figure shows the matrix of shortest paths or distances from a sector i to any other sector j, with darker shades of blue indicating longer distances.

(a) In-Degrees



(b) Out-Degrees

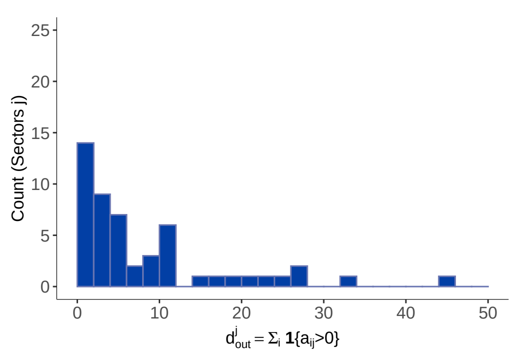


Figure A-3: In-Degrees and Out-Degrees

Notes: The left panel plots in-degrees, equal to the number of non-zero entries by columns of A, which show the number of input-suppliers across sectors. The right panel plots out-degrees, equal to the number of non-zero entries by rows of A, which show the number of customers supplied across sectors.

Table A-1: Output Shares, Inflation Statistics and Network Statistics Across Sectors

Code	Name	Output [%]	Mean π_{it}	StdD π_{it}	Min π_{it}	Max π_{it}	d_i^{in}	wd_i^{in}	d_i^{out}	wd_i^{out}
211	Oil and gas extraction	1.66	0.72	9.61	−32.68	34.30	10	0.30	3	0.97
212	Mining, except oil and gas	0.62	0.48	1.65	−5.11	9.81	11	0.30	6	0.34
213	Support activities for mining	0.46	0.23	1.29	−4.07	7.04	12	0.32	3	0.06
22	Utilities	2.61	0.62	8.19	−47.50	103.11	9	0.39	25	0.69
23	Construction	5.70	0.37	0.77	−1.61	5.84	12	0.41	5	0.13
321	Wood products	0.40	0.33	2.06	−10.02	10.04	7	0.37	5	0.42
327	Nonmetallic mineral products	0.52	0.29	0.42	−0.46	2.61	10	0.44	5	0.24
331	Primary metals	1.32	0.36	2.12	−7.15	9.33	10	0.56	11	1.07
332	Fabricated metal products	1.66	0.30	0.50	−0.85	2.61	9	0.47	24	0.92
333	Machinery	1.79	0.23	0.29	−0.17	2.05	11	0.47	11	0.37
334	Computer and electronic products	2.01	−0.01	0.25	−0.93	1.52	6	0.22	12	0.40
335	Electrical equipment, appliances, and components	0.61	0.28	0.49	−1.62	1.94	10	0.43	7	0.20
3361MV	Motor vehicles, bodies and trailers, and parts	2.44	0.11	0.61	−2.22	4.01	9	0.68	6	0.46
3364OT	Other transportation equipment	1.40	0.19	0.23	−0.57	1.07	12	0.43	3	0.25
337	Furniture and related products	0.32	0.25	0.35	−0.45	2.14	13	0.53	3	0.11
339	Miscellaneous manufacturing	0.90	0.15	0.27	−0.50	2.11	13	0.35	4	0.12
311FT	Food and beverage and tobacco products	4.34	0.28	0.89	−4.65	6.07	7	0.35	7	0.41
313TT	Textile mills and textile product mills	0.28	0.23	0.44	−0.76	1.70	9	0.53	9	0.46
315AL	Apparel and leather and allied products	0.13	0.12	0.25	−0.53	1.12	7	0.42	3	0.20
322	Paper products	0.95	0.26	0.52	−1.12	2.22	10	0.52	11	0.56
323	Printing and related support activities	0.47	0.19	0.44	−0.54	3.31	13	0.44	3	0.08
324	Petroleum and coal products	3.41	0.78	7.42	−33.28	22.76	3	0.76	21	1.00
325	Chemical products	4.01	0.33	0.70	−2.05	2.63	6	0.43	27	1.56
326	Plastics and rubber products	1.05	0.29	0.57	−1.14	2.54	7	0.52	16	0.46
42	Wholesale trade	6.73	0.32	0.88	−1.90	3.88	10	0.22	5	0.09
441	Motor vehicle and parts dealers	0.94	0.33	0.93	−2.72	5.20	6	0.19	0	0.00
445	Food and beverage stores	1.01	0.31	1.39	−2.99	4.06	7	0.20	0	0.00
452	General merchandise stores	1.01	0.23	3.29	−11.76	10.90	8	0.20	0	0.00
4A0	Other retail	3.43	0.25	1.61	−5.99	5.11	6	0.21	0	0.00
481	Air transportation	0.85	0.29	3.08	−9.89	13.25	5	0.36	0	0.00
482	Rail transportation	0.38	0.32	0.83	−2.72	2.77	10	0.40	1	0.01
483	Water transportation	0.27	0.32	1.42	−4.37	8.96	9	0.54	0	0.00
484	Truck transportation	1.41	0.30	0.80	−2.25	4.04	9	0.41	3	0.06
487OS	Other transportation and support activities	0.93	0.28	0.69	−1.40	4.31	9	0.33	6	0.51
511	Publishing industries, except internet (includes software)	1.62	0.08	0.47	−1.90	1.76	9	0.24	1	0.05
513	Broadcasting and telecommunications	3.70	0.05	0.45	−1.24	1.59	7	0.29	9	0.25
514	Data processing, internet publishing, and other information services	0.91	0.02	0.52	−2.23	2.07	9	0.24	3	0.05
521CI	Federal Reserve banks, credit intermediation, and related activities	3.88	0.07	2.59	−9.21	8.19	8	0.28	17	0.53
523	Securities, commodity contracts, and investments	2.54	0.35	1.63	−10.59	6.20	9	0.40	10	0.43
524	Insurance carriers and related activities	3.61	0.17	0.25	−0.14	1.71	3	0.39	11	0.55
ORE	Other real estate	4.51	0.15	0.40	−1.19	1.96	9	0.50	28	1.10
532RL	Rental and leasing services and lessors of intangible assets	1.46	0.18	1.82	−6.50	5.86	8	0.28	20	0.37
5411	Legal services	1.56	0.30	0.59	−0.72	3.27	7	0.17	12	0.17
5412OP	Miscellaneous professional, scientific, and technical services	5.74	0.16	0.51	−0.81	6.28	7	0.21	46	1.68
561	Administrative and support services	3.29	0.14	0.27	−0.69	1.17	7	0.15	34	0.94
562	Waste management and remediation services	0.46	0.24	0.53	−1.77	2.12	10	0.28	1	0.10
621	Ambulatory health care services	4.50	0.09	0.30	−1.10	2.80	7	0.21	2	0.03
622	Hospitals	3.57	0.21	0.34	−0.75	1.96	9	0.30	0	0.00
623	Nursing and residential care facilities	1.09	0.24	0.31	−0.83	1.52	8	0.25	0	0.00
713	Amusements, gambling, and recreation industries	0.60	0.11	0.65	−2.00	2.20	9	0.30	0	0.00
721	Accommodation	0.95	0.22	2.04	−8.85	9.23	8	0.21	0	0.00
	Mean	1.96	0.26	1.35	−5.03	7.14	8.61	0.36	8.61	0.36
	StdD	1.65	0.15	1.94	8.99	14.85	2.25	0.13	10.03	0.41
	Min	0.13	−0.01	0.23	−47.50	1.07	3.00	0.15	0.00	0.00
	Max	6.73	0.78	9.61	−0.14	103.11	13.00	0.76	46.00	1.68

