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

Marko Mlikota

University of Pennsylvania

Job Market Paper

This Version: November 28, 2022

[latest version available here]

Abstract

Many environments in economics feature a cross-section of agents or units linked by a network of bilateral ties. I develop a framework to study dynamics in these cases. It consists of a vector autoregression in which innovations transmit cross-sectionally via bilateral links and which can accommodate general patterns of how network effects of higher order accumulate over time. In a first application, I take the supply chain network of the US economy as given and document how it drives the dynamics of sectoral prices. By estimating the time profile of network effects, the model allows me to go beyond steady state comparisons and study transition dynamics induced by granular shocks. As a result of different positions in the input-output network, sectors differ in both the strength and the timing of their impact on aggregates. In a second application, I discuss how to approximate cross-sectional processes by assuming that dynamics are driven by a network and in turn estimating the latter. The proposed framework offers a sparse, yet flexible and interpretable method for doing so, owing to networks' ability to summarize complex relations among units by relatively few non-zero bilateral links. Modeling industrial production growth across 44 countries, I obtain reductions in out-of-sample mean squared errors of up to 20% relative to a principal components factor model.

JEL codes: C32, C38, D57, E37.

Key words: Granular Shocks, Input-Output Economy, Price Dynamics, High-Dimensional Time Series, Sparsity, Local Factors, Global Business Cycles.

* Correspondence: Department of Economics, University of Pennsylvania, 133 South 36th Street, Philadelphia, PA 19104-6297. Email: mlikota@sas.upenn.edu. I am highly indebted to my advisors, Frank Diebold and Frank Schorfheide, for their invaluable support and guidance. For helpful comments and discussions I would also like to thank Xu Cheng, Wayne Gao, Leon Huetsch, Daniel Lewis, Karen Lewis, Aaron Mora, Joao Ritto, Adrien Wicht as well as seminar participants at the University of Pennsylvania and the 7th Lindau Meeting on Economic Sciences.

Contents

1 Introduction

Numerous economic environments feature a cross-section of agents or 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] such bilateral links imply a comovement of 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 this comovement plays 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, which leads to a static framework and implies that connections of all order play out simultaneously (see e.g. Acemoglu et al. (2012, 2016); Elliott et al. (2014)). 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 (see 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 the frequency at which data is sampled and it remains an empirical question how far a shock travels through the network in one period. Only for studies limited to steady state comparisons, i.e. long-term effects of permanent shocks, the exact timing of network effects is irrelevant. These long-term effects turn out to be the same as the effects in the static framework of contemporaneous linkages. Again, during the period of scientific observation, all spillover and spillback effects fully materialize, to borrow the phrasing from international economics.

Yet, many interesting questions concern transition dynamics rather than steady state comparisons. For example, given a shock experienced by contact-intensive sectors, such as during the onset of the COVID-19 pandemic, firms in other sectors are not only interested in how strong the overall, long-term effect on their marginal costs or output will be, but also when they will first start to feel the impact and how the effect will unfold over time more generally. This possibly includes a changing speed at which indirect, higher-order network effects accumulate over time as well as overreactions and subsequent corrections, as often observed in financial markets. Similar concerns plague various agents in countries all over the world when an important wheat-exporting country is torn by war.

I build an econometric framework which enables me to answer such questions. It consists of a Vector Autoregression (VAR) parameterized based on the assumption that innovations

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

transmit cross-sectionally only via bilateral links. The model can accommodate general patterns on how innovations travel through the network over time, i.e. which connection-orders matter for transmission at which horizons.

The Network-VAR (NVAR) is applicable whenever one is interested in the dynamics of a cross-section linked by bilateral ties, whether or not network data is available. Given a network and a cross-sectional time series supposed to be driven by it, the framework can be used to quantify how network effects of different order unfold over time and come to shape cross-sectional dynamics. In absence of network data, the assumption of innovation transmission via bilateral links yields a sparse, yet flexible and interpretable way of modeling cross-sectional processes, even in high dimensions. I illustrate these two uses of the model with two applications.

In the first application, I take the supply chain network of the US economy as given and study how it shapes the monthly dynamics of Producer Price Indices (PPI) across sectors. I show that the proposed NVAR indeed approximates the process of sectoral prices in an input-output economy with time lags between the production of goods and their subsequent usage as intermediaries in producing other goods. The particular specification suggested by theory leads to the same long-term effects of permanent sectoral price shocks as those reported in the literature on network-induced amplification of granular shocks. By taking an explicit stance on the timing of network effects rather than assuming contemporaneous interactions, the framework can decompose these overall, long-term effects over time and speak to transition dynamics. With network data given, inference on the time profile of network interactions boils down to a pooled-OLS-like regression with covariates that summarize lagged observations using network connections of different order. The appropriate lag and order lengths are inferred from data based on model selection criteria. I show that the set of connection-orders in the VAR-specification is fundamentally related to the frequency of network interactions relative to the frequency of observation.

The results suggest that sectors differ not only in the strength of their impact on aggregate prices, but also in 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 (persistent).

In the second application, rather than taking the network as given, I use the NVAR to model industrial production growth across 44 countries by assuming and in turn estimating an underlying network as relevant for dynamics. This provides a novel perspective on global business cycles as it assumes that the dynamic comovement in economic activity across countries is the result of bilateral connections. The model yields a sparse, yet flexible way of approximating cross-sectional time series 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 – as relevant for dynamics – is estimated and that, given a network, the model can accommodate general patterns of how network effects of different order accrue over time. Estimation is conducted by Least Squares with a Lasso penalty on network links. It involves iterating on the conditional estimator for the timing of network effects given the network – as in the first application – and vice versa.

My theoretical analysis reveals that the NVAR is expected to better capture cross-sectional dynamics than a factor model whenever they are composed of many micro links rather than driven by a few influential units. And indeed, in my application, the NVAR leads to reductions in out-of-sample mean squared errors of up to 20% relative to a principal components factor model, in particular for horizons up to six months. Furthermore, the model returns an estimate for the network as relevant for industrial production dynamics which is roughly in line with expectations and features the US as the most influential country. By explicitly modeling the dependence for any pair of countries, it estimates the whole set of spillover and spillback effects. Differences in these across country pairs are attributed to differences in network-connectedness of different order.

Related Literature At a fundamental level, my work relates to the series of papers by Diebold and Yilmaz (2009, 2014). They map variance decompositions of VARs into weighted, directed and time-varying networks with the goal of understanding dynamic connectedness.[2] 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, making the analysis closely related to Dufour and Renault (1998).

This paper adds to the growing literature on networks in econometrics.[3] 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). Usually, the network is taken as given, but more recent contributions aim at inferring it from observables (e.g. Ahrens and Bhattacharjee (2015); Qu et al. (2021); de Paula et al. (2020)).[4] In contrast, my framework features lagged network effects. The main difference to other

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

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

[4]Outside of SARs, the latter goal is also pursued by Fan et al. (2009); Brownlees et al. (2018); Alidaee et al. (2020), among others. A related literature is concerned with detecting communities and influential units, e.g. Barigozzi et al. (2014); Brownlees and Mesters (2021). Note that in time series models, uncovering contemporaneous dependencies amounts to shock identification (see e.g. Hipp (2020); Dalhaus et al. (2021)).

studies in this category (Knight et al., 2016; Zhu et al., 2017; Yang and Lee, 2019) is that I cast them in an explicit time series model and that my interest lies in studying dynamic, contagion-like innovation transmission through the network. To do so, I generalize the time profile of network effects and in turn conduct inference on it. Furthermore, I do not require network data, but show how to estimate the network as relevant for dynamics. These points also distinguish my work from other studies which model cross-sectional time series by relying on networks. The most prominent is the Global VAR (GVAR) originally proposed in Pesaran et al. (2004). It builds a multi-variable model for the global economy by combining country-level VARs, each of which is estimated by taking as given variables of other countries, averaged using network data such as trade and capital flows.[5] A complementary approach is taken by Barigozzi et al. (2022), who build a framework to forecast a single cross-sectional variable based on factors extracted from multiple observed connection-types. Bykhovskaya (2021) is interested in the temporal evolution of the network itself.

With the first application of the NVAR, I contribute to the wide macroeconomic literature on macro implications of micro shocks and the related literature on multi-sector business cycles under input-output linkages (Horvath, 2000; Foerster et al., 2011; Bouakez et al., 2014). The former shows in particular that disturbances to aggregate Total Factor Productivity (TFP) can be microfounded by idiosyncratic, firm-level shocks in case firms differ in size (Gabaix, 2011) or in their positions in the production network (Acemoglu et al., 2012).[6] Many studies are also interested in the implications of networks at the level of the unit or cross-section as a whole (see e.g. Giroud and Mueller (2019); Giovanni et al. (2018)). In either case, the result that networks amplify idiosyncratic shocks is usually obtained by assuming contemporaneous network interactions. This framework is silent on how networks drive aggregate (and cross-sectional) dynamics.[7] I provide an econometric framework to conduct inference on how networks drive dynamics. In addition, I apply the framework to document empirically how supply chain linkages affect sectoral producer-price dynamics.

With the second application of the NVAR, this paper addresses the vast literature on dimensionality-reduction techniques for modeling high-dimensional processes. The proposed model 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.

[5] Though my analysis is limited to a single variable per cross-sectional unit, an extension of the NVAR to multiple variables is conceptually straightforward. However, in this case precise estimation of the network(s) requires restrictions. See e.g. Mehl et al. (2022), who estimate cross-country-variable-links by interacting a low-dimensional parameter vector with bilateral trade statistics, as suggested my macroeconomic theory.

[6] Intuitively, micro shocks do not wash out as the number of firms increases because large firms or firms with a central position in the supply chain network remain influential.

[7] 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 Section 2.3 for further discussion.

Compared to variable selection methods such as Lasso (Tibshirani, 1996),[8] 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 may matter at different lags.[9] I derive an equivalence result to factor models, whereby the number of factors is given by the number of non-redundant columns in the network adjacency matrix. Correspondingly, NVAR is expected to better capture cross-sectional dynamics whenever these are driven by many seemingly negligible micro-links rather than a few influential units, i.e. when the network adjacency matrix is estimated to be close-to-full rank, yet sparse, in line with the discussion in Boivin and Ng (2006). Sparse factors are rationalized in my framework as locally important units.[10] However, note that the NVAR does not rely on factor extraction but uses lagged observations, more in line with reduced rank regression.[11]

The remainder of this paper is structured as follows. Section 2 presents the proposed econometric framework and discusses its properties. In Section 3, I apply it to study how input-output connections shape the dynamics of sectoral prices in the US economy, taking the network as given. In Section 4, I discuss how to model cross-sectional time series by estimating a sparse network that underlies dynamics, which in turn is put into practice for modeling cross-country industrial production dynamics. Section 5 concludes.

2 Dynamics Derived From Bilateral Links

In this section I present the proposed framework. First I provide some background on bilateral network connections in Section 2.1. Section 2.2 then constructs the general NVAR by building on simple examples of lagged innovation transmission via bilateral links. Section 2.3 discusses the relation to contemporaneous network interactions. Finally, in Section 2.4 I explicitly examine the relation between the frequencies of network interaction and observation. Estimation is discussed in Sections 3.3 and 4.3 for the respective cases with and without network data.

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. If $a_{ij} = 0$, I will say unit i is not connected to unit

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

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

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

[11] As a result of this as well as the absence of contemporaneous interactions, the NVAR does not face any notable identification challenges.

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.[12]

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.

In short, 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, $a_{ij} \equiv (A)_{ij}$, 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.[13] I will refer to this quantity as the Kth-order connection from i to j. A walk from a unit i to itself is called a cycle.[14]

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.

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. By transmission I mean the response of a unit to a disturbance at another unit in partial equilibrium, not taking into account the responses of other units. It is assumed to flow only in one direction through the network. Specifically, the direct link from i to j, a_{ij}, is a vehicle for transmission from j to i. The innovations are allowed to follow an arbitrary process with a mean of zero. In particular, they may be cross-sectionally correlated. For expositional simplicity, I focus on a cross-sectional time series y_t with mean zero.

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

[13] 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.

[14] Usually, a cycle is defined to be a path from i to i, a path being a walk where each intermediary unit is distinct. The differentiation between walks and paths is not relevant for this paper.

If transmission from one unit to another takes one period, and only one period, to materialize, the dynamics of y_t can be represented as follows:

$$y_t = \alpha A y_{t-1} + u_t \ , \quad \alpha \in \mathbb{R} \ . \tag{1}$$

This is a VAR(1) with the VAR-matrix given by the network adjacency matrix A, multiplied by some scalar α which represents the strength of innovation transmission through A. In this case, the one period-ahead expected value of the series for unit i is proportional to a weighted sum of one period-lagged values of the series of all units j to which i is directly linked, with weights given by the strength of direct links a_{ij}:

$$\mathbb{E}_{t-1}[y_{it}] = \alpha \sum_{j=1}^{n} a_{ij} y_{j,t-1} \ .$$[15]

This process is used by Golub and Jackson (2010) in their study of societal opinion formation through friendship ties. Also, Long and Plosser (1983) derive it for sectoral output in a production economy with a one period delay in converting inputs into output. In this context, a sector is expected to produce more (less) than it does on average tomorrow if and only if (a weighted average of) its direct supplier-sectors produced more (less) than they do on average today.

Under this process, the dynamics of y_t, as summarized by Granger causality at different horizons $h = 1, 2, ...$, are shaped by hth order connections encoded in A:

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

As a result, given all other variables y_k , $k \neq j$, y_j is useful in forecasting y_i at horizon h if and only if 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 number and strength 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.

Figure 1 provides an example and depicts the Granger causality pattern for the process in Equation 1 and the network introduced in Section 2.1. In each panel (i, j), the blue line shows $(A^h)_{i,j}$, the network connection from i to j of different order h. The red line represents the decaying series α^h for $\alpha = 0.9$. The purple line shows their product, the GIRF. By definition of the latter, 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.

[15] $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{y_{t-1-j}\}_{j=0}^{\infty}$ is the information set at $t-1$.

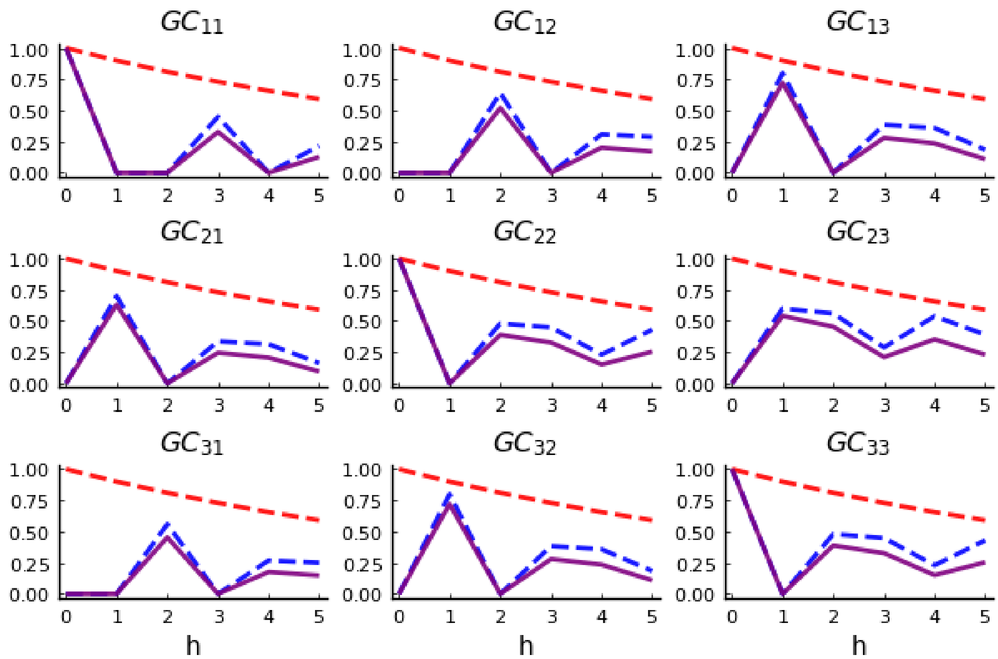

Figure 1: Example Generalized Impulse Responses: NVAR(1, 1)

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

Two points are worth highlighting. First, 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. In the production network example, firm 3 experiences disturbances to its supplier 2 earlier than disturbances to the supplier of its supplier, firm 1. Second, even though in this particular example the diagonal elements in A are all zero and therefore the individual processes do not allow for persistence through own lags, units can experience second round effects to own innovations due to 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, firm 3 will experience further rounds of adjustments because its initial response led to a response of its supplier, firm 2. For larger networks, such second-round responses – to one own's as well as other units' innovations – can surpass the initial response.[16]

This analysis relates to the discussion in Dufour and Renault (1998), who point out that Granger causality can take the form of chains. Specifically, even though 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. The present discussion illustrates that in case innovations transmit only

[16] If a unit i has weak lower-order connections to some unit j but strong higher-order connections, then 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 $\tilde{h} > h + 1$.

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 units (variables).

The assumption that transmission takes exactly one period of observation to materialize appears rather restrictive, for two reasons. First, it rules out less-than-complete transmission at a single lag. As a result, given a price increase by a supplier j, a firm i fully adjusts its own price after one period. A further price adjustment in the next period is 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 from i to j. Second, the above process cannot accommodate the case where innovations travel through the network at a different speed than the frequency in which the process is cast; firm i cannot be impacted by price increases of suppliers situated two positions upstream of i (suppliers of suppliers) earlier than with a lag of two periods, nor can the (initial) response occur later than with a lag of two periods. In the following I present two qualitative ways to extend the simple process above, which in turn lead to the general NVAR model.

