Networks and Business Cycles

Wu Zhu* Yucheng Yang†

Job Market Paper

Click here for the latest version

Abstract

The speed at which the US economy has recovered from recessions ranges from months to years. We propose a model incorporating innovation network, production network, and cross-sectional shock and show that their interactions jointly explain large variations in the recovery speed across recessions in the US.

Besides the production linkages, firms learn insights on production from each other through the innovation network. We show that the shock's sectoral distribution plays a crucial role in its amplification and persistence when the innovation network takes a low-rank structure.

We estimate a state-space model of the cross-sectional technology shock and document a set of new stylized facts on the structure of the innovation network and sectoral distribution of the shock for the US. We show that the specific low-rank network structure and the time-varying sectoral distribution of the shock can well explain the large variation in the recovery speed across recessions in the US. Finally, to emphasize the prevalence of the channel, we explore the application of the theory in asset pricing.

JEL classification: D85, F36, G32, G33, G38

Keywords: Innovation Network, Production Network, Business Cycles, Asset Prices, Technology Spillover, Persistence, and Amplification.

*Corresponding author, Department of Economics, University of Pennsylvania, zhuwu@sas.upenn.edu. This is part of my Job Market Work, I am extremely grateful to my advisors Rakesh Vohra, Linda Zhao, and Frank Schorfheide for their invaluable guidance and support. We also greatly appreciate the comments and suggestions from Mark Aguiar, Harold Cole, Hanming Fang, Wayne Gao, Ben Golub, Mathew Jackson, Ernest Liu, Richard Rogerson, Elisa Rubbo, Kjetil Storesletten, Chris Sims, Mark Watson, Amir Yaron, and seminar participants. All errors are our own.

†Princeton University, Program in Applied and Computational Mathematics, yuchengy@princeton.edu

1 Introduction

The speed at which the US economy has recovered from recessions varies from months to years. Understanding the forces behind a sluggish recovery has been the focus of economists and policymakers [Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997; Brunnermeier, Eisenbach, and Sannikov, 2012; Fernald, Hall, Stock, and Watson, 2017; Bianchi, Kung, and Morales, 2019; Duval, Hong, and Timmer, 2019]. This paper shows that the innovation network and its interactions with the production network and the cross-sectional technology shock explain the large variation in recovery speed from recessions in the US[1]. In this paper, we interpret the technology shock as a shock to the technology progress such as new production methods or processes, and the cross-sectional shock as a vector with each entry being a shock to a sector[2].

Firms rely on each other not just to acquire others' inputs but also insights to improve their productivity. The channel for this is the innovation network, which contains the linkages between firms through knowledge flow [Jaffe, 1986; Bloom, Schankerman, and Van Reenen, 2013; Acemoglu, Akcigit, and Kerr, 2016b; Ahmadpoor and Jones, 2017]. When a firm is exposed to a shock to technology progress, it can be propagated along the input-output chain as well as the innovation network via what firms learn from the new technology. Whether the shock is significantly amplified and becomes persistent depends on the innovation network structure, the interactions between the innovation and production networks, and how the shock propagates through the networks. This paper elucidates these channels and shows that the interactions among the cross-sectional shock, innovation network, and production network play an important role in influencing the recovery speed from recessions.

We propose a dynamic general equilibrium model with multiple sectors incorporating a production network, an innovation network, and cross-sectional shocks. In the economy, sectors are linked through the input-output chain and production technology via the innovation network. We identify sufficient conditions under which the interactions among cross-sectional shocks, the innovation network, and the production network provide a channel through which the initial shocks persist, amplify, and diffuse throughout the economy, yielding a prolonged recovery process when the shocks are adverse. These sufficient conditions are empirically

[1]According to the NBER Business Cycle Dating Committee, there are vast differences in the time that the US economy takes to recover from the adverse shocks. Sometimes, it takes less than one year to recover, while in some episodes, it may take several years to recover from the initial adverse shocks. The reference dates of US business cycles can be found on this webpage.

[2] Consider an economy with J sectors, the cross-sectional shock at period t just refers to a vector $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, ...\epsilon_{Jt})$ with ϵ_{jt} the sector j's exposure to the shock.

measurable, identifiable, and alterable by policy interventions.

Formally, in an economy with J sectors, we first show that the impact of the cross-sectional shocks on future growth can be decomposed into J components, each component includes its amplification and persistence [3]. We show that the amplification can be fully captured by two sufficient statistics - the inner product between the eigenvector centrality of the innovation network and the cross-sectional shocks, and the inner product between sectoral eigenvector centrality of the innovation network and sectoral Katz-centrality in the production network. The first inner product captures how the cross-sectional shock (technology shock) propagates through the innovation network. It is a weighted-average shock with the weights being the sectors' importance in the innovation network [4], suggesting that the direction of the cross-sectional shock matters in the amplification. The second inner-product fully captures the interactions between the innovation and production networks.

The persistence of the initial technology shocks' impact on future growth depends on two forces. Consider a cross-sectional shock. On the one hand, the impact of the shock declines over time due to the depreciation effect if sectors do not learn from each other. On the other hand, sectors can deploy resources to learn and gain insights from the technology shock. This technology spillover could cancel out the depreciation effect. If the spillover effect is sufficiently large, the shocks' impact becomes very persistent.

We show that the spillover effect's strength depends on the sectoral distribution of the shock (i.e., the direction of the shock) and the eigenvalue distribution of the innovation network's adjacency matrix[5]. We first show that all sectors experience the same spillover effect when

[3] We say a time-dependent process $\{x_t, t \in \mathbb{N}\}$ can be decomposed into J components with various persistence and amplification if we can write x_t as

$$x_t = \sum_{k=1}^{J} c_k g_k^t$$

where g_k and the coefficient c_k are referred to as the persistence and the amplification of the k^{th} component, respectively. Consider the case $g_k \in (0,1)$, the k^{th} component becomes very persistent when $g_k \approx 1$.

[4] Here, in an economy with J sectors, consider the cross-sectional shock $\boldsymbol{\epsilon}_t$ at period t and the eigenvector centrality $\boldsymbol{v} = (v_1, v_2, ..., v_J)$ with v_k sector k' importance in innovation network, the inner product is defined as

$$(\boldsymbol{\epsilon}_t, \boldsymbol{v}) = \sum_{k=1}^{J} \epsilon_{kt} v_k.$$

[5] A network can be either represented as a graph or matrix. When we talk about the eigenvalues or eigenvectors of a network, we refer to those associated with its adjacency matrix. For details, see Section 2.

the shock's direction is parallel to an eigenvector of the innovation network. Specifically, all sectors' spillover effect becomes most potent when the shock's direction parallels to the leading eigenvector (i.e., the vector of eigenvector centrality) of the innovation network. In contrast, the spillover effect is weakest when the shock's direction parallels the eigenvector associated with the innovation network's smallest eigenvalue. When the magnitude of the most potent spillover effect roughly equals the depreciation effect, the shock's impact becomes very persistent.

Consider the case where the strongest spillover effect can roughly cancel out the depreciation effect. When the innovation network is low-rank such that the leading eigenvalue is much larger than the remaining ones, the shocks' impact becomes very persistent only if the shock parallels to the eigenvector centrality's direction. In contrast, the impact declines quickly if the shock follows other directions. As a result, the shock direction reveals information on the recovery path of the economy if the innovation network takes a low-rank structure.

Overall, the shock's impact on future growth can be significantly amplified and persistent only if two conditions are satisfied. First, the shock highly correlates with sectors' importance in the innovation network, and the sectors' importance in the innovation network highly correlates with sectors' importance in the production network. Second, the innovation-network structure is such that the most potent technology spillover effect can roughly cancel the depreciation effect. To lift the economy out of the slow recovery, one policy implication of the theory is to bailout important sectors in the innovation network to mitigate their exposures to the adverse shocks.

To evaluate the empirical importance of the channel documented here, we construct a US patent dataset traced back to 1919 and an input-output dataset back to 1951. Based on the patent dataset, we construct a proxy for the technology innovation and estimate the underlying innovation network at the three-digit NAICS level. Using the sectoral input and output data, we construct the production network to proxy for the sectoral input-output linkages. We estimate the remaining undetermined parameters of the innovation network using a state-space model.

We document a set of new facts. First, the innovation network of the US has a low-rank structure so that the leading eigenvalue is much larger in magnitude than the rest. For example, the second-largest eigenvalue is only 20 percent of the leading one. Second, the leading eigenvalue is large enough that the corresponding strongest spillover effect roughly cancels out the depreciation effect. Thus, the shock's impact becomes very persistent when the cross-sectional shock follows the direction of the innovation network's eigenvector centrality. Third, there is a large time-variation in the inner product between the shock and the eigenvector centrality of the innovation network. For example, during the great recession of 2008, important sectors in the innovation network suffered much more than their less important counterparts. This pattern reverses during the recessions of 1991 and 2001 when sectors in the center suffer much less than those in the periphery of the innovation network.

As another application, we examine the implications of the theory on asset prices. Bansal and Yaron [2004] model the expected consumption growth as one with a small but persistent component and refer to it as "long-run risk" of the consumption. They argue that the long-run risk in consumption is the key to several puzzles in the financial markets - equity premium, the risk-free rate, and the market return volatility. However, where the small but persistent component comes from has been a puzzle [Bansal and Yaron, 2004]. Our theory provides a channel to endogenize a time-varying, small but persistent consumption growth component in a networked economy. The persistent component becomes significant when the cross-sectional shock shift to one specific direction - sectors' importance vector.

Literature

This paper contributes to several strands of literature in macro, network economics, and asset pricing. It first contributes to the literature exploring the source of the persistent component of aggregate growth. There are two main narratives, one emphasizes the role of financial friction [Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997; Brunnermeier, Eisenbach, and Sannikov, 2012], another accrue the persistent component to the endogenous TFP [De Ridder and Teulings, 2017; Fernald, Hall, Stock, and Watson, 2017; Queralto, 2019; Bianchi, Kung, and Morales, 2019; Anzoategui, Comin, Gertler, and Martinez, 2019; Duval, Hong, and Timmer, 2019]. Our theory falls in the second category. Comin and Gertler [2006] documents there is a significant medium-term business cycle of post-WWII and attributes it to the endogeneous R&D as a response to high-frequency shocks. Anzoategui, Comin, Gertler, and Martinez [2019] argues that the productivity slowdown of the post-Great Recession of 2007-2009 reflects an endogenous reduction in productivity-enhancing investment activities - the creation of new technology through R&D and the diffusion of technologies via adoption. Bianchi, Kung, and Morales [2019] emphasizes the difference between debt and equity financing and argues that equity (debt) financing shocks are more important for explaining R&D (physical) investment. Thanks to the rare occurrence of sizable adverse aggregate shocks, it is hard to evaluate the importance of the channels proposed in the previous literature. Another main challenge is that they provide us no clues on the persistence and amplification of the initial shock from the cross-sectional information. Unlike those papers which model the economy as a single representative firm, our theory emphasizes the role of network structure - the interactions between the networks and cross-sectional shocks - in amplifying shocks and explaining persistence. The rich interactions between the network structure and the shock enable us to test the channel's importance directly.

In the macroeconometrics and business cycle literature, researchers usually implicitly assume that the aggregate growth contains long-term and short-term components with various loading without further economic justifications, for example, King and Watson [1996]; Hodrick and Prescott [1997]; Baxter and King [1999]; Müller and Watson [2018]. Our theory rationalizes these assumptions by showing that the interaction of cross-sectional shocks with the innovation and production networks allows us to decompose the impact of shocks on aggregate growth into components with various levels of persistence and loadings. The theory provides further insights on the source of the persistence - the technology spillovers, and the source of loadings - the inner products between sectors' importance in innovation network and the cross-sectional shock, and the Katz centrality in the production network.

This article also contributes to the recent literature on innovation network. Bloom, Schankerman, and Van Reenen [2013] proposes a new measure of technology spillover using patent citations across companies. Acemoglu, Akcigit, and Kerr [2016b] describes the innovation network using patent citation of the US since 1976, and document that the network is very stable and sparse and that the upstream sectors can predict the patent issuance of downstream sectors very well. Ahmadpoor and Jones [2017] shows a slow diffusion process of innovation using patents and publications. Our article follows Bloom et al. [2013] in the construction of the innovation network but studies the implications of the innovation network in business cycles in the dynamic context.

Technically, our theory builds on the production network literature Long and Plosser [1983]. Recent studies emphasizing the potential role of idiosyncratic shocks in networks include Horvath et al. [1998]; Horvath [2000]; Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi [2012]; Acemoglu, Akcigit, and Kerr [2016a]; Barrot and Sauvagnat [2016]; Atalay [2017]; Baqaee [2018]; Baqaee and Farhi [2019]. They argue that idiosyncratic shocks to industries' productivities have the potential to generate aggregate fluctuations. For example, Atalay [2017] find that the industry-specific shocks contribute to at least half of the aggregate volatility.

These researches examine the amplification effect in the static context and only focus on the production network. Our paper provides insights into the persistence and amplification of idiosyncratic shocks in the dynamic context by incorporating both production and innovation networks. More importantly, we show that the cross-sectional shock direction reveals no information on the future recovery path if there is no innovation network. However, in an economy with a low-rank innovation network, the shock direction reveals essential information on the economy's recovery path. Thus, the innovation network is an essential element to understand why the direction of the shock matters. Finally, several studies emphasize the role of network structure in propagating shocks. Our contribution is to propose a new set of sufficient statistics to fully summarize network structure's role in amplifying and persisting shocks. Similar tools on eigenvalue decomposition were employed in studying the optimal intervention in networks in a static context Galeotti, Golub, and Goyal [2020].

Finally, this paper provides a new channel to endogenize the long-run risk in a production economy from a networks' perspective. Several recent papers try to rationalize the long-run risk based on production [Garleanu, Panageas, and Yu, 2012; Gârleanu, Kogan, and Panageas, 2012; Kogan, Papanikolaou, and Stoffman, 2013; Kung and Schmid, 2015]. Kung and Schmid [2015] argue that R&D endogenously drives a small, persistent component in productivity that yields long-run uncertainty on economic growth. However, no paper has examined the asset pricing implications of technology networks, while our paper naturally links the predictable components of growth with cross-sectional shocks and asset prices.

The rest of the paper is organized as follows. Section 2 describes the basic setup of the production network, innovation network, preference, and results in general equilibrium. Section 3 discusses the main theoretical results, where we explore the linkage between the persistence of shocks, innovation network, production network, and the cross-sectional shocks. Section 4 presents the main results on model estimation. In section 4.4, we document several facts about the innovation network and its interactions with the production network and technological shock. Section 5 discusses the potential applications of our theory in other puzzles. Section 6 concludes the paper. The technical details, the generalization of our results, and the proofs are in the Appendix.

