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

Marko Mlikota

Geneva Graduate Institute

This Version: 2025-03-20

[latest version: click here]

Abstract

Many environments in economics involve units linked by bilateral ties. I develop an econometric framework that rationalizes the dynamics of cross-sectional variables as the innovation transmission along fixed bilateral links and that can accommodate rich patterns of how network effects of higher order accumulate over time. The proposed Network-VAR (NVAR) can be used to estimate dynamic network effects, with the network given or inferred from dynamic cross-correlations in the data. In the latter case, it also offers a dimensionality-reduction technique for modeling high-dimensional (cross-sectional) processes, owing to networks' ability to summarize complex relations among variables (units) by relatively few bilateral links. In a first application, I show that sectoral output growth in an RBC economy with lagged input-output conversion follows an NVAR. I characterize impulse-responses to TFP shocks in this environment, and I estimate that the lagged transmission of productivity shocks along supply chains can account for a third of the persistence in aggregate output growth. The remainder is due to persistence in the aggregate TFP process, leaving a negligible role for persistence in sectoral TFP. In a second application, I forecast macroeconomic aggregates across OECD countries by assuming and estimating a network that underlies the dynamics. In line with an equivalence result I provide, this reduces out-of-sample mean squared errors relative to a dynamic factor model. The reductions range from -12% for quarterly real GDP growth to -68% for 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, Sparse 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

Numerous economic environments feature 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[1] documents that networks amplify idiosyncratic shocks and generate comovement in cross-sectional variables at a given point in time. For example, the correlation between two sectors' production is higher the closer the two sectors are in the supply chain network, and sectoral disturbances have a stronger effect on aggregates the better connected the sector is.

What is less well understood, however, is how network-induced comovements play out over time. With regard to the timing of network effects, the literature mostly considers two special cases. The first assumes that innovations transmit via bilateral links contemporaneously (e.g. Acemoglu et al. (2012); Elliott et al. (2014)), which leads to a static framework and implies that connections of all order play out simultaneously. The second case posits that network effects materialize exactly one link per period (e.g. Long and Plosser (1983); Golub and Jackson (2010)). These assumptions are useful for theoretical insights or for empirical work involving steady state comparisons – i.e. long-term effects of permanent innovations[2] –, but they are of limited use for empirical studies concerned with transition dynamics.

I build an econometric framework which relates the dynamics of cross-sectional variables to the innovation transmission along bilateral links between cross-sectional units. Assuming uni-directional transmission[3] and time-invariant links, the model can accommodate rich patterns on how innovations travel along bilateral links over time, and, consequently, how network connections of higher order accumulate as time progresses. 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. Correspondingly, stationarity can be characterized in terms of eigenvalues of the network adjacency matrix and roots of an AR process defined by the timing of innovation transmission along a single link. This timing is fundamentally related to the frequency of network interactions relative to the frequency of observation.

The Network-VAR (NVAR) is useful in two rather distinct lines of empirical 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 inferred from dynamic cross-correlations in the data, possibly aided by shrinking towards observed links. With both the network and effect-timing estimated, the NVAR is also applicable as a dimensionality-reduction technique for modeling high-dimensional processes. It assumes that dynamics are generated by innovation

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

[2] Under one period-lagged network interactions, these are equal to the effects in the static framework of contemporaneous linkages. See Section 2.2 and Appendix A.1.

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

transmission along a (small) set of bilateral links among variables. Under a given 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 and prices in a Real Business Cycle (RBC) input-output economy with time lags between the production of goods and their subsequent use as intermediaries in producing other goods. Departing from the model of Long and Plosser (1983), I generalize their one period-lagged input-output conversion (IOC) by assuming that inputs from the past p periods can be used in production and are imperfect substitutes – a stand-in for frictions like storage capacity constraints. Assuming a difference-stationary TFP process with persistent aggregate and idiosyncratic shocks, I characterize impulse-responses and the differing implications of lagged IOC and persistence in exogenous shocks as sources of persistence in aggregate output growth. Using input-output tables and monthly output growth across manufacturing sectors in the US economy, I estimate their respective contribution to business cycles.

Preliminary results suggest that lagged IOC can explain 33% of the persistence in aggregate output growth in this environment. Together with persistence in aggregate TFP, the total autocorrelation can be fully replicated without reliance on persistence in sectoral TFP processes. In addition, under lagged IOC, network positions affect not only the strength of sectoral shocks' effects on aggregates – as documented in existing literature – but also its timing, with no clear relationship between the two. How quickly a shock in a sector affects aggregate output is determined by the sector's importance as an immediate – as opposed to further upstream – supplier to relevant sectors in the economy.

With the second application I show that the NVAR – with the network itself estimated along with the timing of network effects – offers a parsimonious, yet flexible way of approximating high-dimensional (cross-sectional) processes. Parsimony is obtained because dynamics are driven by bilateral links and because variables (units) can be connected even in absence of a direct link between them. As a result, the dynamic comovement of all series can be modeled with relatively few non-zero bilateral links. This is reminiscent of the assumption that longer-term dynamics are driven by a set of shorter-term dynamics, which is upheld by the general class of VARMA(p, q) models. Flexibility is owed to the fact that the network is estimated and that the model can accommodate general patterns of how network effects of higher order accrue over time.

I derive an equivalence result, which suggests that the NVAR is preferred to a factor model if cross-sectional dynamics are composed of many micro links rather than driven by a few influential units. This corresponds to the case of numerous sparse factors with differing sets of non-zero loadings across units, or, equivalently, a sparse, yet high-rank network

adjacency matrix. This is corroborated with applications forecasting monthly industrial production growth, monthly CPI inflation and quarterly GDP growth across OECD countries. Estimating an underlying network as relevant for dynamics by shrinking links to zero, I obtain reductions in out-of-sample mean squared errors of 13%-68% relative to the ex-post best-performing dynamic factor model.

Related Literature This paper adds to the growing literature on networks in econometrics.[4] In particular, there is a large literature on spatial autoregressive (SAR) models. It is mostly concerned with identifying network effects (and effects of other covariates) in a static framework of contemporaneous dependencies (e.g. Manski (1993); Lee (2007); Bramoullé et al. (2009)). In contrast, I analyze lagged network effects. Among other studies that do so (e.g. Knight et al. (2016); Yang and Lee (2019)), Zhu et al. (2017) cast them in an explicit time series model, generalizing the lag length and discussing stationarity and large T-inference. I characterize dynamics in their model as the cross-sectional innovation transmission along network connections of different order. In turn, I further generalize this mapping between network connections and dynamics, which allows me to accommodate richer patterns of how transmission materializes over time. Moreover, I discuss how to conduct inference on the network itself – not only the effect-timing –, which allows me to illustrate the model's merits as a dimensionality-reduction technique – not only as a tool for estimating dynamic network effects.

This study bridges the gap between static SAR models and dynamic VAR models. The NVAR maps networks into VARs by restricting VAR-innovations to transmit via bilateral links among variables (cross-sectional units). This leads to rich patterns of multi-step causality, connecting my work to Dufour and Renault (1998). It contrasts with Diebold and Yilmaz (2009, 2014), who map variance decompositions of VARs into networks with the goal of understanding dynamic connectedness.[5] Relative to other studies restricting time series models using networks (e.g. Pesaran et al. (2004); Barigozzi et al. (2022); **?**); Caporin et al. (2023)),[6] I focus on a single variable per cross-sectional unit and a single type of connection among units, and I entertain the assumption of innovation transmission along bilateral links. This leads to a clear relation between the model's time series properties on the one hand and the network and timing of network effects on the other. It also generates intuitive, analytical expressions for the estimators and allows me to examine the relation to factor models.[7]

With the first application, I address the macroeconomic literature on production networks. It mostly assumes that firms convert inputs into outputs contemporaneously. In this

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

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

[6]Barigozzi et al. (2022) extend the model of Zhu et al. (2017) along a different dimension than I do, allowing for several connection-types to influence dynamics.

[7]Approaches for achieving shock identification using networks are discussed in Hipp (2020) and Dahlhaus et al. (2021). Bykhovskaya (2021) builds a time series model for the evolution of the network itself.

environment, firm- or sector-level shocks transmit contemporaneously along supply chains, generating cross-sectional comovement and amplifying an individual sector's effect on aggregates at a given point in time (Horvath, 2000; Foerster et al., 2011; Acemoglu et al., 2012; Bouakez et al., 2014; Giovanni et al., 2018; Giroud and Mueller, 2019). A notable exception is Long and Plosser (1983), who study an RBC economy with a one period-lag in IOC, which renders the transmission dynamic and generates endogenous business cycles.[8] Carvalho and Reischer (2021) characterize persistence as a function of network properties in this environment.[9] Building on their work, I estimate to what extent business cycless in this model can be rationalized by lagged IOC as opposed to persistence in exogenous productivity shocks. This is analogous to the decomposition of aggregate variance under contemporaneous IOC in Foerster et al. (2011).

With the second application of the NVAR, I address the literature on dimensionality-reduction techniques for time series modeling. By reducing the number of parameters and using shrinkage priors (regularization), the NVAR leverages both ways of addressing the large parameter problem in the Wold representation (see Geweke (1984)). Compared to shrinkage methods like Lasso (Tibshirani, 1996) or Ridge regression,[10] it applies shrinkage to links – which in turn summarize the information in predictors at all lags – rather than to predictors themselves, leading to additional parsimony.[11] Compared to reduced rank regression and factor models (Velu et al., 1986; Stock and Watson, 2002), it finds the linear combination that effectively summarizes the information in lagged values by relying on bilateral links among variables (cross-sectional units). I show analytically that, with the cost of restricting factor dynamics, the NVAR naturally incorporates sparse factors as locally important nodes in the network. This suggests that it can improve upon the poor forecasting performance of factor models in settings with a high dispersion in factor loadings across series (see e.g. Boivin and Ng (2006)),[12] as confirmed by the applications I consider.

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

[8] Under contemporaneous interactions, network effects are static because network effects of all order play out simultaneously. Networks can only amplify existing dynamics – obtained thanks to agents' intertemporal optimization problems in a structural model or due to persistence in shocks – but not drive dynamics themselves. See discussion in Section 2.2.

[9] Based on the relationship they derive, the evolution of the supply chain network of the US economy can account for changes in persistence of aggregate industrial production. They also show that this calibrated model improves upon ARMA models in forecasting aggregate industrial production growth.

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

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

[12] This notably includes the case of sparse factors and units differing in the set of factors they load on. See Onatski (2012) and Freyaldenhoven (2022) for analyses of sparse factors.

2 Lagged Network Effects & Cross-Sectional Dynamics

After providing some basic background on networks in Section 2.1, I present the NVAR in Section 2.2 building on a simple example. Details and proofs are in Appendix A. Estimation is deferred to Section 3.

2.1 Bilateral Connections in Networks

A network is represented by an $n \times n$ adjacency matrix A with elements a_{ij}. I consider a directed, weighted and possibly signed network, which means that $a_{ij} \in [-1, 1]$ shows the sign and strength of the link from cross-sectional unit i to unit j, with $a_{ij} \neq a_{ji}$ possibly. If $a_{ij} = 0$, I say unit i is not connected to unit j. Self-links $a_{ii} \neq 0$ are permitted. The set of bilateral links $\{a_{ij}\}_{i,j=1:n}$ give rise to a plethora of higher-order connections among units, referred to as walks.[13]

Definition 1 (Walk)**.** *A walk from i to j of length K is the product of a sequence of K links $a_{i_k,i_{k+1}}$ between units $i_1, i_2, ..., i_{K-1}$ such that $a_{i_k,i_{k+1}} \neq 0$ for $k = 1 : K-2$, with $i_1 = i$, $i_{K-1} = j$:*

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

Put simply, a walk is the product of bilateral links a_{ij} that lead from unit i to unit j over some intermediary units, all of which are sequentially connected. Just as element (i, j) in the matrix A shows the walk from i to j of length one (direct link), simple matrix algebra reveals that $(A^K)_{ij}$ contains the sum of walks from i to j of length K.[14] I refer to this quantity as the Kth-order connection from i to j.

Consider the following example:

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

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

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

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

2.2 Lagged Innovation Transmission via Bilateral Links

The core assumption that underlies the proposed NVAR is that innovations u_{it} to a process y_{it} transmit cross-sectionally only via bilateral links. I assume that links are fixed and this transmission operates only in one direction through the network; the direct link from i to j, a_{ij}, transmits innovations from j to i. Innovations can be cross-sectionally correlated. To simplify the exposition, I assume $\mathbb{E}[y_{it}] = 0 \; \forall \; i, t$.

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

Consider the following VAR(1) for the cross-sectional time series $y_t = (y_{1t}, ..., y_{nt})'$:

$$y_t = \Phi y_{t-1} + u_t \; , \quad u_t \sim WN(0, \Sigma) \; , \qquad \text{with} \quad \Phi = \alpha A \; , \quad \alpha \in \mathbb{R} \; . \tag{1}$$

If Φ is proportional to an adjacency matrix A containing cross-sectional linkages, one obtains a process that relates the dynamics of the cross-sectional time series y_t to the bilateral links among cross-sectional units.[15] Following Long and Plosser (1983), in Section 4 I derive such a process for sectoral output and prices in an RBC production economy with a one period delay in converting inputs into output. Golub and Jackson (2010) posit it to study societal opinion formation through friendship ties.

Under this process, the one period-ahead expectation of y_{it} is proportional to a weighted sum of one period-lagged values of y_{jt} for all units j to which i is directly linked, with weights given by the strength of direct links a_{ij}: $\mathbb{E}_{t-1}[y_{it}] = \alpha \sum_{j=1}^{n} a_{ij} y_{j,t-1}$. For example, in Long and Plosser (1983), the expected output tomorrow in sector i is a weighted average of the production in its supplier-sectors j today. Dynamics of y_t, as summarized by Granger-causality at horizons $h = 1, 2, ...$, are shaped by hth order network-connections:

$$GC_{ij}^h \equiv \frac{\partial y_{i,t+h}}{\partial y_{jt}} = \frac{\partial y_{i,t+h}}{\partial u_{jt}} = [\alpha^h A^h]_{ij} \; .$$

As a result, given all other variables $y_k \; , k \neq j$, y_j is useful in forecasting y_i at horizon h iff there is an hth order connection from i to j. GC_{ij}^h is also referred to as the Generalized Impulse Response Function (GIRF).[16]

Fig. 1 provides an example. It depicts the GIRFs for the process in Eq. (1) and the network from Section 2.1 for $\alpha = 1$. Each panel (i, j) shows $GC_{ij}^h = [A^h]_{ij}$ for $h = 0 : 6$. By definition, the contemporaneous responses to all but a series' own innovation are zero. From horizon $h = 1$ onwards, GC_{ij}^h equals the network connections from i to j of order h.

[15] As explained in Section 2.2.3, I call this process NVAR(1,1).

[16] It is generalized because it is not concerned with shock identification, but the derivative is taken with respect to potentially correlated reduced form errors in u_t.

