The Bullwhip: Time to Build and Sectoral Fluctuations*

Yan Leng	Ernest Liu	Yifei Ren	Aleh Tsyvinski
UT Austin	*Princeton and NBER*	*Princeton*	*Yale and NBER*

August 2024

Abstract

We develop a theory of sectoral fluctuations due to amplification of demand shocks along supply chains through heterogenous time-to-build and solve the model in closed form. We show that downstream demand shocks get amplified along the sectoral supply chain and result in large fluctuations in upstream sectors creating the sectoral bullwhip effect. Empirically and quantitatively, we show that the bullwhip is significant for all downstream sectors that are important in the consumption bundle of the aggregate consumer.

*We thank Pol Antras, Markus Brunnermeier, Matt Elliott, Matt Jackson, Chad Jones, Stephen Redding, Alireza Tahbaz-Salehi, and seminar participants at Cambridge and Princeton for very helpful comments and suggestions. Ernest Liu acknowledges the support from the Sloan Foundation and the NSF.

1 Introduction

When a wrangler swiftly flicks the grip of a bullwhip, the motion travels down the length of the whip, growing stronger and faster until the tip, or cracker, moves so rapidly it sometimes breaks the sound barrier, resulting in a loud crack. We develop and solve in closed form over primitives a dynamic model of sectoral demand shock propagation along the supply chain where inputs take time to build. We show empirically and quantitatively that amplification of demand shocks through time to build along supply chains—the bullwhip—is a central feature of the economy's response to demand shocks. Metaphorically, flicks of downstream sectors demand shocks get amplified as they travel along the sectoral supply chain, resulting in large cracks in upstream sectors. The main take-away of the paper is that the bullwhip effect is present and important across sectoral supply chains for downstream consumption goods in the US economy.

Our theory of sectoral fluctuations has two key features. The first feature is sectoral demand shocks, as consumers have time-varying, stochastic preferences for final goods. The second feature is heterogeneity of time-to-build delays along the supply chain. Our main interest is in analyzing interactions between demand shocks, heterogeneous delays, and the cross-sectoral linkages of the economy. We fully characterize this model in a perfect information environment, when the history of shocks are known. We then extend the characterization to an imperfect information setting, in which firms cannot distinguish different types of shocks.

The main difficulty in characterizing the effects of sector-specific demand shocks in our heterogeneous time-to-build environment is as follows. Producers choose inputs in anticipation of future demands at all time horizons. The expectation of future demands themselves arise from both the exogenous impulse response of consumer demand shocks—at all time horizons—and endogenous direct and indirect responses to the consumer shocks by other producers—again, at all time horizons. The challenge is due to the need of each producer to consider all time horizons and endogenous actions of producers at different positions in the supply chain who themselves may face different horizon time to build delays. The complexity of the producers' endogenous response arises from the interaction between the network structure and the heterogeneity of both the time-to-build across inputs as well as of the temporal structure of sectoral demand shocks.

Our first contribution is to characterize the solution of this model for general demand shock processes over the model primitives: the demand shocks, the network structure, and the time-to-build heterogeneity. The main object of the analytical characterization is the impulse response of sectoral revenue to sectoral demand shocks affecting different parts of the supply chain. Theorem 1 shows that the impulse response of revenue is given by two sequences. The first sequence is the conditional expectations of future sectoral demands that are determined by the evolution of demand shocks process over time. The second sequence comprises of the weights on the

conditional expectations, which we derive explicitly by calculating the time to build delay across all possible paths of network linkages over a specific time horizon. These weights adjust the conditional expectations according to the time-to-build structure, resulting in different impacts on demand across sectors at various time horizons. Unlike the Leontief inverse, which accounts for all direct and indirect linkages across all time horizons and treat these horizons as being equally important, the weights in our setting differ from each other and are specific to a given time horizon due to the differences in time to build delays between suppliers and users of a particular input. Both of these sequences are derived in terms of the primitives of the economy: the demand shocks, the sectoral linkages, and the time to build structure.

We then specialize the demand shock process positing a stochastic process for sectoral demand that has two key components: a monotonically decaying transitory component, where a demand innovation today is expected to gradually diminish over time, and a persistent hump-shaped component , where an increase in demand today leads to higher demand at a future point before eventually declining. Theorem 2 explicitly determines the impact of persistent versus transitory components of demands. For the transitory component, we apply a similar approach as in the proof for Theorem 1, decomposing the impact into direct and indirect network effects at varying time horizons, where each round of the network effect is discounted by both the discount rate and the rate at which the shock dissipates. The challenge in Theorem 2 lies in deriving the effect of the hump-shaped component because the impact of a persistent shock does not decay monotonically. We show that the impact has an arithmetically increasing component as well as an exponentially decaying component, both of which we calculate explicitly.

As an example to gain intuition, consider an increase in consumer demand for automobiles. How producers respond depends on their positions along the supply chain of automobiles. Downstream producers—those that are close in distance to the shock, such as the producers of tires or glass windows—increase production in response to higher expected demand in the near future. By contrast, because of time-to-build and production delays, producers further upstream—those of semi-conductors or steel alloy, which are not directly used to assemble cars but only indirectly as components for automobile parts—increase production only in response to demand shocks that are anticipated to persist. For producers further distant from the demand shock, their production response depends on the anticipated demand at later times; that is, the more distant time horizon matters. Persistent demand shocks create hump-shaped demand profile over time, meaning an increase in demand for automobiles today implies that demand would be even higher at some future times. This behavior creates the bullwhip effect, as upstream producers in anticipation of future demand respond more than downstream producers that respond to demand today or in the near future. By contrast, transitory demand shocks have time profiles that decay monotonically, implying that the producers closer to the shocks are always more responsive than distant

producers.

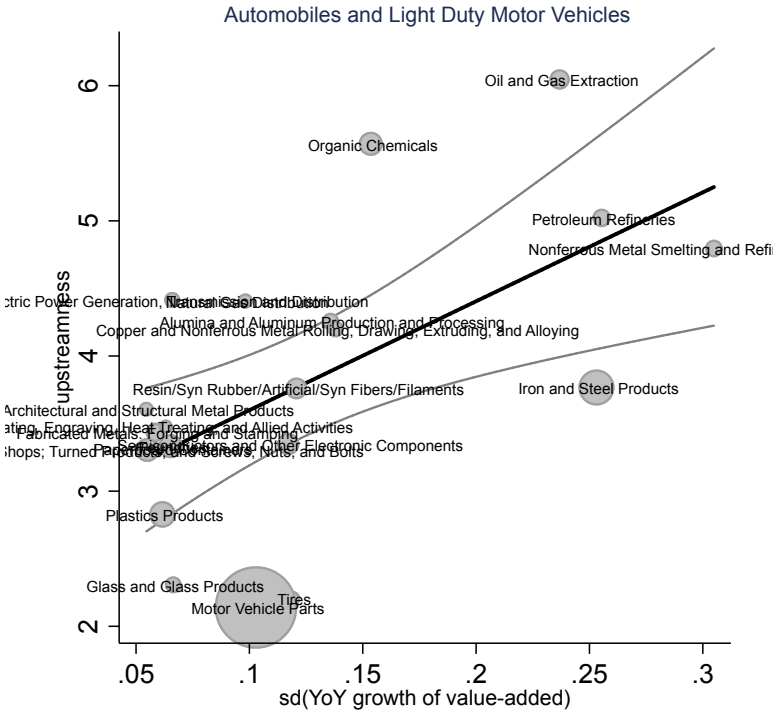
The position of a producer in the supply chain and the distance between producers play an important role in our analysis. A key feature of our model is the notion of delay—how long it takes for a given sector value-added to reach a specific user of its input along the supply chain. With time-to-build, the distance from an input supplier to a producer translates to the relative delay of producing goods between producers. We show that the ratio between a producer's revenue response to a persistent demand shock downstream and the response to a transitory shock is the proper notion of supply chain distance. This ratio captures the weighted average of delays over all possible direct and indirect routes between the input supplier and input user, where the weights are given by the share of the source suppliers output sent along each route. The challenge is to characterize the weights which are themselves endogenous. Our closed-form solution of the impulse responses of producers actions to the persistent shocks allows us to determine these endogenous weights and thus the distance measure in closed form over the primitives of the economy, which we then measure in the data.

We use the theoretical characterization to empirically document existence and importance of the bullwhip effect across US sectors. We show that the bullwhip is significant for all of the downstream sectors that are important in the consumption bundle of the aggregate consumer. As an example, we plot the bullwhip effect for automobiles in Figure 1. Each circle is a sector along the supply chain of automobiles. The Y-axis is the our measure of supply chain distance from each supplier to automobiles; the X-axis the standard deviation in the year-on-year value-added growth, a measure of sectoral volatility. The bullwhip effect manifests in this figure as more upstream sectors in the automobile supply chain have higher volatility. In the paper we establish that this pattern holds across supply chains for all important goods in the consumption bundle.

We extend the model to the environment of imperfect information, where both heterogeneous delays and learning are present. Specifically, we consider an environment in which producers observe only the realized consumer demand and cannot distinguish whether the shock is transitory or persistent. The result augments the analytical characterization of the impulse responses in Theorem 3 with the additional filtering problem faced by the suppliers due to imperfect information. The most interesting aspect of the result is the sectoral response to a sequence of positive or negative demand shocks such as the momentum of the demand shocks. Empirically, we show that the sectoral response to demand shocks depends on the momentum of demand shocks, especially in upstream sectors. That is, upstream value-added is especially responsive to downstream demand shocks when shocks arrive in a sequence with the same sign, after conditioning on the magnitude of the sum.

We then empirically document the time-path of the bullwhip effect along the supply chains

Figure 1. Upstreamness and volatility for the supply chain of automobiles



through the impulse-response of upstream value-added following downstream demand shocks. Specifically, we extract demand shocks to downstream sectors, and analyze how upstream value-added responds at different time horizons. First, we find that innovations to downstream demand create a hump-shaped demand profile in downstream sectors, that is a consequence of demand shock persistence. The peak occurs two quarters after the shock and amplifies the initial shock by about 30%. Half of the shock's initial dissipates after 8 quarters. Second, we find that demand shocks create hump-shaped responses in upstream value-added that gets amplified in sectors higher up along the supply chain. Finally, we construct a measure of momentum in demand shocks, and we show that the supply chain response to demand shocks depend crucially on the momentum, especially in upstream sectors. We conclude that the bullwhip effect is present in the economy, when we pool the bullwhip effects of all sectors together, as well as document significant heterogeneity across the supply chain of different sectors.

Our model builds on two canonical models: Kydland and Prescott (1982) time-to-build environment as the theory of aggregate fluctuations, and Long Jr and Plosser (1983) neoclassical sectoral model where all inputs take exactly one period to build with either aggregate or sectoral productivity shocks. Methodologically, we develop and solve in closed form the model of sectoral demand shocks, incorporating heterogenous time to build frictions. Substantively, our main

interest is in how sectoral demand shocks propagate and are amplified throughout the supply chain, creating the bullwhip effect. This approach thus provides a theory of sectoral fluctuations driven by demand shocks and heterogeneous time-to-build.

Our paper is also inspired by a substantial body of literature on the bullwhip effect in the fields of industrial engineering and operations management (Forrester, 1958, Lee, Padmanabhan and Whang, 1997, 2004). This literature primarily concentrates on the inventory management practices of individual firms, with a body of empirical work consisting of case studies based on firms' different positions within a supply chain. A key theme of this literature is on individual firms' decisions and on how frictions that these firms face – behavioral or managerial biases, coordination failures, forecast errors, or different processes for formation expectations – result in amplification of orders placed.[1] By contrast, we develop and solve an equilibrium theory of sectoral fluctuations, where producers in all sectors, interconnected through the input-output relationships, optimally react to demand shocks subject to technological time to build constraints. We thus show that the bullwhip effect arises as an equilibrium phenomenon in the absence of behavioral, coordination, or other frictions.

2 Model

In this section we describe the economic environment and characterize the equilibrium. The model builds on Long Jr and Plosser (1983) and is a discrete-time dynamic production network featuring time-to-build, i.e., intermediate inputs are required to be supplied in advance. The main departures from Long Jr and Plosser (1983) is the introduction and analysis of sectoral demand shocks and the heterogeneous time-to-build across inputs. For a continuous-time version of the model focusing on productivity shocks, see Liu and Tsyvinski (2024).

2.1 Economic Environment

Preferences There is a representative consumer with preferences

$$V\left(\mathcal{S}_t\right)=\mathbb{E}\left[\sum_{s=t}^{\infty}\beta^{s-t}\left(\sum_{i=1}^{N}\theta_{is}\ln c_{is}-v\left(\bar{\ell}_s\right)\right)|\mathcal{S}_t\right] \tag{1}$$

where $\mathcal{S}_t$ is the set of state variables at the beginning of time t, c_{it} is the consumption of good i, $\{\theta_{it}\}_{\geq 0}$ is the set of demand shocks, and $\bar{\ell}_t$ is total labor supplied.

[1] Its simplest manifestation, often taught in management courses in business schools, is the beer game. MBA students manage orders and inventory for one of four firms: manufacturer, distributors, wholesaler, and retailer. The bullwhip effect arises even in this stylized board game and has been widely documented in a variety of business case studies.

Technology There are N sectors, each constant-returns-to-scale production that uses labor and intermediate goods from other sectors. There is heterogeneous time-to-build across inputs: each input j must be supplied $d_{ij} \in \mathbb{Z}_{>0}$ periods before production takes place in sector i. The time-to-build is supplier-user specific and can be equivalently interpreted either as the shipping delay of sending input j to producer i or the delay during the production of good i. We assume that the input delays are finite $d_{ij} < \infty$.

The production function of sector i is

$$y_{it} = z_{it}\ell_{it}^{\alpha_i}\prod_{j=1}^{N} m_{ij,t-d_{ij}}^{\omega_{ij}}, \qquad \alpha_i + \sum_j \omega_{ij} = 1, \tag{2}$$

where z_{it} is sectoral productivity, ℓ_{it} is the labor input, and $m_{ij,t-d_{ij}}$ is the intermediate good j used in the production of the good i. The elasticities of intermediate inputs $\{\omega_{ij}\}$ define the production network $\mathbf{\Omega} \equiv [\omega_{ij}]'$, where each column of $\mathbf{\Omega}$ is an input user and each row is an input supplier ($'$ denotes the transpose). We assume that $(\boldsymbol{I} - \mathbf{\Omega})^{-1}$ is invertible, which is ensured when the production of any good eventually involves labor along the supply chain (i.e., for any i, $\sum_i \alpha_j \left[\mathbf{\Omega}^k\right]_{ji} > 0$ some k).