2 Model

2.1 Production of Intermediate and Final Products

We consider a production economy with standard input-output linkages following Long and Plosser [1983], which has been widely used to examine the macro implications of input-output linkages [Acemoglu, Akcigit, and Kerr, 2016a; Herskovic, 2018; Baqaee, 2018; Baqaee and Farhi, 2019]. In most cases, previous studies focus on a Cobb-Douglas production technology that allows for an analytical solution, with the exception of Baqaee and Farhi [2019] which examines the non-linearity effect of the micro shocks in a static setting using a second-order Taylor expansion.

Besides the input-output network in the production space, we also incorporate an innovation network to capture the technology linkage between sectors. Suppose there are $[J] = \{1, 2, ...J\}$ sectors for intermediate goods in the economy [6]. Each intermediate good is used as an input for the final consumption good and other intermediate goods. Denote $\boldsymbol{A_t} = (A_{1t}, ..., A_{Jt})$ as a joint process of the productivity driven by technology, and let $a_{it} = \log(A_{it})$. Denote by Y_{it} the output of sector i, I_{it} the composite input of sector i to produce its products, and X_{ijt} the input of sector i from sector j. At time t, sector i combines its own technology and the outputs of other sectors as inputs to produce

$$Y_{it} = A_{it} I_{it}^{\eta}, \;\; s.t. \;\; I_{it} = \Big[\sum_{j \in [J]} \theta_{ij} X_{ijt}^{1-1/v_i} \Big]^{\frac{1}{1-1/v_i}} \tag{1}$$

with $\eta \in (0, 1)$ captures decreasing return to scale. The production of the composite input I_{it} exhibits constant elasticity of substitution (CES). v_i is the elasticity of substitution for the production technology [7]. If $v_i > 1$, the inputs used by firm i are substitutes to each other. An increase in the price of X_{ijt} would induce firm i to substitute away from input j and reduce the share of i's expenditure on j. When $v_i < 1$, the inputs of firm i are complements to each other. Firm i can not flexibly substitute away input j as a response to price of X_{ijt}, leading to a rise in the share of i's expenditure on j when the price of X_{ijt} rises. When $v = 1$, the production technology is reduced to Cobb-Douglas form with $I_{it} = \prod_j X_{ijt}^{\theta_{ij}}$. The

[6] In the theory part, we use a firm or sector interchangeably. In the empirical part, we would provide evidence at both the firm and industry level.

[7] Specifically, at the optimum of firm i,

$$v_i = -\frac{\partial \log(X_{ijk}/X_{ikt})}{\partial \log(P_j/P_k)}$$

share of firm i's expenditure on input j at time t is constant. For the case of $v_i = 1$, $\forall i \in [J]$, we can obtain a closed-form solution to the whole system since the system is log-linear with constant sector shares [Baqaee and Farhi, 2019]. For the general case, we provide an analytical solution to the aggregate output.

Let $\boldsymbol{P}_t = (P_{1t}, ..., P_{Jt})$ be the price vector at period t for the J intermediate products, with P_{it} the price of intermediate good i. The cash flow or dividend from firm $i \in [J]$ at t is

$$D_{it} = \max_{X_{ijt}, I_{it}} P_{it} A_{it} I_{it}^{\eta} - \sum_j P_{jt} X_{ijt}, \;\; s.t. \;\; I_{it} = \Big[\sum_{j \in [J]} \theta_{ij} X_{ijt}^{1-1/v_i} \Big]^{\frac{1}{1-1/v_i}}. \tag{2}$$

Let λ_{it} be the shadow price of the composite input I_{it}, i.e., the Lagrange multiplier of the constraint in Equation (2). The first order condition implies

$$I_{it} : \eta P_{it} Y_{it} = \lambda_{it} I_{it} \Longrightarrow I_{it} = \Big[\frac{\eta A_{it} P_{it}}{\lambda_{it}} \Big]^{\frac{1}{1-\eta}} \tag{3}$$

and

$$X_{ijt} : P_{jt} = \lambda_{it} \theta_{ij} I_{it}^{1/v_i} X_{ijt}^{-1/v_i} \Longrightarrow X_{ijt} = I_{it} \Big[\frac{\theta_{ij} \lambda_{it}}{P_{jt}} \Big]^{v_i}. \tag{4}$$

To write down $I_{it}, X_{ijt}, j \in [J]$ as functions of price and other parameters, we need to pin down expression of shadow price λ_{it}. Plugging Equation 4 into Equation 2, we have

$$\begin{aligned} \lambda_{it}^{1-v_i} &= \sum_j \theta_{ij}^{v_i} P_{jt}^{1-v_i}, \quad \text{if } v_i \neq 1 \\ \lambda_{it} &= \prod_j (\frac{P_{jt}}{\theta_{ij}})^{\theta_{ij}}, \quad \text{if } v_i = 1. \end{aligned} \tag{5}$$

Given the dividend flow D_{it}, The value of firm $i \in [J]$ satisfies

$$V_{it} = D_{it} + E_t M_{t,t+1} V_{it+1} \tag{6}$$

where $M_{t,t+1}$ is the Stochastic Discount Factor (SDF) between periods t and $t + 1$[8]. For details on $M_{t,t+1}$, see Section 2.3.

The final consumption goods are produced by competitive firms with technology $C_t = \prod_i c_{it}^{\alpha_i}$, where c_{it} is the amount of good in sector i used to produce the final consumption good at t.

[8] In the literature, the SDF is also referred to as state price density.

The producer of final consumption goods solves for

$$\max_{C_t, c_{it}} C_t - \sum_i P_{it} c_{it}, \quad \text{with } C_t = \prod_i c_{it}^{\alpha_i} \tag{7}$$

At optimality, $\prod_j P_{jt}^{\alpha_j} = \prod_j \alpha_j^{\alpha_j}$ and $P_{it} c_{it} = \alpha_i C_t$. In this paper, we always normalize the price of final consumption good to be 1.

2.2 Innovation Network and Arrival Rate

In this section, we model the technology process $\boldsymbol{A}_t$ in a reduced form so that we can quickly dive into our main insights. In Appendix A.1, we provide a micro-founded model to justify this reduced form. Let $\Delta a_{it} = a_{it} - a_{it-1}$, $\boldsymbol{a}_t = (a_{1t}, ..., a_{Jt})'$, and $\Delta \boldsymbol{a}_t = (\Delta a_{1t}, ..., \Delta a_{Jt})$. We model the process a_{it} as an arrival process [Aghion and Howitt, 1992]:

$$\Delta a_{it} = \mu_{it} + \epsilon_{it}^A, \tag{8}$$

where μ_{it} is the arrival rate of new innovations for i between $t-1$ and t, and ϵ_{it}^A is the shock to the realization of innovation, with $\mathbb{E}[\epsilon_{it}^A] = 0$.

The main theme of this subsection is to model the underlying arrival rate of the innovation $\boldsymbol{\mu}_t = (\mu_{1t}, ..., \mu_{Jt})$. Sector i can learn insights from the new technology of other sectors, promoting the arrival rate of its own future innovation. We model the learning process as[9]

$$\mu_{it+1} = (1-\rho)\mu_{it} + \sum_j W_{ij} \Delta a_{jt} + \epsilon_{it}^u \tag{9}$$

In appendix A.1, we provide a heuristic micro foundation to this process with endogenous match and search, and R&D decision. $\boldsymbol{\epsilon}_t^u = (\epsilon_{1,t}^u, ..., \epsilon_{J,t}^u)$ is a joint stationary process with $\mathbb{E}_t \boldsymbol{\epsilon}_t^u = 0, \forall i \in [J]$, which will be further specified later. $(1-\rho)\mu_{it}$ is used to capture the depletion effect of new ideas as in Bloom, Jones, Van Reenen, and Webb [2020]. That is, in a model without learning ($W_{ij} = 0, \forall i, j$), the arrival rate of the new technology declines at the rate ρ. The term $\sum_j W_{ij} \Delta a_{jt}$ captures the technology diffusion among sectors. Here, we assume that only the innovation of last period, $\sum_j W_{ij} \Delta a_{jt}$, contributes to the arrival rate of new technology. The contribution of historical knowledge to $\mu_{i,t+1}$ are fully captured by

[9] This setup shares similar spirit to the empirical work by Acemoglu, Akcigit, and Kerr [2016b] where they use patent citation to estimate the sector-to-sector innovation network, and document that the patent issuance in upstream sectors can well predict the patent issuance in the downstream sectors in the innovation network in ten years horizon.

μ_{it}. In appendix, we extend our model to more general cases by allowing firms to learn from the historical innovations according to

$$\mu_{it+1} = (1-\rho)\mu_{it} + \sum_j W_{ij}\varphi(L)\Delta a_{jt} + \epsilon_{it}^u, \tag{10}$$

where $\varphi(L) = \sum_{s\geq 0}\varphi_s L^s$, with the lag-operator L and $\sum_{s\geq 0}\varphi_s = 1$[10].

In matrix notation, we write equations 8 and 9 as

$$\begin{aligned} \Delta \boldsymbol{a}_t &= \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t^A \\ \boldsymbol{\mu}_{t+1} &= (1-\rho)\boldsymbol{\mu}_t + \boldsymbol{W}\Delta\boldsymbol{a}_t + \boldsymbol{\epsilon}_t^u \end{aligned} \tag{11}$$

Consider several special cases of the process 11. Suppose there is no realization shock, $\boldsymbol{\epsilon}_t^A = 0$, the process is reduced to $\Delta\boldsymbol{a}_{t+1} = (1-\rho)\Delta\boldsymbol{a}_t + \boldsymbol{W}\Delta\boldsymbol{a}_t + \boldsymbol{\epsilon}_t^u$., with $\Delta\boldsymbol{a}_t = \boldsymbol{\mu}_t$.

1. Suppose there is no technology spillover, $W_{ij} = 0, \forall i, j \in [J]$. The process is reduced to $\Delta\boldsymbol{a}_{t+1} = (1-\rho)\Delta\boldsymbol{a}_t + \boldsymbol{\epsilon}_t^u$, which is a standard setup as in Onatski and Ruge-Murcia [2013]; Atalay [2017].

2. If we further assume $\rho = 1$, then the process is reduced to $\Delta\boldsymbol{a}_{t+1} = \boldsymbol{\epsilon}_t^u$ which is examined by Foerster, Sarte, and Watson [2011]

Assumption 2.1 *To guarantee the stationarity of the process $\boldsymbol{\mu}_t$, we assume $\lambda_{max}(\boldsymbol{W}) \leq \rho$, where $\lambda_{max}(\boldsymbol{X})$ is the largest eigenvalue of matrix $\boldsymbol{X}$.*

2.3 Consumer

The representative consumer chooses consumption, C_t, and the share-holding on j, ϕ_{jt}, to maximize her life time utility with Epstein-Zin preference [Epstein and Zin, 1989][11].

$$\begin{aligned} U_t &= \max_{C_t,\phi_{jt}} \left[(1-\delta)C_t^{1-1/\psi} + \delta(E_t[U_{t+1}^{1-\gamma}])^{\frac{1-1/\psi}{1-\gamma}}\right]^{\frac{1}{1-1/\psi}} \\ C_t + \sum_j \phi_{jt}(V_{jt} - D_{jt}) &= \sum_j \phi_{jt-1}V_{jt} \end{aligned} \tag{12}$$

where $\delta \in (0,1)$ is the discount rate of time preference. γ is the coefficient of relative risk averse and ψ is the elasticity of inter-temporal substitution (IES). We use E-Z preference

[10] That is, firms learn from historical innovation with distributed lags, for details on distributed lags model, see Griliches [1967]

[11] We use the general Epstein-Zin preference but not its special case, time-separable preference, such that we can explore the asset pricing implications of our theory.

that is standard in asset pricing literature so that the far future consumption growth can be reflected in the current asset price. Denote $\theta = \frac{1-\gamma}{1-1/\psi}$, and the sign of θ is determined by the magnitude of ϕ and γ. When $\gamma = \frac{1}{\psi}$, the utility function is simplified to the usual time-separable preference with constant relative risk averse (CRRA). Denote $W_t = \sum_j \phi_{jt-1} V_{jt}$ as the wealth at the beginning of the period, $G_{t+1} = C_{t+1}/C_t$ as the growth of consumption, and $R_{m,t+1} = \frac{W_{t+1}}{W_t - C_t}$ as the return of the wealth. Epstein and Zin [1989] shows that the gross return of any asset i satisfies

$$E_t\left[\delta^\theta G_{t+1}^{-\theta/\psi} R_{m,t+1}^{\theta-1} R_{i,t+1}\right] = 1 \tag{13}$$

with SDF $M_{t,t+1} = \delta^\theta G_{t+1}^{-\theta/\psi} R_{m,t+1}^{\theta-1}$, and the logarithm of SDF takes

$$m_{t+1} = \theta \log(\delta) - \theta/\psi \Delta c_{t+1} + (\theta - 1) r_{m,t+1} \tag{14}$$

with Δc_{t+1} and $r_{m,t+1}$ as the log of consumption growth and return of aggregate wealth.

2.4 General Equilibrium

Definition 2.1 *A general equilibrium is a set of price and choice vectors -* $\boldsymbol{P_t} = (P_{1t}, ..., P_{Jt})$, $X_{ijt}, i, j \in [J]$, Y_{it}, ϕ_{it}, $c_{it}, i \in [J]$,*and* C_t *such that:*

1. *Given the price* $\boldsymbol{P}_t$, X_{ijt} *and* $Y_{it}, i, j \in [J]$, *solve producers' problem 2.*
2. *Given the price* $\boldsymbol{P}_t$, $c_{it}, i \in [J]$ *solve producers' problem for final consumption goods 7.*
3. *Given the price* $\boldsymbol{P}_t, V_{jt}$, *the portfolio* $\phi_{jt}, \forall j \in [J]$, *and* C_t *solve consumer's problem 12.*
4. *All markets for intermediates clear,* $c_{it} + \sum_{j\in[J]} X_{jit} = Y_{it}, \ \forall \ i, t$.
5. *Stock markets clear,* $\phi_{jt} = 1, \forall j, t$.

2.5 Equilibrium Allocation

Here, we characterize the resource allocation in equilibrium. Because there is no fixed capital and all intermediates are perishable, prices of the spot markets, resource allocation across sectors, and final output can be solved statically as functions of $\boldsymbol{A}_t$. Specifically, given the productivity distribution $\boldsymbol{A}_t = (A_{1t}, ..., A_{Jt})$, the intermediate price vector $\boldsymbol{P}_t$, shadow price vector $\boldsymbol{\lambda}_t$, input-output matrix, output vector $\boldsymbol{PY}_t = (P_{1t}Y_{1t}, ..., P_{Jt}Y_{Jt})$, and aggregate output $Y_t = \sum_{j\in[J]} P_{jt} Y_{jt}$ can be determined.

Definition 2.2 *Define the input-output matrix in equilibrium as*

$$\tilde{\mathbf{\Theta}}_t = (\tilde{\theta}_{ijt})_{J\times J} \tag{15}$$

with $\tilde{\theta}_{ijt} = \frac{P_{jt}X_{ijt}}{P_{it}I_{it}}$ *as the input reliance of sector* i *on sector* j.