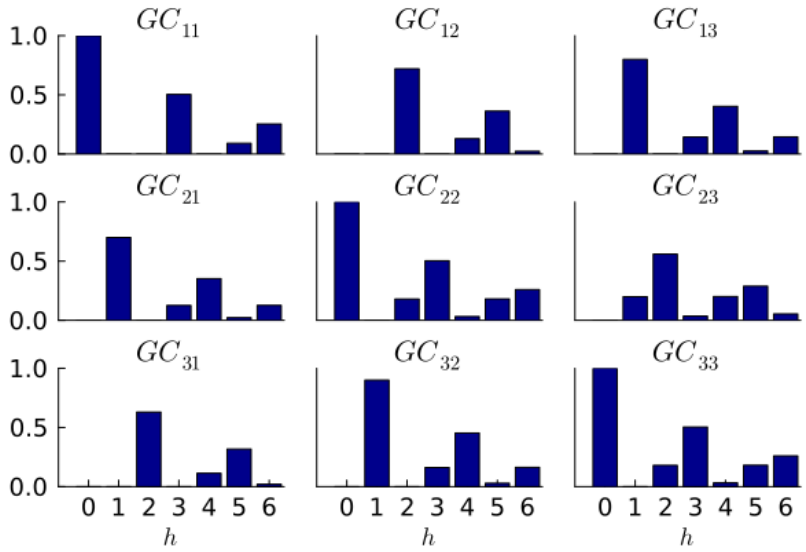


Figure 1: Example Generalized Impulse Responses: NVAR$(1, 1)$

Notes: Panel (i, j) shows $GC^h_{ij} = [A^h]_{ij}$. Note that $GC^0_{ij} = 1$ if $i = j$ and zero otherwise.

Two points are worth highlighting. First, lagged network interactions can be a source of persistence, at the individual as well as aggregate level.[17] This is behind the endogenous business cycles in Long and Plosser (1983) and contrasts with dynamics under contemporaneous network interactions. Nevertheless, dynamics under the two environments are related, as discussed in Section 2.2.4. Second, under this source of persistence, network-connections between i and j shape not only the strength of the impulse-response, but also its timing. For example, while unit 2 is directly linked to unit 1 and therefore experiences the latter's innovation with a lag of one period, unit 3 only has an indirect, second-order connection to unit 1 and is therefore impacted by its innovation only after two periods.

These results relate to the discussion in Dufour and Renault (1998), who point out that Granger-causality can take the form of chains. 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. They examine conditions under which noncausality at a given horizon implies noncausality at higher horizons. If innovations transmit only via bilateral links, these generally non-trivial conditions boil down to the existence of network connections of relevant order between the concerned variables (cross-sectional units).

While useful for theoretical insights, the process in Eq. (1) is of limited use empirically as it entertains a very restrictive mapping between network connections and observed dynamics in y_t. First, it assumes that innovations travel through the network at the speed of one link

[17] Even without autocorrelation in disturbances u_{jt}, the responses of $\{y_{it}\}_{i=1}^n$ are persistent because disturbances u_{jt} reach connected units $i \neq j$ only with a lag. Even without self-links – $a_{jj} = 0$ –, the response of unit j itself is persistent as well so long as there are cycles in the network, i.e. walks from j to j.

per period. This is restrictive, as the frequency of observations might not coincide with the frequency of network interactions. For instance, it implies that a sector i's output does not react to innovations at suppliers situated two positions upstream of i ("suppliers of suppliers") but after two (observational) periods. Second, it assumes complete transmission at a single lag. For example, after an innovation at a supplier j, a sector i fully adjusts its production after one period. Subsequent adjustments only occur if j is also a supplier to other suppliers of i, whose adjustments to j induce further adjustments of i (or if the innovation is persistent; see Section 4.1). In the following, I extend the simple process above along both of these two dimensions.

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

Let the cross-sectional time series $\tilde{y}_\tau$ evolve according to

$$\tilde{y}_\tau = \alpha_1 A\tilde{y}_{\tau-1} + ... + \alpha_p A\tilde{y}_{\tau-p} + \tilde{u}_\tau \ , \quad \tilde{u}_\tau \sim WN(0, \Sigma) \ , \tag{2}$$

with $\alpha = (\alpha_1, ..., \alpha_p)' \in \mathbb{R}^p$. This is a VAR($p$) where each autoregressive matrix is proportional to the same network adjacency matrix A. It is a particular version of the network-autoregressive process in Zhu et al. (2017). I dub this process NVAR($p, 1$). Compared to Eq. (1), setting $p > 1$ allows connections of order lower than h to affect dynamics at horizon h.

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

The proof in Appendix A.1 establishes that the GIRF is of the form

$$\frac{\partial \tilde{y}_{i,\tau+h}}{\partial \tilde{u}_{j,\tau}} = c^h_{\underline{k}}(\alpha) \left[A^{\underline{k}}\right]_{ij} + ... + c^h_h(\alpha) \left[A^h\right]_{ij} \ . \tag{3}$$

The coefficients $\{c^h_k(\alpha)\}_{k=\underline{k}:h}$ are polynomials of $\{\alpha_l\}_{l=1:p}$ and show the importance of different connection-orders k for the impulse response at a given horizon h. For a pair (i, j) with only a direct link – $a_{ij} \neq 0$ and $(A^k)_{ij} = 0 \ \forall \ k > 1$ –, this boils down to $\frac{\partial \tilde{y}_{i,\tau+h}}{\partial \tilde{u}_{j,\tau}} = \alpha_h a_{ij}$ for $h = 1 : p$ and zero otherwise. Hence, Eq. (2) specifies that $\tilde{y}_\tau$ is driven by lagged network interactions, whereby innovation transmission along a bilateral link takes p periods to fully materialize. The parameters $\{\alpha_l\}_{l=1:p}$ determine how this transmission materializes over time and, consequently, how transmission along higher-order connections accumulates

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

as time progresses.[19] Note that the transmission is assumed to be the same for all unit pairs (i, j) and invariant over time.

The process $\tilde{y}_\tau$ evolves at frequency τ, which I shall call the network interaction frequency.[20] It might not coincide with the frequency of observation. In particular, if data is observed at a lower frequency than network interactions occur – as is likely the case for macroeconomic series –, then dynamics at horizon h can be driven by connections of order higher than h, as several rounds of transmission can happen in one period of observation. In addition, this leads to network-induced cross-sectional correlation in observed innovations even in absence of correlation in $\tilde{u}_\tau$. In the following, I formalize this idea. Let the observed data be $\{y_t\}_{t=1:T}$ and assume $\tilde{u}_\tau \sim N(0, \Sigma)$.

2.2.3 Time-Aggregation of Lagged Innovation Transmission: NVAR(p, q)

Stock Variables Suppose y_t is a stock variable and evolves as

$$\begin{aligned} \tilde{y}_\tau &= \alpha_1 A\tilde{y}_{\tau-1} + ... + \alpha_p A\tilde{y}_{\tau-p} + \tilde{u}_\tau \ , \quad \tau = 1 : T_\tau \ , \text{[21]} \\ y_{\tau/q} &= \tilde{y}_\tau \quad \text{if } \tau/q \in \mathbb{N} \ , \end{aligned} \tag{4}$$

for some $q \in \mathbb{Q}_{++}$ and T_τ s.t. $T = |\{(1 : T_\tau)/q\} \cap \mathbb{N}|$.[22] I dub this process NVAR(p, q).

Thereby, q indicates the frequency of network interactions relative to the frequency of observation. If $q = 1$, the two coincide, and, trivially, y_t and $\tilde{y}_\tau$ are the same NVAR$(p, 1)$ process. If $q \in \mathbb{N}\backslash\{1\}$, network interaction frequency is an integer-multiple of observational frequency, which means we observe a snapshot of $\tilde{y}_\tau$ every q periods: $\{y_t\}_{t=1:T} = \{\tilde{y}_{tq}\}_{t=1:T}$. For example, $q = 4$ under monthly observations implies weekly network interactions.

For any NVAR(p, q) with $q \in \mathbb{Q}_{++}$, we can find an observationally equivalent (restricted) NVAR(p^*, q) with $q = 1$ or $q \in \mathbb{N}\backslash\{1\}$. First, consider $q^{-1} \in \mathbb{N}\backslash\{1\}$, i.e. observational frequency is an integer-multiple of the network interaction frequency. For example, $q = 1/3$ under monthly observations implies quarterly network interactions. In this case, all $\tilde{y}_\tau$ are

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

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

[22] The number of elements in the set $1 : T_\tau$ that are integer-multiples of q shall equal 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.

observed. Under Normality of $\tilde{u}_\tau$, y_t follows an NVAR$(p^*, 1)$, with $p^* = p/q \in \mathbb{N}$:

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

In the previous example, the observed, monthly series depends on its value three months ago, six months ago, etc., up to $3p$ months ago.

For any other $q \in \mathbb{Q}_{++}$, we can write $q = q_1 q_2$ with $q_1^{-1} \in \mathbb{N}$ and $q_2 \in \mathbb{N}$.[24] Consequently, we observe every q_2 periods a snapshot of an NVAR$(p/q_1, 1)$ process: $\{y_t\}_{t=1:T} = \{\tilde{y}^*_{tq_2}\}_{t=1:T}$, where $\tilde{y}^*_{\tau^*} \sim$ NVAR$(p/q_1, 1)$. For example, under monthly observations, $q = 4/3$ implies that network interactions occur at tri-weekly frequency, and we observe every fourth period a snapshot of a weekly process that depends on its value three weeks ago, six weeks ago, etc. Analogously, network interactions occur every five weeks under $q = 4/5$ and every four days under $q = 30/4$.

The time-aggregation of lagged transmission patterns generalizes the mapping from network connectedness to observed dynamics. Intuitively, in an NVAR(p, q), q rounds of transmission can occur in one period of observation. As a result, connection-orders higher than h can matter for transmssion at horizon h. Furthermore, observed innovations $u_t = y_t - \mathbb{E}_{t-1}[y_t]$ can be cross-sectionally correlated even if the innovations at network interaction frequency, $\tilde{u}_\tau$, are not.[25]

Proposition 2 (Granger-Causality in NVAR(p, q))**.**
Let the stock variable y_t follow an NVAR(p, q) for $q \in \mathbb{N}\backslash\{1\}$ with $\alpha_l \neq 0$ for $l = 1 : p$. Then $\tilde{y}_j$ Granger-causes $\tilde{y}_i$ at horizon h iff there exists a connection from i to j of at least one order $k \in \{\underline{k}, \underline{k}+1, ..., hq\}$, where $\underline{k} = ceil(hq/p)$.

Flow Variables If y_t is a flow variable, we can write $\{y_t\}_{t=1}^T = \{\tilde{y}_{tq} + ... + \tilde{y}_{tq-q+1}\}_{t=1}^T$ provided that $q \in \mathbb{N}$, i.e. the network interaction frequency either coincides with the observational frequency ($q = 1$) or is an integer-multiple thereof.[26] Dynamics of y_t are represented by the state space system

$$\tilde{y}_\tau = \alpha_1 A \tilde{y}_{\tau-1} + ... + \alpha_p A \tilde{y}_{\tau-p} + \tilde{u}_\tau \ , \quad \tau = 1 : T_\tau \ , \tag{5}$$

$$y_{\tau/q} = \tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1} \quad \text{if } \tau/q \in \mathbb{N} \ .$$

[23] Under distributional equivalence of u_t and $\tilde{u}_\tau$, $\mathbb{E}[y_t y_{t-h}] = \mathbb{E}[\tilde{y}_\tau \tilde{y}_{\tau-hq}] \ \forall \ h$, and under Normality, the first two moments fully characterize the process.

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

[25] This correlation is affected by connections of order $k \leq q-1$. Intuitively, this is because u_t is composed of $\{\tilde{y}_{tq}, ..., \tilde{y}_{tq-q+1}\}$. As a result, $\text{Cov}(u_{it}, u_{jt})$ is affected by partial derivatives $\partial \tilde{y}_{k,\tau+h} / \partial \tilde{y}_{o,\tau}$ for $h = 1 : (q-1)$ and all units $k, o = 1 : n$. A more formal treatment is given in Appendix A.2.

[26] Analogous calculations apply if $y_t = (\tilde{y}_\tau + ... + \tilde{y}_{\tau-q+1})/q$.

As shown in Appendix A, analogous conclusions about the dynamics of y_t can be drawn as in the case of stock variables. For $q \notin \mathbb{N}$, no state space representation can be found for y_t (without assumptions that allow us to convert a flow variable from lower to higher frequency).

2.2.4 Relation to Contemporaneous Innovation Transmission

The econometric literature on Spatial Autoregressive (SAR) models and the macroeconomic literature on production networks typically assume contemporaneous network interactions. In that case, the implicit assumption is that connections of all order materialize in any given period of observation:

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

The cross-sectional comovement among $\{x_i\}_{i=1}^n$ is rationalized by the contemporaneous transmission of cross-sectionally uncorrelated, idiosyncratic shocks $\{\varepsilon_i\}_{i=1}^n$ along bilateral links.

Contemporaneous interactions are silent on how networks drive dynamics.[27] Nevertheless, models with contemporaneous and lagged network interactions are related: the (contemporaneous) response of x_i to a (transitory or persistent) innovation to x_j under contemporaneous interactions is equal to the long-run response of y_{it} to a persistent innovation to y_{jt} under lagged interactions, provided that y_t is stationary.

Proposition 3 (NVAR$(p, 1)$: Long-Term Response to White Noise-Innovations)**.** *Let $y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t$ and $x = aAx + \varepsilon$ with $a = \sum_{l=1}^p \alpha_l$, and assume y_t is weakly stationary. Then*

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

Both responses are given by element (i, j) of the Leontief inverse $(I - aA)^{-1}$, which is a sufficient statistic for the long-term (or static) cross-sectional comovement. By taking a stance on the time profile of network interactions, y_t shows how any such long-term effect materializes over time, i.e. it goes beyond steady state comparisons and characterizes transition dynamics. This result carries over to an NVAR(p, q), $q \in \mathbb{N}\backslash\{1\}$ (see Appendix A.1).[28] Note that there is a one-to-one mapping between the timing of responses to permanent

[27] At least in absence of further structure, such as provided by a dynamic macroeconomic model with intertemporally linked optimization problems of agents who are impacted by disturbances to x (see e.g. Carvalho (2010, ch. 4)). Even in that case, contemporaneous network interactions can only amplify dynamics but not cause them. See Section 4.1 for a more detailed discussion.

[28] i.e. the long-term response of y_t to a permanent increase in the underlying high-frequency innovation $\tilde{u}_\tau$ equals $(I - \alpha A)^{-1}$.

and temporary shocks, with slowly accumulating long-term responses to permanent shocks implying persistent responses to temporary shocks.[29]

Stationarity of the NVAR$(p, 1)$ can be characterized in terms of eigenvalues of the network adjacency matrix A and the time profile of network effects α, which drastically simplifies checking for stationarity under high dimensions.[30] Expectedly, stationarity of an NVAR(p, q) is implied by stationarity of the underlying NVAR$(p, 1)$ (see Proposition 10).