C.3 Estimation

For reasons outlined in the main text, I obtain the MLE in Section 4 as the posterior mode under a prior for θ which is proportional to a constant. For α, the domain restrictions $\alpha_l \in [0,1]$ and $\sum_{l=1}^{p-1} \alpha_l \leq 1$ lend themselves into a prior distribution which is the product of independent uniform distributions, truncated to the region where $\sum_{l=1}^{p-1} \alpha_l \leq 1$:

$$\alpha_1, ..., \alpha_{p-1} \sim \prod_{l=1}^{p-1} \mathcal{U}(0,1) \, \mathbf{1}\left\{ \sum_{l=1}^{p-1} \alpha_l \leq 1 \right\} .$$

As one can verify, this leads to

$$\begin{aligned} p(\alpha_1, ..., \alpha_{p-1}) &= p(\alpha_1|\alpha_2, ..., \alpha_{p-1}) p(\alpha_2|\alpha_3, ..., \alpha_{p-1}) ... p(\alpha_{p-2}|\alpha_{p-1}) p(\alpha_{p-1}) \\ &= \begin{cases} (p-1)! & \text{if } \sum_{l=1}^{p-1} \alpha_l \leq 1 \\ 0 & \text{otherwise} \end{cases} , \end{aligned}$$

where for $l = 1 : p-2$,

$$p(\alpha_l|\alpha_{l+1}, ..., \alpha_{p-1}) = \begin{cases} l \frac{\left(1 - \sum_{m=l}^{p-1} \alpha_m\right)^{l-1}}{\left(1 - \sum_{m=l+1}^{p-1} \alpha_m\right)^{l-1}} & \text{if } \alpha_l \in \left[0, 1 - \sum_{m=l+1}^{p-1} \alpha_m\right] \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$p(\alpha_{p-1}) = \begin{cases} (p-1)\,(1-\alpha_{p-1})^{p-2} & \text{if } \alpha_{p-1} \in [0,1] \\ 0 & \text{otherwise} \end{cases} .$$