Market Clearing The goods and labor market clearing conditions follow

$$y_{jt} = c_{jt} + \sum_{i=1}^{N} m_{ijt}, \qquad \bar{\ell}_t = \sum_i \ell_{it}. \tag{3}$$

Sectoral Demand and Productivity Shocks We assume the economy is buffeted by sectoral demand shocks and θ_{it} fluctuates from period to period. These sectoral demand shocks are the focus of this study, as we elaborate below.

For expositional simplicity, we abstract away from productivity shocks and assume $z_{it} \equiv z_i$ is time-invariant. As we show below, while stochastic productivity shocks affect output quantities, they do not play an important role in how resources are allocated in terms of shares, e.g., what fraction of workers are employed in each sector, and what fraction of each good j is sent to each producer i using input j. Our analysis will be entirely based on allocation shares; hence, our conclusions continue to hold even when productivities are random.

State Variables State variables at each time t include (1) the set of intermediate inputs $\{m_{ij,t-q}\}_{i,j=1}^{N}$ ordered prior to t that have yet to be used in production ($q = 1, \ldots, d_{ij}$), (2) the realization of sectoral demand $\{\theta_{it}\}$, and (3) any additional observed variables relevant for the expectation of future demand. We assume that endogenous variables at time t are chosen after the current-period demand shocks are realized.

2.2 Equilibrium

To analyze the competitive equilibrium, it is actually more transparent to first analyze the social planner's problem. Because there are no distortions or externalities in the model—time-to-build is a technological feature of the economic environment and is also part of the social planner's problem—the competitive equilibrium is efficient: allocations are identical to the planner's solution, and equilibrium prices correspond to the Lagrange multipliers of the planner's problem up to a choice of numeraire. The concordance between the planner's Lagrange multipliers and equilibrium prices enables us to interpret the planner's solution as equilibrium objects involving prices, such as GDP, value-added, sales, and the Domar weight (i.e., sectoral size measured as the sales-to-GDP ratio).

The planner solves

$$\begin{aligned} V_t\left(\{m_{ij,t-q}\}_{q=1,\ldots,d_{ij}}\right) \quad = \quad & \max_{\{\ell_{it},c_{it},m_{ijt},\bar{\ell}_t\}} \sum_i \theta_{it} \ln c_{it} - v\left(\bar{\ell}_t\right) + \beta \mathbb{E}_t\left[V_{t+1}\left(\{m_{ij,t+1-q}\}_{q=1,\ldots,d_{ij}}\right)\right] \\ & + w_t\left[\bar{\ell}_t - \sum_j \ell_{jt}\right] + \sum_j p_{jt}\left[z_j \ell_{jt}^{\alpha_j} \prod_k m_{jk,t-d_{jk}}^{\omega_{jk}} - c_{jt} - \sum_i m_{ijt}\right], \end{aligned}$$

where w_t is the Lagrange multiplier for the labor market clearing constraint, and p_{jt} is the Lagrange multiplier for the market clearing constraint of output from sector j.

The first-order conditions with respect to c_{jt} and ℓ_{jt} imply

$$\{c_{jt}\} \qquad \theta_{jt} = p_{jt} c_{jt} \tag{4}$$

$$\{\ell_{jt}\} \qquad w_t \ell_{jt} = \alpha_j p_{jt} y_{jt} \tag{5}$$

whereas the envelope condition and the first-order condition with respect to $m_{jk,t-d_{jk}}$ jointly imply

$$\{m_{jkt}\} \qquad p_{kt} m_{jkt} = \omega_{jk} \beta^{d_{jk}} \mathbb{E}_t\left[p_{j,t+d_{jk}} y_{j,t+d_{jk}}\right] \tag{6}$$

As we show in Section 2.2.1 below, the Lagrange multipliers correspond to prices in a decentralized competitive equilibrium when the numeraire is chosen such that the marginal utility of income in each period is normalized to one. We henceforth refer to p_{jt} as the price of good j and w_t as the wage rate.

Equation (4) shows that the total expenditure on consumption of good j is equal to θ_{jt}, the concurrent demand shock for that good. Equation (5) shows that each producer's expenditure on labor inputs is α_j fraction of the revenue, where α_j is the value-added share in the sector. Equation (6) shows that $p_{kt} m_{jk,t}$, the expenditure on each input k by producer j at time t, is ω_{jk} fraction of the expected revenue by the producer in the future period $t + d_{jk}$, discounted at rate β per period. The expectation operator in (6) encapsulates the measurability constraint in the planner's solution.

We build on equations (4)–(6) and the market clearing condition (3) for each sector's output. Note, in particular, that we do not rely on the optimality condition for total labor supplied ($\bar{\ell}_t$). In fact, even though the shape of the labor disutility function $v\left(\bar{\ell}_t\right)$ affects the total labor supplied in each period and thus the *levels* of output, it has no impact on how resources are allocated in terms of shares (e.g., $\ell_{it}/\bar{\ell}_t$, c_{it}/y_{it}, and m_{ijt}/y_{it}), which is what we characterize. Our analysis below thus holds true for generic convex labor supply functions $v\left(\bar{\ell}_t\right)$, such as when the disutility of labor is linear ($v\left(\bar{\ell}_t\right) = \bar{\ell}_t$) and when labor is exogenously endowed.

Let $\gamma_{jt} \equiv p_{jt}y_{jt}$ denote the sector j's revenue at time t. This is a key object for our analysis, as we can express the equilibrium allocation in shares (i.e., the fraction of workers in each sector $\ell_{it}/\bar{\ell}_t$, the share of each sector j's output used for consumption c_{jt}/y_{jt} and as production inputs m_{ijt}/y_{jt}) all as functions of the demand shocks and the current and expected future revenues:

$$\frac{\ell_{jt}}{\bar{\ell}_t} = \frac{\alpha_j\gamma_{jt}}{\sum_i \alpha_i\gamma_{it}}, \qquad \frac{c_{jt}}{y_{jt}} = \frac{\theta_{jt}}{\gamma_{jt}}, \qquad \frac{m_{ijt}}{y_{jt}} = \omega_{jk}\beta\frac{\mathbb{E}_t\left[\gamma_{jt+1}\right]}{\gamma_{jt}}.$$

The GDP, i.e., the total value-added in the economy, is $\sum_i \alpha_i\gamma_{it}$. The Domar weight, defined as the sectoral sales-to-GDP ratio, is

$$\text{(Domar weight)} \quad m_{it} \equiv \frac{p_{it}y_{it}}{\sum_i w_t\ell_{it}} = \frac{\gamma_{it}}{\sum_i \alpha_i\gamma_{it}}.$$

We now solve for sectoral sales $\{\gamma_{it}\}$ as a function of expected future demand. First multiply the goods market clearing condition (3) by p_{jt} on both sides and then substitute out c_{jt} and m_{ijt} using (4) and (6), we obtain

$$\begin{aligned}\gamma_{jt} &= \theta_{jt} + \sum_{i=1}^{N} \beta^{d_{ij}}\omega_{ij}\mathbb{E}_t\left[\gamma_{i,t+d_{ij}}\right] \\ &= \theta_{jt} + \sum_{d=1}^{\infty}\sum_{i=1}^{N} \beta^{d}\omega_{ij}\mathbf{1}_{d_{ij}=d}\mathbb{E}_t\left[\gamma_{i,t+d}\right]\end{aligned}$$

In words, the revenue of sector j at time t consists of the concurrent time-t consumer demand for good j and the demand from other producers. Because of the heterogeneous time-to-build input delays by d_{ij} periods, the expenditure on input j at time t by producer i is equal to the input cost share ω_{ij} times the expected revenue at time $t + d_{ij}$, discounted by $\beta^{d_{ij}}$.

To proceed, we re-write the sectoral revenue equation in vector form. Throughout the paper, boldface letters represent matrices (uppercase) and column vectors (lowercase). Let the $N \times N$ matrix $\mathbf{\Omega}_d \equiv \left[\omega_{ij}\mathbf{1}_{d_{ij}=d}\right]'$ denote the input cost shares for inputs with delay d, and let $\boldsymbol{\theta}_t \equiv [\theta_{it}]$ be the column vector of demand shocks. The collection of matrices $\{\mathbf{\Omega}_d\}_{d=1}^{\infty}$ defines the time-to-build structure over the production network. Note that each column of $\mathbf{\Omega}_d$ represents an input user and each row represents an input supplier.

Using the matrix notation, the vector of revenue can be written as

$$\boldsymbol{\gamma}_t = \boldsymbol{\theta}_t + \sum_{d=1}^{\infty} \beta^d \boldsymbol{\Omega}_d \mathbb{E}_t \left[\boldsymbol{\gamma}_{t+d}\right]$$

By iteratively substituting out future revenue, we obtain the follow result, which solves for the sectoral revenue at time t as a function of the expected future demand across sectors.

Theorem 1. *Let Φ_s denote the set of positive, finite sequences $\phi \equiv (\phi_1, \ldots, \phi_n)$ that sum to s (i.e., $\sum_{k=1}^{n} \phi_k = s$ with $\phi_k > 0$ for all k). Sectoral revenue follows*

$$\boldsymbol{\gamma}_t = \boldsymbol{\theta}_t + \sum_{s=1}^{\infty} \boldsymbol{G}_s \mathbb{E}_t \left[\boldsymbol{\theta}_{t+s}\right], \qquad \text{where } \boldsymbol{G}_s \equiv \beta^s \sum_{\phi \in \Phi_s} \prod_{\phi_j \in \phi} \boldsymbol{\Omega}_{\phi_j}.$$

Theorem 1 characterizes how expected future demand at each time horizon influences the sectoral revenue at time t through the production network. To interpret, note that $\boldsymbol{G}_1 = \beta \boldsymbol{\Omega}_1$ and $\boldsymbol{G}_2 = \beta^2 \left(\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_1^2\right)$.[2] While the concurrent consumer demand $\boldsymbol{\theta}_t$ contributes to a sector's revenue directly, the one-period ahead demand $\boldsymbol{\theta}_{t+1}$ contributes to a sector's revenue only indirectly through other producers to which the focal sector supplies with one period delay (captured by the matrix $\boldsymbol{\Omega}_1$ in $\boldsymbol{G}_1$). The contribution of demand two-period ahead ($\boldsymbol{\theta}_{t+2}$) to a focal sector i's revenue (captured by $\boldsymbol{G}_2$) is through both (a) producers to which the focal sector supplies with a delay of two periods (captured the matrix $\boldsymbol{\Omega}_2$), and (b) any producer k that uses an input j with one period delay, and where sector j uses input i also with one period delay (captured by the matrix $\boldsymbol{\Omega}_1^2$). The total contribution of future consumer demand for good k at time $t+s$ on sector i's revenue at time t is thus the cumulation all possible walks of direct and indirect input-output linkages that delivers output of i to k, accounting for the time-to-build associated with each walk (for instance, $\boldsymbol{G}_3 = \beta^3 \left(\boldsymbol{\Omega}_3 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_1^3\right)$).

Implication: Steady-State with Heterogeneous Time-to-Build Applying Theorem 1 to the steady-state, where $\boldsymbol{\theta}_t = \bar{\boldsymbol{\theta}}$ is time-invariant, we can simplify the summation over $\boldsymbol{G}_s$ and obtain the following result.

Proposition 1. *In steady-state, sectoral revenue follows*

$$\bar{\boldsymbol{\gamma}} = \left(\boldsymbol{I} - \sum_{d=1}^{\infty} \beta^d \boldsymbol{\Omega}_d\right)^{-1} \bar{\boldsymbol{\theta}}.$$

We can contrast this result with the literature. When all input delays takes exactly one period, we obtain as a special case the steady-state result of Long Jr and Plosser (1983): $\bar{\boldsymbol{\gamma}} =$

[2] The only positive sequence that sums to 1 is (1). there are two sequences that sum to 2: $(1,1)$ and (2).

$(\boldsymbol{I} - \beta\boldsymbol{\Omega})^{-1}\bar{\boldsymbol{\theta}}$, where $\boldsymbol{\Omega} \equiv [\omega_{ji}]$. In the static model of Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012), sectoral revenue is $\bar{\boldsymbol{\gamma}} = (\boldsymbol{I} - \boldsymbol{\Omega})^{-1}\bar{\boldsymbol{\theta}}$.[3]

We provide the proof of Proposition 1, as it is informative for subsequent results.

Proof. Let $\boldsymbol{G}_\infty \equiv \boldsymbol{I} + \sum_{s=1}^{\infty}\boldsymbol{G}_s$. To show $\boldsymbol{G}_\infty = \left(\boldsymbol{I} - \sum_{d=1}^{\infty}\beta^d\boldsymbol{\Omega}_d\right)^{-1}$, note

$$\boldsymbol{G}_\infty = \boldsymbol{I} + \beta\boldsymbol{\Omega}_1\boldsymbol{G}_\infty + \beta^2\boldsymbol{\Omega}_2\boldsymbol{G}_\infty + \cdots = \boldsymbol{I} + \left(\sum_{d=0}^{\infty}\beta^d\boldsymbol{\Omega}_d\right)\boldsymbol{G}_\infty.$$

Rearranging yields the result. Note that the invertibility of $\left(\boldsymbol{I} - \sum_{d=0}^{\infty}\beta^d\boldsymbol{\Omega}_d\right)$ follows from the invertibility of $\boldsymbol{I} - \boldsymbol{\Omega}$ and that for all entries, $\left[\sum_{d=0}^{\infty}\beta^d\boldsymbol{\Omega}_d\right]_{ji} \leq \boldsymbol{\Omega}_{ji}$. □

Intuitively, the proof utilizes the recursive relation that the contribution of steady-state demand on revenue in a focal sector j ($\boldsymbol{G}_\infty$) can be decomposed into the direct ($\boldsymbol{I}$) and indirect effects through each sector i that j supplies to, taking into account the time-to-build between the j and i. The recursion lies in the fact that the contribution of demand on each sector i is also captured by $\boldsymbol{G}_\infty$.

Implication: When Sectoral Demand Follows $ARMA\,(p,q)$ Theorem 1 hows that the equilibrium revenue in sector i at time t reflects the expected future demand $\mathbb{E}_t\left[\theta_{k,t+s}\right]$ at each time horizon in every other sector k, through each possible walks of direct and indirect input-output linkages that delivers output of i to k, accounting for the time-to-build associated with each walk.

We now provides a characterization of equilibrium revenue when sectoral demand follows a stationary and invertible $ARMA\,(p,q)$. The process determines the expected future demand $\mathbb{E}_t\left[\theta_{k,t+s}\right]$ under rational expectations. The analysis reveals an interesting parallel between (1) how a current demand shock affects the expected future demand through the autoregression of the demand shock process and (2) how the time-to-build and the network structure determines the importance of demand at different horizons.