If innovation transmission takes two periods to materialize fully, the process of y_t can be written as

$$y_t = \alpha_1 A y_{t-1} + \alpha_2 A y_{t-2} + u_t \;, \quad \alpha_1, \alpha_2 \in \mathbb{R} \;. \tag{2}$$

This is a VAR(2) where both VAR-matrices are given by the adjacency matrix A, multiplied by scalars α_1 and α_2, respectively. If α_1, α_2 are both positive, their sum takes the role of α in the example before and shows the overall strength of transmission, while their relative size determines how the latter is split over the two lags. They are allowed to be negative. For example, $\alpha_1 > 0, \alpha_2 < 0$ signifies an initial overreaction and subsequent correction of unit i's series after an innovation at one of the units to which it is connected.

Assuming $\alpha_1, \alpha_2 \neq 0$, we obtain the result that y_j Granger-causes y_i at horizon h if and only if there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}+1, ..., h\}$, where $\underline{k} = ceil(h/2)$ and $ceil(x)$ rounds $x \in \mathbb{Q}$ up to the next integer.[17] Compared to the process in Equation 1, the lagged innovation transmission allows connections of order lower than h to matter at any given horizon $h > 1$. For example, firm i can still be reacting to a disturbance at its supplier sector j even after two periods, regardless of their higher-order connections. Furthermore, by setting $\alpha_1 = 0$, this specification can accommodate a network interaction frequency which is lower than that at which data is observed. In that case, rather than traveling one link a period, as in the process in Equation 1, an innovation travels one link every two observational periods. In the example, firm i would be affected by a disturbance to its supplier sector j only after two periods.

Innovation transmission through the network might also take place at a frequency higher than that of observation. For example, if two rounds of transmission materialize within

[17] See proof in Appendix A.1.

one period of observation, second-order connections matter for dynamics already one period ahead. This is captured by the following process:

$$y_t = (\alpha_{11}A + \alpha_{12}A^2)y_{t-1} + u_t \ , \quad \alpha_{11}, \alpha_{12} \in \mathbb{R} \ . \tag{3}$$

Analogously to before, α_{11} and α_{12} indicate the relative strength of first- and second-order connections at horizon $h = 1$, respectively. It is easy to see that here y_j Granger-causes y_i at horizon h if and only if there exists a connection from i to j of at least one order $k \in \{h, h+1, ..., 2h\}$. In the example, firm i can be affected by suppliers of suppliers with a lag of only one observational period.

Combining these ingredients yields a model which entertains the assumption that innovations transmit only via bilateral links and can accommodate general patterns on how this transmission materializes over time:

$$y_t = \sum_{l=1}^{p} \Phi_l y_{t-l} + u_t \ , \quad \Phi_l = \sum_{g=1}^{q} \alpha_{lg} A^g \ , \quad \alpha_{lg} \in \mathbb{R} \ . \tag{4}$$

In this process, assuming again $\alpha_{lg} \neq 0 \ \forall \ l, g$, y_j Granger-causes y_i at horizon h if and only if there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}+1, ..., hq\}$, where $\underline{k} = ceil(h/p)$ (see Proposition 1 in Appendix A.1). To prove this, I establish that the GIRF is of the form

$$\frac{\partial y_{i,t+h}}{\partial u_{j,t}} = c^h_{\underline{k}}(\alpha) \left[A^{\underline{k}}\right]_{ij} + ... + c^h_{hq}(\alpha) \left[A^{hq}\right]_{ij} \ . \tag{5}$$

The coefficients $\{c^h_k(\alpha)\}_{k=\underline{k}:hq}$ are polynomials in $\{\alpha_{lh}\}_{l=1:p,h=1:q}$ and show the importance of different connection-orders for the impulse response at any one horizon h. As a result, dynamics in this process depend on the strength of network connections of relevant order as determined by p, q and $\alpha = \{\alpha_{lh}\}_{l=1:p,h=1:q}$. The parameter p determines how many connections of order lower than h matter at horizon h, q determines how many connections of order higher than h come into play, while the role of α is to amplify or suppress connections of certain order at certain horizons. I will dub this model NVAR(p, q).

2.3 Contemporaneous Innovation Transmission via Bilateral Links

The proposed NVAR abstracts from contemporaneous network interactions, which feature prominently in 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:

$$\tilde{y}_t = A\tilde{y}_t + \varepsilon_t = (I - A)^{-1}\varepsilon_t = (A + A^2 + A^3 + ...)\varepsilon_t \ .^{18}$$

In that literature, contemporaneous interactions rationalize the cross-sectional comovement among $\{\tilde{y}_{it}\}_{i=1}^n$ as the network-induced amplification of cross-sectionally uncorrelated, idiosyncratic shocks ε_{it}. 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 if the interest lies in quantifying overall connectedness via networks, but they are silent on how networks drive dynamics.[19] Nevertheless, models with contemporaneous and lagged network interactions are related. By Proposition 7 in Appendix A.4, the (contemporaneous) response of y_{it} to a (transitory or persistent) innovation to y_{jt} under contemporaneous interactions is equal to its long-run response to a persistent innovation to y_{jt} under lagged interactions in a corresponding NVAR$(p, 1)$. Specifically, for

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \qquad \text{and} \qquad \tilde{y}_{(t)} = \alpha A \tilde{y}_{(t)} + \tilde{u}_{(t)} \ , \ \ \alpha = \sum_{l=1}^{p} \alpha_l \ ,$$

we have

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

provided the processes are stationary.[20] Both responses are given by element (i, j) of the Leontief inverse $(I - \alpha A)^{-1}$, which is a sufficient statistic for the (long-term) cross-sectional comovement of interest. The difference between the two processes is that y_t contains information on how any such long-term effect materializes over time. To provide this information, it needs to take a stance on the time profile of network interactions. In contrast, the timing of interactions is irrelevant if the interest lies only in steady state comparisons rather than full transition dynamics.

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.

[18]Note that if such contemporaneous interactions are combined with the lagged ones featured in the NVAR(p, q), connections of all order higher than h matter for Granger causality at any horizon h, regardless of the values of p and q, albeit in such a specific, restricted way.

[19]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 $\tilde{y}_t$. 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.

[20]A sufficient condition is $|\alpha| < 1/|\lambda|$, where λ is the largest (in absolute value) Eigenvalue of A. If $\alpha_l \geq 0 \, \forall \, l$, this condition is both necessary and sufficient. Stationarity is discussed in Appendix A.2.

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.

2.4 Time Aggregation of Lagged Transmission Patterns

The NVAR embodies two ideas. First, innovation transmission through the network is not instantaneous but happens with a lag (of one or several periods). Second, the frequency of such network interactions can differ from the frequency of observation. In this section, I explicitly differentiate between these two frequencies and discuss the particular NVAR(p, q) followed by the observed process under lagged innovation transmission and different assumptions on the relative frequencies of network interactions and observation.

Let the cross-sectional time series x_τ follow an NVAR$(p^*, 1)$:

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_{p^*} A x_{\tau-p^*} + v_\tau \ , \quad \delta = (\delta_1, ..., \delta_{p^*}) \in \mathbb{R}^{p^*} \ .$$

This means that x_τ is shaped by lagged network interactions, whereby innovation transmission takes p^* periods to fully materialize. Suppose we observe $\{y_t\}_{t=1}^T = \{x_{q^*t}\}_{t=1}^T$, where q^* denotes the frequency of network interactions relative to the frequency of observation. If $q^* = 1$, these two coincide and we in fact observe x_τ. Trivially, y_t follows the same NVAR$(p^*, 1)$ as x_τ. To deal with the cases $q^* \neq 1$, I assume that x_τ is a stock variable (like prices) such that its values at lower frequency are just snapshots of its values at higher frequency rather than (weighted) sums of realizations during an interval of time, which would be the case if x_τ was a flow variable (like output).

If $q^* < 1$ with $1/q^* \in \mathbb{N}$, network interactions happen at a lower frequency than that of observation. In this case, y_t follows an NVAR$(p, 1)$ with $p = p^*/q^*$, whereby every $(1/q^*)$th lag has a non-zero coefficient in front of it:

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \ , \quad \alpha_l = \begin{cases} \delta_{lq^*} & \text{if } l \text{ is multiple of } 1/q^* \\ 0 & \text{otherwise} \end{cases} \ , \quad u_t \sim v_\tau \ .$$ [21]

For example, with monthly observations, $q^* = 1/3$ signifies that network interactions occur at quarterly frequency. As a result, the observed monthly series depends on its value three months ago, six months ago, etc., up to $3p^*$ months ago.

If $q^* > 1$, $q^* \in \mathbb{N}$, network interactions take place at a higher frequency than that of observation. Then, provided that x_τ is stationary, the process for y_t can be approximated

[21] As can be easily verified, under distributional equivalence of u_t and v_τ, $\mathbb{E}[y_t y_{t-h}] = \mathbb{E}[x_\tau x_{\tau-hq^*}] \ \forall \ h$.

arbitrarily well by an NVAR(p, q) with restricted parameters α and autocorrelated innovations. For simplicity, consider $p^* = 3$ and suppose we observe x_τ every $q^* = 2$ periods. Under monthly observations, this means that network interactions happen bi-weekly. The case for general p^* and q^* is discussed in Appendix A.3. The process of x_τ is

$$x_\tau = \delta_1 A x_{\tau-1} + \delta_2 A x_{\tau-2} + \delta_3 A x_{\tau-3} + v_\tau \ , \quad \delta_1, \delta_2, \delta_3 \in \mathbb{R} \ .$$

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

$$\begin{aligned} x_\tau = & \left[\delta_2 A + \delta_1^2 A^2\right] x_{\tau-2} + \left[(\delta_1\delta_2 + 2\delta_1\delta_3) A^2\right] x_{\tau-4} \\ & + v_\tau + \delta_1 A v_{\tau-1} + (\delta_3 A + \delta_1\delta_2 A^2) v_{\tau-3} + \text{terms in } x_{\tau-6}, x_{\tau-7} \ . \end{aligned}$$

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

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

where $y_{t-l} = x_{\tau-2l}$ for $l = 0, 1, 2$, $u_t = [v_\tau', v_{\tau-1}']'$ stacks all the errors which occurred between the periods of observation $t-1$ and t and analogously $u_{t-1} = [v_{\tau-2}', v_{\tau-3}']'$. The matrices are given by

$$\Phi_1 = \delta_2 A + \delta_1^2 A^2 \ , \quad \Phi_2 = (\delta_1\delta_2 + 2\delta_1\delta_3) A^2 \ , \quad \Theta_0 = [I_n, \delta_1 A] \ , \quad \Theta_1 = \left[0_n, \delta_3 A + \delta_1\delta_2 A^2\right] \ .$$

The exact process $\{y_t\}_{t=1}^T = \{x_{q^*t}\}_{t=1}^T$ 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 p.[22] Given p, we get $q = pq^* - p + 1$ (see Appendix A.3).

Note that time aggregation preserves the property that long-term responses to persistent disturbances are equivalent to responses in the corresponding model with contemporaneous interactions. Specifically, the long-term response of the observed series y_t to a permanent increase in the underlying high-frequency innovation v_τ equals $(I - dA)^{-1}$, with $d = \sum_{l=1}^{p^*} \delta_l$.[23] This discussion reveals another way to think about contemporaneous network effects. They arise in the limit, as the frequency of network interactions relative to the frequency of observation goes to infinity. Therefore, in empirical analyses with contemporaneous network effects, one needs to look at data of sufficiently low frequency.

In the following, I consider two distinct applications of the NVAR. Section 3 is concerned with the case where the network is given and the interest lies in quantifying its impact on cross-sectional dynamics. In contrast, Section 4 discusses how the assumption of shock transmission via bilateral links can be useful for parsimoniously modeling cross-sectional time series by estimating a sparse network that captures the observed dynamic comovement.

[22] See Appendix A.2 for a proof that stationarity of y_t follows from stationarity of x_τ, and vice versa.
[23] See Appendix A.4 for details.

3 Input-Output Links and the Dynamics of Prices

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 introduced in the previous section to analyze the dynamic aspects of this comovement. 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 input-output economy in the Real Business Cycle (RBC) tradition, which I discuss in Section 3.1. The long-term price responses to permanent disturbances in my framework are equal to their responses in the static framework with contemporaneous network interactions. I contribute to the literature by documenting how these long-run responses materialize over time and how this time profile of transmission depends on sectors' positions in the input-output network. After theoretically motivating the analysis in Section 3.1, I discuss the data in Section 3.2 and the estimation procedure in Section 3.3, before presenting the results in Section 3.4.

3.1 Theory

This section extends a benchmark input-output economy by introducing time lags in input-output conversion and shows that the resulting process of sectoral prices can be approximated by 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 B.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

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

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

where z_{it} denotes TFP in sector i and w_t is the price of labor. No restrictions on the process of z_{it} are made. 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 who 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 B.1. Let $x_{t,t-h}^{ij}$ 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 to outputs in the same period when they are produced and purchased, i.e. $x_{ijt} = x_{t,t}^{ij} = x_t^{ij}$. 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 adjacency 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. $x_{ijt} = x_{t,t-1}^{ij} = x_{t-1}^{ij}$, we obtain that sectoral prices approximately follow an NVAR(1,1):

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

where $k_t^{p1} = k^{p1} - (\iota - b)ln\,(G_t^w)$. Thereby, k^{p1} is a vector of constants, ι is a vector of ones, $b = (b_1, ..., b_n)'$ contains sectoral labor shares and $G_t^w = w_t/w_{t-1}$ 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 p^*th lag are zero.

As shown in Appendix B.1, this economy leads to almost the same steady state as the above economy with contemporaneous network interactions.[24] However, while the latter is

[24]Differences vanish as the discount factor $\beta \to 1$.

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, in particular 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 $x^{ij}_{t,t-2}$.[25] 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 the Long and Plosser (1983) economy above, the presumption is that storage is done by the buyer. We then have

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

In the Cobb-Douglas case $r \to 0$, 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.[26]

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 positive 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, $r \to 1$. 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 $r \to 1$.

[25] Therefore, the amount of good j purchased at time t can be used in production at periods $t+1$ and $t+2$: $x^{ij}_t = x^{ij}_{t+1,t} + x^{ij}_{t+2,t}$.
[26] Note that $G^w_t G^w_{t-1} = \frac{w_t}{w_{t-2}}$ is the wage growth from $t-2$ to t.

To sum up, under general lags in input-output conversion, the log of sectoral prices, x_τ, at some model-frequency evolves according to an NVAR$(p^*, 1)$,

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_{p^*} A x_{\tau-p^*} + v_\tau \ ,$$

with $\delta_l \geq 0 \ \forall \ l$ and $\sum_{l=1}^{p^*} \delta_l = 1$. Also, $\sum_j a_{ij} < 1 \ \forall \ i$. These restrictions imply that the process is stationary.[27] In the empirical analysis that follows, I allow the model-frequency to differ from the observational frequency, as in Section 2.4, and infer their relation from the data by model selection criteria.

A difference to the (unrestricted) NVAR$(p, 1)$ from the previous section stands out: the domain restrictions $\delta_1, ..., \delta_p \geq 0$ imply that the impulse response to a shock in sector j has the 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.

3.2 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 3.1 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

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

where $sales_{j \to i}$ 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

[27] 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 Corollary 1, derived from Propositions 2 and 3, all of which are in Appendix A.2. The intuition is that for an NVAR$(p^*, 1)$, as well as for the NVAR(p, q) derived from it by time-aggregation, walks of order k in the GIRF at any horizon h are multiplied by products of δ_ls such that the sum of their exponents is k. The domain restrictions above imply that products of δ_ls remain bounded above by 1, while $lim_{k \to \infty} A^k = 0$.

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.[28]

The corresponding time series data on 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 B.2.

Most of the raw log PPI series show a clear upward trend. To render the series stationary, I estimate and subtract a linear trend and a seasonality component.[29] 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.

For the purposes of descriptive statistics only, I set links smaller than 0.01 to 0, as in Carvalho (2014). As a result, network density, defined as the fraction of non-zero links out of the total number of possible links, falls from 73.55% to 16.88%. This reveals that the majority of links are quite weak. As expected for the lower level of disaggregation in the present analysis, the network density is much higher than the 3% reported for the finer level of 417 sectors in Carvalho (2014). A tabular illustration of the network is provided in Appendix B.2.

As illustrated in the left panel of Figure 2, the weighted in-degree, $wd_i^{in} \equiv \sum_j a_{ij}$, lies 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, $wd_j^{out} \equiv \sum_i a_{ij}$, a measure of the reliance of other sectors on the input supplied by a sector j. This plot provides evidence that most sectors are specialized input-suppliers, while there are also a few general-purpose suppliers. This point is further supported by non-weighted out-degrees, $d_j^{out} \equiv \sum_i \mathbf{1}\{a_{ij} > 0\}$, which show the number of customers of a given sector. They are depicted alongside in-degrees in Figure A-3 in Appendix B.2. All statistics are tabulated by sector in Table A-2.

[28] As discussed in Appendix B.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 $(\delta_1\beta + \delta_2\beta^2)^{-1}$. For general CES aggregation, this constant is also a function of the elasticity r. For now I abstract from these differences in the proper calibration of a_{ij}. In other words, the analysis assumes $\beta \to 1$.

[29] 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.

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

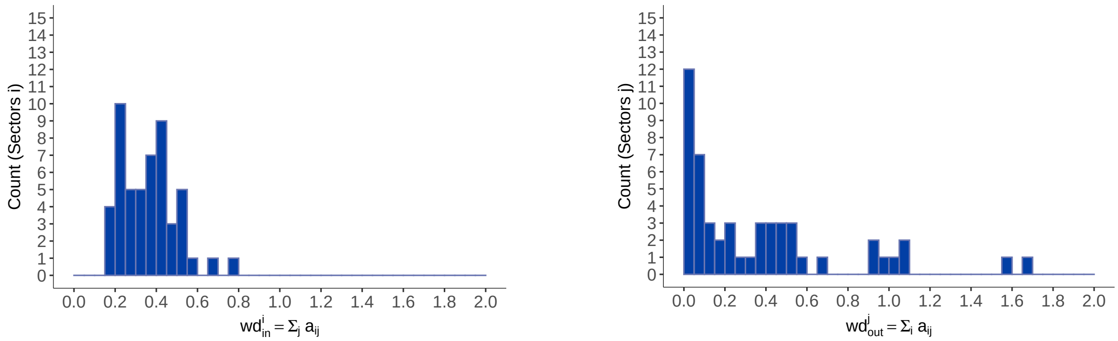

Figure 2: Weighted In-Degrees & Out-Degrees

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.