To draw from $p(\alpha_1, ..., \alpha_{p-1})$, one can draw α_{p-1} from its marginal distribution and iteratively draw $\alpha_{p-2}, ..., \alpha_1$ from the conditionals. In each of these steps, efficient drawing from these non-standard distributions is implemented using the inverse-cdf method; to draw $y_i \sim f(y)$, it draws $x_i \sim \mathcal{U}(0,1)$ and finds y_i so that $\int_{-\infty}^{y_i} f(y)dy = x_i$. In the present case, this yields

$$\alpha_l|(\alpha_{l+1}, ..., \alpha_{p-1}) = \left(1 - \sum_{m=l+1}^{p-1} \alpha_m\right) \left[1 - (1-x_l)^{1/l}\right] , \quad x_l \sim \mathcal{U}(0,1) , \quad l = 1 : p-2$$

and $\alpha_{p-1} = 1 - (1 - x_{p-1})^{1/(p-1)}$, $x_{p-1} \sim \mathcal{U}(0,1)$.

For the parameters σ_i, only the lower bound of the prior distribution is determined by the domain restriction $\sigma_i > 0$. The choice of the upper bound may appear non-trivial. If it is too low, it might restrict the posterior mode, if it is too large, posterior sampling will be inefficient. However, one can choose a lower bound low enough to ensure efficient computation and still allow the SMC sampler to go beyond the upper bound in the search for σ_i associated with high likelihood values in case it is needed by not enforcing the upper bound as a domain restriction. In case the posterior contains draws that do not obey the boundary of the prior, the resulting marginal data density (MDD) will be wrong, but can be adjusted easily ex-post. Let $\bar{s}_i$ be the upper bound in the prior draws for σ_i, and take any $\bar{d}_i$. We have

$$\begin{aligned} p(Y) &= \int p(Y|\theta)p(\theta)d\theta \\ &= \int p(Y|\alpha,\sigma)p(\alpha)p(\sigma)d(\alpha,\sigma) \\ &= (p-1)! \prod_{i=1}^{n} \frac{1}{\bar{s}_i} \int p(Y|\alpha,\sigma)d(\alpha,\sigma) \ , \end{aligned}$$

so that the (estimated) $ln\ p(Y)$ returned by the SMC sampler,

$$ln\ p(Y) = ln\ (p-1)! - \sum_{i=1}^{n} ln\ \bar{s}_i + ln \int p(Y|\alpha,\sigma)d(\alpha,\sigma) \ ,$$

can be adjusted to reflect an effective uniform prior for σ_i with an upper bound different than the one used to initialize the sampler. For example, to go from $\mathcal{U}(0,\bar{s}_i)$ to $\mathcal{U}(0,\bar{d}_i)$, one adds $ln\ \bar{s}_i - ln\ \bar{d}_i$. To be able to use MDD as a model selection device, the prior for σ and hence the effective prior upper bounds $\{\bar{d}_i\}_{i=1:n}$ should be the same for all different models indexed by (q,p). I choose $\bar{s}_i = 5\mathbb{V}[y_{it}]$ and abstract from the re-scaling of the MDD because its exact value is not of importance in the present analysis, only the relative values for different models.

In absence of more precise prior information (and in particular due to the wide priors for σ_i), the SMC algorithm would take a long time to converge if the proposal distribution is taken to be the prior (likelihood tempering). To ameliorate this issue, I use the model tempering variant of the SMC from Mlikota and Schorfheide (2022) and implement a mock model to construct a proposal distribution that tilts the prior draws for σ towards values that are more compatible with high likelihood values conditional on the prior draws for α. To do so, I obtain a consistent estimator for $\sigma|\alpha$ using the method of moments applied to the variance of the high-frequency process x_τ.[65] We know that $\mathbb{V}[y_t] = \mathbb{V}[x_{tq}]$. The latter can be computed as a function of Σ:

$$\mathbb{V}[x_\tau] = \Phi_1 \mathbb{V}[x_\tau]\Phi_1' + ... + \Phi_p \mathbb{V}[x_\tau]\Phi_p' + \Sigma \ ,$$