Specifically, we work with the $AR\,(\infty)$ representation of the $ARMA\,(p,q)$ process[4]. Suppose demand follows

$$\theta_{it} = \sum_{s=1}^{\infty}\delta_s\theta_{it-s} + e_{it}. \tag{7}$$

where $\boldsymbol{e}_t$ is mean zero and mean-independent of past shocks, and $\{\delta_s\}$ are the autoregressive coefficients.

[3]For expositional simplicity, our baseline model features a time-to-build of at least one period. We could extend the baseline model to incorporate static inputs without time-to-build (i.e., $d_{ij} = 0$), thereby nesting the static production network model as a special case. In that case, Proposition 1 extends naturally to $\bar{\boldsymbol{\gamma}} = \left(\boldsymbol{I} - \sum_{d=0}^{\infty}\beta^d\boldsymbol{\Omega}_d\right)^{-1}\bar{\boldsymbol{\theta}}$.

[4]The representation $AR\,(\infty)$ exists for all stationary and invertible $ARMA\,(p,q)$ processes (see, e.g., Box, Jenkins, Reinsel and Ljung, 2015).

In this setting, the impulse-response of expected future demand to current demand innovation is

$$\frac{\partial \mathbb{E}_t\left[\theta_{i,t+s}\right]}{\partial e_{it}} = \sum_{\phi \in \Phi_s} \prod_{\phi_j \in \phi} \delta_{\phi_j}, \tag{8}$$

where, as in Theorem 1, Φ_s is the set of positive, finite sequences $\phi \equiv (\phi_1, \ldots, \phi_n)$ that sum to s. That is, e_{it}'s impact on $\mathbb{E}_t\left[\theta_{i,t+1}\right]$ is δ_1, its impact on $\mathbb{E}_t\left[\theta_{i,t+2}\right]$ is $\delta_1^2 + \delta_2$, and on $\mathbb{E}_t\left[\theta_{i,t+3}\right]$ is $(\delta_1^3 + \delta_2\delta_1 + \delta_1\delta_2 + \delta_3)$, and so on. In words, the impact of a demand shock e_{it} on the expected future demand at time $t+s$ has a recursive structure: it is the sum of its impact on demand $\mathbb{E}_t\left[\theta_{i,t+k}\right]$ for all $0 \leq k \leq s$, and then the impact of $\mathbb{E}_t\left[\theta_{i,t+k}\right]$ on $\mathbb{E}_t\left[\theta_{i,t+s}\right]$ through the autoregressive demand process.

Note in particular the parallel between equation (8) and the object $\boldsymbol{G}_s$ in Theorem 1. Both constructs enumerates over all time gaps that sum to s. In $\boldsymbol{G}_s$, the gap is due to the time-to-build delays; in equation (8), the gap is due to the autoregressive structure of demand.

Proposition 2. *When demand follows (7), the revenue response to a demand shock is*

$$\frac{\partial \boldsymbol{\gamma}_t}{\partial \boldsymbol{e}_t} = \boldsymbol{I} + \sum_{s=1}^{\infty} \beta^s \left(\sum_{\phi \in \Phi_s} \prod_{\phi_j \in \phi} \delta_{\phi_j} \right) \left(\sum_{\phi \in \Phi_s} \prod_{\phi_j \in \phi} \boldsymbol{\Omega}_{\phi_j} \right),$$

where Φ_s is the set of positive, finite sequences $\phi \equiv (\phi_1, \ldots, \phi_n)$ that sum to s.

2.2.1 Decentralizing the Planner's Solution as the Competitive Equilibrium

We now setup the competitive equilibrium to decentralize the planner's solution, and we show that the Lagrange multipliers in the planner's solution indeed correspond to prices. Readers uninterested in this mapping can directly skip to Section 3, where we analyze how demand shocks propagate along supply chains over time and interact with the heterogenous time-to-build.

Consumer Problem At each time t, the consumer solves

$$W_t\left(a_t\right) = \max_{\left\{c_{it}, \bar{\ell}_t\right\}} \sum_i \theta_{it} \ln c_{it} - v\left(\bar{\ell}_t\right) + \beta \mathbb{E}_t\left[W_{t+1}\left(a_{t+1}\right)\right]$$

subject to the budget constraint

$$\sum_i p_{it} c_{it} + a_{t+1}/R_t = w_t \bar{\ell}_t + a_t + \Pi_t,$$

where R_t is the interest rate, a_t captures savings in terms of a numeraire bond (in zero net supply), p_{it} is the price of good i, w_t is the wage rate, and Π_t is the firm profits. We choose the numeraire so that the marginal utility of income in any period t (i.e., the Lagrange multiplier on the budget constraint in each period) is always equal to one. Given the normalization, the intra-temporal

Euler equation implies

$$\theta_{it} = p_{it} c_{it},$$

and the inter-temporal Euler equation implies

$$1 = \beta R_t.$$

Producer Problem The representative producer in sector i solves

$$\max_{m_{ijt}} \mathbb{E}_t \left[\sum_s \Lambda_{t+s} \max_{\ell_{it+s}} \left(p_{it+s} y_{it+s} - w_{t+s} \ell_{it+s} \right) - p_{jt} m_{ijt} \right]$$

where Λ_{t+1} is the stochastic discount factor. Note that given log-linear utility, the stochastic discount factor is equal to β in all states—across all realizations of the productivity or demand shocks.

Planner's Lagrange Multipliers By comparing the equilibrium conditions with the planner's solution, it is evident that the planner's Lagrange multipliers coincide with equilibrium prices.

Remark on the Choice of Numeraire and Our Characterization The object $\boldsymbol{\gamma}_t$ in the planner's solution coincides with the vector of sectoral revenues in the decentralized equilibrium with the marginal utility of income normalized to one in each period. This normalization is equivalent to choosing a numeraire in each period and is therefore inconsequential for analyzing the relative revenue across sectors at each time (i.e., γ_{it}/γ_{jt}). Theorem 1 further implies that the relative sectoral revenue is invariant to the labor supply function $v(\cdot)$ and sectoral productivities. The labor supply function and sectoral productivities do affect quantities, but they do not affect relative revenue across sectors. This property significantly simplifies our analysis. The property has been known to hold in static production network models with log-linear preferences and production functions (Liu and Tsyvinski 2024) as well as in dynamic production network models with log-linear adjustment costs (Liu and Tsyvinski 2024). Our analysis here establishes the property and exploits it in a dynamic production network model with heterogeneous time-to-build.

3 The Supply Chain Response to Demand Shocks

In this section we analyze the equilibrium response to demand shocks along the supply chain. We begin by applying our earlier characterization to a parsimonious stochastic process for sectoral demand that has two components: a monotonically decaying AR(1) component and a hump-shaped AR(2) component. The stochastic process defines the conditional expectation over future demand based on current information $\mathbb{E}_t[\boldsymbol{\theta}_{t+s}]$, and we utilize Theorem 1 to analyze how different

sectoral demand shocks and their implied paths of expected future demand interact with the heterogeneous time-to-build and shape the equilibrium response along the supply chain. Section 3.1 describes the demand shock process and derives the impulse responses. Section 3.2 discusses several implications.

In section 3.3 we describe several extensions. Our baseline specification assumes the monotone-decay and the hump-shaped demand shocks are governed by the same rate of convergence to the steady-state, and that economic agents can perfectly distinguish between these two components. In section 3.3.1 we extend the analysis to a setting where the two components have heterogeneous rates of convergence. In section 3.3.2 we solve for the impulse response under imperfect information in a vertical supply chain, where only the past history of realized demand is observed and agents have to learn about the type of demand shocks.

3.1 Two Components of Demand Shocks

We assume sectoral demand is mean-reverting and has two components: a monotonically decaying component and a hump-shaped component. Specifically, sectoral demand θ_{it} follows

$$\theta_{it} - \bar{\theta}_i = \rho\left(\theta_{it-1} - \bar{\theta}_i\right) + x_{it} + \epsilon_{it}, \tag{9}$$

$$x_{it} = \rho x_{i,t-1} + u_{it}. \tag{10}$$

The parameter $\rho < 1$ captures the rate at which sectoral demand mean-reverts to the steady-state or long-run level $\bar{\theta}_i$. There are two sources of demand shocks: ϵ_{it} and u_{it}, both are mean zero and mean independent over-time. These shocks may be correlated across sectors, and we allow their variances to depend on past shocks in order to ensure that sectoral demand stays non-negative. Note that in our baseline specification the two types of demand shocks share the same persistence parameter ρ. In section 3.3.1 below we extend our analysis with heterogeneous ρ's.

The monotonically decaying, or transitory, component of demand shocks is captured by ϵ_{it}, and the hump-shaped, or persistent, component is captured by u_{it}. For expositional simplicity, we refer to ϵ_{it} as transitory shocks and u_{it} as persistent shocks. In the absence of the persistent shock u_{it}, the transitory shock ϵ_{it} is AR(1). In the absence of the transitory shock ϵ_{it}, the persistent shock u_{it} is AR(2).[5]

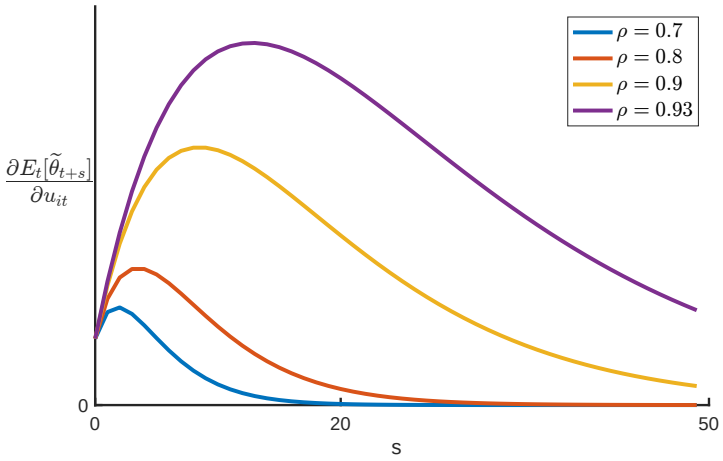
Consider the response of expected demand $\mathbb{E}_t\left[\theta_{i,t+s}\right]$ to an impulse (ϵ_{it}, u_{it}), starting from steady-state at $t-1$ (i.e., $(\theta_{i,t-1}, x_{i,t-1}) = \left(\bar{\theta}_i, 0\right)$). Then

$$\mathbb{E}_t\left[\theta_{i,t+s} - \bar{\theta}_i\right] = \underbrace{\rho^s \epsilon_{it}}_{\substack{\text{response to a} \\ \text{temporary shock}}} + \underbrace{(s+1)\,\rho^s u_{it}}_{\substack{\text{response to a} \\ \text{persistent shock}}} \quad \text{for all } s \geq 0. \tag{11}$$

[5]Specifically, in the absence of ϵ_{it}, demand follows $\theta_{it} - \bar{\theta}_i = 2\rho\left(\theta_{i,t-1} - \bar{\theta}_i\right) - \rho^2\left(\theta_{i,t-2} - \bar{\theta}_i\right) + u_{it}$.

The impulse-response to a transitory shock ϵ_{it} monotonically declines over time at rate ρ. By contrast, for any $\rho > 0.5$, the impulse-response to a persistent u_{it} is hump-shaped: $(s+1)\rho^s u_{it}$ increases over time s for $s \leq \bar{s} \equiv \frac{2\rho-1}{1-\rho}$, and decreases for time $s > \bar{s}$. The impulse-response $\mathbb{E}_t[\theta_{i,t+s}]$ to a shock $(\epsilon_{it}, u_{it}) = (0, u)$ for different levels of persistence ρ are shown in Figure 2, with the initial impact u scaled such that the sum of the response over time ($\sum_{s=0}^{\infty} \mathbb{E}_t[\theta_{i,t+s}]$) is constant across different ρ's. As the figure shows, the impulse-response is hump-shaped, and under a higher persistence parameter ρ, the effects last longer, with the peak realized later in time.

Figure 2. The impulse response $\partial \mathbb{E}_t[\theta_{i,t+s}] \Big/ \partial u_{it}$ of future demand following a persistent shock, for different levels of ρ



Impulse Response under Perfect Information We begin by assuming agents can separately observe the persistent and transitory components of demand. The conditional expectation is thus defined as $\mathbb{E}_t[\cdot] \equiv \left[\cdot \middle| \{\boldsymbol{\theta}_{t-s}, \boldsymbol{x}_{t-s}\}_{s\geq 0}\right]$. Given the stochastic process of demand shocks, the information set consisting of the history of state variables $\{\boldsymbol{\theta}_{t-s}, \boldsymbol{x}_{t-s}\}_{s\geq 0}$ is identical to the information set consisting of the history of shocks $\{\boldsymbol{\epsilon}_{t-s}, \boldsymbol{u}_{t-s}\}_{s\geq 0}$, and we use the two information sets interchangeably.

We let $\widetilde{\cdot}$ denote deviations from steady-state levels, i.e., $\widetilde{\boldsymbol{\theta}}_t \equiv \boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}$. We can derive the conditional expectations as

$$\mathbb{E}_t\left[\widetilde{\boldsymbol{\theta}}_{t+s}\right] = \rho^s\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + (s+1)\rho^s \widetilde{\boldsymbol{x}}_t. \tag{12}$$

On the right-hand side, the first term captures the a transitory component of demand deviation from the steady-state levels; this component decays at rate ρ. The second term captures the persistent component, which is hump-shaped in s: for $\rho > \frac{1}{2}$, $(s+1)\rho^s$ first increases and then decreases in s, peaking at time $s \equiv \lceil \frac{2\rho-1}{1-\rho} \rceil$ periods into the future.

Substituting the conditional expectation into Theorem 1, we obtain our first main result of how persistent and transitory demand shocks affect sectoral revenue. We first state and prove the result under general network and time-to-build structure; we then provide interpretation through a series of specific examples.