Proposition 4 (Stationarity of NVAR$(p, 1)$)**.**
Let $y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t$ with $u_t \sim WN$, assume $\alpha_l \neq 0$ for at least one l, and define $a = \sum_{l=1}^{p} |\alpha_l|$.

Then, y_t is weakly stationary (WS) iff for all eigenvalues λ_i of A, the matrix

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

has all eigenvalues inside the unit circle. For real λ_i, this is equivalent to WS of the univariate AR(p) process $x_t = \lambda_i\alpha_1 x_{t-1} + ... + \lambda_i\alpha_p x_{t-p} + v_t$.

Moreover, a sufficient condition for WS of y_t is is that $|\lambda_i| < 1/a$ for all eigenvalues λ_i of A. Under $\alpha_l \geq 0 \; \forall \; l$, this condition is both necessary and sufficient.

3 Inference

The NVAR can be used to estimate dynamic network effects or to model high-dimensional (cross-sectional) processes more generally, say for forecasting purposes. In Section 3.1, I discuss the estimation of the time profile of network effects α conditional on the network A. In Section 3.2, I discuss joint inference on (α, A).

3.1 Timing of Network Effects

The first part of this section is devoted to the estimation of α in the NVAR$(p, 1)$. The second part deals with the case when data is observed at a higher frequency than network interactions take place: NVAR(p, q) with $q \in \mathbb{N}\backslash\{1\}$.[31] Details are in Appendix B.1.

[29]For any VAR, the long-term response to a permanent shock is equal to the cumulative responses to temporary shocks. Therefore, the (fraction of the) long-term response which materialized until horizon h is equal to the area under the IRF to a temporary shock until horizon h (as a fraction of the total area).

[30]Proposition 4 is a combination of Corollary 1 and Proposition 9 in Appendix A.3. Intuitively, the conditions in Proposition 4 ensure that $\lim_{k\to\infty} c(\alpha, k)A^k = 0$, for any polynomial in α of order k, $c(\alpha, k)$. This in turn ensures that long-term effects of shocks go to zero as the horizon increases (see Proposition 1).

[31]Recall the discussion in Section 2.2.3: flow variables can only accommodate $q = 1$ or $q \in \mathbb{N}\backslash\{1\}$, whereas any NVAR$(p, q)$ with $q \in \mathbb{Q}_{++}\backslash\{1\}$ for stock variables can be written as an NVAR(p^*, q) with $q \in \mathbb{N}\backslash\{1\}$.

NVAR$(p,1)$ The NVAR$(p,1)$ from Eq. (2),[32]

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

can be written as a linear regression:

$$y_t = X_t \alpha + u_t \ , \quad \text{or} \quad y_{it} = x_{it}' \alpha + u_{it} \ , \tag{6}$$

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

$$X_t = \begin{bmatrix} x_{1t}' \\ ... \\ x_{nt}' \end{bmatrix} = \left[A y_{t-1}, A y_{t-2}, ..., A y_{t-p} \right] .^{33} \tag{7}$$

Defining $\Sigma = \mathbb{V}[u_t]$, we obtain the following Least Squares (LS) estimator for α:

$$\hat{\alpha}|\Sigma = \left[\sum_{t=1}^{T} X_t' \Sigma^{-1} X_t \right]^{-1} \left[\sum_{t=1}^{T} X_t' \Sigma^{-1} y_t \right] \ . \tag{8}$$

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

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

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

The joint ML estimator $(\hat{\alpha}, \hat{\Sigma})$ is obtained by iterating on the above expression for $\hat{\alpha}|\Sigma$ and $\hat{\Sigma}|\alpha = \frac{1}{T} \sum_{t=1}^{T} u_t u_t'$ until convergence (see Meng and Rubin (1993)). This is also the posterior mode under a Uniform prior for (α, Σ).

[32] Because $q = 1$, $\tilde{y}_\tau$ and $\tilde{u}_\tau$ are written as y_t and u_t.

[33] The dependence of X_t on A is suppressed because A is taken as given.

NVAR(p,q), $q \in \mathbb{N}\backslash\{1\}$ Consider the NVAR(p,q) for a stock variable y_t from Eq. (4):

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

The following discussion holds likewise for flow variables.

An estimator for α can be obtained by data augmentation. However, point identification is not guaranteed. For example, under $q = 2$ and $p = 1$, the observed process follows

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

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).[34] 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 not bijective.[35] For a general AR(p) and $q = 2$, the vector $(\alpha_1, \alpha_3, ...)$ is identified (jointly) up to sign (see Appendix 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 $p(\alpha, \tilde{Y}_{1:T_\tau}|Y_{1:T})$.[36] 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 Estimation: Network & Effect-Timing

Whether one is interested in estimating dynamic network effects or approximating dynamics of cross-sectional processes, in many cases network data might be missing or it appears

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

[35] As discussed in Appendix A.2, the observed process can be approximated arbitrarily well as a VARMA with coefficient-matrices equal to polynomials of $\{\alpha_l A\}_{l=1}^p$.

[36] 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 Appendix B.1, and see Schorfheide and Song (2015) for a discussion of the analogous case of a mixed-frequency VAR.

restrictive to condition on a single network. This section discusses joint estimation of (α, A). Again the first part deals with an NVAR$(p,1)$, while the second part discusses the case of an NVAR(p,q) with $q \in \mathbb{N}\backslash\{1\}$. Details are in Appendix B.2.

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

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

can also be written as the linear regression

$$y_t = A z_t + u_t \ , \quad \text{or} \ \ Y = ZA' + U \ , \tag{9}$$

whereby $z_t = \sum_{l=1}^{p} \alpha_l y_{t-l} = \big[y_{t-1}, y_{t-2}, ..., y_{t-p}\big]\alpha$, 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 $\alpha_1 > 0$ in the true data generating process.

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:

$$\bar{A}' = \arg \underset{A}{min} \ \sum_{t=1}^{T} (y_t - A z_t)' \, \Sigma^{-1} \, (y_t - A z_t) + \lambda_a \sum_{i,j=1}^{n} (a_{ij} - b_{ij})^2 \ . \tag{10}$$

As $\lambda_a \to \infty$, we impose $A = B$. As $\lambda_a \to 0$, A is inferred from the data alone, ignoring B. 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.[37]

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

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

[37] 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.

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.[38]

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. (10), 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 Appendix 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), $q \in \mathbb{N}\backslash\{1\}$ 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 Input-Output Links & Sectoral Output Dynamics

Long and Plosser (1983) show that firms' lagged conversion of inputs into outputs can generate endogenous business cycles. This contrasts with the models used in most studies, where business cycles are obtained as a result of persistence in exogenous shocks. Yet, Carvalho and Reischer (2021) show that the persistence obtained in the framework of Long and Plosser (1983) matches well empirical measures of persistence, suggesting that the endogenization of business cycles through lagged input-output conversion (henceforth IOC) is not only theoretically attractive, but has empirical merit as well.

[38] 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.

This application provides a further empirical test of Long and Plosser (1983)'s hypothesis. Departing from their model, I generalize the lags in IOC by assuming that firms' production requires inputs produced in the past p periods. This leads to sectoral output growth and inflation evolving at some model-frequency as an NVAR$(p, 1)$, which translates into an NVAR(p, q) for some $q \in \mathbb{N}$ at the 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 application-setup, data and estimation procedure in Section 4.2. Section 4.3 discusses the preliminary results, which focus on sectoral prices and assume i.i.d. productivity shocks in order to shed light on the relation between lagged IOC and persistence in isolation. Details are in Appendix C.

4.1 Business Cycles Through Lagged Input-Output Conversion

The following analysis is based on Carvalho and Tahbaz-Salehi (2019), who discuss a static economy. Derivations are provided in Appendix C.1.

Assume there are n sectors, in each of which a representative firm produces a differentiated good i by combining labor services $l_{i\tau}$ 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 $c_{i\tau} = \gamma_i \frac{w_\tau}{p_{i\tau}}$.[39] This result holds even if households have access to a storage technology, as market clearing under represen-

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

tative 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. $x_{ij\tau} = x^{ij}_\tau$. 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

$$\tilde{y} = k^y + A\tilde{y} + \tilde{z} \ , \tag{11}$$

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 $\tilde{y}$ (see e.g. Acemoglu et al. (2012)), any autocorrelation in $\tilde{y}$ is inherited from that in $\tilde{z}$.

To analyze dynamics under lagged IOC, I assume perfect foresight.[40] If, as in Long and Plosser (1983), it takes one period to convert purchased inputs into output, then $x_{ij\tau} = x^{ij}_{\tau-1}$ 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.[41] 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

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

[41] 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}).

time τ.[42] This shortcut stands for frictions like storage capacity constraints and time-varying input delivery times.[43] In the Cobb-Douglas case $r \to 0$, sectoral output approximately follows an NVAR(2,1):

$$\tilde{y}_\tau = k_\tau^{y2} + \alpha_1 A\tilde{y}_{\tau-1} + \alpha_2 A\tilde{y}_{\tau-2} + \tilde{z}_\tau \ , \tag{12}$$

where once again k_τ^{y2} varies over time only to the extent that the numéraire w_τ changes in value. Under a more general elasticity of substitution $r \in [0, 1)$,[44] the analogous result is obtained by log-linearizing around the steady state (see Appendix 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 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.[45]

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_τ^a and $e_{i\tau}$ evolving as AR(1) processes:

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

$(1 - \rho_a L)e_\tau^a = \varepsilon_\tau^a \sim WN$ and $(1 - \rho_i L)e_{i\tau} = \varepsilon_{i\tau} \sim WN$. Under contemporaneous IOC (Eq. (11)), this leads to

$$\Delta\tilde{y}_\tau = (I - A)^{-1} \left[\gamma + \lambda e_\tau^a + e_\tau\right] \ . \tag{13}$$

This equation points to two sources of persistence in output growth: persistence in the aggregate TFP process e_τ^a and persistence in idiosyncratic TFP processes $\{e_{i\tau}\}_{i=1}^n$. We

[42] 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 $\tau + 2$: $x_\tau^{ij} = x_{\tau+1,\tau}^{ij} + x_{\tau+2,\tau}^{ij}$.

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

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

[45] Berman and Plemmons (1979, p. 37) show that for an element-wise nonnegative matrix with row sums strictly smaller than 1, the absolute value of the largest eigenvalue is strictly less than 1. Stationarity then follows by Proposition 4.

obtain the following IRFs:

$$\frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial \varepsilon_\tau^j} = \left[(I-A)^{-1}\right]_{ij} \rho_j^h \,, \quad \frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial \varepsilon_\tau^a} = \left[(I-A)^{-1}\right]_{i\cdot} \lambda \rho_a^h \,.$$

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 $(I-A)^{-1}$. By summing up connections of all order from i to j, it shows the importance of sector j in sector i's supply chain. In turn, the response of aggregate output growth, $\sum_{i=1}^n v_i \Delta \tilde{y}_{i,\tau+h}$, depends on the weighted average of sector j's importance in the supply chains of all other sectors i, where the weights $\{v_i\}_{i=1}^n$ indicate sectors' shares of aggregate output.

Lagged IOC can be an additional source of persistence. Under Eq. (12), we get

$$\Delta \tilde{y}_\tau \approx \gamma + \alpha_1 A \Delta \tilde{y}_{\tau-1} + \alpha_2 A \Delta \tilde{y}_{\tau-2} + \lambda e_\tau^a + e_\tau \,. \tag{14}$$

and

$$\begin{aligned} \frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial \varepsilon_\tau^j} &= \sum_{l=0}^{h} \frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial e_{\tau+h-l}^j} \frac{\partial e_{\tau+h-l}^j}{\partial \varepsilon_\tau^j} \\ &= \sum_{l=0}^{h} \left\{ c_{\underline{k}(l)}^l(\alpha) \left[A^{\underline{k}(l)}\right]_{ij} + ... + c_l^l(\alpha) \left[A^l\right]_{ij} \right\} \rho_j^{h-l} \,, \end{aligned} \tag{15}$$

where $\underline{k}(l) = ceil(l/p)$ with $p = 2$ in this case, as follows from Proposition 1 and Eq. (3). Under $\rho_j = 0$, the expression simplifies to that in Eq. (3). Under $\alpha_1 = 1$, it simplifies to $\partial \Delta \tilde{y}_{i,\tau+h} / \partial \varepsilon_\tau^j = \sum_{l=0}^h \left[A^h\right]_{ij} \rho_j^{h-l}$.

The differing implications for persistence of ρ_j and $\alpha = (\alpha_1, \alpha_2)'$ are illustrated best by considering the IRF of two sectors $i \neq j$ that share only a first- or only a second-order connection, respectively. This is done in Fig. 2, for $\rho_j = .6$ and once for $\alpha = (.8, .2)'$ (top row), once for $\alpha = (.2, .8)'$ (bottom row). The left plots illustrate the effect of α on the timing of transmission of a single shock along network connection-orders. Under $\alpha = (.8, .2)'$, the shock travels faster through the network and, for example, induces the largest response of second-order connections after two and three periods, whereas it occurs only after three and four periods under $\alpha = (.2, .8)'$. In line with that, the middle panel shows that α determines how the initial shock's transmission along a direct link is spread out over horizons 1 and 2 (blue bars), while ρ_j determines the persistence of sector i's response following each of these two transmission rounds (ever lighter gray bars). Under $\alpha = (.8, .2)'$, the response starts at horizon 1 and is highest at horizons 1 and 2, whereas under $\alpha = (.2, .8)'$, it effectively starts at horizon 2 and is highest at horizons 2 and 3. The right panel shows that the impact of α on the timing of the peak-response is even more pronounced for units who only share a second-order connection. Under $\alpha = (.8, .2)'$, the response starts at horizon 2 and is highest at horizons 2 and 3, whereas under $\alpha = (.2, .8)'$, it effectively starts at horizon 3 and is

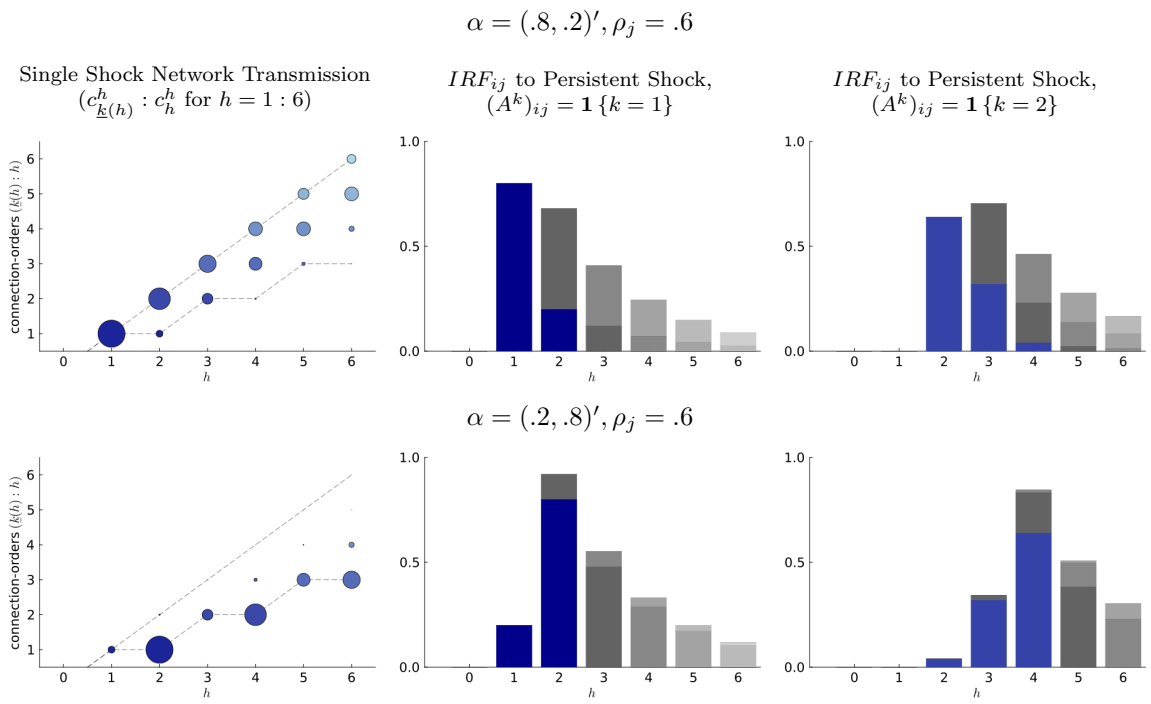


Figure 2: Composition of Impulse Responses under NVAR$(2, 1)$ with AR(1)-Shocks

Notes: The top plots refer to an NVAR$(2, 1)$ with $\alpha = (.8, .2)'$, the bottom plots to $\alpha = (.2, .8)'$. The left panel depicts the coefficients $c^h_{\underline{k}(h)} : c^h_h$ from Eq. (3) for $h = 1 : 6$, 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 middle panel plots the resulting IRF $\partial y_{i,t+h}/\partial y_{j,t}$ for a pair (i, j) with only a first-order connection to a persistent AR(1) shock $\varepsilon_{j,t}$ with $\rho_j = .6$ (see Eq. (15)). The right panel repeats this for a pair with only a second-order connection.

highest at horizons 4 and 5.