[65] As opposed to that, using the process for y_t, the conditional MLE for $\sigma|\alpha$ would only be available using data augmentation due to the presence of MA errors whose dimension surpasses that of y_t.

which in turn implies

$$vec(\mathbb{V}[x_\tau]) = \left(\sum_{l=1}^{p} \Phi_l \otimes \Phi_l\right) vec(\mathbb{V}[x_\tau]) + vec(\Sigma)$$
$$= \left(\sum_{l=1}^{p} \alpha_l^2\right) (A \otimes A)\, vec(\mathbb{V}[x_\tau]) + vec(\Sigma) \ .$$

Overall, we get

$$vec(\hat{\Sigma}_{MM}|\alpha) = \left[I - \left(\sum_{l=1}^{p} \alpha_l^2\right)(A \otimes A)\right] vec(\hat{\mathbb{V}}_{MM}[x_\tau]) \ , \quad \hat{\mathbb{V}}_{MM}[x_\tau] = \hat{\mathbb{V}}_{MM}[y_t] = \frac{1}{T}\sum_{t=1}^{T} y_t y_t' \ .$$

I then construct the likelihood for the mock model as the density of independent Inverse Gamma distributions for σ_i with a mode at $\hat{\sigma}_{i,MM}|\alpha = (\hat{\Sigma}_{MM}|\alpha)_{ii}$. This means that the proposal distribution is the product of the prior for α and Inverse Gamma distributions for $\sigma_i|\alpha$.

I also use the adaptive tempering method proposed by Cai et al. (2021), which ensures a precise estimation of the posterior in the present case in which the distance between the proposal and posterior distributions is difficult to assess. Finally, to implement the algorithm under the presence of the tight domain restrictions for α, I consider a transformation of the parameters in the mutation step of the SMC algorithm. Define the function g s.t. $\check{\theta} = g^{-1}(\theta)$ is generated by taking logs of σ_i and computing $\gamma_l = ln\ \alpha_l/\alpha_p$ for $\alpha_1, ..., \alpha_{p-1}$. Note that both are one-to-one mappings and ensure that the transformed parameters can fall everywhere on the real line. As a result, no draws in the mutation step are rejected because of domain violations. I use a Random Walk Metropolis Hastings (RWMH) algorithm in the mutation step. Even though the proposal density for the transformed draws is symmetric, for the original parameters it is not. The mutation step needs to be adjusted to reflect this. Overall, the mutation of particle i in iteration n of the SMC algorithm is performed as follows:

Algorithm 1 (Particle Mutation in SMC Algorithm)**.**

1. *Given particle* θ_{n-1}^i*, set* $\theta_n^{i,0} = \theta_{n-1}^i$.
2. *For* $m = 1 : N_{MH}$*:*
   - *Compute* $\check{\theta}_n^{i,m-1} = g^{-1}(\theta_n^{i,m-1})$ *and draw*

$$\check{v}|\theta_n^{i,m-1} \sim \check{q}(\check{v}|\theta_n^{i,m-1}) = N(\check{\theta}_n^{i,m-1}, c_n^2\Sigma_n) = N(g^{-1}(\theta_n^{i,m-1}), c_n^2\Sigma_n) \ .$$

- *Set*

$$\theta_n^{i,m} = \begin{cases} v = g(\check{v}) & w.p. \quad \alpha(v|\theta_n^{i,m-1}) \\ \theta_n^{i,m-1} & otherwise \end{cases} ,$$

where

$$\alpha(v|\theta_n^{i,m-1}) = min \left\{ 1, \frac{p(Y|v)p(v)/q(v|\theta_n^{i,m-1})}{p(Y|\theta_n^{i,m-1})p(\theta_n^{i,m-1})/q(\theta_n^{i,m-1}|v)} \right\} .$$

The densities $q(v|\theta_n^{i,m-1})$ *and* $q(\theta_n^{i,m-1}|v)$ *are obtained using analogous density transformations starting from* $q(\check{v}|\theta_n^{i,m-1})$ *and* $q(\check{\theta}_n^{i,m-1}|v)$*, respectively;*

$$q(v|\theta_n^{i,m-1}) = \check{q}(g^{-1}(v)|\theta_n^{i,m-1})|J(v)| ,$$

where the Jacobian matrix $J(\theta)$ *is block diagonal with*

$$J_{11}(\theta) = \begin{bmatrix} \alpha_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \alpha_{p-1}^{-1} \end{bmatrix} + \alpha_p^{-1}\iota\iota' , \quad J_{22}(\theta) = \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \sigma_N^{-1} \end{bmatrix} .$$[66]