Theorem 2. *Under perfect information (i.e., agents separately observe the persistent and transitory components of demand shocks up to the current period), with demand characterized by equations (9) and (10), sectoral revenue at time t follows*

$$\widetilde{\boldsymbol{\gamma}}_t = \boldsymbol{G}^{\epsilon}_{\infty}\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + \boldsymbol{G}^{x}_{\infty}\widetilde{\boldsymbol{x}}_t, \tag{13}$$

where $\boldsymbol{G}^{\epsilon}_{\infty}$ captures the impact of demand's transitory component on revenue, and $\boldsymbol{G}^{x}_{\infty}$ captures the impact of demand's persistent component, and

$$\boldsymbol{G}^{\epsilon}_{\infty} \equiv \left(\boldsymbol{I} - \sum_{d=1}^{\infty}(\rho\beta)^d \boldsymbol{\Omega}_d\right)^{-1}, \qquad \boldsymbol{G}^{x}_{\infty} \equiv \boldsymbol{G}^{\epsilon}_{\infty}\left(\sum_{d=1}^{\infty} d\,(\rho\beta)^d \boldsymbol{\Omega}_d\right)\boldsymbol{G}^{\epsilon}_{\infty} + \boldsymbol{G}^{\epsilon}_{\infty}.$$

The impulse-response functions are:

$$\frac{\partial \widetilde{\boldsymbol{\gamma}}_{t+s}}{\partial \boldsymbol{\epsilon}_t} = \rho^s \boldsymbol{G}^{\epsilon}_{\infty}, \qquad \frac{\partial \widetilde{\boldsymbol{\gamma}}_{t+s}}{\partial \boldsymbol{u}_t} = s\rho^s \boldsymbol{G}^{\epsilon}_{\infty} + \rho^s \boldsymbol{G}^{x}_{\infty}.$$

Proof. Equation (12) and Theorem 1 jointly imply that

$$\widetilde{\boldsymbol{\gamma}}_t = \widetilde{\boldsymbol{\theta}}_t + \sum_{s=1}^{\infty}\boldsymbol{G}_s\left(\rho^s\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + (s+1)\,\rho^s\widetilde{\boldsymbol{x}}_t\right).$$

Let $\widetilde{\boldsymbol{\gamma}}_t = \boldsymbol{G}^{\epsilon}_{\infty}\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + \boldsymbol{G}^{x}_{\infty}\widetilde{\boldsymbol{x}}_t$. The matrix $\boldsymbol{G}^{\epsilon}_{\infty}$ summarizes the impact of demand's transitory component on revenue, and $\boldsymbol{G}^{x}_{\infty}$ captures the impact of the persistent component. We derive these two components separately.

To derive $\boldsymbol{G}^{\epsilon}_{\infty}$, we can follow the same logic as in the proof for Proposition 1 and decompose the impact of demand's transitory component recursively as the direct and indirect network effects, where each round of the network effect is summarized by $\boldsymbol{G}^{\epsilon}_{\infty}$, discounted to the presented by both the discount rate (β) and also the rate at which the shock dissipates (ρ). Specifically, note

$$\begin{aligned}
\boldsymbol{G}^{\epsilon}_{\infty} &= \boldsymbol{I} + \rho\beta\boldsymbol{\Omega}_1 + (\rho\beta)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + (\rho\beta)^3\left(\boldsymbol{\Omega}_1^3 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_3\right) + \cdots \\
&= \boldsymbol{I} + \rho\beta\boldsymbol{\Omega}_1\boldsymbol{G}^{\epsilon}_{\infty} + (\rho\beta)^2\boldsymbol{\Omega}_2\boldsymbol{G}^{\epsilon}_{\infty} + \cdots \\
&= \boldsymbol{I} + \left(\sum_{d=1}^{\infty}(\rho\beta)^d\boldsymbol{\Omega}_d\right)\boldsymbol{G}^{\epsilon}_{\infty},
\end{aligned}$$

and re-arranging yields the that $\boldsymbol{G}^{\epsilon}_{\infty} = \left(\boldsymbol{I} - \sum_{d=1}^{\infty}(\rho\beta)^d\boldsymbol{\Omega}_d\right)^{-1}$.

It is slightly more involved to derive $\boldsymbol{G}^{x}_{\infty}$ because the impact of a persistent shock does not decay monotonically; instead, the impact has an arithmetically increasing component as well as

an exponentially decaying component. We can write $\boldsymbol{G}^x_\infty$ as

$$\boldsymbol{G}^x_\infty = \boldsymbol{I} + 2\rho\beta\boldsymbol{\Omega}_1 + 3\left(\rho\beta\right)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + 4\left(\rho\beta\right)^3\left(\boldsymbol{\Omega}_1^3 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_3\right) + \cdots$$

1. The first term $\boldsymbol{I}$ captures the impact of the persistent component $\widetilde{\boldsymbol{x}}$ on time-t revenue through concurrent demand.

2. The second term $2\rho\beta\boldsymbol{\Omega}_1$ captures the impact of $\widetilde{\boldsymbol{x}}_t$ on time-t revenue through expected demand at $t+1$. The impact of $\widetilde{\boldsymbol{x}}$ on $t+1$ demand is captured by 2ρ. The impact of $t+1$ demand on time-t revenue goes through input-output linkages with one-period delay and is captured by $\boldsymbol{\Omega}_1$.

3. The third term $3\left(\rho\beta\right)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right)$ captures the impact of $\widetilde{\boldsymbol{x}}_t$ on time-t revenue through expected demand at $t+2$. The impact of $\widetilde{\boldsymbol{x}}$ on $t+2$ demand is captured by $3\rho^2$. The $t+2$ demand on time-t revenue goes through input-output linkages with two periods delay and is captured by $\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2$. The term $\boldsymbol{\Omega}_1^2$ captures the linkages from inputs supplied by the focal sector to the sector experiencing the demand shock through another intermediate producer, with two walks in total and one-period delay along each walk. The term $\boldsymbol{\Omega}_2$ captures the linkages from the focal sector directly to the sector experiencing the demand shock, with one walk of two periods delay.

Each successive term in $\boldsymbol{G}^x_\infty$ summarizes the impact of $\widetilde{\boldsymbol{x}}_t$ on time-t revenue through demand at a future time $t+s$, accounting for all possible direct and indirect network linkages with a delay of s periods in total. To derive $\boldsymbol{G}^x_\infty$ in closed-form, we re-write it recursively.

Define $\boldsymbol{Z} \equiv \sum_{d=1}^{\infty} d\left(\rho\beta\right)^d \boldsymbol{\Omega}_d$. We can then write $\boldsymbol{G}^x_\infty$ as

$$\begin{aligned}
\boldsymbol{G}^x_\infty &= \boldsymbol{I} + \rho\beta\boldsymbol{\Omega}_1\left(2\boldsymbol{I} + 3\rho\beta\boldsymbol{\Omega}_1 + 4\left(\rho\beta\right)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + \cdots\right) \\
&\quad + \left(\rho\beta\right)^2\boldsymbol{\Omega}_2\left(3\boldsymbol{I} + 4\rho\beta\boldsymbol{\Omega}_1 + 5\left(\rho\beta\right)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + \cdots\right) \\
&\quad + \left(\rho\beta\right)^3\boldsymbol{\Omega}_3\left(4\boldsymbol{I} + 5\rho\beta\boldsymbol{\Omega}_1 + 6\left(\rho\beta\right)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + \cdots\right) \\
&= \boldsymbol{I} + \rho\beta\boldsymbol{\Omega}_1\left(\boldsymbol{G}^\epsilon_\infty + \boldsymbol{G}^x_\infty\right) + \left(\rho\beta\right)^2\boldsymbol{\Omega}_2\left(2\boldsymbol{G}^\epsilon_\infty + \boldsymbol{G}^x_\infty\right) + \left(\rho\beta\right)^3\boldsymbol{\Omega}_3\left(3\boldsymbol{G}^\epsilon_\infty + \boldsymbol{G}^x_\infty\right) + \cdots \\
&= \boldsymbol{I} + \boldsymbol{Z}\boldsymbol{G}^\epsilon_\infty + \sum_{d=1}^{\infty}\left(\rho\beta\right)^d\boldsymbol{\Omega}_d\boldsymbol{G}^x_\infty \\
&= \left(\boldsymbol{I} - \sum_{d=1}^{\infty}\left(\rho\beta\right)^d\boldsymbol{\Omega}_d\right)^{-1}\left(\boldsymbol{I} + \boldsymbol{Z}\boldsymbol{G}^\epsilon_\infty\right) \\
&= \boldsymbol{G}^\epsilon_\infty + \boldsymbol{G}^\epsilon_\infty\boldsymbol{Z}\boldsymbol{G}^\epsilon_\infty
\end{aligned}$$

To derive the impulse-response, equation (11) implies that

$$\partial\widetilde{\boldsymbol{\theta}}_{t+s}\Big/\partial\boldsymbol{\epsilon}_t = \rho^s, \qquad \partial\widetilde{\boldsymbol{x}}_{t+s}\Big/\partial\boldsymbol{\epsilon}_t = 0,$$

$$\partial\widetilde{\boldsymbol{\theta}}_{t+s}\Big/\partial\boldsymbol{u}_t = (s+1)\,\rho^s, \quad \text{and} \quad \partial\widetilde{\boldsymbol{x}}_{t+s}\Big/\partial\boldsymbol{\epsilon}_t = \rho^s.$$

Hence

$$\frac{\partial\widetilde{\boldsymbol{\gamma}}_{t+s}}{\partial\boldsymbol{\epsilon}_t} = \boldsymbol{G}^{\epsilon}_{\infty}\frac{\partial\widetilde{\boldsymbol{\theta}}_{t+s}}{\partial\boldsymbol{\epsilon}_t} = \rho^s\boldsymbol{G}^{\epsilon}_{\infty},$$

$$\begin{aligned}\frac{\partial\widetilde{\boldsymbol{\gamma}}_{t+s}}{\partial\boldsymbol{u}_t} &= \boldsymbol{G}^{\epsilon}_{\infty}\frac{\partial\widetilde{\boldsymbol{\theta}}_{t+s}}{\partial\boldsymbol{u}_t} + \left(\boldsymbol{G}^{x}_{\infty} - \boldsymbol{G}^{\epsilon}_{\infty}\right)\frac{\partial\widetilde{\boldsymbol{x}}_{t+s}}{\partial\boldsymbol{u}_t} \\ &= (s+1)\,\rho^s\boldsymbol{G}^{\epsilon}_{\infty} + \rho^s\boldsymbol{G}^{\epsilon}_{\infty}\left(d\sum_{d=1}^{\infty}(\rho\beta)^d\,\boldsymbol{\Omega}_d\right)\boldsymbol{G}^{\epsilon}_{\infty}, \qquad \text{as desired.}\end{aligned}$$

□

3.2 Implications of Theorem 2

We now discuss several implications of Theorem 2.

3.2.1 Upstream's Amplified Response to Persistent Downstream Demand Shocks

The bullwhip effect refers to the phenomenon where a small demand shock downstream leads to amplified response in upstream sectors. Theorem 2 implies that the bullwhip effect cannot occur when demand shocks are transitory and may occur only if demand shocks are persistent. Intuitively, how producers respond to a downstream demand shock depends on their positions along the supply chain. Producers that are close in network distance to the shock increase production in response to higher expected demand in the near future, but upstream producers—because of time-to-build—increase production only in response to demand shocks that are anticipated to persist. In our context, transitory shocks decay monotonically over time, and the revenue response of upstream producers are therefore dampened relative to that of downstream producers. On the other hand, a persistent demand shock creates a hump-shaped demand profile over time and thus may generate the bullwhip effect, eliciting an amplified response in upstream sectors.

Formally, the revenue response to a concurrent transitory shock is

$$\partial\widetilde{\boldsymbol{\gamma}}_t\Big/\partial\boldsymbol{\epsilon}_t = \boldsymbol{I} + \rho\beta\boldsymbol{\Omega}_1 + (\rho\beta)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + (\rho\beta)^3\left(\boldsymbol{\Omega}_1^3 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_3\right) + \cdots \tag{14}$$

and the response to a concurrent persistent shock is

$$\partial\widetilde{\boldsymbol{\gamma}}_t\Big/\partial\boldsymbol{u}_t = 1\times\boldsymbol{I} + 2\times\rho\beta\boldsymbol{\Omega}_1 + 3\times(\rho\beta)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right) + 4\times(\rho\beta)^3\left(\boldsymbol{\Omega}_1^3 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_3\right) + \cdots \tag{15}$$

Each successive term in these summations capture the revenue response due to demand in a subsequent period. For a transitory shock, the scalars associated with each term forms an exponentially decaying sequence: $1, \rho\beta, (\rho\beta)^2, \ldots$ The further distant a sector is from the shock,

the lower is the impact of the shock on sectoral revenue. By contrast, the scalar sequence for a persistent shock has a hump-shaped profile $1, 2\rho\beta, 3(\rho\beta)^2, \ldots$, creating the scope for amplified response in upstream sectors.

In the special case where each input has one period delay only, $\boldsymbol{G}^{\epsilon}_{\infty} = (\boldsymbol{I} - \rho\beta\boldsymbol{\Omega}_1)^{-1}$ and $\boldsymbol{G}^{x}_{\infty} = (\boldsymbol{I} - \rho\beta\boldsymbol{\Omega}_1)^{-2}$, and equation (15) becomes

$$\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{u}_t = (\boldsymbol{I} - \rho\beta\boldsymbol{\Omega}_1)^{-2} = \boldsymbol{I} + 2\rho\beta\boldsymbol{\Omega}_1 + 3(\rho\beta\boldsymbol{\Omega}_1)^2 + 4(\rho\beta\boldsymbol{\Omega}_1)^3 + \cdots$$

where the Leontief-inverse-squared is the matrix analogous to the scalar summation formula $\sum_{t=0}^{\infty}(t+1)\rho^t = (1-\rho)^{-2}$.

3.2.2 Example: A Vertical Network

To get more intuitions, consider a vertical supply chain with $N = 4$ sectors. Sector 1 produces with labor only; each successive sector $i \geq 2$ produces using input $i-1$ with intermediate elasticity $\omega_{i,i-1} = 1$. Each good i takes d_i periods to build. All goods may be consumed. We say sector j is upstream to sector i if $j < i$, as sector j's output eventually becomes sector i's input.

Given the network structure, we can write the impulse-response of sectoral revenue response to a transitory demand shock as

$$\frac{\partial\widetilde{\boldsymbol{\gamma}}_t}{\partial\boldsymbol{\epsilon}_t} \times \boldsymbol{\epsilon}_t = \begin{bmatrix} 1 & (\rho\beta)^{d_1} & (\rho\beta)^{d_1+d_2} & (\rho\beta)^{d_1+d_2+d_3} \\ 0 & 1 & (\rho\beta)^{d_2} & (\rho\beta)^{d_2+d_3} \\ 0 & 0 & 1 & (\rho\beta)^{d_3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

and to a persistent demand shock as

$$\frac{\partial\widetilde{\boldsymbol{\gamma}}_t}{\partial\boldsymbol{u}_t} \times \boldsymbol{u}_t = \begin{bmatrix} 1 & (1+d_1)(\rho\beta)^{d_1} & (1+d_1+d_2)(\rho\beta)^{d_1+d_2} & (1+d_1+d_2+d_3)(\rho\beta)^{d_1+d_2+d_3} \\ 0 & 1 & (1+d_2)(\rho\beta)^{d_2} & (1+d_2+d_3)(\rho\beta)^{d_2+d_3} \\ 0 & 0 & 1 & (1+d_3)(\rho\beta)^{d_3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$$

The sensitivity of sector j's revenue with respect to a transitory demand shock in sector $i \geq j$ is $(\rho\beta)^d$, where d is the total periods of delay for any value-added in sector j to be used as inputs in sector i. By contrast, when the demand shock is persistent, the sensitivity is $(d+1)(\rho\beta)^d$.

