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

Marko Mlikota

Geneva Graduate Institute

This Version: 2026-01-23

[latest version: click here]

Abstract

Many economic environments involve units linked by a network. I develop an econometric framework that derives the dynamics of cross-sectional variables from the lagged innovation transmission along bilateral links and that can accommodate general patterns of how higher-order network effects accumulate over time. The proposed NVAR rationalizes the Spatial Autoregression as the limit under an infinitely high frequency of lagged network interactions. The factor-representation of the NVAR suggests that at the cost of restricting factor dynamics, it naturally incorporates sparse factors as locally important nodes in the network. The NVAR can be used to estimate dynamic network effects. When the network is estimated as well, it also offers a dimensionality-reduction technique for modeling high-dimensional processes. In a first application, I show that sectoral output in a Real Business Cycle-economy with lagged input-output conversion follows an NVAR. In turn, I estimate that the dynamic transmission of productivity shocks along supply chains accounts for 61% of persistence in aggregate output growth, leaving minor roles for autocorrelation in exogenous productivity processes. In a second application, I forecast macroeconomic aggregates across OECD countries by estimating a network behind global business cycle dynamics. This reduces out-of-sample mean squared errors for one-step ahead forecasts relative to a dynamic factor model by -12% (quarterly real GDP growth) to -68% (monthly CPI inflation).

JEL codes: C32, E32, E37.

Key words: Vector Autoregression, Spatial Autoregression, Dynamic Network Effects, Input-Output Economy, Business Cycles, High-Dimensional Time Series, Local Factors.

* Correspondence: Department of International Economics, Geneva Graduate Institute (IHEID), 1202 Geneva, Switzerland. Email: marko.mlikota@graduateinstitute.ch. This paper constitutes part of my doctoral dissertation at the University of Pennsylvania. I am highly indebted to my advisors, Frank Diebold and Frank Schorfheide, for their support and guidance. I thank Shuxiao Zhou for research assistance. For helpful comments and discussions I am also grateful to Xu Cheng, Marco Del Negro, Thorsten Drautzburg, Simon Freyaldenhoven, Wayne Gao, Domenico Giannone, Bo Honoré, Leon Huetsch, Daniel Lewis, Laura Liu, Geert Mesters, Aaron Mora, Serena Ng, Michael Pollmann, Giorgio Primiceri, Joao Ritto, Elisa Rubbo, Adrien Wicht as well as seminar and conference participants at Penn, 7th Lindau Meeting on Economic Sciences, SNDE Symposium 2023, Philadelphia Fed, and NBER Summer Institute 2023 ("Forecasting & Empirical Methods") (among many others).

1 Introduction

In economics, we often encounter a cross-section of units linked by a network of bilateral ties, such as sectors connected through supply chains or individuals acquainted to each other. A large theoretical and empirical literature documents that networks amplify idiosyncratic shocks and generate comovement in cross-sectional variables at a given point in time.[1]

How network-induced comovements unfold over time, however, is less well understood. The literature predominantly assumes contemporaneous network interactions (e.g. Bramoullé et al. (2009); Acemoglu et al. (2012); Elliott et al. (2014)), embodied by the Spatial Autoregressive (SAR) model. This implies that an innovation contemporaneously transmits along network connections of all orders. The resulting static framework is useful for "networked" steady state comparisons. Among the few studies that consider lagged network interactions, most posit that network effects materialize exactly one link per period (e.g. Long and Plosser (1983); Golub and Jackson (2010); Elhorst (2012)). While useful for analyzing the qualitative implications of the laggedness of network interactions, this assumption is of limited use for empirical work concerned with "networked" transition dynamics. Zhu et al. (2017) generalize the lag length, but the generality of their empirical framework and its relation to structurally motivated models of contemporaneous or single-lagged network interactions remain unclear.

I construct an econometric framework that derives the dynamics of cross-sectional variables from the lagged innovation transmission along bilateral links between cross-sectional units: the Network-VAR (NVAR). Like existing studies, I assume uni-directional transmission[2] and time-invariant links. Unlike existing studies, the model can accommodate general patterns on how innovations travel along links over time and, consequently, how network connections of higher order accumulate as time progresses. To obtain this generality, I distinguish between the frequency of network interactions and the frequency of observation. As the former diverges to infinity, the impulse-response to contemporaneous high-frequency disturbances and the networked covariance among observations approach their counterparts from the SAR model. The same impulse-response is also obtained when considering long-term effects of permanent innovations. By taking a stance on the timing of network effects, the NVAR goes beyond such steady state comparisons and characterizes transition dynamics.

In the NVAR, the interdependence of observations y_{it} and $y_{j,t-h}$ arises as the interplay of the temporal distance between periods t and $t-h$ and the cross-sectional distance between units i and j encoded by the network. Under a sparse network, this dependence is modeled parsimoniously; the dynamic comovement among all series is rationalized by the dynamic innovation transmission along a few bilateral links among units. This is akin to the assumption that longer-term dynamics are driven by a set of shorter-term dynamics, upheld by the

[1] See work surveyed in Carvalho and Tahbaz-Salehi (2019); Bramoullé et al. (2016) and the following literature review.

[2] i.e. either downstream or upstream, the distinction being relevant only for directed networks.

general class of VARMA(p, q) models. The dynamics under the NVAR can be represented by a Dynamic Factor Model (DFM) whereby the number of factors equals the rank of the network adjacency matrix. At the cost of restricting their dynamics, the NVAR naturally incorporates sparse factors as locally important nodes in the network.

The NVAR is useful in two distinct kinds of work with cross-sectional time series. By conducting inference on the timing of innovation transmission along bilateral links, it can be used to estimate dynamic network effects. Thereby, the network can be taken as given or it can be inferred from dynamic cross-correlations in the data, possibly aided by shrinking towards observed links. When both the network and effect-timing are estimated, the NVAR is also applicable as a dimensionality-reduction technique for modeling high-dimensional processes. Conditional on a network, inference on the timing of network effects boils down to a linear regression with covariates that summarize lagged observations using bilateral links. Joint inference is implemented by iterating on analytically available conditional estimators, with Bayesian as well as frequentist interpretations. I illustrate each of these two model uses with a respective application.

In the first application, I show that the NVAR approximates the process of sectoral output in a Real Business Cycle (RBC) input-output economy with lagged input-output conversion (IOC). In turn, I evaluate the potential of lagged IOC to generate endogenous business cycles, as suggested theoretically by Long and Plosser (1983). I generalize their one period-lagged IOC by assuming that inputs from the past p periods are imperfectly substitutable in the production at any time t – a stand-in for a range of frictions that prevent just-in-time input-sourcing – and by allowing the model frequency to differ from the observational frequency. I characterize impulse-responses under a difference-stationary TFP process with persistent aggregate and sectoral shocks. The analysis illustrates the differing implications of lagged IOC and persistence in exogenous shocks for the dynamics of sectoral and aggregate output; the former formalizes the idea that idiosyncratic changes in, say, energy production take time to feed through the networked economy, while the latter merely prolongs the effects of each round of transmission. Using input-output tables and monthly output growth across 23 manufacturing sectors in the US economy, I estimate their respective contributions to persistence in aggregate output growth.

The results suggest that lagged IOC explain 61% of the autocorrelation in aggregate output growth. With persistence in aggregate TFP, this number grows to 93%, leaving little room for persistence in idiosyncratic sectoral TFP. My analysis reveals that shocks to more central sectors not only affect aggregate output more strongly in the long run – as established by Acemoglu et al. (2012) –, but they also tend to materialize more sluggishly. This relationship, however, is far from clear-cut; for example, owing to its position at the top of supply chains, a TFP shock in the primary metals sector materializes more slowly than a shock to the food and beveredge sector, despite similar long-term effects.

In the second application, I investigate the merits of the NVAR as a dimensionality-reduction technique for forecasting high-dimensional (cross-sectional) processes. I consider monthly industrial production growth, monthly CPI inflation and quarterly GDP growth across OECD countries. Under a sparse network, the NVAR models the dynamics parsimoniously; the dynamic comovement among all series is rationalized by the dynamic innovation transmission along a small set of bilateral links among countries. In line with that, I estimate the network by shrinking links to zero using a Lasso-penalty. I also consider alternative specifications that apply a Ridge-penalty and shrink to observable links.

In my setting, the NVAR reduces out-of-sample mean squared errors by 13%-68% relative to the ex-post best-performing DFM. This corroborates my equivalence result: the NVAR better forecasts these series, whose cross-country dynamics seem to be driven by many micro links rather than a few influential countries. This corresponds to a setting with numerous sparse factors and differing sets of non-zero loadings across units or, equivalently, a sparse, yet high-rank network adjacency matrix.

Contribution There is a large econometric literature on spatial and network-interactions (Manski, 1993; Kelejian and Prucha, 2010; Anselin, 2003; Lee, 2007; Bramoullé et al., 2009; Shalizi and Thomas, 2011; Kuersteiner and Prucha, 2020; de Paula et al., 2024). It is concerned with identifying spatial/peer-effects and accounting for heterogeneity in cross-sectional and panel data-regressions.[3] It predominantly assumes contemporaneous network interactions, as exemplified by the canonical SAR model. Among the few studies that consider lagged interactions (Elhorst, 2012; Knight et al., 2016), Zhu et al. (2017) cast them in an an explicit time series model, generalizing the lag length and discussing stationarity and large T-inference. I characterize Granger causality in their environment as the cross-sectional innovation transmission along network connections of different order, formalizing the idea that network effects materialize dynamically over time. In turn, I generalize this mapping between network connections and dynamics, which allows me to derive the SAR model as the limit of an underlying process driven by high-frequency lagged network interactions.[4] Like Pinkse et al. (2002); Lam and Souza (2020); Qu et al. (2021); de Paula et al. (2024) do in the static environment, I discuss inference on the network itself – not only the effect-timing –, which institutes the NVAR as a dimensionality-reduction technique – not only a tool for estimating dynamic network effects – and motivates my equivalence result to factor models.

Many studies incorporate networks in time series analyses. Following Diebold and Yılmaz (2009, 2014), some researchers interpret quantities of a given time series model – typically a VAR – using network analysis (Billio et al., 2012; Anufriev and Panchenko, 2015; Chen et al., 2025). Other studies restrict time series models using networks (Carvalho and West, 2007; Chudik and Pesaran, 2011; Davis et al., 2016; Ahelegbey et al., 2016; Zhu et al., 2017;

[3] A separate strand devises methods for dyadic data, i.e. the case when edges rather than nodes in the network are observed (Graham, 2020).

[4] Barigozzi et al. (2025); Maung (2022); Armillotta and Fokianos (2023) extend Zhu et al. (2017) along different dimensions than I do.

Barigozzi and Brownlees, 2018; Billio et al., 2019; Caporin et al., 2021; Martin et al., 2025).[5] Among them, the Global VAR (GVAR) (Pesaran et al., 2004; Chudik and Pesaran, 2014) restricts static and dynamic interdependencies across units in a panel-VAR using exogenous network-weights – e.g. bilateral trade –, enabling the estimation of many interconnected VAR models under a weak exogeneity assumption.[6] The NVAR also belongs to this second category; it restricts innovations in a (cross-sectional) VAR to transmit along bilateral links among variables (units), which leads to rich patterns of multi-step causality as in Dufour and Renault (1998). In contrast to the GVAR and most "networked" time series approaches above, I focus on a single variable per cross-sectional unit and a single type of connections among them, and I entertain the assumption that innovations transmit along bilateral links. This allows me to characterize a range of theoretical properties – including the relation to static SAR models and DFMs –, to structurally motivate the NVAR with an RBC production economy, and to efficiently conduct inference by relying on analytical conditional estimators.

When the network is estimated parsimoniously, the NVAR also addresses the literature on dimensionality-reduction methods for time-series modeling. By reducing the number of parameters and applying shrinkage priors, it leverages both approaches available for addressing the large parameter problem in the Wold representation (Geweke, 1984). Compared to standard shrinkage methods (Litterman, 1986; Tibshirani, 1996), it applies shrinkage to links – which in turn summarize the information in predictors at all lags – rather than to predictors themselves. Compared to reduced rank regression (Velu et al., 1986) or factor models (Geweke, 1977; Stock and Watson, 2002), it finds the linear combination to summarize the information in predictors by relying on network connections of different order and, ultimately, on bilateral links among units. While many of the network-restricted time series approaches mentioned above appear to have a reduced-rank representation – consider the GVAR's construction of weighted averages of other countries' variables –, the simple setup of the NVAR allows me to characterize this representation explicitly. Owing to the equivalence result that compares the NVAR to the DFM – corroborated by their relative performance forecasting the three processes I consider –, the NVAR contributes to the long-standing literature on sparse factors (Boivin and Ng, 2006; Onatski, 2012; Bai and Ng, 2019; Fan et al., 2022; Anatolyev and Mikusheva, 2022; Freyaldenhoven, 2025). Interestingly, in my application, a sparse network is preferred to a dense one, suggesting the presence of sparse rather than weak factors and running counter to the results of Giannone et al. (2021).

With the other application of the NVAR, I contribute to the series of efforts to microfound aggregate dynamics using production networks (see Horvath (1998, 2000); Dupor (1999); Shea (2002); Carvalho (2010); Foerster et al. (2011); Acemoglu et al. (2012); di Giovanni et al. (2014) and survey of Carvalho and Tahbaz-Salehi (2019)). These studies rely

[5] Holly and Petrella (2012); Dahlhaus et al. (2021); De Graeve and Schneider (2023); Chodorow-Reich et al. (2025) exploit networks for shock-identification in VARs. Bykhovskaya (2023) models the temporal evolution of a weighted network, which belongs to the realm of dyadic regressions (Graham, 2020) and (dynamic) network-formation rather than -effects.

[6] See Canova and Ciccarelli (2013) for an overview of panel-VAR approaches.

on contemporaneous IOC – mirroring the prevalence of contemporaneous interactions in econometrics and in the related study of spatial economies (Desmet and Parro, 2025) – and show that the production network greatly contributes to aggregate volatility. A notable exception is Long and Plosser (1983), who show that lagged IOC endogenizes business cycles – i.e. persistence in aggregates –, without relying on autocorrelation in exogenous shocks, non-rational expectations, or other frictions. Their analysis points to a more prominent role for production networks than as mere amplification devices of existing dynamics; it formalizes (qualitatively) the notion that sectoral shocks take time to feed through the economy, an idea well-accepted in the trade literature (Alessandria et al., 2010; Liu and Tsyvinski, 2024; Antràs and Tubdenov, 2025). Carvalho and Reischer (2021) characterize persistence in the Long and Plosser (1983)-economy and show that its implied evolution due to observed changes in the US input-output network accounts well for empirical measures of the changing persistence of aggregate output growth. Their analysis suggests that the endogenization of business cycles through lagged IOC is not only theoretically attractive, but has empirical merit as well. I quantify the empirical salience of Long and Plosser (1983)'s hypothesis by taking an RBC economy with general lags in IOC to the data. While lagged IOC leads to the same steady state as under contemporaneous IOC, it decomposes the long-term responses to granular TFP shocks, with distinct transition dynamics from those generated by autocorrelated TFP processes. By conducting inference on the timing of IOC along with the autocorrelations and variance of exogenous TFP processes, I quantify the contribution of lagged IOC to the entire autocorrelation function of aggregate output growth, supplementing and extending earlier decompositions of aggregate volatility (Foerster et al., 2011; di Giovanni et al., 2014; De Graeve and Schneider, 2023).

Outline The rest of this paper is structured as follows. The model and its properties are discussed in Section 2. Section 3 treats inference. In Section 4, I study how input-output connections affect output dynamics in the US economy. In Section 5, I illustrate the merits of the NVAR as a dimensionality-reduction technique. Section 6 concludes.

2 Dynamics via Lagged Network Effects: Theory

After providing some basic background on networks in Section 2.1, I present the NVAR and its main properties in Section 2.2, and I discuss further properties in Section 2.3. Details and proofs are in Section A. The discussion of inference is deferred to Section 3.

2.1 Bilateral Connections in Networks

A network is represented by an $n \times n$ adjacency matrix A with elements a_{ij}. I consider a directed, weighted and possibly signed network; $a_{ij} \in [-1, 1]$ shows the sign and strength of the link from unit i to unit j, with $a_{ij} \neq a_{ji}$ possibly. If $a_{ij} = 0$, I say unit i is not connected to unit j. Self-links are permitted: $a_{ii} \neq 0$. The set of bilateral links $\{a_{ij}\}_{i,j=1:n}$ give rise to a myriad of higher-order (hyper-dyadic) links among units, referred to as walks.[7]

Definition 1 (Walk)**.** *A walk from i to j of length $K \geq 2$ is*

$$a_{u_1,u_2,...,u_{K+1}} \equiv \prod_{k=1}^{K} a_{u_k,u_{k+1}} \ , \quad u_1 = i \ , \ \ u_2 = j \ .$$

A walk is a product of bilateral links that lead from unit i to unit j over intermediary units. It is non-zero if all of these units are sequentially connected. Simple matrix algebra reveals that $[A^K]_{ij}$ contains the sum of walks from i to j of length K. I refer to this quantity as the Kth-order connection from i to j. Consider the following example:

$$A = \begin{bmatrix} 0 & 0 & .8 \\ .7 & 0 & .2 \\ 0 & .9 & 0 \end{bmatrix} \ , \ \ A^2 = \begin{bmatrix} 0 & .72 & 0 \\ 0 & .18 & .56 \\ .63 & 0 & .18 \end{bmatrix} \ , \ \ A^3 = \begin{bmatrix} .50 & 0 & .14 \\ .13 & .50 & .04 \\ 0 & .16 & .50 \end{bmatrix} \ .$$

Even though unit 3 is not directly connected to unit 1 ($a_{31} = 0$), there exists a second-order connection via unit 2 ($a_{32}a_{21} \neq 0$). 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

Consider a stationary cross-sectional time series $y_t = (y_{1t}, ..., y_{nt})'$ with mean zero. Under an NVAR, the dynamics of y_t are driven by the lagged transmission of innovations u_{it} along bilateral links a_{ij} among cross-sectional units $i, j = 1 : n$. Throughout, I assume that links are fixed over time and that transmission is uni-directional: the direct link from i to j, a_{ij}, transmits innovations from j to i. Innovations u_t may be cross-sectionally correlated.

2.2.1 Single Lag in Innovation Transmission: NVAR(1,1)

Consider a VAR(1) with an autoregressive matrix proportional to the adjacency matrix A:

$$y_t = \alpha A y_{t-1} + u_t \ , \tag{1}$$

for $\alpha \in \mathbb{R}$. Under this process, the one step-ahead expectation of y_{it} is proportional to the weighted sum of one period-lagged values of y_{jt} for all units j to which i is directly linked: $\mathbb{E}_{t-1}[y_{it}] = \alpha \sum_{j=1}^{n} a_{ij} y_{j,t-1}$. For example, in the RBC input-output economy of Long and Plosser (1983), the expected output in sector i tomorrow is a weighted average

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

of the production in its supplier-sectors j today. In Golub and Jackson (2010), individual i's expected opinion tomorrow is a weighted average of their friends' opinions today. For reasons clarified below, I dub this process NVAR(1,1).

Under Eq. (1), dynamics of y_t at horizon $h \geq 1$ are driven by hth-order network connections. More precisely, assuming $\alpha \neq 0$, y_j Granger-causes y_i at horizon h iff there exists an hth-order connection from i to j:[8]

$$GC^h_{ij} \equiv \frac{\partial y_{i,t+h}}{\partial y_{jt}} = \frac{\partial y_{i,t+h}}{\partial u_{jt}} = [\alpha^h A^h]_{ij} \ , \quad h \geq 1 \ .$$

Fig. 1 illustrates these Granger-causality dynamics – also referred to as Generalized Impulse-Response Functions (GIRFs)[9] – for the network from Section 2.1 under $\alpha = 1$. It points to two insights. First, lagged network interactions generate persistence; even white noise-innovations $\{u_{jt}\}_{j=1}^n$ lead to persistent reactions of y_t because each u_{jt} affects connected units $i \neq j$ with a lag.[10] This result is behind the endogenous business cycles in Long and Plosser (1983)'s economy with lagged input-output conversion. Second, under this type of persistence, network-connections affect not only the strength of impulse-responses, but also their timing. For example, while unit 2 is directly linked to unit 1 and therefore experiences the latter's innovation with a lag of one period, unit 3 only has a second-order connection to unit 1 and therefore reacts only after two periods. This result relates to Dufour and Renault (1998), who point out that Granger-causality can take the form of chains.[11] In the case of a (cross-sectional) time series driven by lagged innovation transmission along bilateral links, their insight emerges naturally and their generally non-trivial conditions for Granger-(non)causality boil down to the presence or absence of network connections of relevant order between the relevant variables (units). As I show in the following, this holds even under a more general transmission-timing than considered in Eq. (1).

While the process in Eq. (1) reveals useful theoretical insights in both macro- (Long and Plosser, 1983) and microeconomics (Golub and Jackson, 2010), it is too restrictive for most empirical work. It assumes that innovations transmit at the speed of one link per period and that transmission materializes completely within a single period. This implies, for instance, that in response to news gathered by a friend of a friend, j, an individual i adjusts their opinion in two periods, not earlier, not later. Similarly, an innovation to a sector j's output affects the output of other sectors i located two positions downstream ("customers of customers") in exactly two periods. In either case, earlier responses are ruled

[8]In other words, given all other variables y_{kt} for $k \neq j$, y_{jt} is useful in forecasting $y_{i,t+h}$ iff there exists an hth-order connection from i to j.

[9]The GIRF is "generalized" because it disregards shock identification, but considers the propagation of reduced form errors u_t over future time periods $t + h$, $h \geq 1$.

[10]Furthermore, each individual y_{jt} reacts persistently to its own white noise-innovation u_{jt} as long as there is a cycle in the network, i.e. a walk from j to j. Notably, this can hold even without a self-link – $a_{jj} = 0$.

[11]Specifically, even if a series x_t does not Granger-cause a series y_t at horizon 1, under the presence of a third series z_t, x_t might Granger-cause y_t at higher horizons as the causality could run from x_t to z_t to y_t.

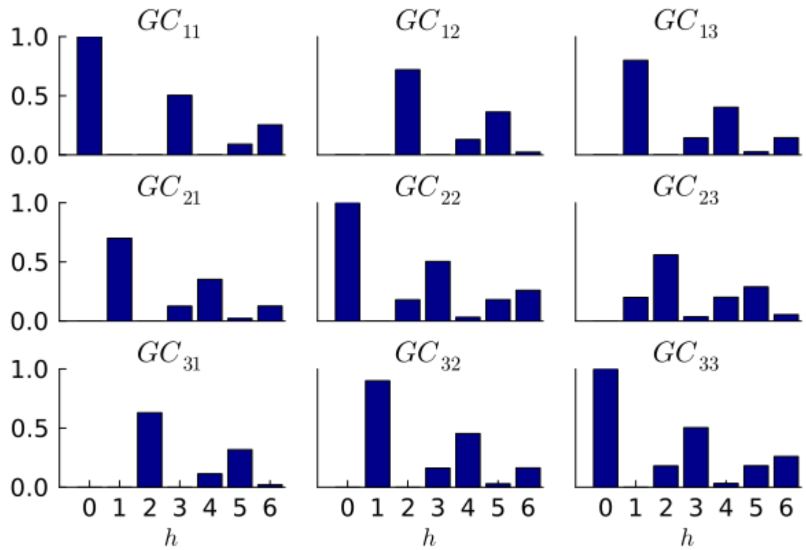

Figure 1: Example of Generalized Impulse-Responses under an NVAR$(1, 1)$

Notes: Panel (i, j) shows $GC^h_{ij} = [\alpha^h A^h]_{ij}$ under $\alpha = 1$. Note that $GC^0_{ij} = 1$ if $i = j$ and zero otherwise, i.e. by definition, the contemporaneous responses to all but a series' own innovation are zero.

out altogether, and later adjustments only occur if i has a third- or higher-order connection to j. In the following, I generalize the timing of lagged network effects by extending the simple process above along two dimensions.

2.2.2 Multiple Lags in Innovation Transmission: NVAR$(p, 1)$

Let the cross-sectional time series y_t evolve according to a VAR(p), where each autoregressive matrix is proportional to the same network adjacency matrix A:

$$y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t \; , \tag{2}$$

with $\alpha = (\alpha_1, ..., \alpha_p)' \in \mathbb{R}^p$. I dub this process NVAR$(p, 1)$. It simplifies the process presented in Zhu et al. (2017) by not distinguishing self- and cross-links, by abstracting from restrictions on the rows of A, and by disregarding the intercept and covariates. These simplifications prove useful for characterizing a range of proprties in this section, for conducting inference in Section 3, and for taking a macroeconomic model of lagged input-output conversion to the data in Section 4.

Compared to Eq. (1), under the NVAR$(p, 1)$ with $p > 1$, dynamics at horizon h are affected by lower-order connections:

Proposition 1 (Granger-Causality in NVAR$(p, 1)$)**.**
Let y_t evolve as in Eq. (2). Then, Granger-causality from y_j to y_i at horizon h is a linear combination of connections from i to j of order $k \in \{\underline{k}, \underline{k}+1, ..., h\}$, where $\underline{k} = ceil(h/p)$.[12]

Under $\alpha_l > 0 \; \forall \; l$ and $a_{ij} \geq 0$, Proposition 1 implies that y_j Granger-causes y_i at horizon h iff there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}+1, ..., h\}$. The proof of Proposition 1 in Section A.1 establishes that the GIRF is of the form

$$\frac{\partial y_{i,t+h}}{\partial u_{j,t}} = c^h_{\underline{k}}(\alpha) \left[A^{\underline{k}}\right]_{ij} + ... + c^h_h(\alpha) \left[A^h\right]_{ij} \; . \tag{3}$$

Each coefficient c^h_k for $k = \underline{k} : h$ is a polynomial of $\{\alpha_l\}_{l=1:p}$ and shows the importance of connection-order k for the impulse-response at horizon h. If i has a first- and no higher-order connections to j, then $\partial y_{i,t+h}/\partial u_{j,t} = \alpha_h a_{ij}$ for $h = 1 : p$ and zero otherwise. Hence, the NVAR$(p, 1)$ specifies that dynamics are driven by the transmission of innovations along direct links lagged over p periods. α determines how the transmission is spread out over these p periods and, consequently, how transmission along higher-order connections accrues as time progresses. Note that the transmission is assumed to be the same for all unit pairs (i, j) and invariant over time.

2.2.3 Multiple Rounds of Innovation Transmission: NVAR(p, q)

To further generalize the timing of innovation transmission, suppose that we observe $\tilde{y}_\tau \sim$ NVAR$(p, 1)$ every $q \in \mathbb{N}$ periods; let the observed series y_t evolve according to the state space system

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

I dub this process NVAR(p, q). q indicates the relation between the frequency of network interactions, at which $\tilde{y}_\tau$ evolves, and the frequency at which data $\{y_t\}_{t=1:T}$ is observed. If $q = 1$, the two coincide and, trivially, y_t and $\tilde{y}_\tau$ are the same NVAR$(p, 1)$ process. Instead, if $q > 1$, then network interactions occur at higher frequency than data is observed. For example, under yearly observations, $q = 4$ implies quarterly network interactions. As a result, multiple rounds of transmission occur in a single observational period and dynamics at horizon h are affected by higher-order connections, $k > h$:

Proposition 2 (Granger-Causality in NVAR(p, q))**.**
Let y_t evolve as in Eq. (4) for some $q \in \mathbb{N}$. Then, Granger-causality from $\tilde{y}_j$ to $\tilde{y}_i$ at horizon h is a linear combination of connections from i to j of order $k \in \{\underline{k}, \underline{k}+1, ..., hq\}$, where $\underline{k} = ceil(hq/p)$.[13]

[12] $ceil(x)$ rounds $x \in \mathbb{Q}$ up to the next integer.

[13] Assuming again $\alpha_l > 0 \; \forall \; l$ and $a_{ij} \geq 0$, Proposition 2 implies that $\tilde{y}_j$ Granger-causes $\tilde{y}_i$ at horizon h iff there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}+1, ..., hq\}$.

Eq. (4) implicitly assumes that $\tilde{y}_\tau$ is a stock variable; every q periods we observe a snapshot of it. To accommodate a flow variable $\tilde{y}_\tau$, we can model y_t as

$$\begin{aligned} \tilde{y}_\tau &= \alpha_1 A \tilde{y}_{\tau-1} + ... + \alpha_p A \tilde{y}_{\tau-p} + \tilde{u}_\tau \ , \\ y_{\tau/q} &= (\tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1})/q \quad \text{if } \tau/q \in \mathbb{N} \ , \end{aligned} \tag{5}$$

with similar properties as under Eq. (4).[14]

By restricting the NVAR(p,q) in Eq. (4), we can accommodate more intricate relations between the frequencies of network interactions and observation; for any $q \in \mathbb{Q}_{++}$, we can write $q = q_1 q_2$ with $q_2, q_1^{-1} \in \mathbb{N}$, and we can model y_t as an NVAR(p^*, q_2) with $p^* = p/q_1 \in \mathbb{N}$ and with restricted parameters $\{\alpha_l\}_{l=1:p^*}$ to comply with the stated interpretations of p and q. For example, under monthly observations, $q = 4/3$ signifies that network interactions occur (roughly) every three weeks. This amounts to observing every 4 periods a snapshot of a weekly process that depends on its value in the past p 3-week-periods, i.e. on its value three weeks ago, six weeks ago, and so forth, until $p^* = 3p$ weeks ago. Section A.4 elaborates.

2.3 Further Properties of the NVAR

Stationarity In the NVAR, the interdependence of y_{it} and $y_{j,t-h}$ is an interplay of the cross-sectional distance between units i and j – encoded by the network – and the temporal distance between periods t and $t-h$. Correspondingly, Corollary 1 characterizes stationarity in terms of eigenvalues of the network adjacency matrix A and roots of an AR process shaped by the timing of innovation transmission along a single link, α.[15] Under $p > 1$, this simplifies checking for stationarity, especially for large n. Note that the second statement requires the AR(p) process $(1-\phi_1 L - ... - \phi_p L^p)x_t = v_t$ with $\phi_l = \alpha_l \lambda_i$ to be stationary for each eigenvalue λ_i of A, though AR-coefficients are typically required to be real-valued.

Corollary 1 (NVAR(p,q): Stationarity)**.**
Let y_t evolve as in Eq. (4) or Eq. (5) for some $q \in \mathbb{N}$. Let $\tilde{u}_\tau \sim WN(0, \Sigma)$, suppose $\alpha_l \neq 0$ for at least one l, and define $a = \sum_{l=1}^p |\alpha_l|$.