3. Set $\theta_n^i = \theta_n^{i,N_{MH}}$.

Note that because $\check{q}(g^{-1}(v)|\theta_n^{i,m-1}) = \check{q}(g^{-1}(\theta_n^{i,m-1})|v)$ is symmetric, we obtain

$$\alpha(v|\theta_n^{i,m-1}) = min \left\{ 1, \frac{p(Y|v)p(v)}{p(Y|\theta_n^{i,m-1})p(\theta_n^{i,m-1})} \frac{|J(\theta_n^{i,m-1})|}{|J(v)|} \right\} ,$$

and one can show that

$$ln \frac{|J(\theta_n^{i,m-1})|}{|J(v)|} = \sum_{j=1}^{N} \left[ln\ \sigma_j - ln\ \sigma_{j,n}^{i,m-1} \right] + ln\ |J_{11}(\theta_n^{i,m-1})| - ln\ |J_{11}(v)| .$$

C.4 Results

D Forecasting Global Industrial Production Growth

[66] ι denotes a column vector of ones, and I write a capital N for the cross-sectional sample size to distinguish it from the iteration of the SMC algorithm, n.

Table A-2: Model Selection

Log MDD

		p					
		$1q$	$2q$	$3q$	$4q$	$5q$	$6q$
q	1/3			19079			19044
	1/2		19384		18768		18690
	1	20153	20056	19675	19879	18899	20218
	2	17546	19570	19248	20142	18662	19636
	4	18517	19808	19754	19655	18904	19301

BIC

		p					
		$1q$	$2q$	$3q$	$4q$	$5q$	$6q$
q	1/3			-38568			-38487
	1/2		-38849		-38194		-38089
	1	-40964	-40799	-39650	-39944	-38060	-41060
	2	-36597	-39556	-39073	-40832	-37206	-39420
	4	-38096	-39883	-39648	-39274	-37806	-38659

AIC

		p					
		$1q$	$2q$	$3q$	$4q$	$5q$	$6q$
q	1/3			-38939			-38865
	1/2		-39220		-38572		-38474
	1	-41335	-41177	-40036	-40336	-38460	-41467
	2	-36975	-39948	-39480	-41253	-37640	-39868
	4	-38488	-40304	-40098	-39753	-38313	-39194

Notes: The values for q (from top to bottom) refer to quarterly, bi-monthly, monthly, bi-weekly and weekly network interactions, respectively, while $p = mq$ implies that the last m months matter for dynamics.

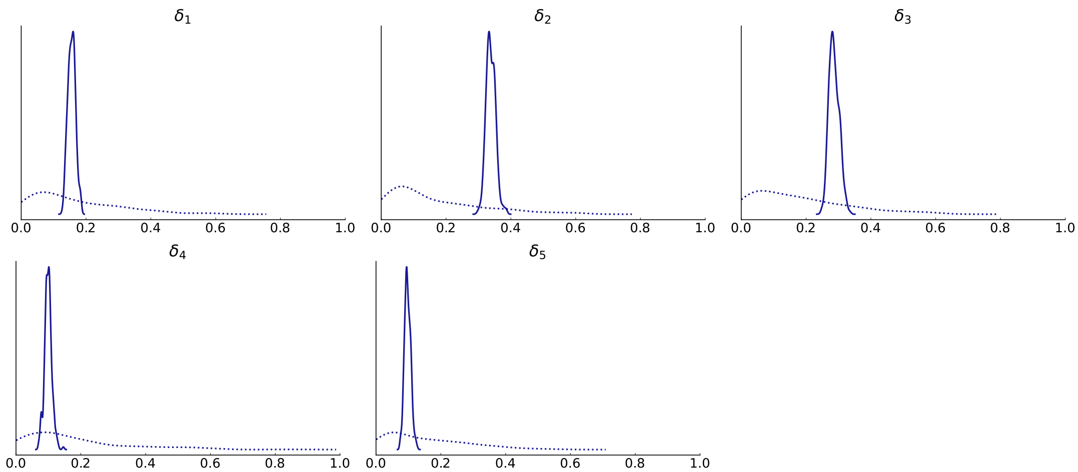


Figure A-4: Marginal Priors and Posteriors For α

Notes: The solid line shows the posterior, the dotted line the prior.