Consider a transitory shock ϵ_4 to the most downstream sector 4, and we examine the last column of the matrix $\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t$ for the sectoral revenue response. Sector 4's revenue responds one-for-one to the shock. The response of sector 3 is dampened by $(\rho\beta)^{d_3} < 1$. The response of sector 2 is further dampened to $(\rho\beta)^{d_2+d_3}$. The further a producer is from the transitory shock—further upstream—the less revenue response to the shock.

Now instead consider a persistent shock u_4 to the most downstream sector 4, and we examine the last column of $\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{u}_t$ for sectoral revenue response. Sector 4's revenue responds one-for-one to the shock. The response of sector 3 is $(1+d_3)(\rho\beta)^{d_3}$ and may be amplified. The response of sector 2, which is further upstream to sector 3, is $(1+d_2+d_3)(\omega\rho\beta)^{d_2+d_3}$ and may be further amplified relative to sector 3. For a sector sufficiently upstream, the response will become dampened and eventually converges to zero (as $\lim_{s\to\infty}(s+1)(\rho\beta)^s = 0$), as the distance between the sector and the final demand shock could become so large such that, due to the long time-to-build, the sector does not even respond to persistent demand shocks.

3.2.3 A Bilateral Measure of Supply Chain Delays

Consider sector i's revenue response to sector j's persistent demand shock relative to that of a transitory shock. Specifically, define

$$\xi_{ij} \equiv \begin{cases} \left[\frac{\partial\widetilde{\boldsymbol{\gamma}}_t}{\partial \boldsymbol{u}_t}\right]_{ij} \Bigg/ \left[\frac{\partial\widetilde{\boldsymbol{\gamma}}_t}{\partial \boldsymbol{\epsilon}_t}\right]_{ij} & \text{if } \left[\frac{\partial\widetilde{\boldsymbol{\gamma}}_t}{\partial \boldsymbol{\epsilon}_t}\right]_{ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

ξ_{ij} is the element-wise ratio between the matrices capturing the response to persistent and transitory shocks. In the vertical supply chain example, the matrix $\boldsymbol{\Xi} \equiv \left[\xi_{ij}\right]$ is

$$\boldsymbol{\Xi} = \begin{bmatrix} 1 & 1+d_1 & 1+d_1+d_2 & 1+d_1+d_2+d_3 \\ 0 & 1 & 1+d_2 & 1+d_2+d_3 \\ 0 & 0 & 1 & 1+d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The ij-the entry, ξ_{ij}, is equal to one plus the delay between production of each input i to when the input is eventually used in sector j, and zero if input i is never used (even indirectly) in the production of j. It is therefore a measure of the pair-wise input delay between an input supplier and an input user.

We now show that this interpretation of ξ_{ij} as a bilateral measure of input delays extends to a general network structure. Specifically, consider sector i's revenue response to a transitory shock in sector j. Whenever $\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{\epsilon}_t\right]_{ij}$ is non-zero, we divide the ij-th entry on the right-hand side of equation (14) by the left-hand side to obtain

$$1 = \frac{[\boldsymbol{I}]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{\epsilon}_t\right]_{ij}} + \frac{[\rho\beta\boldsymbol{\Omega}_1]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{\epsilon}_t\right]_{ij}} + \frac{\left[(\rho\beta)^2\left(\boldsymbol{\Omega}_1^2+\boldsymbol{\Omega}_2\right)\right]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{\epsilon}_t\right]_{ij}} + \frac{\left[(\rho\beta)^3\left(\boldsymbol{\Omega}_1^3+\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2+\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_3\right)\right]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial \boldsymbol{\epsilon}_t\right]_{ij}} + \cdots \quad (16)$$

Sector i responds to sector j's transitory demand shock by raising production and sending its output along direct or indirect network routes to be used as an input in sector j. Each term on the

right-hand side of (16) captures the fraction of the output response that eventually becomes sector j's input with a specific time delay. The first term is the direct response to i's own demand shock. The second term is the revenue response to inputs sold with one-period delay to producers that experience the demand shock. The third term is sector i's share of revenue response to demand shocks two periods later, and so on. In summary, each term in the infinite sequence reflects the fraction of sector i's sales response due to sales at further distant positions in the supply chain, and the sequence sums to one.

We can write ξ_{ij} as

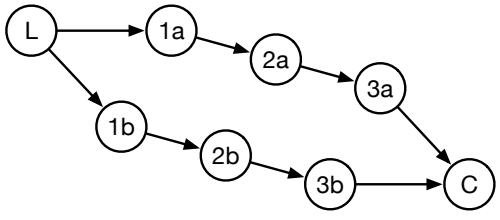
$$\begin{aligned}\xi_{ij} &= 1 \times \frac{[\boldsymbol{I}]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t\right]_{ij}} + 2 \times \frac{[\rho\beta\boldsymbol{\Omega}_1]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t\right]_{ij}} + 3 \times \frac{\left[(\rho\beta)^2\left(\boldsymbol{\Omega}_1^2 + \boldsymbol{\Omega}_2\right)\right]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t\right]_{ij}} \\ &\quad + 4 \times \frac{\left[(\rho\beta)^3\left(\boldsymbol{\Omega}_1^3 + \boldsymbol{\Omega}_1\boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_3\right)\right]_{ij}}{\left[\partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t\right]_{ij}} + \cdots\end{aligned}$$

ξ_{ij} is therefore one plus the average delay between when good i is produced and when it becomes an input in sector j, weighted by the additional output of i sent along each route in response to a transitory demand shock to j.

Note that ξ_{ij} depends on ρ, the autocorrelation coefficient of demand shocks. As $\rho \to 1$, a transitory AR(1) shock ϵ_{it} leads to a permanent shift in the level of demand.[6] Hence, $\lim_{\rho\to 1} \partial\widetilde{\boldsymbol{\gamma}}_t \Big/ \partial\boldsymbol{\epsilon}_t = \partial\bar{\boldsymbol{\gamma}} \Big/ \partial\bar{\boldsymbol{\theta}}$ provides the steady-state changes in revenue following steady-state demand changes, and $\lim_{\rho\to 1} \xi_{ij}$ captures the average delay from an input supplier i to an input user j along each network route using the steady-state response of goods flow along each route to demand changes in j as the weights.

Consider an example network below. The network has two supply chains labeled "a" and "b", each with three sectors.

Figure 3. Two production chains



$$\boldsymbol{\Xi} = \begin{array}{c} \\ \text{1a} \\ \text{2a} \\ \text{3a} \\ \text{1b} \\ \text{2b} \\ \text{3b} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{1a} & \text{2a} & \text{3a} & \text{1b} & \text{2b} & \text{3b} \end{array} \\ \begin{bmatrix} 1 & 1+d_{1a} & 1+d_{1a}+d_{2a} & 0 & 0 & 0 \\ 0 & 1 & 1+d_{2a} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1+d_{1b} & 1+d_{1b}+d_{2b} \\ 0 & 0 & 0 & 0 & 1 & 1+d_{2b} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Consider $3a$ as the focal sector. The delay from $2a$ to $3a$ is simply d_{2a}. The delay from $1a$ to $3a$ is $d_{1a} + d_{2a}$, as input $1a$ has to first go through producer $2a$. The delay from any sector along

[6] A persistent AR(2) shock $\boldsymbol{u}_t$ becomes a permanent change in the growth of demand as $\rho \to 1$.

the "b" chain is unconnected to $3a$ and thus has a delay measure of zero. The matrix $\boldsymbol{\Xi}$ therefore reveals the bilateral supply chain distance measured by production delays.

Note that our measure ξ_{ij} differs from the standard Leontief inverse measure. The latter captures a notion of network dependence—the network-adjusted value of each input used to produce a unit value of each good, whether in steady-state ($\partial\bar{\boldsymbol{\gamma}}\Big/\partial\bar{\boldsymbol{\theta}} = \left(\boldsymbol{I} - \sum_{d=1}^{\infty}\beta^d\boldsymbol{\Omega}_d\right)^{-1}$) or in response to transitory demand shocks ($\partial\widetilde{\boldsymbol{\gamma}}_t\Big/\partial\boldsymbol{\epsilon}_t = \left(\boldsymbol{I} - \sum_{d=1}^{\infty}(\rho\beta)^d\boldsymbol{\Omega}_d\right)^{-1}$). Instead, our bilateral measure ξ_{ij} captures a notion of network *distance* due to delays. Our measure ξ_{ij} can therefore be used to identify the ordering of inputs along each good-specific supply chains, as evident from the example in Figure 3 of two production chains. We implement this measure empirically in later sections.

Related to ξ_{ij} is the Upstreamness measure ψ_i of Antràs, Chor, Fally and Hillberry (2012), which is defined as the average distance in terms of production stages from a sector i to the final consumer demand in a static model. In our dynamic context, ψ_i as defined by Antràs et al. (2012) is equivalent to the average delay from input i to all sectors j, weighted by the steady-state size of each sector j (as $\rho \to 1$), in a special case where all input delays are one period only. Specifically $\psi_i = \lim_{\rho\to 1}\sum_j \xi_{ij}\bar{\gamma}_j$. The key distinctions are twofolds: first, ξ_{ij} generalizes the upstreamness measure by Antràs et al. (2012) to accommodate different lengths of production delays; second, ξ_{ij} is defined pair-wise and measures the delay from each input not to final demand as a whole but to each input-using industry separately (also see Alfaro, Chor, Antras and Conconi, 2019 for a pair-wise extension of Antràs et al., 2012's in a static model).

3.2.4 Persistent Demand Shocks Create Amplified Volatility in Upstream Sectors

In sections 3.2.1 and 3.2.2 we have shown that upstream sectors' revenues are more responsive to persistent demand shocks. These shocks have hump-shaped impulse responses, and an increase in demand at time t implies an even higher expected future demand, to which upstream is more responsive due to time-to-build.

In this section, we formally derive the volatility of sectoral revenue in term of period-to-period changes and show that the same intuition implies that under persistent shocks, upstream sectors tend to have higher volatility. We do so via two steps. First, we show that our derivation of Theorem 2 extends naturally for calculating the variance of one-period-ahead sectoral revenue conditional on the current information.

Proposition 3. *Under perfect information, the sectoral revenue volatility is:*

$$Var_t\left(\Delta\widetilde{\boldsymbol{\gamma}}_{t+1}\right) = \boldsymbol{G}^{\epsilon}_{\infty}\boldsymbol{\Sigma}^{\epsilon}_t\boldsymbol{G}^{\epsilon'}_{\infty} + \boldsymbol{G}^{x}_{\infty}\boldsymbol{\Sigma}^{u}_t\boldsymbol{G}^{x'}_{\infty},$$

where $\boldsymbol{\Sigma}^{\epsilon}_t$ and $\boldsymbol{\Sigma}^{x}_t$ are the variance-covariance matrices of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{u}_t$, respectively, and $\boldsymbol{G}^{\epsilon}_{\infty}, \boldsymbol{G}^{x}_{\infty}$

are as defined in Theorem 2.

As in Theorem 2, the matrix $\boldsymbol{G}^{\epsilon}_{\infty}$ captures the impact of demand's transitory component on revenue, and $\boldsymbol{G}^{x}_{\infty}$ captures the impact of demand's persistent component. Correspondingly, $\boldsymbol{G}^{\epsilon}_{\infty}\boldsymbol{\Sigma}^{\epsilon}\boldsymbol{G}^{\epsilon'}_{\infty}$ and $\boldsymbol{G}^{x}_{\infty}\boldsymbol{\Sigma}^{u}\boldsymbol{G}^{x'}_{\infty}$ are the quadratic forms capturing demand's two components on the variance of the one-period-ahead revenue conditional on current period information. In a parallel fashion to our analysis in sections 3.2.1 and 3.2.2, Lemma 3 implies that transitory demand shocks from downstream sectors create dampened volatility in the future (one period ahead) revenue of upstream sectors, persistent, hump-shaped demand shocks create amplified volatility for the future revenue of upstream sectors. Note that this statement is distinct from the fact that upstream sectors may have higher volatility, which may be simply because they supply to many downstream sectors and therefore accumulate the impact of many downstream demand shocks. Instead, Lemma 3 implies that the persistent demand shocks faced by a downstream sector can create amplified effects on upstream revenue even if there is only a single path of input-output linkages between the two sectors.

To see Lemma 3's implication as clearly as possible, consider again the vertical supply chain example in section 3.2.2. When only the most downstream sector 4 faces uncertain demand—with variance $\sigma^2_{\epsilon_4}$ and $\sigma^2_{u_4}$ for the transitory and persistent shocks, respectively—the conditional variance of one-period-ahead revenue $Var_t\left(\widetilde{\gamma}_{i,t+1}\right)$ is $Var_t\left(\widetilde{\gamma}_{4,t+1}\right) = \sigma^2_{\epsilon_4} + \sigma^2_{u_4}$ in the most downstream sector 4 and is

$$Var_t\left(\widetilde{\gamma}_{1,t+1}\right) = (\rho\beta)^{2(d_1+d_2+d_3)}\,\sigma^2_{\epsilon_4} + (1+d_1+d_2+d_3)^2\,(\rho\beta)^{2(d_1+d_2+d_3)}\,\sigma^2_{u_4}$$

in the most upstream sector 1. The effect of transitory demand shocks on upstream revenue is dampened (by a factor of $(\rho\beta)^{2(d_1+d_2+d_3)}$), whereas the effect of persistent demand shocks may be magnified.

3.3 Extensions

3.3.1 Heterogeneous Rates of Convergence for Demand Shock Components

Our analysis thus far focuses on the case where both the transitory and persistent components of demand shocks have the same parameter ρ capturing the rate of convergence. In this section, we generalize our results in section 3 thus far to the case where the two demand shock components have heterogeneous rates of convergence.

Specifically, we replace the demand shock process described by equations (9) and (10) with the following:

$$\theta_{it} - \bar{\theta}_i = \rho_\epsilon\left(\theta_{it-1} - \bar{\theta}_i\right) + x_{it} + \epsilon_{it}, \tag{17}$$

$$x_{it} = \rho_x x_{i,t-1} + u_{it}. \tag{18}$$

where ρ_ϵ captures the rate at which demand reverts back to the steady-state under transitory shocks and ρ_x parametrizes the rate of convergence under persistent shocks.