Denote sale share of sector i as $s_{it} = \frac{P_{it}Y_{it}}{\sum_{j\in[J]} P_{jt}Y_{jt}}$ that is a measure of sector i's importance in production space, the sale share vector as $\boldsymbol{s}_t = (s_{1t}, ..., s_{Jt})'$, and $\boldsymbol{\alpha} = (\alpha_1, .., \alpha_J)$. In the following, all quantities are evaluated in equilibrium.

Proposition 2.1 *In equilibrium, the sale share satisfies* $\boldsymbol{s}_t = (1-\eta)[\boldsymbol{I} - \eta\tilde{\mathbf{\Theta}}_t']^{-1}\boldsymbol{\alpha}$ *.When the production technology is Cobb-Douglas, i.e.,* $v_i = 1, \forall i \in [J]$, *we have* $\tilde{\mathbf{\Theta}}_t = \mathbf{\Theta}$, $\boldsymbol{s}_t = \boldsymbol{s}$, *with* $\mathbf{\Theta} = (\theta_{ij})_{J\times J}$ *and* $\boldsymbol{s} = (1-\eta)[\boldsymbol{I} - \eta\tilde{\mathbf{\Theta}}']^{-1}\boldsymbol{\alpha}$.

The proof for the general case is in appendix A.2. There are two things worth mentioning. First, the share s_{it} measures the importance of sector i in production space and is defined recursively as the weighted average of the importance of sectors who rely on sector i's output as input. To make this point more clearly, we can write $s_{it} = (1-\eta)\alpha_i + \sum_j s_{jt}\tilde{\theta}_{jit}$. $\boldsymbol{s}_t$ is also called the Katz centrality [Katz, 1953; Bonacich, 1987; Bonacich and Lloyd, 2001]. Second, when the nested production function takes the form of Cobb-Douglas, the substitution effect exactly cancels out the income effect when the input prices change. Consequently, the expenditure share of sector i on its input j is constant, i.e., $\tilde{\theta}_{ijt} = \theta_{ij}$, and the importance of each sector is constant over time. In general, the input-output matrix $\tilde{\mathbf{\Theta}}_t$, and share $\boldsymbol{s}_t$ depends on $\boldsymbol{A}_t$. For details, please see the appendix A.2.

Definition 2.3 *We define the adjusted input sparsity of sector* j *as*

$$N^{\theta}_{jt} = \sum_{i\in[J]} \tilde{\theta}_{jit}\log(\tilde{\theta}_{jit}) + \frac{v_j}{1-v_j}\sum_{i\in[J]} \tilde{\theta}_{jit}\log\left(\frac{\tilde{\theta}_{jit}}{\theta_{ji}}\right), \forall v_j \neq 1$$

with $N^{\theta}_{jt} = \sum_{i\in[J]} \theta_{ji}\log(\theta_{ji})$, $\forall v_j = 1$.

Intuitively, the adjusted input sparsity captures the input diversity of the sector and the extent to which sector j can substitute its inputs away from each other. For the Cobb-Douglas case, the adjusted input sparsity reduces to the usual one [Herskovic, 2018].[12].

[12] In Figure 11, we use the BEA input-output data together with sectoral gross output data to estimate the non-linear effect of adjusted sparsity on aggregate economic growth. More specifically, we decompose the adjusted sparsity into the component proposed by Herskovic [2018] and the remaining part, and estimate them separately.

The following proposition shows that, even in a general nested CES production economy, the growth of gross output can be decomposed into three components - concentration term, sparsity term, and the Hulten term [Hulten, 1978].

PROPOSITION 2.2 *Denote* $\boldsymbol{N}_t^\theta = (\boldsymbol{N}_{1t}^\theta, ..., \boldsymbol{N}_{Jt}^\theta)$, *in equilibrium.*

1. *Gowth of consumption is the same as that of output,* $\Delta c_{t+1} = \Delta y_{t+1}$.

2. *The gross output and prices take the following form:*

$$\begin{aligned} \log(Y_t) &= \boldsymbol{s}_t'\Big[-\log(\boldsymbol{s}_t) + \frac{\eta}{1-\eta}\boldsymbol{N_t^\theta} + \frac{1}{1-\eta}\log(\boldsymbol{A}_t)\Big] + \frac{\eta}{1-\eta}log(\eta) + \boldsymbol{\alpha}'\boldsymbol{\alpha} \\ \log(\boldsymbol{P}_t) &= (\boldsymbol{I} - \eta\tilde{\boldsymbol{\Theta}}_t')^{-1}\Big[(1-\eta)\log(\boldsymbol{s_t}) + (1-\eta)\boldsymbol{1}\log(Y_t) - \eta\boldsymbol{N}_t^\theta - \log(\boldsymbol{A}_t) - \eta\log(\eta)\boldsymbol{1}\Big]. \end{aligned} \tag{16}$$

When $v_i = 1, \forall i \in [J]$, *we have*

$$\begin{aligned} \log(Y_t) &= \boldsymbol{s}'\Big[-\log(\boldsymbol{s}) + \frac{\eta}{1-\eta}\boldsymbol{N^\theta} + \frac{1}{1-\eta}\log(\boldsymbol{A}_t)\Big] + \frac{\eta}{1-\eta}log(\eta) + \boldsymbol{\alpha}'\boldsymbol{\alpha} \\ \log(\boldsymbol{P}_t) &= (\boldsymbol{I} - \eta\tilde{\boldsymbol{\Theta}}')^{-1}\Big[(1-\eta)\log(\boldsymbol{s}) + (1-\eta)\boldsymbol{1}\log(Y_t) - \eta\boldsymbol{N_t}^\theta - \log(\boldsymbol{A}_t) - \eta\log(\eta)\boldsymbol{1}\Big]. \end{aligned} \tag{17}$$

The proof for the general case is in appendix A.2. The intuition of the first part in Proposition 2.2 is that the consumption is equal to total dividend, which is a constant fraction of the total output. We emphasize the new results of this proposition - the aggregate output can always be decomposed into three components even with general nested CES production.

1. The first term, $\frac{1}{1-\eta}\boldsymbol{s}_t'\log(\boldsymbol{A}_t)$, is the usual linear term captured by Hulten's theorem [Hulten, 1978], [Baqaee and Farhi, 2019].

2. All the non-linear effects due to technology progress can be fully captured by two sufficient statistics - the concentration index $N_t^c = -\boldsymbol{s}_t'\log(\boldsymbol{s}_t)$, and the general adjusted sparsity $N_t^s = \boldsymbol{s}_t'\boldsymbol{N}_t^\theta$. The aggregate output increases with the concentration and the adjusted sparsity.

3. The concentration term reflects the resource reallocation across sectors while the sparsity term reflects the resource reallocation within sectors. Specifically, when resources are more concentrated in one sector, the concentration N_t^c declines, and thus aggregate output drops.

Note that both the adjusted sparsity and the concentration depends on $\boldsymbol{A}_t$. Thus, the impact of the technology shock $\Delta \boldsymbol{a}_{t+1}$ on aggregate output works through three channels - the usual Hulten channel $\boldsymbol{s}_t' \log(\boldsymbol{A}_t)$, the concentration channel $N_t^c = -\boldsymbol{s}_t' \log(\boldsymbol{s}_t)$ capturing resource reallocation across sectors, and the sparsity channel $N_t^s = \boldsymbol{s}_t' \boldsymbol{N}_t^\theta$ capturing the resource reallocation across inputs within the sectors. Previous literature has mostly taken first-order approximation by simply assuming Cobb-Douglas technology [Hulten, 1978; Long and Plosser, 1983; Acemoglu et al., 2016a; Herskovic, 2018]. Under the Cobb-Douglas assumption, the concentration index and the adjusted sparsity are constant, with no effect of technological shock on these two terms. Through taking second-order log-linear approximation, Baqaee and Farhi [2019] find that the usual first-order log-linear approximation significantly underestimates the macro effect of micro shocks, due to missing reallocation effects. Different from them, we find that all of the non-linear terms can be sufficiently captured by the concentration and adjusted sparsity, which can be estimated from real data. This enables us to directly measure the relative contribution of the concentration and adjusted sparsity terms to aggregate fluctuations [Yang and Zhu, 2020].

3 Networks, Amplification, and Persistence

This section presents our main results on the interactions of cross-sectional shock, innovation network, and production network. These interactions allow us to decompose the effect of the initial shock on future growth into J components. Each component includes its amplification and persistence. We further show that persistence can be captured by the eigenvalues of the innovation network, while two sufficient statistics can characterize the amplification. This decomposition clearly shows how the structure of the cross-sectional shocks and networks matter in amplifying the shock and making it persistent.

3.1 Basic Results

We consider the Cobb-Douglas case that enables us to get a closed-form solution[13]. Let $g_{t+1} = \Delta y_{t+1}$, which is the growth of aggregate output:

$$g_{t+1} = \frac{1}{1-\eta} \boldsymbol{s}' \Delta \boldsymbol{a}_{t+1}. \tag{18}$$

[13] For the general case, we still can make a similar decomposition through log-linear approximation.

The forward expectation of period $t+\tau$ at period t would be

$$\mathbb{E}_t g_{t+\tau} = \frac{1}{1-\eta}\mathbb{E}_t \boldsymbol{s}'\Delta\boldsymbol{a}_{t+\tau} = \frac{\boldsymbol{s}'}{1-\eta}\mathbb{E}_t\boldsymbol{\mu}_{t+\tau}, \tag{19}$$

where $\mathbb{E}_t[\cdot]$ is the expectation conditional on $\{A_s, s\le t\}$. Equation 19 follows from 11 and the condition that $\mathbb{E}_t\boldsymbol{\epsilon}^A_{t+\tau} = 0, \forall\tau\ge 1$. From equation 11, we have

$$\mathbb{E}_t\boldsymbol{\mu}_{t+\tau} = [(1-\rho)\boldsymbol{I}+\boldsymbol{W}]\mathbb{E}_t\boldsymbol{\mu}_{t+\tau-1} = [(1-\rho)\boldsymbol{I}+\boldsymbol{W}]^\tau\boldsymbol{\mu}_t. \tag{20}$$

Thus, an initial shock to $\boldsymbol{\mu}_t$ will change the growth prospects of period $t+\tau$, the persistence of this effect depends on the structure of the innovation network $\boldsymbol{W}$[14]. Suppose there is an initial shock to the arrival rate: $\boldsymbol{\mu}_t \to \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t$, the associated effect of this initial shock on the growth and the arrival rate at $t+\tau$ is denoted as $\delta g_{t+\tau}$ and $\delta\boldsymbol{\mu}_{t+\tau}$[15].

Proposition 3.1

$$\mathbb{E}_t\delta\boldsymbol{\mu}_{t+\tau} = [(1-\rho)\boldsymbol{I}+\boldsymbol{W}]^\tau\boldsymbol{\epsilon}_t$$

Under Cobb-Douglas, the effect on the growth is

$$\mathbb{E}_t\delta g_{t+\tau} = \frac{\boldsymbol{\epsilon}'_t}{1-\eta}[(1-\rho)\boldsymbol{I}+\boldsymbol{W'}]^\tau\boldsymbol{s}_t.$$

Alternatively, we have that $\mathbb{E}_t\delta g_{t+\tau} = \boldsymbol{\epsilon'}_t[(1-\rho)\boldsymbol{I}+\boldsymbol{W'}]^\tau[\boldsymbol{I}-\eta\tilde{\boldsymbol{\Theta}}']^{-1}\boldsymbol{\alpha}$.

3.1.1 Symmetric Technology Diffusion Matrices

We first consider a symmetric innovation network which has been implictly assumed in empirical studies $\boldsymbol{W}$ Jaffe [1986]; Jaffe, Trajtenberg, and Henderson [1993]; Bloom, Schankerman, and Van Reenen [2013] to deliver our main intuition. The asymmetric case will be discussed the subsection 3.1.2.

Definition 3.1 *Denote the i^{th} largest eigenvalue of matrix $\boldsymbol{W'}$ as $\lambda_i(\boldsymbol{W'})$, and the corresponding orthonormal eigenvector as $\boldsymbol{v}_i$, i.e. $\boldsymbol{W'}\boldsymbol{v}_i = \lambda_i(\boldsymbol{W'})\boldsymbol{v}_i$ such that $(\boldsymbol{v}_i,\boldsymbol{v}_j)=\delta_{ij}$.*

[14] It is also easy for us to calculate the conditional volatility of growth prospect

$$Var_t(\boldsymbol{\mu}_{t+s}) = \sum_{j=1}^{s}(\boldsymbol{I}+\tilde{\boldsymbol{W}})^{s-j}(\sigma_u^2+\sigma_A^2\boldsymbol{W}\boldsymbol{W'})(\boldsymbol{I}+\tilde{\boldsymbol{W}})^{s-j} \tag{21}$$

[15] Here, δ is an operator indicating a change, don't confuse it with a scaler. This shock can be either from the realization shock of innovation $\boldsymbol{z}_t^A$ or $\boldsymbol{z}_t^u$

Here, $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the inner product of vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and δ_{ij} is the Kronecker delta. Note that, in general case, we always have $\lambda_i(\boldsymbol{W}) = \lambda_i(\boldsymbol{W}')$.

LEMMA 3.1 *Suppose $\boldsymbol{W}$ is symmetric, we have $\boldsymbol{v}_1 \geq 0$, i.e., $v_{1i} \geq 0,\ \forall i \leq J$.*

LEMMA 3.2 *Suppose $\boldsymbol{W}$, $\lambda_i(\boldsymbol{W})$, and $\boldsymbol{v}_i$ are defined as in 3.1, then,*

$$((1-\rho)\boldsymbol{I} + \boldsymbol{W}')^s \boldsymbol{v}_i = (1 + \lambda_i(\boldsymbol{W}) - \rho)^s \boldsymbol{v}_i$$

The symmetry of $\boldsymbol{W}$ implies that $(\boldsymbol{v}_i, i \in [J])$ spans the whole J dimensional linear space. Thus, we can decompose the arrival rate as a linear combination of the eigenvectors $\boldsymbol{v}_i$.

LEMMA 3.3 *Suppose $\boldsymbol{W}$ is symmetric, we have $\boldsymbol{\mu}_t = \sum_{i=1}^{J} (\boldsymbol{\mu}_t, \boldsymbol{v}_i)\boldsymbol{v}_i$ with the loadings on $\boldsymbol{v}_i$ be $(\boldsymbol{\mu}_t, \boldsymbol{v}_i)$. Similarly, we have $\boldsymbol{s}_t = \sum_{i=1}^{J} (\boldsymbol{s}_t, \boldsymbol{v}_i)\boldsymbol{v}_i$.*

PROPOSITION 3.2 *If $\boldsymbol{W}$ is symmetric,*

$$\mathbb{E}_t \boldsymbol{\mu}_{t+\tau} = \sum_{i=1}^{J} (1 + \lambda_i(\boldsymbol{W}) - \rho)^\tau (\boldsymbol{\mu}_t, \boldsymbol{v}_i)\boldsymbol{v}_i$$

Several things are worth mentioning. First, we refer to, $\boldsymbol{v}_1$, the eigenvector associated with the leading eigenvalue, as the eigenvector centrality of the network [Bonacich and Lloyd, 2001; Bonacich, 2007; Allen et al., 2019]. In our setting, it is intuitive to interpret $v_{1,i}$ as sector i's importance in creating and transmitting technology insights in technology space. To see it clearly, we note that $v_{1,i} = \frac{1}{\lambda_1(\boldsymbol{W})} \sum_j v_{1,j} W_{ji}$. That is, sector i is important if sectors who heavily learn insights from i are important.