Summing up, α determines the timing of the initial TFP shock transmission from a supplier- to a customer-sector along a direct link, hence also the timing of its transmission along any higher-order link and, as a result, also the extent to which the initial response occurs later for sectors with higher relative to sectors with more immediate supply-chain relations to the shocked sector j. In turn, ρ_j determines the persistence of a sector i's response following each of these transmission rounds, without influencing the timing of the initial response, but possibly delaying the peak response (see top right plot in Fig. 2) In addition, ρ_j rationalizes the persistence of a sector j's response to its own TFP shocks without regard to network connections, whereas α requires cycles in the input-output network, i.e. walks from j to j. The interplay of these forces for the example network from Section 2.1 is illustrated by the full set of IRFs in Fig. A-1.

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. Thereby, 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. This result is unsurprising, given that any TFP level yields the same steady state in both economies

(disregarding differences in A between the two). For $\rho_j = 0$ and $\rho_a = 0$, this is implied by Proposition 3, but it holds more generally:

Proposition 5 (NVAR$(p,1)$: Long-Term Response to Persistent Innovations)**.** *Let $y_\tau = \alpha_1 A y_{\tau-1} + ... + \alpha_p A y_{\tau-p} + u_\tau$ and $x_\tau = aAx_\tau + u_\tau$ with $a = \sum_{l=1}^{p} \alpha_l$. Assume $(1-\rho_j L)u_{j\tau} = \varepsilon_{j\tau} \sim WN$ with $\rho_j \in [0,1)$, and assume α and A satisfy the conditions from Proposition 4 s.t. y_t is weakly stationary. Then*

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

To take the processes in Eq. (13) and Eq. (14) 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):

$$\Delta y_{\tau/q} = \Delta \tilde{y}_\tau + ... + \Delta \tilde{y}_{\tau-q+1} \quad \text{for } \tau/q = 1, ..., T , \tag{16}$$

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.[46] As the meaning of q depends on the frequency of observations, 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. (13) and (16) and Eqs. (14) and (16), 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.

[46] See Appendix C.1 as well as Section 2.2.3 and Appendix A.2 for more detailed discussions of these points. In the limit as $q \to \infty$, the long-term response referenced in Proposition 5 – a function of all connection-orders – materializes after a single observational period.

I compute log differences of monthly industrial production (IP) indices across 116 manufacturing and mining sectors in the US economy provided by the Federal Reserve Board.[47] 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

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

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.[48] 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 differences in the meaning and proper calibration of a_{ij} between contemporaneous and lagged IOC, effectively assuming $\beta \to 1$. 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 Appendix C.2.

For the specification with lagged IOC, I consider $p = \{1, 2, 3, 4, 5, 6\}$ and $q \in \{1, 2, 3, 4\}$, implying network interaction frequencies of a month, two weeks, 10 days and one week. 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 $\lambda_1 = 1$, redefining λ. 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: $\theta = (\alpha', \gamma', \lambda', \rho_a, \rho', \sigma^{2\prime})'$, 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.[49]

The inference from Section 3 is not applicable because of the autocorrelation and factor structure of TFP processes and because of the restrictions on α under lagged IOC. The likelihood of both models can be evaluated with a Kalman filter, as stated in Appendix C.3. I consider Bayesian inference on θ under uniform priors on the respective domains.[50] As a result, the posterior mode equals the Maximum Likelihood estimator. The posterior is

[47] The same data is used in Foerster et al. (2011) and Carvalho and Reischer (2021).

[48] 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.

[49] 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 $p = 6$.

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

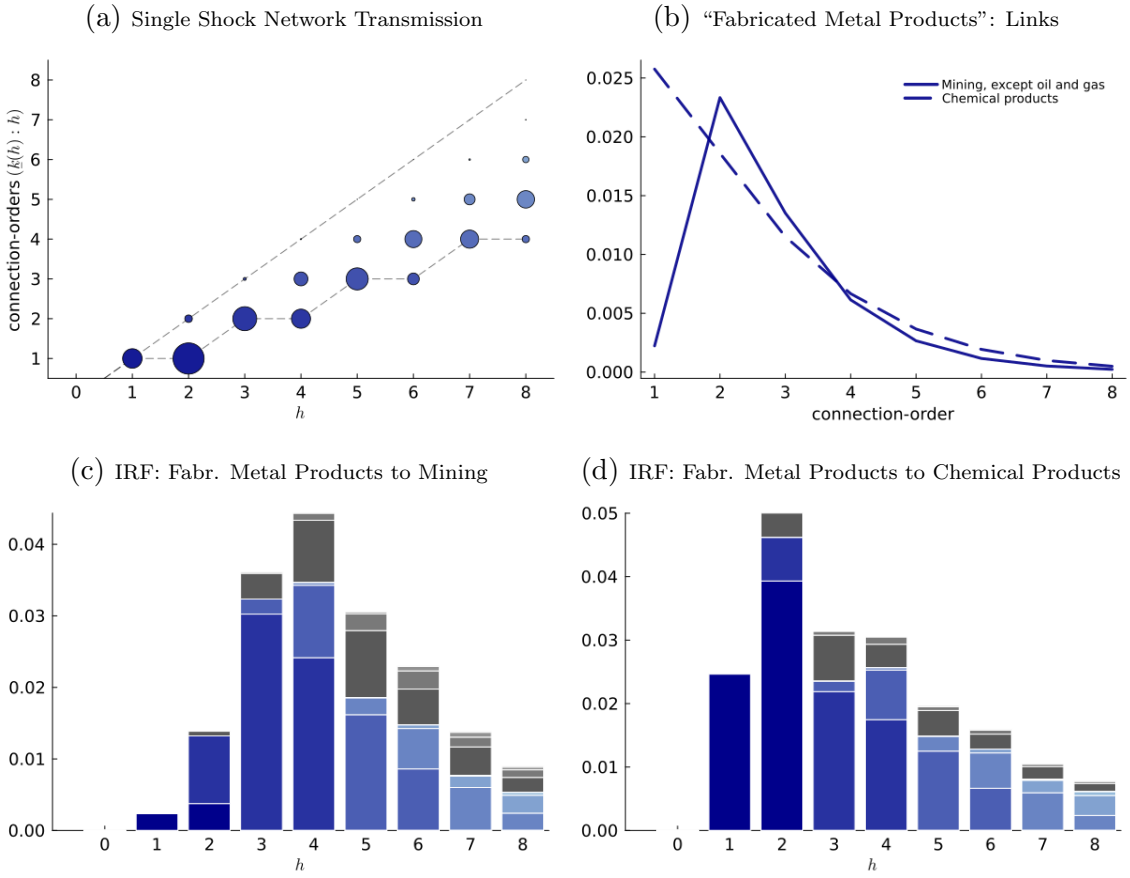


Figure 3: Impulse Responses to Idiosyncratic TFP Shocks

Notes: The top left panel depicts the coefficients $c^h_{\underline{k}(h)} : c^h_h$ from Eq. (3) 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.

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

The following, preliminary results are based on the model with lagged IOC under $p = 2$ and $q = 1$, applied at the more coarse aggregation level of 23 manufacturing sectors.

Fig. 3 illustrates the composition of impulse responses, analogously as Fig. 2 does for the example from Section 4.1. 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,

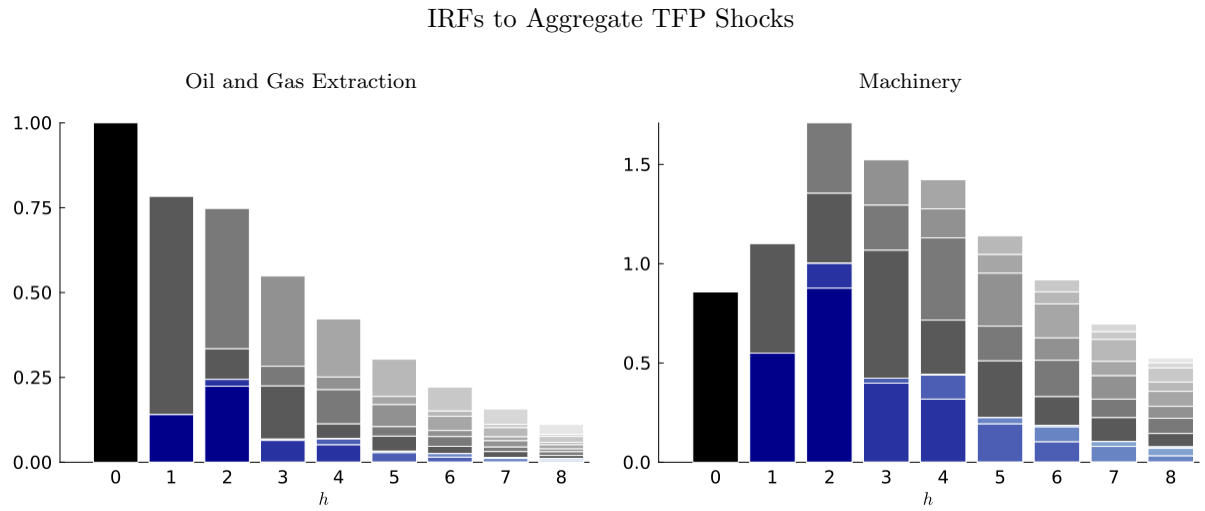


Figure 4: Impulse Responses to Aggregate TFP Shocks

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".

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 (transitory) response of output growth. Nevertheless, the magnitudes are rather small. This is expected; after all, these are 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 rather positioned at the top of supply 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.

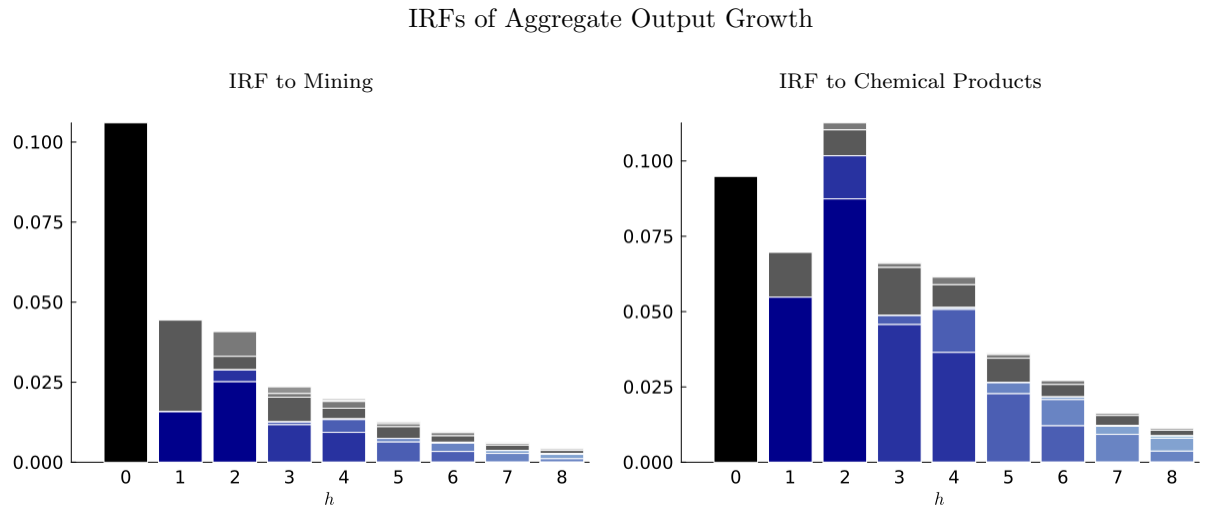


Figure 5: Impulse Responses of Aggregate Output Growth

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".

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.

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 5, 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 in terms of the strength of their impact on aggregate output, but also by the timing of this impact. 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 similarly small 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.

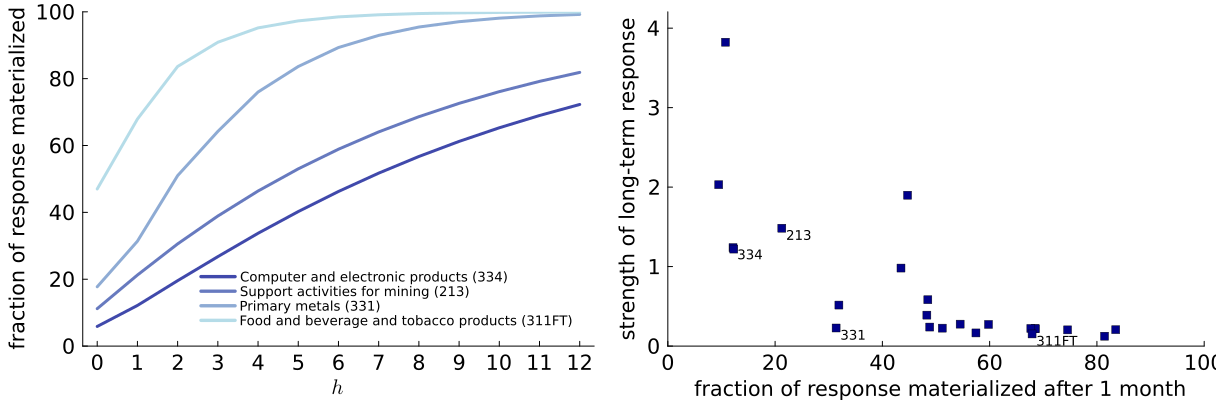


Figure 6: Size and Timing of IRFs of Aggregate Output Growth

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.