We can write the expected future demand as

$$\mathbb{E}_t\left[\widetilde{\boldsymbol{\theta}}_{t+s}\right] = \rho_\epsilon^s\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + \frac{\rho_\epsilon^{s+1} - \rho_x^{s+1}}{\rho_\epsilon - \rho_x}\widetilde{\boldsymbol{x}}_t$$

which generalizes equation (12) of the baseline model. We now generalize Theorems 2 and Proposition 3 to this setting.

Proposition 4. *Under perfect information, with demand characterized by equations (17) and (18), sectoral revenue at time t follows*

$$\widetilde{\boldsymbol{\gamma}}_t = \boldsymbol{G}_\infty^\epsilon\left(\widetilde{\boldsymbol{\theta}}_t - \widetilde{\boldsymbol{x}}_t\right) + \boldsymbol{G}_\infty^x\widetilde{\boldsymbol{x}}_t, \tag{19}$$

where $\boldsymbol{G}_\infty^\epsilon$ captures the impact of demand's transitory component on revenue, and $\boldsymbol{G}_\infty^x$ captures the impact of demand's persistent component, and

$$\boldsymbol{G}_\infty^\epsilon \equiv \left(\boldsymbol{I} - \sum_{d=1}^{\infty}\left(\rho_\epsilon\beta\right)^d\boldsymbol{\Omega}_d\right)^{-1}, \qquad \boldsymbol{G}_\infty^x \equiv \frac{\rho_\epsilon}{\rho_\epsilon - \rho_x}\boldsymbol{G}_\infty^\epsilon - \frac{\rho_x}{\rho_\epsilon - \rho_x}\left(\boldsymbol{I} - \sum_{d=1}^{\infty}\left(\rho_x\beta\right)^d\boldsymbol{\Omega}_d\right)^{-1}.$$

The impulse-response functions are:

$$\frac{\partial\mathbb{E}_t\left[\widetilde{\boldsymbol{\gamma}}_{t+s}\right]}{\partial\boldsymbol{\epsilon}_t} = \rho_\epsilon^s\boldsymbol{G}_\infty^\epsilon, \qquad \frac{\partial\mathbb{E}_t\left[\widetilde{\boldsymbol{\gamma}}_{t+s}\right]}{\partial\boldsymbol{u}_t} = \boldsymbol{G}_\infty^\epsilon\frac{\rho_\epsilon^{s+1} - \rho_x^{s+1}}{\rho_\epsilon - \rho_x} + \left(\boldsymbol{G}_\infty^x - \boldsymbol{G}_\infty^\epsilon\right)\rho_x^s.$$

The revenue volatility is $Var_t\left(\Delta\widetilde{\boldsymbol{\gamma}}_{t+1}\right) = \boldsymbol{G}_\infty^\epsilon\boldsymbol{\Sigma}_t^\epsilon\boldsymbol{G}_\infty^{\epsilon\prime} + \boldsymbol{G}_\infty^x\boldsymbol{\Sigma}_t^u\boldsymbol{G}_\infty^{x\prime}$.

3.3.2 Imperfect Information

Our analysis thus far demonstrates that downstream sectors' revenue responds more to transitory demand shocks, whereas upstream sectors' revenue may respond more to persistent demand shocks. What if agents in the economy only observes the demand $\widetilde{\boldsymbol{\theta}}_t$ but not the separate components? Intuitively, when a demand shock arises at time t, it is difficult for agents to accurately assess its impact on future demand. Agents only learn over-time about the nature of shocks: as an unexpectedly positive demand stays elevated over time, agents may gradually shift their beliefs towards the shock being persistent. Hence, agents in this environment solve an additional filtering problem due to imperfect information. Sectoral revenue response exhibits history dependence—the response is stronger if a positive demand shock comes after a sequence of positive shocks, i.e., the momentum of demand shocks—especially in upstream sectors.

We demonstrate this result in the context of the 4-sector vertical supply chain example. Suppose only sector 4 supplies to the consumer and experiences demand shocks. For notational clarity we remove the sector 4 subscript when discussing demand θ_t and shocks (ϵ_t, u_t). Suppose

that agents observe only the history of demand $\left\{\widetilde{\theta}_{t-s}\right\}_{s\geq 0}$ but not the transitory and persistent components separately. The conditional expectation operator is then defined as $\mathbb{E}_t[\cdot] \equiv \mathbb{E}\left[\cdot|\left\{\widetilde{\theta}_{t-s}\right\}_{s\geq 0}\right]$. In this setting, agents, including the social planner, need to solve a filtering problem and forecast future demand based the history of demand shocks. Only $\widetilde{\theta}_t$ is observable; x_t, the permanent component of current demand, is a hidden state.

The exogenous state variables at time t continues to be $\left(\widetilde{\theta}_t, \widetilde{x}_t\right)$, as in the baseline model with perfect information. Under imperfect information, there is an additional, endogenous state variable that captures the agents' belief of the hidden state $\widetilde{x}_t$ based on the history of demand up to time t. We refer to the mean of the belief $\hat{x}_t \equiv \mathbb{E}_t\left[\widetilde{x}_t|\left\{\widetilde{\theta}_{t-s}\right\}_{s\geq 0}\right]$ as the nowcast. We posit that learning follows a Kalman filter:

$$\hat{x}_{t+1} = \rho\hat{x}_t + \kappa_{t+1}\left(\widetilde{\theta}_{t+1} - \rho\left(\widetilde{\theta}_t + \hat{x}_t\right)\right) \tag{20}$$

where the Kalman gain $\kappa_{t+1} \in [0,1]$ is a function of the history of demand shocks and the variance of the two components of demand shocks. The nowcast based on the Kalman filter is optimal when demand shocks are Gaussian; in our setting, demand shocks must be bounded and thus must be non-Gaussian; nevertheless, the Kalman filter still provides the best linear nowcast in the mean-square-error sense (Humpherys, Redd and West, 2012). Note that under perfect information, or if all shocks are persistent, $\kappa_t = 1$, in which case agents can accurately infer that the persistent demand shock component is $x_{t+1} = \widetilde{\theta}_{t+1} - \rho\widetilde{\theta}_t$. A lower κ_{t+1} therefore implies that agents under-reacts to new information.

To demonstrate how under-reaction and learning interacts with the supply chain, we now derive the path of sectoral revenue $\widetilde{\gamma}_t$, given $\widetilde{\theta}_0 = x_0$ initially and shutting down all shocks for time $t > 0$. We also examine how the initial nowcast $\hat{x}_0$, which is based on the cumulative learning from past shocks, affects the revenue path.

First, note that, in the absence of any further shocks from time t onwards, the law of motions for the state variables are:

$$\widetilde{\theta}_s = (s+1)\,\rho^s\widetilde{x}_0, \qquad \widetilde{x}_s = \rho^s\widetilde{x}_0.$$

The law of motion for the nowcast $\hat{x}_s$ based on realized demand until each time s is

$$\hat{x}_s = \widetilde{x}_s + \rho^s\left(\hat{x}_0 - \widetilde{x}_0\right)\prod_{u=1}^{s}\left(1-\kappa_u\right). \tag{21}$$

Equation (21) implies that the gap between the nowcast $\hat{x}_s$ and the actual state variable $\widetilde{x}_s$ shrinks to zero over time faster than the rate ρ with which the state variable $\widetilde{x}_s$ converges to zero. This is because there is learning based on new information, i.e., the path of observed demand $\widetilde{\theta}_s$. The rate at which $\hat{x}_s$ converges to $\widetilde{x}_s$ depends on the rate of learning. Recall that the κ's capture

the rate of under-reaction to news; learning is faster when the κ's are closer to 1 and slower when the κ's are closer to zero.

Note also that the initial gap between the nowcast and the actual state variable, $\hat{x}_0 - \widetilde{x}_0$, persists over time. The initial nowcast $\hat{x}_0$ is based on learning from past shocks. The law of motion (20) implies that $\hat{x}_0$ tends to be larger when the surprises in past demand $(\widetilde{\theta}_{t+1} - \rho\left(\widetilde{\theta}_t + \hat{x}_t\right))$ were consistently positive. This implies that the path of sectoral revenue exhibit history dependence.

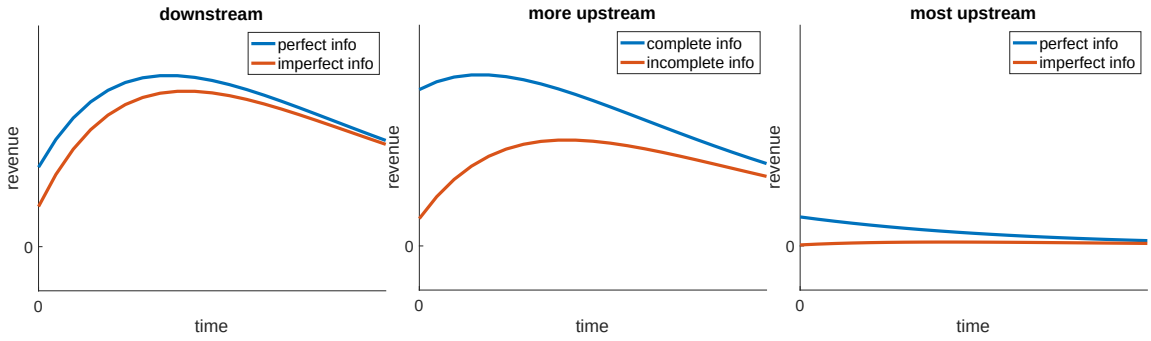
To see the implication of learning more explicitly, note that the difference between the revenue paths due to imperfect information, normalized by the rate ρ at which $\widetilde{x}_s$ converges to zero, is

$$\frac{\widetilde{\gamma}_s^{\text{imperfect info}} - \widetilde{\gamma}_s^{\text{perfect info}}}{\rho^s} = (\hat{x}_0 - \widetilde{x}_0) \prod_{u=1}^{s} (1 - \kappa_u) \times \begin{bmatrix} (1 + d_1 + d_2 + d_3)(\rho\beta)^{d_1+d_2+d_3} \\ (1 + d_2 + d_3)(\rho\beta)^{d_2+d_3} \\ (1 + d_3)(\rho\beta)^{d_3} \\ 1 \end{bmatrix}.$$

To the extent that the initial nowcast is below the true persistent component of demand ($\hat{x}_0 - \widetilde{x}_0 < 0$), sectoral revenue also under-reacts to demand. As demand converges back to the steady-state at rate ρ, the under-reaction in revenue converges to zero at a rate faster than ρ due to learning, governed by the Kalman gains κ_s. The revenue response features history dependence, as $\hat{x}_0$ reflects learning from past demand realizations. Most importantly, the history dependence exhibit non-monotonicity along the supply chain: starting from the most downstream sector, the history dependence may get amplified as one goes upstream—because of the increasing sequence $1 + d_1 + d_2 + \cdots$—but the effect eventually dampens as one goes further upstream because of the decaying sequence $(\rho\beta)^{1+d_1+d_2+\cdots}$.

As an illustration, Figure 4 demonstrates the revenue paths in a vertical supply chain, under perfect (blue line) and imperfect information (red line) for a downstream sector (left panel), a relatively upstream sector (middle panel), and a further upstream sector (right panel).

Figure 4. The revenue paths under perfect and imperfect information for a vertical supply chain, with initial state $\theta_0 = x_0, \hat{x}_0 < x_0$, and no shocks after time 0.



4 Empirics

We now turn to the empirical analysis with three goals. First, we provide evidence of the bullwhip effect at the sectoral level. That is, within a sectoral supply chain, upstream sectors tend to exhibit higher volatility than downstream sectors. Second, using quarterly data on industrial production, we extract demand shocks to downstream sectors and demonstrate the impulse-response of these shocks on sectoral production along the supply chain. We show that demand shocks tend to be hump-shaped, consistent with our notion that shocks are persistent, and that they elicit especially large response in upstream sectors when there is momentum in the demand shocks. Third, we estimate the process of demand shocks in the data and demonstrate that persistent demand shocks for downstream sectors account for a quantitatively significant share of sectoral fluctuations along the supply chains.

4.1 Data

We use data from three sources. First is the Industrial Production (IP) data published by the Federal Reserve Board (FRB). The data set has panel information on the sectoral value-added for the 114 manufacturing industries over the period 1972-2023. A sector in the data roughly corresponds to a four-digit industry classified by the North American Industry Classification System (NAICS) (Foerster, Sarte and Watson, 2011). The original data is at the monthly frequency. To smooth out measurement errors and high frequency shocks, we average the monthly data to construct quarterly sectoral value-added in our baseline specifications. We report specifications based on monthly data as robustness checks. The data report sectoral value-added as shares, meaning the value-added of each sector as a share of the total value-added of the 114 sectors. Because our analysis focuses on the relative sectoral size instead of the aggregate output, the relative value-added data is well-suited for our analysis.

To construct our bilateral measure of supply chains, we need information on both input-output linkages and input time-to-build delays. The second data source we use is the 2007 BEA input-output (IO) use table. From the IO table we construct input cost shares for inputs ω_{ij}. We aggregate sectors in the IO tables to concord with the IP data, resulting in 114 sectors in both data sets and 127 non-IP sectors. We also extract from the IO table additional information such as the value-added intensity (i.e., value-added to revenue ratio), and the consumer expenditure shares on each sector.

Third, to get the information on time-to-build delays, we rely on the U.S. Census M3 survey of manufacturers' shipments, inventories, and orders. The data set provides broad-based, monthly statistical data on economic conditions in the manufacturing sector. We follow Liu and Tsyvinski (2024) and construct the backlog ratio for each input, i.e., the ratio between the stock value of

unfilled orders and the flow value of goods delivered, as a measure of supply chain delays. For each sector, we compute the average backlog ratio between the years 2010 and 2019. We truncate the sample before 2020 in order to exclude the spike in backlogs due to the COVID-19 pandemic. We impute the backlog ratio using the sample average respectively for durable goods sectors and non-durable goods sectors that are not in the M3 survey. For those sectors that only has the aggregate series in M3 survey, we use aggregate number for those sectors. Sectors with the highest backlog ratios are Aerospace Products and Parts and Railroad Rolling Stock (32.3 months), ship and boat building (11.3 months), Medical Equipment and Supplies (9.6 months). The standard deviation is 3.89 months. We record a time-to-build delay of d quarters for any inputs with backlog ratios between d and $d+1$ quarters, with d bounded below at 1 quarter and capped at 4 quarters.