If $|\lambda_i| < 1/a$ for all eigenvalues λ_i of A, then y_t is weakly stationary. Under $\alpha_l \geq 0 \ \forall \ l$, the implication is both-sided. Moreover, y_t is weakly stationary iff for all eigenvalues λ_i of A, the $p \times p$ matrix below has all eigenvalues inside the unit circle:

$$\begin{bmatrix} \alpha_1 \lambda_i & ... & \alpha_p \lambda_i \\ I_{p-1} & & 0 \end{bmatrix} \ .$$

[14] See Section A.4 and Section 2.3. Analogous calculations apply if $y_t = \tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1}$.

[15] Corollary 1 follows from Propositions 8, 9 and 13. Intuitively, the conditions ensure that $\lim_{k\to\infty} c(\alpha,k)A^k = 0$, for any polynomial in α of order k, $c(\alpha,k)$. By Proposition 1, this ensures that the effects of disturbances vanish for higher horizons.

Relation to SAR Model In all of economics, the literature overwhelmingly considers contemporaneous network interactions as embodied by the Spatial Autoregressive (SAR) model. If $x = (x_1, ..., x_n)'$ is the cross-sectional variable of interest, it posits

$$x = Ax + v \ . \tag{6}$$

Provided that $|\lambda_i| < 1$ for all eigenvalues λ_i of A, we can write $x = (I - A)^{-1}v = (A + A^2 + A^3 + ...)v$. Hence, the implicit assumption is that connections of all order materialize in a single period. In line with that, Corollary 2 establishes that the response of x_i to an innovation v_j in the SAR model from Eq. (6) is equal to the long-run response of y_{it} to a persistent innovation $\tilde{u}_{j\tau}$ in the NVAR(p, q) from Eq. (4) or Eq. (5).[16]

Corollary 2 (NVAR(p, q): Long-Term Response to Persistent WN-Innovations)**.** *Let y_t evolve as in Eq.* (4) *or Eq.* (5) *for some $q \in \mathbb{N}$, and let $x = aAx + v$ with $a = \sum_{l=1}^{p} \alpha_l$. Assume y_t is weakly stationary. Then,*

$$\lim_{H \to \infty} \left[\frac{\partial y_{t+H}}{\partial \tilde{u}_{tq}} + \frac{\partial y_{t+H}}{\partial \tilde{u}_{tq+1}} + ... + \frac{\partial y_{t+H}}{\partial \tilde{u}_{(t+H)q}} \right] = \frac{\partial x}{\partial v} = (I - aA)^{-1} \ .$$

Both responses are given by the Leontief inverse $(I - aA)^{-1}$, which is a sufficient statistic for the long-term cross-sectional comovement. By taking a stance on the timing and frequency of network interactions, the NVAR shows how any such long-term effects materialize over time; it goes beyond steady state comparisons and characterizes transitional dynamics.[17] α temporally decomposes the spatial autoregressive parameter a.

The long-run is defined in terms of the frequency of network interactions. To justify the use of the SAR at the expense of the NVAR, then, we need this frequency to be high enough and the data frequency to be low enough. Corollary 3 formalizes this idea; if we define the observed process y_t as the sum of an underlying process $\tilde{y}_\tau$ evolving at the higher frequency of network interactions, then, as this frequency diverges to infinity, y_t's response to a high-frequency innovation that occurs within observational period t converges to the response of x_i to an innovation v_j in the SAR model from Eq. (6).[18]

Corollary 3 (NVAR(p, q): Contemporaneous Response to Within-Period Innovation)**.** *Let y_t evolve as in Eq.* (5) *for some $q \in \mathbb{N}\backslash\{1\}$, and consider $y_t q$. Then, for $g \in 0 : (q-1)$,*

$$\lim_{g \to \infty} q \frac{\partial y_t}{\partial \tilde{u}_{\tau - g}} = (I - aA)^{-1} \ .$$

[16] Corollary 2 follows from Propositions 5 and 11.

[17] Note the correspondence between the timing of transitional dynamics and the timing of responses to temporary innovations; as for any VAR, the long-term response to a permanent innovation is equal to the cumulative response to temporary innovations.

[18] Corollary 3 follows from Propositions 5 and 12, applied for $\rho_j = 0$.

Difference to Exogenous Persistence Instead of relying on lagged network interactions, one can also introduce transitional dynamics to the SAR model from Eq. (6) through autocorrelated innovations v. This approach is predominant in macroeconomic studies of input-output-economies; under fairly standard assumptions (see e.g. Carvalho and Tahbaz-Salehi (2019)), sectoral production and prices evolve according to Eq. (6), whereby v is Total Factor Productivity (TFP), which – across all areas of macroeconomics – is typically assumed to follow an autoregressive process.[19]

Lagged network interactions and autocorrelated innovations lead to different kinds of dynamics. As first illustrated in Fig. 1 for the NVAR(1,1), and as discussed throughout this section for the general NVAR(p, q), lagged network interactions relate impulse-responses to network-connections of different order, whereby the precise mapping is determined by α and q. As a consequence, how strongly y_i reacts to an impulse to y_j depends on the strength of network-connections from i to j of relevant order, and how fast it reacts depends on how many lower- as opposed to higher-order connections the two units share. For instance, in the environment of Golub and Jackson (2010) this implies that an individual adjusts their opinion faster after news shared by closer than more distant friends. Similarly, in the context of Long and Plosser (1983) it means that a sector contracts its production faster after a negative productivity shock to more immediate suppliers than to ones located further upstream.

In contrast, autocorrelated innovations lead to exponentially decaying impulse-responses. This shape is preserved when they are coupled with contemporaneous network interactions; if $x_t = aAx_t + v_t$ and $(1 - \rho_j L)v_{jt} = \epsilon_{jt}$, then

$$\frac{\partial x_{i,t+h}}{\partial \epsilon_{jt}} = \left[(I - aA)^{-1}\right]_{ij} \rho_j^h \ . \tag{7}$$

The strength of connections from i to j merely scales the exponentially decaying impulse-response; network connections merely amplify exogenous dynamics.

When the two sources of dynamics are combined, their qualitatively distinct contributions remain. Consider an NVAR$(p, 1)$, and let $(1 - \rho_j L)u_{jt} = \varepsilon_{jt}$. Then, using Proposition 1,

$$\begin{aligned}
\frac{\partial y_{i,t+h}}{\partial \varepsilon_{jt}} &= \sum_{l=0}^{h} \frac{\partial y_{i,t+h}}{\partial u_{j,t+h-l}} \frac{\partial u_{j,t+h-l}}{\partial \varepsilon_{j,t}} \\
&= \sum_{l=0}^{h} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \frac{\partial u_{j,t+h-l}}{\partial \varepsilon_{j,t}} = \sum_{l=0}^{h} \left\{ c^l_{\underline{k}(l)}(\alpha) \left[A^{\underline{k}(l)}\right]_{ij} + ... + c^l_l(\alpha) \left[A^l\right]_{ij} \right\} \rho_j^{h-l} \ ,
\end{aligned} \tag{8}$$

for $\underline{k}(l) = ceil(l/p)$. Under $\rho_j = 0$, this expression simplifies to Eq. (3). Under $p = 1$ and $\alpha_1 = 1$, it simplifies to $\sum_{l=0}^{h} \left[A^h\right]_{ij} \rho_j^{h-l}$. While α determines the timing of innovation

[19] It can be split into an idiosyncratic and an aggregate component; see Section 4.

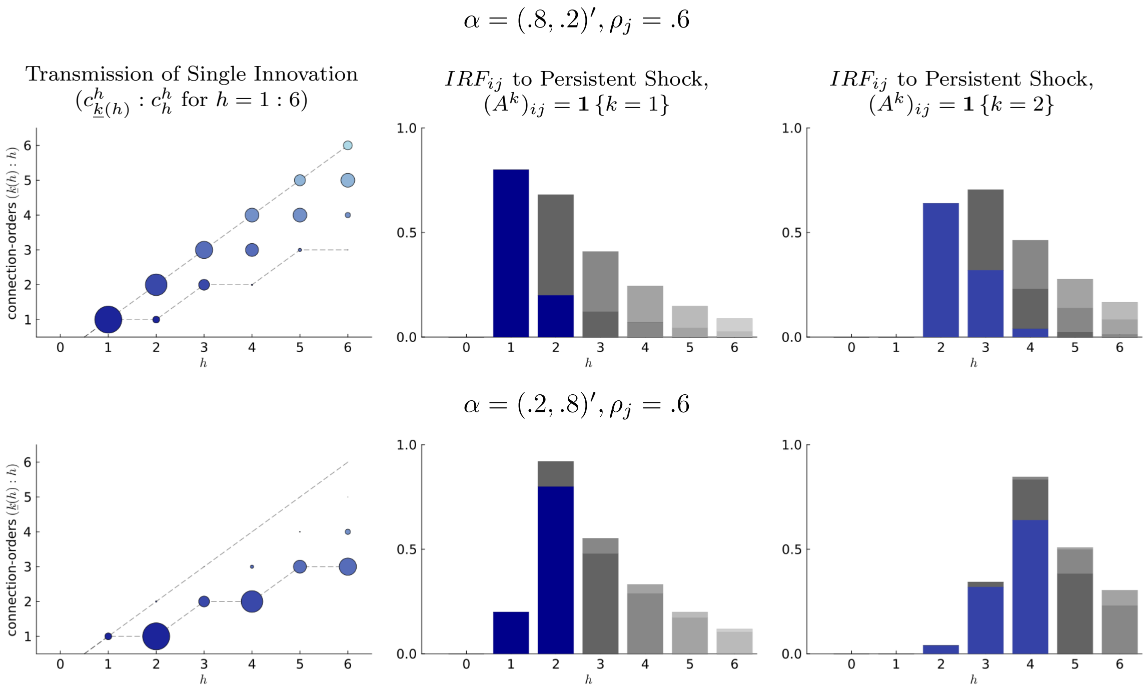

Figure 2: Composition of Impulse-Responses under NVAR$(2, 1)$ with AR(1)-Innovations

Notes: The top row refers to $\alpha = (.8, .2)'$, the bottom row to $\alpha = (.2, .8)'$. The left panel depicts the coefficients $c^h_{\underline{k}(h)} : c^h_h$ from Eq. (3) for $h = 1 : 6$, which illustrate how an innovation transmits along connection-orders (y-axis) over time (x-axis; horizons). The dashed lines bound the relevant connection-orders in an NVAR$(2, 1)$: $h \in \underline{k}(h) : h$. The middle panel plots the IRF from Eq. (8) under $\rho_j = .6$ for a pair (i, j) with only a first-order connection. The right panel repeats this for a pair with only a second-order connection.

transmission along direct links and, therefore, also the timing of transmission along all higher-order connections from i to j, ρ_j determines the persistence of y_i's response after each round of transmission.

Fig. 2 illustrates for an NVAR$(2, 1)$ and $\rho_j = .6$, once under $\alpha = (.8, .2)'$ (top row), once under $\alpha = (.2, .8)'$ (bottom row). The left panel depicts how connection-orders accumulate over horizons for these α, while the middle and right panels show the resulting impulse-responses for two sectors that share only a first- or only a second-order connection, respectively. Under $\alpha = (.8, .2)'$, the innovation travels faster through the network than under $\alpha = (.2, .8)'$; in the former case, the transmission along a direct link (darkest blue dots and bars) is strongest after one period, in the latter after two periods. Consequently, under $\alpha = (.8, .2)'$, the transmission along second-order connections (lighter blue dots and bars) is strongest in the second period, under $\alpha = (.2, .8)'$ in the fourth period. The exogenous persistence ensures that the fraction ρ_j of any transmission at horizon h is carried over to horizon $h + 1$, a fraction ρ_j^2 to horizon $h + 2$, etc. (ever lighter gray bars).

Even under autocorrelated innovations, lagged and contemporaneous network interactions yield the same long-run responses. This is shown by Corollary 4.[20]

[20]Corollary 4 follows from Propositions 6 and 11. In contrast to Corollary 2, it considers the long-run rather

Corollary 4 (NVAR(p,q): Long-Term Response to Persistent AR(1)-Innovations)**.** *Let y_t evolve as in Eq. (4) or Eq. (5) for some $q \in \mathbb{N}$, and let $x_t = aAx_t + v_t$, with $a = \sum_{l=1}^{p} \alpha_l$. Assume y_t and x_t are weakly stationary, and assume $(1-\rho_j L)u_{jt} = \varepsilon_{jt} \sim WN$ and $(1-\rho_j L)v_{jt} = \epsilon_{jt} \sim WN$, with $\rho_j \in [0,1)$. Then,*

$$\lim_{H\to\infty}\left[\frac{\partial y_{i,t+H}}{\partial \tilde{u}_{j,tq}} + ... + \frac{\partial y_{i,t+H}}{\partial \tilde{u}_{j,(t+H)q}}\right] = \lim_{H\to\infty}\left[\frac{\partial x_{i,t+H}}{\partial \epsilon_{j,t}} + ... + \frac{\partial x_{i,t+H}}{\partial \epsilon_{j,t+H}}\right] = \frac{[(I-aA)^{-1}]_{ij}}{1-\rho_j} \ .$$

Networked Contemporaneous Correlation Trivially, contemporaneous network interactions lead to contemporaneous correlation among $\{x_i\}_{i=1:n}$; under Eq. (6), we have $\mathbb{V}[x] = \mathbb{V}[(I-A)^{-1}v]$. In this case, $\text{Cov}(x_i, x_j)$ reflects the bilateral exposure between i and j and their mutual exposure to third units, whereby exposure is determined by network connections of all order.[21] De Graeve and Schneider (2023) exploit this insight for shock identification in the context of a production economy.

Networked contemporaneous correlation is obtained as soon as the network interaction-frequency exceeds the frequency of observation, even under finite values of the former. In an NVAR(p,q), we can define the "observable innovations" $u_t = y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{\tilde{y}_{(t-1)q}, \tilde{y}_{(t-1)q-1}, ...\}$. Under $q=1$, $u_t = \tilde{u}_\tau$ inherits the variance of the exogenous innovations $\tilde{u}_\tau$. Under $q>1$, u_t is a linear combination of past $\tilde{u}_\tau$, and its variance reflects units' bilateral exposure and mutual exposure to third units, whereby exposure is determined by network connections of orders $k \in 1:(q-1)$, along which the underlying innovations $\tilde{u}_\tau$ travel within a single period of observation. For example, if y_t evolves as in Eq. (5) for $q=2$, we have $u_t = \tilde{u}_\tau + (I+\alpha_1 A)\tilde{u}_{\tau-1}$. If $\tilde{u}_\tau \sim WN(0,\Sigma)$ with $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$, then

$$\text{Cov}(u_{it}, u_{jt}) = 2\sigma_i^2 \mathbf{1}\{i=j\} + \alpha_1 a_{ij}\sigma_j^2 + \alpha_1 a_{ji}\sigma_i^2 + \alpha_1^2 \sum_{k=1}^{n} a_{ik}a_{jk}\sigma_k^2 \ .$$

Thus, u_{it} and u_{jt} comove based on a_{ij}, a_{ji} and $\{a_{ik}a_{jk}\}_{k=1:n}$. Propositions 14 and 15 state $\mathbb{V}[u_t]$ for the NVAR(p,q) processes in Eqs. (4) and (5) and general $q>1$.

Corollary 3 above establishes that an SAR for x is obtained in the limit when writing x as the sum of an underlying process driven by lagged network interactions that evolves at an ever higher frequency. In addition, by Proposition 3, the Normality-assumption $v \sim N(0,\Sigma)$ in the SAR model can be justified by temporally independent high-frequency innovations

than contemporaneous response of x, since the implicit assumption in the SAR model with autocorrelated innovations is that they evolve at observational frequency.

[21] We have $\mathbb{V}[x] = \mathbb{V}[(I+A+A^2+...)v]$, provided that $|\lambda_i|<1$ for all eigenvalues λ_i of A. Assuming that $\mathbb{V}[v] = \Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$, we get

$$\text{Cov}(x_i, x_j) = \sigma_i^2 \mathbf{1}\{i=j\} + \sigma_j^2 \sum_{h=1}^{\infty}\left[A^h\right]_{ij} + \sigma_i^2 \sum_{h=1}^{\infty}\left[A^h\right]_{ji} + \sum_{o=1}^{n}\sigma_o^2 \sum_{h_i=1}^{\infty}\sum_{h_j=1}^{\infty}\left[A^{h_i}\right]_{io}\left[A^{h_j}\right]_{jo} \ .$$

$\tilde{u}_\tau$ with $\mathbb{V}[\tilde{u}_\tau] = q^{-1}\Sigma$. Concretely, for the NVAR(p,q) from Eq. (5) and for $a = \sum_{l=1}^{p} \alpha_l$, Proposition 3 shows that $\sqrt{q}u_t \overset{d}{\to} N\left(0, \mathbb{V}[(I - aA)^{-1}\tilde{u}_\tau]\right)$ as $q \to \infty$, provided the high-frequency innovations $\tilde{u}_\tau$ are strictly stationary.

Proposition 3 (NVAR(p,q): Limit Distribution of "Observable Innovations")**.**
Let y_t evolve as in Eq. (5) for some $q \in \mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $\tilde{u}_\tau \sim WN(0,\Sigma)$ is temporally independent. Define $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{\tilde{y}_{\tau-q}, \tilde{y}_{\tau-q-1}, ...\}$. Also, let $a = \sum_{l=1}^{p} \alpha_l$. Then, as $q \to \infty$,

$$\sqrt{q}u_t \overset{d}{\to} N(0, \Gamma_*\Sigma\Gamma_*') \ , \quad \Gamma_* = (I - aA)^{-1} \ .$$

Relation to Dynamic Factor Model The NVAR is a restricted VAR in which innovations transmit across series along constant, bilateral links (in one direction). Relative to a VAR(p) with $p > 1$, the NVAR$(p,1)$ induces parsimony; e.g. for each y_{it}, the n links $\{a_{ij}\}_{j=1:n}$ in $A_{i\cdot}$ summarize the information in the np-dimensional set $\{y_{j,t-l}\}_{j=1:n,l=1:p}$:

$$y_{it} = x_{it}'\alpha + u_{it} \quad \text{and} \quad y_t = X_t\alpha + u_t \ , \tag{9}$$

where $x_{it} = \sum_{j=1}^{n} a_{ij}y_{j,t-l}$ is the ith row of $X_t = [Ay_{t-1}, ..., Ay_{t-l}]$. If, in addition, the network A is sparse – as is the case across a wide range of applications –, then the NVAR rationalizes the dynamic comovement among all $\{y_{it}\}_{i=1:n}$ by the dynamic innovation transmission along few bilateral links among $i = 1:n$.

The Dynamic Factor Model (DFM) also reduces dimensionality by summarizing a large set of predictors by a few linear combinations. Proposition 4 shows that the NVAR$(p,1)$ can be written as a DFM, while a DFM with restricted factor dynamics can be written as an NVAR$(p,1)$ in the limit as $n \to \infty$.[22] Together with Eq. (9) above, it suggests that, at the cost of restricting factor dynamics, the NVAR allows the linear combinations of predictors to vary more flexibly across units, to the point that it naturally acommodates sparse factors as (linear combinations of) locally important units in the network.[23]

Proposition 4 (NVAR$(p,1)$-Factor Model Equivalence Result)**.**
Let y_t evolve as in Eq. (2). Let r be the rank of A. Then, we can write

$$y_t = \Lambda f_t + u_t \ , \quad \textit{with} \ \ f_t \in \mathbb{R}^r \ .$$

[22]The second part of Proposition 4 restricts $f_t \in \mathbb{R}^r$ to follow an NVAR$(p,1)$. This is less restrictive than it may seem; for $p = 1$, it requires $f_t \sim VAR(p)$, while under $r = 1$ it requires $f_t \sim AR(p)$. Also, note that, as usual, the factor representation is not unique, and we can re-scale A and ϕ to ensure $a_{ij} \in [-1,1]$.

[23]The DFM constructs $\{\mathbb{E}_{t-1}[y_{it}]\}_{i=1}^{n}$ as *different* ($\Lambda_{i\cdot}$) linear combinations of *the same* factors f_t, which in turn are linear combinations of observables. In contrast, the NVAR constructs $\{\mathbb{E}_{t-1}[y_{it}]\}_{i=1}^{n}$ as *the same* (α) linear combination of *different* covariates x_{it}, which in turn are linear combinations of observables.

Conversely, let $y_t = \Lambda f_t + \xi_t$*, with* $f_t \in \mathbb{R}^r$*. Assume* $f_t = \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t$*, with* $\Phi_l = \phi_l \Phi$ *for each* $l = 1 : p$ *and some* Φ*. Then, as* $n \to \infty$*, we can write*

$$y_t = \phi_1 A y_{t-1} + ... + \phi_p A y_{t-p} + u_t \ , \quad u_t = \Lambda \eta_t + \xi_t \ ,$$

where $A = \Lambda\Phi(W\Lambda)^{-1}W$ *and* W *is any* $r \times n$ *matrix with distinct rows.*

3 Dynamics via Lagged Network Effects: Inference

The NVAR can be used to estimate dynamic (lagged) network effects, as determined by α, the time profile of network effects. When the network A is estimated as well, it can also be used as a dimensionality-reduction technique useful for forecasting (cross-sectional) processes. In Section 3.1, I discuss inference on α, treating A as given. In Section 3.2, I discuss joint inference on (α, A).Details are in Section B.

3.1 Timing of Network Effects

NVAR($p,1$) The NVAR$(p,1)$ from Eq. (2) can be written as the linear regression in Eq. (9). Defining $\Sigma = \mathbb{V}[u_t]$, we obtain the Least Squares (LS) estimator for α:

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

As usual, under $u_t \sim N(0,\Sigma)$, it is also the (conditional) Maximum Likelihood (ML) estimator and the posterior-mean and -mode under a Uniform prior for $\alpha|\Sigma$. Under $\Sigma = I$, it yields the OLS estimator, which takes a "pooled" form:

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

Section B.1 establishes consistency and asymptotic Normality of $\hat{\alpha}_{OLS}$ under large T, large n and large (n,T) asymptotics. The derivations under large n assume that the observed network adjacency matrix A_n converges to some limit A_* so that, for example, $\frac{1}{n}\sum_{i=1}^{n}\left(A_{n,i\cdot} y_{t-l}\right)' u_{it} \xrightarrow{p} \mathbb{E}\left[\left(A_{*,i\cdot} y_{t-l}\right)' u_{it}\right]$.

We can estimate Σ standardly by $\hat{\Sigma}|\alpha = \frac{1}{T}\sum_{t=1}^{T} u_t u_t'$. The joint ML estimator $(\hat{\alpha}, \hat{\Sigma})$ is obtained by iterating on $\hat{\alpha}|\Sigma$ and $\hat{\Sigma}|\alpha$ until convergence (see Meng and Rubin (1993)).

NVAR(p,q) In case of an NVAR(p,q), as defined in Eq. (4), an estimator for α can be obtained by data augmentation.[24] However, point identification is not guaranteed. For

[24] The discussion holds likewise for the NVAR(p,q) from Eq. (5).

example, under $q = 2$ and $p = 1$, the observed process follows

$$y_t = \alpha_1^2 A^2 y_{t-1} + u_t \ , \quad \text{with} \ \ u_t = \tilde{u}_{2t} + \alpha_1 A \tilde{u}_{2t-1} \ ,$$

which suggests that α_1 is identified only up to sign. The identification problem is akin to estimating an AR(p) observed every q periods, discussed in Palm and Nijman (1984).[25] It is due to the fact that the mapping between the parameters α in the high-frequency process $\tilde{y}_\tau$ and the parameters in the observed process y_t is generally not bijective. For a general AR(p) and $q = 2$, the vector $(\alpha_1, \alpha_3, ...)$ is identified (jointly) up to sign (see Section B.1).

In the application in Section 4, the restrictions on α imposed by economic theory render α point-identified, as suggested by a unique posterior mode under Uniform priors. In other cases, one could follow the suggestion of Palm and Nijman (1984) and inform the estimation of α with a prior. This is facilitated by the clear interpretation of $\{\alpha_l\}_{l=1:p}$; it is the GIRF for units i and j that only share a first-order connection (see Section 2.2) and indicates how innovations transmit along a single link over time.

Conditioning on $\tilde{\Sigma}$, the posterior $p(\alpha|Y_{1:T})$ can be obtained using the Gibbs sampler of Carter and Kohn (1994). Treating the unobserved data in $\tilde{Y}_{1:T_\tau}$ as parameters, it iteratively draws from $p(\tilde{Y}_{1:T_\tau}|\alpha, Y_{1:T})$ and $p(\alpha|\tilde{Y}_{1:T_\tau})$ to obtain a sample from $p(\alpha, \tilde{Y}_{1:T_\tau}|Y_{1:T})$.[26] Under a Uniform prior for α, the resulting posterior mode of α converges to the ML estimator obtained using the Expectation-Maximization (EM) algorithm. To estimate $\tilde{\Sigma}$ as well, an additional iteration step is added to the Gibbs sampler. Using a uniform prior, we get an Inverse-Wishart conditional posterior for $\Sigma|\tilde{Y}_{1:T_\tau}, \alpha$ with mode $\frac{1}{T_\tau}\sum_{\tau=1}^{T_\tau} \tilde{u}_\tau \tilde{u}_\tau'$.

3.2 Joint Inference: Network & Effect-Timing

When interest lies in dynamic network effects, joint estimation of (α, A) is useful because network data may be difficult to collect or it may appear restrictive to condition the analysis on available network data. In addition, estimating (α, A) jointly enables us to use the NVAR as a dimensionality-reduction technique.

NVAR($p, 1$) The NVAR($p, 1$) from Eq. (2) can also be written as the linear regression

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

[25]It is also similar to estimating continuous time models using discrete time data (see e.g. Phillips (1973)).

[26]The particular state space model where in some periods τ no data is observed implies that in these periods the updating-step of the Kalman filter is skipped and the updated distribution of states equals their predicted distribution. See Section B.1, and see Schorfheide and Song (2015) for a discussion of the analogous case of a mixed-frequency VAR.

where $z_t = \sum_{l=1}^{p} \alpha_l y_{t-l} = \left[y_{t-1}, y_{t-2}, ..., y_{t-p}\right]\alpha$, and the $T \times n$ matrices Y, Z and U stack y_t, z_t and u_t along rows, respectively. To simplifty notation, I suppress the dependence of z_t on α and that of X_t on A.

To render (α, A) jointly identified, I normalize $\alpha_1 = 1$, with appropriate redefinitions of α as well as y_t, z_t and X_t. Relative to the alternative of restricting the norm of α, this normalization facilitates analytical calculations and asymptotic analysis, but it requires $\alpha_1 \neq 0$ in the true data generating process.[27]

Suppose data on a network B with elements b_{ij} is available. Under independent priors $a_{ij} \sim N(b_{ij}, \lambda_a^{-1})$, we obtain a matrix-variate Normal conditional posterior for A:

$$A \mid (Y, \alpha, \Sigma, B, \lambda_a) \sim MN\left(\bar{A}', \Sigma, \bar{U}_A\right) \ , \quad \bar{U}_A = \left[Z'Z + \lambda_a \Sigma\right]^{-1} \ , \ \ \bar{A} = \bar{U}_A \left[Z'Y + \lambda_a B'\Sigma\right] \ .$$

Its mean and mode, $\bar{A}'$, is the conditional optimizer for A under a LS objective function with a Ridge-penalty:

$$\bar{A}' = \arg\min_{A} \sum_{t=1}^{T} \left(y_t - A z_t\right)' \Sigma^{-1} \left(y_t - A z_t\right) + \lambda_a \sum_{i,j=1}^{n} (a_{ij} - b_{ij})^2 \ . \tag{12}$$

As $\lambda_a \to \infty$, we impose $A = B$. As $\lambda_a \to 0$, we ignore B and infer A from the data alone. No domain restrictions on A are imposed because any parameter value (α, A) can be rescaled to yield $a_{ij} \in [-1, 1] \ \forall \ i, j$, so that A can be interpreted as a network.[28]

Under a Laplace prior, the conditional posterior of A and its mode – the Lasso estimator – is available analytically when imposing $a_{ij} \geq 0$ and shrinking to $b_{ij} = 0$. We get

$$A \mid (Y, \alpha, \Sigma, \lambda_a) \sim MN\left(\bar{A}', \Sigma, \bar{U}_A\right) \ , \quad \bar{U}_A = \left(Z'Z\right)^{-1} \ , \quad \bar{A} = \bar{U}_A \left[Z'Y - \lambda_a \iota \iota' \Sigma\right] \ ,$$

truncated to $\mathbb{R}_+^{n^2}$. Drawing from this distribution or computing its mode is computationally feasible only for a diagonal Σ, which renders the distribution of each row i of A independent across i. Under $\Sigma = I$, we get

$$A_{i\cdot} \mid (Y, \alpha, \Sigma = I, \lambda_a) \sim N\left((\bar{A}')_{i,\cdot}, \bar{U}_A\right) \ , \quad \text{truncated to } \mathbb{R}_+^n \ .$$

A draw from this distribution is obtained using Gibbs sampling by iteratively drawing from the Normal conditional densities $a_{ij} \mid (A_{i,-j}, Y, \alpha, \Sigma = I, \lambda_a)$ for $j = 1 : n$. Its mode is computed by iterating on the latters' modes.[29]

---

[27] In case $a_{ij} \geq 0$ is restricted, it requires $\alpha_1 > 0$.

[28] To enforce $a_{ij} \in [0, 1]$ even under low λ_a, $a_{ij} \geq 0$ must be imposed. This leads to the high-dimensional Normal posterior being truncated to $\mathbb{R}_+^{n^2}$ and considerably complicates the analysis, as both computing the mode and drawing from this distribution is computationally intensive.

[29] Taken together, these two results mean that we can draw from the distribution of $A \mid (Y, \alpha, \Sigma, \lambda_a)$ or compute its mode by iterating on each column of A given all other columns.

Given Σ and λ_a, the joint posterior $p(\alpha, A|Y)$ is obtained by Gibbs sampling, iteratively drawing from the conditional posteriors $p(A|Y, \alpha, B, \Sigma, \lambda_a)$ and $p(\alpha|Y, A, \Sigma)$. To estimate Σ as well, an additional step is added to draw from $p(\Sigma|\alpha, A, Y)$. Under uniform priors for α and Σ, the posterior mode of $p(\alpha, A, \Sigma|Y)$ is equal to the GLS estimator $(\hat{\alpha}, \hat{A}, \hat{\Sigma})$ of the objective function in Eq. (12), obtained by iterating until convergence on the three respective conditional estimators. Fixing $\Sigma = I$, we obtain the OLS estimator of (α, A), for which consistency and asymptotic Normality under $T \to \infty$ are established in Section B.2. The choice of λ_a for predictive purposes as well as the possibility to construct B as a combination of multiple link-types is discussed in Section 5.