The autocorrelation of aggregate output growth at the posterior mean is estimated to be 0.85. In a hypothetical environment without persistence in exogenous shocks, this number drops to 0.28. As a result, lagged IOC can account for about a third of aggregate persistence. Together with persistence in the aggregate TFP process, this number increases to 0.83, leaving little room for persistence in sectoral TFP process in explaining business cycles in this RBC environment.

5 Parsimonious Networks & Cross-Country Macroeconomic Dynamics

The previous section examined a case where a cross-sectional time series is arguably driven by one particular, observed network and interest lies in quantifying how network effects materialize over time. In this section, I consider the problem of forecasting a rather high-dimensional, cross-sectional time series and, for the most part, I assume that no network data is available. By rationalizing the dynamic comovement of all units' series as the dynamic innovation transmission along a few fixed bilateral links among units, the NVAR offers a parsimonious, yet flexible method for forecasting cross-sectional – and possibly also general – high-dimensional time series.

I first discuss the merits of the NVAR as a tool for modeling high-dimensional time series and the relation to alternative methods – in particular the factor model – in Section 5.1. The application to cross-country industrial production growth is set up in Section 5.2, with results presented in Section 5.3. Details are in Appendix D.

5.1 Innovation Transmission Through Parsimonious Networks

Consider the problem of approximating the dynamics of a cross-sectional time series $y_t \in \mathbb{R}^n$. Even for intermediate n, an unrestricted VAR(p) is not feasible. The NVAR(p, q) offers a parsimonious, flexible and interpretable alternative.

Consider first the case of $q = 1$. Parsimony is owed to two considerations. First, by construction, the NVAR posits that innovations transmit cross-sectionally only via bilateral links. As a result, relative to a VAR(p), for $p > 1$, the NVAR($p, 1$) reduces the number of parameters in the autoregressive matrices from pn^2 to n^2+p-1. A single network adjacency matrix A with n^2 parameters is used to summarize the information in the high-dimensional vector of potential covariates – lags of y_{it}, $i = 1 : n$ – using first-order network connections:

$$y_t = X_t\alpha + u_t \ , \quad X_t = [Ay_{t-1}, ..., Ay_{t-l}] \ , \tag{17}$$

where x_{it} is the ith row of X_t. In turn, the $p-1$ free parameters in α determine the importance of different lags in x_{it} for the dynamics of y_{it}. In addition, a rationale for shrinking A to zero and increasing parsimony is provided by the fact that the dynamic comovement of two units can be captured with higher-order connections, even in absence of direct links between them. With A shrunk to zero, the dynamic comovement of y_{it} and y_{jt} for all unit-pairs (i, j) is modeled as the result of innovation transmission along a small set of bilateral links. This is akin to the assumption that longer-term dynamics are driven by a set of shorter-term dynamics, which is upheld by the general class of VARMA(p, q) models.

Flexibility is obtained because the n^2 links in A are estimated and because the NVAR(p, q) allows for rich mappings between connection-orders and dynamics. Relative to $q = 1$, taking $q < 1$ with $q^{-1} \in \mathbb{N}$ captures dynamics at lower frequency and avoids placing too much weight on the possibly noisy recent past. With the cost of introducing noise in the estimation process through filtering, taking $q > 1$ with $q \in \mathbb{N}$ circumvents the restrictions implied by a parsimonious A by incorporating higher-order connections into short-term dynamics. The possibility of taking $q = q_1q_2 \in \mathbb{Q}_{++}$ with $q_1^{-1} \in \mathbb{N}$ and $q_2 \in \mathbb{N}$ generalizes and moderates these two qualitative choices.

By reducing the number of parameters and using shrinkage priors (applying regularization), the NVAR leverages both ways of addressing the large parameter problem in the Wold representation (see Geweke (1984)). Rather than applying shrinkage to the parameters in a VAR(p), the NVAR applies shrinkage to links in A, which in turn summarize the information in the predictors at all lags $y_t, ..., y_{t-p}$ (see Eq. (17)). This leads to additional parsimony and implies that for any series y_{it}, the same linear combinations of predictor-variables $\{y_{j,t-l}\}_{j=1:n}$ enter the linear projection $\mathbb{E}_{t-1}[y_{it}]$ at all lags $l = 1 : p$. Nevertheless, dynamics at different horizons h are driven by different linear combinations, i.e. different connection-orders, as determined by α (see Propositions 1 and 2).

Like factor models and reduced rank regressions, the NVAR reduces dimensionality by summarizing a large set of predictors by a few linear combinations. In contrast to them, it allows these linear combinations to differ more flexibly across units, to the point that it naturally acommodates sparse factors as (linear combinations of) locally important units in the network.[51] However, this comes at the cost of restricting factor dynamics, as illustrated by Proposition 6, which formalizes the comparison of the NVAR and the (dynamic) factor model.

Proposition 6 (NVAR$(p,1)$-Factor Model Equivalence Result)**.**
Let y_t follow an NVAR$(p,1)$: $y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t$. It holds that

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

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

Conversely, let $y_t = \Lambda f_t + \xi_t$ with $f_t = \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t$ being r-dimensional, and assume $\Phi_l = \phi_l \Phi$, $l = 1 : p$. As $n \to \infty$, we obtain

$$y_t = \phi_1 A y_{t-1} + ... + \phi_p A y_{t-p} + u_t$$

for $u_t = \Lambda \eta_t + \xi_t$ and $A = \Lambda \Phi (W\Lambda)^{-1} W$, where W is any $r \times n$ matrix with distinct rows.[52]

Based on this result, the NVAR is expected to capture cross-sectional dynamics in finite samples better than the factor model in the presence of numerous sparse factors or, equivalently, in case of a sparse, yet close-to-full-rank network adjacency matrix A. Intuitively, this occurs when dynamics are driven by many bilateral links rather than a few influential units. More generally, the NVAR is expected to improve upon the poor forecasting performance of factor models in the case of high dispersion of factor loadings across series (see Boivin and Ng (2006)).[53] [54]

[51] The factor model constructs $\{\mathbb{E}_{t-1}[y_{it}]\}_{i=1}^n$ as *different* linear combinations of *the same* factors, which in turn are linear combinations of observables. In contrast, the NVAR models $\{\mathbb{E}_{t-1}[y_{it}]\}_{i=1}^n$ as *the same* linear combination of *different* constructed covariates x_{it}, which in turn are linear combinations of observables.

[52] The second part of the proposition restricts the dynamic evolution of the r factors to an NVAR$(p,1)$. For $p = 1$, this is equivalent to $f_t \sim VAR(p)$, while under $r = 1$ it comes equal to $f_t \sim AR(p)$. As usual, these factor representations are not unique. Relatedly, we can re-scale A and ϕ to ensure $a_{ij} \in [-1, 1]$.

[53] This notably includes the case of sparse factors, i.e. loading-vectors with zero and non-zero entries.

[54] As Boivin and Ng (2006) point out, even if the number of factors is selected separately for each series, the forecasts of series that depend on less dominant factors are nevertheless more noisy than those for series that depend on the most dominant factors because including more estimated factors adds noise to the forecasts.

5.2 Data, Application-Setup & Estimation

To validate the NVAR's merit for forecasting suggested by theory, I compare its performance to that of the benchmark-Dynamic Factor Model (DFM) of Geweke and Zhou (1996) in forecasting a range of macroeconomic time series across countries. The estimation of the DFM is outlined in Appendix D.4. Two types of NVARs are estimated, applying both shrinkage with a Normal prior/L2-penalty and an Exponential prior/L1-penalty, as outlined in Section 3.2. They are labeled NVAR-R and NVAR-L in the following.

The models are used to forecast monthly industrial production (IP) growth, monthly CPI inflation and quarterly real GDP growth across OECD member countries, applicants and partners.[55] The series are obtained from the IMF's IFS database, and YoY growth rates are used. Limiting attention to the pre-COVID time frames 2001:M1 - 2019:M12 and 2001:Q1 - 2019:Q4, respectively, and deleting countries with missings or more than two consecutive constants yields datasets with $n = 42$, $n = 34$ and $n = 39$ observations. The series are deseasonalized by subtracting fitted values from a regression on period-dummies, after which the resulting series are standardized to have mean zero and standard deviation one.

The forecasting performance is assessed using expanding windows with 24 and 16 forecasting origins, from 2017:M12 to 2019:M11 and 2015:Q4 to 2019:Q3, respectively. 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. This yields $T = 203$ and $T = 60$, respectively.

The models are compared based on mean squared errors (MSEs) across forecasting horizons, whereby the mean is taken over countries and forecasting origins. For the DFM, I consider $p = 1 : 4$ lags and $r = 10$ factors, after which the ex-post best-performing DFM is selected as the one that minimizes the cumulative MSE over the first three horizons. This renders the assessment independent of methods to select the number of factors ex-ante. For the NVARs, I also consider $p = 1 : 4$ lags, and I mostly focus on $q = 1$, the case referred to by Proposition 6. The hyperparameter λ_a controlling the degree of shrinkage is selected according to two methods. The first finds its marginal posterior mode (MPM) using a hierarchical modeling approach with a uniform hyperprior for λ_a. The resulting λ_a maximizes the marginal data density (MDD) and hence one-step ahead predictive ability (see Giannone et al. (2015)).[56] For NVAR-L, a second approach maximizes the Bayesian Information Criterion (BIC) by counting the number of non-zero elements in $\hat{A}(\lambda)$, as suggested in Zou et al. (2007) for other Lasso-applications. For NVAR-R, an analogous heuristic method is obtained by maximizing the conditional MDD $p(Y|\lambda_a, B, \alpha, \Sigma)$, derived in Appendix D.2.

[55] 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).

[56] For this purpose, the Gibbs sampler from Section 3.2 is augmented with a step where λ_a is drawn from its conditional posterior, derived in Appendix D.2.

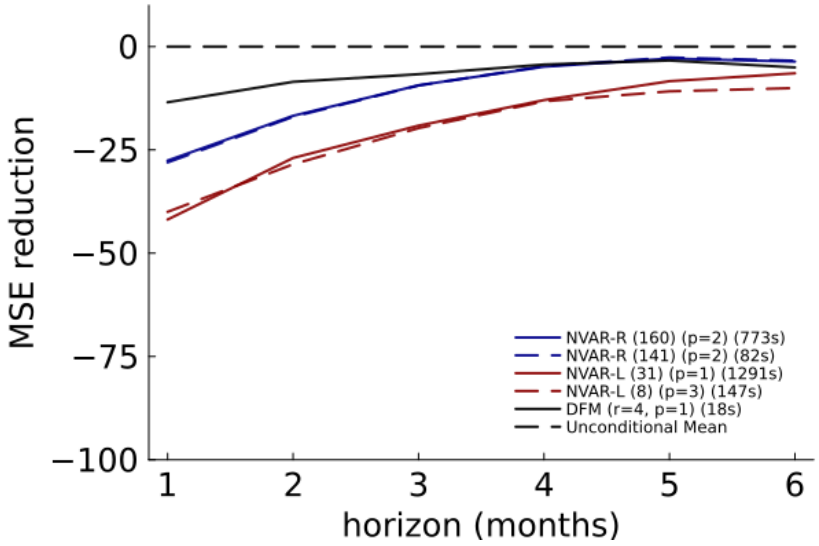


Figure 7: Forecasting: NVAR$(p, 1)$ vs. Factor Model, Industrial Production Growth

Notes: The plot depicts the out-of-sample Mean Squared Errors generated by different versions of the NVAR$(p, 1)$ and those generated the ex-post best Dynamic Factor Model, all relative to those generated by an unconditional mean forecast. All forecasts refer to those obtained under the posterior mode.

These criteria are maximized using a grid-search. Unless otherwise stated, the links are shrunk to $B = 0$.

5.3 Results

Fig. 7 illustrates the results for industrial production 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 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.[57]

[57]The models' performances under different different choices of p and r are shown in Fig. A-2 and Fig. A-3.

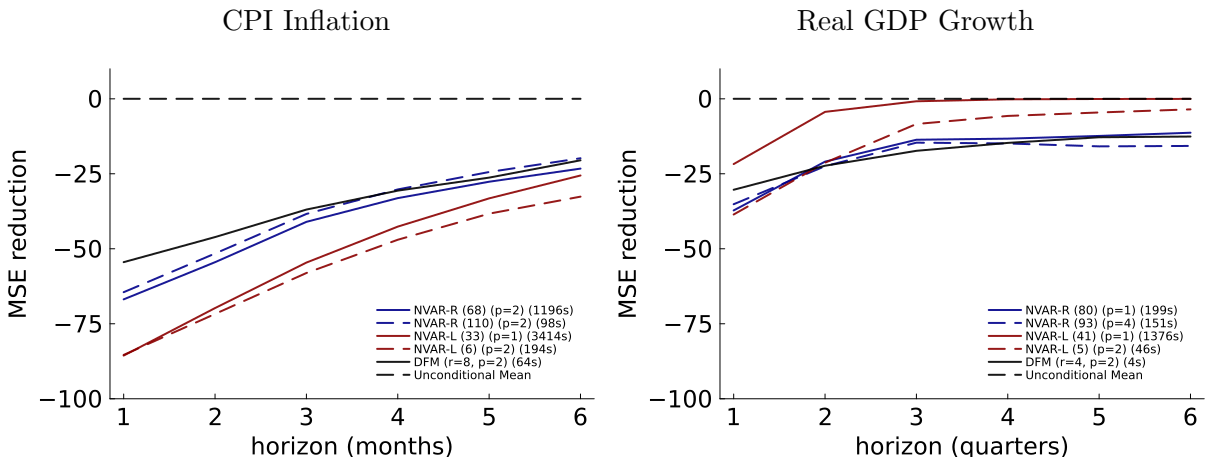


Figure 8: Forecasting: NVAR$(p, 1)$ vs. Factor Model, CPI Inflation & Real GDP Growth

Notes: The plot depicts the out-of-sample Mean Squared Errors generated by different versions of the NVAR$(p, 1)$ and those generated the ex-post best Dynamic Factor Model, all relative to those generated by an unconditional mean forecast. All forecasts refer to those obtained under the posterior mode.

The results in Fig. 7 are obtained using the respective posterior mode, i.e. frequentist point estimator. As shown in Fig. A-2, 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.[58] 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 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.[59]

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

[58] 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. $Ay_{t-1} + \alpha_2 Ay_{t-2} \approx (1 + \alpha_2)Ay_{t-1}$. 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.

[59] 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.

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%.