Second, $\lambda_i(\boldsymbol{W}) - \rho$ captures the declining rate of the shock in the innovation network if the shock is parallel to $\boldsymbol{v}_i$. To see that clearly, suppose the initial shock $\boldsymbol{\epsilon}_t$ is proportional to $\boldsymbol{v}_i$, the associated effect of the initial shock on the arrival rate of next period is $[1-\rho+\lambda_i(\boldsymbol{W})]\boldsymbol{\epsilon}_t$. The intuition is straightforward. On one hand, the arrival rate of innovation declines by $\rho\boldsymbol{\epsilon}_t$ due to the depreciation effect [Bloom et al., 2020]. On the other hand, the arrival rate of innovation will be promoted by $\boldsymbol{W}\boldsymbol{\epsilon} = \lambda_i(\boldsymbol{W})\boldsymbol{\epsilon}_t$ due to the technology spillover from learning if the shock is proportional to $\boldsymbol{v}_i$. Overall, the net effect would be $(\lambda_i(\boldsymbol{W}) - \rho)\boldsymbol{\epsilon}_t$.

Third, the strength of the technology spillover depends on the shock's direction. When the shock follows the eigenvector centrality (the eigenvector associated with the leading eigenvalue of the innovation network), the technology spillover is the strongest and will be promoted by $\lambda_1(\boldsymbol{W})$. However, when the shock follows the eigenvector associated with the smallest eigenvalue, the spillover effect becomes the weakest. Overall, the direction of the shock matters in the strength of the spillover effect. Finally, we have a singular value decomposition of the symmetric matrix, $\boldsymbol{W'} = \sum_j \lambda_j(\boldsymbol{W})\boldsymbol{v}_j\boldsymbol{v}_j'$, with $\boldsymbol{v}_i'\boldsymbol{v}_j = \delta_{ij}$.

If we decompose the initial shock into a linear combination of $\boldsymbol{v}_i, i \leq J$, the effect of the initial shock on future arrival rate can be written as a decaying linear combinations of $(\boldsymbol{\epsilon}_t, \boldsymbol{v}_i)\boldsymbol{v}_i$. The component parallel to $\boldsymbol{v}_1$ declines at the slowest rate, $\rho - \lambda_1(\boldsymbol{W})$. Proposition 3.2 shows that the i^{th} component of the initial shock declines with the rate $\rho - \lambda_i(\boldsymbol{W})$.

PROPOSITION 3.3 *The expected effect of the cross-sectional shock $\boldsymbol{\epsilon}_t$ on future growth can be written as*

$$\mathbb{E}_t \delta g_{t+\tau} = \frac{1}{1-\eta}\sum_{i=1}^{J}(1-(\rho-\lambda_i(\boldsymbol{W})))^{\tau}(\boldsymbol{\epsilon}_t, \boldsymbol{v}_i)(\boldsymbol{v}_i, \boldsymbol{s}_t). \tag{22}$$

3.1.2 Asymmetric Technology Diffusion Matrices

When the matrix $\boldsymbol{W}$ is asymmetric, we assume the matrix is diagonalizable and its eigenvectors almost surely span the whole J dimension space. The assumption holds in empirical data as shown in section 4.4. Thus, we can take eigenvector decomposition

$$\boldsymbol{W'}\boldsymbol{v}_i = \lambda_i(\boldsymbol{W})\boldsymbol{v}_i, \ i \leq J \tag{23}$$

with $\rho > Re(\lambda_1(\boldsymbol{W})) > ... > Re(\lambda_J(\boldsymbol{W}))$. In section 4.4, we document that all eigenvalues are approximately real in the sense that the imaginary part of each eigenvalue is negligible compared to its real part. Under the diagonalizable assumption, we have results similar to 3.2 and 3.1

PROPOSITION 3.4 *Suppose $\boldsymbol{W}$ satisfies 23, we have*

$$\mathbb{E}_t \delta g_{t+\tau} = \frac{1}{1-\eta}\sum_{i=1}^{J}(1-\rho+\lambda_i(\boldsymbol{W}))^{\tau}(\boldsymbol{\epsilon}_t, \boldsymbol{v}_i)(\boldsymbol{v}_i, \boldsymbol{s}_t) \tag{24}$$

In macro econometrics or finance, researchers usually think of growth as a process consisting of several components with various frequencies (persistence) and loadings [King and Watson, 1996; Baxter and King, 1999; Müller and Watson, 2018], proposition 3.4 rationalizes this decomposition by showing that we can decompose the effect of the shock on growth prospect into J components with various levels of persistence and loadings. The main difference is that we can only decompose the expected shock's impact on the growth prospect into components but not the growth itself. The proposition further shows the sources of persistence and loadings.

3.2 Network Structure, Persistence, and Amplification

In this subsection, we discuss in detail the importance of the innovation structure and the direction of the cross-sectional in amplifying and persisting the effect on growth.

PROPOSITION 3.5 *PERSISTENCE*

1. *If $\rho >> \lambda_1(\boldsymbol{W})$, the depreciation effect dominates the strongest technology spillover effect, and the shock's impact declines exponentially.*

2. *If $\rho \approx \lambda_1(\boldsymbol{W})$, consider two cases:*

   2.1 *If the innovation network is low-rank so that $\lambda_1(\boldsymbol{W}) >> \lambda_i(\boldsymbol{W}), i \geq 2$, when the shock is parallel to $\boldsymbol{v}_1$, its effect on future growth declines linearly; however, when the shock is orthogonal to $\boldsymbol{v}_1$, its effect on future growth declines exponentially.*

   2.2 *If the innovation network is high-rank that $\lambda_1(\boldsymbol{W}) \approx \lambda_2(\boldsymbol{W})... = \lambda$. The effect of the shock always declines by $\rho - \lambda$ no matter the direction of the shock.*

We emphasize the importance of part 2 in the theorem 3.5, where the strongest spillover effect roughly cancels the depreciation effect. The theorem indicates that if the innovation network is low-rank, the shock's direction reveals useful information on the shock's persistence. Only when the shock follows a specific direction (leading eigenvector), the impact of the shock become persistent. However, when the shock follows other directions, the impact of the shock will decline quickly. The intuition is that the spillover effect becomes most potent only when the shock following the eigenvector centrality.

On the contrary, when the network is not low-rank. In the extreme case, all the eigenvalues are the same denoted as λ. The spillover effect will be the same no matter the direction of the shock. As a result, the direction of the shock reveals no information on the persistence of the impact. In fact, under this extreme case, we have

$$\mathbb{E}_t \delta g_{t+\tau} = \frac{1}{1-\eta}(1-(\rho-\lambda))^{\tau}\delta g_t \tag{25}$$

as long as the initial aggregate effect of the shocks are the same, the future pathes of various shocks will be exactly the same.

In the empirical portion of the paper, we show that the innovation network of the US has a low-rank structure. Besides, the strongest spillover effect roughly equals the depreciation effect. Thus, only when the shock follows the direction of eigenvector centrality, the impact becomes persistent empirically.

The amplification of the i^{th} component is captured by two inner products - the inner product between the shock and the i^{th} eigenvector of the innovation network, and the inner product between the i^{th} eigenvector of the innovation network and the Katz-centrality of the production network. The first inner product captures the amplification effect when shocks are propagated in the innovation network, while the second captures the interactions between the innovation and production networks. Consider the technology shocks to the oil production and exploration sector and the cloud computing sector to illustrate the idea. The oil sector is much larger and more important than the cloud sector in the production network [16]. However, few sectors learn insights from the oil sector's innovation, but many sectors gain insights to prompt their technology progress from the cloud's. These sectors may account for a significant share in production and be important in the innovation network. As a result, a shock to the cloud sector can have a more substantial and persistent impact on aggregate future growth than a similar shock to the oil sector. In the production network literature, the amplification of the technology shock can be captured by the Domar weight [Liu, 2019]. In our setting, the inner product between the Katz- centrality in the production network and the eigenvector centrality in the innovation network can be viewed as an alternative Domar weight but adjusted by the sector's importance in the innovation network.

3.3 Illustration: An Example

To emphasize the role of cross-sectional shocks in amplifying and persisting the effects of shocks, we present a simple example. The basic insight is that the cross-sectional shock reveals useful information on the recovery speed of the economy in the future beyond the aggregate shock. Consider an economy with three sectors $J = 3$ with Cobb-Douglas production technology and symmetric input-output production network. Thus, in equilibrium,

[16] In production network literature, a sector's importance in the production network is usually measured as its sale share [Baqaee and Farhi, 2019; Liu, 2019]. In the US, the oil sector's sale share is much larger than the cloud sector.

$\boldsymbol{s}_t = (1/3, 1/3, 1/3)$. We further set the matrix representation of the innovation network as

$$\boldsymbol{W} = \begin{bmatrix} 0.327 & 0.067 & 0.067 \\ 0.067 & 0.0467 & 0.0467 \\ 0.067 & 0.0467 & 0.0467 \end{bmatrix}$$

$\rho = 0.4$ and $\eta = 0.35$. W is chose to partly reflect the pattern in real data and partly simplify the analysis. The eigenvalues are $\lambda_1(\boldsymbol{W}) = 0.36, \lambda_2(\boldsymbol{W}) = 0.06$, and $\lambda_3(\boldsymbol{W}) = 0$. Correspondingly, $\rho - \lambda_1(\boldsymbol{W}) = 0.04$, $\rho - \lambda_2(\boldsymbol{W}) = 0.34$, and $\rho - \lambda_2(\boldsymbol{W}) = 0.40$. The associated eigenvectors are $\boldsymbol{v}_1 = \frac{1}{3\sqrt{2}}(4, 1, 1), \boldsymbol{v}_2 = \frac{1}{3}(-1, 2, 2), \boldsymbol{v}_3 = \frac{1}{2}(0, 1, -1)$. Consider two scenarios of cross-sectional shocks.

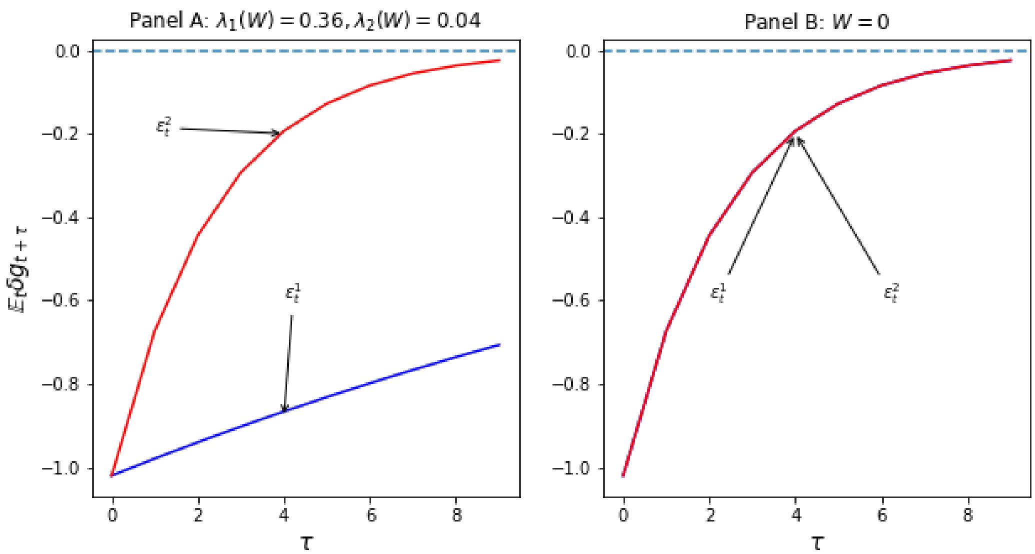
- Scenario 1: $\boldsymbol{\epsilon^1} = -1.5\boldsymbol{v}_1$.
- Scenario 2: $\boldsymbol{\epsilon^2} = -2.1\boldsymbol{v}_2$.

Under scenario 1, the cross-sectional shock is parallel to the eigenvector centrality $\boldsymbol{v}_1$, sectors in the center of the innovation network (sector 1) suffer much more than those in the periphery of the network (sectors 2 and 3). In contrast, under scenario 2, the cross-sectional shock is parallel to $\boldsymbol{v}_2$, sectors (sectors 2 and 3) in the periphery of the innovation network suffer more than those in the center of the innovation network (sector 1). In our setting, the aggregate effects on the current growth under the two scenarios are the same since

$$\frac{1}{1-\eta}(\boldsymbol{s}_t, \boldsymbol{\epsilon}^1) = \frac{1}{1-\eta}(\boldsymbol{s}_t, \boldsymbol{\epsilon}^2) = -1.0.$$

However, the impact of the cross-sectional shock $\boldsymbol{\epsilon}^1$ is much more persistent than that of $\boldsymbol{\epsilon}^2$. Figure 1 shows the effect of the shock on the growth in the next ten periods. As shown in the figure 1, the aggregate effects of the two shocks are the same at period 0 with a decline in aggregate output by 1%, while the effects on future growth are very different. Under scenario 2, the negative effect shrinks sharply to -0.2% after four periods. However, under scenario 1, the economy recovers very slowly and exhibits a staggering -0.75% growth even after ten periods.

Figure 1: The Effects of Cross-sectional Shocks on Expected Future Growth



Note: This figure shows the importance of cross-sectional shocks in persisting the effects of shocks on future growth. Panel A shows the shows the recovery path with various shocks when innovation network takes low-rank structure. Panel B shows the recovery path with various shocks when there is no innovation network.

4 Model Estimation and Evaluation

The basic propositions 3.4 and 3.5 show the network structure and the sectoral distribution of the shock play a crucial role in the amplification and persistence. The persistence of the shock depends on the relative magnitude of the technology spillover effect and the depreciation effect, while the amplification effect depends on $(\boldsymbol{v}_i, \boldsymbol{\epsilon}_t)$ and $(\boldsymbol{v}_i, \boldsymbol{s})$. In this section, we describe in detail the estimation of parameters in the innovation process 11 and use different datasets to estimate the technology shock from the model. Basically, in the data, we find the innovation network takes a low-rank structure and the strongest spillover effect roughly cancels out the depreciation effect, that is, $\lambda_1(\boldsymbol{W}) \approx \rho$ and $\lambda_1(\boldsymbol{W}) \gg \lambda_i(\boldsymbol{W}), i \geq 2$. This finding suggest that the direction of the shock reveals useful information on the recovery path of the economy as discussed in Section 3.3. On the amplification, we find that there is a large time-variation in $(\boldsymbol{v}_1, \boldsymbol{\epsilon}_t)$ and stable $(\boldsymbol{v}_1, \boldsymbol{s})$.