The distance or shortest path from sector i to sector j is the lowest order at which i is linked to j. It measures how closely connected sector i is to sector j. 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. These relatively low numbers provide evidence of the small-world nature of the input-output network, which the literature attributes to a small number of hub-like sectors (general purpose suppliers in the case of an input-output network) and which is a common feature of networks in economics.[30]

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 Figure 3. 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 PPI with lagged values of sector j's PPI is declining with the distance from sector i to sector j. This is illustrated by the remaining lines in Figure 3, 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.[31]

The left panel in Figure 4 depicts the raw PPI series for a few sectors. It provides evidence of considerable heterogeneity in price dynamics across sectors, even disregarding the highly

[30] See Figure A-2 for a tabular representation of distances between sectors.

[31] Note that Figure 3 plots mean correlations by distance and masks plenty of heterogeneity across sector-pairs.

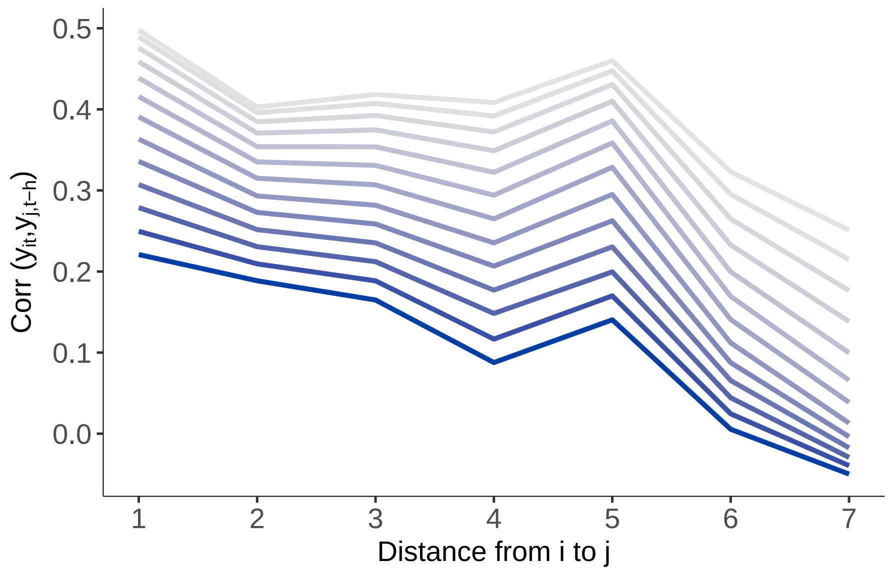

Figure 3: Network Distance and the Correlation of Sectoral Inflation

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.

volatile energy-related sectors. The mean, standard deviation and range of sectoral PPI changes can be found in Table A-2. 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 Figure 4 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.[32] 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. The latter is included in the analysis because it contains potentially valuable information on how price shocks transmit through the input-output network.

3.3 Estimation

The general NVAR(p, q) from Section 2 is

$$y_t = \sum_{l=1}^{p} \left(\sum_{g=1}^{q} \alpha_{lg} A^g \right) y_{t-l} + u_t \ .$$

[32] 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.

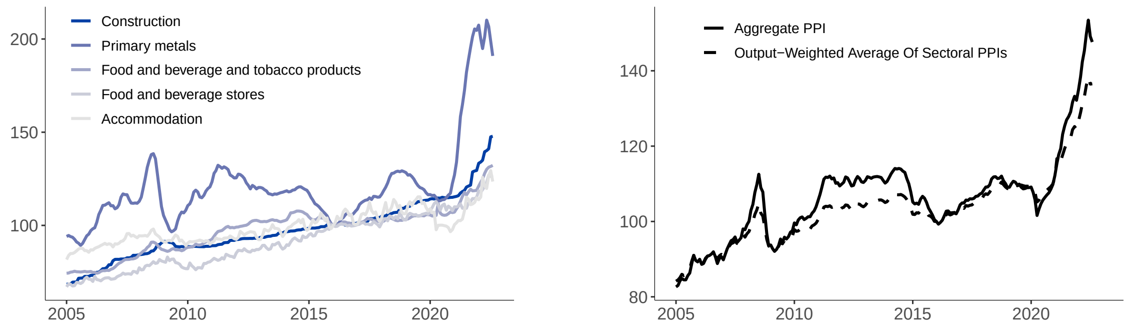

Figure 4: Sectoral and Aggregate PPI

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.

In absence of further imposed structure, analyzing how a given network A shapes the dynamics of y_t requires estimating the coefficients $\{\alpha_{lg}\}_{l=1:p,g=1:q}$. They provide information on how innovations u_{it} propagate cross-sectionally over time through the network.

The model from Section 3.1 does, however, impose additional structure on the estimation problem. It states that sectoral prices at some model-frequency follow an NVAR$(p^*, 1)$:

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_{p^*} A x_{\tau-p^*} + v_\tau \ ,$$

with $\delta_l \geq 0 \ \forall \ l$ and $\sum_{l=1}^{p^*} \delta_l = 1$. Rather than estimating α, I build on the theoretical framework and estimate $(\delta_1, .., \delta_{p^*})$. The unrestricted estimation of $\alpha|A$ is discussed in Appendix B.3. It amounts to a linear regression with regressors generated by summarizing information in lagged values of y_t using the network adjacency matrix A.

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 2.4, let $\{y_t\}_{t=1}^T = \{x_{q^*t}\}_{t=1}^T$ 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.

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 $\theta = (\delta, \sigma)$, where $\delta = (\delta_1, .., \delta_{p^*-1})$ and $\sigma = (\sigma_1, ..., \sigma_n)$. Because of the restriction $\sum_{l=1}^{p^*} \delta_l = 1$, I drop δ_{p^*} from δ and impose the domain restrictions $\delta_l \in [0, 1]$ for $l = 1 : p^* - 1$ and $\sum_{l=1}^{p^*-1} \delta_l \leq 1$. In turn, $\delta_{p^*} = 1 - \sum_{l=1}^{p^*-1} \delta_l$ with $\delta_{p^*} \in [0, 1]$.

For $q^* \leq 1$, the likelihood $p(Y|\theta)$ can be evaluated directly. For $q^* > 1$, the process has a linear-Gaussian state space representation and $p(Y|\theta)$ can be evaluated using the Kalman filter.[33] 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 δ. Additionally, for $q^* > 1$, likelihood evaluation is costly because it is a nonlinear function of θ. For the same reason, any attempt at obtaining $\hat{\theta}_{ML}$ using nonlinear optimization would suffer from local optima issues. Therefore, I consider a Bayesian implementation of the MLE. 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.[34] I use independent, Uniform priors for $\{\delta_l\}_{l=1:p^*-1} \in [0,1]^{p^*-1}$, truncated to satisfy the additional domain restriction $\sum_{l=1}^{p^*-1} \delta_l \leq 1$. The resulting distribution is derived in Appendix B.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 $\hat{\sigma}_{i,ML}$.

3.4 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-3).[35] The following analysis is based on this preferred NVAR$(6,1)$.

Table 2 reports the estimation results for δ. 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. With some estimates close to the boundary of the parameter space, standard asymptotic theory for extremum estimators breaks down and it is difficult to assess the precision of the MLE in the frequentist sense. Instead, I report the 95% Bayesian Highest Posterior Density (HPD) sets, which together with the peaked marginal posteriors,

[33] See Section 2.4 and Appendix A.3 for discussion on the process for observables y_t implied by the NVAR$(p^*,1)$ for x_τ and see Appendix B.3 for its state space representation under $q^* > 1$.

[34] See Herbst and Schorfheide (2015) for a general discussion of the SMC algorithm and Appendix B.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.

[35] Note that with monthly data, models with higher network interaction frequency than monthly ($q^* > 1$) are at a disadvantage because estimation under time aggregation introduces noise. In an ideal setting, data would be sampled at the highest possible frequency and only specifications with $q^* \leq 1$ would be assessed.

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.

shown in Figure A-4 in Appendix B.4, illustrate that δ 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.

Table 2: Estimation Results: δ

	MLE	Mean	Low	High
δ_1	0.1550	0.1557	0.1370	0.1745
δ_2	0.3460	0.3382	0.3168	0.3605
δ_3	0.2816	0.2865	0.2644	0.3129
δ_4	0.0915	0.0991	0.0785	0.1174
δ_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). As discussed in Section 2.2, the impulse response of y_t at horizon h comprises supply chain connections of order $k \in \underline{k} : h$, with $\underline{k} = ceil(h/6)$:

$$\frac{\partial y_{i,t+h}}{\partial u_{j,t}} = \left[\frac{\partial y_{t+h}}{\partial u_t}\right]_{ij} = c^h_{\underline{k}}(\delta) \left[A^{\underline{k}}\right]_{ij} + ... + c^h_h(\delta) \left[A^h\right]_{ij} \ . \tag{6}$$

[36]

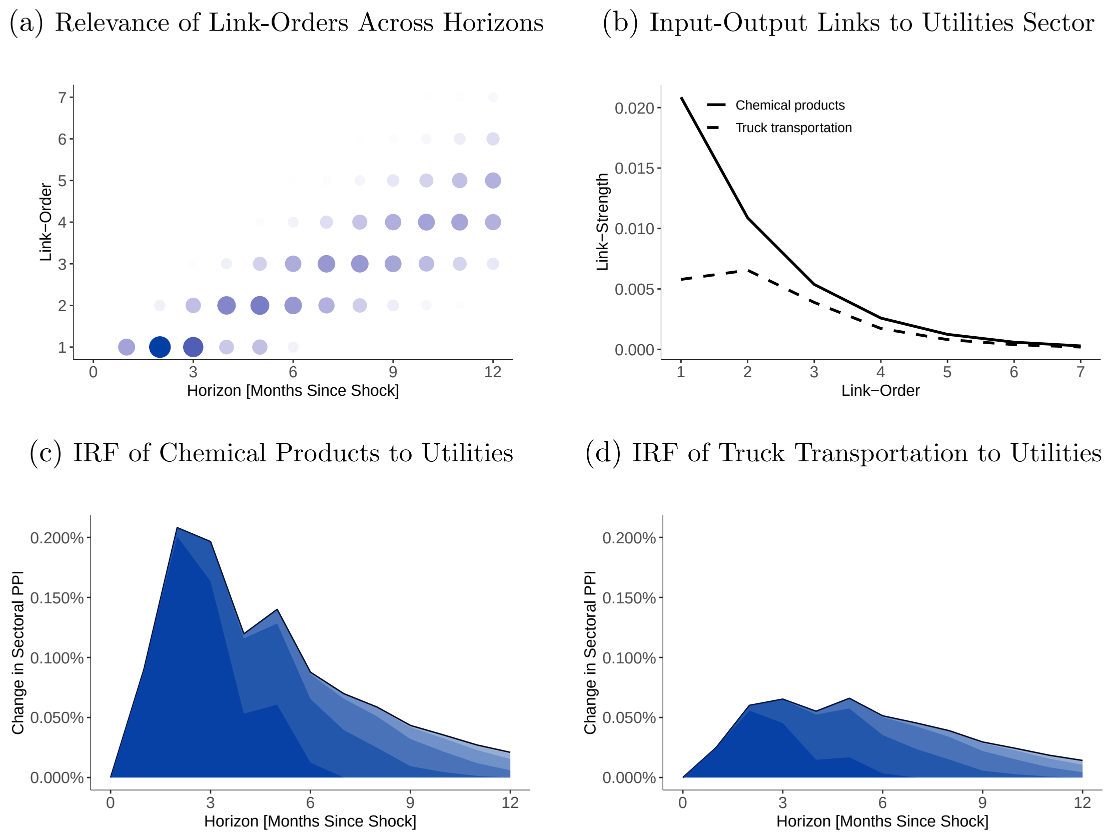

Figure 5: Impulse Responses: Transmission of Price Shocks via Supply-Chain Links

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.

The coefficients $\{c_k^h(\delta)\}_{k=\underline{k}:h}$ are functions of δ 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 δ, these coefficients are constant across time and sector-pairs. They are obtained using Algorithm 1 in Appendix A.3.

Figure 5 illustrates this composition of impulse responses. The dots in the top left panel depict the coefficients $\{c_k^h\}_{k=\underline{k}:h}$ with connection-orders k on the y-axis and horizons h on the x-axis. Larger values are represented by larger and darker dots. As stated above, under $\delta_l > 0$ for $l = 1:6$, at a given horizon h, orders $\underline{k} : h$ matter. 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 $\{\delta_l\}_{l=1:6}$ 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 Figure 5 shows the strength of network connections of different

[36]See Proposition 1 and the discussion on IRFs in Appendix A.4.

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 Equation 6 makes clear, such network-connections from a sector i to a sector j are the second building block of impulse responses in the NVAR.

The lower panels of Figure 5 illustrate the resulting impulse responses. The different shades of blue depict the individual terms $c_k^h(\delta)\left[A^k\right]_{ij}$, which show the contribution of network-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 since the latter share similarly strong higher-order connections to the utilities sector.

In sum, the stronger the connections from sector i to sector j, the more pronounced will be the response of sector i's PPI to a price shock in sector j. Also, 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.

Thanks to the literature on granular origins of business cycles, we know that the effects of sectoral price shocks on aggregate prices are stronger for sectors with more central positions in the supply chain network. As shown in Section 2.3, 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. The comparative advantage of the present framework is that it allows us to study how the effects of a shock unfold over time. In the following, I will 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 Figure 6 shows a few time profiles of aggregate PPI responses to sectoral price shocks. It suggest 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

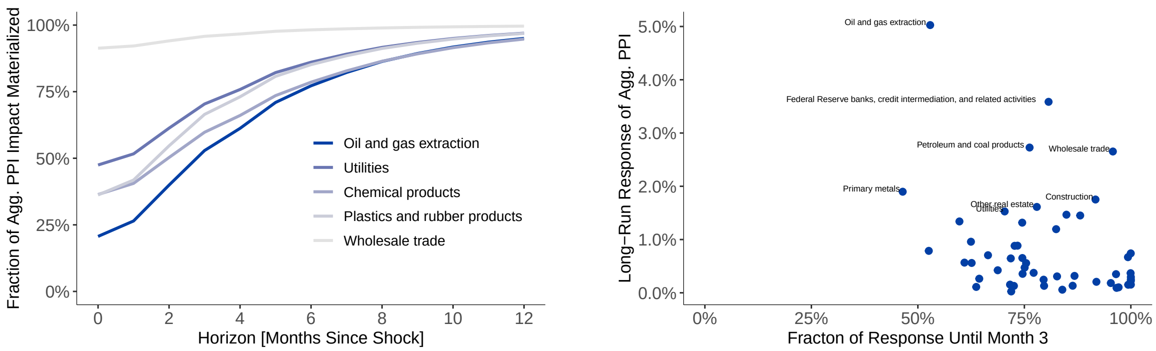

Figure 6: Size and Timing of Aggregate PPI Response to Sectoral Shocks

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.

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 Figure 6 plots the strength of aggregate PPI responses against the fractions which materialize in the first quarter after the shocks to the respective sectors' prices. 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.[37] 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.

4 Global Industrial Production Dynamics

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 such spillover and spillback effects 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

[37] 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 2.3.

with factor models, which in this context posit that it is the result of exposure to a few influential countries.

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. I first provide some intuition on the merits and limitations of the NVAR as a tool for approximating cross-sectional time series in Section 4.1. The relation to factor models, arguably the most popular tool for modeling high-dimensional time series in macroeconomics, is analyzed in more detail in Section 4.2. In Section 4.3 I discuss the estimation of the parameters in the NVAR, this time including the network adjacency matrix. Finally, the results of the application to cross-country industrial production growth are presented in Section 4.4.

4.1 Modeling Cross-Sectional Processes by Sparse Networks

Consider the problem of approximating the process 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,

$$y_t = \sum_{l=1}^{p} \Phi_l(\alpha, A) y_{t-l} + u_t \ , \quad \Phi_l = \sum_{g=1}^{q} \alpha_{lg} A^g \ , \quad \alpha_{lg} \in \mathbb{R} \ , \quad a_{ij} \in [0,1] \ , \tag{7}$$

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 – given by lagged values of y_t – is compressed into a low-dimensional vector of regressors that summarizes this information using network connections of different order. To see this, note that the NVAR(p, q) above can be re-written as

$$y_t = X_t \alpha + u_t \ , \quad X_t = \left[\tilde{y}^1_{t-1}, \tilde{y}^2_{t-1}, ..., \tilde{y}^q_{t-1}, \tilde{y}^1_{t-2}, ..., \tilde{y}^q_{t-p}\right] \ , \quad \tilde{y}^g_{t-l} \equiv A^g y_{t-l} \ . \tag{8}$$

The $n \times pq$ matrix X_t summarizes the information in lags 1 to p of y_t using network connections of order 1 to q. 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. 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.

There are two ways in which the assumption of bilateral links shaping the dynamics at all lags could be restrictive. First, units may be affected by common forces outside of the network. To some extent, this is captured by potential cross-sectional correlation in the disturbances u_{it}. It could further be captured by including additional covariates into the equation. Second, the restriction that the effects of lagged values of y_{jt} on y_{it} are proportional to a linear combination of connections from i to j of different order may be too limiting. However, this appears unlikely given the flexibility of the NVAR(p,q).