6 Conclusion

In this paper, I develop the Network-VAR (NVAR); an econometric framework that rationalizes the dynamics of a cross-sectional variable as the dynamic innovation transmission along bilateral links among cross-sectional units. After characterizing dynamics in relation to network connections and inference of both the timing of network effects and the network itself, I consider two applications. First, consistent with an RBC economy with lagged input-output conversion, I use the NVAR to characterize and estimate the latter's contribution to business cycles, i.e. to persistence in aggregate output growth. Second, I use the NVAR as a dimensionality-reduction technique for forecasting cross-country macroeconomic aggregates.

Rather than assuming time-invariant links, an important methodological step forward would be to develop a tractable framework that links dynamic network effects to dynamic network formation, allowing units to adjust their their connections in reaction to innovations transmitted through the network. This would permit the study of economic growth through network formation in a dynamic context and of the dynamic consequences of network destruction, such as witnessed through geoeconomic fragmentation in international trade. Methodologically interesting in this context is the study of cointegration relationships as a function of network connectedness.

Furthermore, the NVAR could be augmented to accommodate heterogeneous propagation patterns across units or over time. Relatedly, the application of the NVAR for forecasting time series in very high dimensions would benefit from refinements of the crude shrinkage priors used in this paper.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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," *Manuscript, Bank of England*.

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

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

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

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

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

Fan, J., S. Hong, and F. Parro (2023): "Learning and Expectations in Dynamic Spatial Economies," *NBER Working Paper Series, No. 31504*.

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

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

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

Geweke, J. (1984): "Inference and Causality in Economic Time Series Models," *Handbook of Econometrics*, 11, 1101–1144.

Geweke, J. and G. Zhou (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The 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.

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

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

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

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

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

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

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

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

Knight, M. I., M. A. Nunes, and G. P. Nason (2016): "Modelling, Detrending and

Decorrelation of Network Time Series," *Manuscript, University of Bristol*.

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

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

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

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

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, 1415–1435.

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

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

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

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

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

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

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

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

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

Appendix

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

Marko Mlikota
Geneva Graduate Institute

A NVAR: Theory

A.1 Granger-Causality & Generalized Impulse-Responses

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

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

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

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

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

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

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

$h = 1$. Assume it is true for h. For $h+1$ we have

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

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

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

Proposition 2 (Granger-Causality in NVAR(p,q))**.**
Let y_t be a stock variable and follow an NVAR(p,q) for $q \in \mathbb{N}\backslash\{1\}$,

$$\tilde{y}_\tau = \alpha_1 A\tilde{y}_{\tau-1} + ... + \alpha_p A\tilde{y}_{\tau-p} + \tilde{u}_\tau \ , \quad \tau = 1:T_\tau \ ,$$
$$y_{\tau/q} = \tilde{y}_\tau \quad if\, \tau/q \in \mathbb{N} \ ,$$

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

Proof: This is a direct consequence of Proposition 1 and the stated definition of y_t, which allow us to write

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

For a flow variable, $\partial y_{t+h}/\partial y_t$ can be composed of network-connections of order $k \in \{ceil((q(h-1)+1)/p), ..., hq\}$, depending on which of the terms in $y_t = \tilde{y}_{tq} + ... + \tilde{y}_{tq-q+1}$ is responsible for the change in y_t. If it is the first term, we get

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

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

Proposition 3 (Long-Term Response in NVAR$(p,1)$)**.**
Let y_t follow an NVAR$(p,1)$:

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

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

$$x = aAx + \varepsilon \ .$$

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

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

Proof: First, note that

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

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

Turning to y_t, note that under stationarity

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

To find the impulse response function for y_t, write it in companion form as

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

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

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

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

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

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

Note that

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

Therefore,

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

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

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

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

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

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

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

which implies

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

The first element gives

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

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

Note that these long-term responses to permanent innovations are not affected by time-aggregation. Specifically, for a stock variable $y_t \sim \text{NVAR}(p,q)$, $q \in \mathbb{N}\backslash\{1\}$, we can write $\{y_t\}_{t=1}^T = \{\tilde{y}_{tq}\}_{t=1}^T$ and we have

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

For a flow variable, we have $\{y_t\}_{t=1}^T = \{\tilde{y}_{tq} + ... + \tilde{y}_{(t-1)q+1}\}_{t=1}^T$ and

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

If y_t is obtained by averaging instead of summing up, q cancels out.

A.2 Time-Aggregation of Lagged Transmission Patterns

NVAR(p,q), $q \in \mathbb{N}\backslash\{1\}$ at Observational Frequency The following discussion suggests that, if $\tilde{y}_\tau$ is stationary, the dynamics of $y_t \sim \text{NVAR}(p,q)$, $q \in \mathbb{N}\backslash\{1\}$ from the state space representations in Eq. (4) and Eq. (5) can be approximated arbitrarily well by the restricted VARMA process

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

by taking p^* large. The coefficients γ_{lg} are polynomials of $\{\alpha_l\}_{l=1:p}$ (or zero), and $q^* \leq p^*q$. $\eta_t = (\tilde{u}'_\tau, ..., \tilde{u}'_{\tau-q+1})'$ stacks all high-frequency innovations occurring in-between periods of observation $t-1$ and t.[60]

To illustrate, consider an NVAR(3, 2). First, let y_t be a stock variable. W.l.o.g., let the particular realization at period τ, $\tilde{y}_\tau$, be observed. Inserting sequentially for the non-observed $\tilde{y}_{\tau-1}$ and $\tilde{y}_{\tau-3}$, we get

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

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

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

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

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

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

Let y_t instead be a flow variable. Starting from $y_t = \tilde{y}_\tau + \tilde{y}_{\tau-1}$, iteratively inserting for $\tilde{y}_\tau$ and $\tilde{y}_{\tau-1}$ and adding and subtracting an $\tilde{y}_{\tau-3}$-term so as to align the coefficients in front of $\tilde{y}_{\tau-2}$ and $\tilde{y}_{\tau-3}$, we obtain

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

Inserting for the terms in the second row and adding and subtracting a $\tilde{y}_{\tau-5}$-term, we get

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

which can be written as

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

[60] Furthermore, the $n \times nq$ matrices $\{\Theta_l\}_{l=0}^{p^*-1}$ are made up of $n \times n$ blocks, all of which are polynomials in A of the form of Φ_l. Provided $p \geq q$ and assuming $\alpha_l \neq 0 \ \forall \ l$, Φ_1 is a linear combination of $\{A, A^2, ..., A^q\}$, while $\Theta_0 = [\Theta_{0,0}, \Theta_{0,1}, ..., \Theta_{0,q-1}]$ whereby $\Theta_{0,k}$ is a linear combination of $\{A^0, A^1, ..., A^k\}$.

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

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

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

Note that $q^* \leq p^* q$ in Eq. (A.1), i.e. the largest power of A in any Φ_l cannot exceed $p^* q$. To see this, note that starting from the equation for $\tilde{y}_\tau$ and iteratively inserting for each $\tilde{y}_{\tau-j}$, $j = 1 : h-1$, one would obtain $\partial \tilde{y}_{\tau+h} / \partial \tilde{y}_\tau$ as the coefficient-matrix in front of $\tilde{y}_{\tau-h}$. From Proposition 1, we know it consists of powers of A in the set $k \leq h$. Continuing this way, the coefficient-matrix in front of $y_{t-l} = \tilde{y}_{(t-l)q}$ would contain powers of A not higher than lq. In contrast to these GIRF-calculations, when performing the time-aggregation (for stock variables), every qth $\tilde{y}_\tau$ is not inserted for, which means that powers of A accumulate slower. As a result, Φ_l, the coefficient-matrix in front of y_{t-l}, contains powers of A not higher than lq, which is maximized for $l = p^*$. Analogous reasoning applies to flow variables.

Contemporaneous Correlation in Observed Innovations Define the innovation at observation frequency as the error in the linear projection of y_t on information available at time $t-1$: $u_t \equiv y_t - \mathbb{E}_{t-1}[y_t] = \Theta_0 \eta_t$, with $\eta_t = (\tilde{u}'_\tau, ..., \tilde{u}'_{\tau-q+1})'$. Even with cross-sectionally (and temporally) uncorrelated innovations $\tilde{u}_\tau$, u_t is cross-sectionally correlated. Let $\text{Cov}(\tilde{u}_{i\tau}, \tilde{u}_{j\tau}) = \mathbf{1}\{i = j\}\sigma_i^2$ and consider first $q = 2$ and a stock variable y_t. Then

$$\mathbb{V}[u_t] = \Theta_0 \mathbb{V}[\eta_t] \Theta_0' = [I_n, \alpha_1 A] \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} I_n \\ \alpha_1 A' \end{bmatrix} = \Sigma + \alpha_1^2 A \Sigma A' \ ,$$

for $\Sigma = diag(\sigma_1^2, ..., \sigma_n^2)$. This reveals that $\text{Cov}(u_{it}, u_{jt}) = \mathbf{1}\{i = j\}\sigma_i^2 + \alpha_1^2 \sum_{k=1}^n a_{ik} a_{jk} \sigma_k^2$. For flow variables, we have $\Theta_0 = [I_n, I_n + \alpha_1 A]$, which leads to

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

i.e. $\text{Cov}(u_{it}, u_{jt}) = \mathbf{1}\{i = j\} 2\sigma_i^2 + \alpha_1(a_{ij}\sigma_j^2 + a_{ji}\sigma_i^2) + \alpha_1^2 \sum_{k=1}^n a_{ik} a_{jk} \sigma_k^2$. This means that $\text{Cov}(u_{it}, u_{jt})$ is determined by common exposures to third units (and bilateral exposure), exposure meaning first order connections.[61] Under $q = 3$, exposure is determined by first and

[61] 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: $\text{Cov}(u_{it}, u_{jt}) \neq 0$ 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

second order connections. For example, for a stock variable with $q = 3$, $\Theta_0 = [I_n, \alpha_1 A, \alpha_2 A + \alpha_1^2 A^2]$ and

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

A.3 Stationarity

Proposition 7 (Stationarity of NVAR$(1,1)$)**.**
Let y_t follow an NVAR(1,1):

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

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

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

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

We have

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

where the last equivalence follows from

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

Proposition 8.
Let y_t follow an NVAR$(p,1)$:

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

bilateral exposure: $a_{ij} \neq 0 \mid a_{ji} \neq 0 \;\Rightarrow\; \text{Cov}\,(u_{it}, u_{jt}) \neq 0$.

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

$$y_t^* = aAy_{t-1}^* + u_t^* \ .$$

Then, y_t is weakly stationary if y_t^* is weakly stationary. If in addition $\alpha_l \geq 0$ for all l, this implication is both-sided.

Proof: Let

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

First, assume that α_l can be negative. 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

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

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

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

a contradiction.

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

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

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

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

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

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

If in addition $\alpha_l \geq 0 \; \forall \; l$, then this condition is both necessary and sufficient.

Proposition 9 (Stationarity of NVAR$(p, 1)$ II)**.**
Let y_t follow an NVAR$(p, 1)$:

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

with $u_t \sim WN$ and $\alpha_l \neq 0$ for at least one l.

Then, y_t is weakly stationary iff for all eigenvalues λ_i of A, the matrix

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

has all eigenvalues inside the unit circle.

Note that for real λ_i, this is equivalent to weak stationarity of the univariate AR(p) process

$$x_t = \lambda_i \alpha_1 x_{t-1} + ... + \lambda_i \alpha_p x_{t-p} + v_t \; .$$

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

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

Proposition 10 (Stationarity Preservation Under Time-Aggregation)**.**
Let $\tilde{y}_\tau$ follow an NVAR(p, 1)

$$\tilde{y}_\tau = \alpha_1 A\tilde{y}_{\tau-1} + ... + \alpha_p A\tilde{y}_{\tau-p} + \tilde{u}_\tau \ ,$$

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

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

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

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. $Cov(y_t, y_{t-h}) = Cov(\tilde{y}_{tq}, \tilde{y}_{(t-h)q}) = Cov(\tilde{y}_{(t-l)q}, \tilde{y}_{(t-l-h)q}) = Cov(y_{t-l}, y_{t-l-h}) \ \forall \ l, h \ ,$

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

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

which is the definition of stationarity for z_t. $\blacksquare$

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 11 (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) $\frac{1}{n}\sum_{i=1}^{n} \left(A_{n,i\cdot} y_{t-l}\right)' \left(A_{n,i\cdot} y_{t-k}\right) \xrightarrow{p} \mathbb{E}\left[\left(A_{*,i\cdot} y_{t-l}\right)' \left(A_{*,i\cdot} y_{t-k}\right)\right]$*; and*
   (b) $\frac{1}{n}\sum_{i=1}^{n} \left(A_{n,i\cdot} y_{t-l}\right)' u_{it} \xrightarrow{p} \mathbb{E}\left[\left(A_{*,i\cdot} y_{t-l}\right)' u_{it}\right]$.
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* $n \to \infty$*,*

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

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 12 (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),

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

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, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'u_t \xrightarrow{d} N\left(0, \mathbb{V}[X_t'u_t]\right)$, as $\mathbb{E}[X_t'u_t] = 0$. Thereby, $\mathbb{V}[X_t'u_t] = \mathbb{E}[X_t'u_tu_t'X_t] = \mathbb{E}[X_t'\Sigma X_t]$ 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 13 (Large (n,T) Consistency & Asymptotic Normality of $\hat{\alpha}_{OLS}$)**.**
Suppose either i) the conditions in Proposition 11 hold, or ii) the conditions in Proposition 12 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 $\tilde{y}_\tau \sim \text{NVAR}(p,1)$ 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) $\tilde{x}_\tau = \alpha_1 \tilde{x}_{\tau-1} + ... + \alpha_p \tilde{x}_{\tau-p} + \breve{u}_\tau$ 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

$$\begin{bmatrix} \gamma_1 \\ \gamma_3 \\ ... \\ \gamma_{\underline{m}} \end{bmatrix} = \overline{A} \begin{bmatrix} \gamma_1 \\ \gamma_3 \\ ... \\ \gamma_{\underline{m}} \end{bmatrix} + \underline{A} \begin{bmatrix} \gamma_0 \\ \gamma_2 \\ ... \\ \gamma_{\overline{m}} \end{bmatrix} = (I - \overline{A})^{-1} \underline{A} \begin{bmatrix} \gamma_0 \\ \gamma_2 \\ ... \\ \gamma_{\overline{m}} \end{bmatrix} , \tag{A.2}$$

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

$$\begin{bmatrix}\gamma_0 - \sigma^2 \\ \gamma_2 \\ ... \\ \gamma_{\overline{m}}\end{bmatrix} = \underline{B} \begin{bmatrix}\gamma_1 \\ \gamma_3 \\ ... \\ \gamma_{\underline{m}}\end{bmatrix} + \overline{B} \begin{bmatrix}\gamma_0 \\ \gamma_2 \\ ... \\ \gamma_{\overline{m}}\end{bmatrix}, \tag{A.3}$$

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.2) and Eq. (A.3) 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

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

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

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

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* $[s_\tau^m]_1$ *from* $N\left(\left[s_{\tau|\tau+1}\right]_1, \left[P_{\tau|\tau+1}\right]_{11}\right)$ *with*
   - $s_{\tau|\tau+1} = s_{\tau|\tau} + P_{\tau|\tau} R_{1\cdot}' \left(R_{1\cdot} P_{\tau|\tau} R_{1\cdot}' + \Sigma\right)^{-1} \left(\left[s_{\tau+1}^m\right]_1 - \left[s_{\tau+1|\tau}\right]_1\right) \,,$
   - $P_{\tau|\tau+1} = P_{\tau|\tau} - P_{\tau|\tau} R_{1\cdot}' \left(R_{1\cdot} P_{\tau|\tau} R_{1\cdot}' + \Sigma\right)^{-1} R_{1\cdot} P_{\tau|\tau} \,.$

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 Kim and Nelson (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

$$\begin{aligned} p(A|Y, \alpha, \Sigma, B, \lambda_a) &\propto p(Y|\alpha, A, \Sigma)p(A|B, \lambda_a) \\ &\propto exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - Az_t)'\Sigma^{-1}(y_t - Az_t)\right\} exp\left\{-\frac{1}{2}\lambda_a \sum_{i,j=1}^{n}(a_{ij} - b_{ij})^2\right\} \\ &= exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}(Y - ZA')'(Y - ZA')\right]\right\} exp\left\{-\frac{1}{2}\lambda_a tr[(A-B)'(A-B)]\right\}, \\ &\propto exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left[A(Z'Z + \lambda_a\Sigma)A' - 2A(Z'Y + \lambda_a B'\Sigma)\right]\right]\right\}, \end{aligned}$$ [62]

which lets us deduce that

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

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. (10). Therefore, its conditional minimizer is the mode of $p(A|Y, \alpha, \Sigma)$. The (joint) minimzer of the objective function in Eq. (10) is the mode of $p(\alpha, A|Y, \Sigma, B, \lambda_a)$ under a uniform prior for α.