NVAR(p, q) As in the estimation of $\alpha|A$, if the network interaction frequency is higher than the observation frequency, joint inference on (α, A) can be conducted by relying on data augmentation. As before, identification is not guaranteed. In fact, relative to the estimation of $\alpha|A$ the problem is likely worsened, even if A may be tightly shrunk to a known (or sparsely parameterized) network B. However, the estimation of (α, A) in the application in Section 5 of this paper is judged based on forecasting performance, not its ability to deliver point identification.

4 Business Cycles by Lagged Input-Output Conversion

Long and Plosser (1983) show that time lags between the production of goods and their subsequent use as intermediaries for producing other goods can generate endogenous business cycles. In this application, I empirically quantify the importance of their proposed channel. I generalize their RBC economy with one-period lagged IOC by assuming that firms' production requires inputs produced in the past p periods. This leads to sectoral output growth evolving at some model-frequency as an NVAR$(p, 1)$, which translates into an NVAR(p, q) for some $q \in \mathbb{N}$ at my monthly frequency of observation. Thereby, A is the input-output matrix, u_t contains sectoral productivity processes, while $\{\alpha_l\}_{l=1:p}$ show how input-sourcing is spread out over the p periods. By estimating α, q and the persistence in u_t, I use the NVAR to quantify the extent to which business cycles in this framework are due to lagged IOC as opposed to persistence in exogenous productivity processes.

After theoretically motivating the analysis in Section 4.1, I discuss the setup, data and estimation in Section 4.2. Section 4.3 presents the results. Details are in Section C.

4.1 Theory

The following analysis is based on Carvalho and Tahbaz-Salehi (2019) – who discuss a static economy – and the Appendix to Acemoglu et al. (2016). Derivations are in Section C.1.

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

$$\Pi_{i\tau} = p_{i\tau} y_{i\tau} - w_\tau l_{i\tau} - \sum_{j=1}^{n} p_{j\tau} x_\tau^{ij} \ , \quad \text{with } \ y_{i\tau} = z_{i\tau} l_{i\tau}^{b_i} \prod_{j=1}^{n} x_{ij\tau}^{a_{ij}} \ ,$$

where $b_i > 0$, $a_{ij} \geq 0$ and $b_i + \sum_{j=1}^{n} a_{ij} = 1$. $z_{i\tau}$ denotes Total Factor Productivity (TFP) in sector i, $\{p_{i\tau}\}_{i=1}^{n}$ are the prices of the n different goods, and w_τ is the price of labor. $x_{ij\tau}$ is the amount of good j used in the production at time τ. As discussed below, it can differ from the amount of good j purchased in period τ, x_τ^{ij}.

In this environment, prices are entirely determined by supply. To characterize output, I assume the presence of a representative household who supplies one unit of labor inelastically and exhibits log-preferences over the n goods:

$$\underset{\{c_{i\tau}\}_{i=1}^{n}}{max} \ \sum_{i=1}^{n} \gamma_i \ ln(c_{i\tau}/\gamma_i) \ , \quad \text{s.t.} \ \sum_{i=1}^{n} p_{i\tau} c_{i\tau} = w_\tau \ ,$$

where $\sum_{i=1}^{n} \gamma_i = 1$. The first-order condition (FOC) yields $c_{i\tau} = \gamma_i \frac{w_\tau}{p_{i\tau}}$.[30] This result holds even if households have access to a storage technology, as market clearing under representative households in a closed economy implies that households spend their whole period τ income, w_τ, on consumption.

Different assumptions on the timing of IOC lead to different dynamics of sectoral prices and output in this economy. Typically, it is assumed that inputs are converted into outputs in the same period when they are purchased, i.e. $x_{ij\tau} = x_\tau^{ij}$. Dropping time subscripts in this static environment, define $\tilde{y} = ln(y)$, $y = (y_1, ..., y_n)$ and analogously for $\tilde{z}$. In equilibrium, sectoral output satisfies

$$\tilde{y} = k^y + A\tilde{y} + \tilde{z} \ , \tag{13}$$

where k^y is a vector of constants, and $a_{ij} = p_j x^{ij}/(p_i y_i)$ is value of good j purchases by sector i as a fraction of the value of sector i's output. In this environment, while the network A amplifies idiosyncratic TFP shocks and therefore affects the variance of $\tilde{y}$, any autocorrelation in $\tilde{y}$ is inherited from that in $\tilde{z}$ (see e.g. Carvalho (2010); Acemoglu et al. (2012)).

To analyze dynamics under lagged IOC, I assume perfect foresight.[31] If, as in Long and Plosser (1983), it takes one period to convert purchased inputs into output, then $x_{ij\tau} = x_{\tau-1}^{ij}$

[30] Hence, γ_i is the share of good i in households' expenditures.

[31] As discussed in Fan et al. (2023), this assumption is standard for modeling dynamic spatial economies, which are closely related to dynamic network economies.

and sectoral output approximately follows an NVAR(1, 1):

$$\tilde{y}_\tau = k^{y1}_\tau + A\tilde{y}_{\tau-1} + \tilde{z}_\tau \ ,$$

where the time variation in k^{y1}_τ is due to time variation in the numéraire w_τ. So long as $\beta < 1$, in steady state, prices are higher and output is lower than in the economy with contemporaneous IOC. Also, the meaning of $a_{ij} = \beta^{-1} p_j x^{ij}/(p_i y_i)$ changes slightly. These differences vanish as $\beta \to 1$. More importantly, while the former, static economy is always in steady state, the economy with one period-lagged IOC features transition dynamics; after a disturbance to $\tilde{z}_\tau$, $\tilde{y}_\tau$ only asymptotically converges to steady state.[32] It is due to these transition dynamics that this framework can generate endogenous business cycles, i.e. autocorrelation in $\tilde{y}_\tau$ even in absence of autocorrelation in $\tilde{z}_\tau$.

To take the economy with lagged IOC to the data, I generalize the lag length by assuming that firms require inputs produced in the past p periods for production. For ease of exposition, let $p = 2$. Let $x_{ij\tau}$ aggregate quantities of input j purchased at different periods in the past using a Constant Elasticity of Substitution-aggregator:

$$x_{ij\tau} = \left[\alpha_1 (x^{ij}_{\tau,\tau-1})^r + \alpha_2 (x^{ij}_{\tau,\tau-2})^r\right]^{1/r} \ , \quad \alpha_1, \alpha_2 \geq 0 \ , \quad \alpha_1 + \alpha_2 = 1 \ ,$$

where $x^{ij}_{\tau,\tau-h}$ denotes the use of good j purchased at time $\tau - h$ in the production of good i at time τ.[33] This shortcut stands for frictions like delivery costs and -lags and storage capacity constraints (Khan and Thomas, 2007; Alessandria et al., 2010; Liu and Tsyvinski, 2024; Antràs and Tubdenov, 2025).[34] In the Cobb-Douglas case $r \to 0$, sectoral output approximately follows an NVAR(2,1):

$$\tilde{y}_\tau = k^{y2}_\tau + \alpha_1 A\tilde{y}_{\tau-1} + \alpha_2 A\tilde{y}_{\tau-2} + \tilde{z}_\tau \ , \tag{14}$$

where once again k^{y2}_τ varies over time only to the extent that the numéraire w_τ changes in value. Under a more general elasticity of substitution $r \in [0, 1)$,[35] the analogous result is obtained by log-linearizing around the steady state (see Section C.1). This specification nests the one period-lagged economy, which is obtained under $\alpha_2 = 0$. Relative to that case, $\alpha_2 > 0$ increases prices and decreases output in steady state, provided that $\beta < 1$. Also, we have $a_{ij} = \left[\beta\alpha_1 + \beta^2\alpha_2\right]^{-1} (p_j x^{ij})/(p_i y_i)$. As $\beta \to 1$, the links a_{ij} retain their interpretation as the output shares of different inputs j in the production of good i. The parameters (α_1, α_2) show the shares of an input j purchased at different periods in the past in the overall usage

[32] Relatedly, as discussed in Section 2.2, the response of $\tilde{y}_{(\tau)}$ to a change in $\tilde{z}_{(\tau)}$ in the static economy corresponds to the long-term response of $\tilde{y}_\tau$ to a permanent change in $\tilde{z}_\tau$ in this dynamic economy (disregarding the slightly changed meaning of a_{ij}).

[33] This means that a good perishes after two periods (with regard to its suitability as an input in production). Therefore, the amount of good j purchased at time τ can be used in production at periods $\tau + 1$ and $\tau + 2$: $x^{ij}_\tau = x^{ij}_{\tau+1,\tau} + x^{ij}_{\tau+2,\tau}$.

[34] As in the Long and Plosser (1983)-economy above, the presumption is that storage is done by the buyer.

[35] This notably excludes complementarity ($r < 0$) and perfect substitutability ($r = 1$).

of input j in the production of good i. Their homogeneity means that the time profile of input-sourcing is assumed to be constant over time and across input-output pairs (i, j). The restrictions $\alpha_1, \alpha_2 \geq 0$, $\alpha_1 + \alpha_2 = 1$ and $\sum_{j=1}^{n} a_{ij} < 1 \; \forall \; i$ imply that $\tilde{y}_\tau$ is stationary as long as $\tilde{z}_\tau$ is.[36]

The theory above does not restrict the process of (log) sectoral TFP $\tilde{z}_{i\tau}$. In general, it may display secular growth and persistent shocks of both aggregate and idioyncratic nature. Consider a difference-stationary specification with aggregate and idiosyncratic TFP processes e^a_τ and $e_{i\tau}$ evolving as AR(1) processes as in Foerster et al. (2011):

$$\Delta \tilde{z}_{i\tau} = \gamma_i + \delta_i e^a_\tau + e_{i\tau} \; ,$$

$(1 - \rho_a L) e^a_\tau = \varepsilon^a_\tau \sim WN$ and $(1 - \rho_i L) e_{i\tau} = \varepsilon_{i\tau} \sim WN$. Under contemporaneous IOC (Eq. (13)), this leads to

$$\Delta \tilde{y}_\tau = (I - A)^{-1} \left[\gamma + \delta e^a_\tau + e_\tau \right] \; . \tag{15}$$

We obtain IRFs $\partial \Delta \tilde{y}_{i,\tau+h} / \partial \varepsilon_{j,\tau}$ and $\partial \Delta \tilde{y}_{i,\tau+h} / \partial \varepsilon^a_\tau$ akin to those in Eq. (7) in Section 2.3. This points to two sources of persistence in output growth: persistence in the aggregate TFP process e^a_τ and persistence in idiosyncratic TFP processes $\{e_{i\tau}\}_{i=1}^{n}$. The role of input-output links is limited to amplification; the response of output growth in sector i to a TFP shock in sector j is scaled by element (i, j) of the Leontief-inverse $(I - A)^{-1}$, as in Carvalho (2010, Fig. 7 and 8). By summing up connections of all order from i to j, the latter shows the importance of sector j in sector i's supply chain.

Lagged IOC can be an additional source of persistence. Under Eq. (14), we get

$$\Delta \tilde{y}_\tau \approx \gamma + \alpha_1 A \Delta \tilde{y}_{\tau-1} + \alpha_2 A \Delta \tilde{y}_{\tau-2} + \delta e^a_\tau + e_\tau \; , \tag{16}$$

which leads to IRFs $\partial \Delta \tilde{y}_{i,\tau+h} / \partial \varepsilon_{j,\tau}$ and $\partial \Delta \tilde{y}_{i,\tau+h} / \partial \varepsilon^a_\tau$ akin to those in Eq. (8) in Section 2.3. As discussed in Section 2.3, lagged IOC and TFP shocks' autocorrelation imply distinct dynamics; α relates the timing at which a sector's output is first impacted by a TFP shock in another sector and the strength of the ensuing response to network connections of different order, whereas ρ_a and $\{\rho_j\}_{j=1:n}$ induce an exponentially decaying response after every round of networked transmission.

Under a difference-stationary specification for log sectoral TFP $\tilde{z}_\tau$, TFP shocks have temporary effects on output growth, but persistent effects on output levels. By Corollary 4, the long-term response of (log) output to a TFP shock – equal to the cumulative response of output growth – is the same under contemporaneous and lagged IOC, in line with the fact

[36] Berman and Plemmons (1994, ch. 2) 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.

that any TFP level yields the same steady state in both economies (disregarding differences in A between the two).

To take the processes in Eq. (15) and Eq. (16) to the data, one has to take a stance on what a period in the theoretical models above signifies. Let $\{\Delta y_t\}_{t=1}^T$ be observed output growth. As it is a flow variable, the model-frequency must be either equal to the observational frequency or an integer-multiple thereof (see Section 2.2.3):

$$\Delta y_{\tau/q} = \Delta\tilde{y}_\tau + ... + \Delta\tilde{y}_{\tau-q+1} \quad \text{for} \;\; \tau/q = 1, ..., T \;, \tag{17}$$

and for $q \in \mathbb{N}$. Other things equal, under a higher q the economy approaches faster the new steady state level of output following a TFP shock. Relatedly, under lagged IOC, it also means that the IRFs at any single horizon increasingly depend on higher-order connections, and it leads to network-induced cross-sectional correlation in innovations at observational frequency, whereas for $q = 1$ the correlation is entirely due to aggregate TFP shocks.[37] As the meaning of q depends on the frequency of observation, its choice is discussed in Section 4.2.

4.2 Application-Setup, Data & Estimation

I quantify the relative contributions of these two (three) drivers of aggregate persistence as seen through the lens of the theory of real business cycles with contemporaneous and lagged IOC. To do so, I estimate the state space models characterized by Eqs. (15) and (17) and Eqs. (16) and (17), respectively, based on industrial production growth across US manufacturing sectors, while calibrating the links a_{ij} using input-output data. The analysis first seeks to determine whether there is a role for lagged IOC at all by comparing the data fit of specifications with lagged IOC to that with contemporaneous IOC based on model selection criteria. Presuming that one of the former is preferred, the role of lagged IOC can be quantified by computing the change in the autocorrelation implied by the estimated model with lagged IOC when the persistence in TFP shocks is set to zero.

I compute log differences of monthly industrial production (IP) indices across 23 manufacturing and mining sectors in the US economy provided by the Federal Reserve Board.The indices are available from January 2005 through August 2022. To eliminate seasonal patterns, I regress each series on month-dummies, take the residuals and add back the mean. To construct A, I use annual data on input-output (IO) matrices provided by the Bureau of Economic Analysis (BEA). I take the input-output matrix for 2010. Following the theory in Section 4.1, links a_{ij} are calibrated as

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

[37] See Section C.1 as well as Section 2.2.3 for more detailed discussions of these points. In the limit as $q \to \infty$, the long-term response referenced in Proposition 6 – a function of all connection-orders – materializes after a single observational period.

where $sales_{j \longrightarrow i}$ is the total value of goods and services purchased by sector i from sector j as determined by the corresponding entry in the BEA's "use" table.[38] The value of a_{ij} shows how many dollars worth of output of sector j sector i needs to purchase in order to produce one dollar's worth of its own output. I abstract from the differences in the calibration of a_{ij} between contemporaneous and lagged IOC, which renders the analysis valid as $\beta \rightarrow 1$. More importantly, the calibration assumes that firms' input shares reported for the course of a year are equal to those at higher frequency intervals. Details on the matching of IP and IO data is provided in Section C.2.

For the specification with lagged IOC, I consider $p = \{1, 2, 3, 4, 5, 6\}$ and $q \in \{1, 2, 3\}$, implying network interaction frequencies of a month, two weeks and 10 days. Under contemporaneous IOC, I take $q = 1$, as it already refers to the limit case of an infinitely high frequency of network interactions. In both cases, I consider Normal AR(1) processes for idiosyncratic and aggregate TFP: $(1 - \rho_a L)e^a_\tau = \varepsilon^a_\tau \sim N(0, \sigma^2_a)$ and $(1 - \rho_i L)e_{i\tau} = \varepsilon_{i\tau} \sim N(0, \sigma^2_i)$. To separately identify both TFP processes, I normalize $\sigma^2_a = 1$ and $\delta_1 = 1$, re-defining δ. To accommodate the restrictions $\alpha_l \geq 0 \;\; \forall \, l$ and $\sum_{l=1}^{p} \alpha_l = 1$, I drop α_p from α and impose the domain restrictions $\alpha_l \in [0, 1]$ for $l = 1 : p - 1$ and $\sum_{l=1}^{p-1} \alpha_l \leq 1$. Under lagged IOC, this yields $p - 1 + 4n$ unknown parameters: $\theta = (\alpha', \gamma', \delta', \rho_a, \rho', \sigma^{2\prime})'$, where ρ and σ^2 stack $\{\rho_i\}_{i=1}^n$ and $\{\sigma^2_i\}_{i=1}^n$, respectively. Under contemporaneous IOC, α is dropped from θ, leading to $4n$ parameters.[39]

The inference from Section 3 is not applicable because of the autocorrelation and factor structure of TFP processes and due to the restrictions on α under lagged IOC. The likelihood of both models can be evaluated with a Kalman filter, as stated in Section C.3. I consider Bayesian inference on θ under uniform priors on the respective domains.[40] As a result, the posterior mode equals the Maximum Likelihood estimator. The posterior is obtained numerically using a Sequential Monte Carlo algorithm, which – as a by-product – estimates the marginal likelihood and therefore enables model selection.

4.3 Results

Table 1 reports the Marginal Data Density (MDD) for different specifications of the model with lagged IOC. For comparison, its value under contemporaneous IOC is -11,590. This number is beaten by all but a few lagged IOC specifications with $q = 1$. The most preferred specification features $q = 2$ and $p = 5$, i.e. bi-weekly network interactions where innovations travel 2.5 months along single input-output links. The subsequent analysis is based on this preferred specification.

[38]This is in line with the literature. See Acemoglu et al. (2016) for example. Carvalho and Tahbaz-Salehi (2019) discuss the IO data in more detail.

[39]With $n = 116$ sectors and $T = 211 - p$ periods available for estimation, this yields an observations-to-parameters ratio of 52.5 under $p = 1$ and 50.7 under $p = 6$.

[40]Specifically, $\rho_i, \rho_a \in [0, 1)$ and $\sigma^2_i, \sigma^2_a \in \mathbb{R}_{++}$ for $i = 1 : n$, $\delta \in \mathbb{R}^n$ and $\alpha \in [0, 1]^{p-1} \cap \{\alpha : ||\alpha||_1 \leq 1\}$.

Table 1: Model Selection: Log MDD

		p					
		1	2	3	4	5	6
q	1	-12058	-11975	-11680	-11466	-11450	-11729
	2	-11223	-11509	-11317	-11180	-11109	-11353
	3	-11257	-11174	-11406	-11185	-11204	-11219

Notes: The table shows the log Marginal Data Density (MDD) across model specifications. The values for q (from top to bottom) refer to network interaction frequencies of a month, two weeks and 10 days, respectively. The log MDD under contemporaneous IOC is -11590.

Fig. 3 illustrates the composition of impulse responses, analogously as Fig. 2 does for two units that share only a first- or second-order connection. The top-left panel shows the estimated temporal propagation of TFP shocks along supply chain linkages of different order. Remarkably, within the first two months after the shock, the impact is mostly limited to direct customer-sectors. After that, the shock spreads somewhat more quickly to higher-order connections.

The top-right panel shows the strength of network connections of different order from the sector "Fabricated Metal Products" to the sectors "Mining (except oil and gas)" and "Chemical Products", respectively. Firms producing fabricated metals depend on chemical products directly as well as indirectly in their supply chain. In contrast, they rely on mining products only indirectly, though this higher-order dependence is of a similar magnitude.

The lower panels of Fig. 3 illustrate the resulting impulse responses to a respective one standard deviation idiosyncratic TFP shock to Chemical Products and Mining. As a result of its stronger direct reliance on chemicals, the response of the growth in the production of fabricated metals to a TFP shock in the chemical sector materializes much faster than does the response to a TFP shock in mining. Yet, the two are of a similar magnitue at their peaks.

The responses refer to percentage point increases in sectoral output growth. To interpret the magnitudes, recall that the data is in monthly frequency and that the (persistent) response of the level of sectoral output is obtained by summing up the illustrated (transitory) response of output growth. Expectedly, the magnitudes are rather small, as they refer to responses of sectoral output to idioyncratic TFP shocks in a single other sector.

Larger responses are obtained when considering aggregate TFP shocks. On top of their direct effect on the output growth of all sectors, the latter have an indirect effect, as the supply chain network amplifies initial effects. Fig. 4 shows the respones of output growth in the sectors "Oil and Gas Extraction" and "Machinery" to a one standard deviation shock to aggregate TFP. The oil and gas extraction sector is positioned rather at the top of supply

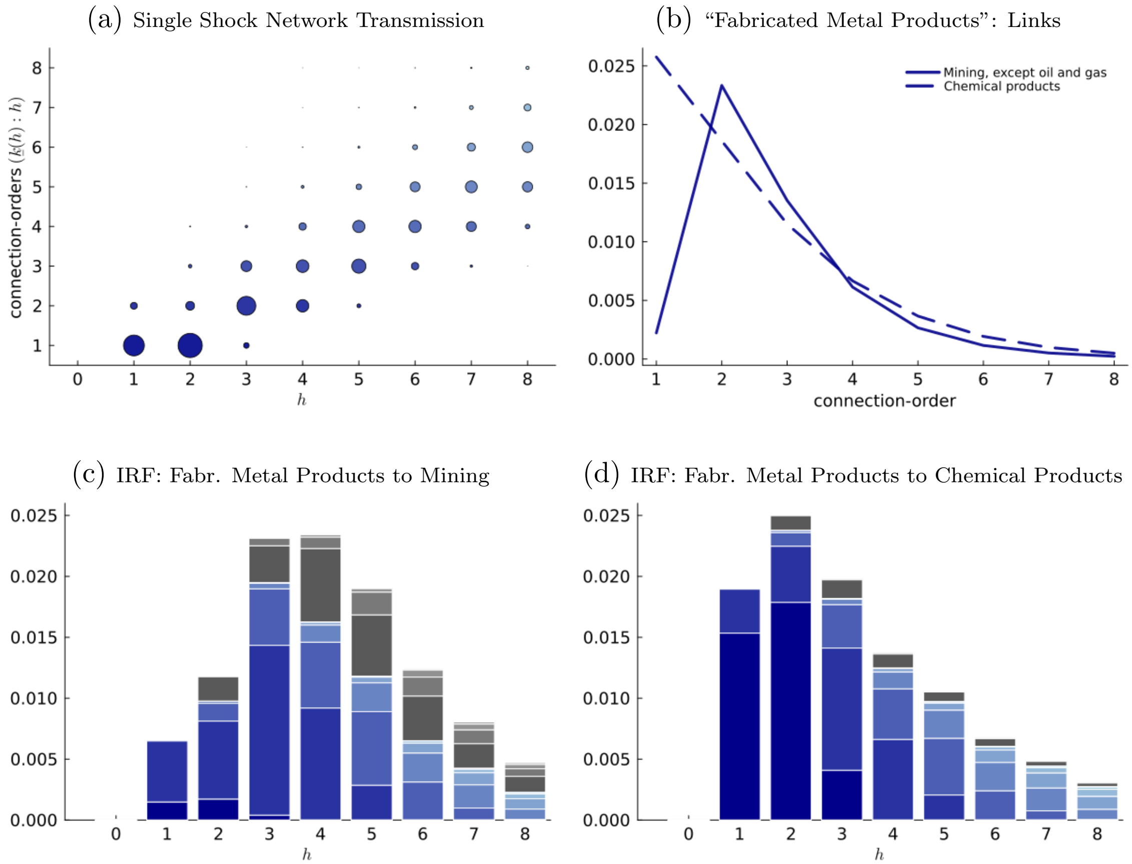

Figure 3: Impulse Responses to Idiosyncratic TFP Shocks

Notes: The top left panel depicts the coefficients $c^h_{\underline{k}(h)} : c^h_h$ from Eq. (8) for $h = 1 : 6$, illustrating how a single shock transmits along network connection-orders (y-axis) over time (x-axis; horizons). The dashed lines show the bounds on which connection-orders can matter in an NVAR(2, 1): $h \in \underline{k}(h) : h$. The top right panel shows the supply chain connections of different order from "Fabricated Metal Products" to "Mining (except oil and gas)" and "Chemical Products", and the bottom panels show the resulting IRFs to a TFP shock of one standard deviation.

chains. Due to its weak reliance on other sectors as suppliers, its response to aggregate TFP shocks is only weakly amplified by supply chain connections. In contrast, after a similar initial response, the machinery sector experiences a hump-shaped response due to second-order transmission operating via its supplier-sectors. The relevant network-quantity for these IRFs is a sector's weighted reliance on all other sectors as suppliers, with weights given by their exposures to the aggregate TFP shock, δ (see Eq. (16)).

A similar reasoning explains the differing IRFs of aggregate output growth to sectoral TFP shocks. Fig. 5 shows these responses to TFP shocks in "Mining (except oil and gas)" and "Chemical Products", respectively. As the chemical sector sits on the top of supply chains, it leads to a much more persistent increase in aggregate industrial production than does an increase in the TFP in the mining industry. The relevant network-quantity for these IRFs is all sectors' weighted reliance on the particular sector in question as a supplier, with weights given by sectors' contribution to aggregate output growth.

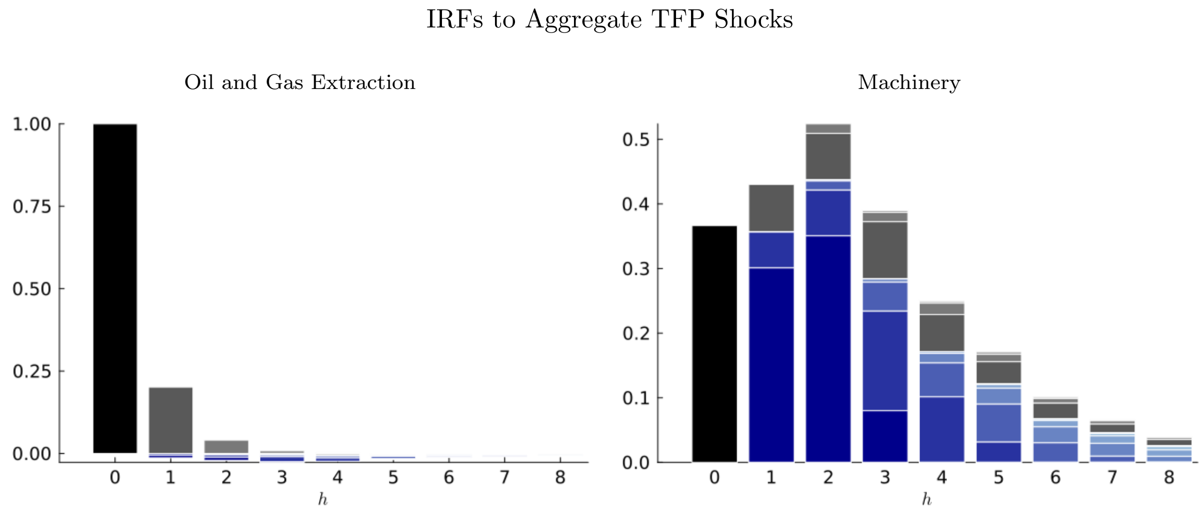

Figure 4: Impulse Responses to Aggregate TFP Shocks

Notes: The left panel shows the impulse-response of the production growth in "Oil and Gas Extraction" to a one standard deviation aggregate TFP shock. The right panel does the same for the sector "Machinery".

Existing studies use a static framework of contemporaneous input-output conversion to show that the effects of sectoral TFP shocks on aggregate output are stronger for sectors with more central positions in the supply chain network. By Proposition 6, the present analysis leads to the same long-run effects, but it sheds light on the transition dynamics. The left panel of Fig. 6 shows the time profiles of the response of aggregate industrial production to TFP shocks in different sectors. It illustrates that, due to different positions in the supply chain network, under lagged IOC sectors differ not only in terms of the strength of their impact on aggregate output, but also by its timing. Sectors at the bottom of supply chains, such as the food and beverage sector, have a much more immediate effect on aggregate output than sectors that act as important suppliers to other sectors in the economy.

Although stronger effects tend to take more time to realize, there is no clear relationship between the strength and timing of the response of aggregate output to sectoral TFP shocks. This is illustrated by the right panel of Fig. 6. For example, a TFP shock in the food and beverage sector has a similar long-term effect on aggregate output as a shock to primary metals, yet the latter materializes much more sluggishly. One month after the TFP shock in the primary metals sector, a similar fraction of the long-term effect on aggregate output has materialized as in the case of a TFP shock to mining support activities, yet the latter are estimated to lead to a stronger long-term effect.

The autocorrelation of aggregate output growth at the posterior mean is estimated to be 0.389. In a hypothetical environment without persistence in exogenous shocks, this number drops to 0.237. As a result, lagged IOC can account for about two thirds of aggregate persistence. Together with persistence in the aggregate TFP process, this number increases to 0.365. In contrast, if only persistence in sectoral TFP processes is added, the autocorrelation

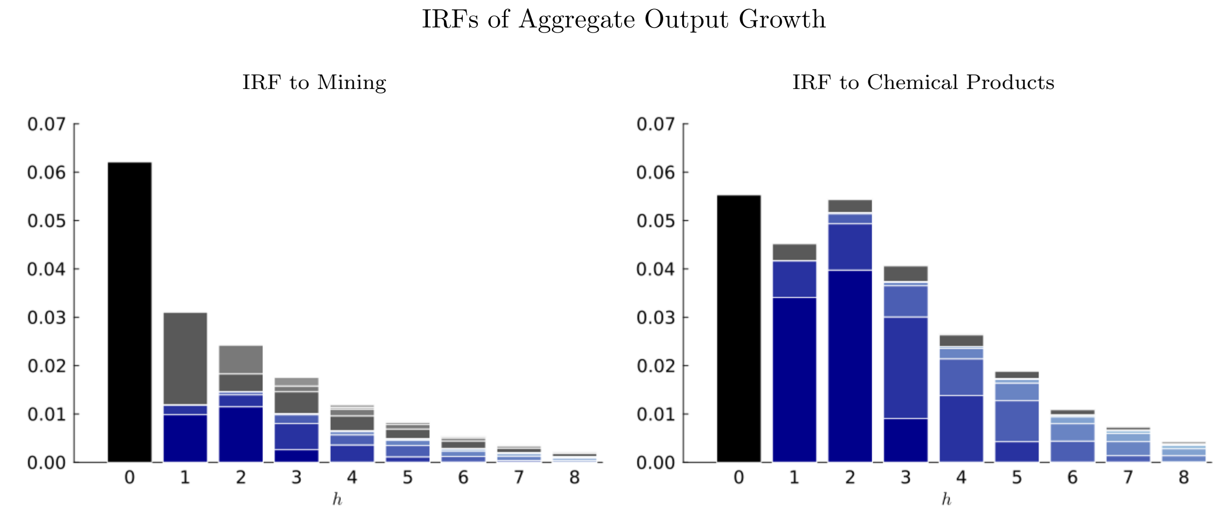

Figure 5: Impulse Responses of Aggregate Output Growth

Notes: The left panel shows the impulse-response of the aggregate growth in industrial production to a one standard deviation shock to the TFP in "Mining (except oil and gas)". The right panel does the same for the response to a TFP shock in "Chemical Products".

increases only to 0.276. Overall, these results indicate that lagged IOC and a single, persistent aggregate TFP process can account wel for business cycles in this RBC environment.