Flexibility is obtained because the adjacency matrix A and hence connections of different order are estimated as well as the fact that the parameters $\{\alpha_{lg}\}_{l=1:p,g=1:q}$ can capture a very general pattern of which connection-orders matter at which horizons. The latter point is discussed theoretically in Section 2.2 and documented empirically for the case of price dynamics and input-output links in Section 3.4. The general NVAR(p,q) in Equation 7 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 $\alpha_{lk}A^k$ 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. Note that the first point is qualified by the requirement that $A, A^2, ..., A^k$ are linearly independent, which requires k not to be too large. The Cayley-Hamilton theorem gives the upper bound for the highest power q. It states that for any $n \times n$ matrix A, A^n can be expressed as a linear combination of lower powers of A. Therefore, $q \leq n-1$.

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: variable selection methods, shrinkage estimators, and factor models and reduced rank regression.[38] Variable selection methods like Lasso or boosting aim at finding the most important predictors by excluding less relevant ones. Instead of imposing outright exclusion restrictions, shrinkage methods such as Ridge regression or Minnesota-type priors do so 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. This leads to a model that naturally incorporates weak factors and adds to the interpretability of the estimated process. 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. Relatedly, it entertains the additional sparsity assumption that for every y_{it}, the same linear combinations of predictors y_{jt} matter at all lags (see Equation 8),

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

although possibly in different ways, as captured by differences in $(\alpha_{l1}, ..., \alpha_{lq})$ across lags l. This additional restriction can become important in higher dimensions.[39]

Note that in principle the NVAR could be applied for general, not necessarily cross-sectional time series. However, then the assumption of an adjacency matrix with all positive entries becomes untenable. While this is permissible from a network perspective, estimation becomes non-trivial. I leave this case for future research and limit the analysis to cross-sectional time series.

4.2 Relation to Factor Models

In this section, I derive an equivalence result between an NVAR$(p, 1)$ and a factor model. It supports the subsequent discussion on the environment in which an NVAR is expected to better capture cross-sectional dynamics than a factor model.

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

[39] Note that more coarse levels of cross-sectional granularity offer a rationale to apply shrinkage rather than selection to the bilateral links in A because with less units, the assumption that two units do not share a direct link becomes less tenable. Generally, elements in A could be estimated using an elastic net approach.

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

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

Assuming that Λ has all positive entries, 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 Appendix C.1).

Summing up, and more generally, an NVAR$(p,1)$ yields a particular factor model, where the number of factors is given by the number of non-redundant columns in A. Conversely, for large n, a factor model for y_t can be cast as an NVAR$(p,1)$ – with the number of factors again equal to the rank of A – provided that the factor loadings are all positive and that the factors themselves evolve according to an NVAR$(p,1)$. Note that for $p=1$, the latter condition is equivalent to saying that the factors evolve according to a VAR(1), while in case of a single factor, it just requires the factor to follow an AR(p). The assumption of positive factor loadings Λ is the factor model-equivalent to the assumption of positive links a_{ij} in A and is tenable in many cross-sectional environments. It implies that the sign of the response of y_{it} to a disturbance in y_{jt} is the same for all units i. It is violated, for example, in case an increase in economic activity in one country leads to a contraction in other countries, or if a price increase in one sector leads to a price decrease in other sectors.

Sparse or weak factors are rationalized by the NVAR as locally important nodes. Examples are sectors which supply only a subset of other sectors in the economy or countries that trade only with a subset of other countries. In many cases, we expect A to be sparse, yet of close-to-full rank. For example, for most sectors we can find at least one other sector whose output or price-setting behavior depends on that of the sector in question. Similarly, in the study of global business cycles, for most countries we can find at least one other country whose output depends on the economic activity in the country in question.

Based on these insights, the NVAR is expected to better capture cross-sectional dynamics in cases where the latter are composed of many seemingly neglibile micro links rather than driven by a few influential units, i.e. if A is close to full rank. It is an empirical question to what extent these micro links can be captured by a few latent factors. Yet, 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 $\Lambda_{i\cdot}$ 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.[40] This dispersion notably includes the case of weak factors, as captured by a sparse loading matrix Λ or a sparse adjacency matrix A in the case of an NVAR. Note that in the context of 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 (see causality chain discussion in Section 2.2). Therefore, the NVAR is 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 enables the analysis of spillover and spillback effects as it estimates the whole set of IRFs. Third, the estimated network offers a possible method for shock identification even in high dimensions, the assumption being that the same bilateral links that rationalize lagged innovation transmission are also behind contemporaneous shock transmission.

4.3 Estimation

This section discusses the joint estimation of (α, A) in the NVAR. For now, the analysis is limited to an NVAR$(p, 1)$,

$$y_t = \sum_{l=1}^{p} \alpha_l A y_{t-l} + u_t \ , \quad \alpha \equiv (\alpha_1, ..., \alpha_p) \in \mathbb{R}^p \ , \quad a_{ij} \in [0, 1] \ , \quad \mathbb{V}[u_t] = \Sigma \ .$$

although some results I discuss hold for $q > 1$ as well.

Note that (α, A) are jointly identified only up to scale. To render them identified, I normalize $||\alpha||_1 = 1$. Note that this requires $||\alpha||_1 > 0$ to hold in the true data-generating

[40] More specifically, while they suggest to select the number of factors for each series separately in order to improve forecasts, the forecasts for series that depend on less dominant factors will nevertheless be 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.

process. To deal with the possibility that $||\alpha||_1 a_{ij} > 1$ for some (i, j), I change the domain of a_{ij} to $\mathbb{R}_+$ and, if needed, re-scale the estimates to interpret the estimated A as a network. Appendix C.1 contains more details on identification and normalization.

Consistent with the motivation above, I consider Least Squares (LS) estimation of (α, A, Σ) under a Lasso-penalty for elements a_{ij} in A. For generality, I also consider a Ridge-penalty for α. Shrinkage of α becomes important for estimating a general NVAR(p, q) as different powers of A could be close to multicollinear.[41] This leads to the following optimization problem:

$$\underset{\alpha,A}{min}\ \frac{1}{nT}\sum_{t=1}^{T} u_t(\alpha, A)'\Sigma^{-1}u_t(\alpha, A) + \tilde{\lambda}\sum_{i,j=1}^{n} a_{ij} + \tilde{\varphi}\sum_{l,g}\alpha_{lg}^2 \quad \text{s.t.} \quad a_{ij} \geq 0\ , \quad ||\alpha||_1 = 1\ , \tag{9}$$

with

$$u_t(\alpha, A) = y_t - \sum_{l=1}^{p}\alpha_l A y_{t-l} = y_t - X_t\alpha = y_t - Az_t\ .$$

In this expression, $X_t(A) = A[y_{t-1}, ..., y_{t-p}]$ is an $n \times p$ matrix and $z_t(\alpha) = [\alpha_1 y_{t-1}, ..., \alpha_p y_{t-p}]$ is an $n \times 1$ vector. The hyperparameters $\tilde{\varphi}$ and $\tilde{\lambda}$ control the shrinkage of α and sparsity of A, respectively. Under Ordinary Least Squares (OLS), we get the conditional estimators

$$\hat{\alpha}_{LS}|A = \left[\varphi I + \sum_{t=1}^{T} X_t' X_t\right]^{-1}\left[\sum_{t=1}^{T} X_t' y_t\right]\ , \tag{10}$$

$$\hat{a}_{ij,LS}|(\alpha, A_{i,-j}) = max\{0, \breve{a}_{ij}\}\ \ , \quad \breve{a}_{ij} = \frac{\sum_{t=1}^{T}(y_{it} - A_{i,-j}z_{-j,t})z_{jt} - \lambda}{\sum_{t=1}^{T} z_{jt}^2}\ , \tag{11}$$

where $\varphi = nT\tilde{\varphi}$ and $\lambda = \frac{nT}{2}\tilde{\lambda}$. $A_{i,-j}$ is the ith row and all except the jth column of A. $z_{-j,t}$ is defined analogously. Below, I wil write $A_{\cdot,-j}$ to denote the matrix obtained by deleting the jth column of A.

There is no analytical expression for the unconditional estimator $(\hat{\alpha}_{LS}, \hat{A}_{LS})$. However, it is obtained efficiently by iterating on the conditional estimators $\hat{\alpha}_{LS}|A$ and $\hat{A}_{\cdot,j,LS}|(A_{\cdot,-j}, \alpha)$ until convergence as outlined in Meng and Rubin (1993). Appendix C.2 discusses the procedure in more detail. In turn,

$$\hat{\Sigma}_{LS} = \frac{1}{T}u_t(\hat{\alpha}_{LS}, \hat{A}_{LS})u_t(\hat{\alpha}_{LS}, \hat{A}_{LS})'\ .$$

Provided u_t is Normal, the LS estimator of (α, A, Σ) in the general NVAR(p, q) is equal to the posterior mode under the priors $\alpha \sim N(0, \varphi^{-1}I)$, $a_{ij} \sim$ Exponential(λ) and an improper

[41] The rationale is the same as under the estimation of α when A is given, which is discussed in Section B.3.

prior for Σ: $p(\Sigma) \propto c$ (see again Appendix C.2). By standard arguments, analytical expressions for the conditional posteriors and posterior modes of $\alpha|A, \Sigma$ and $\Sigma|\alpha, A$ are available. The conditional posterior and posterior mode of $A|\alpha, \Sigma$ can only be solved for under $q = 1$. In this case, Bayesian estimation can be implemented efficiently using Gibbs sampling. In contrast, estimation under $q > 1$ cannot be conducted by iterating on conditional posteriors - in case of Bayesian estimation - or posterior modes - in case of LS estimation.

4.4 Results

In this section, I apply the NVAR$(p, 1)$ with (α, A) estimated by LS to model the dynamics of monthly industrial production growth across countries. This shows that even this restricted version of the proposed NVAR(p, q) is a viable alternative to factor models for modeling cross-sectional dynamics. Also, I demonstrate that an IRF-analysis akin to that in Section 3 can be conducted, despite the lack of network data.

Industrial production 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-4 in Appendix C.3.

To assess forecasting performance, I first estimate the 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.

The NVAR$(p, 1)$ is estimated as outlined in the previous section. I select the optimal degree of sparsity in A based on BIC by counting the number of non-zero elements in $\hat{A}(\lambda)$ (see Zou et al. (2007)). No shrinkage to α is applied ($\varphi = 0$). 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. The number of factors is selected based on the information criterion developed in Bai and Ng (2002). Forecasts are constructed by fitting a VAR(p) for the factors.

The results of the forecasting exercise are shown in Figure 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 results for alternative choices of p are similar (see Figure A-6). The NVAR$(p, 1)$ yields a substantial reduction in MSE compared to the factor model. This holds in particular for forecasts up to six months ahead.

On top of forecasting performance, an additional advantage of the NVAR is the high degree of interpretability it offers. In particular, it returns an estimate for the network as relevant for industrial production dynamics as well as for the full set of spillover and spillback

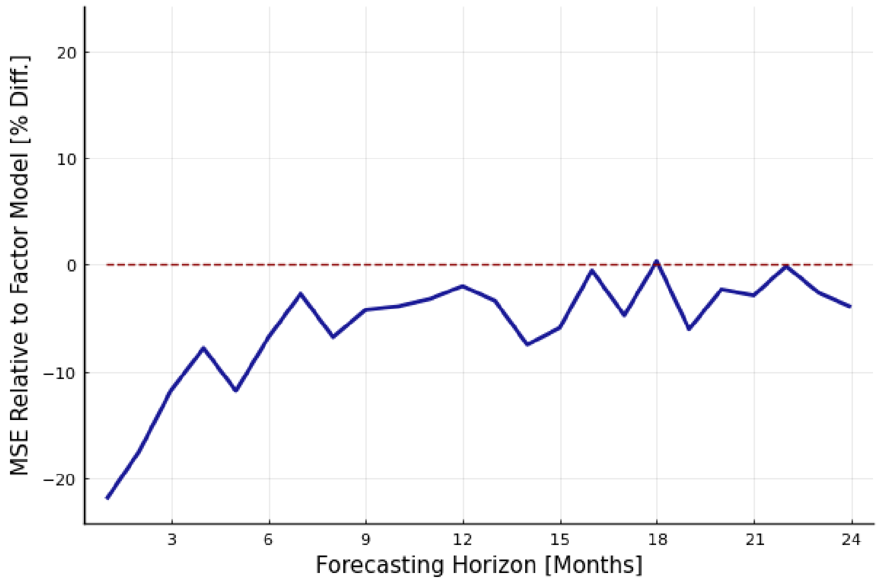

Figure 7: Out-of-Sample Forecasting Performance: NVAR$(4, 1)$ vs. Factor Model

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.

effects 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 the (conditional) estimator for a_{ij} is truncated to $\mathbb{R}_+$ (see Equation 11). In the estimated network, 22% of the links are non-zero. Excluding links below 0.05, this density drops to 11%. Yet, $\hat{A}$ 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 Figure 7, this number is lower, but always stays at rather high levels.

Figure 8 shows weighted out-degrees, $\hat{wd}^j_{out} = \sum_i \hat{a}_{ij}$, a measure of country j's influence on industrial production dynamics of other countries in the sample. Without any information beyond the industrial production series across countries, the NVAR estimates the most influential country to be the United States, in line with intuition. 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 because of the 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, or restrict the link of Germany to Slovenia, say, to be no greater than the link from Slovenia to Germany.

In the top left panel of Figure 9, I illustrate the propagation pattern of innovations to

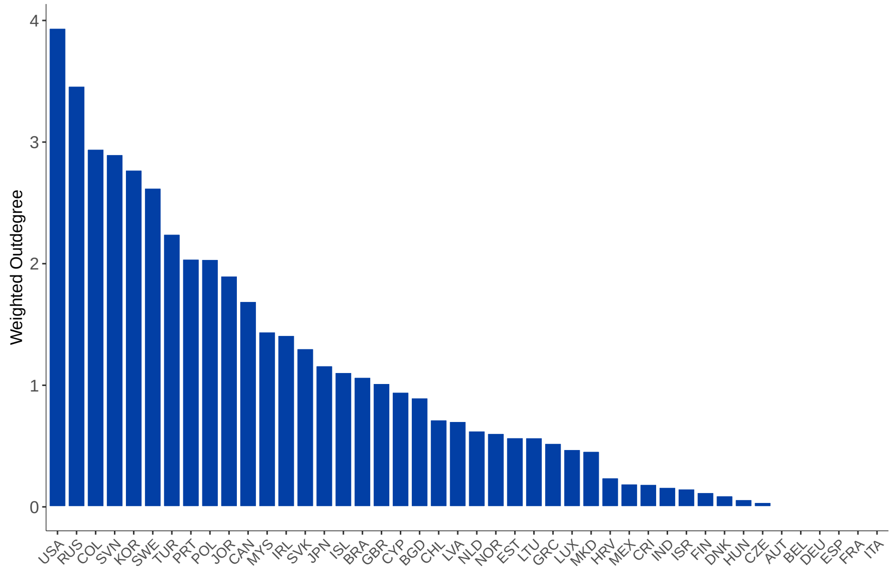

Figure 8: Weighted Out-degrees in the Estimated Network

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

industrial production growth through the network as captured by $\hat{\alpha}$. As before in Figure 5, the dots show which connection orders matter for innovation propagation at which horizon, with the strength reflected by the dots' size and shading. As opposed to the case of Figure 5, however, the coefficients in α are not restricted to be positive in this application. To distinguish positive from negative transmission via network connections, I show the former in blue and the latter in red. The plot suggests that following an incrase in a country's industrial production 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 Figure 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 Finnish industrial production on that of the US comes only from higher-order connections and is weaker overall. As a result, the lower panels of Figure 9 show that industrial production in Germany responds much faster and stronger to an increase in US industrial production 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. Even before network effects come to a complete halt, the impulse response dies out as the positive transmission through higher link-orders starts to balance with the corrective, negative transmission via lower

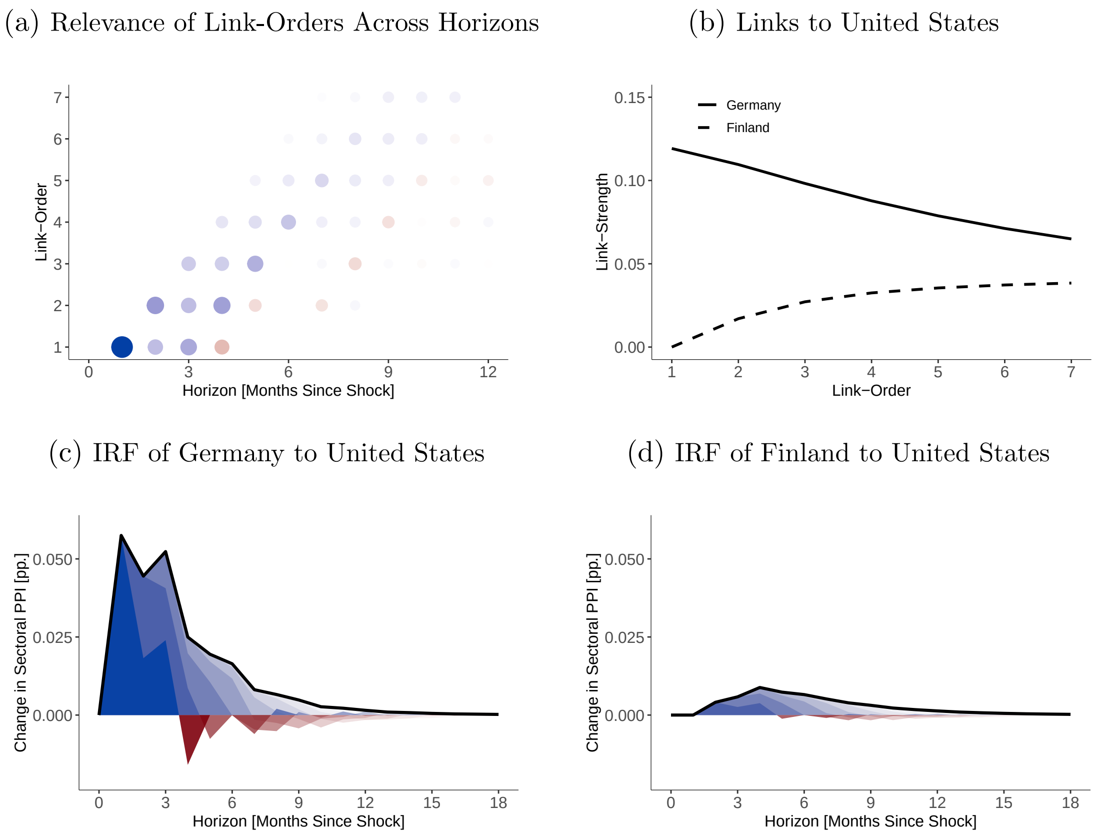

Figure 9: Impulse Responses: Network-Induced Transmission of Economic Activity

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.

orders.[42]

5 Conclusion

In this paper I develop an econometric framework for the dynamics of a cross-section linked by a network of bilateral ties. In a first application, I use it to document how supply chain linkages affect the dynamics of sectoral prices in the US economy. In a second application, I use it as a dimensionality-reduction technique for modeling cross-country industrial production growth.