Our model uses the relative sectoral revenue as the outcome variable for measuring size. In our empirical analysis, we capture the relative sectoral revenue by adjusting the value-added in each sector from IP data using the sector's value-added intensity (i.e., value-added to revenue ratio) from the BEA IO table.

4.2 Higher Volatility in Upstream Sectors Across Supply Chains

We provide evidence for the bullwhip effect across supply chains of the US production network. That is, upstream sectors within a supply chain tend to have higher volatility.

We measure sectoral volatility using the standard deviation of the year-on-year (YoY) growth rate of sectoral value-added (as a share of total value-added of the IP industries) during our sample period. The YoY growth rate is chosen to account for the seasonality in sectoral value-added fluctuations.

The main challenge in documenting the bullwhip effect is to identify the supply chains in the US production network. To do so, we rely on our bilateral measure of supply chain delays ξ_{ij}, following the procedure described in section 3.2.3.

We start by identifying a set $\mathcal{D}$ of 25 "downstream" sectors that are important in the consumption bundle, following Antràs et al. (2012).[7] These sectors are listed in Table 1; they supply most of their output directly to the final consumer instead of other producers. Broadly speaking, these sectors fall into the categories of transportation, food, and other household consumption goods.

We then identify the supply chain for each of these downstream sectors. First, for each downstream sector $j \in \mathcal{D}$, we keep a set of sectors $\mathcal{U}_j$ that are considered as direct or indirect suppliers

[7] Specifically, we first select the 30 most downstream sectors according to the "upstreamness" measure by Antràs et al. (2012). We then drop "Other Miscellaneous Manufacturing", "Construction Machinery", "Agricultural Implements", "Industrial Machinery" and "Support Activities for Mining", as these five sectors do not account for significant consumer expenditure and are thus not important for the consumer demand.

Table 1. 25 downstream sectors

Apparel	Coffee and Tea
Ship and Boat Building	Other Transportation Equipment
Newspaper Publishers	Heavy Duty Trucks
Tobacco	Office And Other Furniture
Soft Drinks and Ice	Aerospace Products and Parts
Bakeries and Tortilla	Breweries
Motor Vehicle Bodies and Trailers	Other Food Except Coffee and Tea
Carpet and Rug Mills	Pharmaceuticals and Medicines
Soap, Cleaning Compounds, and Toilet Preparation	Navigational, Measuring, Electromedical, and Control Instruments
Periodical, Book, and Other Publishers	Sugar and Confectionery Products
Automobiles and Light Duty Motor Vehicles	Animal Slaughtering and Processing
Major Electrical Household Appliances	Medical Equipment and Supplies
Fruit and Vegetable Preserving and Specialty Foods	

to j: a sector i is in the set $\mathcal{U}_j$ if producing 100 dollar of j's output requires at least one dollar of input i, directly or indirectly. Formally, this means that the ij-th entry of the Leontief inverse $L_{ij} \equiv (I - \Omega)^{-1}_{ij}$ is greater than 1%. As our model features only non-durable inputs—the model does not have capital—we further restrict $\mathcal{U}_j$ to include only non-durable sectors, defined as those inputs for which no more than 50% of the final use goes into fixed capital formation in the BEA IO table.

Next, for each downstream sector $j \in \mathcal{D}$, we order the set of suppliers to j using our bilateral measure of supply chain delays ξ_{ij}, following the procedure described in section 3.2.3. In the baseline, we set $\rho = \beta = 1$ when calculating ξ_{ij}, but our results are robust to using realistic quarterly discount factor $\beta = 0.98$ and the rate of decay parameter $\rho = 0.8$ estimated from the data (as we discuss later in section 4.4). The resulting measure captures the average number of quarters in the time-to-build delays for input i to reach producer j, accounting for all the direct and indirect network linkages from i to j. Within the supply chain of each good j, we interpret sector $i1$ as more upstream to $i2$ if there is a longer delay from input $i1$ to producer j than from $i2$.

Take "automobiles and light duty motor vehicles" ("automobiles" for short) and "motor vehicle bodies and trailers" ("trailers" for short) as two examples of downstream sectors. Despite the fact that these two sectors share many common suppliers—such as such as "rubber products", "paperboard containers", "semiconductors", "iron and steel products", and "electric power generation and transmission"—our measure can isolate the intricate differences in their supply chains. Specifically, our measure identifies that the sectors most immediately upstream to automobiles

are "motor vehicle parts", "audio and video equipment", "tires", and "glass and glass products". By contrast, those that are most immediately upstream to trailers include "major electrical household appliances", "millwork", "rubber products excluding tires", and "paints and coatings".

Figure 5 shows the relationship between upstreamness, as measured by supply chain delays, and volatility, along the sector-specific supply chain for 9 of the 25 downstream sectors (we produce the same figure for each of the remaining 16 downstream sectors in the appendix). Specifically, each panel represents the supply chain associated with a downstream sector, as indicated by the subtitle (e.g., the top-left panel is for the supply chain of automobiles). Within each panel, a circle represents an input supplier along the supply chain (e.g., the bottom circle in the top-left panel is "motor vehicle parts"). The Y-axis captures the time-to-build delays from the supplier to the downstream producer, i.e., a notion of upstreamness. The X-axis is the volatility in the YoY value-added growth of the supplier. The size of each circle represents the share of the total downstream cost that is spent on each input, directly or indirectly.

The key takeaway from Figure 5 is the significant upward-sloping relationship between sectoral volatility and upstreamness. This is an empirical pattern that is consistently evident across the supply chains for all 25 downstream sectors. In terms of magnitude, pooling across 25 supply chains we find that going from 25 to 75 percentile in terms of upstreamness is associated with 6.2 percentage points higher volatility in value-added growth. This effect is pronounced across transportation, food, households consumption goods, and all other categories of downstream goods.

4.3 Demand Shocks and Impulse Response Along the Supply Chains

We now use local projections to demonstrate that demand shocks tend to be hump-shaped, and we estimate the impulse-response of sectoral value-added along the supply chain to downstream demand shocks. We show that upstream sector's response is especially sensitive to the momentum of demand shocks, consistent with learning.

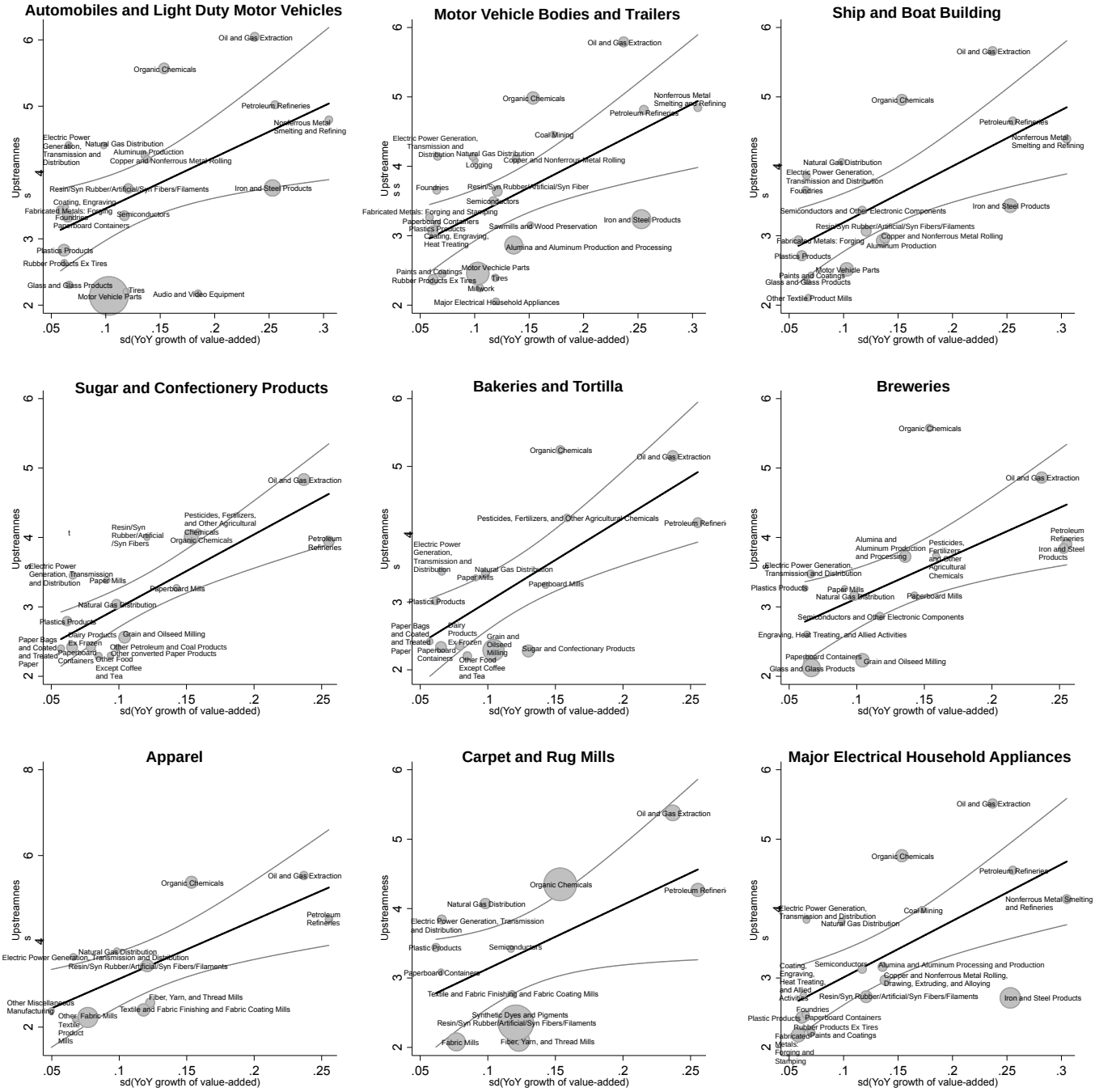
First, to account for the heterogeneous size of sectors, we standardize sectoral value-added and use the resulting z-score for our analysis.

To identify the demand shock of the 25 downstream sectors, we project each sector's value-added on its own lags, as well as sector-fixed effects μ_i and time-fixed effects μ_t, and we extract the residual, i.e., the demand innovation r_{it}, as the demand shock for sector i at time t:

$$VA_{it}^{down} = \sum_{s=1}^{p} \gamma_s VA_{i,t-1}^{down} + \mu_i + \mu_t + r_{it}, \quad i \in \mathcal{D}.$$

In the baseline specification, we choose $p = 1$, and we report longer lags as robustness.

Figure 5. Supply chain and volatility



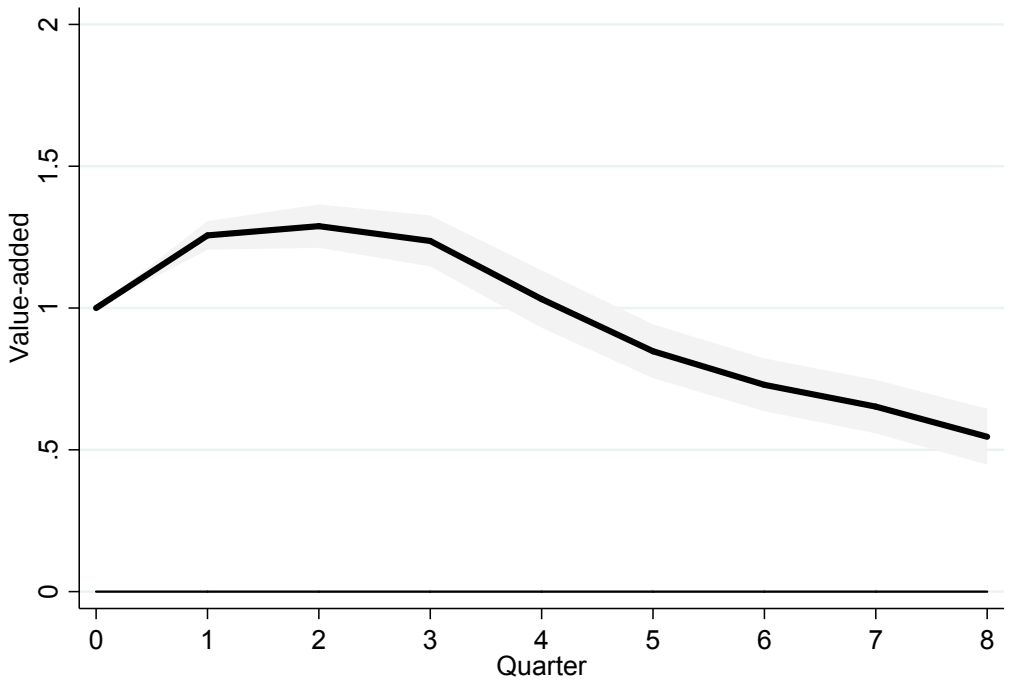
We first estimate the impulse-response of a downstream sector's value-added to its own demand shocks $\hat{r}_{it}$:

$$VA^{down}_{i,t+h} = \alpha^h r_{it} + X_{it} + \zeta_{it}, \tag{22}$$

$$\text{with controls} \quad X_{jt} = \left\{ \left\{ VA^{down}_{j,t-s} \right\}_{s=1}^{p}, \mu_j, \mu_t \right\}.$$

Figure 6 reports the coefficient α^h, displaying the estimates from Equation 22 of the impact of a demand shock from downstream sectors in quarter t on their own value-added share in quarter $t+h$ for a 2-year (8-quarter) horizon. Figure 6 captures a clear hump-shaped, mean-reverting impulse response function, consistent with our focus in section 3 on persistent shocks, which are hump-shaped and mean-reverting. One unit of demand shock in standard deviation of value-added from downstream sectors raises its own concurrent value-added by about one unit by our construction, while the cumulative response reaches its peak (1.289) two quarters later. The value-added decreases after two quarters and decays to about half of its initial response in about two years.

Figure 6. Cumulative impulse response



We then use the shocks $\hat{r}_{it}$ to estimate the impulse-response of upstream sectors for each production chain. Because there is a large number of sectors along each supply chain, we group sectors according to their supply chain positions. Specifically, for the supply chain associated with each good j, we divide suppliers (indexed by i) into 4 groups according to quartiles of our upstreamness/production delay measure ξ_{ij}. To be notationally consistent with our 4-sector vertical example, we designate suppliers in the most upstream quartile into group 1 and those suppliers closest to the downstream sector as group 4. Notice that if the same sector supplies to different

downstream goods, it can occupy different positions along the supply chains and may be categorized into different groups across supply chains. We drop "Oil and Gas Extraction" from our analysis as the sector is part of the supply chain of every downstream good.