5 Dimensionality-Reduction by Innovation Transmission through Parsimonious Networks

In Section 4, a process y_t is driven by an observed network, and we quantify how network effects materialize over time. In this section, I forecast a rather high-dimensional series y_t and, for the most part, I assume that no network data is available. When (α, A) are jointly estimated and the estimation of A is regularized, the NVAR becomes useful as a dimensionality-reduction technique. It rationalizes the dynamic comovement among all $\{y_{it}\}_{i=1:n}$ by the dynamic innovation transmission along a few bilateral links among units $i = 1 : n$.

In Section 5.1, I discuss the potential of the NVAR to reduce dimensionality, building on Section 3.2. Section 5.2 then sets up the application to forecast macroeconomic aggregates across countries, and Section 5.3 presents the results. Details are in Section D.

5.1 NVAR-Estimation for Dimensionality-Reduction

Even for intermediate n, an unrestricted VAR(p) poorly forecasts $y_t \in \mathbb{R}^n$. Relative to it, under $p > 1$, the NVAR$(p, 1)$ reduces the number of parameters in the autoregressive matrices from pn^2 to $n^2 + p - 1$, owing to the assumption that innovations transmit cross-sectionally

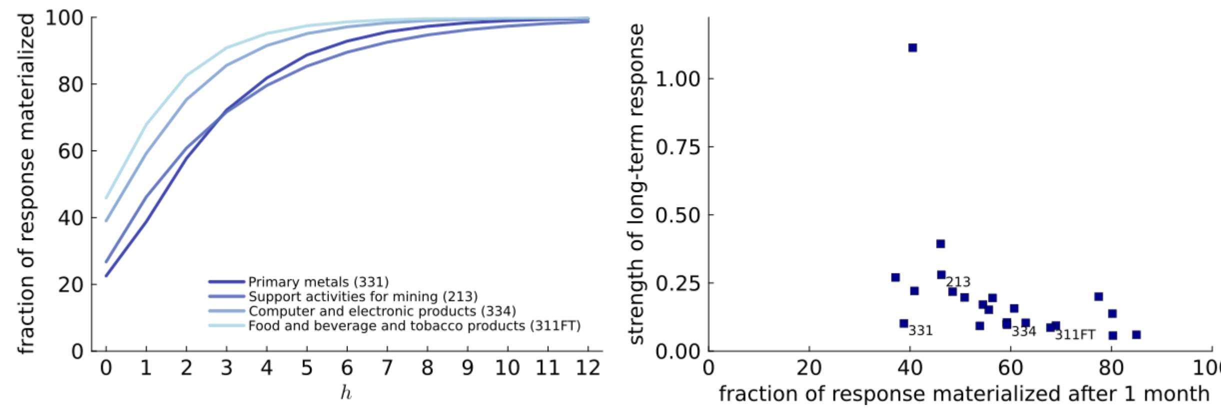

Figure 6: Size and Timing of IRFs of Aggregate Output Growth

Notes: The left panel shows the time profile of the effect of sectoral TFP shocks on aggregate industrial production for a few selected sectors. The right panel relates the strength of the effects to their timing. The shock sizes are equal to one standard deviation of the respective sectoral TFP shock.

only via bilateral links; rather than freely relating y_{it} to $\{y_{j,t-l}\}_{j=1:n,l=1:p}$, it uses the n links $\{a_{ij}\}_{i=1:n}$ to compress this information to $x_{i,t-l} = \sum_{j=1}^{n} a_{ij} y_{j,t-l}$, and in turn the $p-1$ free parameters in α determine the importance of different lags of x_{it} for the dynamics of y_{it}.[41] If, in addition, A is estimated parsimoniously, then the NVAR leverages both approaches available to address the large parameter problem in the Wold representation (Geweke, 1984): it reduces the number of parameters and applies shrinkage priors (regularization). Regularization of A is motivated by the sparseness of real world-networks across a wide range of applications and by the fact that the dynamic comovement of two series y_{it} and y_{jt} can be captured with higher-order connections between i and j without a direct link between them (see Propositions 1 and 2).

Section 3.2 discussed the estimation of A under a Normal prior (L2-penalty) and under an Exponential prior (L1-penalty and restricting $a_{ij} \geq 0$). In the following, these two approaches are labeled NVAR-R and NVAR-L. The joint posterior of (α, A) is obtained by iteratively drawing from the conditional posteriors of $\alpha|A$ and $A|\alpha$, the joint posterior mode – the Ridge-/Lasso-regularized LS-/ML-estimator – by iterating on the conditional posterior modes.

When the NVAR-R and NVAR-L are applied for predictive purposes, selecting the hyperparameter λ_a – i.e. the degree of shrinkage – becomes particularly important. Following Giannone et al. (2015), setting a hyperparameter to its marginal posterior mode (MPM) under a Uniform hyperprior maximizes the marginal data density (MDD) and, hence, the one-step ahead predictive ability. Under NVAR-R and NVAR-L, respectively, we obtain the

[41] Even though $\mathbb{E}_{t-1}[y_{it}]$ contains the same linear combinations of $\{y_{j,t-l}\}_{j=1:n}$ at all lags $l = 1:p$, dynamics at higher horizons h are driven by different linear combinations, since higher-order network connections accumulate over time (see Proposition 1).

conditional posteriors

$$\lambda_a \mid (A, B) \sim G\left(\frac{n^2+2}{2}\ ,\ \frac{tr\left[(A-B)'(A-B)\right]}{2}\right) \quad \text{and} \quad \lambda_a \mid A \sim G\left(n^2+1\ ,\ \iota' A \iota\right)\ .$$

We get the joint posterior of (α, A, λ_a) by augmenting the Gibbs sampler from Section 3.2 with a step to draw λ_a from this conditional posterior. In turn, we arrive at the marginal posterior of λ_a by simply considering its posterior draws in isolation.

While well-grounded in theory, the approach of Giannone et al. (2015) is computationally expensive in the present environment, as it requires obtaining the full posterior.[42] For NVAR-L, a heuristic approach is to maximize the Bayesian Information Criterion (BIC) suggested by Zou et al. (2007) for other Lasso-applications. This involves counting the number of non-zero elements in $\hat{A}(\lambda_a)$. For NVAR-R, an analogous criterion is the conditional MDD $p(Y|\lambda_a, B, \alpha, \Sigma)$, which is derived in Section D.1 and can be maximized when evaluated at $\hat{\alpha}$ (and $\hat{\Sigma}$).

5.2 Forecasting-Setup

To validate the NVAR's merit as a dimensionality-reduction technique, I forecast a range of macroeconomic time series across countries and compare its performance to that of the Dynamic Factor Model (DFM) of Geweke and Zhou (1996):

$$y_t = \Lambda f_t + u_t\ , \quad f_t = \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t\ , \tag{18}$$

where $u_t \sim N(0, I_n)$, $\eta_t \sim N(0, I_r)$ and $\Lambda_{1:r,\cdot}$ is lower-triangular with positive diagonal elements. The equivalence result in Proposition 4 suggests that, relative to the DFM, the NVAR restricts factor dynamics, but allows the linear combinations of predictors to vary more flexibly across units and acommodates sparse factors as (linear combinations of) locally important units in the network. Therefore, the NVAR is expected to capture cross-sectional dynamics in finite samples better than the factor model in the presence of many sparse factors or, equivalently, in case of a sparse, yet close-to-full-rank network adjacency matrix. Intuitively, this is the case of dynamics driven by many bilateral links rather than a few influential units. More generally, Proposition 4 suggests the NVAR could improve upon the poor forecasting performance of factor models under a high dispersion of factor loadings across series (see Boivin and Ng (2006)).[43]

[42] This is particularly disadvantageous if only the posterior mode estimator is of interest. Even if the full posterior of (α, A) evaluated at the optimal λ_a is desired, however, it is a drawback, as it requires conducting posterior sampling twice. The disadvantage disappears only if one is interested in the marginal posterior of (α, A) under a Uniform prior for λ_a (model-averaging rather than -selection).

[43] This notably includes the case of sparse factors, i.e. loading-vectors with zero and non-zero entries. As Boivin and Ng (2006) point out, selecting the number of factors separately for each series is a poor remedy: series that depend on less dominant factors are still poorly forecasted, as including more estimated factors

To investigate these hypotheses, I use the NVAR and DFM to forecast monthly industrial production (IP) growth, monthly CPI inflation and quarterly real GDP growth across OECD-member countries, -applicants and -partners.[44] The series are obtained from the IMF's IFS database, and YoY growth rates are used. I limit attention to the pre-COVID periods 2001:M1 - 2019:M12 and 2001:Q1 - 2019:Q4, respectively, and I delete countries with missings or more than two consecutive constants. This yields datasets with $n = 42$, $n = 34$ and $n = 39$ observations. The series are de-seasonalized by subtracting fitted values from a regression on period-dummies, and the resulting series are standardized.

Comparison is based on mean squared errors (MSEs) at different forecasting horizons, whereby the mean is taken over countries as well as forecasting origins. I consider expanding windows with 24 and 16 origins, respectively, ranging from 2017:M12 to 2019:M11 and from 2015:Q4 to 2019:Q3. To reduce the computational burden, the models are fully estimated only at the first origin, after which parameters are fixed and, if required, hidden states are re-estimated by conditioning on these initially estimated parameters. At the first origin, we have $T = 203$ and $T = 60$ observations, respectively.

I limit the analysis to point forecasts obtained under the posterior mode of (α, A) in the NVAR-R and NVAR-L as presented in Section 3.2 – the Ridge-/Lasso-regularized ML-/LS-estimator –, and the posterior mode of (Λ, Φ), $\Phi = [\Phi_1, ..., \Phi_p]'$ under a Uniform prior in the DFM – the ML-/LS-estimator (see Section D.3). For the DFM, I consider $p = 1 : 4$ lags and $r = 10$ factors. In turn, I select the ex-post best-performing specification that minimizes the cumulative MSE over the first three horizons, and I compare all NVAR specifications against this benchmark. This renders the assessment of the NVAR's suitability for forecasting independent of methods to select the number of factors ex-ante. For the NVAR, I also consider $p = 1 : 4$ lags, and I mostly focus on $q = 1$ – the case referred to by Proposition 4 . In both NVAR models, I set $\Sigma = I$, and, unless otherwise stated, I shrink A to $B = 0$. The degree of shrinkage, as embodied by λ_a, is chosen based on two methods. The first takes the MPM and, therefore, maximizes the MDD exactly, while the second approximates this choice by using grid-search to maximize the BIC (in case of the NVAR-L) or the conditional MDD (for the NVAR-R).

5.3 Results

Fig. 7 illustrates the results for IP growth and NVARs with $q = 1$. The ex-post best DFM features $r = 4$ factors and $p = 1$ lags. As shown by the black line, it reduces the MSE of one-step ahead forecasts by 13% relative to forecasting the unconditonal mean of zero. As expected, this improvement vanishes for higher horizons. The NVAR-R (blue

adds noise to the forecasts.

[44]As of August 2024, this includes 45 countries. On top of 34 OECD members, there are 8 applicants (Argentina, Brazil, Bulgaria, Croatia, Indonesia, Peru, Romania and Thailand) and 3 partners (China, India, South Africa).

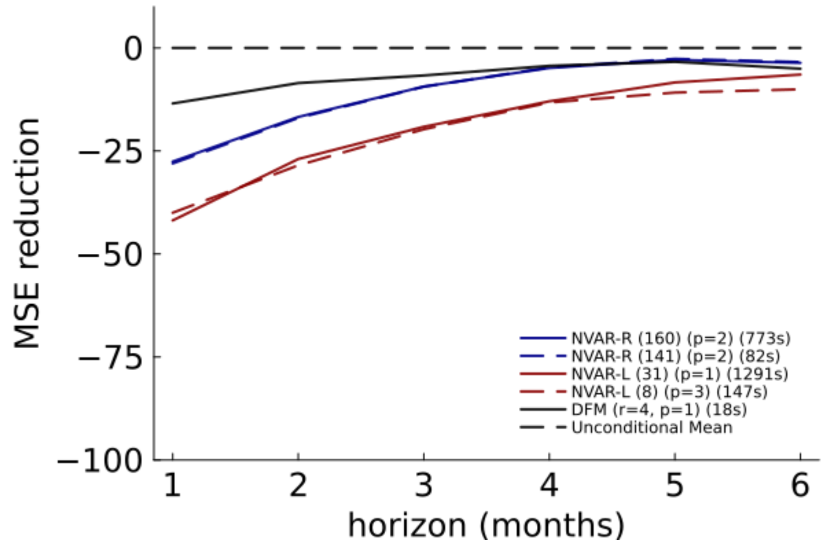

Figure 7: Forecasting: NVAR$(p, 1)$ vs. Factor Model, Industrial Production Growth

Notes: The plot depicts the out-of-sample MSEs generated by the ex-post best-performing DFM, NVAR-R and NVAR-L, all relative to those generated by an unconditional mean forecast. All forecasts refer to those obtained under the posterior mode.

lines) further reduces the MSE at the first horizon, amounting to -27% relative to the same benchmark (-16% relative to the DFM), whereby the noticeable improvement relative to the DFM persists for the first two horizons. Thereby, both methods to select λ_a yield similar answers – $\lambda_{a,MPM} = 160$ (solid line) and $\lambda_{a,MDD} = 141$ (dashed line) – and indistinguishable forecasting performances. They both lead to $p = 2$ as the ex-post best-performing model. For NVAR-L (red lines), smaller values and differing lag-lengths are selected: $\lambda_{a,MPM} = 31$ with $p = 1$ (solid line) and $\lambda_{a,BIC} = 8$ with $p = 3$ (dashed line). Nevertheless, the forecasting performances are again similar. For the first horizon, they yield MSE reductions of -42% and -40% relative to the unconditional mean (-33% and -30% relative to the DFM). The improvement is long-lasting, reverting to levels obtained under the DFM and NVAR-R only for six-period ahead forecasts.[45]

The results in Fig. 7 are obtained using the respective posterior mode, i.e. frequentist point estimator. As shown in Fig. A-1, the factor models yield similar performances even for forecasts at the posterior mean or the posterior mean forecasts. Selecting the best DFM using these alternative forecasting-types does not change any of the above numbers by more than two percentage points. For NVAR-L, the normalization $\alpha_1 = 1$ is applied, whereas for NVAR-R, $||\alpha||_1 = 1$ is imposed.[46] The computational time needed to obtain the posterior modes of the DFM and the NVAR for a given λ_a are of a similar order of magnitude and

[45]The models' performances under different different choices of p and r are shown in Fig. A-1 and Fig. A-2.

[46]Under the former normalization, NVAR-R yields poor performance, as the free parameters in α are increased to extremely high values, while all elements in A are shrunk accordingly. This is likely because, for highly correlated series with $\mathbb{E}[y_t y'_{t-1}] \approx \mathbb{E}[y_t y'_{t-2}]$, α_{-1} is weakly identified, as e.g. $Ay_{t-1} + \alpha_2 A y_{t-2} \approx (1 + \alpha_2) A y_{t-1}$. This issue does not occur for NVAR-L, likely because it shrinks numerous elements of A all the way to zero. The issue does not occur either when estimating the full posterior of NVAR-R or NVAR-L.

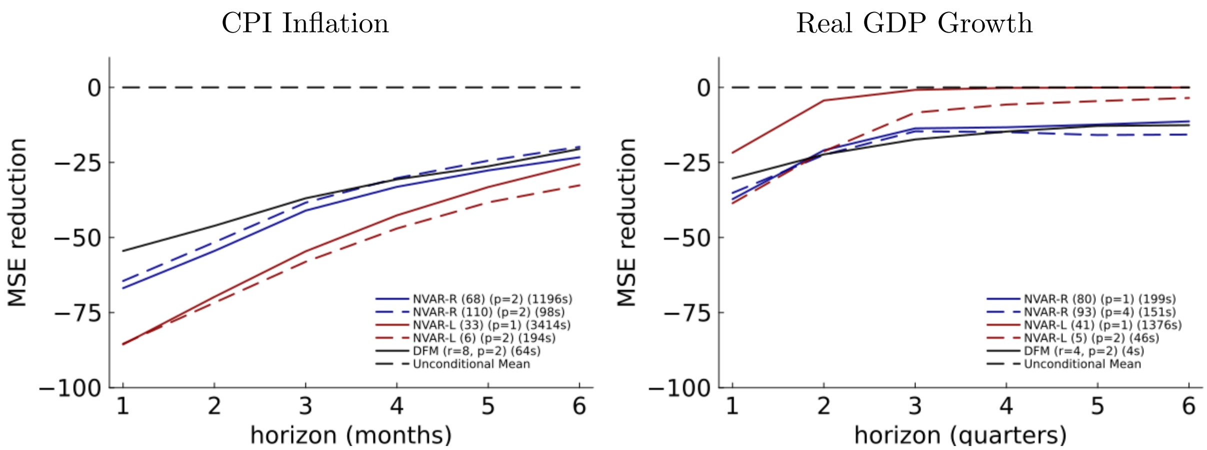

Figure 8: Forecasting: NVAR$(p, 1)$ vs. Factor Model, CPI Inflation & Real GDP Growth

Notes: The plot depicts the out-of-sample MSEs generated by the ex-post best-performing DFM, NVAR-R and NVAR-L, all relative to those generated by an unconditional mean forecast. All forecasts refer to those obtained under the posterior mode.

amount to 5-10 seconds. However, this burden is considerably increased by the need to re-compute the posterior mode for many different λ_a in the search for the value that maximizes MDD or BIC and the need to compute the full posterior distribution in the search for the MPM of λ_a.[47]

Qualitatively, the same conclusions apply when forecasting monthly CPI inflation. The left panel of Fig. 8 shows that, in this case, the DFM reduces the one-step ahead MSE by 54% relative to the unconditional mean forecast, with a noticeable improvement persisting throughout the first six months. The NVAR-R improves slightly upon this, yielding -67% and -64% (-27% and -22% relative to the DFM). The NVAR-L delivers again the best performance: -86% and -85% (-68% relative to DFM).

When forecasting quarterly GDP growth, the conclusions change. The performance of the best DFM (-30%) is beaten only slightly and only at the first horizon. The NVAR-R models yield -37% and -35% (-10% and -7% relative to DFM), and the NVAR-L with λ_a-selection according to BIC yields -38% (-12% relative to DFM). At longer horizons, the NVAR-R yields a similar performance as the DFM, beating the unconditional mean by about 10-15%, while the reduction by the mentioned NVAR-L model reverts faster to zero. The NVAR-L with λ_a selected by MPM yields -22% and thereby underperforms the DFM by 12%.

[47]For NVAR-R, this yields about 50-100s in the former case and about 10min in the latter case. These times are doubled for NVAR-L.

6 Conclusion

In this paper, I develop the Network-VAR (NVAR) – an econometric framework that rationalizes the dynamics of a cross-sectional variable by the dynamic innovation transmission along bilateral links among cross-sectional units – and I consider two applications. First, I estimate the contribution of lagged input-output conversion to business cycles through the lens of an RBC economy, thereby providing a structural underpinning of the NVAR. Second, I use the NVAR as a dimensionality-reduction technique for forecasting cross-country macroeconomic aggregates.

More work is needed to explore the extent to which the NVAR is useful to address sparse factor-issues and improve upon the forecasting performance of alternative dimensionality-reduction techniques. Its application for forecasting very high-dimensional processes would benefit from refinements of the crude shrinkage priors used in this paper.

Furthermore, rather than assuming time-invariant links, an important methodological step forward would be to jointly study dynamic network effects and -formation. Rather than assuming stationarity, an interesting avenue for further research would be to study networked cointegration. Finally, the NVAR could be augmented to accommodate heterogeneous propagation patterns across units or over time, which in the context of a production economy amounts to endogenizing firms' inventory management. I leave these directions for future research.

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.

Ahelegbey, D. F., M. Billio, and R. Casarin (2016): "Bayesian Graphical Models for Structural Vector Autoregressive Processes," *Journal of Applied Econometrics*, 31, 357–386.

Alessandria, G., J. P. Kaboski, and V. Midrigan (2010): "Inventories, Lumpy Trade, and Large Devaluations," *American Economic Review*, 100, 2304–2339.

Anatolyev, S. and A. Mikusheva (2022): "Factor models with many assets: Strong factors, weak factors, and the two-pass procedure," *Journal of Econometrics*, 229, 103–126.

Anselin, L. (2003): "Spatial Econometrics," in *A Companion to Theoretical Econometrics*, ed. by B. H. Baltagi, Wiley, 310–330, 1 ed.

Antràs, P. and V. Tubdenov (2025): "Measuring the Average Period of Production," *NBER Working Paper*.

Anufriev, M. and V. Panchenko (2015): "Connecting the dots: Econometric methods for uncovering networks with an application to the Australian financial institutions," *Journal of Banking & Finance*, 61, S241–S255.

Armillotta, M. and K. Fokianos (2023): "Nonlinear Network Autoregression," *The Annals of Statistics*, 51.

Bai, J. and S. Ng (2019): "Rank regularized estimation of approximate factor models," *Journal of Econometrics*, 212, 78–96.

Barigozzi, M. and C. Brownlees (2018): "Nets: Network Estimation for Time Series," .

Barigozzi, M., G. Cavaliere, and G. Moramarco (2025): "Factor Network Autoregressions," *Journal of Business & Economic Statistics*, 1–14.

Berman, A. and R. J. Plemmons (1994): *Nonnegative Matrices in the Mathematical Sciences*, Society for Industrial and Applied Mathematics.

Billio, M., R. Casarin, and L. Rossini (2019): "Bayesian nonparametric sparse VAR models," *Journal of Econometrics*, 212, 97–115.

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

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

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, eds. (2016): *The Oxford Handbook of the Economics of Networks*, Oxford University Press, 1 ed.

Bykhovskaya, A. (2023): "Time Series Approach to the Evolution of Networks: Prediction and Estimation," *Journal of Business & Economic Statistics*, 41, 170–183.

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

Canova, F. and M. Ciccarelli (2013): "Panel vector autoregressive models: a survey," *ECB Working Paper Series*, No. 1507.

Caporin, M., D. Erdemlioglu, and S. Nasini (2021): "Estimating Financial Networks by Realized Interdependencies: A Restricted Autoregressive Approach," *SSRN Electronic Journal*.

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

Carvalho, C. M. and M. West (2007): "Dynamic Matrix-Variate Graphical Models," *Bayesian Analysis*, 2, 69–98.

Carvalho, V. M. (2010): "Aggregate Fluctuations and The Network Structure of Intersectoral Trade," *Manuscript, Universitat Pompeu Fabra.*

Carvalho, V. M. and M. Reischer (2021): "Pathways to Persistence," *Manuscript, University of Cambridge.*

Carvalho, V. M. and A. Tahbaz-Salehi (2019): "Production Networks: A Primer," *Annual Review of Economics*, 11, 635–663.

Cesa-Bianchi, A. and A. Ferrero (2021): "The transmission of Keynesian supply shocks," *SSRN Electronic Journal.*

Chen, J., D. Li, Y.-N. Li, and O. Linton (2025): "Estimating time-varying networks for high-dimensional time series," *Journal of Econometrics*, 105941.

Chodorow-Reich, G., X. Gabaix, and D. Viviano (2025): "Propagation of Shocks in Networks: Identification and Applications," .

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

——— (2014): "Theory and Practice of GVAR Modeling," .

Dahlhaus, T., J. Schaumburg, and T. Sekhposyan (2021): "Networking the Yield Curve: Implications for Monetary Policy," *SSRN Electronic Journal.*

Davis, R. A., P. Zang, and T. Zheng (2016): "Sparse Vector Autoregressive Modeling," *Journal of Computational and Graphical Statistics*, 25, 1077–1096.

De Graeve, F. and J. D. Schneider (2023): "Identifying sectoral shocks and their role in business cycles," *Journal of Monetary Economics*, 140, 124–141.

de Paula, A., I. Rasul, and P. C. Souza (2024): "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition," *Review of Economic Studies*, 1–39.

Desmet, K. and F. Parro (2025): "Spatial Dynamics," *NBER Working Paper*, No. 33443.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82, 1303–1340.

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.

Dupor, B. (1999): "Aggregation and irrelevance in multi-sector models," *Journal of Monetary Economics*, 43, 391–409.

Elhorst, J. P. (2012): "Dynamic spatial panels: models, methods, and inferences," *Journal of Geographical Systems*, 14, 5–28.

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

Fan, J., S. Hong, and F. Parro (2023): "Learning and Expectations in Dynamic Spatial Economies," *NBER Working Paper*.

Fan, J., R. Masini, and M. C. Medeiros (2022): "Bridging factor and sparse models," .

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, 1–38.

Freyaldenhoven, S. (2025): "Identification Through Sparsity in Factor Models: the '1-rotation criterion," *Manuscript, Federal Reserve Bank of Philadelphia*.

Geweke, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, Amsterdam.

——— (1984): "Inference and causality in economic time series models," in *Handbook of Econometrics*, Elsevier, vol. 2, 1101–1144.

Geweke, J. and G. Zhou (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557–587.

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

——— (2021): "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437.

Golub, B. and M. O. Jackson (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," .

Graham, B. S. (2020): "Network data," in *Handbook of Econometrics*, Elsevier, vol. 7, 111–218.

Holly, S. and I. Petrella (2012): "Factor Demand Linkages, Technology Shocks, and the Business Cycle," *The Review of Economics and Statistics*, 94, 948–963.

Horvath, M. (1998): "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*, 1, 781–808.

——— (2000): "Sectoral shocks and aggregate fuctuations," *Journal of Monetary Economics*, 45, 69–106.

Kelejian, H. H. and I. R. Prucha (2010): "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, 157, 53–67.

Khan, A. and J. K. Thomas (2007): "Inventories and the Business Cycle: An Equilibrium Analysis of (S, s) Policies," .

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

Kuersteiner, G. M. and I. R. Prucha (2020): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," *Econometrica*, 88, 2109–2146.

Lam, C. and P. C. Souza (2020): "Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model," *Journal of Business & Economic Statistics*, 38, 693–710.

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

Litterman, R. B. (1986): "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business & Economic Statistics*, 4, 25.

Liu, E. and A. Tsyvinski (2024): "A Dynamic Model of Input–Output Networks," *Review of Economic Studies*, 91, 3608–3644.

Long, Jr., J. B. and C. I. Plosser (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

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

Martin, B., F. S. Passino, M. Cucuringu, and A. Luati (2025): "NIRVAR: Network Informed Restricted Vector Autoregression," .

Maung, K. (2022): "Large Network Autoregressions with Unknown Adjacency Matrix," .

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

Nelson, C. R. and C.-J. Kim (1999): *State-Space Models with Regime Switching*, The MIT Press.

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

Palm, F. C. and T. E. Nijman (1984): "Missing Observations in the Dynamic Regression Model," *Econometrica*, 52, 1415.

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

Phillips, P. (1973): "The problem of identification in finite parameter continuous time models," *Journal of Econometrics*, 1, 351–362.

Pinkse, J., M. E. Slade, and C. Brett (2002): "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, 70, 1111–1153.

Qu, X., L.-f. Lee, and C. Yang (2021): "Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables," *Journal of Econometrics*, 221, 180–197.

Schorfheide, F. and D. Song (2015): "Real-Time Forecasting With a Mixed-Frequency VAR," *Journal of Business & Economic Statistics*, 33, 366–380.

Shalizi, C. R. and A. C. Thomas (2011): "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research*, 40, 211–239.

Shea, J. (2002): "Complementarities and Comovements," .

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: Statistical Methodology*, 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.

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

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

Online Appendix

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

Marko Mlikota

Geneva Graduate Institute

A NVAR: Theory

A.1 NVAR$(p, 1)$: Granger-Causality/GIRFs

Proposition 1 (NVAR$(p, 1)$: Granger-Causality).
Let y_t evolve as in Eq. (2). Then, for $\underline{k} = ceil(h/p)$ and some coefficients $\{c_k^h\}_{k=\underline{k}:h}$, we have

$$\frac{\partial y_{i,t+h}}{\partial u_{j,t}} = c_{\underline{k}}^h \left[A^{\underline{k}}\right]_{ij} + ... + c_h^h \left[A^h\right]_{ij} \ .$$

Proof: Write y_t in companion form as $z_t = F z_{t-1} + e_t$, where $z_t = (y_t', y_{t-1}', ..., y_{t-p+1}')'$ and $e_t = (u_t', 0', ..., 0')'$ are $np \times 1$ vectors, and the $np \times np$ matrix F is

$$F = \begin{bmatrix} \Phi_1 & ... & \Phi_{p-1} & \Phi_p \\ I_n & ... & 0_n & 0_n \\ \vdots & \ddots & & \vdots \\ 0_n & ... & I_n & 0_n \end{bmatrix} \ .$$

We have

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

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

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

(Only the first row of blocks in F^{h+1} are relevant to the argument.) Consider $m \in 1 : (p-1)$ s.t $h+m$ is a multiple of p. Then

$$ceil\left(\frac{h+l-1}{p}\right) = \begin{cases} \frac{h+m}{p} & \text{for } l = 1 : (m+1) \\ \frac{h+m}{p} + 1 & \text{for } l = (m+2) : p \end{cases} .$$

This means that for $l = 1 : (m+1)$, $(F^h)_{1l}$ has powers of A in $\left(\frac{h+m}{p}\right) : h$, while for $l = (m+2) : p$ it has powers in $\left(\frac{h+m}{p} + 1\right) : h$. Then, by the equation above, for $l = 1 : m$, $(F^{h+1})_{1l}$ has powers of A in $\left(\frac{h+m}{p}\right) : (h+1)$, while for $l = 1 : m$ it has powers in $\left(\frac{h+m}{p} + 1\right) : (h+1)$. These sets are both equal to $ceil\left(\frac{h+1+l-1}{p}\right) : (h+1)$ and independent of m. Therefore, the claim holds in all possible cases. ■

Proposition 5 (NVAR$(p,1)$: Long-Term Response to Persistent WN-Innovations)**.**
Let y_t evolve as in Eq. (2), and let $x = aAx + v$, with $a = \sum_{l=1}^{p} \alpha_l$. Assume y_t is weakly stationary. Then,

$$\lim_{H\to\infty} \sum_{h=0}^{H} \frac{\partial y_{i,t+H}}{\partial u_{j,t+h}} = \frac{\partial x_i}{\partial v_j} = \left[(I - aA)^{-1}\right]_{ij} .$$

Proof: First, note that $x = (I - aA)^{-1}v$ and therefore $\partial x/\partial v = (I - aA)^{-1}$. Turning to y_t, under weak stationarity,

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

To find R, write y_t in companion form as $z_t = Fz_{t-1} + e_t$. We have

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

Since $\sum_{h=0}^{\infty} F^h = (I - F)^{-1}$,

$$R = \sum_{l=0}^{\infty} \frac{\partial y_{t+l}}{\partial u_t} = \left((I - F)^{-1}\right)_{11}$$

is given by the $n \times n$ matrix in position (1,1) in the $np \times np$ matrix $(I - F)^{-1}$.