4.1 State Space Model

There are two main challenges in parameter estimation. First, we need some observable variables to proxy for the technology process $\Delta \boldsymbol{a}_t$. We consider three candidates - the patent, TFP, and R&D. All are imperfect proxies for the technology progress. Patents have been widely used in the literature to measure technology innovation. The main problem is that there is a time lag between invention usage and patent issuance released by the USPTO. This problem becomes severe when using the estimated shocks in equation 11 to approximate for $\boldsymbol{\epsilon}_t$ due to lagged issuance. The problem using TFP as a proxy is that TFP reflects not only technology progress but also many others like managerial skills, measurement errors, or residuals due to model misspecification. To handle this problem, we extend our model so that the observable TFP consists of an unobservable technology progress component following 11, a latent variable that reflect others factors influencing the measured TFP like the managerial skills, and an error term. That is, we model the unobservable technology process of the TFP as one with technology learning, and the second part of the TFP as a non-learning process. We then estimate an extended state-space-model. As shown in appendix A.1, R&D growth is a good proxy for the resources deployed by firms to learn from others. It is a good proxy for the technology progress in the future. Besides, the R&D data is too short, only available after 1988. Table 1 shows comparisons of these three datasets.

The second challenge is the matrix $\boldsymbol{W}$ is of high dimension even if we only consider the sector-to-sector innovation network at the three-digit level. Specifically, there are 87 sectors at the three-digit NAICS level. Thus, we have 7569 parameters to be estimated if we impose no restrictions on $\boldsymbol{W}$. Given our short patent dataset between 1926 and 2014, TFP dataset between 1987 and 2018, and R&D dataset between 1988 and 2018, we need to impose additional restrictions on $\boldsymbol{W}$ to estimate the underlying parameters. Specifically, we write $\boldsymbol{W} = \Xi\tilde{\boldsymbol{W}}$ with $\boldsymbol{\Xi} = \text{Diag}(\xi_1, ..., \xi_J)$, $\xi_i = \sum_j W_{ij}$, and $\tilde{W}_{ij} = \frac{W_{ij}}{\xi_i}$. We use patent citation dataset to directly estimate $\tilde{\boldsymbol{W}}$ as in appendix C.1.2.

We write 11 in the form of state space model, and estimate all parameters with an Expectation-Maximization (EM) algorithm. The problem is formulated as follows

$$\text{Measurement Equation: } \Delta \boldsymbol{a}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t^A, \text{ with } \boldsymbol{\epsilon}_t^A \sim \mathcal{N}(0, \boldsymbol{\Sigma}_A) \tag{26}$$

$$\text{State Equation: } \boldsymbol{\mu}_{t+1} = (1-\rho)\boldsymbol{\mu}_t + \varphi(L)\boldsymbol{W}_t \boldsymbol{\Delta a}_t + \boldsymbol{\epsilon}_t^u, \text{ with } \boldsymbol{\epsilon}_t^u \sim \mathcal{N}(0, \boldsymbol{\Sigma}_u) \tag{27}$$

with $\boldsymbol{W}_t = \boldsymbol{\Xi}\tilde{\boldsymbol{W}}_t$, $\boldsymbol{\Sigma}_A = \sigma_A^2 \boldsymbol{I}$, and $\boldsymbol{\Sigma}_u = \sigma_u^2 \boldsymbol{I}$. Here, we write $\boldsymbol{W}_t$ to indicate that we can estimate the time-varying $\tilde{\boldsymbol{W}}_t$ using patent citation dataset. With a little abuse of notation, we denote $\boldsymbol{A} = (1-\rho)\boldsymbol{I}$. $\varphi(L) = \sum_{j\geq 0} \varphi_j L^j$ is to capture the more general learning from history, with L the one-period lag-operator. The observable vector in period t is $\Delta \boldsymbol{a}_t$, the unknown parameters are scalars σ_A, σ_u, ρ, $J \times J$ diagonal matrix $\boldsymbol{\Xi}$.

Under the setting 26, the parameters of the model are denoted as $\boldsymbol{\Theta} = \{\rho, \varphi_j, \boldsymbol{\Xi}, \tilde{\boldsymbol{W}}_t, \boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_u\}$. To overcome the challenges from high dimensionality, we impose further restrictions on the parameters as shown in the table 2. For the $\varphi(L)$, we suppose that $\varphi(L) = \varphi_A \sum_{j\geq 0}(1 - \varphi_A)^j L^j$. In our state-space-model, we set $\varphi_A = 0.05, 0.1, 0.2, \text{and}, 1.0$, and find our estimates are robust across various φ_A.

Denote $\Delta\boldsymbol{a}_{0:t} = (\Delta\boldsymbol{a}_0, ..., \Delta\boldsymbol{a}_t)$, conditional mean and variance of $\boldsymbol{\mu}_t$ given data until time period τ as $\boldsymbol{\mu}_{t|\tau}$ and $\boldsymbol{P}_{t|\tau}$. Denote conditional variance of $\Delta\boldsymbol{a}_{t+1}$ given data until time period τ as $\boldsymbol{F}_{t|\tau}$, the forecast as $\Delta\boldsymbol{a}_{t+1|\tau}$, and the Kalman gain as $\boldsymbol{K}_t$. The EM algorithm to estimate σ_A, σ_u, ρ and $J \times J$ diagonal matrix $\boldsymbol{\Lambda}$ proceed as follows:

E Step Given current guess/estimates[17] of σ_A, σ_u, ρ and $\boldsymbol{\Lambda}$, use Kalman filter and Kalman smoother to calculate the conditional mean $\boldsymbol{\mu}_{t|T}$ and covariance $\boldsymbol{P}_{t|T}$ of latent states $\boldsymbol{\mu}_t$ given all the observed data.

E Step (1): Kalman filter Starting from $t = 0$ with initial guess $\boldsymbol{\mu}_{0|0}, \boldsymbol{P}_{0|0}$, while $t < T$ we calculate:

a. $\boldsymbol{\mu}_{t+1|t} = \boldsymbol{A}\boldsymbol{\mu}_{t|t} + \boldsymbol{W}\Delta\boldsymbol{a}_t$

b. $\boldsymbol{P}_{t+1|t} = \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}' + \boldsymbol{\Sigma}_u$

c. $\Delta\boldsymbol{a}_{t+1|t} = \boldsymbol{\mu}_{t+1|t}$

d. $\boldsymbol{F}_{t+1|t} = \boldsymbol{P}_{t+1|t} + \boldsymbol{\Sigma}_A$

e. $\boldsymbol{P}_{t+1|t+1} = \boldsymbol{P}_{t+1|t} - \boldsymbol{K}_{t+1}\boldsymbol{P}_{t+1|t}$

f. $\boldsymbol{\mu}_{t+1|t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{K}_{t+1}\boldsymbol{v}_{t+1}$ with $\boldsymbol{K}_{t+1} = \boldsymbol{P}_{t+1|t}\boldsymbol{F}_{t+1}^{-1}$.

E Step (2): Kalman Smoother Starting from $t = T$ with $\boldsymbol{\mu}_{T|T}, \boldsymbol{P}_{T|T}$ from the Kalman smoother, while $t >= 0$:

g. $\boldsymbol{L}_t = \boldsymbol{P}_{t|t}\boldsymbol{A}'\boldsymbol{P}_{t+1|t}^{-1}$

h. $\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t} + \boldsymbol{L}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t})$

i. $\boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t} + \boldsymbol{L}_t(\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t})\boldsymbol{L}_t'$

[17] We use the VAR estimates when $\sigma_A \equiv 0$ as the initial guess in our calculation. In later iterations, we always use the maximum likelihood estimators from the M step.

M Step: Given the conditional mean and covariance of latent states $\boldsymbol{\mu}_t$ and all the observed data, we estimate all the parameters $\boldsymbol{\Theta} = (\sigma_A^2, \sigma_u^2, \rho, \boldsymbol{\Xi})$ with maximum likelihood estimation. The estimators are

$$\widehat{\sigma_A^2} = \frac{1}{J(T+1)} Tr(\mathbb{E}(\sum_{t=0}^{T}(\Delta \boldsymbol{a}_t \Delta \boldsymbol{a}_t' - \Delta \boldsymbol{a}_t \boldsymbol{\mu}_t' - \boldsymbol{\mu}_t \Delta \boldsymbol{a}_t' + \boldsymbol{\mu}_t \boldsymbol{\mu}_t'))|\Delta \boldsymbol{a}_{0:T}) \tag{28}$$

$$\begin{aligned}\widehat{\sigma_u^2} =& \frac{1}{JT}\sum_{t=0}^{T-1} Tr(\mathbb{E}((\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)' - (\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)\boldsymbol{\mu}_t' \boldsymbol{A}' \\ & - \boldsymbol{A}\boldsymbol{\mu}_t(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)' + \boldsymbol{A}\boldsymbol{\mu}_t \boldsymbol{\mu}_t' \boldsymbol{A}' |\Delta \boldsymbol{a}_{0:T}))\end{aligned} \tag{29}$$

$$\widehat{\rho} = 1 - \frac{Tr(\sum_{t=0}^{T-1} \mathbb{E}[(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)\boldsymbol{\mu}_t' + \boldsymbol{\mu}_t(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta \boldsymbol{a}_t)' |\Delta \boldsymbol{a}_{0:T}]}{2\sum_{t=0}^{T-1} Tr(\mathbb{E}[\boldsymbol{\mu}_t \boldsymbol{\mu}_t' |\Delta \boldsymbol{a}_{0:T}])}) \tag{30}$$

$$\widehat{\xi}_i = \frac{\sum_{t=0}^{T-1} \mathbb{E}[(\tilde{\boldsymbol{W}}\Delta \boldsymbol{a}_t)_i (\boldsymbol{\mu_{t+1}} - \boldsymbol{A}\boldsymbol{\mu_t})_{\boldsymbol{i}} |\boldsymbol{\Delta a_{0:T}}]}{\sum_{t=0}^{T-1} \mathbb{E}[((\tilde{\boldsymbol{W}}\Delta \boldsymbol{a}_t)(\tilde{\boldsymbol{W}}\Delta \boldsymbol{a}_t)')_{ii} |\Delta \boldsymbol{a}_{0:T}]} \tag{31}$$

and $\widehat{\boldsymbol{\Xi}} = \text{diag}\{\widehat{\xi}_1, ..., \widehat{\xi}_K\}$. The detailed derivation can be found in the Appendix D.1. We iterate the **E Step** with new estimators of parameters, and the **M Step** with new conditional mean and covariance until convergence. The proof of convergence for the EM algorithm in our problem is in Appendix D.3.

Table 1: Comparison of Different Data to Measure Technology Progress

	Data Source	Time Range	Advantage	Shortcoming
Patent Issuance	USPTO	1926-2019	Long time range	Lag in measurement
TFP	BLS/BEA	1987-2018	Sectoral level measure	Short time range; more than innovation
R&D Expenditure	Compustat	1977-2018	Direct R&D measure	Sample bias change over time

This table presents the comparisons among three different datasets to Measure R&D Activities: the patent issuance data, industry level TFP data, and the Compustat R&D expenditure data.

Table 2: Model Setup and Restrictive Assumptions

Model description	Model parameter	Restrictions
Covariance of measurement noise	$\boldsymbol{\Sigma}_A$	$\boldsymbol{\Sigma}_A = \sigma_A^2 \boldsymbol{I}$
Covariance of shocks to innovation	$\boldsymbol{\Sigma}_u$	$\boldsymbol{\Sigma}_u = \sigma_u^2 \boldsymbol{I}$
Lag polynomial for learning	$\varphi(L) = \sum_{j\geq 0} \varphi_j L^j$	$\varphi(L) = \varphi_A \sum_{j\geq 0}(1-\varphi_A)^j L^j$ with $\varphi_A = 0.05, 0.1, 0.2, 1$
Learning matrix	$\boldsymbol{W}_t = \boldsymbol{\Xi}\tilde{\boldsymbol{W}}_t$	(1)$\tilde{\boldsymbol{W}}_t$ directly estimated with patent data (2)$\boldsymbol{\Xi} = \xi \boldsymbol{I}$ in the simplified case.

This table presents the restrictions we impose on the parameters when estimate the model. The restrictions are necessary to overcome the challenges of high dimensionality.

4.2 Estimates with Patent Dataset

Here, we discuss in detail the parameter estimation based on equation 27 using patent issuance data. For details about the estimation using TFP or R&D, see appendix B. From our model, it is natural to interpret $\boldsymbol{A}_t$ as the productivity driven by technology innovation. Denote N_{it} as the number of patents of sector i issued in year t, we proxy for A_{it} as

$$A_{it} = \delta_A \sum_{s \geq 0} (1 - \delta_A)^s N_{it-s}. \tag{32}$$

That is, we assume that the contribution of the patent to the productivity declines over time with depreciate rate δ_A. In estimation, we choose $\delta_A = 0.05$, i.e., patents' value on production depreciates to zero after 20 years, which is consistent with 20 years patent protection. Similar results on estimates are obtained for $\delta_A = 0.1$ or $\delta_A = 0.2$. Δa_{it} is estimated as $\log(A_{it}) - \log(A_{it-1})$.

We first present estimates under a simplified case where $\sigma_A = 0$ and $\boldsymbol{\Xi} = \xi \boldsymbol{I}$. The estimates under the simplified case will be used as an initial guess for the general case in EM algorithm. Under this simplified assumption, the model is reduced to

$$\Delta \boldsymbol{a}_{t+1} = (1 - \rho)\Delta \boldsymbol{a}_t + \lambda \boldsymbol{\Delta} \tilde{\boldsymbol{a}}_t + \boldsymbol{\epsilon}_t^u \tag{33}$$

with $\Delta \tilde{\boldsymbol{a}}_t = \tilde{\boldsymbol{W}}_t \varphi(L) \Delta \boldsymbol{a}_t$ capturing the network effects.

Under the simplified assumption, the state-space-model is reduced to the usual panel regression. Table 3 reports the results with $\varphi_A = 0.05, 0.1, 0.2, \text{and}, 1.0$. Columns 1 and 2 report the results when $\varphi_A = 0.05$. In column 1, we make seemingly OLS regression through pooling all observations together, while column 2 reports the estimates after controlling for the year and sector fixed effect. As shown in column 1, $\rho(= 0.315)$ is very close to $\xi(= 0.322)$, indicating that the depreciation effect will be roughly canceled out by spillover effect when the cross-sectional shock is parallel to the eigenvector centrality of the innovation network. Column 1 shows that $1 - \rho + \xi = 1.007$ that is very close to 1. Similar results are reported in column 2 where $1 - \rho + \xi = 1.002$. The results are quite robust for various other $\varphi_A = 0.1, 0.2, \text{and}\, 1.0$. In the case where $\varphi_A = 1.0$, $1 - \rho + \xi$ declines to 0.912 but still close to 1. One possible explanation is the spillover effect is under biased since sectors can only learn from the latest innovation when $\varphi_A = 1.0$.