Directions for future research in the spirit of the first application abound; many environments in economics feature a cross-section linked by a network, for which oftentimes data is

[42]Note that the interpretative decomposition of impulse responses in Figure 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 values.

available. Thereby, the framework could be modified to accommodate heterogeneous propagation patterns across units or time. By adding covariates, it could assess to what extent dynamics are driven by a particular network as opposed to other forces.

In contrast to its appeal for conducting inference on the time dimension of network effects, the usefulness of the proposed framework as a parsimonious modeling device can only be determined over time by practical applications. A methodological innovation in this realm would be to use covariates on the unit- or unit-pair-level to shrink the estimated network towards a prior benchmark, thereby avoiding an illusion of sparsity.

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.

Ahrens, A. and A. Bhattacharjee (2015): "Two-step lasso estimation of the spatial weights matrix," *Econometrics*, 3, 128–155.

Alidaee, H., E. Auerbach, and M. P. Leung (2020): "Recovering Network Structure from Aggregated Relational Data using Penalized Regression," *Manuscript, Kellogg School of Management, Northwestern University.*

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., C. Brownlees, G. M. Gallo, and D. Veredas (2014): "Disentangling systematic and idiosyncratic dynamics in panels of volatility measures," *Journal of Econometrics*, 182, 364–384.

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.*

Brownlees, C. and G. Mesters (2021): "Detecting granular time series in large panels," *Journal of Econometrics*, 220, 544–561.

Brownlees, C., E. Nualart, and Y. Sun (2018): "Realized Networks," *Journal of Applied Econometrics*, 33, 986–1006.

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.*

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

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.*

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

de Paula, A., I. Rasul, and P. C. Souza (2020): "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition," *Manuscript, University College London.*

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., Y. Feng, and Y. Wu (2009): "Network exploration via the adaptive LASSO and SCAD penalties," *Annals of Applied Statistics*, 3, 521–541.

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.*

Gabaix, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

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.

QU, X., L. FEI LEE, AND C. YANG (2021): "Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables," *Journal of Econometrics*, 221, 180–197.

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

This Appendix consists of the following sections:

A Network-VAR

A.1 Granger Causality

Proposition 1 (Granger Causality in NVAR(p,q))**.**
Let y_t follow an NVAR(p,q):

$$y_t = \sum_{l=1}^{p} \Phi_l y_{t-l} + u_t \ , \quad \Phi_l = \sum_{g=1}^{q} \alpha_{lg} A^h \ , \quad \alpha_{lh} \in \mathbb{R} \ ,$$

and assume $\alpha_{lg} \neq 0$ for $l = 1:p, g = 1:q$. Then $\frac{\partial y_{i,t+h}}{\partial y_{j,t}} | \mathcal{F}_t > 0 \Leftrightarrow (A^k)_{ij} > 0$ for at least one $k \in \underline{k} : hq$, $\underline{k} = ceil(h/p)$, 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 $k \in \underline{k} : hq$.

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

$$\frac{\partial y_{t+h}}{\partial y_t} = [I_n, 0_{n\times n(p-1)}] F^h [I_n, 0_{n\times n(p-1)}]' = (F^h)_{11} \ ,$$
$$F = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & I_n & \dots & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & \dots & I_n & 0_n \end{bmatrix} \ , \quad \Phi_l = \sum_{g=1}^{q} \alpha_{lg} A^g \ .$$

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

A.2 Stationarity

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

$$\tilde{x}_t = dA\tilde{x}_{t-1} + u_t \ ,$$

where $u_t \sim WN$. If $d \neq 0$, $\tilde{x}_t$ is stationary if and only if for all Eigenvalues λ_i of A it holds that $|\lambda_i| < 1/|d|$.

Proof: This follows directly from the fact that λ_i is an Eigenvalue of A if and only if $d\lambda_i$ is an Eigenvalue of dA. More formally, let

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

Stationarity of $\tilde{x}_t$ is equivalent to the statment that

$$\begin{aligned} &\forall\, \tilde{\lambda}_i \in \tilde{\mathcal{L}} \ , \quad |\tilde{\lambda}_i| < 1 \\ \Leftrightarrow \quad &\forall\, \tilde{\lambda}_i \in \tilde{\mathcal{L}} \ , \quad |\tilde{\lambda}_i/d| = |\tilde{\lambda}_i|/|d| < 1/|d| \ . \end{aligned}$$

Also, note that

$$|\tilde{\lambda}_i I - dA| = |d(\tilde{\lambda}_i/dI - A)| = d^n|\tilde{\lambda}_i/dI - A| = 0 \Leftrightarrow |\tilde{\lambda}_i/dI - A| = 0 \ ,$$

regardless of the dimension n of A. Therefore, we have that

$$\begin{aligned} &\forall\, \tilde{\lambda}_i \text{ such that } |\tilde{\lambda}_i/dI - A| = 0 \ , \quad |\tilde{\lambda}_i/d| < 1/|d| \\ \Leftrightarrow \quad &\forall\, \lambda_i \in \mathcal{L} \ , \quad |\lambda_i| < 1/|d| \ . \quad ■ \end{aligned}$$

Proposition 3.
Let x_t follow an NVAR($p^*, 1$):

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ ,$$

where $v_t \sim WN$, and assume $\delta_l \geq 0$ for $l = 1 : p$ and $\delta_l > 0$ for at least one l. Define $d = \sum_{l=1}^{p} \delta_l$, and let

$$\tilde{x}_t = dA\tilde{x}_{t-1} + u_t \ .$$

Then, x_t is stationary if and only if $\tilde{x}_t$ is stationary.

Proof: Note that $d > 0$. Let $\mathcal{Z}$ be the set of all z_i that solve the characteristic polynomial

$$|I - \delta_1 A z_i - ... - \delta_p A z_i^p| = |I - (\delta_1 z_i + ... + \delta_p z_i^p)A| = 0 \ .$$

Similarly, let

$$\tilde{\mathcal{Z}} = \{\tilde{z}_i : |I - \tilde{z}_i dA| = 0\} \ .$$

Note that stationarity of x_t is equivalent to the statement that for all $z_i \in \mathcal{Z}$, $|z_i| > 1$, and stationarity of $\tilde{x}_t$ is equivalent to the statement that for all $\tilde{z}_i \in \tilde{\mathcal{Z}}$, $|\tilde{z}_i| > 1$. Therefore, the proof shall show

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

We have

$$\begin{aligned}
& \forall \ \tilde{z}_i \in \tilde{\mathcal{Z}} \ , \quad |\tilde{z}_i| > 1 \\
\Leftrightarrow \ & \forall \ \tilde{z}_i \in \tilde{\mathcal{Z}}, \quad |d\tilde{z}_i| = d|\tilde{z}_i| > d \\
\Leftrightarrow \ & \forall \ z_i \in \mathcal{Z} \ , \quad |\delta_1 z_i + ... + \delta_p z_i^p| > d \\
\Leftrightarrow \ & \forall \ z_i \in \mathcal{Z} \ , \quad |z_i| > 1 \ .
\end{aligned}$$

To show the last (both-sided) 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,

$$|\delta_1 z_i + ... + \delta_p z_i^p| \leq |(\delta_1 + ... + \delta_p) z_i| = |dz_i| = d|z_i| \leq d \ ,$$

a contradiction. Conversely,

$$\begin{aligned}
& \forall \ z_i \in \mathcal{Z} \ , \quad |z_i| > 1 \\
\Rightarrow \ & \forall \ z_i \in \mathcal{Z} \ , \quad |\delta_1 z_i + ... + \delta_p z_i^p| > |(\delta_1 + ... + \delta_p) z_i| = |dz_i| = d|z_i| > d \ . \quad \blacksquare
\end{aligned}$$

Corollary 1 (Stationarity of NVAR($p^*, 1$) With Positive Coefficients)**.**
Let x_t follow an NVAR($p^*, 1$):

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ ,$$

where $v_t \sim WN$, and assume $\delta_l \geq 0 \ \forall \ l = 1 : p$ and $\delta_l > 0$ for at least one l. Define $d = \sum_{l=1}^{p} \delta_l$. Then x_t is stationary if and only if for all Eigenvalues λ_i of A it holds that $|\lambda_i| < 1/d$.

Proposition 4.
Let x_t follow an NVAR($p^*, 1$):

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ ,$$

where $v_t \sim WN$, and assume $\delta_l \neq 0$ for at least one l. Define $d = \sum_{l=1}^{p} |\delta_l|$, and let

$$\tilde{x}_t = dA\tilde{x}_{t-1} + u_t \ .$$

Then, x_t is stationary if $\tilde{x}_t$ is stationary.

Proof: As in the proof of Proposition 3, let

$$\begin{aligned} \mathcal{Z} &= \{z_i : |I - \delta_1 A z_i - ... - \delta_p A z_i^p| = |I - (\delta_1 z_i + ... + \delta_p z_i^p)A| = 0\} \ , \\ \tilde{\mathcal{Z}} &= \{\tilde{z}_i : |I - \tilde{z}_i dA| = 0\} \ . \end{aligned}$$

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

$$\begin{aligned} & \forall \ \tilde{z}_i \in \tilde{\mathcal{Z}} \ , \quad |\tilde{z}_i| > 1 \\ \Leftrightarrow \ & \forall \ \tilde{z}_i \in \tilde{\mathcal{Z}}, \quad |d\tilde{z}_i| = d|\tilde{z}_i| > d \\ \Leftrightarrow \ & \forall \ z_i \in \mathcal{Z} \ , \quad |\delta_1 z_i + ... + \delta_p z_i^p| > d \\ \Rightarrow \ & \forall \ z_i \in \mathcal{Z} \ , \quad |z_i| > 1 \ . \end{aligned}$$

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,

$$|\delta_1 z_i + ... + \delta_p z_i^p| \leq |\delta_1 z_i| + ... + |\delta_p z_i^p| \leq |\delta_1 z_i| + ... + |\delta_p z_i| \leq (|\delta_1| + ... + |\delta_p|)|z_i| = d|z_i| \leq d \ ,$$

a contradiction. ■

Corollary 2 (Stationarity of NVAR($p^*, 1$), Sufficient Condition)**.**
Let x_t follow an NVAR($p^*, 1$):

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ ,$$

where $v_t \sim WN$, and assume $\delta_l \neq 0$ for at least one l. Define $d = \sum_{l=1}^{p} |\delta_l|$. Then x_t is stationary if for all Eigenvalues λ_i of A it holds that $|\lambda_i| < 1/d$.

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

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ ,$$

where $v_t \sim WN$ and $\delta_l \neq 0$ for at least one l. Then, x_t is stationary if and only if the univariate process

$$\check{x}_t = \lambda_i \delta_1 \check{x}_{t-1} + ... + \lambda_i \delta_p \check{x}_{t-p} + \check{v}_t$$

is stationary for all Eigenvalues λ_i of A.

Proof: Stationarity of x_t is equivalent to the statement that for all Eigenvalues l_i of

$$F = \begin{bmatrix} \delta_1 A & \delta_2 A & ... & \delta_{p-1} A & \delta_p A \\ I_n & 0_n & ... & 0_n & 0_n \\ 0_n & I_n & ... & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & ... & I_n & 0_n \end{bmatrix}$$

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

$$\begin{aligned} & |l_i I - F| = 0 \\ \Leftrightarrow \quad & \left| l_i^p I - l_i^{p-1} \delta_1 A - ... - l_i \delta_{p-1} A - \delta_p A \right| = 0 \\ \Leftrightarrow \quad & l_i^{n(p-1)} \left| l_i I - \left(\delta_1 + \delta_2 / l_i + ... + \delta_p / l_i^{p-1} \right) A \right| = 0 \\ \Leftrightarrow \quad & \left(l_i^{p-1} \left(\delta_1 + \delta_2 / l_i + ... + \delta_p / l_i^{p-1} \right) \right)^n \left| \frac{l_i}{\delta_1 + \delta_2 / l_i + ... + \delta_p / l_i^{p-1}} I - A \right| = 0 \\ \Leftrightarrow \quad & \left| \frac{l_i}{\delta_1 + \delta_2 / l_i + ... + \delta_p / l_i^{p-1}} I - A \right| = 0 \ . \end{aligned}$$

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 $l_i / \left(\delta_1 + \delta_2 / l_i + ... + \delta_p / l_i^{p-1} \right)$ is an Eigenvalue of A.

Conversely, given an Eigenvalue λ_i of A, all Eigenvalues l_i that solve

$$l_i^p - l_i^{p-1}\lambda_i\delta_1 - ... - l_i\lambda_i\delta_{p-1} - \lambda_i\delta_p = 0$$

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

Proposition 6 (Stationarity Preservation Under Time-Aggregation).
Let x_τ follow an NVAR($p^*, 1$)

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_p A x_{\tau-p} + v_t \ ,$$

where $v_t \sim WN$. Consider the time series y_t defined by $\{y_t\}_{t=1}^T = \{x_{tq^*}\}_{t=1}^T$ for some $q^* \geq 1, q^* \in \mathbb{N}$ or for some $q^* < 1, 1/q^* \in \mathbb{N}$. Then y_t is stationary if and only if x_τ is stationary.

Proof: Stationarity of x_τ is defined by the two conditions

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

This implies 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.

Conversely, the stationarity of y_t implies that

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

Without further assumptions on the process of x_τ, its mean and variance could be different in observed and non-observed periods, and more generally the covariance could be different over time for displacements other than q^* (i.e. for displacements which are not observed). However, if x_τ follows an NVAR($p^*, 1$), then the fact that first and second moments are the same every q^*th period implies that they are the same in every period τ. ■

Corollary 3 (Stationarity of Time-Aggregated NVAR($p^*, 1$))**.**
Let x_τ follow an NVAR($p^*, 1$)

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_p A x_{\tau-p} + v_t \ ,$$

where $v_t \sim WN$, and assume $\delta_l \neq 0$ for at least one l. Let $d = \sum_{l=1}^{p} |\delta_l|$ and define y_t by $\{y_t\}_{t=1}^{T} = \{x_{tq^*}\}_{t=1}^{T}$ for some $q^* \geq 1, q^* \in \mathbb{N}$ or for some $q^* < 1, 1/q^* \in \mathbb{N}$.
Then, y_t is stationary if and only if the univariate process

$$\check{x}_t = \lambda_i \delta_1 \check{x}_{t-1} + ... + \lambda_i \delta_p \check{x}_{t-p} + \check{v}_t$$

is stationary for all Eigenvalues λ_i of A.
Furthermore, a sufficient condition for stationarity of y_t is that for all Eigenvalues λ_i of A it holds that $|\lambda_i| < 1/d$. If in addition $\delta_1, ..., \delta_p \geq 0$, this condition is both necessary and sufficient.

A.3 Time Aggregation of an NVAR($p^*, 1$)

Let x_τ follow an NVAR($p^*, 1$):

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_{p^*} A x_{\tau-p^*} + v_\tau \ .$$

If we observe a snapshot of x_τ every $q^* > 1$ periods, the observed process $\{y_t\}_{t=1}^{T} = \{x_{q^*t}\}_{t=1}^{T}$ can be approximated by an NVAR(p, q) with $\alpha = \{\alpha_{lg}\}_{l=1:p, g=1:q}$ being a function of $\delta = \{\delta_l\}_{l=1:p^*}$ and with errors following an MA with $p-1$ lags:

$$\begin{aligned}
y_t &= \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + ... + \Theta_{p-1} u_{t-p+1} \ , \\
\Phi_l(\delta) &= \sum_{g=1}^{q} \alpha_{lg}(\delta) A^g \ , \\
\Theta_0(\delta) &= \left[I_n, \sum_{g=1}^{q} \beta_{0g}^{2}(\delta) A^g, ..., \sum_{g=1}^{q} \beta_{q^*g}^{q^*}(\delta) A^g \right] \ , \\
\Theta_l(\delta) &= \left[0_n, \sum_{g=1}^{q} \beta_{lg}^{lq^*+2}(\delta) A^g, ..., \sum_{g=1}^{q} \beta_{lg}^{(l+1)q^*}(\delta) A^g \right] \ , \quad l = 1 : p-1 \ ,
\end{aligned}$$

where $u_t = [v'_\tau, v'_{\tau-1}, ..., v'_{\tau-q^*+1}]' = [v'_{q^*t}, v'_{q^*t-1}, ..., v'_{q^*t-q^*+1}]'$. The approximation error involves terms in $x_{\tau-pq^*-1}, ..., x_{\tau-pq^*-p}$, which under stationarity vanish for p large enough. In the following I show how to obtain $\{\Phi_l(\delta)\}_{l=1}^{p}$ and $\{\Theta_l(\delta)\}_{l=0}^{p-1}$.