Let $M = (I - F)^{-1}$ and partition it into p^2 blocks of dimension $n \times n$, denoted by

$\{M_{lm}\}_{l,m=1:p}$. Then, $R = M_{11}$. We have

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

$$= \begin{bmatrix} M_{11} & M_{12} & \dots & M_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p1} & M_{p2} & \dots & M_{pp} \end{bmatrix} \begin{bmatrix} I-\alpha_1 A & -\alpha_2 A & -\alpha_3 A & \dots & -\alpha_{p-1}A & -\alpha_p A \\ -I_n & I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & -I_n & I_n & \dots & 0_n & 0_n \\ \vdots & & \ddots & \ddots & & \vdots \\ 0_n & 0_n & \dots & -I_n & I_n & 0_n \\ 0_n & 0_n & \dots & 0_n & -I_n & I_n \end{bmatrix}$$

Comparing the left- and right-hand sides for block $(1, p)$, we get

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

which implies $M_{1p} = M_{11}\alpha_p A$. For block $(1, l)$, we get

$$0_n = -M_{11}\alpha_l A + M_{1l} - M_{1,l+1} \ , \quad l \in 2:(p-1) \ ,$$

which implies

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

The first element gives

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

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

Proposition 6 (NVAR$(p, 1)$: Long-Term Response to Persistent AR(1)-Innovations)**.** *Let y_t evolve as in Eq. (2), and let $x_t = aAx_t + v_t$, with $a = \sum_{l=1}^{p}\alpha_l$. Assume y_t and x_t are weakly stationary, and assume $(1-\rho_j L)u_{jt} = \varepsilon_{jt} \sim WN$ and $(1-\rho_j L)v_{jt} = \epsilon_{jt} \sim WN$, with $\rho_j \in [0, 1)$. Then,*

$$\lim_{H\to\infty}\sum_{h=0}^{H}\frac{\partial y_{i,t+H}}{\partial \varepsilon_{j,t+h}} = \lim_{H\to\infty}\sum_{h=0}^{H}\frac{\partial x_{i,t+H}}{\partial \epsilon_{j,t+h}} = \frac{1}{1-\rho_j}\left[(I-aA)^{-1}\right]_{ij} \ .$$

Proof: It holds that

$$\frac{\partial x_{t+h}}{\partial \epsilon_t} = \sum_{l=0}^{h}\frac{\partial x_{t+h}}{\partial v_{t+h-l}}\frac{\partial v_{t+h-l}}{\partial \epsilon_t} \quad \text{and} \quad \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \sum_{l=0}^{h}\frac{\partial y_{t+h}}{\partial u_{t+h-l}}\frac{\partial u_{t+h-l}}{\partial \varepsilon_t} \ .$$

By Proposition 5, under stationarity of x_t and y_t, we also know that

$$\frac{\partial x_t}{\partial v_t} = \lim_{H\to\infty} \sum_{l=0}^{H} \frac{\partial y_{t+l}}{\partial u_t} = (I - aA)^{-1} \ .$$

For x_t, $\partial x_{t+h}/\partial v_{t+h-l} = 0$ for $l > 1$. We then have

$$\lim_{H\to\infty} \sum_{h=0}^{H} \frac{\partial x_{t+H}}{\partial \epsilon_{t+h}} = \sum_{h=0}^{\infty} \frac{\partial x_{i,t+h}}{\partial \epsilon_{j,t}} = \sum_{h=0}^{\infty} \frac{\partial x_{i,t+h}}{\partial v_{j,t+h}} \frac{\partial v_{j,t+h}}{\partial \epsilon_{j,t}} = \sum_{h=0}^{\infty} \left[(I - aA)^{-1}\right]_{ij} \rho_j^h = \frac{\left[(I - aA)^{-1}\right]_{ij}}{1 - \rho_j} \ .$$

For y_t, we have

$$\begin{aligned}
\lim_{H\to\infty} \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial \varepsilon_{j,t}} &= \lim_{H\to\infty} \sum_{h=0}^{H} \sum_{l=0}^{h} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \rho_j^{h-l} \\
&= \lim_{H\to\infty} \sum_{l=0}^{H} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \sum_{h=l}^{H} \rho_j^{h-l} \\
&= \lim_{H\to\infty} \sum_{l=0}^{H} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \frac{1 - \rho_j^{H-l+1}}{1 - \rho_j} \\
&= \frac{1}{1 - \rho_j} \lim_{H\to\infty} \sum_{l=0}^{H} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} - \lim_{H\to\infty} \sum_{l=0}^{H} \frac{\rho_j^{H-l+1}}{1 - \rho_j} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \ .
\end{aligned}$$

By Proposition 5, the first term equals $(1 - \rho_j)^{-1} \left[(I - aA)^{-1}\right]_{ij}$.

The second term equals zero. To see this, break it up as follows:

$$\lim_{H\to\infty} \sum_{l=0}^{H-1} \frac{\rho_j^{H-l+1}}{1 - \rho_j} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} + \lim_{H\to\infty} \frac{\rho_j}{1 - \rho_j} \frac{\partial y_{i,t+H}}{\partial u_{j,t}} \ .$$

The latter term equals zero since y_t is stationary. The former term can be written as

$$\frac{1}{1 - \rho_j} \lim_{H\to\infty} \rho_j^{H+1-c} \sum_{l=0}^{H-1} \rho_j^{c-l} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \ \forall\ c > 0 \ \ .$$

Let $c = H(1 - \delta)$ for $\delta > 0$ small. Since $\lim_{H\to\infty} \rho_j^{H+1-c} = 0$, it remains to be shown that the limt of the second term in this product is finite, in which case the limit of products is equal to the product of their limits, which is zero.

Let $\tilde{\rho}_j = \rho_j^{c/(H-1)-1}$. For δ small enough, $\tilde{\rho}_j \in [0,1)$. We then have

$$\begin{aligned}
\lim_{H\to\infty} \sum_{l=0}^{H-1} \rho_j^{c-l} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} &= \lim_{H\to\infty} \sum_{l=0}^{H-1} \left(\rho_j^{c/l-1}\right)^l \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \\
&\leq \lim_{H\to\infty} \sum_{l=0}^{H-1} \tilde{\rho}_j^l \frac{\partial y_{i,t+l}}{\partial u_{j,t}} \\
&\leq \lim_{H\to\infty} \sum_{l=0}^{H-1} \frac{\partial y_{i,t+l}}{\partial u_{j,t}} = \left[(I-aA)^{-1}\right]_{ij} < \infty \quad \blacksquare
\end{aligned}$$

A.2 NVAR$(p,1)$: Stationarity

Proposition 7 (NVAR$(1,1)$: Stationarity)**.**
Let $y_t = aAy_{t-1} + u_t$. Assume $u_t \sim WN$ and $a \neq 0$.

Then y_t is weakly stationary iff $|\lambda_i| < 1/|a|$ for all eigenvalues λ_i of A.

Proof: Let $\mathcal{L}_A$ and $\mathcal{L}$ be the sets of eigenvalues of A and aA, respectively:

$$\begin{aligned}
\mathcal{L}_A &= \{\lambda_i : |\lambda_i I - A| = 0\} \ , \\
\mathcal{L} &= \{l_i : |l_i I - aA| = 0\} \ .
\end{aligned}$$

The pairs of eigenvalues l_i and λ_i are related by the identity $\lambda_i = l_i/a$:

$$|l_i I - aA| = |a(l_i/aI - A)| = a^n |l_i/aI - A| = 0 \Leftrightarrow |l_i/aI - A| = 0 \ .$$

We have

$$\begin{aligned}
y_t \text{ is weakly stationary} \quad &\Leftrightarrow \quad \forall\, l_i \in \mathcal{L} \ , \quad |l_i| < 1 \\
&\Leftrightarrow \quad \forall\, l_i \in \mathcal{L} \ , \quad |l_i/a| = |l_i|/|a| < 1/|a| \\
&\Leftrightarrow \quad \forall\, \lambda_i \in \mathcal{L}_A \ , \quad |\lambda_i| < 1/|a| \ . \quad \blacksquare
\end{aligned}$$

Proposition 8 (NVAR$(p,1)$: Stationarity I)**.**
Let y_t evolve as in Eq. (2). Assume $u_t \sim WN$ and $\alpha_l \neq 0$ for at least one l, and define $a = \sum_{l=1}^{p} |\alpha_l|$.

If $|\lambda_i| < 1/a$ for all eigenvalues λ_i of A, then y_t is weakly stationary. Under $\alpha_l \geq 0\ \forall\, l$, the implication is both-sided.

Proof: Consider the NVAR(1,1) $y_t^* = aAy_{t-1}^* + u_t^*$. By Proposition 7, we know

$$y_t^* \text{ is weakly stationary} \quad \Leftrightarrow \quad |\lambda_i| < 1/|a| \quad \forall\, \lambda_i \in \mathcal{L}_A = \{\lambda_i : |\lambda_i I - A| = 0\}\ .$$

It also holds that

$$y_t^* \text{ is weakly stationary} \quad \Leftrightarrow \quad |z_i^*| > 1 \quad \forall\, z_i^* \in \mathcal{Z}^* = \{z_i^* : |I - z_i^* aA| = 0\}\ .$$

where $\mathcal{Z}^*$ is the set of roots z_i^* of the lag polynomial $(1 - aAL)$. Analogously, let

$$\mathcal{Z} = \{z_i : |I - \alpha_1 A z_i - ... - \alpha_p A z_i^p| = |I - (\alpha_1 z_i + ... + \alpha_p z_i^p)A| = 0\}$$

be the set of roots z_i of the lag polynomial $(1 - \alpha_1 AL - \alpha_2 AL^2 - ... - \alpha_p AL^p)$. The proof shall show

$$\forall\, z_i^* \in \mathcal{Z}^*, \quad |z_i^*| > 1 \quad \Rightarrow \quad \forall\, z_i \in \mathcal{Z}, \quad |z_i| > 1\ .$$

We have

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

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

$$\begin{aligned} |\alpha_1 z_i + ... + \alpha_p z_i^p| &\leq |\alpha_1 z_i| + ... + |\alpha_p z_i^p| \\ &\leq |\alpha_1 z_i| + ... + |\alpha_p z_i| \\ &\leq (|\alpha_1| + ... + |\alpha_p|)|z_i| = a|z_i| \leq a\ , \end{aligned}$$

a contradiction. If $\alpha_l \geq 0\ \forall\, l$, the last implication is both-sided:

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

Proposition 9 (NVAR(p, 1): Stationarity II)**.**
Let y_t evolve as in Eq. (2). Assume $u_t \sim WN$ and $\alpha_l \neq 0$ for at least one l.

Then, y_t is weakly stationary iff for all eigenvalues λ_i of A the $p \times p$ matrix

$$\begin{bmatrix} \alpha_1\lambda_i & ... & \alpha_p\lambda_i \\ I_{p-1} & & 0 \end{bmatrix}$$

has all eigenvalues inside the unit circle.

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

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

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

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

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

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

are eigenvalues of F. This equation is the characteristic polynomial for eigenvalues of the matrix

$$F_X = \begin{bmatrix} \alpha_1\lambda_i & ... & \alpha_p\lambda_i \\ I_{p-1} & & 0 \end{bmatrix} \; . \quad \blacksquare$$

A.3 NVAR$(p, 1)$: Relation to Dynamic Factor Model

Proposition 4 (NVAR$(p,1)$-Factor Model Equivalence Result)**.**
Let y_t evolve as in Eq. (2). Then we can write

$$y_t = \Lambda f_t + u_t \ ,$$

where $f_t = C[\alpha_1 y_{t-1} + ... + \alpha_p y_{t-p}] \in \mathbb{R}^r$, r is the rank of A, and Λ and C are full-rank matrices that satisfy $A = \Lambda C$.

Conversely, let $y_t = \Lambda f_t + \xi_t$, with $f_t \in \mathbb{R}^r$. Assume $f_t = \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t$ with $\Phi_l = \phi_l \Phi$ for all l and some Φ. Then, as $n \to \infty$, we can write

$$y_t = \phi_1 A y_{t-1} + ... + \phi_p A y_{t-p} + u_t \ , \quad u_t = \Lambda \eta_t + \xi_t \ ,$$

where $A = \Lambda \Phi (W\Lambda)^{-1} W$ and W is any $r \times n$ matrix with distinct rows.

Proof: The argument works for any p. For expositional simplicity, let $p = 2$. The NVAR$(2,1)$ can be written as

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

Given $r = rank(A)$, we can find $n \times r$ and $r \times n$ matrices B and C, both of full rank, such that $A = BC$. In turn, we can write

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

where $\Lambda = B$ and $f_t = C[\alpha_1 y_{t-1} + ... + \alpha_p y_{t-p}]$.

Conversely, let

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

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

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

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

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

where the $r \times n$ matrix W stacks $w^{k\prime}$ along rows. In turn, we can solve for $f_t = (W\Lambda)^{-1} W y_t$.

As this equation holds for all t, we can re-write the process for y_t as

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

with $u_t = \Lambda\eta_t + \xi_t$. If $\Phi_1 = \phi_1\Phi$ and $\Phi_2 = \phi_2\Phi$ for some ϕ_1, ϕ_2, Φ, this simplifies to

$$\begin{aligned} y_t &= \Lambda\Phi(W\Lambda)^{-1}W[\phi_1 y_{t-1} + \phi_2 y_{t-2}] + u_t \\ &= \phi_1 A y_{t-1} + ... + \phi_p A y_{t-p} + u_t \end{aligned}$$

for $A = \Lambda\Phi(W\Lambda)^{-1}W$ of rank r. ■

A.4 NVAR(p, q)

Generality of $q \in \mathbb{N}$

Throughout the paper and in the remainder of this section, I discuss the NVAR(p, q) under $q = 1$ and $q \in \mathbb{N}\backslash\{1\}$. In Section 2.2.3, I mention that we can accommodate any $q \in \mathbb{Q}_{++}$ by constructing a restricted NVAR$(p*, q^*)$ with $q^* \in \mathbb{N}$, at least in the case of Eq. (4), where $\tilde{y}_\tau$ is interpreted as a stock variable.

Consider first $q^{-1} \in \mathbb{N}\backslash\{1\}$, i.e. observational frequency is an integer-multiple of the network interaction frequency. For example, under monthly observations, $q = 1/3$ indicates quarterly network interactions, and p signifies over how many past quarters transmission is spread out. In line with Eq. (4), $\tilde{y}_\tau$ is observed in each period τ, which means that $\tilde{y}_\tau$ must evolve at observational frequency. To make sense of $q^{-1} \in \mathbb{N}$ then, we can let the stock $\tilde{y}_\tau$ follow an NVAR$(p^*, 1)$ with $p^* = p/q$ s.t. it depends on its value from q until p/q observational periods ago, which correspond to the last p periods at network interaction frequency:

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

In the previous example, the observed monthly series depends on its value in the past p quarters, i.e. on its value three months ago, six months ago, etc., up to $3p$ months ago.[48]

Consider next the general case: $q \in \mathbb{Q}_{++}$. We can write $q = q_1 q_2$ with $q_1^{-1} \in \mathbb{N}$ and $q_2 \in \mathbb{N}$.[49] For example, under monthly observations, $q = 4/3$ implies that network interactions

[48]Note that this assumes that transmission happens instantaneously at the end of each (network interaction-) period. Alternative definitions of a smoother transmission inevitably lead towards abandoning the paradigm of discrete time.

[49]Note that q_2 is the least common multiple of q and 1, whereas q_1 is their greatest common denominator.

occur every three weeks, $q = 4/5$ that they occur every five weeks, and $q = 30/4$ that they occur every four days. We can model this case as a combination of the $q \in \mathbb{N}$ and $q^{-1} \in \mathbb{N}$ cases: we observe every $q_2 \in \mathbb{N}$ periods a snapshot of an NVAR$(p^*, 1)$ process with $p^* = p/q_1 \in \mathbb{N}$ and with parameters restricted as in the preceding paragraph. For example, under monthly observations and $q = 4/3$, this amounts to observing every fourth period a snapshot of a weekly process that depends on its value three weeks ago, six weeks ago, etc.

If y_t is observed for T periods – $t = 1 : T$ –, then $\tilde{y}_\tau$ evolves for T_τ (network interaction-)periods – $\tau = 1 : T_\tau$. Thereby, T_τ satisfies $T = |\{(1 : T_\tau)/q\} \cap \mathbb{N}|$; the number of elements in the set $1 : T_\tau$ that are integer-multiples of q equals T. This yields $T_\tau = qT$ under $q \in \mathbb{N}$, and $T_\tau = T$ under $q^{-1} \in \mathbb{N}$. For other $q \in \mathbb{Q}_{++}$, we have $T_\tau = q_2 T$, where q_2 is least common multiple of q and 1.

NVAR(p, q): Granger-Causality/GIRFs and Stationarity

Proposition 2 (NVAR(p, q): Granger-Causality I)**.**
Let y_t evolve as in Eq. (4) for some $q \in \mathbb{N} \backslash \{1\}$.

Then, for $\underline{k} = ceil(hq/p)$ and some coefficients $\{c^h_k\}_{k=\underline{k}:hq}$, we have

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

Proof: By the definition of y_t,

$$\frac{\partial y_{i,t+h}}{\partial y_{j,t}} = \frac{\partial \tilde{y}_{i,(t+h)q}}{\partial \tilde{y}_{j,tq}} = \frac{\partial \tilde{y}_{i,\tau+hq}}{\partial \tilde{y}_{j,\tau}} = \frac{\partial \tilde{y}_{i,\tau+hq}}{\partial \tilde{u}_{j,\tau}} \ .$$

The statement follows then from Proposition 1. ■

Proposition 10 (NVAR(p, q): Granger-Causality II)**.**
Let y_t evolve as in Eq. (5) for some $q \in \mathbb{N} \backslash \{1\}$.

Then, for $\underline{k} = ceil(hq/p)$ and some coefficients $\{c^h_k\}_{k=\underline{k}:hq}$, we have

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

where R is a linear combination of A^k for some $k \notin \underline{k} : hq$.

Proof: We have

$$y_{t+h} = (\tilde{y}_{(t+h)q} + ... + \tilde{y}_{(t+h)q-q+1})/q \; ,$$
$$y_t = (\tilde{y}_{tq} + ... + \tilde{y}_{tq-q+1})/q \; .$$

Suppose the first (latest) term, $\tilde{y}_{tq}$ is responsible for the change in y_t. We then have

$$\frac{\partial y_{t+h}}{\partial \tilde{y}_{tq}} = \frac{\partial(\tilde{y}_{(t+h)q} + ... + \tilde{y}_{(t+h)q-q+1})/q}{\partial \tilde{y}_{tq}} = \frac{\partial \tilde{y}_{\tau+hq}}{\partial \tilde{y}_{\tau}} + ... + \frac{\partial \tilde{y}_{\tau+hq-q+1}}{\partial \tilde{y}_{\tau}} \; ,$$

and, by Proposition 1, connection-orders $k \in \{ceil((q(h-1)+1)/p), ..., hq\}$ matter. Analogous calculations show that if the last (earliest) term, $\tilde{y}_{tq-q+1}$, is responsible for the change in y_t, connection-orders $k \in \{ceil(hq/p), ..., hq+q-1\}$ matter, while the cases in-between lead to sets contained in the union of these two sets. Therefore, regardless of which term is responsible for the change in y_t, $\partial y_{i,t+h}/\partial u_{j,t}$ is a linear combination of A^k for connection-orders k in the intersection of these two sets,

$$\{ceil(hq/p), ..., hq\} = \{ceil((q(h-1)+1)/p), ..., hq\} \cap \{ceil(hq/p), ..., hq+q-1\} \; . \quad \blacksquare$$

Proposition 11 (NVAR(p,q): Preservaton of Long-Term Response to Persistent Innovations under Time-Aggregation)**.**
Let y_t evolve as in Eq. (4) *or Eq.* (5) *for some $q \in \mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $(1-\rho_j L)\tilde{u}_{j\tau} = \tilde{\varepsilon}_{j\tau} \sim WN$, with $\rho_j \in [0,1)$. Then,*

$$\lim_{H\to\infty} \sum_{h=0}^{Hq} \frac{\partial y_{t+H}}{\partial \tilde{\varepsilon}_{tq+h}} = \lim_{H\to\infty} \sum_{h=0}^{H} \frac{\partial \tilde{y}_{\tau+H}}{\partial \tilde{\varepsilon}_{\tau+h}} \; .$$

Proof: If y_t evolves as in Eq. (4), we have $y_{t+H} = \tilde{y}_{(t+H)q}$, and so

$$\lim_{H\to\infty} \sum_{h=0}^{Hq} \frac{\partial y_{t+H}}{\partial \tilde{\varepsilon}_{tq+h}} = \lim_{H\to\infty} \sum_{h=0}^{Hq} \frac{\partial \tilde{y}_{(t+H)q}}{\partial \tilde{\varepsilon}_{tq+h}} = \lim_{H\to\infty} \sum_{h=0}^{Hq} \frac{\partial \tilde{y}_{\tau+Hq}}{\partial \tilde{\varepsilon}_{\tau+h}} = \lim_{H\to\infty} \sum_{h=0}^{H} \frac{\partial \tilde{y}_{\tau+H}}{\partial \tilde{\varepsilon}_{\tau+h}} \; .$$

If y_t evolves as in Eq. (5), we have $y_{t+H} = q^{-1}(\tilde{y}_{(t+H)q} + ... + \tilde{y}_{(t+H)q-q+1})$, and the same result is obtained:

$$\begin{aligned}\lim_{H\to\infty}\sum_{h=0}^{Hq}\frac{\partial y_{t+H}}{\partial\tilde{\varepsilon}_{tq+h}} &= \lim_{H\to\infty}\sum_{h=0}^{Hq}q^{-1}\sum_{k=0}^{q-1}\frac{\partial\tilde{y}_{(t+H)q-k}}{\partial\tilde{\varepsilon}_{tq+h}}\mathbf{1}\{h\leq Hq-k\}\\ &= \lim_{H\to\infty}q^{-1}\sum_{k=0}^{q-1}\sum_{h=0}^{Hq-k}\frac{\partial\tilde{y}_{(t+H)q-k}}{\partial\tilde{\varepsilon}_{tq+h}}\\ &= \lim_{H\to\infty}\sum_{h=0}^{Hq-k}\frac{\partial\tilde{y}_{(t+H)q-k}}{\partial\tilde{\varepsilon}_{tq+h}}\\ &= \lim_{H\to\infty}\sum_{h=0}^{H}\frac{\partial\tilde{y}_{\tau+H}}{\partial\tilde{\varepsilon}_{\tau+h}}\ .^{50}\quad\blacksquare\end{aligned}$$

Proposition 12 (NVAR(p,q): Contemporaneous Response to Within-Period Innovation)**.** *Let y_t evolve as in Eq.* (5) *for some $q\in\mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $(1-\rho_j L)\tilde{u}_{j\tau}=\tilde{\varepsilon}_{j\tau}\sim WN$, with $\rho_j\in[0,1)$. Let $\tilde{\varepsilon}_\tau=(\tilde{\varepsilon}_{1\tau},...,\tilde{\varepsilon}_{n\tau})'$.*

Then, for $g\in 0:(q-1)$,

$$\lim_{g\to\infty}q\frac{\partial y_t}{\partial\tilde{\varepsilon}_{\tau-g}}=\lim_{g\to\infty}\sum_{h=0}^{g}\frac{\partial\tilde{y}_{\tau+g}}{\partial\tilde{\varepsilon}_{\tau+h}}\ .$$

Proof: We consider the response of $qy_t=\sum_{k=0}^{q-1}\tilde{y}_{tq-k}$ to a (high-frequency) innovation $\tilde{\varepsilon}_{tq-g}$, $g\in 0:(q-1)$ that occurs within observational period t. We have

$$q\frac{\partial y_t}{\partial\tilde{\varepsilon}_{tq-g}}=\sum_{k=0}^{q-1}\frac{\partial\tilde{y}_{tq-k}}{\partial\tilde{\varepsilon}_{tq-g}}\mathbf{1}\{k\leq g\}=\sum_{k=0}^{q-1}\frac{\partial\tilde{y}_{\tau+g-k}}{\partial\tilde{\varepsilon}_{\tau}}\mathbf{1}\{k\leq g\}=\sum_{k=0}^{g}\frac{\partial\tilde{y}_{\tau+g-k}}{\partial\tilde{\varepsilon}_{\tau}}=\sum_{h=0}^{g}\frac{\partial\tilde{y}_{\tau+h}}{\partial\tilde{\varepsilon}_{\tau}}\ .\quad\blacksquare$$

Proposition 13 (NVAR(p,q): Preservation of Stationarity under Time-Aggregation)**.** *For some $q\in\mathbb{N}\backslash\{1\}$, let y_t evolve as in Eq.* (4) *and let y_t^* evolve as in Eq.* (5)*. Assume $\tilde{u}_\tau\sim WN(0,\Sigma)$.*

Then, y_t and y_t^* are weakly stationary iff $\tilde{y}_\tau$ is weakly stationary.

Proof: Weak stationarity of $\tilde{y}_\tau$ is defined by the conditions

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

[50] The indicator function ensures that only responses to contemporaneous or past impulses are summed-up.

3. $\mathbb{V}[\tilde{y}_\tau] = Cov(\tilde{y}_\tau, \tilde{y}_\tau) < \infty$.

They imply that

1. $\mathbb{E}[y_t] = \mathbb{E}[\tilde{y}_{tq}] = \mathbb{E}[\tilde{y}_{(t-l)q}] = \mathbb{E}[y_{t-l}] \ \forall\ l$,
2. $Cov(y_t, y_{t-h}) = Cov(\tilde{y}_{tq}, \tilde{y}_{(t-h)q}) = Cov(\tilde{y}_{(t-l)q}, \tilde{y}_{(t-l-h)q}) = Cov(y_{t-l}, y_{t-l-h}) \ \forall\ l, h$,
3. $\mathbb{V}[y_t] = Cov(y_t, y_t) < \infty$,

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

1. $\mathbb{E}[y^*_t] = \mathbb{E}[\tilde{y}_{tq} + ... + \tilde{y}_{(t-1)q+1}]/q = \mathbb{E}[\tilde{y}_{(t-l)q} + ... + \tilde{y}_{(t-l-1)q+1}]/q = \mathbb{E}[y^*_{t-l}] \ \forall\ l$,
2. $$\begin{aligned} Cov(y^*_t, y^*_{t-h}) &= Cov(\tilde{y}_{tq} + ... + \tilde{y}_{(t-1)q+1}\ ,\ \tilde{y}_{(t-h)q} + ... + \tilde{y}_{(t-h-1)q+1})/q^2 \\ &= Cov(\tilde{y}_{(t-l)q} + ... + \tilde{y}_{(t-l-1)q+1}\ ,\ \tilde{y}_{(t-l-h)q} + ... + \tilde{y}_{(t-l-h-1)q+1})/q^2 \\ &= Cov(y^*_{t-l}, y^*_{t-l-h}) \ \forall\ l, h \ , \end{aligned}$$
3. $\mathbb{V}[y^*_t] = Cov(y^*_t, y^*_t) < \infty$,

which is the definition of weak stationarity for y^*_t.