For the general case without restrictions on $\boldsymbol{\Xi} = \text{diag}(\xi_1, \xi_2, ..., \xi_J)$, we estimate the state-space-model using EM algorithm. Table 4 reports the results. Panel A reports the EM estimates with homogeneous learning efficiency,i.e., $\xi_i = \xi$,while Panel B reports the EM estimates with heterogeneous learning efficiency. Panel A reports results for various φ_A. Column 1 shows results with $\varphi_A = 0.05$, we can see $1 - \rho + \xi (= 0.986)$ that is very close to 1. Furthermore, all results are robust across all columns. In panel B, we present the EM estimates without restriction on $\boldsymbol{\Xi}$. Besides the average of $\xi_i, i \in [J]$, standard deviation, 25^{th} and 75^{th} percentiles of $\xi_i, i \in [J]$ are reported. In the final row, we report the $1-\rho+Ave(\xi_i)$, where $Ave(\xi_i)$ is the average of ξ_i. As shown in the table, $1 - \rho + Ave(\xi_i)$ is very close to 1 and roust across various φ_A.

Table 3: Parameters of the Innovation Network

Panel Regression with assumption $\sigma_A = 0$ and $\boldsymbol{\Lambda} = \lambda \boldsymbol{I}$								
	$\varphi_A = 0.05$		$\varphi_A = 0.1$		$\varphi_A = 0.2$		$\varphi_A = 1.0$	
$1-\rho$	0.685***	0.643***	0.678***	0.641***	0.671***	0.639***	0.668***	0.641***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
λ	0.3218***	0.359*	0.299***	0.401***	0.287***	0.364***	0.244***	0.140*
	(0.0224)	(0.192)	(0.020)	(0.135)	(0.0019)	(0.104)	(0.017)	(0.072)
Sector Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
Year Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
σ_u	0.0442		0.0441		0.0441		0.0441	
R-Square	0.552	0.412	0.553	0.413	0.553	0.413	0.551	0.412

This table presents the parameter estimates under a simplified assumption that $\sigma_A = 0$. Under this assumption, our estimate equation is reduced to

$$\Delta \boldsymbol{a}_{t+1} = (1-\rho)\Delta \boldsymbol{a}_t + \lambda \tilde{\boldsymbol{W}} \varphi(L) \Delta \boldsymbol{a}_t + \boldsymbol{\epsilon}_t^u$$

with $\varphi(L) = \varphi_A \sum_{j \geq 0} (1-\varphi_A)^j L^j$. Columns 1-2 report the estimates with $\varphi_A = 0.05$. Besides reporting the estimates of $1-\rho$,λ, and σ_u based on our model, we also report the results with controlling for the sector and year fixed effects. Columns 3-8 report similar results under various value of φ_A. Robust standard errors in parentheses with *** p<0.01, ** p<0.05, * p<0.1.

Table 4: EM Estimates of the Innovation Network with Patent Data

Panel A: EM estimates with assumption that $\boldsymbol{\Xi} = \xi \boldsymbol{I}$				
	$\varphi_A = 0.05$	$\varphi_A = 0.1$	$\varphi_A = 0.2$	$\varphi_A = 1.0$
$1-\rho$	0.823	0.818	0.814	0.817
ξ	0.163	0.149	0.140	0.130
σ_u	0.0322	0.0324	0.0325	0.0316
σ_A	0.0237	0.0235	0.0234	0.0243
$1-\rho+\xi$	0.986	0.967	0.958	0.947
Panel B: EM estimates with general $\boldsymbol{\Xi}$				
	$\varphi_A = 0.05$	$\varphi_A = 0.1$	$\varphi_A = 0.2$	$\varphi_A = 1.0$
$1-\rho$	0.791	0.779	0.770	0.780
Mean of $\xi_j,\ j \in [J]$	0.198	0.187	0.182	0.162
Standard Dev of $\xi_j,\ j \in [J]$	0.160	0.141	0.131	0.113
25th percentile of $\xi_j,\ j \in [J]$	0.099	0.090	0.093	0.095
75th percentile of $\xi_j,\ j \in [J]$	0.298	0.264	0.266	0.238
σ_u	0.0332	0.0336	0.0339	0.0324
σ_A	0.0227	0.0222	0.0220	0.0235
$1-\rho+\xi(\xi = \text{Mean of } \xi_j,\ j \in [J])$	0.989	0.966	0.952	0.942

This table presents the parameter estimates using EM algorithm (for details, please see the appendix). In Panel A, we impose an assumption that $\boldsymbol{\Xi} = \xi \boldsymbol{I}$ - all sectors share the same parameter ξ. In Panel B, we remove this restriction and allow for heterogeneity in ξ across sectors. For the general case that $\boldsymbol{\Xi} = diag(\xi_1, ..., \xi_J)$ in Panel B, we also report the mean, standard deviation, 25th and 75th percentiles. In both panels, columns 1-4 report the results with $\varphi_A = 0.05$, 0.1,0.2, and 1.0.

4.3 Estimates with TFP Dataset

Since TFP not only captures the technology progress but also many others that influence firms' productivity like managerial skills. We slightly extend our model to

$$\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_t + \boldsymbol{m}_t, \tag{34}$$

with $\tilde{\boldsymbol{a}}_t = (\tilde{a}_{1t}, ..., \tilde{a}_{Jt})$ the log of the observable TFP at the three-digit NAICS level, $\boldsymbol{a}_t = (a_{1t}, ..., a_{Jt})$ the component driven by technology progress, and $\boldsymbol{m}_t = (m_{1t}, ..., m_{Jt})$ the component beyond the technology like managerial skills. We further assume $\boldsymbol{m}_t$ follows AR(1), $\boldsymbol{m}_t = \rho_m \boldsymbol{m}_{t-1} + \boldsymbol{\epsilon}_t^m$. That is, different from the technology side, there is no spillover effect for the process $\boldsymbol{m}_t$. For details on the estimation using TFP, see appendix B.

Since the TFP data is only available after 1987, much shorter than than the patent data, we further assume that $\boldsymbol{\Xi} = \xi \boldsymbol{I}$. Thus we only need to estimate $\rho, \rho_m, \xi, \sigma_u^2, \sigma_a^2, \sigma_m^2$. Using the state space estimation algorithm in Appendix D, the estimation results are reported in Table 5. Similar to the estimation results with patent data, $1-\rho+\xi$ is 0.976 that is very close to 1.

With the parameter estimates, we can recover the technology shock to the process $\boldsymbol{\epsilon}_t^u$, and investigate its inner product with the eigenvector centrality of the innovation network over time. In Figure 9, we plot the inner product of the technology shocks and the leading eigenvector of the innovation network over time. The grey shadow area indicate the NBER recessions. We can see that during the Great Recession of 2008, the negative TFP shock gets mostly amplified through the innovation network.

Overall, the estimate indicates that the strongest technology spillover effect will cancel out the depreciation effect. Consequently, when the cross-sectional shock highly correlates with the eigenvector centrality of the innovation network, the effect of the shock on future growth will become very persistent and amplified. If the shock is adversary, we will expect the economy to experience a prolonged recovery in the following years.

Table 5: EM Estimates of the Innovation Network with TFP Data

EM estimates with assumption that $\boldsymbol{\Xi} = \xi \boldsymbol{I}$		
	Estimates	Standard error
ρ	0.2227	0.006
ξ	0.1990	0.023
ρ_m	0.5062	0.006
σ_u	0.0001	2.14×10^{-5}
σ_a	0.0011	2.43×10^{-5}
σ_m	4.754×10^{-8}	4.43×10^{-5}
$1-\rho+\xi$	0.9763	-

This table presents the parameter estimates using EM algorithm (for details, please see the appendix). For simplicity, we impose an assumption that $\boldsymbol{\Xi} = \xi \boldsymbol{I}$ - all sectors share the same parameter ξ.

4.4 Empirical Facts

In the previous section, we show $\lambda_1(\boldsymbol{W}) \approx \rho$ - the strongest technology spillover roughly cancels out the depreciation effect. This sectional documents several facts about the structure of the innovation network, the coincidence between the innovation and production network, and the inner product between the sectors' importance in the innovation network and the cross-sectional shock.

Fact 1. The innovation network is low-rank in the sense that $\lambda_1(\boldsymbol{W}) >> \lambda_i(\boldsymbol{W}),\ i \geq 2$.

Fact 2. The innovation network has non-negligible overlapping over the production network in the sense that $(\boldsymbol{v}_1, \boldsymbol{s}_t) >> (\boldsymbol{v}_i, \boldsymbol{s}_t),\ i \geq 2$.

Fact 3. Across recessions in the US, recessions with slow recovery are those witnessing sizable negative shock to sectors in the center of the innovation network.

Based on our basic results in 3.4, fact 1 combined with $\lambda_1(\boldsymbol{W})$ suggests that the shock's impact will be persistent only if the shock follows the eigenvector centrality of the innovation. If the shock is orthogonal to the eigenvector centrality, the shock will decline exponentially roughly at the rate of ρ (since $\lambda_2(\boldsymbol{W}) << \rho$).

Fact 2 suggests that the amplification $(\boldsymbol{v}_1, \boldsymbol{s})(\boldsymbol{v}_1, \boldsymbol{\epsilon}_t)$ will be non-negligible if the shock follows the sectoral importance in the innovation network. Fact 3 is the key that is the main implication of our theory. That is, when the innovation network takes a low-rank structure, the strongest spillover effect roughly equals the depreciation effect, and $(\boldsymbol{v}_1, \boldsymbol{s})$ are non-negligible, the amplification and persistence depends on $(\boldsymbol{v}_1, \boldsymbol{\epsilon}_t)$.

4.4.1 Sparsity of the Innovation Network

This section documents the structure of the standardize innovation network $\tilde{\boldsymbol{W}}$, similar patterns are observed for $\boldsymbol{W}$ based on state-space model estimation. Figure 2 shows the heatmap of the matrix $\tilde{\boldsymbol{W}}$ of year 2014 at the three-digit NAICS level. First, there are several industries (e.g., Sector 334: Computer and Electronic Product Manufacturing; Sector 541: Professional, Scientific, and Technical Services[18]) playing a key role in providing technology insights to nearly all other sectors. Second, most sectors digest the insights from others but providing little knowledge to others(e.g., sectors between 441 and 512 in NAICS code).

Fact 1.1: The innovation matrix is highly asymmetric. Some sectors play critical roles in yielding insights to others, while others mainly digest knowledge.

[18] The industry description correspondence to three-digit NAICS code can be found in the NAICS manual on this webpage, last accessed on Jan 6, 2020.

Figure 3 shows histogram distribution on the eigenvalues of $\tilde{\boldsymbol{W}}$ in 2014. Note that the eigenvalues of the standardized matrix can be complex due to high asymmetry. The left panel shows the histogram on real part of eigenvalues, while the right panel shows the corresponding imaginary part. There are two basic facts:

> Fact 1.2: On the real part of the eigenvalues, $\lambda_1(\tilde{\boldsymbol{W}}) >> \lambda_i(\tilde{\boldsymbol{W}})$, $i \geq 2$. Furthermore, the eigenvalues are approximately real.

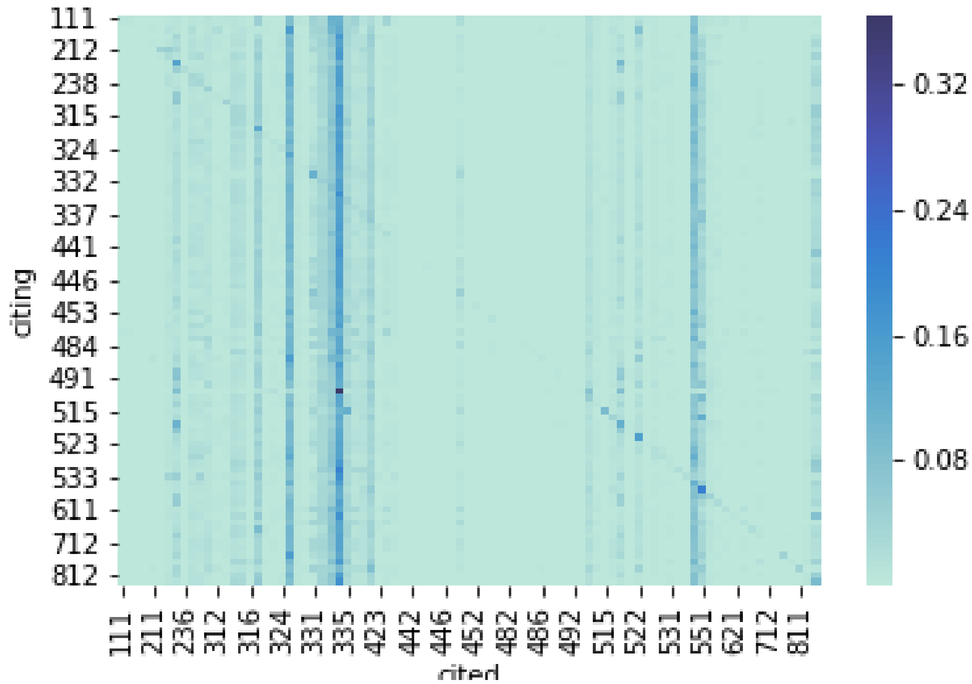
As shown in the left panel of figure 3, the distribution of the eigenvalues is low-rank that the largest eigenvalue is much larger than the rest in absolute value. Specifically, the second largest eigenvalue is roughly 20% of the largest one. Second, the imaginary part of the eigenvalue is tiny compared to its real part. As shown in the right panel of figure 3, the imaginary part is of the order of 10^{-5} that is negligible either relative to the real part or in the magnitude itself.