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

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

[62] 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.[63] 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. (10), 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

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

(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

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

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

[63] 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 14 (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:

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

converges in probability to zero because, under condition 1,

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

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 $z_t(k\alpha_*) = kz_t(\alpha_*)$.[64]

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,

$$\begin{aligned}\sqrt{T}Q^{(1)}_{n,T}(\theta_0; Y) &= \sqrt{T}\left[\begin{matrix}\frac{\partial Q_{n,T}(\theta;Y)}{\partial \alpha} \\ \frac{\partial Q_{n,T}(\theta;Y)}{\partial \overrightarrow{A}}\end{matrix}\right]\Bigg|_{\theta=\theta_0} \\ &= -2\begin{bmatrix} \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t'(y_t - X_t\alpha_*) \\ \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(y_t - \overrightarrow{A}_* z_t)z_t'] - \sqrt{T}\tilde{\lambda}[\overrightarrow{A_*} - B]\end{bmatrix} \xrightarrow{d} N(0, M)\ ,\end{aligned}$$

because

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

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,

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

[64] 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

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

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

$$\frac{\partial Q_{n,T}(\theta;Y)}{\partial \overrightarrow{A}\partial \overrightarrow{A}'} = \frac{2}{nT}\sum_{t=1}^{T}\begin{bmatrix}z_t' \otimes z_{1t}I_n \\ \vdots \\ z_t' \otimes z_{nt}I_n\end{bmatrix} \xrightarrow{p} \frac{2}{n}\begin{bmatrix}\mathbb{E}[z_t' \otimes z_{1t}I_n] \\ \vdots \\ \mathbb{E}[z_t' \otimes z_{nt}I_n]\end{bmatrix} \equiv H_{22} \ .^{65}$$

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.

[65] 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 Input-Output Links & Sectoral Output Dynamics

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

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

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

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

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

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 $\tilde{\lambda} = ln(\lambda)$. 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

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

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 $\beta \to 1$.

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:

$$\begin{aligned} V_i\left(\{\check{x}^{ij}_{t,t-2},\check{x}^{ij}_{t,t-1}\}_{j=1}^n\right) = \underset{\substack{l_{it},\{x^{ij}_t,\\ x^{ij}_{t,t-1},x^{ij}_{t,t-2}\}_j}}{max} \quad & \Pi_{it} + \beta V_i\left(\{\check{x}^{ij}_{t+1,t-1},\check{x}^{ij}_{t+1,t}\}_{j=1}^n\right) \\ \text{s.t.} \quad & \check{x}^{ij}_{t,t-2} = x^{ij}_{t,t-2}\ , \\ & \check{x}^{ij}_{t+1,t} = x^{ij}_{t}\ , \\ & \check{x}^{ij}_{t,t-1} = x^{ij}_{t,t-1} + x^{ij}_{t+1,t-1}\ . \end{aligned}$$

The problem can be written more compactly as

$$\begin{aligned} V_i\left(\{x^{ij}_{t,t-2},\check{x}^{ij}_{t,t-1}\}_{j=1}^n\right) = \underset{\substack{l_{it},\{x^{ij}_t,\\ x^{ij}_{t+1,t-1}\}_j}}{max} \Bigg[& p_{it}z_{it}l_{it}^{b_i}\prod_{j=1}^n\left[\alpha_1\left(\check{x}^{ij}_{t,t-1}-x^{ij}_{t+1,t-1}\right)^r+\alpha_2\left(x^{ij}_{t,t-2}\right)^r\right]^{\frac{a_{ij}}{r}} \\ & -w_tl_{it}-\sum_{j=1}^n p_{jt}x^{ij}_t\Bigg] + \beta V_i\left(\{x^{ij}_{t+1,t-1},x^{ij}_t\}_{j=1}^n\right) \end{aligned}$$

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

$$l_{it} = b_i\frac{p_{it}y_{it}}{w_t}\ , \quad x^{ij}_{t,t-1} = \beta\alpha_1 a_{ij}\frac{p_{it}y_{it}}{p_{j,t-1}}\ , \quad x^{ij}_{t,t-2} = \beta^2\alpha_2 a_{ij}\frac{p_{it}y_{it}}{p_{j,t-2}}\ .$$

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

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

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 $k^{p2}_i = -b_i ln(b_i) - \sum_{j=1}^n a_{ij}\left[\alpha_1 ln(\beta\alpha_1 a_{ij}) + \alpha_2 ln(\beta^2\alpha_2 a_{ij})\right.$

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

$$a_{ij} = \left[\beta\alpha_1 + \beta^2\alpha_2\right]^{-1} (p_j x^{ij})/(p_i y_i) \ .^{66}$$

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:

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

[66] 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,

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

Impulse-Response Results

For the processes in Eq. (13) and Eq. (14), we have:

$$\frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial \varepsilon_\tau^j} = \sum_{l=0}^{h} \frac{\partial \Delta \tilde{y}_{i,\tau+h}}{\partial e_{\tau+h-l}^j} \frac{\partial e_{\tau+h-l}^j}{\partial \varepsilon_\tau^j} .$$

Thereby, $\partial e^j_{\tau+h-l}/\partial \varepsilon^j_\tau = \rho_j^{h-l}$. Under Eq. (13), $\partial \Delta\tilde{y}_{i,\tau+h}/\partial e^j_{\tau+h-l} = [(I-A)^{-1}]_{ij}$ for $l=0$ and zero otherwise. Under Eq. (14), the expression is characterized by Eq. (3).

Proposition 5 (NVAR$(p,1)$: Long-Term Response to Persistent Innovations)**.** *Let $y_\tau = \alpha_1 A y_{\tau-1} + ... + \alpha_p A y_{\tau-p} + u_\tau$ and $x_\tau = aAx_\tau + u_\tau$ with $a = \sum_{l=1}^p \alpha_l$. Assume $(1-\rho_j L)u_{j\tau} = \varepsilon_{j\tau} \sim WN$ with $\rho_j \in [0,1)$, and assume α and A satisfy the conditions from Proposition 4 s.t. y_t is weakly stationary. Then*

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

Proof: For x_τ, we have

$$\sum_{h=0}^{H} \frac{\partial x_{i,\tau+h}}{\partial u_{j,\tau}} = \sum_{h=0}^{H} \left[(I-A)^{-1}\right]_{ij} \rho_j^h = \left[(I-A)^{-1}\right]_{ij} \sum_{h=0}^{H} \rho_j^h = \frac{1}{1-\rho_j} \left[(I-aA)^{-1}\right]_{ij} .$$

Taking the limit as $H\to\infty$ proves the statement, as $\lim_{H\to\infty} \sum_{h=0}^{H} \rho_j^h = \frac{1}{1-\rho_j}$.

For y_τ, we have

$$\begin{aligned}
\sum_{h=0}^{H} \frac{\partial y_{i,\tau+h}}{\partial u_{j,\tau}} &= \sum_{h=0}^{H} \sum_{l=0}^{h} \frac{\partial y_{i,\tau+l}}{\partial u_{j,\tau}} \rho_j^{h-l} \\
&= \sum_{l=0}^{H} \frac{\partial y_{i,\tau+l}}{\partial u_{j,\tau}} \sum_{h=l}^{H} \rho_j^{h-l} \\
&= \sum_{l=0}^{H} \frac{\partial y_{i,\tau+l}}{\partial u_{j,\tau}} \sum_{h=0}^{H-l} \rho_j^{h} .
\end{aligned}$$

For any l, $\lim_{H\to\infty} \sum_{h=0}^{H-l} \rho_j^h = \frac{1}{1-\rho_j}$. Therefore,

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

where the last equality follows from Proposition 3. ■

$q>1$ Now consider the response of $\Delta\tilde{y}_{i,t+h} = \sum_{k=0}^{q-1} \Delta\tilde{y}_{i,tq+h-k}$ to a (high-frequency) TFP shock in period t. As any observational period t consists of q periods at network interaction frequency, we can consider shocks to ε^j_τ, $\varepsilon^j_{\tau-1}$, ..., $\varepsilon^j_{\tau-(q-1)}$. For simplicity, I take $\alpha_1 = 1$. First, consider $h=0$. If the shock occurred in period τ (the last high-frequency period

within observational period t), we obtain

$$\frac{\partial \Delta \tilde{y}_{it}}{\partial \varepsilon_{\tau}^{j}} = \sum_{k=0}^{q-1} \frac{\partial \Delta \tilde{y}_{i,tq-k}}{\partial \varepsilon_{tq}^{j}} = \frac{\partial \Delta \tilde{y}_{i,tq}}{\partial \varepsilon_{tq}^{j}} = \sum_{l=0}^{0} \left[A^{l}\right]_{ij} \rho^{0-l} = \mathbf{1}\left\{i=j\right\} \ ,$$

as $\partial \Delta \tilde{y}_{i,tq-k}/\partial \varepsilon_{tq}^{j} = 0$ for $k > 0$ and $A^{0} = I$. Similarly, we get

$$\begin{aligned}
\frac{\partial \Delta \tilde{y}_{it}}{\partial \varepsilon_{\tau-1}^{j}} &= \mathbf{1}\left\{i=j\right\}(1+\rho_j) + \left[A^{1}\right]_{ij} \ , \\
\frac{\partial \Delta \tilde{y}_{it}}{\partial \varepsilon_{\tau-2}^{j}} &= \mathbf{1}\left\{i=j\right\}(1+\rho_j+\rho_j^2) + \left[A^{1}\right]_{ij}(1+\rho_j) + \left[A^{2}\right]_{ij} \ .
\end{aligned}$$

More generally, for $g \in \{0,1,...,q-1\}$, we have

$$\begin{aligned}
\frac{\partial \Delta y_{it}}{\partial \varepsilon_{\tau-g}^{j}} &= \sum_{k=0}^{q-1} \frac{\partial \Delta \tilde{y}_{i,tq-k}}{\partial \varepsilon_{tq-g}^{j}} = \sum_{k=0}^{g} \frac{\partial \Delta \tilde{y}_{i,tq+g}}{\partial \varepsilon_{tq}^{j}} \\
&= \sum_{k=0}^{g} \sum_{l=0}^{k} \left[A^{l}\right]_{ij} \rho_j^{k-l} = \sum_{l=0}^{g} \left[A^{l}\right]_{ij} \sum_{k=l}^{g} \rho_j^{k-l} \\
&= \left[A^{0}\right]_{ij} \sum_{k=0}^{g} \rho_j^{k} + \left[A^{1}\right]_{ij} \sum_{k=1}^{g} \rho_j^{k-1} + ... + \left[A^{g}\right]_{ij} \sum_{k=g}^{g} \rho_j^{k} \\
&= \left[A^{0}\right]_{ij} \sum_{k=0}^{g} \rho_j^{k} + \left[A^{1}\right]_{ij} \sum_{k=0}^{g-1} \rho_j^{k} + ... + \left[A^{g}\right]_{ij} \sum_{k=0}^{0} \rho_j^{k} \\
&= \sum_{l=0}^{g} \rho_j^{l} \sum_{r=0}^{g-l} \left[A^{r}\right]_{ij} \ .
\end{aligned}$$

As $g \to \infty$, this converges to $\frac{1}{1-\rho_j}\left[(I-A)^{-1}\right]_{ij}$ (same argument as in proof of Proposition 5).

For $h \geq 1$, we get

$$\begin{aligned}
\frac{\partial \Delta y_{i,t+h}}{\partial \varepsilon_{\tau}^{j}} &= \sum_{k=0}^{q-1} \frac{\partial \Delta \tilde{y}_{i,(t+h)q-k}}{\partial \varepsilon_{tq}^{j}} \\
&= \sum_{k=0}^{q-1} \sum_{l=0}^{hq-k} \left[A^{l}\right]_{ij} \rho_j^{hq-k-l} \\
&= \sum_{l=0}^{hq} \left[A^{l}\right]_{ij} \sum_{k=0}^{min\{q-1,hq-m\}} \rho_j^{hq-k-l} \ .
\end{aligned}$$

Under $h = 1$, this gives

$$\begin{aligned}
\frac{\partial \Delta y_{i,t+h}}{\partial \varepsilon_\tau^j} &= \left[A^0\right]_{ij} \sum_{k=0}^{q-1} \rho_j^{q-k} + \sum_{l=1}^{q} \left[A^l\right]_{ij} \sum_{k=0}^{q-l} \rho_j^{q-k-l} \\
&= \left[A^0\right]_{ij} \sum_{k=1}^{q} \rho_j^{k} + \left[A^1\right]_{ij} \sum_{k=0}^{q-1} \rho_j^{k} + ... + \left[A^q\right]_{ij} \sum_{k=0}^{0} \rho_j^{k} \\
&= \rho_j^0 \sum_{r=0}^{q} \left[A^r\right]_{ij} + \sum_{l=1}^{q} \rho_j^l \sum_{r=0}^{q-l} \left[A^r\right]_{ij} \\
&\to \left[(I-A)^{-1}A\right]_{ij} + \frac{\rho_j}{1-\rho_j} \left[(I-A)^{-1}\right]_{ij} \quad \text{as} \;\; q \to \infty \, .
\end{aligned}$$

Under $h = 2$, it yields

$$\begin{aligned}
\frac{\partial \Delta y_{i,t+h}}{\partial \varepsilon_\tau^j} &= \sum_{l=0}^{q} \left[A^l\right]_{ij} \sum_{k=0}^{q-1} \rho_j^{2q-k-l} + \sum_{l=q+1}^{2q} \left[A^l\right]_{ij} \sum_{k=0}^{2q-l} \rho_j^{2q-k-l} \\
&= \sum_{l=0}^{q} \left[A^l\right]_{ij} \sum_{k=q+1-l}^{2q-l} \rho_j^{k} + \sum_{l=q+1}^{2q} \left[A^l\right]_{ij} \sum_{k=0}^{2q-l} \rho_j^{k} \\
&= \sum_{l=0}^{q} \rho_j^l \sum_{r=q+1-l}^{2q-l} \left[A^r\right]_{ij} + \sum_{l=q+1}^{2q} \rho_j^l \sum_{r=0}^{2q-l} \left[A^r\right]_{ij} \\
&\to 0 \quad \text{as} \;\; q \to \infty \, ,
\end{aligned}$$

as expected, given that under $q \to \infty$, the long-term response materializes fully after one observational period. Analogous derivations show that, as $q \to \infty$,

$$\begin{aligned}
&\frac{\partial \Delta y_{i,t+1}}{\partial \varepsilon_{\tau-1}^j} \to \left[(I-A)^{-1}A^2\right]_{ij} + \rho_j \left[(I-A)^{-1}A\right]_{ij} + \frac{\rho_j^2}{1-\rho_j} \left[(I-A)^{-1}\right]_{ij} \, , \\
&\frac{\partial \Delta y_{i,t+2}}{\partial \varepsilon_{\tau-1}^j} \to 0 \, .
\end{aligned}$$

C.2 Data

...