Let v_h be the set which collects the $n \times n$ matrices in front of $\{x_{\tau-1}, x_{\tau-2}, ...\}$ in the linear projection of x_τ on the information set at $t-h$, $\mathcal{F}_{\tau-h} = \{x_{\tau-h-j}, v_{\tau-h-j}\}_{j=0}^{\infty}$, which

I denote as $x_{\tau|\tau-h} \equiv x_\tau|\ \mathcal{F}_{\tau-h}$. For example, $v_1 = \{\delta_1 A, \delta_2 A, ..., \delta_p A, 0_n, 0_n, ...\}$. Let w_h collect the $n \times n$ matrices in front of $\{v_\tau, v_{\tau-1}, ...\}$ in the forecasting error at horizon h, $v_{\tau|\tau-h} \equiv x_\tau - x_{\tau|\tau-h}$. For $h = 1$, we have $w_1 = \{I, 0_n, 0_n, ...\}$. The procedure of iteratively inserting for $x_{\tau-l}$ for all $l = 1, 2, ..., h-1$ would decompose x_τ as $x_\tau = x_{\tau|\tau-h} + u_{\tau|\tau-h}$, i.e. it would lead to v_h and w_h. Doing so for all $l = 1, 2, ...$ except multiples of q^*, we get the information necessary to construct the equation for y_t, which includes the linear projection of y_t on the infinite past, $y_{t|t-1} \equiv y_t|\ \mathcal{G}_{t-1}$, $\mathcal{G}_{t-1} = \{y_{t-1-j}, u_{t-1-j}\}_{j=0}^{\infty}$, as well as the one-step ahead forecasting error $u_{t|t-1} \equiv y_t - y_{t|t-1} = \Theta_0 u_t$.

Algorithm 1 formalizes this iterative procedure performed in the time aggregation of x_τ. Iteration h inserts the equation $\delta_1 A x_{\tau-h} + ... + \delta_p A x_{\tau-h-p+1} + u_{\tau-h-p+1}$ to eliminate $x_{\tau-h+1}$ in case h is not a multiple q^* and leaves the equation unchanged if h is a multiple of q^*, in which case $x_{\tau-h}$ is observed.[43] Note that $v_1[h : h+p-1] = \{\delta_1 A, ..., \delta_p A\}$ and $w_1[1] = I$ yield the matrices in the original equation for $x_\tau = x_{\tau|\tau-1} + v_{\tau|\tau-1}$.

Algorithm 1 (Time Aggregation: Obtaining $\{\Phi_l(\delta)\}_{l=1}^{p}$ and $\{\Theta_l(\delta)\}_{l=0}^{p-1}$)**.**

1. *Let*

$$v_1 = \{\delta_1 A, \delta_2 A, ..., \delta_p A, 0_n, 0_n, ...\}\ ,$$
$$w_1 = \{I_n, 0_n, 0_n, ...\}\ \ .$$

2. *For $h = 2 : pq^*$, given v_{h-1} and w_{h-1}, obtain v_h and w_h as follows:*

   a) *If h is a multiple of q^*,*

$$v_h = v_{h-1}\ , \quad w_h = w_{h-1}\ .$$

   b) *If h is not a multiple of q^*,*

$$v_h[l] = \begin{cases} v_{h-1}[l] & \textit{for } l = 1 : h-2 \\ 0_n & \textit{for } l = h-1 \\ v_{h-1}[h-1]v_1[l-h+1] + v_{h-1}[l] & \textit{for } l = h : h+p-1 \\ 0_n & \textit{for } l \geq h+p \end{cases}\ ,$$

$$w_h[l] = \begin{cases} w_{h-1}[l] & \textit{for } l = 1 : h-1 \\ v_{h-1}[h-1]w_1[1] & \textit{for } l = h \\ 0_n & \textit{for } l \geq h+1 \end{cases}\ .$$

[43] This holds if we define τ and t such that $x_\tau = y_t$ is observed.

3. *The $n \times n$ matrices in front of $y_{t-1}, ..., y_{t-p}$ in $y_{t|t-1}$ are the elements $q^*, 2q^*, ..., pq^*$ in v_{pq^*}:*

$$\Phi_l = v_{pq^*}[lq^*] \ , \quad l = 1 : p \ .$$

The $n \times n$ matrices which make up the $n \times nq^*$ matrices in front of $u_{t-1}, ..., u_{t-p+1}$ in $y_{t|t-1}$ as well as the matrix in front of $u_t = u_{t|t-1}$ are contained in w_{pq^*}:

$$\Theta_0 = \Big[w_{pq^*}[1], ..., w_{pq^*}[q^*] \Big] \ ,$$
$$\Theta_l = \Big[0_n, w_{pq^*}[lq^* + 2], ..., w_{pq^*}[(l+1)q^*] \Big] \ , \ \ l = 2 : p - 1 \ .$$

Note that the linear projection of y_t on $\mathcal{G}_{t-1}$ involves infinitely many lags in y_t and in u_t. Nevertheless, for a given cutoff lag p, it suffices to carry out the iterative procedure for $h = 2, 3, ..., pq^*$ as v_{pq^*} already contains the matrices in front of $y_{t-1}, y_{t-2}, ..., y_{t-p}$ and w_{pq^*} contains the matrices in front of $u_t, u_{t-1}, ..., u_{t-p+1}$.

Also, note that in practice, because each element in v_h and w_h is a polynomial in A, rather than carrying out the costly multiplications of these polynomials in each iteration, I keep track of the coefficients obtained when multiplying two polynomials and only in the end I construct the polynomials of interest in the relevant positions of v_{pq^*} and w_{pq^*}. These coefficients are products of elements in $\delta = \{\delta_l\}_{l=1}^{p^*}$.

Finally, note that for an approximation order p chosen by the researcher, one obtains an NVAR(p, q) with $q = pq^* - (p - 1)$. To see this, note that by proposition 1, the largest power of A in $\frac{\partial x_\tau}{\partial x_{\tau-h}} | \ \mathcal{F}_{\tau-h}$ is h. As mentioned above, if the iterative procedure in step 2b of Algorithm 1 was carried out for all h including multiples of q^*, v_h would contain the matrices in front of $x_{\tau-h}, x_{\tau-h-1}, ...$ in the projection of x_τ on $\mathcal{F}_{\tau-h}$, i.e. $v_h[h] = \frac{\partial x_\tau}{\partial x_{\tau-h}} | \ \mathcal{F}_{\tau-h}$. By Proposition 1 we know the highest power of A in $v_{lq*}[lq*] = \frac{\partial x_\tau}{\partial x_{\tau-lq*}} | \ \mathcal{F}_{\tau-lq*}$ would be lq^*. As opposed to this case, the time aggregation performed above skips every q^*th iteration, which means that $\Phi_l = v_{lq*}[lq*]$ has the largest power in A equal to $lq^* - (l-1)$. It is largest for $l = p$.

A.4 Impulse Responses

Proposition 7 (Long-Term Response in NVAR($p^*, 1$)).
Let x_t follow an NVAR($p^*, 1$):

$$x_t = \delta_1 A x_{t-1} + ... + \delta_p A x_{t-p} + v_t \ .$$

Define $d = \sum_{l=1}^{p} \delta_l$, and let

$$\tilde{x}_t = dA\tilde{x}_{t-1} + u_t \ ,$$
$$\hat{x}_t = dA\hat{x}_t + w_t \ .$$

Assume x_t and $\tilde{x}_t$ are stationary. Then, the long-term response of x_t to a permanent increase in v_t is equivalent to the long-term response of $\tilde{x}_t$ to a permanent increase in u_t, i.e.

$$\tilde{R} \equiv lim_{h\to\infty} \left[\frac{\partial \tilde{x}_{t+h}}{\partial u_t} + \frac{\partial \tilde{x}_{t+h}}{\partial u_{t+1}} + ... + \frac{\partial \tilde{x}_{t+h}}{\partial u_{t+h}} \right] = lim_{h\to\infty} \left[\frac{\partial x_{t+h}}{\partial v_t} + \frac{\partial x_{t+h}}{\partial v_{t+1}} + ... + \frac{\partial x_{t+h}}{\partial v_{t+h}} \right] \equiv R \ .$$

Both, in turn, are equivalent to the contemporaneous reponse of $\hat{x}_t$ to a disturbance in w_t, $\partial\hat{x}_t/\partial w_t$.

Proof: First of all, note that

$$\hat{x}_t = (I - dA)^{-1} w_t \ ,$$

which means that $\partial\hat{x}_t/\partial w_t = (I - dA)^{-1}$.

Turning to x_t and $\tilde{x}_t$, note that under stationarity

$$R = \lim_{h\to\infty} \sum_{j=0}^{h+1} \frac{\partial x_{t+h}}{\partial v_{t+h-j}} = \lim_{h\to\infty} \sum_{j=0}^{h+1} \frac{\partial x_{t+j}}{\partial v_t} = \sum_{j=0}^{\infty} \frac{\partial x_{t+j}}{\partial v_t} \ ,$$

and analogously for $\tilde{R}$. It is easy to see that

$$\tilde{R} = \sum_{j=0}^{\infty} \frac{\partial \tilde{x}_{t+j}}{\partial u_t} = \sum_{j=0}^{\infty} (dA)^j = (I - dA)^{-1} \ .$$

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

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

where $z_t = [x'_t, x'_{t-1}, ..., x'_{t-p+1}]'$ and $e_t = [v'_t, 0', ..., 0']'$ are np-dimensional vectors, and the

$n \times n$ matrix F is defined as

$$F = \begin{bmatrix} \delta_1 A & \delta_2 A & \dots & \delta_{p-1} A & \delta_p A \\ I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & I_n & \dots & 0_n & 0_n \\ \vdots & & \ddots & & \vdots \\ 0_n & 0_n & \dots & I_n & 0_n \end{bmatrix} .$$

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 $(F^h)_{11}$:

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

Note that

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

Therefore,

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

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

$$\begin{aligned} I &= M(I - F) \\ &= \begin{bmatrix} M_{11} & M_{12} & M_{13} & \dots & M_{1,p-1} & M_{1p} \\ M_{21} & M_{22} & M_{23} & \dots & M_{2,p-1} & M_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{p1} & M_{p2} & M_{p3} & \dots & M_{p,p-1} & M_{pp} \end{bmatrix} \begin{bmatrix} I - \delta_1 A & -\delta_2 A & -\delta_3 A & \dots & -\delta_{p-1} A & -\delta_p A \\ -I_n & I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & -I_n & I_n & \dots & 0_n & 0_n \\ \vdots & & \ddots & \ddots & & \vdots \\ 0_n & 0_n & \dots & -I_n & I_n & 0_n \\ 0_n & 0_n & \dots & 0_n & -I_n & I_n \end{bmatrix} \end{aligned}$$

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

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

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

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

which implies

$$M_{12} = M_{11}\delta_2 A + M_{13} = M_{11}\delta_2 A + M_{11}\delta_3 A + M_{14} = ... = M_{11}(\delta_2 + ... + \delta_p)A \ .$$

The first element gives

$$I_n = M_{11}(I - \delta_1 A) - M_{12} = M_{11}\left(I - (\delta_1 + \delta_2 + ... + \delta_p)A\right) = M_{11}(I - dA) \ ,$$

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

Note that the fact that x_t may not be observed every period does not change its long-term response to a permanent increase in v_t (see below).

Impulse Responses in an NVAR(p, q)

Let y_t follow an NVAR(p, q). Its IRF can be computed using its companion form, as outlined in the proof of Proposition 1. The latter shows that the impulse response at any horizon h is a polynomial in A with powers ranging from $\underline{k} = ceil(h/p)$ to hq. Overall,

$$\frac{\partial y_{t+h}}{\partial u_t} = (F^h)_{11} = c_{\underline{k}}(\alpha)\left[A^{\underline{k}}\right]_{ij} + ... + c_{hq}(\alpha)\left[A^{hq}\right]_{ij} \ .$$

The coefficients $c_k(\delta)$ can be obtained using Algorithm 1, ignoring step 2a) (see discussion in Section A.3).

Impulse Responses Under Time-Aggregation of an NVAR($p^*, 1$)

Let x_τ follow an NVAR($p^*, 1$) and denote the observed series by $\{y_t\}_{t=1}^T = \{x_{q^* t}\}_{t=1}^T$. As discussed in Section 2.4, if $q^* \leq 1, 1/q^* \in \mathbb{N}$, y_t follows an NVAR($p, 1$) with $p = p^*/q^*$, which can be characterized analytically. Impulse responses are computed as outlined above. If instead $q^* > 1, q^* \in \mathbb{N}$, y_t follows an NVAR(p, q) with restricted parameters and MA-errors. In this case, the error u_t in the observed process y_t contains all high-frequency errors v_τ which occurred between the periods of observation $t-1$ and t. As a result, the impulse responses for y_t can be computed not only over different horizons of observation h, but also for disturbances that occurred at different high-frequency periods within the same horizon of observation h. In other words, the usual impulse response, $\partial y_{t+h}/\partial u_t$, is composed of

$$\frac{\partial y_{t+h}}{\partial u_t} = \left[\frac{\partial y_{t+h}}{\partial v_{tq^*}}, \frac{\partial y_{t+h}}{\partial v_{tq^*-1}}, ..., \frac{\partial y_{t+h}}{\partial v_{tq^*-q^*+1}}\right] \ .$$

It holds that

$$\frac{\partial y_{t+h}}{\partial v_{tq^*-l}} = \frac{\partial x_{(t+h)q^*}}{\partial v_{tq^*-l}} = \frac{\partial x_{tq^*+hq^*+l}}{\partial v_{tq^*}} = \frac{\partial x_{\tau+hq^*+l}}{\partial v_{\tau}} \ .$$

Therefore, we obtain

$$\frac{\partial y_{t+h}}{\partial v_{tq^*-l}} = \frac{\partial x_{\tau+h^*}}{\partial v_{\tau}} = (F_x^{hq^*+l})_{11} \ .$$

In practice, we can compute $(F_x^{h^*})_{11}$ for $h^* = 0, 1, 2, ...$, and given $(F_x^{h^*})_{11}$, obtain $\partial y_{t+h}/\partial v_{tq^*-l} = (F^{h^*})_{11}$ for $h = floor(h^*/q^*)$, $l = h^* - hq^*$.

By the same token, long-term responses of y_t to permanent disturbances to v_τ, starting from period $\tau = tq^* - l$, are given by

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

B Input-Output Network and Price Dynamics

B.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: $x_{ijt} = x^{ij}_{t,t} = x^{ij}_t$. 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

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

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.

The representative household's problem is

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

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:

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

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

In this case, the amount of good j purchased at time $t-1$ is used in production at time t: $x_{ijt} = x^{ij}_{t,t-1} = x^{ij}_{t-1}$. I will write x^{ij}_{t-1} for this quantity. Firm i's problem is

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

This leads to the following value function:

$$V_i\left(\left\{x^{ij}_{t-1}\right\}_{j=1}^n\right) = \max_{l_{it}, \{x^{ij}_t\}_{j=1}^n} p_{it} z_{it} l_{it}^{b_i} \prod_{j=1}^{n} \left(x^{ij}_{t-1}\right)^{a_{ij}} - w_t l_{it} - \sum_{j=1}^{n} p_{jt} x^{ij}_{t-1} + \beta V_i\left(\left\{x^{ij}_{t}\right\}_{j=1}^n\right) \ .$$

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

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

Note that in steady state, the latter expression yields $a_{ij} = \beta^{-1}(p_j x^{ij})/(p_i y_i)$, 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 $k_{it}^{p1} = -\left[b_i ln(b_i) + \sum_{j=1}^{n} a_{ij} ln(\beta a_{ij}) + (1-b_i) ln(G_t^w)\right]$ 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 $c_{it} = \gamma_i w_t / p_{it}$. 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

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

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

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 $\beta \to 1$.

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

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

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 $r \to 0$, 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

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

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

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

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

For output we have, as before, $ln\left(y_i\right) = ln\left(\lambda_i\right) - ln\left(p_i/w\right)$. 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 $\beta \to 1$.

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

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

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

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

where

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

and

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

B.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.

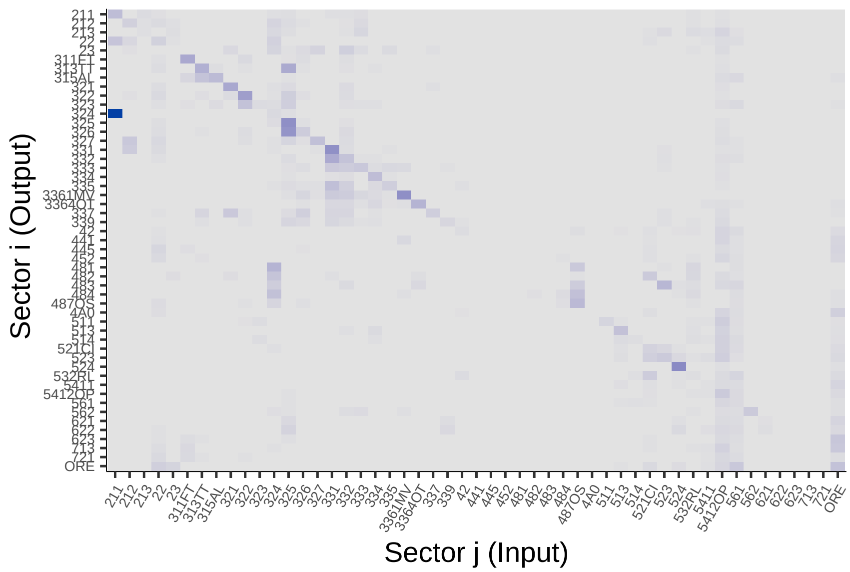

Figure A-1: Tabular Representation of the Input-Output Matrix

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

Table A-1: Quarterly Correlation at Different Months

	Q1	Q2	Q3	Q4
Month 1	0.9697	0.9723	0.9836	0.9737
Month 2	0.9670	0.9642	0.9847	0.9694
Month 3	0.9675	0.9709	0.9793	0.9695

Notes: The table shows the cross-sectional mean of the three-month-lagged autocorrelation of log PPI computed separately for each month of the year. No difference between end-of-quarter months and remaining months is noticeable.