To show the concentration of the eigenvalue distribution for the matrix $\tilde{\boldsymbol{W}}$, we define a concentration measure as

$$Conc_j(\tilde{\boldsymbol{W}}) = \frac{\sum_{i \leq j} |\lambda_i(\tilde{\boldsymbol{W}})|}{\sum_i |\lambda_{i=1}^J(\tilde{\boldsymbol{W}})|}$$

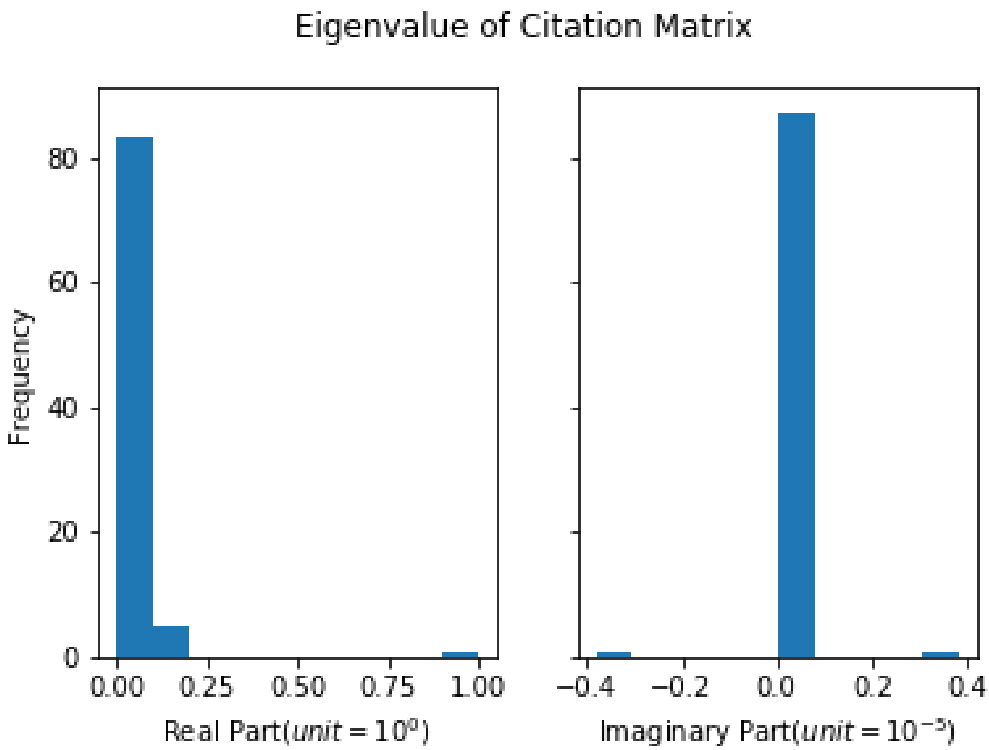
with $|\cdot|$ the norm of a complex number. $Conc_j(\tilde{\boldsymbol{W}})$ measures the fraction of the largest j eigenvalues to all eigenvalues. Figure 4 shows the time-varying $Conc_j(\tilde{\boldsymbol{W}}), j \in 1, 5, 10$ from 1951 to 2014. Over years, the concentration is quite stable. The largest eigenvalue contributes to 1/3, the largest five eigenvalues contribute to 60%, and the largest ten eigenvalues contribute to 75%, of the $\sum_j |\lambda_j(\tilde{\boldsymbol{W}})|$. In summary, one only needs to focus on the first several largest eigenvalues and the projection of the shock onto the associated eigenvectors if ρ and ξ are non-negligible.

Figure 2: Sparsity of Technology Spillover across Technology Classes



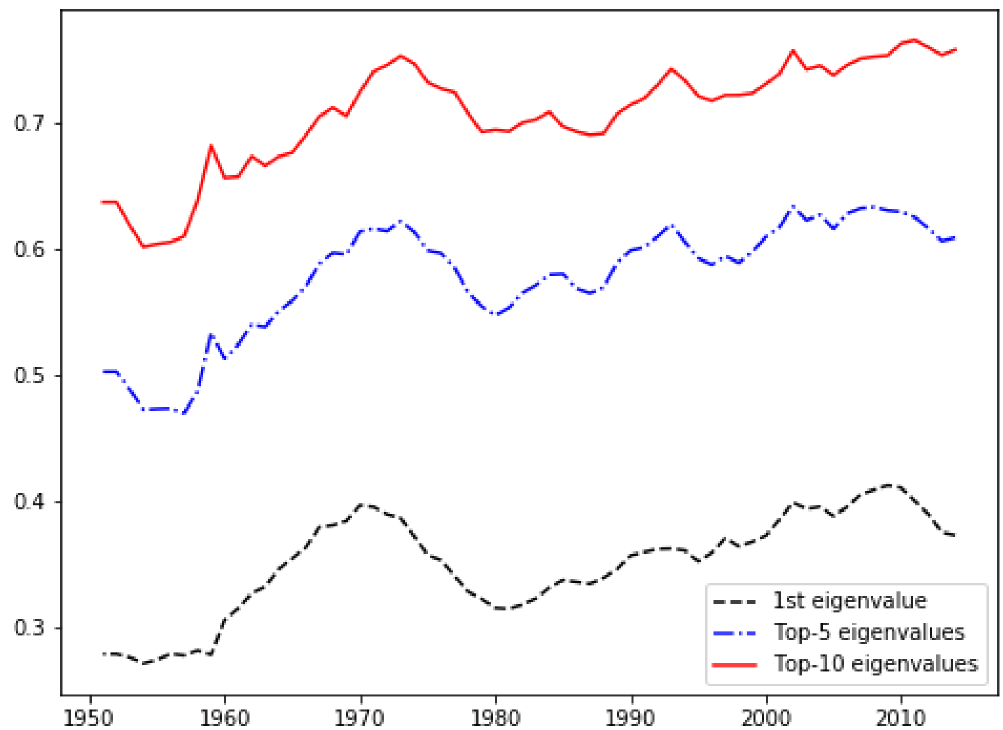
Note: This figure illustrates the knowledge diffusion matrix $\tilde{\boldsymbol{W}}$ in 2014. In year 2014, there are 87 sectors at three-digit NAICS level. X-axis represents sectors with knowledge flow out, and Y-axis represents sectors learning from others. The color indicates the magnitude of the knowledge flow $\tilde{W}_{ij}$, the deeper the color at (i, j) is, the larger the $\tilde{W}_{ij}$ is. For example, as shown in the figure, sector 334 (Computer and Electronic Product Manufacturing, see the NAICS Manual on this webpage for detail) and sector 541 (Professional, Scientific, and Technical Services) are very important in generating knowledge to others since nearly all other sectors intensively learn from it.

Figure 3: Distribution of Eigenvalues of the Standardized Innovation Network



Note: This figure illustrates the sparsity of the eigenvalue distribution of innovation diffusion matrix $\tilde{\boldsymbol{W}}$ in 2014. We standardize the diffusion matrix such that the largest eigenvalue is one. In year 2014, there are 87 sectors at three-digit NAICS level, and the matrix $\tilde{\boldsymbol{W}}$ is constructed at three-digit NAICS level. The left-panel of the figure shows the real parts of the eigenvalues and the right-panel shows the imaginary parts of the eigenvalues. We can see that the largest eigenvalue is much larger than the others in magnitude. The eigenvalues of the diffusion matrix are nearly real since the imaginary components are much smaller in magnitude compared to the real parts.

Figure 4: Concentration of the Eigenvalue Distribution for the Innovation Network



Note: This figure illustrates the sparsity on the eigenvalue distribution of innovation diffusion matrix $\tilde{\boldsymbol{W}}$ over years. We construct our innovation network each year from 1951 to 2014, and then standardize the matrix such that the largest eigenvalue of the diffusion matrix is one each year (for detail on the construction of $\tilde{\boldsymbol{W}}$, see the appendix C.1). Specifically, we calculate the concentration of eigenvalues as $Conc_i = \frac{\sum_{j \leq k} \xi_j(\tilde{\boldsymbol{W}})}{\sum_{j \leq k} \lambda_j(\tilde{\boldsymbol{W}})}, j \in \{1, 5, 10\}$ that measures the relative magnitude of the largest k eigenvalues to the rest.

4.5 Coincidence of Innovation and Production Networks

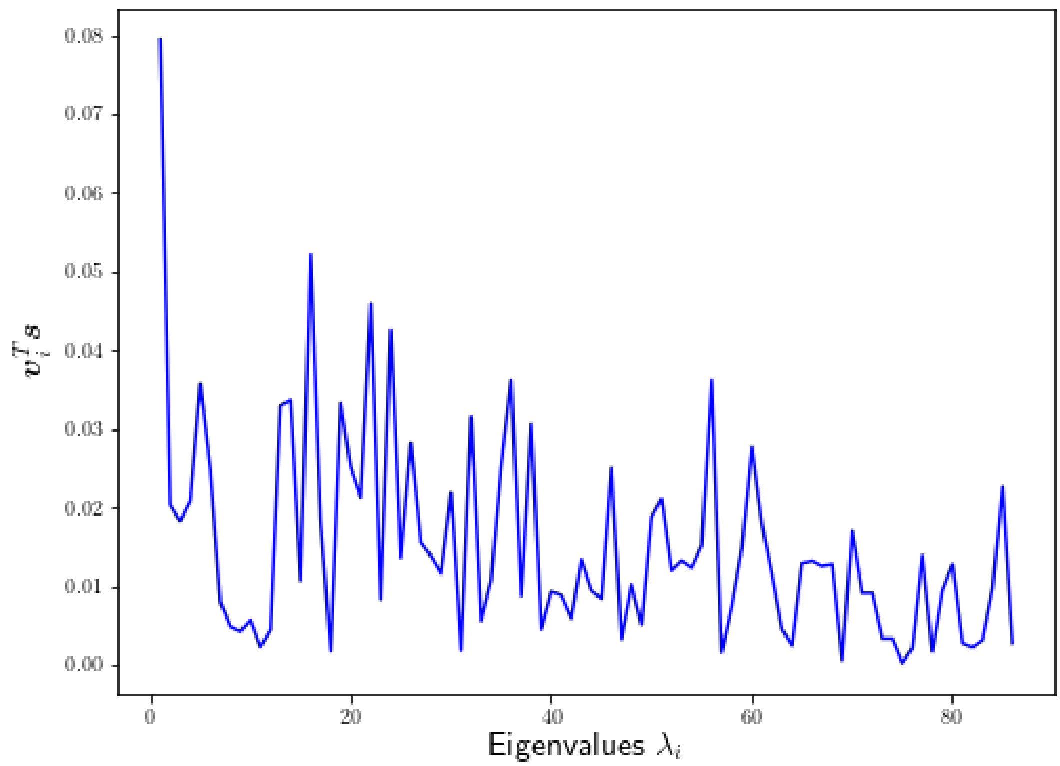
Proposition 3.4 shows that the amplification effect of the shock crucially depends on the inner product $(\boldsymbol{v}_i, \boldsymbol{s}_t)$ when the shock follows $\boldsymbol{v}_i$'s direction. This subsection documents the stylized facts on the inner product. As shown in 3.4 and 5.3, the amplification effect of the initial shock depends on the $(\boldsymbol{v}_i, \boldsymbol{s}), i \in [J]$. If $(\boldsymbol{v}_1, \boldsymbol{s}) \approx 0$, the net effect will still be small due to the tiny loading of the persistent component even though $\rho - \lambda_1(\boldsymbol{W}) \approx 0$. This section indicates that this is not the case.

> Fact 2.1: $(\boldsymbol{v}_1, \boldsymbol{s}) >> (\boldsymbol{v}_i, \boldsymbol{s}), i \geq 1$, this pattern is quite stable over time.

> Fact 2.2: $(\boldsymbol{v}_1, \boldsymbol{s})$ increases steadily from 0.06 in 1947 to 0.10 in 1980, and then declines to 0.08 recently.

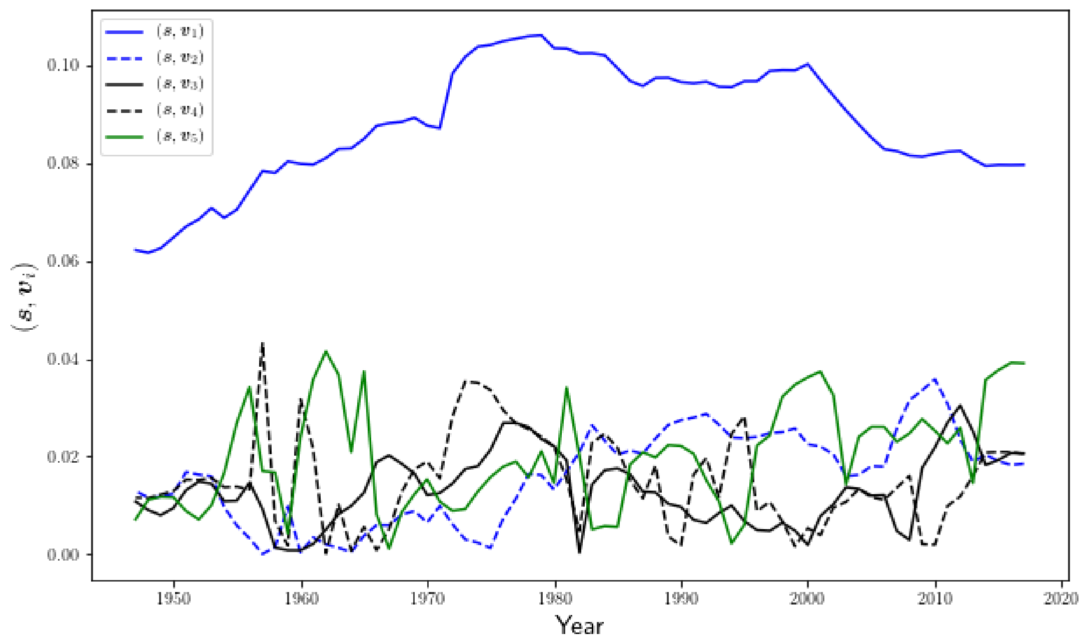
Note that if sectors were indifferent in the innovation network, $(\boldsymbol{v}_1, \boldsymbol{s})$ would have been $1/J \approx 0.011 << 0.06$. Thus, overall, important sectors in the innovation network are more likely to be important in the production network. Figure 5 shows $(\boldsymbol{v}_i, \boldsymbol{s}_t), i \in [J]$, of the year 2014. It shows that $(\boldsymbol{v}_1, \boldsymbol{s}_t)$ is much sizable than the other $(\boldsymbol{v}_i, \boldsymbol{s}_t), i \neq 1$. To check the trend of $(\boldsymbol{v}_i, \boldsymbol{s}_t)$ over time. Figure 6 shows the inner product between the sector share and the eigenvectors associated with the largest five eigenvalues from 1947 to 2017. Over the past 70 years, among all the inner products, $(\boldsymbol{v}_1, \boldsymbol{s}_t)$ is the largest one and much sizable than the rest. Another interesting finding is $(\boldsymbol{v}_1, \boldsymbol{s}_t)$ increases steadily from roughly 0.06 at the beginning of the 1950s to more than 0.1 in the 1980s and then declines to roughly 0.08 recently.

Figure 5: Coincidence between the Innovation and Production Networks



Note: This figure illustrates the coincidence of the eigenvectors of innovation diffusion matrices and firms' importances in production network - the inner product $(\boldsymbol{v}_i, \boldsymbol{s})$ - based on the knowledge diffusion matrix and production network in 2014 (for details, please see the appendix C.1). Both the innovation diffusion matrix and the output-share vector are constructed at three-digit NAICS level. $\boldsymbol{v}_i$ is the eigenvector associated with the *ith* largest eigenvalue of the knowledge diffusion matrix. $\boldsymbol{s}$ is the vector output share (Katz Centrality in the production network). For the eigenvectors associated with the first several largest eigenvalues, the inner product $(\boldsymbol{v}_1, \boldsymbol{s})$ is much larger than the rest.