The other direction is proved in contrapositive form: if $\tilde{y}_\tau$ is not weakly stationary, than y_t and y^*_t are not, either. Write $\tilde{y}_\tau$ in companion form as $\check{y}_\tau = F\check{y}_{\tau-1} + \check{u}_\tau$. If $\tilde{y}_\tau$ is not weakly stationary, then $\lim_{h\to\infty} F^h = \infty$. Hence, $\mathbb{V}[\check{y}_\tau] = F\mathbb{V}[\check{y}_{\tau-1}]F' + \check{\Sigma}$ diverges to infinity as $\tau \to \infty$. (This holds if $\check{y}_\tau$ starts in the infinite past and also if it has been initialized at some $\check{y}_0 = (\tilde{y}'_0, ..., \tilde{y}'_{-p+1})'$ with mean $\mathbb{E}[\check{y}_0]$ and variance $\mathbb{V}[\check{y}_0] < \infty$.) The same holds then for $\mathbb{V}[\tilde{y}_\tau]$, the upper-left $n \times n$ block in the $np \times np$ matrix $\mathbb{V}[\check{y}_\tau]$. In turn, the same holds for

$$\begin{aligned} &\mathbb{V}[y_t] = \mathbb{V}[\tilde{y}_{tq}] \ , \\ &\mathbb{V}[y^*_t] = \mathbb{V}[\tilde{y}_{tq} + ... + \tilde{y}_{(t-1)q+1}]/q^2 = \sum_{l=0}^{q-1} \mathbb{V}[\tilde{y}_{tq}]/q^2 + \sum_{l,k=0,l\neq k}^{q-1} Cov(\tilde{y}_{tq-l}, \tilde{y}_{tq-k})/q^2 \quad \blacksquare \end{aligned}$$

NVAR(p, q): Networked Correlation of "Observable Innovations"

Proposition 14 (NVAR(p, q): Correlated "Observable Innovations" I)**.**
Let y_t evolve as in Eq. (4) for some $q \in \mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $\tilde{u}_\tau \sim WN(0, \Sigma)$. Define $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{\tilde{y}_{\tau-q}, \tilde{y}_{\tau-q-1}, ...\}$. Then,

$$\mathbb{V}[u_t] = \Sigma + \sum_{h=1}^{q-1} \Theta_h \Sigma \Theta'_h \ , \quad \Theta_h = \partial\tilde{y}_{\tau+h}/\partial\tilde{y}_\tau \ .$$

Proof: By sequentially inserting for $\tilde{y}_{\tau-1}, \tilde{y}_{\tau-2}$, etc. in the expression for $\tilde{y}_\tau$, we get its VMA(∞)-representation:

$$\tilde{y}_\tau = (I + \Theta_1 L + \Theta_2 L^2 + ... + \Theta_{q-1} L^{q-1} + \Theta_q L^q + ...)\tilde{u}_\tau \ ,$$

where $\Theta_h = \partial \tilde{y}_\tau / \partial \tilde{y}_{\tau-h} = \partial \tilde{y}_{\tau+h} / \partial \tilde{y}_\tau$. Combining this with the definition of $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}] = \tilde{y}_\tau - \mathbb{E}[\tilde{y}_\tau|\mathcal{F}_{t-1}]$, y_t and $\mathcal{F}_{t-1}$, we get

$$u_t = (I + \Theta_1 L + ... + \Theta_{q-1} L^{q-1})\tilde{u}_\tau \ .$$

(Any $\tilde{u}_{\tau-h}$ is a function of $\tilde{y}_{\tau-h}$ and more distant lags of $\tilde{y}_\tau$, which means that $\tilde{u}_{\tau-q}$ and earlier terms are contained in $\mathcal{F}_{t-1}$.) Since $\tilde{u}_\tau \sim WN(0, \Sigma)$,

$$\mathbb{V}[u_t] = \Sigma + \sum_{h=1}^{q-1} \Theta_h \Sigma \Theta_h' \ . \quad \blacksquare$$

If $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$, then, by Proposition 1, for some constants $\{d^h_{h_i,h_j}\}_{h_i,h_j=\underline{k}(h):h}$,

$$\text{Cov}(u_{it}, u_{jt}) = \sigma_i^2 \mathbf{1}\{i = j\} + \sum_{o=1}^{n} \sigma_o^2 \sum_{h=1}^{q-1} \sum_{h_i,h_j=\underline{k}(h)}^{h} d^h_{h_i,h_j} \left[A^{h_i}\right]_{io} \left[A^{h_j}\right]_{jo} \ .$$

Proposition 15 (NVAR(p,q): Correlated "Observable Innovations" II)**.**
Let y_t evolve as in Eq. (5) for some $q \in \mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $\tilde{u}_\tau \sim WN(0, \Sigma)$. Define $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{\tilde{y}_{\tau-q}, \tilde{y}_{\tau-q-1}, ...\}$. Then,

$$\mathbb{V}[u_t] = \frac{1}{q^2}\left(\sum_{h=0}^{q-1} \Gamma_h \Sigma \Gamma_h'\right) \ , \quad \Gamma_h = \sum_{m=0}^{h} \Theta_m \ , \quad \Theta_h = \partial \tilde{y}_{\tau+h} / \partial \tilde{y}_\tau \ .$$

Proof: By sequentially inserting for $\tilde{y}_{\tau-1}, \tilde{y}_{\tau-2}$, etc. in the expression for $\tilde{y}_\tau$, we get its VMA(∞)-representation:

$$\tilde{y}_\tau = (I + \Theta_1 L + \Theta_2 L^2 + ... + \Theta_{q-1} L^{q-1} + \Theta_q L^q + ...)\tilde{u}_\tau \ ,$$

where $\Theta_h = \partial \tilde{y}_\tau / \partial \tilde{y}_{\tau-h} = \partial \tilde{y}_{\tau+h} / \partial \tilde{y}_\tau$. Therefore,

$$y_t = \frac{1}{q}(\tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1}) = \frac{1}{q}(I + \Gamma_1 L + \Gamma_2 L^2 + ... + \Gamma_{q-1} L^{q-1} + \Gamma_q L^q + ...)\tilde{u}_\tau \ ,$$

where $\Gamma_h = I + \sum_{m=1}^{h} \Theta_m$. Combining this with the definitions of $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, y_t

and $\mathscr{I}_{t-1}$, we get

$$u_t = \frac{1}{q}(I + \Gamma_1 L + ... + \Gamma_{q-1} L^{q-1})\tilde{u}_\tau \ .$$

In turn, since $\tilde{u}_\tau \sim WN(0, \Sigma)$,

$$\mathbb{V}[u_t] = \frac{1}{q^2}\left(\Sigma + \sum_{h=1}^{q-1} \Gamma_h \Sigma \Gamma_h'\right) \ . \quad \blacksquare$$

If $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$, then, by Proposition 1, for some constants $\{d^h_{h_i,h_j}\}_{h_i,h_j=1:h}$ and $\{d^h_{h_i}\}_{h_i=1:h}$,

$$\begin{aligned}\mathrm{Cov}(u_{it}, u_{jt}) = \frac{1}{q}\sigma_i^2 \mathbf{1}\{i=j\} + &\frac{1}{q^2}\sum_{o=1}^{n}\sigma_o^2 \sum_{h=1}^{q-1}\sum_{h_i,h_j=1}^{h} d^h_{h_i,h_j}\left[A^{h_i}\right]_{io}\left[A^{h_j}\right]_{jo} \\ &+ \frac{1}{q^2}\sigma_j^2 \sum_{h=1}^{q-1}\sum_{h_i=1}^{h} d^h_{h_i}\left[A^{h_i}\right]_{io}\left[A^{h_i}\right]_{ij} \\ &+ \frac{1}{q^2}\sigma_i^2 \sum_{h=1}^{q-1}\sum_{h_j=1}^{h} d^h_{h_j}\left[A^{h_j}\right]_{io}\left[A^{h_j}\right]_{ji} \ .\end{aligned}$$

To illustrate Propositions 14 and 15, consider an NVAR(p, q) for $q = 2$. If $y_t = \tilde{y}_\tau$ for $t = \tau/q \in \mathbb{N}$ (as in Eq. (4)), then $u_t = \tilde{u}_\tau + \alpha_1 A \tilde{u}_{\tau-1}$, with

$$\mathbb{V}[u_t] = \Sigma + \alpha_1^2 A \Sigma A' \ .$$

This reveals that $\mathrm{Cov}(u_{it}, u_{jt}) = \mathbf{1}\{i=j\}\sigma_i^2 + \alpha_1^2 \sum_{k=1}^n a_{ik} a_{jk} \sigma_k^2$. If instead $y_t = (\tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1})/q$ for $t = \tau/q \in \mathbb{N}$ (as in Eq. (5)), then $u_t = \frac{1}{2}\tilde{u}_\tau + \frac{1}{2}(I + \alpha_1 A)\tilde{u}_{\tau-1}$, with

$$\mathbb{V}[u_t] = \frac{1}{2}\Sigma + \frac{1}{4}\alpha_1^2 A \Sigma A' + \frac{1}{4}\alpha_1\left(\Sigma A + (\Sigma A)'\right) \ ,$$

which leads to $\mathrm{Cov}(u_{it}, u_{jt}) = \mathbf{1}\{i=j\}\frac{1}{2}\sigma_i^2 + \frac{1}{4}\alpha_1(a_{ij}\sigma_j^2 + a_{ji}\sigma_i^2) + \frac{1}{4}\alpha_1^2 \sum_{k=1}^n a_{ik} a_{jk} \sigma_k^2$. In the former case, $\mathrm{Cov}(u_{it}, u_{jt})$ is determined by common exposure to third units, while in the latter case bilateral exposure matters as well. In either case, exposure means first-order connections.[51] Under $q = 3$, exposure is determined by first- and second-order connections:

[51] To understand the different results for stock and flow variables, assume for simplicity that there are no self-links: $a_{ii} = 0 \ \forall \ i$. For a stock variable, $y_{it} = \tilde{y}_{i,\tau}$, which implies that u_{it} only depends on $\tilde{u}_{i,\tau}$, not on $\tilde{u}_{i,\tau-1}$: we have $u_{it} = \tilde{u}_{i,\tau} + \alpha_1 \sum_{j=1}^n a_{ij}\tilde{u}_{j,\tau-1}$. As a result, the only (possibly) common terms in u_{it} and u_{jt} are the one-period lagged high-frequency innovations of third units, $\tilde{u}_{k,\tau-1}$, and any comovement between u_{it} and u_{jt} is due to common exposure to these units: $\mathrm{Cov}(u_{it}, u_{jt}) \neq 0$ iff $\exists \ k$ s.t. $a_{ik}, a_{jk} \neq 0$. In contrast, for a flow variable, $y_{it} = \tilde{y}_{i,\tau} + \tilde{y}_{i,\tau-1}$, which implies that u_{it} also depends on $\tilde{u}_{i,\tau-1}$: we have $u_{it} = \tilde{u}_{i,\tau} + \tilde{u}_{i,\tau-1} + \alpha_1 \sum_{j=1}^n a_{ij}\tilde{u}_{j,\tau-1}$. Besides common exposure to third units, comovement is also due to bilateral exposure: $a_{ij} \neq 0 \mid a_{ji} \neq 0 \ \Rightarrow \ \mathrm{Cov}(u_{it}, u_{jt}) \neq 0$.

e.g. for a stock variable, we have $u_t = \tilde{u}_\tau + \alpha_1 A \tilde{u}_{\tau-1} + (\alpha_2 A + \alpha_1^2 A^2)\tilde{u}_{\tau-2}$, with

$$\mathbb{V}[u_t] = \Sigma + (\alpha_1^2 + \alpha_2)A\Sigma A' + \alpha_1^2\alpha_2 \left(A\Sigma(A^2)' + \left[A\Sigma(A^2)'\right]' \right) + \alpha_1^4 A^2 \Sigma (A^2)' \ .$$

Proposition 3 (NVAR(p,q): Limit Distribution of "Observable Innovations")**.**
Let y_t evolve as in Eq. (5) *for some $q \in \mathbb{N}\backslash\{1\}$. Assume y_t is weakly stationary and $\tilde{u}_\tau \sim WN(0,\Sigma)$ is temporally independent. Define $u_t \equiv y_t - \mathbb{E}[y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1} = \{\tilde{y}_{\tau-q}, \tilde{y}_{\tau-q-1}, ...\}$. Also, let $a = \sum_{l=1}^p \alpha_l$. Then, as $q \to \infty$,*

$$\sqrt{q} u_t \xrightarrow{d} N(0, \Gamma_* \Sigma \Gamma_*') \ , \quad \Gamma_* = (I - aA)^{-1} \ .$$

Proof: By the proof of Proposition 15, we know

$$u_t = \frac{1}{q}\Gamma_h \tilde{u}_{\tau-h} \ , \quad \Gamma_h = \sum_{m=0}^{h} \Theta_m \ , \quad \Theta_h = \partial \tilde{y}_{\tau+h} / \partial \tilde{y}_\tau \ .$$

Define $\tilde{u}_h \equiv \Gamma_h \tilde{u}_{\tau-h}$ with $\mathbb{E}[\tilde{u}_h] = 0$ and $V_h \equiv \mathbb{V}[\tilde{u}_h] = \Gamma_h \Sigma \Gamma_h'$. Consider $d'u_t = \frac{1}{q}\sum_{h=0}^{q-1} d'\tilde{u}_h$ for some $d \in \mathbb{R}^n$. By Lyapunov's Central Limit Theorem,

$$\frac{q}{s_q} d' u_t = \frac{1}{s_q} \sum_{h=0}^{q-1} d' \tilde{u}_h \xrightarrow{d} N(0,1) \ ,$$

where $s_q^2 = \sum_{h=0}^{q-1} d' V_h d$. We know $s_q^2/q \to d'V_* d$, $V_* = \Gamma_* \Sigma \Gamma_*'$ (shown below). Therefore, by the above and by Slutsky's theorem,

$$\sqrt{q} d' u_t = \frac{s_q}{\sqrt{q}} \frac{q}{s_q} d' u_t \xrightarrow{d} N(0, d' V_* d) \ .$$

As this argument applies for arbitrary d, by the Cramer-Wold theorem, $\sqrt{q} u_t \xrightarrow{d} N(0, V_*)$.

It remains to show that $s_q^2/q = \frac{1}{q}\sum_{h=0}^{q-1} \sigma_h \to d'V_* d \equiv \sigma_*$, with $\sigma_h = d' V_h d$. By Proposition 5, as $h \to \infty$, $\Gamma_h \to \Gamma_*$. Therefore, $V_h \to \Gamma_* \Sigma \Gamma_*$, and $\sigma_h \to \sigma_*$. In other words, $\forall\, \delta > 0$, $\exists\, H$ s.t. $|\sigma_h - \sigma_*| < \delta \ \forall\, h > H$. Also, for $H \leq q-1$,

$$\begin{aligned} |s_q^2/q - \sigma_*| &= \left| \frac{1}{q} \sum_{h=0}^{q-1} (\sigma_h - \sigma_*) \right| \\ &\leq \frac{1}{q} \sum_{h=0}^{H-1} |\sigma_h - \sigma_*| + \frac{1}{q} \sum_{h=H}^{q-1} |\sigma_h - \sigma_*| \\ &\leq \frac{1}{q} \sum_{h=0}^{H-1} |\sigma_h - \sigma_*| + \delta \ . \end{aligned}$$

The first term vanishes as $q \to \infty$. Hence, $\forall\ \delta > 0$, $\lim_{q\to\infty} |s_q^2/q - \sigma_*| \leq \delta$, i.e. $s_q^2/q \to \sigma_*$. ■

B NVAR: Inference

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

In all of the derivations in this section, A is taken as given, and the explicit conditioning on it is omitted for notational simplicity.

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

The OLS estimator for α from Section 3.1 is given by

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

Proposition 16 (Large n Consistency & Asymptotic Normality of $\hat{\alpha}_{OLS}$)**.**
Suppose

1. *Model is specified correctly:* $y_{it} = x_{n,it}'\alpha_* + u_{it}$.
2. $\mathbb{E}_{t-1}[u_{it}] = 0$.
3. *The observed network adjacency matrix* A_n *converges to some limit* A_* *in the sense that* $\forall\ t$ *and* $l,k = 1:p$, *as* $n \to \infty$,
   (a) $\frac{1}{n}\sum_{i=1}^{n} (A_{n,i\cdot} y_{t-l})' (A_{n,i\cdot} y_{t-k}) \xrightarrow{p} \mathbb{E}\left[(A_{*,i\cdot} y_{t-l})' (A_{*,i\cdot} y_{t-k})\right]$*; and*
   (b) $\frac{1}{n}\sum_{i=1}^{n} (A_{n,i\cdot} y_{t-l})' u_{it} \xrightarrow{p} \mathbb{E}\left[(A_{*,i\cdot} y_{t-l})' u_{it}\right]$.
4. $\mathbb{E}_{t-1}[u_{it}u_{is}] = \sigma^2$ *if* $t = s$ *and zero otherwise.*
5. $\forall\ t$ *and* $l,k = 1:p$, *as* $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (A_{n,i\cdot} y_{t-l})' u_{it} \xrightarrow{d} N\left(\mathbb{E}\left[(A_{*,i\cdot} y_{t-l})' u_{it}\right], \mathbb{V}\left[(A_{*,i\cdot} y_{t-l})' u_{it}\right]\right) .$$

Under conditions 1 - 3, $\hat{\alpha}_{OLS} \xrightarrow{p} \alpha_*$ *as* $n \to \infty$. *Under conditions 1 - 5,*

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

By condition 1,

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

Condition 3 ensures that

$$\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{n,it}x'_{n,it} \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_{*,it}x'_{*,it}] ,$$
$$\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{n,it}u_{it} \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_{*,it}u_{it}]$$

are defined. By condition 2 and the Law of Iterated Expectations (LIE), $\mathbb{E}[x_{*,it}u_{it}] = 0$. As usual, assembling these pieces by Slutsky's theorem yields consistency.

To establish asymptotic Normality, write

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

Condition 5 and Slutsky's theorem ensure that

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{n,it}u_{it} \xrightarrow{d} N\left(0, \mathbb{V}\left[\frac{1}{T}\sum_{t=1}^{T} x_{*,it}u_{it}\right]\right) ,$$

as $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_{*,it}u_{it}\right] = 0$. By condition 4 and LIE,

$$\mathbb{V}\left[\frac{1}{T}\sum_{t=1}^{T} x_{*,it}u_{it}\right] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} x_{*,it}u_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T} x_{*,is}u_{is}\right)'\right] = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[x_{*,it}x'_{*,is}u_{it}u_{is}] = \frac{\sigma^2}{T}\mathbb{E}[x_{*,it}x'_{*,it}]$$

Slutsky's theorem then yields asymptotic Normality with mean zero and variance $\frac{\sigma^2}{T}\mathbb{E}[x_{it}x'_{it}]^{-1}$.

Proposition 17 (Large T Consistency & Asymptotic Normality of $\hat{\alpha}_{OLS}$)**.**
Suppose

1. *Model is specified correctly:* $y_t = X_t\alpha_* + u_t$.
2. $\mathbb{E}_{t-1}[u_t] = 0$.
3. y_t *is ergodic and strictly stationary (SS).*
4. $\mathbb{E}_{t-1}[u_t u'_t] = \Sigma$.

Under conditions 1 - 3, $\hat{\alpha}_{OLS} \xrightarrow{p} \alpha_*$ as $T \to \infty$. Under conditions 1 - 4,

$$\sqrt{T}(\hat{\alpha}_{OLS} - \alpha_*) \xrightarrow{d} N\left(0\ ,\ \mathbb{E}[X_t'X_t]^{-1}\mathbb{E}[X_t'\Sigma X_t]\mathbb{E}[X_t'X_t]^{-1'}\right)\ .$$

By condition 1,

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

By the Weak Law of Large Numbers (WLLN) for ergodic and SS time series (condition 3),

$$\frac{1}{T}\sum_{t=1}^{T}(A_n y_{t-l})'(A_n y_{t-k}) \xrightarrow{p} \mathbb{E}\left[(A_n y_{t-l})'(A_n y_{t-k})\right]$$

so that $\frac{1}{T}\sum_{t=1}^{T} X_t'X_t \xrightarrow{p} \mathbb{E}[X_t'X_t]$. By the same condition and condition 2, $\frac{1}{T}\sum_{t=1}^{T} X_t'u_t \xrightarrow{p} 0$. This establishes consistency.

To establish asymptotic Normality, write

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

By the Central Limit Theorem (CLT) for ergodic and SS time series, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'u_t \xrightarrow{d} N(0, \mathbb{V}[X_t'u_t])$, as $\mathbb{E}[X_t'u_t] = 0$. Thereby, $\mathbb{V}[X_t'u_t] = \mathbb{E}[X_t'u_t u_t'X_t] = \mathbb{E}[X_t'\Sigma X_t]$ by LIE and conditions 2 and 4. Slutsky's theorem then yields asymptotic Normality with mean zero and variance $\mathbb{E}[X_t'X_t]^{-1}\mathbb{E}[X_t'\Sigma X_t]\mathbb{E}[X_t'X_t]^{-1'}$. If $\Sigma = \sigma^2 I$, the latter boils down to $\sigma^2\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}']^{-1}$. If in addition we can write $\mathbb{E}[\sum_{i=1}^{n} x_{it}x_{it}'] = n\mathbb{E}[x_{it}x_{it}']$, it becomes $\frac{\sigma^2}{n}\mathbb{E}[x_{it}x_{it}']^{-1}$.

Proposition 18 (Large (n, T) Consistency & Asymptotic Normality of $\hat{\alpha}_{OLS}$)**.**
Suppose either i) the conditions in Proposition 16 hold, or ii) the conditions in Proposition 17 as well as the following two conditions hold:

1. $\Sigma = \sigma^2 I$
2. $\sum_{i=1}^{n} \mathbb{E}[x_{n,it}x_{n,it}'] = n\mathbb{E}[x_{*,it}x_{*,it}']$

Then, $\sqrt{nT}(\hat{\alpha}_{OLS} - \alpha_*) \xrightarrow{d} N\left(0, \sigma^2\mathbb{E}[x_{*,it}x_{*,it}']^{-1}\right)$ as $(n, T) \to \infty$.

NVAR(p,q), $q \in \mathbb{N}\backslash\{1\}$: Identification

With A given, the problem of identifying α under $\tilde{y}_\tau \sim$ NVAR$(p,1)$ and $\{y_t\}_{t=1}^T = \{\tilde{y}_{tq}\}_{t=1}^T$ for $q \in \mathbb{N}\backslash\{1\}$ is akin to identifying α in the AR(p) $\tilde{x}_\tau = \alpha_1\tilde{x}_{\tau-1} + ... + \alpha_p\tilde{x}_{\tau-p} + \check{u}_\tau$ when the univariate process $\tilde{x}_\tau$ is observed every q periods: $\{x_t\}_{t=1}^T = \{\tilde{x}_{tq}\}_{t=1}^T$. For example, under $p = 1$ and $q = 2$, we have

$$y_t = \alpha_1^2 A^2 y_{t-1} + u_t \ , \quad \text{and} \quad x_t = \alpha_1^2 x_{t-1} + e_t \ ,$$

respectively, and in both cases α_1 is identified only up to sign. While characterization of the identified set remains elusive for the former case for all but $(p = 1, q = 2)$, the latter case provides insights for $q = 2$ and general p. For further discussion, see Palm and Nijman (1984).

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

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

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

Let $\underline{m}$ be the largest odd number in $1 : m$ and $\overline{m}$ the largest even one. For the non-observed $\{\gamma_h\}_{h=1,3,...}$, we have

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

and therefore

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

where only α_l for l even appear in $\overline{A}$ (and its elements are linear in α), and only α_l for l odd appear in $\underline{A}$. For the observed $\{\gamma_h\}_{h=0,2,...}$, we have

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

and therefore

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

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

Posterior Derivations: (α, Σ)

This section derives the conditional (full-sample) posteriors $p(\alpha|\tilde{Y}_{1:T_\tau}, \tilde{\Sigma})$, $p(\tilde{\Sigma}|\tilde{Y}_{1:T_\tau}, \alpha)$, $p(\beta_\alpha|\alpha, \lambda_\alpha)$ and $p(\lambda_\alpha^{-1}|\alpha, \beta_\alpha)$ in an NVAR$(p, 1)$. To simplify notation, I ignore the possibility that $\tilde{Y}_{1:T_\tau}$ has been obtained from a data augmentation step and write $Y_{1:T}$, X_t, u_t and Σ for $\tilde{Y}_{1:T_\tau}$ $\tilde{X}_\tau$, $\tilde{u}_\tau$ and $\tilde{\Sigma}$.

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

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

where $u_t = y_t - \sum_{l=1}^{p} \alpha_l A y_{t-l} = y_t - X_t\alpha$. I write this likelihood in short as $p(Y|\alpha, \Sigma)$.

Under a Uniform prior for α, – $p(\alpha) \propto c$ –, we get

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

which shows that

$$\alpha \mid (Y, \Sigma) \sim N\left(\bar{\alpha}, \bar{V}_\alpha\right) \ ,$$

with

$$\bar{V}_\alpha = \left[\sum_{t=1}^{T} X_t' \Sigma^{-1} X_t\right]^{-1} \ , \quad \bar{\alpha} = \bar{V}_\alpha \left[\sum_{t=1}^{T} X_t' \Sigma^{-1} y_t\right] \ .$$

Under a uniform prior for Σ, we get

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

where U is $T \times n$ and stacks u_t along rows. This shows that

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

The mode of $p(\alpha, \Sigma|Y)$ is equal to the Generalized LS estimator $(\hat{\alpha}, \hat{\Sigma})$, obtained by iterating on the conditional estimators $\hat{\alpha}|\Sigma = \left[\sum_{t=1}^{T} X_t'\Sigma^{-1}X_t\right]^{-1}\left[\sum_{t=1}^{T} X_t'\Sigma^{-1}y_t\right]$ and $\hat{\Sigma}|\alpha = \frac{1}{T}\sum_{t=1}^{T} u_t u_t'$ until convergence.

NVAR(p,q), $q \in \mathbb{N}\backslash\{1\}$: Data Augmentation

The usual formulas for the Kalman filter and Carter & Kohn simulation smoother simplify for the particular state space model characterizing the NVAR. This can be exploited for computational efficiency.

Given an $np \times 1$ vector x, let $[x]_1 = x_{1:n}$ contain the first n elements, $[x]_{-1}$ all but the first n elements, and $[x]_{-p}$ all but the last n elements. Similarly, given an $np \times np$ matrix X, let

$$X = \begin{bmatrix} [X]_{1,1} & [X]_{1,-1} \\ [X]_{-1,1} & [X]_{-1,-1} \end{bmatrix} = \begin{bmatrix} [X]_{-p,-p} & [X]_{-p,p} \\ [X]_{p,-p} & [X]_{p,p} \end{bmatrix} = \begin{bmatrix} [X]_{-p,\cdot} \\ [X]_{p,\cdot} \end{bmatrix} = \begin{bmatrix} [X]_{\cdot,-p} & [X]_{\cdot,p} \end{bmatrix} \ ,$$

where $[X]_{1,1}$ and $[X]_{p,p}$ are $n \times n$, $[X]_{p,\cdot}$ is $n \times (np - p)$ and $[X]_{\cdot,p}$ is $(np - p) \times rn$.

For a stock variable y_t, the NVAR(p,q) for $q \in \mathbb{N}\backslash\{1\}$ leads to the state space model

$$\begin{aligned} y_t &= [I_n, 0_{n\times np-n}]s_\tau \quad \text{if } t = \tau/q \in \mathbb{N} \ , \\ s_\tau &= R s_{\tau-1} + v_\tau \ , \quad v_\tau \sim N(0,\Sigma_v) \ , \quad \text{for } \tau = 1:T_\tau \ , \end{aligned}$$

where $s_t = (\tilde{y}'_\tau, \tilde{y}'_{\tau-1}, ..., \tilde{y}'_{\tau-p+1})'$ and $v_\tau = (u'_\tau, 0, ..., 0)'$ are $np \times 1$, and

$$R = \begin{bmatrix} \alpha_1 A, \alpha_2 A, ..., \alpha_p A \\ I_{np-n\times np-n} \quad 0_{np-n\times n} \end{bmatrix} \quad \text{and} \quad \Sigma_v = \begin{bmatrix} \Sigma_u \quad 0_{n\times np-n} \\ 0_{np-n\times np} \end{bmatrix}$$

are $np \times np$. For notational simplicity, write $\tilde{y}_\tau$ as x_τ.

Algorithm 1 (Kalman Filter for NVAR(p,q), $q \in \mathbb{N}\backslash\{1\}$, for Stock Variables)**.**

1. *Initialize* $s_{0|0} = 0$ *and* $P_{0|0} = \sum_{l=0}^{h} R^l \Sigma_v R^{l\prime}$ *for* h *large.*

2. *For* $\tau = 1 : T_\tau$*, given* $s_{\tau-1|\tau-1}$ *and* $P_{\tau-1|\tau-1}$*,*

   (a): Forecast s_τ*: compute* $s_{\tau|\tau-1}$ *and* $P_{\tau|\tau-1}$ *as*

   - $\left[s_{\tau|\tau-1}\right]_1 = R_{1,\cdot} s_{\tau-1|\tau-1} \quad , \qquad \left[s_{\tau|\tau-1}\right]_{-1} = \left[s_{\tau-1|\tau-1}\right]_{-p} \ ,$
   - $\left[P_{\tau|\tau-1}\right]_{11} = R_{1,\cdot} P_{\tau-1|\tau-1} R'_{1,\cdot} + \Sigma_u \quad , \qquad \left[P_{\tau|\tau-1}\right]_{1,-1} = \left[P_{\tau|\tau-1}\right]'_{-1,1} \ ,$
     $\left[P_{\tau|\tau-1}\right]_{-1,1} = \left[P_{\tau-1|\tau-1}\right]_{-p,\cdot} R'_{1,\cdot} \quad , \qquad \left[P_{\tau|\tau-1}\right]_{-1,-1} = \left[P_{\tau-1|\tau-1}\right]_{-p,-p} \ .$

   (b): Forecast x_τ*: if* $\tau/q \in \mathbb{N}$*, compute* $x_{\tau|\tau-1}$ *and* $F_{\tau|\tau-1}$ *as*

   - $x_{\tau|\tau-1} = \left[s_{\tau|\tau-1}\right]_1 \ ,$
   - $F_{\tau|\tau-1} = \left[P_{\tau|\tau-1}\right]_{11} \ .$

   If $\tau/q \notin \mathbb{N}$*, skip this step.*

   (c): Given observation x_τ*, update forecast for* s_τ*: if* $\tau/q \in \mathbb{N}$*, compute* $s_{\tau|\tau}$ *and* $P_{\tau|\tau}$ *as*

   - $s_{\tau|\tau} = s_{\tau|\tau-1} + \left[P_{\tau|\tau-1}\right]_{\cdot,1} F^{-1}_{\tau|\tau-1}(x_\tau - x_{\tau|\tau-1}) \ ,$
   - $P_{\tau|\tau} = P_{\tau|\tau-1} - \left[P_{\tau|\tau-1}\right]_{\cdot,1} F^{-1}_{\tau|\tau-1} \left[P_{\tau|\tau-1}\right]_{1,\cdot} \ .$

   If $\tau/q \notin \mathbb{N}$*, let* $s_{\tau|\tau} = s_{\tau|\tau-1}$ *and* $P_{\tau|\tau} = P_{\tau|\tau-1}$*.*

Thereby, $R_{1,\cdot} = [\alpha_1 A, \alpha_2 A, ..., \alpha_p] A = (\alpha_1, ..., \alpha_p) \otimes A$, and $\left[P_{\tau|\tau-1}\right]_{1,\cdot} = \left[P_{\tau|\tau-1}\right]'_{\cdot,1}$.

For a flow variable y_t, the NVAR(p, q) for $q \in \mathbb{N}\backslash\{1\}$ leads to the state space model

$$y_t = \Psi s_\tau \quad \text{if } t = \tau/q \in \mathbb{N} \ ,$$
$$s_\tau = R s_{\tau-1} + v_\tau \ , \quad v_\tau \sim N(0, \Sigma_v) \ , \quad \text{for } \tau = 1 : T_\tau \ ,$$

where s_τ, v_τ, R and Σ_v are analogous to above, but with dimensions np' instead of np, where $p' = \max\{p, q\}$. If $p' > p$, we set $\alpha_l = 0$ for $l = (p+1) : p'$. Also, $\Psi = [I_n, ..., I_n, 0_{n\times n(p'-q)}]$ is $n \times np'$. Step (b) in the Kalman filter changes to

- $x_{\tau|\tau-1} = \Psi s_{\tau|\tau-1} = \sum_{l=1}^{q} \left[s_{\tau|\tau-1}\right]_l \ ,$
- $F_{\tau|\tau-1} = \Psi P_{\tau|\tau-1} \Psi' = \sum_{l=1}^{q} \sum_{k=1}^{q} \left[P_{\tau|\tau-1}\right]_{lk} \ ,$

and step (c) changes to

- $s_{\tau|\tau} = s_{\tau|\tau-1} + P_{\tau|\tau-1} \Psi' F_{\tau|\tau-1}^{-1} (x_\tau - x_{\tau|\tau-1}) \ ,$
- $P_{\tau|\tau} = P_{\tau|\tau-1} - P_{\tau|\tau-1} \Psi' F_{\tau|\tau-1}^{-1} \Psi P_{\tau|\tau-1} \ ,$

where $P_{\tau|\tau-1}\Psi' = \sum_{l=1}^{q} \left[P_{\tau|\tau-1}\right]_{\cdot,l} = \left(\Psi P_{\tau|\tau-1}\right)'$.