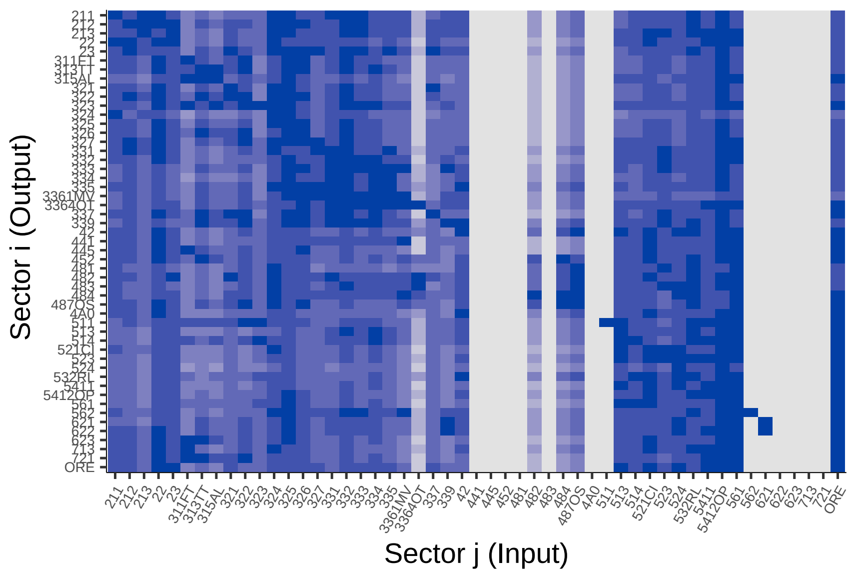

Figure A-2: Tabular Representation of Shortest Paths

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

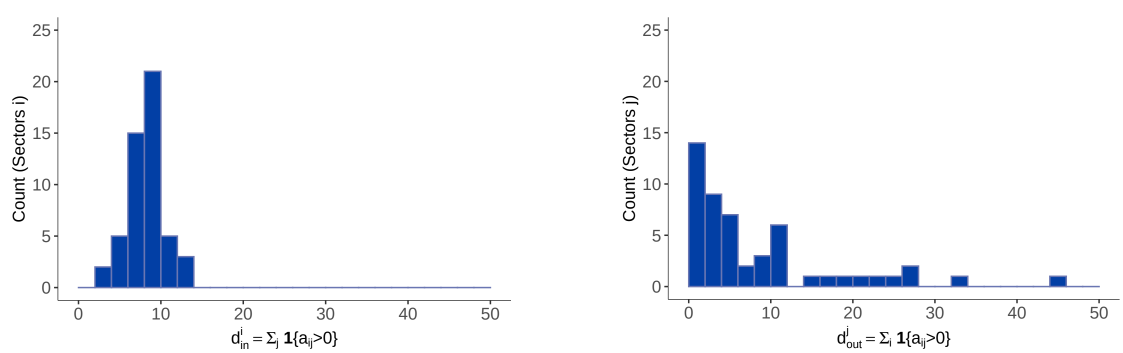

Figure A-3: In-Degrees and Out-Degrees

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-2: 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

B.3 Estimation of the Network-VAR Given Network Data

Estimation of $\alpha|A$ in Unrestricted NVAR(p, q)

This section outlines the estimation of $(\alpha, \Sigma)|A$ in the general NVAR(p, q),

$$y_t = \sum_{l=1}^{p} \left(\sum_{g=1}^{q} \alpha_{lg} A^g \right) y_{t-l} + u_t \ , \quad \mathbb{V}[u_t] = \Sigma.$$

This model can alternatively be written as

$$y_{it} = x_{it}' \alpha + u_{it} \ ,$$

where $\alpha = (\alpha_1', ..., \alpha_p')$, $\alpha_l = (\alpha_{l1}, ..., \alpha_{lq})'$ and $x_{it} = (x_{it}^{1'}, ..., x_{it}^{p'})'$, $x_{it}^l = (\tilde{y}_{it-l}^1, ..., \tilde{y}_{t-l}^q)'$, and where $\tilde{y}_{it-l}^g$ is the ith element of the $n \times 1$ vector $\tilde{y}_{t-l}^g \equiv A^g y_{t-l}$. Because the network A is taken as given, the dependence of $\tilde{y}_{t-l}^g$ on A is suppressed. This is a linear regression with regressors generated by summarizing information in lagged values of y_t using the network adjacency matrix A. In matrix form,

$$y_t = X_t \alpha + u_t \ . \tag{A.1}$$

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

$$X_t = \begin{bmatrix} x_{1t}' \\ \cdots \\ x_{nt}' \end{bmatrix} = \left[A y_{t-1}, A^2 y_{t-1}, ..., A^q y_{t-1}, A y_{t-2}, ..., A^q y_{t-p} \right] .$$

Least Squares (LS) estimation of α with a Ridge-penalty to potentially induce shrinkage yields the following optimization problem:

$$\underset{\alpha}{min} \ \frac{1}{nT} \sum_{t=1}^{T} (y_t - X_t \alpha)' \, \Sigma^{-1} \, (y_t - X_t \alpha) + \tilde{\varphi} \sum_{l=1}^{p} \sum_{g=1}^{q} \alpha_{lg}^2 \ .$$

In turn, $\hat{\Sigma}|\alpha = \frac{1}{T} \sum_{t=1}^{T} u_t u_t'$. We obtain

$$\hat{\alpha}_{LS}|\Sigma = \bar{V}_\alpha \left[\sum_{t=1}^{T} X_t' \Sigma^{-1} y_t \right] \ , \bar{V}_\alpha = \left[\tilde{\varphi} I + \sum_{t=1}^{T} X_t' \Sigma^{-1} X_t \right]^{-1} \ .$$

Note that under $\tilde{\varphi} = 0$ and $\Sigma = I$, the estimator for α takes the form of a pooled OLS estimator, as expected based on the model representation in equation A.1:

$$\hat{\alpha}_{LS}|(\Sigma = I, \tilde{\varphi} = 0) = \left(\sum_{t=1}^{T} X_t' X_t\right)^{-1} \sum_{t=1}^{T} X_t' \tilde{y}_t = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} x_{it}'\right)^{-1} \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} y_{it}\right) .$$

In Appendix C.2, I derive $\hat{\alpha}_{LS}|\Sigma$ as the posterior mode under the prior $\alpha \sim N(0, \varphi^{-1} I)$ with $\varphi = nT\tilde{\varphi}$ and Normality of u_t. With the conditional posteriors available, Bayesian estimation of $\alpha, \Sigma|A$ can be implemented straightforwardly using Gibbs sampling.

If α is estimated without shrinkage ($\tilde{\varphi} = 0$), we require $\sum_{t=1}^{T} X_t' X_t$ to be of full rank:

$$\lambda_{min}\left(\mathbb{E}\left[X_t' X_t\right]\right) \geq c > 0 , \quad \text{for some } c \in \mathbb{R}_+ ,$$

where $\lambda_{min}(M)$ denotes the smallest Eigenvalue of the matrix M. Due to the stochastic process of innovations u_t, we expect $y_t, ..., y_{t-p}$ to be linearly independent. Given the structure of X_t, this means that the above requirement demands q to be such that the matrices $A, A^2, ..., A^q$ are linearly independent. The Cayley-Hamilton theorem gives the upper bound $q \leq n-1$. It states that for any $n \times n$ matrix A, A^n can be expressed as the linear combination of lower powers of A. In practice, A could be such that even for small q one runs into issues of almost perfect multicollinearity in the estimation of α, which provides a rationale for applying shrinkage by setting $\tilde{\varphi} > 0$.

Estimation of $\delta|A$ in (Potentially) Time-Aggregated NVAR($p^*, 1$)

This section is concerned with estimating $(\delta, \Sigma)|A$, where δ is the vector of coefficients in the NVAR($p^*, 1$) for x_τ, whereby a snapshot of x_τ is observed every q^* periods: the observed process is $\{y_t\}_{t=1}^T = \{x_{q^* t}\}_{t=1}^T$. First I discuss how to obtain the likelihood of $\{y_t\}_{t=1}^T$ for general q^*. Then I discuss the Bayesian implementation of the MLE used to obtain the results in Section 3.4.

Likelihood Evaluation The process for x_τ is

$$x_\tau = \delta_1 A x_{\tau-1} + ... + \delta_{p^*} A x_{\tau-p^*} + v_\tau .$$

The observed process is $\{y_t\}_{t=1}^T = \{x_{q^* t}\}_{t=1}^T$. If $q^* \leq 1$, $1/q^* \in \mathbb{N}$, y_t follows an NVAR($p, 1$) with $p = p^*/q^*$. In this case, there are analytical expressions for the LS objective function and likelihood.

If instead $q^* > 1$, $q^* \in \mathbb{N}$, the process for y_t can be approximated arbitrarily well by an NVAR(p, q) with errors following an MA($p-1$):

$$y_t = \Phi_1 y_{t-1} + ... + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + ... + \Theta_{p-1} u_{t-p+1} .$$

where $u_t = [v'_\tau, v'_{\tau-1}, ..., v'_{\tau-q^*+1}]' = [v'_{q^*t}, v'_{q^*t-1}, ..., v'_{q^*t-q^*+1}]'$. The dependence of $\{\Phi_l(\delta)\}_{l=1}^p$ and $\{\Theta_l(\delta)\}_{l=0}^{p-1}$ on $\delta = (\delta_1, ..., \delta_{p^*})$ is suppressed for notational simplicity. Appendix A.3 outlines how to obtain these matrices.

Assume $v_\tau \sim N(0, \Sigma)$ and let $z_t = (z'_{1t},\ z'_{2t}, ..., z'_{pt})'$. The process for $\{y_t\}_{t=1}^T$ can be cast in state space form as

$$\begin{aligned} z_t &= F z_{t-1} + T u_t \ , \quad u_t \sim N(0, Q) \ , \\ y_t &= M z_t = z_{1t} \ , \end{aligned}$$

where $M = [I_n, 0_n, ..., 0_n]$ and

$$F = \begin{bmatrix} \Phi_1 & I_n & 0_n & \dots & 0_n \\ \Phi_2 & 0_n & I_n & & \\ \vdots & \vdots & & \ddots & \\ \Phi_{p-1} & 0_n & & & I_n \\ \Phi_p & 0_n & \dots & & 0_n \end{bmatrix} , \quad T = \begin{bmatrix} \Theta_0 \\ \Theta_1 \\ \vdots \\ \Theta_{p-1} \end{bmatrix} , \quad Q = \begin{bmatrix} \Sigma & 0_n & \dots & 0_n \\ 0_n & \Sigma & & \\ \vdots & & \ddots & \\ 0_n & & & \Sigma \end{bmatrix} .$$

The likelihood can be evaluated using the Kalman filter. To obtain the conditional likelihood, the filter is initialized as follows. Partition F into p blocks of dimension $n \times n$. Let $F_{-1,\cdot}$ be the $n(p-1) \times np$ matrix obtained by deleting the first n rows of F, let $F_{-1,1}$ be the first n columns of it, and let $F_{-1,-1}$ be the $n(p-1) \times n(p-1)$ matrix obtained by deleting the first n rows as well as the first n columns of F. Correspondingly, define $z_{-1,t} = (z'_{2t}, ..., z'_{pt})'$ so that $z_t = (z'_{1,t}, z'_{-1,t})'$. The object of interest is the distribution

$$z_0 | y_0, y_{-1}, ..., y_{-p+1} \equiv z_0 | y_{0:-p+1} \sim N(\mu_z, \Sigma_z) \ .$$

We know $\mathbb{E}[z_{1,0} | y_{0:-p+1}] = y_0$. For the remaining part of μ_z, we get

$$\mathbb{E}[z_{-1,0} | y_{0:-p+1}] = F_{-1,\cdot} \begin{bmatrix} y_{-1} \\ \mathbb{E}[z_{-1,-1} | y_{0:-p+1}] \end{bmatrix} = F_{-1,1} y_{-1} + F_{-1,-1} \mathbb{E}[z_{-1,-1} | y_{0:-p+1}] \ .$$

We can iterate on this expression starting from $\mathbb{E}[z_{-1,-p+1} | y_{0:-p+1}]$ on the right-hand side. Note that because of the nature of $F_{-1,-1}$, this gives an exact expression for $\mathbb{E}[z_{-1,0} | y_{0:-p+1}]$ as a function of $y_{-1}, ..., y_{-p+1}$.

Turning to the variance, y_0 is known and therefore the first n rows and the first n columns of Σ_z are zero. For the remaining part, we get

$$\mathbb{V}[z_{-1,0} | y_{0:-p+1}] = T_{-1} Q T'_{-1} + F_{-1,-1} \mathbb{V}[z_{-1,-1} | y_{0:-p+1}] F'_{-1,-1} \ ,$$

which can again be iterated on starting from $\mathbb{V}[z_{-1,-p+1} | y_{0:-p+1}]$ on the right-hand side.

Bayesian Implementation of MLE For reasons outlined in the main text, I obtain the MLE for the application in Section 3 as the posterior mode under a prior for θ which is proportional to a constant. For δ, the domain restrictions $\delta_l \in [0,1]$ and $\sum_{l=1}^{p^*-1} \delta_l \leq 1$ lend themselves into a prior distribution which is the product of independent uniform distributions, truncated to the region where $\sum_{l=1}^{p^*-1} \delta_l \leq 1$:

$$\delta_1, ..., \delta_{p^*-1} \sim \prod_{l=1}^{p^*-1} \mathcal{U}(0,1) \, \mathbf{1}\left\{ \sum_{l=1}^{p^*-1} \delta_l \leq 1 \right\} .$$

As one can verify, this leads to

$$\begin{aligned} p(\delta_1, ..., \delta_{p^*-1}) &= p(\delta_1|\delta_2, ..., \delta_{p^*-1})p(\delta_2|\delta_3, ..., \delta_{p^*-1})...p(\delta_{p^*-2}|\delta_{p^*-1})p(\delta_{p^*-1}) \\ &= \begin{cases} (p^*-1)! & \text{if } \sum_{l=1}^{p^*-1} \delta_l \leq 1 \\ 0 & \text{otherwise} \end{cases} , \end{aligned}$$

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

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

and

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

To draw from $p(\delta_1, ..., \delta_{p^*-1})$, one can draw δ_{p^*-1} from its marginal distribution and iteratively draw $\delta_{p^*-2}, ..., \delta_1$ 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

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

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

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

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

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

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

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 δ. To do so, I obtain a consistent estimator for $\sigma|\delta$ using the method of moments applied to the variance of the high-frequency process x_τ.[44] We know that $\mathbb{V}[y_t] = \mathbb{V}[x_{tq^*}]$. The latter can be computed as a function of Σ:

$$\mathbb{V}[x_\tau] = \Phi_1 \mathbb{V}[x_\tau]\Phi_1' + ... + \Phi_{p^*}\mathbb{V}[x_\tau]\Phi_{p^*}' + \Sigma \ ,$$

[44]As opposed to that, using the process for y_t, the conditional MLE for $\sigma|\delta$ 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

$$\begin{aligned} vec(\mathbb{V}[x_\tau]) &= \left(\sum_{l=1}^{p^*} \Phi_l \otimes \Phi_l\right) vec(\mathbb{V}[x_\tau]) + vec(\Sigma) \\ &= \left(\sum_{l=1}^{p^*} \delta_l^2\right) (A \otimes A)\, vec(\mathbb{V}[x_\tau]) + vec(\Sigma) \;. \end{aligned}$$

Overall, we get

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

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

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 δ, I consider a transformation of the parameters in the mutation step of the SMC algorithm. Define the function g s.t. $\breve{\theta} = g^{-1}(\theta)$ is generated by taking logs of σ_i and computing $\gamma_l = ln\ \delta_l/\delta_{p^*}$ for $\delta_1, ..., \delta_{p^*-1}$. 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 2 (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* $\breve{\theta}_n^{i,m-1} = g^{-1}(\theta_n^{i,m-1})$ *and draw*

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

- *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

$$J_{11}(\theta) = \begin{bmatrix} \delta_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \delta_{p^*-1}^{-1} \end{bmatrix} + \delta_{p^*}^{-1}\iota\iota' , \quad J_{22}(\theta) = \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \sigma_N^{-1} \end{bmatrix} .^{45}$$

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

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

B.4 Results

[45] 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-3: 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.