To demonstrate learning from past demand shocks, we need a measure of momentum in sectoral demand. We adopt the "information continuity" (IC) measure of Da, Gurun and Warachka (2014), which is designed to capture momentum by measuring the consistency in the sign of shocks in the recent history. When past demand shocks were all positive or all negative, the IC measure takes the value of one; when post shocks alternative between positives and negatives, IC takes the value of zero. When the positive net change over a given window of history is driven by a large positive shock netting out many small negative shocks, the IC measure takes negative value and could approach -1. Formally, we construct IC by looking back 8 quarters and calculate the fraction of positive (%pos) and negative (%neg) demand shocks in each downstream sector:

$$IC_{j,t} = sign\left(VA^{down}_{jt} - VA^{down}_{j,t-8}\right) \times [\,\%\text{ pos } - \%\text{ neg }].$$

The mean IC in our sample is 0.325 with standard deviation 0.318, minimum IC value is -0.5 and maximum IC value is 1.

We estimate the impulse-response of sectors along the supply chain to downstream demand shocks using the following specification:

$$VA^{up,g}_{i,t+h} = \beta^{h,g} \sum_{w_{ij}\geq 0.01} w_{ij} r_{jt} + \delta^{h,g} \sum_{w_{ij}\geq 0.01} w_{ij} r_{jt} \times IC_{jt} + \sum_{w_{ij}\geq 0.01} w_{ij} IC_{jt} + Z^g_{it} + \epsilon_{it}, \quad (23)$$

$$\text{with controls} \quad Z^g_{it} = \left\{ \left\{ \sum_{w_{ij}\geq 0.01} w_{ij} VA^{down}_{j,t-s}, VA^{up,g}_{i,t-s} \right\}_{s=1}^{p}, \mu_i, \mu_t \right\}.$$

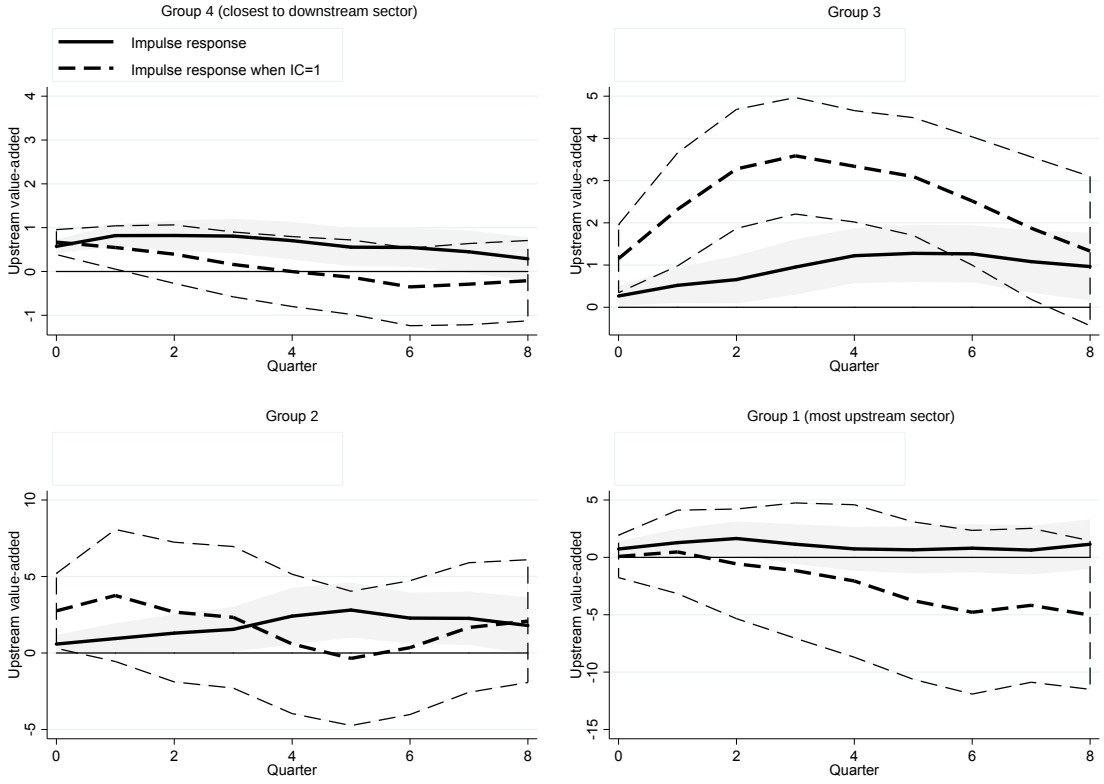
Specifically, we project demand shock from downstream sector j onto the revenue of sectors within each group g of the supply chain for good j—where, recall, the groups are assigned in terms of upstreamness, with group 1 being the most upstream and group 4 the most downstream. We also include as regressors the interaction between the demand shock from each downstream sector with the momentum of shocks in each sector. Because a producer may belong to the distinct supply chains of different goods, we combine the demand shock from downstream sectors by adding and weighting the shocks using $w_{ij} = L_{ij} \times \text{std}\left(\gamma_{j,t}\right) / \text{std}\left(\gamma_{i,t}\right)$. The weights are the Leontief-inverse between suppliers and downstream input users, which captures the importance of a supplier for that downstream good, adjusting for the fact that value-added variables are standardized. Doing so avoids repeated observations, as each producer appears only once on the left-hand side.

We plot the coefficients $\beta^{h,g}$ and $\beta^{h,g} + \delta^{h,g}$ across time h for different supply chain groups h

in Figure 7. Dashed lines show the estimates of $\beta^{h,g} + \delta^{h,g}$, which is also the impulse response function when $IC_j = 1$ for all j. Solid lines show the estimates of $\beta^{h,g}$, which is also the impulse response function when $IC_j = 0$ for all j. The bar shows the 90% confidence interval of these estimates. The differences between dashed lines and solid lines demonstrates the importance of shock moment for the impulse-response, i.e., evidence of learning.

Figure 7 shows that t as one goes from group 4—the group of suppliers closest to the downstream sectors—to group 1—the most upstream suppliers—the gap between the dashed and solid lines first widens and then closes. For group 3, this gap also first widens over time and then eventually closes. These patterns are consistent with our theory's implication, as discussed in section 3.3.2.

Figure 7. Cumulative impulse response



4.4 Demand Shocks' Quantitative Importance for Sectoral Fluctuations

In this section, we estimate the demand shock process from the data and demonstrate the quantitative importance of persistent demand shocks in a time-to-build environment for explaining sectoral fluctuations along the supply chain. Specifically, we estimate a variance-covariance ma-

trix Σ_u for persistent shocks and Σ_ϵ for transitory shocks. We assume a common rate of decay ρ across sectors, and we assume the shocks are correlated across sectors following a factor structure, with one factor for each demand shock component:

$$\mathbf{\Sigma}_u = \boldsymbol{\lambda}_u \sigma_u^F \boldsymbol{\lambda}_u^{'} + \widetilde{\mathbf{\Sigma}}_u, \quad \mathbf{\Sigma}_\epsilon = \boldsymbol{\lambda}_\epsilon \sigma_\epsilon^F \boldsymbol{\lambda}_\epsilon^{'} + \widetilde{\mathbf{\Sigma}}_\epsilon,$$

where $\boldsymbol{\lambda}_u$ is the 25×1 vector of loadings of persistent demand shocks on the persistent factor, $\boldsymbol{\lambda}_\epsilon$ is the vector of loadings on the transitory factor, and $\widetilde{\mathbf{\Sigma}}_u$ and $\widetilde{\mathbf{\Sigma}}_\epsilon$ are diagonal matrices capturing idiosyncratic shocks across sectors.

Following Aguiar and Gopinath (2007), we estimate all parameters using GMM through minimizing the squared difference between the theoretical and data moments. We choose the variance and auto-covariance of sectoral revenue as the moments to match.

Table 2 shows the calibration results. We find $\rho = 0.802$. For the ease of interpretation, we report the kalman gain in the table instead of the variance-covariance matrices of the two shock components. A higher kalman gain indicates a larger share of a downstream sector's demand variation tend to be persistent. We find that "Automobiles and Light Duty Motor Vehicles", "Major Electrical Household Appliances", "Heavy Duty Trucks" and "Motor Vehicle Bodies and Trailers" are downstream sectors most exposed to persistent shocks.

Counterfactuals: Upstream Sectoral Fluctuations during 2020-2023 We now compute counterfactuals using our estimated model. Specifically, we perform the exercises of changing all demand shocks to be persistent (or changing all shocks to be transitory) in selected downstream sectors—holding the total demand variability constant—and we investigate the change in the implied sectoral volatility along the associated supply chains.

Table 3 shows the results. When all demand shocks are persistent in the the four downstream sectors of the transportation category, the volatility of the 27 sectors in the transportation supply chains almost doubles. By contrast, when only transitory shocks are present, the volatility in the supply chain decrease by 10%.

We conduct the exercise for three other categories of sectors: food products (8 downstream sectors and 21 along the supply chain), household goods (4 downstream sectors and 29 along the supply chain), and medical products (2 downstream sectors and 4 along the supply chain). Across the board, making demand shocks in downstream sectors more persistent significantly contribute to heightened volatility along the supply chain, whereas making demand shocks more transitory reduces volatility. The magnitude of the effects are significantly larger for the transportation category, especially as demand shocks become more persistent. We think this is partly because these sectors have longer supply chains—with more average time-to-build delays, as shown in the last column—and thus their suppliers' revenues are more sensitive to the nature (persistent

Table 2. Estimation results

$\rho = 0.802$	Kalman gain
Soft Drinks and Ice	0.01
Carpet and Rug Mills	0.02
Breweries	0.02
Office And Other Furniture	0.05
Tobacco	0.06
Navigational/Measuring/Electromedical/Control Instruments	0.08
Other Transportation Equipment	0.08
Coffee and Tea	0.08
Periodical, Book, and Other Publishers	0.09
Aerospace Products and Parts	0.12
Fruit and Vegetable Preserving and Specialty Foods	0.12
Sugar and Confectionery Products	0.15
Animal Slaughtering and Processing	0.17
Medical Equipment and Supplies	0.25
Newspaper Publishers	0.29
Soap, Cleaning Compounds, and Toilet Preparation	0.29
Bakeries and Tortilla	0.35
Other Food Except Coffee and Tea	0.37
Apparel	0.38
Ship and Boat Building	0.59
Automobiles and Light Duty Motor Vehicles	0.71
Major Electrical Household Appliances	0.72
Heavy Duty Trucks	0.86
Motor Vehicle Bodies and Trailers	0.94

vs. transitory) of demand shocks.

Table 4 conducts an related exercise, but this time we change the nature of demand shocks in all 25 downstream sectors—again holding the total demand variability constant—and we examine the implied changes in volatility of four selected upstream sectors that supply broadly to many producers and are part of many supply chains. When all downstream demand shocks are persistent, the "Oil and Gas Extraction" sector experiences 36% increase in volatility. The volatility increases by 18% in "Electric Power Generation, Transmission and Distribution", by 30% in "Organic Chemicals", and by 25% in "Petroleum Refineries". When all downstream demand shocks are transitory, sectoral volatility declines in all four upstream sectors.

Table 3. Counterfactual change in supply chain volatility based on the nature of downstream demand shocks

Group	Downstream Sectors	Change in supply chain volatility (holding downstream demand variance fixed)		Mean delay
		Only persistent shocks	Only transitory shocks	
Transportation (27 sectors along the supply chain)	Ship and Boat Building	113%	-12%	3.35
	Automobiles and Light Duty Motor Vehicles	116%	-12%	3.79
	Heavy Duty Trucks	92%	-10%	3.89
	Motor Vehicle Bodies and Trailers	110%	-12%	3.59
Food (21 sectors)	Soft Drinks and Ice	32%	-3%	3.36
	Breweries	26%	-2%	3.38
	Coffee and Tea	30%	-3%	3.68
	Fruit and Vegetable Preserving and Specialty Foods	26%	-3%	3.43
	Sugar and Confectionary Products	25%	-2%	3.15
	Animal Slaughtering and Processing	32%	-3%	4.15
	Bakeries and Tortilla	19%	-2%	3.21
	Other Food Except Coffee and Tea	25%	-3%	3.15
Household goods (29 sectors)	Apparel	21%	-2%	3.46
	Carpet and Rug Mills	19%	-2%	3.36
	Major Electrical Household Appliances	10%	-1%	3.27
	Office and Other Furniture	14%	-1%	3.05
Medical (4 sectors)	Pharmaceuticals and Medicines	60%	-11%	3.68
	Medical Equipment and Supplies	18%	-1%	3.62

Table 4. Counterfactual change in upstream volatility based on the nature of downstream demand shocks

Selected upstream sectors	Only persistent shocks	Only transitory shocks
Oil and Gas Extraction	36%	-16%
Electric Power Generation, Transmission and Distribution	18%	-13%
Organic Chemicals	30%	-14%
Petroleum Refineries	25%	-12%

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Aguiar, Mark and Gita Gopinath, "Emerging market business cycles: The cycle is the trend," *Journal of political Economy*, 2007, *115* (1), 69–102.

Alfaro, Laura, Davin Chor, Pol Antras, and Paola Conconi, "Internalizing Global Value Chains: A Firm-Level Analysis," *Journal of Political Economy*, 2019, *127* (2), 508–559.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry, "Measuring the Upstreamness of Production and Trade Flows," *American Economic Review*, 2012, *102* (3), 412–416.

Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2015.

Da, Zhi, Umit G Gurun, and Mitch Warachka, "Frog in the Pan: Continuous Information and Momentum," *The review of financial studies*, 2014, *27* (7), 2171–2218.

Forrester, Jay W, "Industrial Dynamics: A Major Breakthrough for Decision Makers," *Harvard business review*, 1958, *36* (4), 37–66.

Humpherys, Jeffrey, Preston Redd, and Jeremy West, "A Fresh Look at the Kalman Filter," *SIAM review*, 2012, *54* (4), 801–823.

Jr, John B Long and Charles I Plosser, "Real Business Cycles," *Journal of Political Economy*, 1983, *91* (1), 39–69.

Kydland, Finn E and Edward C Prescott, "Time to Build and Aggregate Fluctuations," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1345–1370.

Lee, Hau L, Venkata Padmanabhan, and Seungjin Whang, "Information Distortion in A Supply Chain: The Bullwhip Effect," *Management science*, 1997, *43* (4), 546–558.

_ , Vineet Padmanabhan, and Seungjin Whang, "Comments on Ïnformation distortion in a supply chain: The bullwhip effect¨," *Management science*, 2004, *50* (12_supplement), 1887–1893.

Liu, Ernest and Aleh Tsyvinski, "A Dynamic Model of Input–Output Networks," *Review of Economic Studies*, 2024, p. rdae012.