Algorithm 2 (Carter and Kohn (1994) Simulation Smoother for NVAR(p, q), $q \in \mathbb{N}\backslash\{1\}$)**.**

1. *Run the Kalman filter to get* $\{s_{\tau|\tau}, s_{\tau|\tau-1}, P_{\tau|\tau}, P_{\tau|\tau-1}\}_{\tau=1}^{T_\tau}$.
2. *Draw* $\left[s_{T_\tau}^m\right]_1$ *from* $N\left(\left[s_{T_\tau|T_\tau}\right]_1, \left[P_{T_\tau|T_\tau}\right]_{11}\right)$.
3. *For* $\tau = T_\tau - 1, ..., 0$*, given draw* $\left[s_{\tau+1}^m\right]_1$ *from* $N\left(\left[s_{\tau+1|\tau+2}\right]_1, \left[P_{\tau+1|\tau+2}\right]_{11}\right)$*, draw* $\left[s_\tau^m\right]_1$ *from* $N\left(\left[s_{\tau|\tau+1}\right]_1, \left[P_{\tau|\tau+1}\right]_{11}\right)$ *with*
   - $s_{\tau|\tau+1} = s_{\tau|\tau} + P_{\tau|\tau} R_{1.}' \left(R_{1.} P_{\tau|\tau} R_{1.}' + \Sigma\right)^{-1} \left(\left[s_{\tau+1}^m\right]_1 - \left[s_{\tau+1|\tau}\right]_1\right) \ ,$
   - $P_{\tau|\tau+1} = P_{\tau|\tau} - P_{\tau|\tau} R_{1.}' \left(R_{1.} P_{\tau|\tau} R_{1.}' + \Sigma\right)^{-1} R_{1.} P_{\tau|\tau} \ .$

Relative to the notation used for the Kalman filter, this is with a slight abuse of notation, as $s_{\tau|\tau+1} \neq \mathbb{E}\left[s_\tau | X_{1:\tau}, s_{\tau+1}\right]$ but $s_{\tau|\tau+1} = \mathbb{E}\left[s_\tau | X_{1:\tau}, [s_{\tau+1}]_1\right]$, and similarly $P_{\tau|\tau+1} = \mathbb{V}\left[s_\tau | X_{1:\tau}, [s_{\tau+1}]_1\right]$. See Nelson and Kim (1999, p. 194) for the adjustments of the CKSS required when s_t is a companion form-VAR(1).

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

Posterior Derivations for A

Normal Prior Under independent priors $a_{ij} \sim N(b_{ij}, \lambda_a^{-1})$, the conditional posterior of $A|(\alpha, \Sigma, B, \lambda_a)$ is

$$\begin{aligned} p(A|Y, \alpha, \Sigma, B, \lambda_a) &\propto p(Y|\alpha, A, \Sigma)p(A|B, \lambda_a) \\ &\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - Az_t)'\,\Sigma^{-1}\,(y_t - Az_t)\right\} exp\left\{-\frac{1}{2}\lambda_a \sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2\right\} \\ &= exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\,(Y - ZA')'\,(Y - ZA')\right]\right\} exp\left\{-\frac{1}{2}\lambda_a tr[(A-B)'(A-B)]\right\} , \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left[A(Z'Z + \lambda_a\Sigma)A' - 2A(Z'Y + \lambda_a B'\Sigma)\right]\right]\right\} ,^{52} \end{aligned}$$

which lets us deduce that

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

and therefore

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

Note that $-\log\ p(A|Y, \alpha, \Sigma, B, \lambda_a)$ is proportional to the LS objective function with a Ridge-penalty in Eq. (12). Therefore, its conditional minimizer is the mode of $p(A|Y, \alpha, \Sigma)$. The (joint) minimzer of the objective function in Eq. (12) is the mode of $p(\alpha, A|Y, \Sigma, B, \lambda_a)$ under a uniform prior for α.

Exponential Prior Consider the alternative prior $a_{ij} \sim \text{Exponential}(\lambda_a)$. It leads to the conditional posterior

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

[52] Note that $\sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2 = vec(A-B)'vec(A-B) = tr[(A-B)'(A-B)]$. Also, I use the results that $tr[AB] = tr[BA]$, $tr[A] = tr[A']$ and $c\, tr[A] = tr[cA]$.

where ι is an n-dimensional vector of ones.[53] This leads to

$$A' \mid (Y, \alpha, \Sigma, \lambda_a) \sim MN\left(\bar{A}, \bar{U}_A, \bar{V}_A\right) \;, \quad \text{truncated to } \mathbb{R}_+^{n^2} \;,$$

with $\bar{U}_A = (Z'Z)^{-1}$, $\bar{A} = \bar{U}_A\left[Z'Y - \lambda_a \iota\iota'\Sigma\right]$ and $\bar{V}_A = \Sigma$. Alternatively, this can be written as

$$vec(A') \mid (Y, \alpha, \Sigma, \lambda_a) \sim N\left(vec(\bar{A}), \bar{V}_A \otimes \bar{U}_A\right) \;, \quad \text{truncated to } \mathbb{R}_+^{n^2} \;.$$

Note that $-\log\ p(A|Y, \alpha, \Sigma, \lambda_a)$ is proportional to a LS objective function analogous to Eq. (12), but imposing restrictions $a_{ij} \geq 0$ and using a Lasso-penalty $\lambda_a \sum_{i,j=1}^n |a_{ij}| = \lambda_a \sum_{i,j=1}^n a_{ij}$ to shrink a_{ij} to zero.

This expression simplifies under $\Sigma = I$:

$$A_{i\cdot} \mid (Y, \alpha, \Sigma = I, \lambda_a) \sim N\left((\bar{A}')_{i,\cdot}, \bar{U}_A\right) \;, \quad \text{truncated to } \mathbb{R}_+^n \;,$$

independent across rows i. One can draw from this distribution using Gibbs sampling, iterating on the conditional densities

$$a_{ij} \mid (A_{i,-j}, Y, \alpha, \Sigma = I, \lambda_a) \sim N(\mu^{a_{ij}}, \Sigma^{a_{ij}}) \;, \quad \text{truncated to } \mathbb{R}_+$$

for $j = 1 : n$, whereby

$$\mu^{a_{ij}} = (\bar{A}')_{ij} + (\bar{U}_A)_{j,-j}(\bar{U}_A)^{-1}_{-j,-j}(A_{i,-j} - (\bar{A}')_{i,-j}) \quad \text{and} \quad \Sigma^{a_{ij}} = (\bar{U}_A)_{jj} - (\bar{U}_A)_{j,-j}(\bar{U}_A)^{-1}_{-j,-j}(\bar{U}_A)_{-j,j} \;.$$

Analogously, the mode of this distribution is obtained by iterating on the conditional modes

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

(see Meng and Rubin (1993)). Doing so for all rows i yields the mode of $p(A|Y, \alpha, \Sigma = I, \lambda_a)$, which is the conditional OLS estimator of A.

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

Let $\theta = (\alpha, A)$. As elaborated on above, the OLS estimator solves

$$\hat{\theta} = \arg\min_{\theta \in \Theta} Q_{n,T}(\theta; Y) \;,$$

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

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

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

Proposition 19 (Large T Consistency & Asymptotic Normality of $\hat{\theta} = (\hat{\alpha}, \hat{A})$)**.**
Take $\Theta = [-c, c]^{p-1+n^2}$ for $c > 0$ large such that $\Theta \subset \mathbb{R}^{p-1+n^2}$ is compact, and suppose

1. *$\tilde{\lambda}_{n,T} = o(T^{-\frac{1}{2}})$.*
2. *y_t is ergodic and strictly stationary (SS).*
3. *$\mathbb{E}[X_t'X_t]$ and $\mathbb{E}[z_t z_t']$ are of full rank.*
4. *Model is specified correctly: $y_t = A\tilde{X}_t\alpha + u_t$.*
5. *$\mathbb{E}_{t-1}[u_t] = 0$.*
6. *$\mathbb{E}_{t-1}[u_t u_t'] = \Sigma$.*

Under conditions 1 - 3, $\hat{\theta} = (\hat{\alpha}, \hat{A}) \xrightarrow{p} \theta_0$ as $T \to \infty$. Under conditions 1 - 5,

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

with H and M defined below.

By conditions 1 and 2, $Q_{n,T}(\theta; Y)$ converges uniformly in probability to the limit objective function $Q(\theta) = \frac{1}{n}\mathbb{E}\left[u_t(\theta)'u_t(\theta)\right]$, which is continuous on Θ:

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

converges in probability to zero because, under condition 1,

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

where $\tilde{c} = \max_{i,j}(c + b_{ij})^2$, while under condition 2, $\frac{1}{T}\sum_{t=1}^{T} u_t(\theta)'u_t(\theta) \xrightarrow{p} \mathbb{E}\left[u_t(\theta)'u_t(\theta)\right]$ by WLLN for ergodic and SS time series. Finally, under condition 3, $Q(\theta)$ is uniquely minimized by $\theta_0 = (\alpha_*, A_*)$ defined by the first-order conditions (FOC)

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

Note that for c large enough, we necessarily get a solution $\theta_0 \in int(\Theta)$. Without the imposed normalization, θ_0 would not be unique, as for any (α_*, A_*) that solves the above, $(k\alpha_*, k^{-1}A_*)$ for any $k \in \mathbb{R}$ does, too, because $X_t(k^{-1}A_*) = k^{-1}X_t(A_*)$ and $z_t(k\alpha_*) = kz_t(\alpha_*)$.[54]

Write $\overrightarrow{A}$ for $vec\,(A)$. Note that $\sqrt{T}\tilde{\lambda} \to 0$ by condition 1. By condition 2 and the CLT for ergodic and SS time series,

$$\sqrt{T}Q^{(1)}_{n,T}(\theta_0; Y) = \sqrt{T}\begin{bmatrix} \frac{\partial Q_{n,T}(\theta;Y)}{\partial \alpha} \\ \frac{\partial Q_{n,T}(\theta;Y)}{\partial \overrightarrow{A}} \end{bmatrix}\Bigg|_{\theta=\theta_0}$$

$$= -2\begin{bmatrix} \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'(y_t - X_t\alpha_*) \\ \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\overrightarrow{(y_t - A_* z_t)z_t'}] - \sqrt{T}\tilde{\lambda}[\overrightarrow{A_* - B}] \end{bmatrix} \overset{d}{\to} N(0, M)\ ,$$

because

$$-\frac{2}{n}\begin{bmatrix} \mathbb{E}\left[X_t'(y_t - X_t\alpha_*)\right] \\ \mathbb{E}[\overrightarrow{(y_t - A_* z_t)z_t'}] \end{bmatrix} = -\frac{2}{n}\begin{bmatrix} \mathbb{E}\left[X_t' u_t\right] \\ \mathbb{E}[\overrightarrow{u_t z_t'}] \end{bmatrix} = 0$$

by conditions 3 and 4. Using conditions 4, 5 and 6 as well as LIE,

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

Furthermore, using again the WLLN for ergodic and SS time series as well as conditions 4 and 5,

$$Q^{(2)}_{n,t}(\theta_0; Y) = \begin{bmatrix} \frac{\partial Q_{n,T}(\theta;Y)}{\partial \alpha \partial \alpha'} & \frac{\partial Q_{n,T}(\theta;Y)}{\partial \alpha \partial \overrightarrow{A}'} \\ \frac{\partial Q_{n,T}(\theta;Y)}{\partial \overrightarrow{A} \partial \alpha'} & \frac{\partial Q_{n,T}(\theta;Y)}{\partial \overrightarrow{A} \partial \overrightarrow{A}'} \end{bmatrix} \overset{p}{\to} \begin{bmatrix} H_{11} & H_{21}' \\ H_{21} & H_{22} \end{bmatrix} \equiv H \quad \forall\, \theta \overset{p}{\to} \theta_0\ ,$$

[54]With α_l dropped, it still holds that $X_t(k^{-1}A_*) = k^{-1}X_t(A_*)$ and $z_t(k\alpha_*) = kz_t(\alpha_*)$, but the y_t in the expression for $\alpha_*|A_*$ is in fact $y_t - Ay_{t-l}$, while it is unchanged in the expression for $A_*|\alpha_*$. This renders the solution unique.

with

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

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

$$\frac{\partial Q_{n,T}(\theta;Y)}{\partial\overrightarrow{A}\partial\overrightarrow{A}'} = \frac{2}{nT}\sum_{t=1}^{T}\begin{bmatrix}z_t'\otimes z_{1t}I_n \\ \vdots \\ z_t'\otimes z_{nt}I_n\end{bmatrix} \xrightarrow{p} \frac{2}{n}\begin{bmatrix}\mathbb{E}[z_t'\otimes z_{1t}I_n] \\ \vdots \\ \mathbb{E}[z_t'\otimes z_{nt}I_n]\end{bmatrix} \equiv H_{22} \ .$$[55]

Consistency and asymptotic Normality also apply under a Lasso-penalty for A, although no analytical expression for the conditional estimator can be found in that case. Under $a_{ij} \geq 0$, only consistency goes through as A_* is (likely) not interior.

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

C Business Cycles by Lagged IOC

C.1 Theory

RBC Model with Contemporaneous Input-Output Conversion

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

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

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

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

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

$$\tilde{p}_i = k_i^p + \sum_{j=1}^n a_{ij} \tilde{p}_j - \tilde{z}_i \quad \Leftrightarrow \quad \tilde{p} = k^p + A\tilde{p} - \tilde{z} \ ,$$

where $\tilde{p}_i = ln(p_i/w)$ and $\tilde{z}_i = ln(z_i)$. 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.

The representative household's problem is

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

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

Market clearing for good j requires $y_j = c_j + \sum_{i=1}^n x^{ij}$. 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:

$$\lambda_j \equiv \frac{y_j p_j}{w} = \gamma_j + \sum_{i=1}^n a_{ij} \lambda_i \quad \Leftrightarrow \quad \lambda = \gamma + A'\lambda \ .$$

As a result, $\lambda = (I - A')^{-1}\gamma$. The Domar weight of sector i reflects its importance as a supplier to relevant sectors in the economy, where relevance is defined by households' expenditure shares $\{\gamma_j\}_{j=1}^n$: $\lambda_i = \sum_{j=1}^n \gamma_j l_{ji}$. l_{ji} is element (j,i) of the Leontief-inverse $(I-A)^{-1}$. It sums up connections of all order from a sector j to a sector i and therefore shows how important sector i is in j's supply chain. Using the definition of λ_i, we have $\tilde{y} = \tilde{\lambda} - \tilde{p}$, where $\tilde{y} = ln(y)$ and $\tilde{\lambda} = ln\left(\lambda\right)$. Combining this with the equation for $\tilde{p}$ yields

$$\tilde{y} = k^y + A\tilde{y} + \tilde{z} \ ,$$

with $k^y = (I-A)\tilde{\lambda} - k^p$. For the sake of completeness, labor market clearing requires $\sum_{i=1}^n l_i = 1$ and gives $w = \sum_{i=1}^n b_i p_i y_i$.

RBC Model with Single-Lag Input-Output Conversion

Assume good j used in production at time t is purchased at time $t-1$: $x_{ijt} = x_{t-1}^{ij}$. Firm i's value function is then:

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

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

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

Plugging these expressions into the production function and taking logs gives

$$\tilde{p}_{it} = k_{it}^{p1} + \sum_{j=1}^n a_{ij} \tilde{p}_{j,t-1} - \tilde{z}_{it} \quad \Leftrightarrow \quad \tilde{p}_t = k_t^{p1} + A\tilde{p}_{t-1} - \tilde{z}_t \ ,$$

where again $\tilde{p}_{it} = ln\left(\frac{p_{it}}{w_t}\right)$ and $\tilde{z}_{it} = ln(z_{it})$. Also, $k_{it}^{p1} = k_i^{p1} - (1-b_i)\tilde{G}_{t,t-1}^w$ with $k_i^{p1} = -\left[b_i ln(b_i) + \sum_{j=1}^n a_{ij} ln(\beta a_{ij})\right]$ and $\tilde{G}_{t,t-1}^w = ln(G_{t,t-1}^w)$, $G_{t,t-1}^w = w_t / w_{t-1}$.

Provided that in every period t households spend all their period t income, w_t, we again get $c_{it} = \gamma_i w_t / p_{it}$. By market clearing of good j, then,

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

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

$$\lambda_{jt} \equiv \frac{y_{jt} p_{jt}}{w_t} = \gamma_j + \sum_{i=1}^n \beta a_{ij} G_{t+1,t}^w \lambda_{i,t+1} \quad \Leftrightarrow \quad \lambda_t = \gamma + \beta G_{t+1,t}^w A' \lambda_{t+1} \ .$$

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

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

For output $\tilde{y}_t = \tilde{y}_t - \tilde{p}_t$, we obtain

$$\tilde{y}_t = k^{y1}_t + A\tilde{y}_{t-1} + \tilde{z}_t \ ,$$

where $k^{y1}_t = \tilde{\lambda}_t - A\tilde{\lambda}_{t-1} - k^{p1}_t$ and again $\tilde{y}_t = ln(y_t)$ and $\tilde{\lambda}_t = ln(\lambda_t)$.

In the steady state (SS) with $\tilde{z}_t = \tilde{z} \; \forall \; t$ we get

$$\lambda = (I - \beta A')^{-1}\gamma \ , \quad \tilde{p} = (I - A)^{-1}(k^{p1} - \tilde{z}) \ , \quad \tilde{y}_t = (I - A)^{-1}(k^{y1} + \tilde{z}) \ ,$$

where k^{p1} contains elements k^{p1}_i, and $k^{y1} = (I - A)\tilde{\lambda} - k^{p1}$. Relative to the static economy above, the meaning of a_{ij} changes slightly: in SS, it equals

$$a_{ij} = \beta^{-1}(p_j x^{ij})/(p_i y_i) \ .$$

Taking this into account, the SS value of λ is unchanged. $\tilde{p}$ is slightly higher than in the static economy, as $k^{p1}_i = k^p_i - (1 - b_i)ln(\beta) > k^p_i$. For the same reason, $\tilde{y}$ is slightly lower. These differences vanish as $\beta \rightarrow 1$.

RBC Model with Multiple-Lags Input-Output Conversion

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

$$\max_{\substack{\{l_{it}, \{x^{ij}_t, x^{ij}_{t,t-1}, \\ x^{ij}_{t,t-2}\}^n_{j=1}\}^\infty_{t=0}}} \sum_{t=0}^{\infty} \beta^t \Pi_{it} \quad \text{s.t.} \quad x^{ij}_t = x^{ij}_{t+1,t} + x^{ij}_{t+2,t} \; \forall \; t, i, j \ ,$$

$$\Pi_{it} = p_{it} z_{it} l^{b_i}_{it} \prod_{j=1}^{n} \left[\alpha_1 \left(x^{ij}_{t,t-1}\right)^r + \alpha_2 \left(x^{ij}_{t,t-2}\right)^r\right]^{\frac{a_{ij}}{r}} - w_t l_{it} - \sum_{j=1}^{n} p_{jt} x^{ij}_t \ .$$

For each input j, the firm chooses how much to buy in period t, x^{ij}_t, and how to distribute the purchased amount for production over periods $t+1, t+2$. Abstracting from perfect substitutability allows me to ignore the boundary constraints $l_{it}, x^{ij}_{t+1,t}, x^{ij}_{t+2,t} \geq 0 \; \forall \; t, i, j$.

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

$$V_i\left(\{\breve{x}^{ij}_{t,t-2},\breve{x}^{ij}_{t,t-1}\}^n_{j=1}\right)=\max_{\substack{l_{it},\{x^{ij}_t,\\ x^{ij}_{t,t-1},x^{ij}_{t,t-2}\}_j}}\ \Pi_{it}+\beta V_i\left(\{\breve{x}^{ij}_{t+1,t-1},\breve{x}^{ij}_{t+1,t}\}^n_{j=1}\right)$$
$$\text{s.t.}\qquad \breve{x}^{ij}_{t,t-2}=x^{ij}_{t,t-2}\ ,\quad \breve{x}^{ij}_{t+1,t}=x^{ij}_t\ ,\quad \breve{x}^{ij}_{t,t-1}=x^{ij}_{t,t-1}+x^{ij}_{t+1,t-1}\ .$$

The problem can be written more compactly as

$$V_i\left(\{x^{ij}_{t,t-2},\breve{x}^{ij}_{t,t-1}\}^n_{j=1}\right)=\max_{\substack{l_{it},\{x^{ij}_t,\\ x^{ij}_{t+1,t-1}\}_j}}\ \left[p_{it}z_{it}l^{b_i}_{it}\prod_{j=1}^n\left[\alpha_1\left(\breve{x}^{ij}_{t,t-1}-x^{ij}_{t+1,t-1}\right)^r+\alpha_2\left(x^{ij}_{t,t-2}\right)^r\right]^{\frac{a_{ij}}{r}}\right.$$
$$\left.-w_tl_{it}-\sum_{j=1}^n p_{jt}x^{ij}_t\right]+\beta V_i\left(\{x^{ij}_{t+1,t-1},x^{ij}_t\}^n_{j=1}\right)$$

In each period t, and for each input j, a firm only chooses how how much to buy in period t – to be used 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 Inputs Purchased in Past Under $r\to 0$, we have $x_{ijt}=\left(x^{ij}_{t,t-1}\right)^{\alpha_1}\left(x^{ij}_{t,t-2}\right)^{\alpha_2}$ and the optimality conditions yield

$$l_{it}=b_i\frac{p_{it}y_{it}}{w_t}\ ,\quad x^{ij}_{t,t-1}=\beta\alpha_1a_{ij}\frac{p_{it}y_{it}}{p_{j,t-1}}\ ,\quad x^{ij}_{t,t-2}=\beta^2\alpha_2a_{ij}\frac{p_{it}y_{it}}{p_{j,t-2}}\ .$$

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

$$\tilde{p}_{it}=k^{p2}_t+\sum_{j=1}^n a_{ij}\left[\alpha_1\tilde{p}_{j,t-1}+\alpha_2\tilde{p}_{j,t-2}\right]-\tilde{z}_t\quad\Leftrightarrow\quad \tilde{p}_t=k^{p2}_t+\alpha_1A\tilde{p}_{t-1}+\alpha_2A\tilde{p}_{t-2}-\tilde{z}_t\ ,$$

where $k^{p2}_{it}=k^{p2}_i-(1-b_i)\left[\alpha_1\tilde{G}^w_{t,t-1}+\alpha_2\tilde{G}^w_{t,t-2}\right]$ and $k^{p2}_i=-b_iln(b_i)-\sum_{j=1}^n a_{ij}\left[\alpha_1ln(\beta\alpha_1a_{ij})+\alpha_2ln(\beta^2\alpha_2a_{ij})\right.$

The market clearing condition for good j is now

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

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

$$\lambda_{jt} = \gamma_j + \beta\alpha_1 G^w_{t,t-1} \sum_{i=1}^{n} a_{ij}\lambda_{i,t+1} + \beta^2\alpha_2 G^w_{t,t-2} \sum_{i=1}^{n} a_{ij}\lambda_{i,t+2} \ .$$

When stacked for all i, one could solve forward to obtain λ_t. For output we get then

$$\tilde{y}_t = k^{y2}_t + \alpha_1 A\tilde{y}_{t-1} + \alpha_2 A\tilde{y}_{t-2} + \tilde{z}_t \ ,$$

where $k^{y2}_t = \tilde{\lambda}_t - \alpha_1 A\tilde{\lambda}_{t-1} - \alpha_2 A\tilde{\lambda}_{t-2} - k^{p2}_t$.

In the SS with $\tilde{z}_t = \tilde{z} \ \forall \ t$ we get

$$\lambda = (I - (\beta\alpha_1 + \beta^2\alpha_2)A')^{-1}\gamma \ , \quad \tilde{p} = (I - A)^{-1}(k^{p2} - \tilde{z}) \ , \quad \tilde{y} = (I - A)^{-1}(k^{y2} + \tilde{z}) \ .$$

In this economy, we have

$$a_{ij} = \left[\beta\alpha_1 + \beta^2\alpha_2\right]^{-1} (p_j x^{ij})/(p_i y_i) \ .$$ [56]

Again, $\tilde{\lambda}$ is unaltered relative to the static economy, while $\tilde{p}$ increases and $\tilde{y}$ decreases, owing to the increase in $k^{p2}_i = k^p_i - (1 - b_i)ln\left((\beta\alpha_1)^{\alpha_1}(\beta^2\alpha_2)^{\alpha_2}\right) > k^p_i$. Differences to the economy with one period-lagged input-output conversion vanish as $\alpha_1 \to 1$, and differences to the economy with contemporaneous input-output conversion vanish as $\beta \to 1$ and either $\alpha_1 \to 1$ or $\alpha_1 \to 0$. Decreasing α_1 starting from $\alpha_1 = 1$ decreases k^{p2}_i and therefore increases prices and decreases output.

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

$$l_{it} = b_i \frac{y_{it}p_{it}}{w_t} \ , \quad x^{ij}_{t,t-1} = \left[a_{ij}\alpha_1\beta \frac{y_{it}p_{it}}{p_{jt-1}x^r_{ijt}}\right]^{\frac{1}{1-r}} \ , \quad x^{ij}_{t,t-2} = \left[a_{ij}\alpha_2\beta^2 \frac{y_{it}p_{it}}{p_{jt-2}x^r_{ijt}}\right]^{\frac{1}{1-r}} \ ,$$

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

$$x_{ijt} = a_{ij}p_{it}y_{it}\left[n_1 p_{j,t-1}^{-\frac{r}{1-r}} + n_2 p_{j,t-2}^{-\frac{r}{1-r}}\right]^{\frac{1-r}{r}} \ , \quad n_1 = \alpha_1(\alpha_1\beta)^{\frac{r}{1-r}} \ , \ n_2 = \alpha_2(\alpha_2\beta^2)^{\frac{r}{1-r}} \ .$$

In turn, inserting this equation back for x_{ijt} in the expressions above yields:

$$x^{ij}_{t,t-1} = a_{ij}\frac{p_{it}y_{it}}{p_{j,t-1}} \frac{n_1^{-r}}{n_1 + n_2\left(\frac{p_{j,t-1}}{p_{j,t-2}}\right)^{\frac{r}{1-r}}} \ , \quad x^{ij}_{t,t-2} = a_{ij}\frac{p_{it}y_{it}}{p_{j,t-2}} \frac{n_2^{-r}}{n_1\left(\frac{p_{j,t-1}}{p_{j,t-2}}\right)^{-\frac{r}{1-r}} + n_2} \ .$$

[56] Note that $x^{ij} \neq x_{ij}$. In terms of x_{ij}, we have $a_{ij} = x_{ij}p_j/(p_i y_i)(\beta\alpha_1)^{-\alpha_1}(\beta^2\alpha_2)^{-\alpha_2}$.

Inserting the expression for x_{ijt} into the production function gives

$$1 = z_{it} b_i^{b_i} p_{it}^* \prod_{j=1}^{n} a_{ij}^{a_{ij}} \left[n_1 \frac{G_{t,t-1}^w}{p_{j,t-1}^*}^{\frac{r}{1-r}} + n_2 \frac{G_{t,t-2}^w}{p_{j,t-2}^*}^{\frac{r}{1-r}} \right]^{a_{ij}\frac{1-r}{r}} ,$$

where $p_{it}^* = p_{it}/w_t$. Linearizing around the SS characterized below leads to

$$\hat{p}_{it}^* = c_{it}^{p2} + \sum_{j=1}^{n} a_{ij} \left[\chi_1 \hat{p}_{j,t-1}^* + \chi_2 \hat{p}_{j,t-2}^* \right] - \hat{z}_{it} \quad \Leftrightarrow \quad \hat{p}_t^* = c_t^{p2} + \chi_1 A \hat{p}_{t-1}^* + \chi_2 A \hat{p}_{t-2}^* - \hat{z}_t ,$$

where $c_{it}^{p2} = -(1-b_i) \left[\chi_1 \hat{G}_{t,t-1}^w + \chi_2 \hat{G}_{t,t-2}^w \right]$, $\chi_1 = n_1/(n_1+n_2)$ and $\chi_2 = 1-\chi_1$. Inserting the optimal choices of c_{jt}, $x_{t+1,t}^{ij}$ and $x_{t+2,t}^{ij}$ into the market clearing condition for good j yields

$$\lambda_{jt} = \gamma_j + \sum_{i=1}^{n} a_{ij} \frac{n_1^{-r}}{n_1 + n_2 \left(\frac{p_{i,t}}{p_{i,t-1}} \right)^{\frac{r}{1-r}}} G_{t+1,t}^w \lambda_{i,t+1} + \sum_{i=1}^{n} a_{ij} \frac{n_2^{-r}}{n_1 \left(\frac{p_{i,t}}{p_{i,t-1}} \right)^{-\frac{r}{1-r}} + n_2} G_{t+2,t}^w \lambda_{i,t+2} .$$

Using the relation $\hat{y}_t = \hat{\lambda}_t - \hat{p}_t^*$, we get

$$\hat{y}_t = c_t^{y2} + \chi_1 A \hat{y}_{t-1} + \chi_2 A \hat{y}_{t-2} + \hat{z}_t ,$$

where $c_t^{y2} = c_t^{p2} + \hat{\lambda}_t - \chi_1 A \hat{\lambda}_{t-1} - \chi_2 A \hat{\lambda}_{t-2}$.

In SS,

$$\lambda = \left(I - \frac{n_1^{-r} + n_2^{-r}}{n_1 + n_2} A' \right)^{-1} \gamma , \quad \tilde{p} = (I - A)^{-1} (k^{p3} - \tilde{z}) , \quad \tilde{y} = (I - A)^{-1} (k^{y3} + \tilde{z}) ,$$

where k^{p3} contains elements $k_i^{p3} = -b_i ln(b_i) - \sum_{j=1}^{n} a_{ij} ln(a_{ij}) - (1-b_i) \frac{1-r}{r} ln(n_1 + n_2)$ and $k^{y3} = (I - A)\tilde{\lambda} - k^{p3}$. Also,

$$a_{ij} = \left[\frac{n_1^{-r} + n_2^{-r}}{n_1 + n_2} \right]^{-1} (p_j x^{ij})/(p_i y_i) .$$

C.2 Data

...