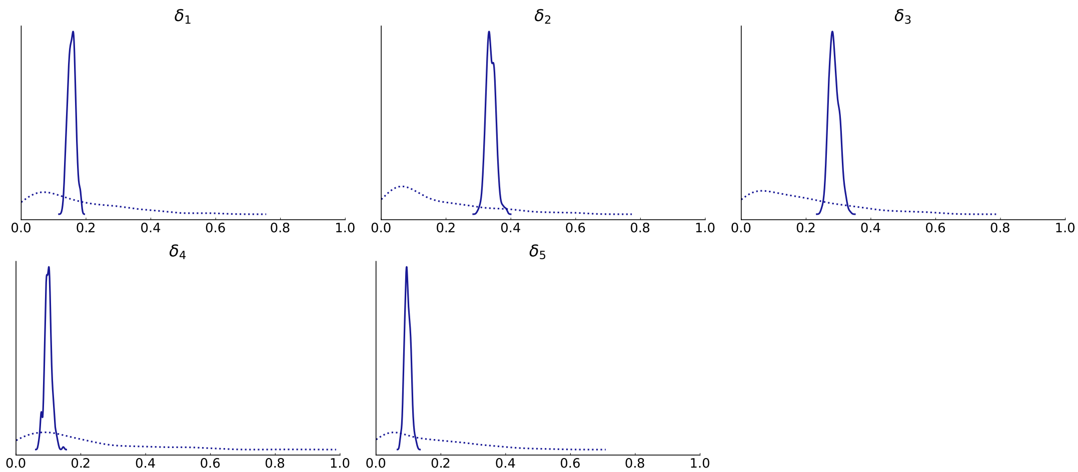

Figure A-4: Marginal Priors and Posteriors For δ

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

C Process Approximation by Sparse Networks

C.1 Identification and Normalization

Without further assumptions, (α, A) is not identified. For simplicity, consider the case of $p = 1$ and $q = 1$. Then

$$y_t = \alpha_{11} A y_{t-1} + u_t \ ,$$

and for any $(\hat{\alpha}_{11}, \hat{A})$, an observationally equivalent process that satisfies the same domain restriction for the elements of A is obtained with $(c^{-1}\hat{\alpha}_{11}, c\hat{A})$ for any $c \in (0, \bar{c})$ with $\bar{c} = [sup_{i,j}\ \hat{a}_{ij} + \varepsilon]^{-1}$ for $\varepsilon > 0$ small. Generally, for any $(\hat{\alpha}, \hat{A})$, an observationally equivalent process is obtained with $(\tilde{\alpha}, \tilde{A})$ defined by

$$\tilde{A} = c\hat{A} \ , \quad \tilde{\alpha}_{lg} = c^{-g}\hat{\alpha}_{lg} \quad \text{for } l = 1:p, \ g = 1:q \ ,$$

where as before $c \in (0, \bar{c})$ with $\bar{c} = [sup_{i,j}\ \hat{a}_{ij} + \varepsilon]^{-1}$ for $\varepsilon > 0$ small.

To identify (α, A), at least one element in α or A needs to be normalized. For example, consider fixing $\alpha_{11} = 1$. Because the elements in A are restricted to be positive, this means one requires $\alpha_{11} > 0$ to hold in the true data generating process. Alternative normalizations are possible. To estimate the NVAR$(p, 1)$ in Section 4.4, I normalize $||\alpha||_1 = 1$. One could also normalize one of the elements in A to some positive number, reasoning that two units (i, j) ought to share a direct and directed link.

One related subtlety arises. For many normalizations, the constraint $a_{ij} \in [0, 1)$ becomes potentially too restrictive. For example, under $\alpha_{11} = 1$ this is because some elements in the matrix $\alpha_{11}A$ might be larger than 1. To deal with this, one can first estimate the model without an upper bound for a_{ij}. If needed, the parameters can then be re-scaled along the lines outlined above in order to preserve the network interpretation. More specifically, let $(\breve{\alpha}, \breve{A})$ be the solution for the problem with restrictions $a_{ij} \geq 0$ and $\alpha_{11} = 1$ fixed. Given $(\breve{\alpha}, \breve{A})$, define $(\hat{\alpha}, \hat{A})$ as

$$\hat{\alpha}_{11} = \underset{i,j}{sup}\, \breve{a}_{ij} + \varepsilon \quad , \quad \hat{A} = \frac{1}{\hat{\alpha}_{11}} \breve{A} \quad , \quad \hat{\alpha}_{lg} = \hat{\alpha}_{11}^{g} \breve{\alpha}_{lg} \text{ for } l = 1:p,\; g = 1:q \;,\; (l,g) \neq (1,1)$$

for some $\varepsilon > 0$ small.

C.2 Estimation of the Network-VAR

Least Squares Estimator as Posterior Mode

Estimating (α, A, Σ) in the general NVAR(p, q) using (Generalized) Least Squares (LS) with a Lasso penalty term to induce sparsity in A and a Ridge penalty term to induce shrinkage in α amounts to the following optimization problem:

$$\underset{\alpha,A}{min} \frac{1}{nT} \sum_{t=1}^{T} u_t(\alpha, A)' \Sigma^{-1} u_t(\alpha, A) + \tilde{\lambda} \sum_{i,j=1}^{n} a_{ij} + \tilde{\varphi} \sum_{l,g} \alpha_{lg}^2 \quad \text{s.t. } a_{ij} \geq 0 \;, \tag{A.2}$$

with

$$u_t(\alpha, A) = y_t - \sum_{l=1}^{p} \left(\sum_{g=1}^{q} \alpha_{lg} A^g \right) y_{t-l} \;,$$

and with (at least) one element in or statistic of α or A normalized (see Appendix C.1). In turn, $\hat{\Sigma}_{LS} | \alpha, A = \frac{1}{T} \sum_{t=1}^{T} u_t(\alpha, A) u_t(\alpha, A)'$.

I suppress the dependence of u_t on (α, A) for notational simplicity. Under Normality of u_t, the (conditional) likelihood function associated with this problem is

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

Consider the prior

$$\alpha \sim N(0, \varphi^{-1} I) \ , \quad a_{ij} \sim \text{Exponential}(\lambda) \ , \quad p(\Sigma) \propto c \ .$$

The posterior of (α, A, Σ) is then

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

where ι is an n-dimensional vector of ones and therefore $\iota' A \iota = \sum_{i,j=1}^{n} a_{ij}$. Under $\varphi = nT\tilde{\varphi}$ and $\lambda = \frac{nT}{2}\tilde{\lambda}$, the negative of the logarithm of this posterior is proportional to the objective function in the LS minimization problem in equation A.2 above. Therefore, with the proper re-scaling of the penalty parameters, the conditional posterior mode of $\alpha, A|Y, \Sigma$ is equal to $(\hat{\alpha}_{LS}, \hat{A}_{LS})|\Sigma$, while the mode of $p(\Sigma|Y, \alpha, A)$ is equal to $\hat{\Sigma}_{LS}|\alpha, A$. These posteriors and their modes are derived below.

Conditional Posteriors

To find the conditional posterior $p(\alpha|Y, A, \Sigma)$, it is convenient to rewrite u_t as in Section B.3 as $u_t = y_t - X_t(A)\alpha$, where the $n \times pq$ matrix X_t depends only on A and the data:

$$X_t = \left[\tilde{y}_{t-1}^1, \tilde{y}_{t-1}^2, ..., \tilde{y}_{t-1}^q, \tilde{y}_{t-2}^1, ..., \tilde{y}_{t-p}^q \right] \ , \quad \tilde{y}_{t-l}^g \equiv A^g y_{t-l} \ .$$

Under the stated prior, the conditional posterior of $\alpha|A, \Sigma$ is

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

which shows that

$$\alpha|Y, A, \Sigma \sim N\left(\bar{\alpha}, \bar{V}_\alpha\right) \ , \quad \text{with} \quad \bar{V}_\alpha = \left[\varphi I + \sum_t X_t' \Sigma^{-1} X_t \right]^{-1} \ , \quad \bar{\alpha} = \bar{V}_\alpha \left[\sum_t X_t' \Sigma^{-1} y_t \right] \ .$$

For $p(\Sigma|Y,\alpha,A)$, we obtain

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

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

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

To find $p(A|Y,\alpha,\Sigma)$ for $q = 1$, write the model as

$$y_t = Az_t + u_t \ , \quad u_t \sim N(0,\Sigma) \ ,$$

with the $n \times 1$ vector z_t given by $z_t = [\alpha_{11}y_{t-1} + ... + \alpha_{p1}y_{t-p}]$. Defining the $T \times n$ matrices Y, Z and U, this can be written as

$$Y = ZA' + U \ .$$

In turn,

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

To simplify exposition, define $\check{A} = A'$. Using the rules $a'Ba = tr[Baa']$, $tr[AB] = tr[BA]$

and $tr[A] = tr[A']$,[46] we get

$$\begin{aligned} p(A|Y,\alpha,\Sigma) &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(Y'Y - \breve{A}'Z'Y - Y'Z\breve{A} + \breve{A}'Z'Z\breve{A}\right) + 2\lambda\breve{A}\iota\iota'\right]\right\} \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(Y'Y - \breve{A}'Z'Y - Y'Z\breve{A} + \breve{A}'Z'Z\breve{A}\right) + 2\lambda\iota\iota'\breve{A}\right]\right\} \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(Y'Y - \breve{A}'Z'Y - Y'Z\breve{A} + \breve{A}'Z'Z\breve{A} + 2\lambda\Sigma\iota\iota'\breve{A}\right)\right]\right\} \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left[\breve{A}'Z'Z\breve{A} - 2\breve{A}'Z'Y + 2\lambda\breve{A}'\iota\iota'\Sigma\right]\right]\right\} \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left[\breve{A}'Z'Z\breve{A} - 2\breve{A}'\left(Z'Y - \lambda\iota\iota'\Sigma\right)\right]\right]\right\} , \end{aligned}$$

which implies

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

Estimation of an NVAR($p,1$)

Bayesian estimation of the NVAR($p,1$) can be implemented by Gibbs sampling, i.e. iteratively drawing from the conditional posteriors. Analogously, the OLS estimator $(\hat{\alpha}_{OLS}, \hat{A}_{OLS}, \hat{\Sigma}_{OLS})$ can be obtained by iterating on conditional OLS estimators as outlined in Meng and Rubin (1993). These are given by

$$\begin{aligned} \hat{\alpha}_{OLS}|A &= \left[\tilde{\varphi}I + \sum_t X_t(A)'X_t(A)\right]^{-1}\left[\sum_t X_t(A)'y_t\right] , \\ \hat{a}_{ij,LS}|(A_{i,-j},\alpha) = max\{0,\breve{a}_{ij}\} \quad , &\quad \breve{a}_{ij} = \frac{\sum_{t=1}^T (y_{it} - A_{i,-j}z_{-j,t})z_{jt} - \lambda}{\sum_{t=1}^T z_{jt}^2} . \end{aligned}$$

Under OLS, conditional on α, the estimators $\hat{a}_{ij,LS}$ are independent across i. This leads to the following procedure:

Algorithm 3 (OLS Estimation of (α, A) in NVAR($p,1$))**.**

1. *Let* $\alpha^{(0)} = (1,0,...,0)'$ *and* $A^{(0)} = 0_n$.
2. *For* $m = 1,2,...,$ *do the following:*

[46] I also make use of the symmetry of Σ and the fact that $tr[A] + tr[B] = tr[A+B]$.

a) *Given* $\alpha^{(m-1)}$ *and* $A^{(m-1)}_{\cdot,k}$ *for columns* $k = 2 : n$ *of* A*, update the estimate for the first column of* A *by computing*

$$\{\hat{a}_{i1,LS}\}_{i=1}^{n} | (A^{(m-1)}_{\cdot,-1}, \alpha^{(m-1)})$$

b) *Repeat the above iteratively for columns* $j = 2 : n$*: given* $\alpha^{(m-1)}$, $A^{(m)}_{\cdot,k}$ *for columns* $k = 1 : j-1$ *of* A *and* $A^{(m-1)}_{\cdot,k}$ *for columns* $k = j+1 : n$ *of* A*, update the estimate for column* j *of* A *by computing*

$$\{\hat{a}_{ij,LS}\}_{i=1}^{n} | (A^{(m)}_{\cdot,1:j-1}, A^{(m-1)}_{\cdot,j+1:n}, \alpha^{(m-1)})$$

c) *Compute* $\alpha^{(m)} | A^{(m)}$.

d) *Let* $c = ||\alpha^{(m)}||_1$*. Implement the normalization by dividing* $\alpha^{(m)}$ *by* c *and multiplying* $A^{(m)}$ *by* c.

e) *If max* $\left\{||\alpha^{(m)} - \alpha^{(m-1)}||, ||\Sigma^{(m)} - \Sigma^{(m-1)}||\right\} < \varepsilon$ *for some* ε *small, break. Otherwise, continue.*

C.3 Data and Results

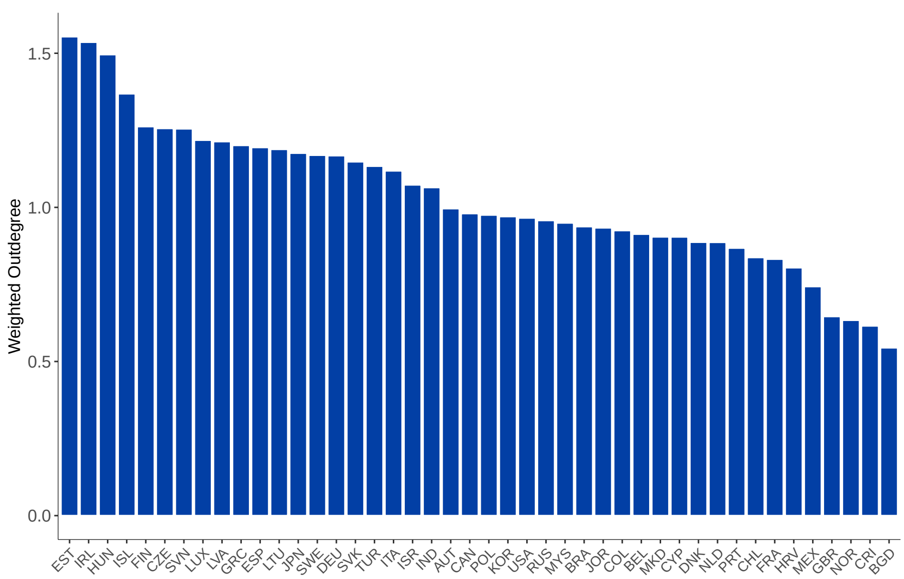

Figure A-5: Weighted Indegrees in the Estimated Network

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

Table A-4: 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

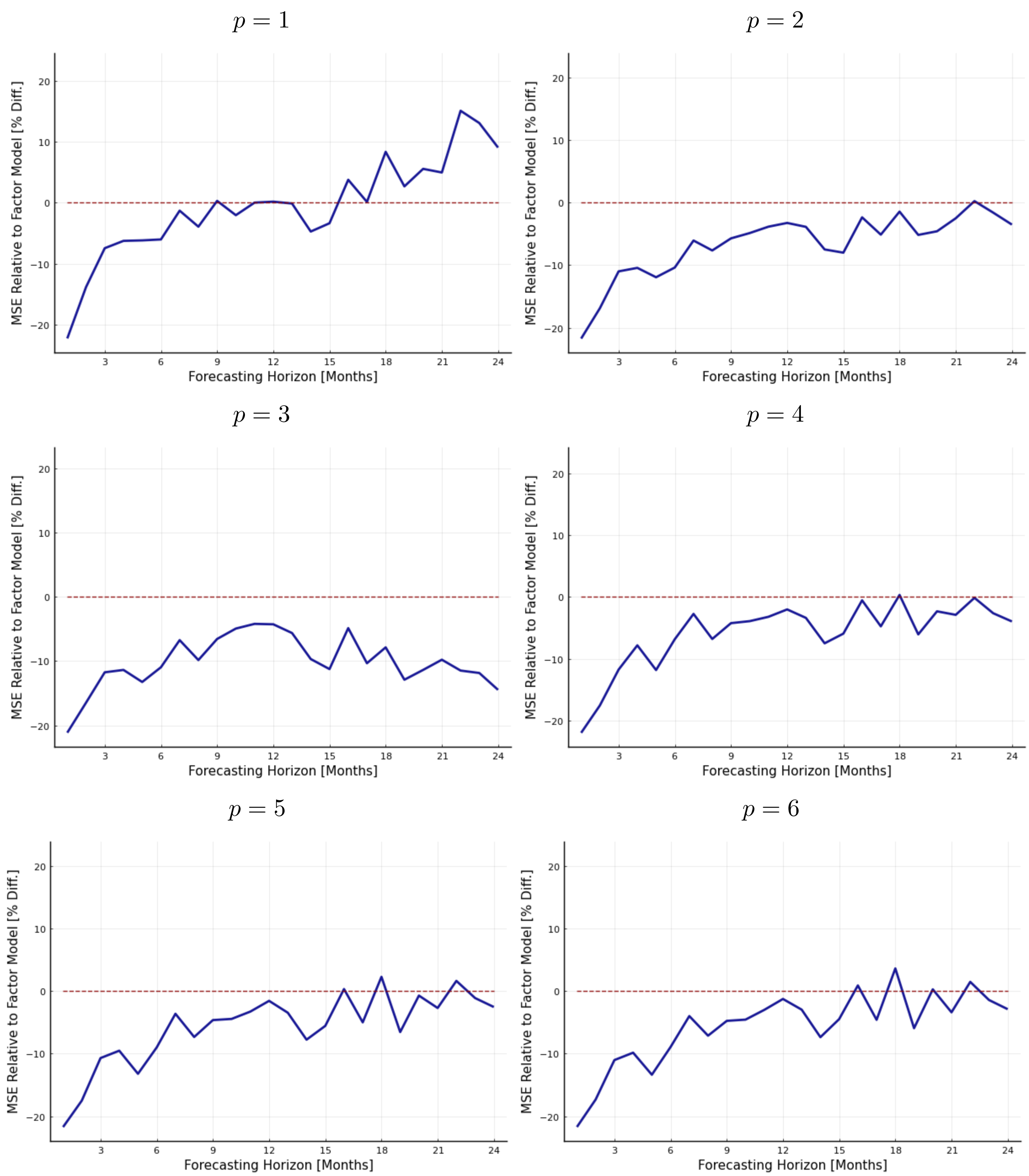

Figure A-6: Out-of-Sample Forecasting Performance: NVAR$(p, 1)$ vs. Factor Model

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.