D.1 NVAR$(p, 1)$ & Factor Model: Equivalence Result

For expositional simplicity, the equivalence result is shown for an NVAR$(2, 1)$. The extension to general p is straightforward. The NVAR$(2, 1)$ can be written as

$$y_t = A[\alpha_1 y_{t-1} + \alpha_2 y_{t-2}] + u_t \ .$$

Let r denote the rank of A. We can find $n \times r$ and $r \times n$ matrices B and C, both of full rank, such that $A = BC$. In turn, the NVAR$(2, 1)$ can be represented as a factor model with r factors:

$$y_t = BC[\alpha_1 y_{t-1} + \alpha_2 y_{t-2}] + u_t = \Lambda f_t + u_t \ .$$

The $n \times r$ matrix of loadings Λ is given by B, while factor k is given by $f_{kt} = \alpha_1 C_{k\cdot} y_{t-1} + \alpha_2 C_{k\cdot} y_{t-2}$, where $C_{k\cdot}$ denotes the kth row of C. Note that this factor representation is not unique, as an observationally equivalent process is obtained by writing $A = BC = BQQ^{-1}C = \tilde{B}\tilde{C}$ for any $r \times r$ full-rank matrix Q.

Conversely, let y_t permit a factor structure, with r factors evolving dynamically according to a VAR(2):

$$y_t = \Lambda f_t + \xi_t \ , \quad f_t = \Phi_1 f_{t-1} + \Phi_2 f_{t-2} + \eta_t \ .$$

Using an argument similar to the one in Cesa-Bianchi and Ferrero (2021), take r distinct vectors of weights $w^k = (w_1^k, ..., w_n^k)$, $k = 1 : r$, and consider weighted averages of $\{y_{it}\}_{i=1}^n$ of the form

$$\sum_{i=1}^{n} w_i^k y_{it} = \sum_{i=1}^{n} w_i^k \Lambda_{i\cdot} f_t + \sum_{i=1}^{n} w_i^k \xi_{it} \ .$$

For n large enough, $\bar{\xi}_t^k \equiv \sum_{i=1}^{n} w_i^k \xi_{it} \sim O_p(n^{-1/2})$ is negligible and we can write

$$W y_t = W \Lambda f_t \ ,$$

where the $r \times n$ matrix W stacks $w^{k\prime}$ along rows. In turn, we can solve for $f_t = (W\Lambda)^{-1} W y_t$. As this equation holds for all t, we can re-write the process for y_t as

$$\begin{aligned} y_t &= \Lambda \left(\Phi_1 f_{t-1} + \Phi_2 f_{t-2} + \eta_t \right) + \xi_t \\ &= \Lambda \Phi_1 (W\Lambda)^{-1} W y_{t-1} + \Lambda \Phi_2 (W\Lambda)^{-1} W y_{t-2} + u_t \ , \end{aligned}$$

with $u_t = \Lambda \eta_t + \xi_t$. If the dynamic evolution of the r factors is restricted to an NVAR$(2,1)$, then $\Phi_1 = \phi_1 \Phi$ and $\Phi_2 = \phi_2 \Phi$ for some ϕ_1, ϕ_2, Φ, and the above equation simplifies to

$$y_t = \Lambda \Phi (W\Lambda)^{-1} W [\phi_1 y_{t-1} + \phi_2 y_{t-2}] + u_t \ .$$

This equation implies that y_t follows an NVAR$(2,1)$ with adjacency matrix $A = \Lambda \Phi (W\Lambda)^{-1} W$. A has rank r and can be written as $A = BC$ with $B = \Lambda Q$, $C = Q^{-1} \Phi (W\Lambda)^{-1} W$ for any $r \times r$ orthogonal matrix Q. Note that we can re-scale A and (ϕ_1, ϕ_2) in case an element in A exceeds unity (see Section 3.2).