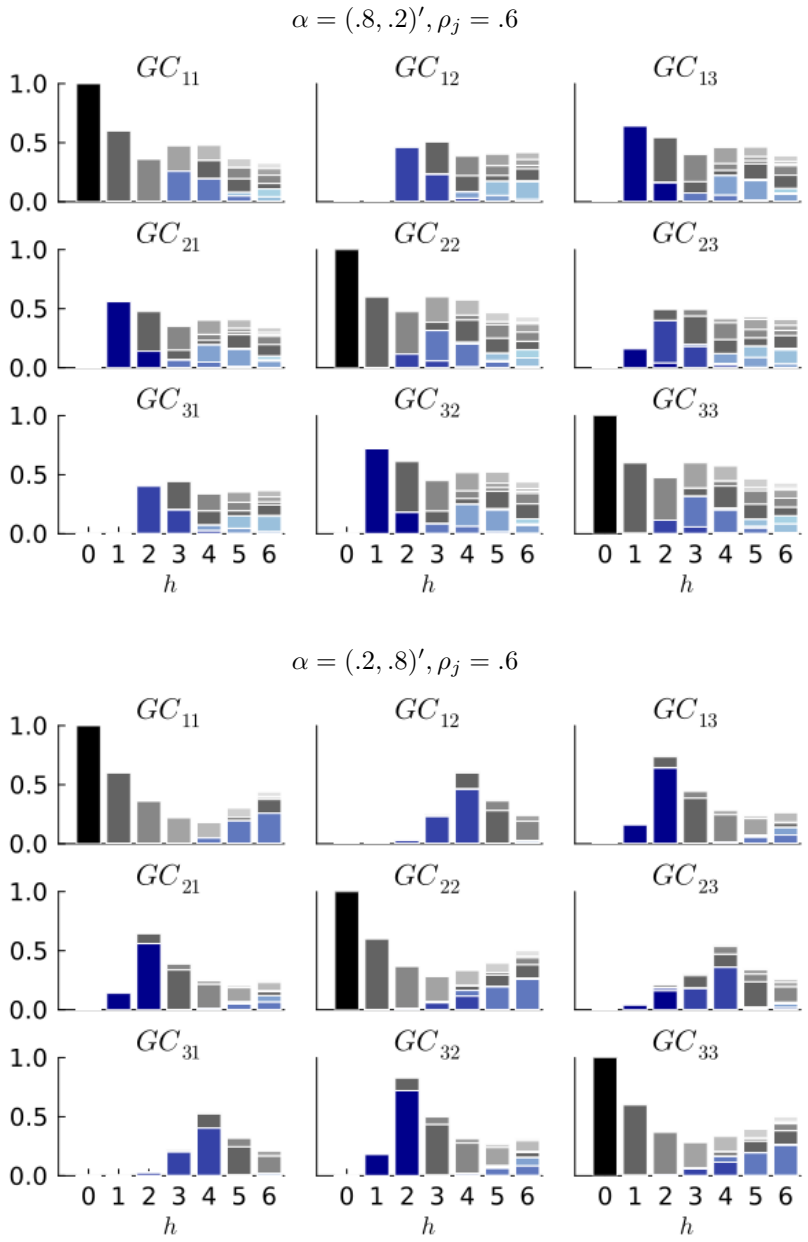


Figure A-1: Example Impulse Responses: NVAR(2, 1) with AR(1)-Shocks

Notes: The top plots refer to an NVAR(2, 1) with $\alpha = (.8, .2)'$, the bottom plots to $\alpha = (.2, .8)'$. Each panel (i, j) shows $\partial y_{i,t+h}/\partial \varepsilon_t^j$ for the example network from Section 2.1, decomposed according to the transmisson of the initial shock along different connection orders (ever lighter shades of blue) and the persistence of the response after each transmission round for $\rho_j = .6$ (ever lighter shades of gray), in line with Eq. (15) and Fig. 2.

C.3 Estimation

Likelihood Evaluation

Under $q = 1$, the model with contemporaneous IOC (Eqs. (13) and (16)) 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 $(2n+1) \times 1$, $\Psi = \big[I_n, 0\big]$, 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. (14) and (16)) 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 Appendix B).

As in the estimation of the NVAR with uncorrelated errors (see Appendix B), the Kalman filter iterations can be performed more efficiently by using the particular structure of the state space matrices to simplify the computations. Under contemporaneous IOC, partition s_t in three parts with length n, n and 1. Under lagged IOC, partition s_τ into $\tilde{p}+2$ parts, all of length n except part $\tilde{p}+2$, which is of length 1. We arrive at the following algorithms:

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

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

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.[67] 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

[67] 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.

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 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 Parsimonious Networks & Cross-Country Macroeconomic Dynamics

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

Proposition 6 (NVAR$(p,1)$-Factor Model Equivalence Result)**.**
Let y_t follow an NVAR$(p,1)$: $y_t = \alpha_1 A y_{t-1} + ... + \alpha_p A y_{t-p} + u_t$. We can write this process as

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

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

Conversely, let $y_t = \Lambda f_t + \xi_t$ with $f_t = \Phi_1 f_{t-1} + ... + \Phi_p f_{t-p} + \eta_t$ being r-dimensional, and assume $\Phi_l = \phi_l \Phi$, $l = 1 : p$. It holds that

$$y_t = \phi_1 A y_{t-1} + ... + \phi_p A y_{t-p} + u_t$$

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

Proof: For expositional simplicity, let $p = 2$. The extension to general p is straightforward.

The NVAR$(2,1)$ can be written as

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

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

$$Wy_t = W\Lambda f_t \ ,$$

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

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

with $u_t = \Lambda\eta_t + \xi_t$. If $\Phi_1 = \phi_1\Phi$ 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. ■

D.2 Hyperparameter Selection in NVAR

Marginal Posterior Mode of λ_a To obtain the marginal posterior mode (MPM) of λ_a, it is treated as a parameter and the Gibbs sampling iterations outlined in Section 3.2 and Appendix B.2 are augmented with steps drawing from the conditional posterior of λ_a.

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

$$\begin{aligned} p(\lambda_a | A, B) &\propto p(A | B, \lambda_a) \\ &\propto \lambda_a^{n^2/2} exp \left\{ -\frac{1}{2} \lambda_a \sum_{i,j=1}^n (a_{ij} - b_{ij})^2 \right\} \ , \end{aligned}$$

which shows that

$$\lambda_a \mid (A, B) \sim G \left(\frac{n^2}{2} + 1 \ , \ \frac{1}{2} \sum_{i,j=1}^n (a_{ij} - b_{ij})^2 \right) \ .$$

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 Marginal Data Density 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 Marginal Data Density (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.3 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

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

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

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

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.4 Factor Model Estimation

Consider the following 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

$$\begin{aligned} y_t &= [\Lambda, 0, ..., 0]s_t + u_t\ , & u_t &\sim N(0, I)\ , \\ s_t &= Rs_{t-1} + v_t\ , & v_t &\sim N(0, \Sigma_v)\ , \end{aligned}$$

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*

   - $\left[s_{t|t-1}\right]_1 = R_{1,\cdot}s_{t-1|t-1} \quad , \qquad \left[s_{t|t-1}\right]_{-1} = \left[s_{t-1|t-1}\right]_{-p} \ ,$
   - $\left[P_{t|t-1}\right]_{11} = R_{1,\cdot}P_{t-1|t-1}R'_{1,\cdot} + \Sigma_\eta \quad , \qquad \left[P_{t|t-1}\right]_{1,-1} = \left[P_{t|t-1}\right]'_{-1,1} \ ,$
     $\left[P_{t|t-1}\right]_{-1,1} = \left[P_{t-1|t-1}\right]_{-p,\cdot} R'_{1,\cdot} \quad , \qquad \left[P_{t|t-1}\right]_{-1,-1} = \left[P_{t-1|t-1}\right]_{-p,-p} \ .$

   (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* $\{s_{t|t}, s_{t|t-1}, P_{t|t}, P_{t|t-1}\}_{t=1}^T$.

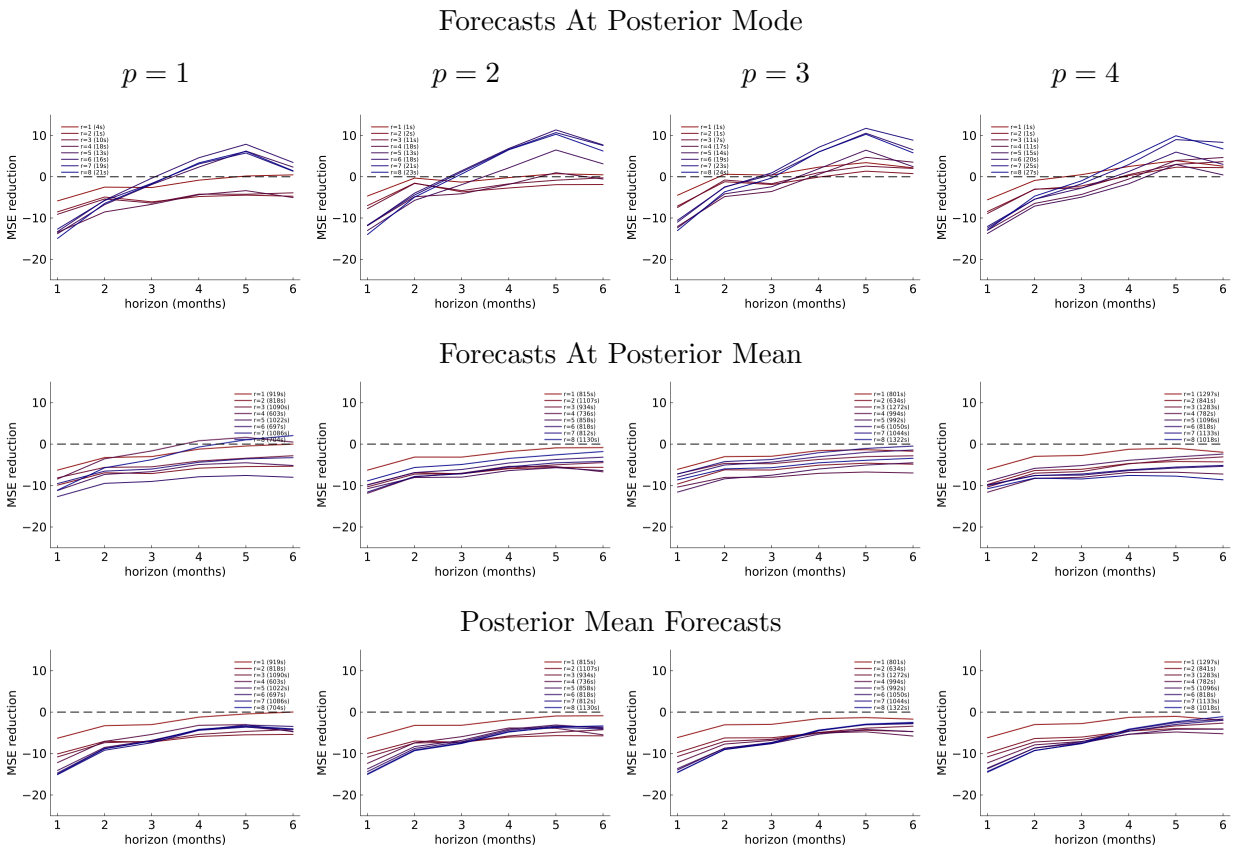


Figure A-2: Forecasting: Factor Model, Industrial Production Growth

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.

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*

   - $s_{t|t+1} = s_{t|t} + P_{t|t}R_{1.}'\left(R_{1.}P_{t|t}R_{1.}' + \Sigma\right)^{-1}\left(\left[s_{t+1}^m\right]_1 - \left[s_{t+1|t}\right]_1\right)$,
   - $P_{t|t+1} = P_{t|t} - P_{t|t}R_{1.}'\left(R_{1.}P_{t|t}R_{1.}' + \Sigma\right)^{-1}R_{1.}P_{t|t}$.

Analogous comments apply as for Algorithm 2.

D.5 Application Details

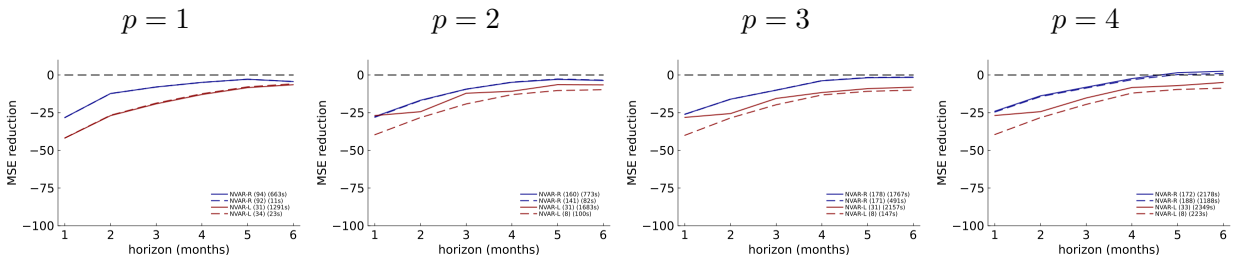


Figure A-3: Forecasting: NVAR$(p, 1)$, Industrial Production Growth

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.