C.3 Estimation

Likelihood Evaluation

Under $q = 1$, the model with contemporaneous IOC (Eqs. (15) and (17)) yields the following state space form:

$$s_t = \Phi_0 + \Phi_1 s_{t-1} + v_t \ , \quad \Delta y_t = \Psi s_t \ ,$$

where $s_t = (\Delta y_t', e_t', e_t^a)'$ and $v_t = (0, \varepsilon_t', \varepsilon_t^a)' \sim N(0, \Sigma_v)$ are $(2n+1) \times 1$, $\Psi = \left[I_n, 0\right]$, and

$$\Phi_0 = \begin{bmatrix} L\gamma \\ 0 \\ 0 \end{bmatrix} \ , \quad \Phi_1 = \begin{bmatrix} 0 & L & L\lambda \\ 0 & P & 0 \\ 0 & 0 & \rho_a \end{bmatrix} \ , \quad \Sigma_v = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma & 0 \\ 0 & 0 & \sigma_a^2 \end{bmatrix} \ ,$$

and we further define $L = (I - A)^{-1}$ to be the Leontief-inverse and P and Σ to be diagonal matrices containing $(\rho_1, ..., \rho_n)$ and $(\sigma_1^2, ..., \sigma_n^2)$, respectively.

To write the model with lagged IOC (Eqs. (16) and (17)) in state space form, let $\tilde{p} = \max\{p, q\}$ and $\alpha_l = 0$ for $l = (p+1) : \tilde{p}$. Define the $(n\tilde{p} + n + 1) \times 1$ vectors $s_\tau = (\Delta\tilde{y}_\tau', ..., \Delta\tilde{y}_{\tau-\tilde{p}+1}', e_\tau', e_\tau^a)'$ and $v_\tau = (0, ..., 0, \varepsilon_\tau', \varepsilon_\tau^a)'$. We have the analogous state space form as above, with

$$\Phi_0 = \begin{bmatrix} \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix} \ , \quad \Phi_1 = \begin{bmatrix} \alpha_1 A & \dots & \alpha_{\tilde{p}-1} A & \alpha_{\tilde{p}} A & I & \lambda \\ & I_{np-n} & & 0 & 0 & 0 \\ & 0 & & 0 & P & 0 \\ & 0 & & 0 & 0 & \rho_a \end{bmatrix} \ , \quad \Sigma_v = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma & 0 \\ 0 & 0 & 0 & \sigma_a^2 \end{bmatrix} \ .$$

Under $q = 1$, we observe $\Delta y_t = \Delta\tilde{y}_\tau$, meaning that $\Psi = [I_n, 0]$. Under $q > 1$, we have observe $\Delta y_t = \sum_{l=0}^{q-1} \Delta\tilde{y}_{\tau-l}$ – i.e. $\Psi = [I_n, ..., I_n, 0]$ – only for periods $t = \tau/q \in \mathbb{N}$ (see Section B).

Prior-Construction & -Drawing

The Uniform prior for $\alpha \in [0,1]^{p-1} \cap \{\alpha : \sum_{l=1}^{p-1} \alpha_l \leq 1\}$ can be broken up as follows:

$$p(\alpha_1, ..., \alpha_{p-1}) = p(\alpha_1|\alpha_2, ..., \alpha_{p-1}) p(\alpha_2|\alpha_3, ..., \alpha_{p-1}) ... p(\alpha_{p-2}|\alpha_{p-1}) p(\alpha_{p-1}) \ ,$$

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

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

and

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

To draw from $p(\alpha_1, ..., \alpha_{p-1})$, I draw α_{p-1} from its marginal distribution and iteratively draw $\alpha_{p-2}, ..., \alpha_1$ from the conditionals, using the inverse-cdf method in each step: to draw $y_i \sim f(y)$, I draw $x_i \sim \mathcal{U}(0,1)$ and find y_i so that $\int_{-\infty}^{y_i} f(y)dy = x_i$. In the present case, this yields

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

for $l = 1 : (p-2)$, and $\alpha_{p-1} = 1 - (1-x_{p-1})^{1/(p-1)}$, $x_{p-1} \sim \mathcal{U}(0,1)$.

For the parameters $(\sigma_a^2, \sigma^{2\prime})' \in \mathbb{R}_{++}^{n+1}$, an upper bound has to be chosen so as to specify a proper Uniform prior distribution, since draws from it are needed to initialize the SMC sampler. A high upper bound is desirable so as to avoid domain restrictions in areas with non-trivial likelihood values. However, for efficient sampling, lower values are preferred. I choose a low upper bound to initialize the sampler, but allow the algorithm to disrespect it in search for parameter-values associated with high likelihood values and therefore high posterior mass. Ex-post, I redefine the upper bound as the largest posterior draw among $(\sigma_a^2, \sigma^{2\prime})'$ across all models. The resulting MDDs could be re-scaled to take this into account, but this is not necessary, as all models' MDD is "off" by the same proportionality constant, which means that the ranking is unaffected.[57] I set $\bar{\sigma}_i^2 = 0.5\mathbb{V}[\Delta \tilde{y}_{it}]$ and $\bar{\sigma}_a^2 = \max_i \bar{\sigma}_i^2$.

SMC Algorithm & Parameter Transformation

I use the adaptive tempering variant of the SMC algorithm (see Cai et al. (2021)), which ensures a precise estimation of the posterior even as the distance between the prior and posterior distributions is difficult to assess. For efficient sampling even under tight domain restrictions, I reparameterize the parameters in the mutation step of the SMC algorithm as follows. Define the function g s.t. $\check{\theta} = g^{-1}(\theta)$ is generated by transforming α_l into $\ln \alpha_l/\alpha_p$ for $l = 1 : (p-1)$, σ_i^2 into $\ln \sigma_i^2$ for $i = 1 : n$, and ρ_i into $\ln \rho_i/(1-\rho_i)$ for $i = 1 : n$ and $i = a$. These are one-to-one mappings and ensure that $\check{\theta} \in \mathbb{R}$. As a result, no draws in the mutation

[57] Let $s = (\sigma_a^2, \sigma^{2\prime})' \in [0, \bar{s}]^{n+1}$. We get the MDD

$$p(Y) = \int p(Y|\theta_{-s}, s) p(\theta_{-s}) p(s) d(\theta_{-s}, s) = \bar{s}^{-(n+1)} \int p(Y|\theta_{-s}, s) p(\theta_{-s}) d(\theta_{-s}, s) .$$

If the upper bound is ex-post rescaled to s^*, the MDD needs to be corrected by multiplying by $(\bar{s}/s^*)^{n+1}$. The analogous holds with heterogeneous upper bounds.

step are rejected because of domain violations. To account for this reparameterization, the acceptance probabilities in the mutation of particle i in iteration n are adjusted as follows:

Algorithm 3 (Particle Mutation in SMC Algorithm)**.**

1. *Given particle* θ_{n-1}^i*, set* $\theta_n^{i,0} = \theta_{n-1}^i$.
2. *For* $m = 1 : N_{MH}$*:*
   - *Compute* $\check{\theta}_n^{i,m-1} = g^{-1}(\theta_n^{i,m-1})$ *and draw*
   $$\check{v}|\theta_n^{i,m-1} \sim \check{q}(\check{v}|\theta_n^{i,m-1}) = N(\check{\theta}_n^{i,m-1}, c_n^2\Sigma_n) = N(g^{-1}(\theta_n^{i,m-1}), c_n^2\Sigma_n)\ .$$
   - *Set*
   $$\theta_n^{i,m} = \begin{cases} v = g(\check{v}) & w.p. \quad \alpha(v|\theta_n^{i,m-1}) \\ \theta_n^{i,m-1} & otherwise \end{cases} ,$$
   where
   $$\alpha(v|\theta_n^{i,m-1}) = min\ \left\{1, \frac{p(Y|v)p(v)/q(v|\theta_n^{i,m-1})}{p(Y|\theta_n^{i,m-1})p(\theta_n^{i,m-1})/q(\theta_n^{i,m-1}|v)}\right\}\ .$$
   The densities $q(v|\theta_n^{i,m-1})$ *and* $q(\theta_n^{i,m-1}|v)$ *are obtained using analogous density transformations starting from* $q(\check{v}|\theta_n^{i,m-1})$ *and* $q(\check{\theta}_n^{i,m-1}|v)$*, respectively; e.g.*
   $$q(v|\theta_n^{i,m-1}) = \check{q}(g^{-1}(v)|\theta_n^{i,m-1})|J(v)|\ .$$
   where $J(\theta)$ *is the Jacobian matrix.*
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

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

The Jacobian matrix $J(\theta) = \partial\check{\theta}/\partial\theta'$ is diagonal and leads to

$$|J(\theta)| = \left\{\prod_{i=1}^{n}\rho_i(1-\rho_i)\sigma_i^2\right\}^{-1}\left\{\prod_{l=1}^{p-1}\alpha_l\right\}^{-1}\{\rho_a(1-\rho_a)\}^{-1}\ .$$

C.4 Results

D Dimensionality-Reduction by Parsimonious NVAR

D.1 Hyperparameter Selection

Marginal Posterior Mode of λ_a Under a Normal prior for A and a uniform hyperprior for λ_a, we obtain

$$\begin{aligned} p(\lambda_a|A,B) &\propto p(A|B,\lambda_a) \\ &\propto \lambda_a^{n^2/2} exp\left\{-\frac{1}{2}\lambda_a tr\left[(A-B)'(A-B)\right]\right\} \ , \end{aligned}$$

whereby $tr\left[(A-B)'(A-B)\right] = \sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2$. This shows that

$$\lambda_a \mid (A,B) \sim G\left(\frac{n^2}{2}+1\ ,\ \frac{1}{2}tr\left[(A-B)'(A-B)\right]\right) \ .$$

Under an exponential prior for A and a uniform hyperprior for λ_a, we obtain

$$\begin{aligned} p(\lambda_a|A) &\propto p(A|\lambda_a) \\ &\propto \lambda_a^{n^2} exp\left\{-\lambda_a \iota' A \iota\right\} \ , \end{aligned}$$

which shows that

$$\lambda_a \mid A \sim G\left(n^2+1\ ,\ \iota' A \iota\right) \ .$$

Conditional MDD under NVAR-R By Bayes' theorem,

$$p(A|Y,\lambda_a,B,\cdot) = \frac{p(Y|A,\cdot)p(A|\lambda_a,B)p(\lambda_a)}{p(Y|\lambda_a,\cdot)} \ ,$$

where $\cdot$ stands for the remaining parameters affecting the likelihood, (α,Σ). The conditional MDD $p(Y|\lambda_a,\cdot)$ is obtained by rearranging this formula, inserting the known expressions for the densities $p(Y|A,\cdot)$, $p(A|\lambda_a,B)$, $p(\lambda_a)=c$ and $p(A|Y,\lambda_a,B,\cdot)$, and cancelling all terms involving A:

$$\begin{aligned} p(Y|\lambda_a,\cdot) &= \frac{p(Y|A,\cdot)p(A|\lambda_a,B)p(\lambda_a)}{p(A|Y,\lambda_a,B,\cdot)} \\ &= c\frac{(2\pi)^{-\frac{nT}{2}}|\Sigma|^{-\frac{T}{2}} exp\left\{-\frac{1}{2}\sum_{t=1}^{T}u_t'\Sigma^{-1}u_t\right\}(2\pi)^{-\frac{n^2}{2}}\lambda_a^{\frac{n^2}{2}} exp\left\{-\frac{1}{2}\lambda_a\sum_{i,j=1}^{n}(a_{ij}-b_{ij})^2\right\}}{(2\pi)^{-\frac{n^2}{2}}|\bar{V}_A|^{-\frac{n}{2}}|\bar{U}_A|^{-\frac{n}{2}} exp\left\{-\frac{1}{2}tr\left[\bar{U}_A^{-1}(A-\bar{A}')'\bar{V}_A^{-1}(A-\bar{A}')\right]\right\}} \\ &= c(2\pi)^{-\frac{nT}{2}}|\Sigma|^{-\frac{T}{2}}|\bar{V}_A|^{\frac{n}{2}}|\bar{U}_A|^{\frac{n}{2}}\lambda_a^{\frac{n^2}{2}} exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}y_t'\Sigma^{-1}y_t + \lambda_a\sum_{i,j=1}^{n}b_{ij}^2 - tr\left[\bar{U}_A^{-1}\bar{A}\bar{V}_A^{-1}\bar{A}'\right]\right]\right\} \ . \end{aligned}$$

D.2 NVAR-R: Network-Construction Using Multiple Link-Types

Parameterizing B by $b_{ij} = w_{ij}^{b\prime}\beta_b$ with a hyperprior $\beta_b \sim N(0, \lambda_b^{-1}I)$, we get

$$\begin{aligned} p(\beta_b|A, \lambda_a, \lambda_b) &\propto p(A|\beta_b, \lambda_a)p(\beta_b|\lambda_b) \\ &\propto exp\left\{-\frac{1}{2}\lambda_a \left(vec(A) - W^b\beta_b\right)' \left(vec(A) - W^b\beta_b\right)\right\} exp\left\{-\frac{1}{2}\lambda_b\beta_b'\beta_b\right\} \\ &\propto exp\left\{-\frac{1}{2}\left[\beta_b'\left(\lambda_a W^{b\prime}W^b + \lambda_b I\right)\beta_b - 2\beta_b'\left(\lambda_a W^{b\prime}vec(A)\right)\right]\right\} , \end{aligned}$$

where W^b is s.t. $W^b\beta_b = vec(B)$. This shows that

$$\beta_b \mid (A, \lambda_a, \lambda_b) \sim N\left(\bar{\beta}_b, \bar{V}_{\beta_b}\right) , \quad \bar{V}_{\beta_b} = \left[\lambda_a W^{b\prime}W^b + \lambda_b I\right]^{-1} , \quad \bar{\beta}_b = \bar{V}_{\beta_b}\left[\lambda_a W^{b\prime}vec(A)\right] .$$

Further specifying a hyperprior $p(\lambda_b) \propto c$ for λ_b yields

$$\begin{aligned} p(\lambda_b|\beta_b) &\propto p(\beta_b|\lambda_b) \\ &\propto \lambda_b^{k_b/2} exp\left\{-\frac{1}{2}\lambda_b\beta_b'\beta_b\right\} , \end{aligned}$$

where $k_b = dim(\beta_b)$. This implies

$$\lambda_b \mid \beta_b \sim G\left(\frac{k_b}{2} + 1 \, , \, \frac{1}{2}\beta_b'\beta_b\right) .$$

There are two special cases. Under $\lambda_a = 0$, we use a Uniform prior for A, and β_b and λ_b become irrelevant to the estimation problem. As $\lambda_a \to \infty$, we effectively reparameterize A as B, which, if the elements of B are parameterized as $b_{ij} = w_{ij}^{b\prime}\beta_b$, means that the above posterior for β_b changes to

$$\beta_b \mid (Y, \Sigma, \lambda_b) \sim N\left(\bar{\beta}_b, \bar{V}_{\beta_b}\right) ,$$

with

$$\bar{V}_{\beta_b} = \left[\sum_{t=1}^{T} X_t^{b\prime}\Sigma^{-1}X_t^b + \lambda_b I_{k_b}\right]^{-1} , \quad \bar{\beta}_b = \bar{V}_{\beta_b}\left[\sum_{t=1}^{T} X_t^{b\prime}\Sigma^{-1}y_t\right] ,$$

where $X_t^b = [B_1 z_t, B_2 z_t, ..., B_{k_b} z_t]$ and $B_1, ..., B_{k_b}$ are $n \times n$ matrices containing the different link-types in W^b, i.e. $w_{ij}^b = (B_{1,ij}, ..., B_{k_b,ij})'$ and $A = B = B_1\beta_{b,1} + ... + B_{k_b}\beta_{b,k_b}$.

D.3 Factor Model Estimation

Consider the dynamic factor model

$$\begin{aligned} y_t &= \Lambda f_t + u_t \ , & u_t &\sim N(0, \Sigma_u) \ , \\ f_t &= \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t \ , & \eta_t &\sim N(0, \Sigma_\eta) \ , \end{aligned}$$

where y_t is n-dimensional and f_t is r-dimensional. The normalization of Geweke and Zhou (1996) sets $\Sigma_u = I$, $\Sigma_\eta = I$ and takes $\Lambda_{1:r,\cdot}$ to be lower-triangular with positive diagonal elements. The VAR(p) for f_t can also be written as

$$f_t' = x_t^{F\prime}\Phi + \eta_t' \ , \quad \text{or} \quad F = X^F\Phi + N \ ,$$

where $x_t^F = (f_{t-1}', ..., f_{t-p}')'$ is $rp \times 1$, $\Phi = [\Phi_1, ..., \Phi_p]'$ is $rp \times r$, and the matrices F, X^F and N stack f_t', $x_t^{F\prime}$ and η_t' along rows, respectively. The factor model permits the state space representation

$$\begin{aligned} y_t &= [\Lambda, 0, ..., 0] s_t + u_t \ , & u_t &\sim N(0, I) \ , \\ s_t &= R s_{t-1} + v_t \ , & v_t &\sim N(0, \Sigma_v) \ , \end{aligned}$$

where $s_t = (f_t', f_{t-1}', ..., f_{t-p}')'$ and $v_t = (\eta_t', 0, ..., 0)'$ are $rp \times 1$, and

$$R = \begin{bmatrix} \Phi_1, \Phi_2, ..., \Phi_p \\ I_{rp-r\times rp-r} \quad 0_{rp-r\times r} \end{bmatrix} \quad \text{and} \quad \Sigma_v = \begin{bmatrix} I \quad 0_{r\times rp-r} \\ 0_{rp-r\times rp} \end{bmatrix}$$

are $rp \times rp$.

The aim is to find the joint posterior $p(\Phi, \{\lambda^{ii}\}_{i=1}^r, \{\lambda^i\}_{i=2}^n | Y)$, where $Y = Y_{1:T} = \{y_1, ..., y_T\}$, λ^{ii} is the ith diagonal element of $\Lambda_{1:r,\cdot}$ and λ^i is the vector containing the remaining free parameters in $\Lambda_{i,\cdot}$, the ith row of Λ. It is achieved by treating $S = S_{1:T} = \{s_1, ..., s_T\}$ as parameters and obtaining first the posterior $p(\Lambda, \Phi, S|Y)$. A draw from $p(S|Y, \Lambda, \Phi)$ is obtained using the Carter and Kohn (1994) Simulation Smoother, while under Uniform priors for Φ and Λ, we can analytically derive the conditional posteriors $p(\Phi|Y, S)$, $\{p(\lambda^{ii}|Y, S, \lambda^i)\}_{i=1}^r$ and $\{p(\lambda^i|Y, S, \lambda^{ii})\}_{i=2}^n$.

Drawing from $p(\Lambda|Y, S)$ Given that $u_t \sim N(0, I)$, the measurement equation consists of a set of independent regressions

$$y_t^i = f_t^{i\prime}\lambda^i + u_{it} \ , \quad u_{it} \sim N(0, 1) \ , \quad i = 1 : n \ ,$$

where f_t^i is the vector of factors corresponding to λ^i, and

$$y_t^i = \begin{cases} y_{it} - \lambda^{ii} f_{it} & \text{for } i \leq r \\ y_{it} & \text{for } i > r \end{cases} .$$

Under $p(\lambda^i) \propto c$, we get

$$\lambda^i | Y, S, \lambda^{ii} \sim N(\hat{\lambda}^i, (F^{i\prime} F^i)^{-1}) , \quad \hat{\lambda}^i = (F^{i\prime} F^i)^{-1} (F^{i\prime} Y^i) ,$$

where F^i and Y^i stack f_t^i and y_t^i along rows, respectively. Analogously, $p(\lambda^{ii}) \propto \mathbf{1}\{\lambda^{ii} > 0\}$ yields

$$\lambda^{ii} | Y, S, \lambda^i \sim N(\hat{\lambda}^{ii}, (F^{ii\prime} F^{ii})^{-1}) , \quad \hat{\lambda}^{ii} = (F^{ii\prime} F^{ii})^{-1} (F^{ii\prime} Y^{ii}) , \quad \text{truncated to } \mathbb{R}_{++}$$

where F^{ii} and Y^{ii} stack f_{it} and $y_{it} - f_t^{i\prime} \lambda^i$ along rows, respectively.

Drawing from $p(\Phi | Y, S)$ Given that $\Sigma_\eta = I$ is diagonal, the transition equation is also a set of independent regressions,

$$f_{it} = x_t^{F\prime} \Phi_{\cdot,i} + \eta_{it} , \quad \eta_{it} \sim N(0, 1) , \quad i = 1 : r .$$

Under $p(\Phi_{\cdot,i}) \propto c$, we get

$$\Phi_{\cdot,i} | Y, S \sim N\left(\hat{\Phi}_{\cdot,i}, (X^{F\prime} X^F)^{-1}\right) , \quad \hat{\Phi}_{\cdot,i} = (X^{F\prime} X^F)^{-1} X^{F\prime} F_{\cdot,i} .$$

Drawing from $p(S | Y, \Lambda, \Phi)$ The usual formulas for the Kalman filter and Carter and Kohn (1994) simulation smoother simplify for the particular state space model above. Given an $rp \times 1$ vector x, let $[x]_1 = x_{1:r}$ contain the first r elements, $[x]_{-1}$ all but the first r elements, and $[x]_{-p}$ all but the last r elements. Similarly, given an $rp \times rp$ matrix X, let

$$X = \begin{bmatrix} [X]_{1,1} & [X]_{1,-1} \\ [X]_{-1,1} & [X]_{-1,-1} \end{bmatrix} = \begin{bmatrix} [X]_{-p,-p} & [X]_{-p,p} \\ [X]_{p,-p} & [X]_{p,p} \end{bmatrix} = \begin{bmatrix} [X]_{-p,\cdot} \\ [X]_{p,\cdot} \end{bmatrix} = \begin{bmatrix} [X]_{\cdot,-p} & [X]_{\cdot,p} \end{bmatrix} ,$$

where $[X]_{1,1}$ and $[X]_{p,p}$ are $r \times r$, $[X]_{p,\cdot}$ is $r \times (rp - p)$ and $[X]_{\cdot,p}$ is $(rp - p) \times r$.

Algorithm 4 (Kalman Filter for Factor Model)**.**

1. *Initialize* $s_{0|0} = 0$ *and* $P_{0|0} = \sum_{l=0}^{h} R^l \Sigma_v R^{l\prime}$ *for* $h = 10$*, say.*

2. *For* $t = 1 : T$*, given* $s_{t-1|t-1}$ *and* $P_{t-1|t-1}$*,*

(a: Forecast s_t:) compute $s_{t|t-1}$ and $P_{t|t-1}$ as

- $\left[s_{t|t-1}\right]_1 = R_{1,\cdot}s_{t-1|t-1}$, $\quad \left[s_{t|t-1}\right]_{-1} = \left[s_{t-1|t-1}\right]_{-p}$,
- $\left[P_{t|t-1}\right]_{11} = R_{1,\cdot}P_{t-1|t-1}R'_{1,\cdot} + \Sigma_\eta$, $\quad \left[P_{t|t-1}\right]_{1,-1} = \left[P_{t|t-1}\right]'_{-1,1}$,
  $\left[P_{t|t-1}\right]_{-1,1} = \left[P_{t-1|t-1}\right]_{-p,\cdot} R'_{1,\cdot}$, $\quad \left[P_{t|t-1}\right]_{-1,-1} = \left[P_{t-1|t-1}\right]_{-p,-p}$.

(b: Forecast y_t:) compute $y_{t|t-1}$ and $F_{t|t-1}$ as

- $y_{t|t-1} = \Lambda \left[s_{t|t-1}\right]_1$,
- $F_{t|t-1} = \Lambda \left[P_{t|t-1}\right]_{11} \Lambda' + I$.

(c: Update the forecast for s_t given observation y_t:) compute $s_{t|t}$ and $P_{t|t}$ as

- $s_{t|t} = s_{t|t-1} + \left[P_{t|t-1}\right]_{\cdot,1} \Lambda' F^{-1}_{t|t-1}(y_t - y_{t|t-1})$,
- $P_{t|t} = P_{t|t-1} - \left[P_{t|t-1}\right]_{\cdot,1} \Lambda' F^{-1}_{t|t-1} \Lambda \left[P_{t|t-1}\right]_{1,\cdot}$.

Thereby, $R_{1,\cdot} = [\Phi_1, \Phi_2, ..., \Phi_p]$, and $\Lambda \left[P_{t|t-1}\right]_{1,\cdot} = \left(\left[P_{t|t-1}\right]_{\cdot,1} \Lambda'\right)'$.

Algorithm 5 (Carter and Kohn (1994) Simulation Smoother for Factor Model)**.**

1. *Run the Kalman filter to get* $\{s_{t|t}, s_{t|t-1}, P_{t|t}, P_{t|t-1}\}_{t=1}^T$.
2. *Draw* $[s_T^m]_1$ *from* $N\left(\left[s_{T|T}\right]_1, \left[P_{T|T}\right]_{11}\right)$.
3. *For* $t = T-1, ..., 0$, *given draw* $\left[s_{t+1}^m\right]_1$ *from* $N\left(\left[s_{t+1|t+2}\right]_1, \left[P_{t+1|t+2}\right]_{11}\right)$, *draw* $[s_t^m]_1$ *from* $N\left(\left[s_{t|t+1}\right]_1, \left[P_{t|t+1}\right]_{11}\right)$ *with*
   - $s_{t|t+1} = s_{t|t} + P_{t|t}R'_{1\cdot}\left(R_{1\cdot}P_{t|t}R'_{1\cdot} + \Sigma\right)^{-1}\left(\left[s_{t+1}^m\right]_1 - \left[s_{t+1|t}\right]_1\right)$,
   - $P_{t|t+1} = P_{t|t} - P_{t|t}R'_{1\cdot}\left(R_{1\cdot}P_{t|t}R'_{1\cdot} + \Sigma\right)^{-1} R_{1\cdot}P_{t|t}$.

Analogous comments apply as for Algorithm 2.

D.4 Application Details

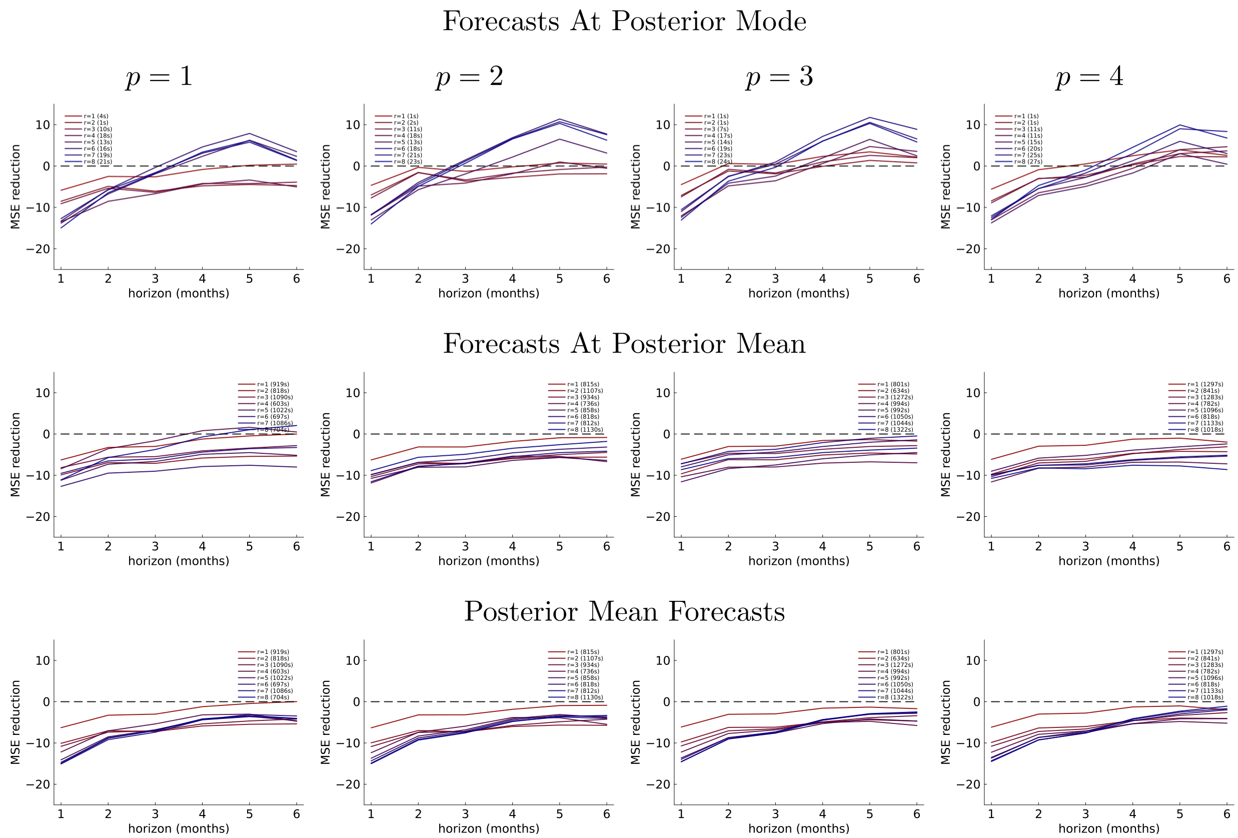

Figure A-1: Forecasting: Factor Model, Industrial Production Growth

Notes: The plot depicts the percentage difference between the out-of-sample Mean Squared Errors generated by the Dynamic Factor Model and those generated by an unconditional mean forecast for different choices of p and r and for different types of forecasts. All forecasts are obtained for industrial production growth.

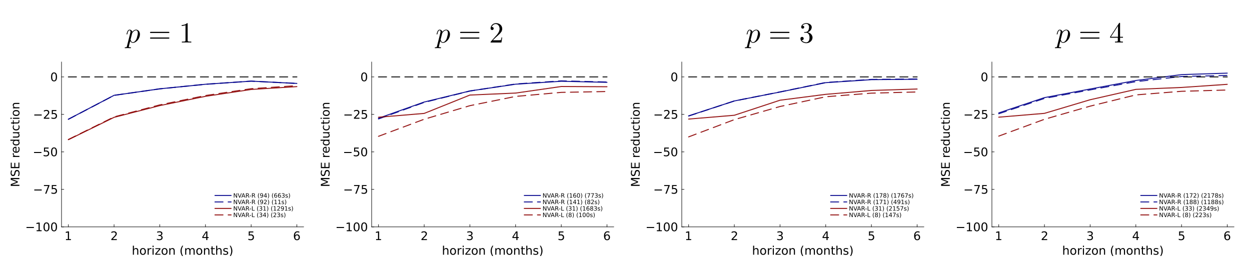

Figure A-2: Forecasting: NVAR$(p, 1)$, Industrial Production Growth

Notes: The plot depicts the percentage difference between the out-of-sample Mean Squared Errors generated by the NVAR$(p, 1)$ and those generated by an unconditional mean forecast for different choices of p, types of shrinkage and hyperparameter selection methods. All forecasts are obtained for industrial production growth using the posterior mode.