D.2 Application Details

Table A-3: Descriptive Data Statistics and Estimated Outdegrees

Code	Name	Mean y_{it}	StdD y_{it}	Min y_{it}	Max y_{it}	$\hat{wd}_i^{out}$
AUT	Austria	−0.14	7.11	−25.92	32.05	0.00
BGD	Bangladesh	2.09	8.51	−33.49	65.66	0.90
BEL	Belgium	0.62	7.53	−22.55	36.17	0.00
BRA	Brazil	−0.74	7.09	−29.14	33.29	1.07
CAN	Canada	0.07	5.58	−20.86	15.40	1.69
CHL	Chile	0.12	4.40	−13.09	12.39	0.72
COL	Colombia	0.45	7.88	−36.73	58.73	2.94
CRI	Costa Rica	−0.60	3.93	−14.67	12.74	0.19
CYP	Cyprus	−0.35	8.40	−37.58	58.53	0.95
CZE	Czech Republic	0.58	8.37	−37.33	51.70	0.04
DEU	Germany	−0.21	7.84	−31.27	35.07	0.00
DNK	Denmark	0.19	6.77	−21.50	22.43	0.09
EST	Estonia	−0.09	10.79	−40.17	31.80	0.57
ESP	Spain	−1.24	7.85	−35.60	51.00	0.00
FIN	Finland	−1.30	6.71	−27.86	21.42	0.12
FRA	France	−0.71	6.50	−35.64	43.58	0.00
GBR	United Kingdom	−0.78	5.81	−26.57	28.56	1.02
GRC	Greece	−0.18	8.52	−25.33	21.84	0.52
HRV	Croatia	−0.17	5.42	−16.18	15.68	0.24
HUN	Hungary	1.11	9.20	−39.83	56.31	0.06
IRL	Ireland	−1.52	16.10	−38.83	59.64	1.41
ISR	Israel	0.08	8.12	−22.43	31.42	0.15
IND	India	−2.13	11.35	−58.94	127.82	0.16
ISL	Iceland	0.29	13.65	−25.78	50.08	1.11
ITA	Italy	−0.56	9.52	−29.31	78.25	0.00
JOR	Jordan	−0.14	7.86	−25.26	26.30	1.90
JPN	Japan	−1.28	8.40	−37.18	26.70	1.16
KOR	Korea, Republic of	−3.03	7.10	−32.63	31.36	2.77
LTU	Lithuania	0.48	9.28	−31.29	33.41	0.57
LUX	Luxembourg	−1.01	8.64	−34.17	42.01	0.47
LVA	Latvia	−0.10	7.95	−31.12	20.58	0.70
MKD	North Macedonia, Republic of	1.11	9.34	−22.59	32.46	0.46
MEX	Mexico	−1.01	5.63	−31.73	34.86	0.19
MYS	Malaysia	−1.54	5.45	−22.03	13.90	1.44
NLD	Netherlands	−0.36	5.78	−21.61	20.80	0.63
NOR	Norway	−2.38	5.32	−15.40	10.80	0.61
POL	Poland	1.24	6.59	−29.50	40.54	2.04
PRT	Portugal	−1.99	6.04	−30.68	36.22	2.04
RUS	Russian Federation	1.74	4.95	−18.98	11.01	3.46
SWE	Sweden	−1.00	6.49	−24.47	20.95	2.62
SVN	Slovenia	0.10	7.06	−27.74	32.12	2.90
SVK	Slovakia	1.97	10.31	−45.26	65.41	1.30
TUR	Turkey	0.58	9.19	−35.66	59.56	2.24
USA	United States	−1.24	4.65	−18.84	14.73	3.94

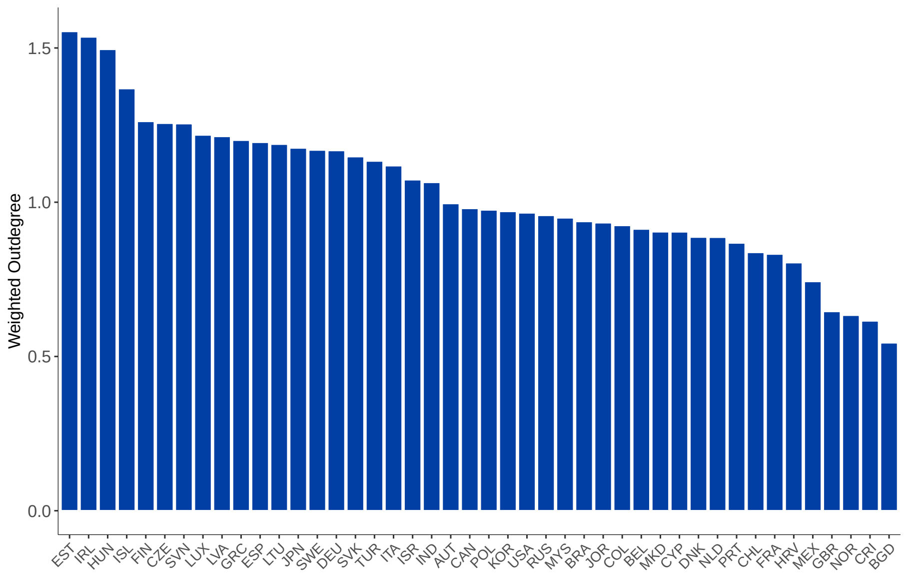


Figure A-5: Weighted Indegrees in the Estimated Network

Notes: The plot depicts the weighted indegrees in the estimated network as relevant for monthly industrial production dynamics across countries.

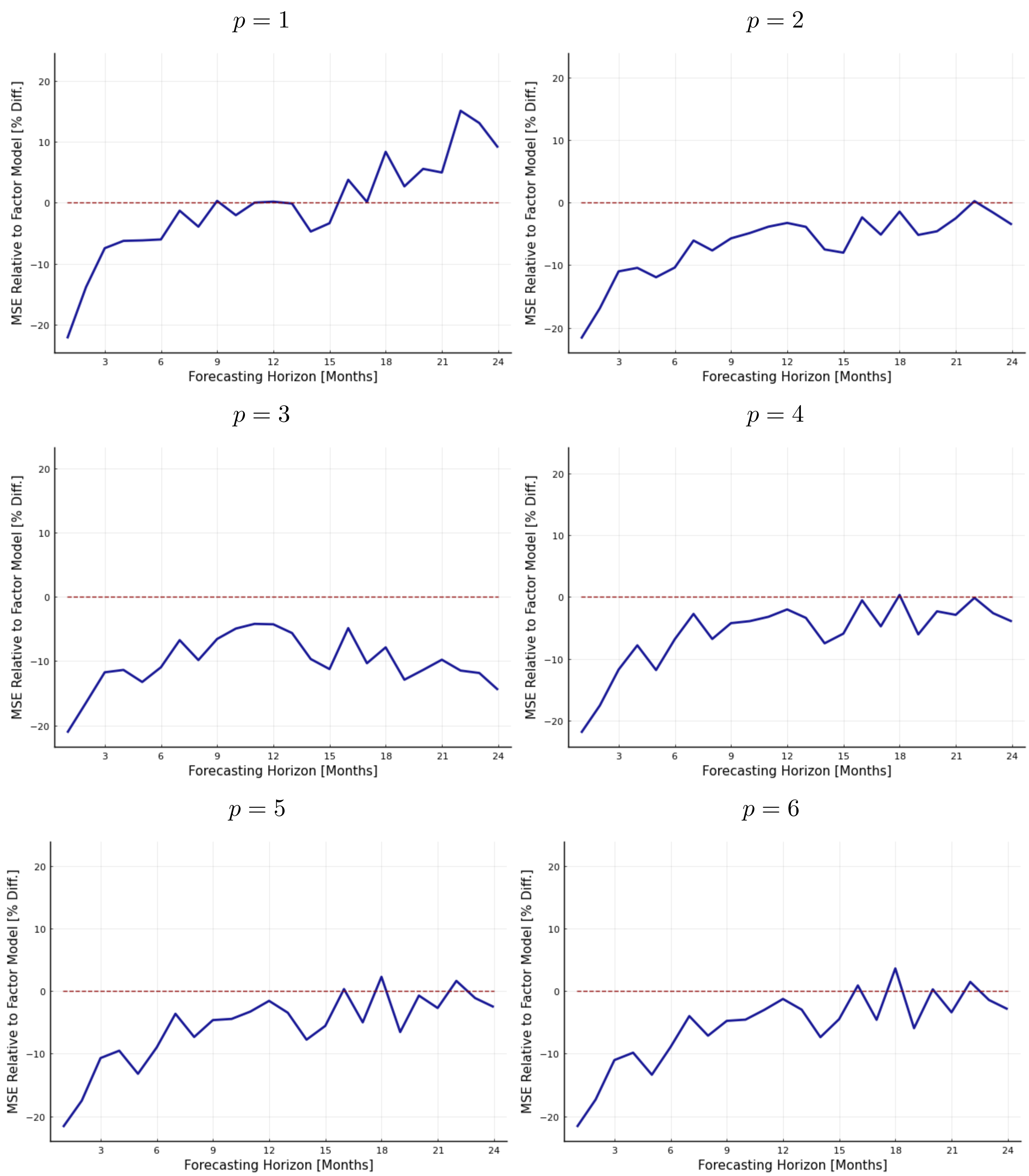


Figure A-6: Out-of-Sample Forecasting Performance: NVAR$(p, 1)$ vs. Factor Model

Notes: The plot depicts the percentage difference between the out-of-sample Mean Squared Errors generated by the NVAR$(p, 1)$ to those generated by the Principal Components Factor Model for different choices of p.