Figure 6: Coincidence of the innovation and Production Network Overtime



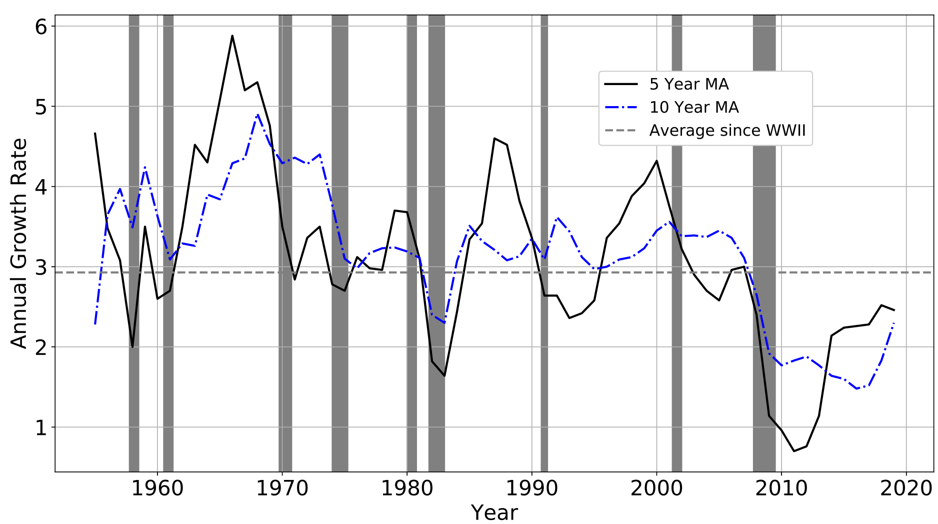
Note: This figure examines the coincidence of the eigenvectors of knowledge diffusion matrices and firms' importances in production network - the inner product $(\boldsymbol{v}_i, \boldsymbol{s})$ - over years from 1947 to 2014 (for details, please see the appendix C.1). At each year t, we construct output share of each sector at three-digit NAICS level based on the sectoral output from BEA or BLS, and the knowledge diffusion matrices at three-digit NAICS level based on patent datasets. The figure shows the time trend of the first five inner products $(\boldsymbol{s}, \boldsymbol{v}_i)$ associated with the largest five eigenvalues of the innovation diffusion matrix each year.

4.6 Coincidence between $\boldsymbol{v}_i$ and Shocks

In the previous sections, we have shown the strongest spillover effect cancels out the depreciation effect ($\boldsymbol{\lambda_1}(\boldsymbol{W} \approx \rho)$, the innovation network takes a low-rank structure ($\lambda_1((W)) >> \lambda_i((W)), \forall i > 1$), and large $(\boldsymbol{v}_1, \boldsymbol{s})$. If the channel in the proposition 3.4 does work, we expect the economy to experience a prolonged recovery if $(\boldsymbol{v}_1, \boldsymbol{\epsilon_t})$ experiences a large drop, that is, important sectors in the innovation network suffer much more than those less important ones.

Figure 7 shows the moving average GDP growth of the US since 1955. There is a considerable variation in the recovery period that the economy takes to move back to its average growth trend. For example, during the great recession of 2008, it takes more than ten years for the economy to move back to its average growth trend. However, in other episodes, the economy quickly recovers back to the growth trend like the recessions in the 1970s triggered by the oil crises.

Figure 7: Moving Average of Annual Real GDP Growth

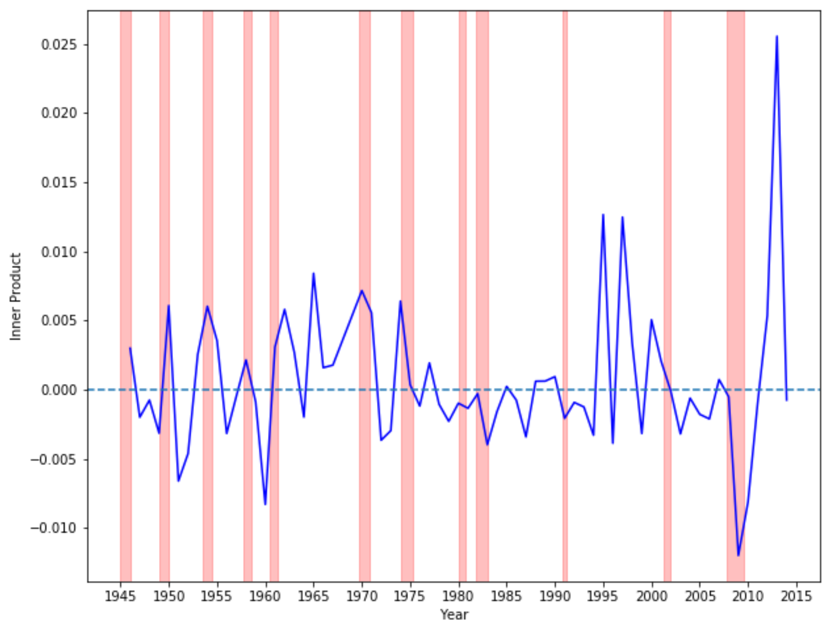


Note: This figure shows the moving average of the annual real GDP growth of the US.

To show the $(\boldsymbol{v}_1, \boldsymbol{\epsilon_t})$ across recessions, we estimate it based on the state-space model. As discussed in 4.1, each measure suffers from problems when we use the patent, TFP, and R&D to proxy for the current or future technological progress. We should be cautious about the interpretations. Specifically, using patent as a measure of the technology progress may be subject to a severe lagging problem, TFP may capture too much beyond the technology progress, while the R&D is a good proxy for the future technological progress but not now.

Figure 8 shows $(\boldsymbol{v}_1, \boldsymbol{\epsilon_t})$ across time but using the patents to proxy for technology progress as in 4.2. As we expected, during the great recession of 2008, the inner-product drops sharply but modest during other recessions. Another concern is the shock to the patent issuance is just a reflection of resources available to the US Patent Office.

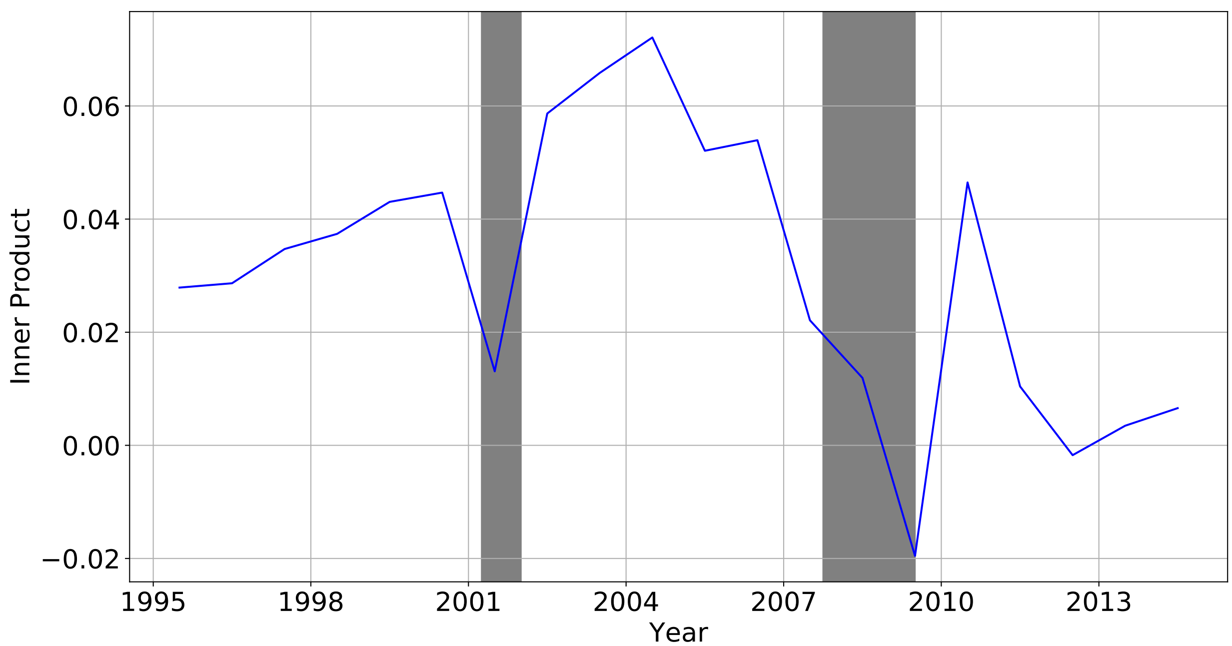
Figure 8: Inner Product between the Eigenvector Centrality and the Shock (Measured by Patent Issuance)



Note: This figure shows the inner product between the sectoral importance and the estimated technology shock. Here, we use the patent issuance to proxy for the technology progress and use the state-space-model to recover the underlying shock as discussed in section 4.2 .

To overcome this problem, Figure 9 plots the inner product between the sectoral importance in the innovation network and the shock. The cross-sectional shock is estimated as in section 4.3, where we explicitly model the TFP process as a combination of technological progress and the process beyond spillover. The shock used in figure 9 is the shock to technology progress. Similarly, we find that $(\boldsymbol{v}_1, \boldsymbol{\epsilon_t})$ experiences a significant drop to -0.02 during the great recession.

Figure 9: Inner Product between the Eigenvector Centrality and the Shock (Measured by TFP)



Note: This figure illustrates the coincidence of technological shocks and the leading eigenvector of the knowledge diffusion matrices over time. Both the innovation diffusion matrix and the technological shocks are constructed at three-digit NAICS level. The grey shadow area plots the NBER recessions. We can see that during the Great Recession, the negative technological shocks get mostly amplified through the knowledge diffusion network.

Does the drop along the direction of the sectoral importance matters at the aggregate level? Let us consider the great recession of the 2008 when $(\boldsymbol{v}_1, \boldsymbol{\epsilon_t}) \approx -0.02$ and $(\boldsymbol{v}_1, \boldsymbol{s}) \approx 0.10$, and use the equation for the first component in

$$\mathbb{E}_t \delta g_{t+\tau} = \frac{1}{1-\eta} \sum_{i=1}^{J} (1 - \rho + \lambda_i(\boldsymbol{W}))^{\tau} (\boldsymbol{\epsilon}_t, \boldsymbol{v}_i)(\boldsymbol{v}_i, \boldsymbol{s}_t) \tag{35}$$

The parameter on the return to scale is $\eta = 0.35$ following Herskovic [2018], and $\rho - \lambda \approx 0.02$. Thus, during the great recession of 2008, this shock leads to a nearly permanent drop in the GDP growth by 0.3 percent, the impact of this shock declines by half after 35 years. This aggregate effect is large compared to 3 percent of the long-run growth for the US. Another point is to compare with the deviation from the long growth trend. As shown in figure 7, ten years later, the moving average of the GDP growth is still lower by 0.4 percent relative to the growth trend. Thus, the 0.3 percent drop accounts for roughly 75% of such deviation.

4.7 Cross-Sectional Evidence across Recessions

In this subsection, we use the growth in R&D expenditure to proxy for the shock to the sector's arrival rate of innovation. This proxy is motivated by our micro foundation A.1, where we show that the R&D expenditure is proportional to the potential knowledge from which firms can learn. Intuitively, R&D expenditure is a good proxy for the arrival rate of future innovations.

Since the quarterly R&D is only available after 1988, we can only compare the cross-sectional change in R&D expenditure in three recent recessions - 1991, 2001, and the Great Recession of 2008. We first document that sectors playing critical roles in the innovation network suffer much worse, measured by the R&D growth, in the Great Recession than that in the recessions of 1991 and 2001. Specifically, we first calculate the R&D expense changes of sectors at the three-digit NAICS level, we then assign each sector into one of the five groups based on its eigenvector centrality, $\boldsymbol{v}_i$, of the last year. We construct dummies $Eigen_{itd}, d = 1, 2, 3, 4, 5$. $Eigen_{itd}$ is 1 if the sector i's eigenvector centrality, at the beginning of period t, falls between $20 \times (d-1)$ percent and $20 \times d$ percent, otherwise 0. To examine how sectors with various importance in the innovation network suffer differently between Great Recessions of 2008 and that of 1991 or 2001, we take the following specification

$$y_{it} = \sum_{d=1}^{5} [\alpha_d Eigen_{itd} + \beta_d I_{t=crisis08} * Eigen_{itd}] + \epsilon_{it} \tag{36}$$

where y_{it} is the R&D growth for sector i in episode t (1991, 2001 or 2008), $I_{t=crisis08}$ is the episode dummy which equals to 1 if the episode is the 2008 crisis and 0 for the 1991 (or 2001) recession. Consider a comparison between the recessions of 2008 and of 2001, α_d is the average R&D growth of sectors in 2001, while β_d measures the difference in the average R&D growth in the recession of 2008 relative to 2001 within the d^{th} group. For example, a negative β_5 indicates that sectors in the center of the innovation network suffer worse in 2008 relative to 2001.

Table 6 shows the basic regression results. Columns 1 and 2 examine the cross-sectional difference of sectors in the exposure to the adversary shocks of 2001 and 2008. Column 1 shows the results with equal weights across sectors, and column 2 shows the results with the previous year's sectoral R&D as weights. There are several things worth emphasizing. First, in 2001, sectors in the periphery of the innovation network (sectors with eigenvector centrality in the bottom 20%) experienced a more significant drop in the R&D growth than those in the center of the innovation network (with eigenvector centrality in the top 20%). However, this pattern reverses during the Great Recession. Sectors in the center of the innovation network experienced a more significant decline in R&D growth than their periphery counterparts. Second, compared to the recession of 2001, sectors with eigenvector centrality in the top 20% experienced a further decline in the R&D growth by 14.8% with equal weights, or 32.8% with R&D weighted in Great Recession. On the contrary, sectors in the bottom 20% witnessed an increase in the R& D growth by 18.7% with equal weights or 12.3% with R& D weighted in the Great Recession.

Columns 3 and 4 show similar results when we compare the recessions of 1991 and 2008. The Great Recession witnessed a much larger drop in R&D growth for sectors in the center of the innovation network, relative to 1991. Specifically, during the Great Recession, sectors with centrality in the top 20% experienced a sharp slide in the R&D growth by 24.9% with equal weights or 26.8% with R&D weighted, relative to the recession of 1991.

Table 6: Sector Centrality and R&D Expense Change over the Recessions

VARIABLES	R&D Change (2001 vs 2008)	R&D Change (2001 vs 2008)	R&D Change (1991 vs 2008)	R&D Change (1991 vs 2008)
Bottom 20% centrality	-0.225***	-0.0711***	0.0332	0.177***
	(0.0639)	(0.00822)	(0.110)	(0.0233)
20-40% centrality	-0.0979	-0.0533***	0.360	-0.0767***
	(0.0910)	(0.00233)	(0.365)	(0.00136)
40-60% centrality	0.0993	-0.0319***	-0.0479	-0.0330
	(0.143)	(0.000884)	(0.0979)	(0.0343)
60-80% centrality	-0.0805	0.00669	-0.191	-0.0205***
	(0.103)	(0.00448)	(0.141)	(0.00374)
Top 20% centrality	0.00134	0.0705***	0.103	0.0103***
	(0.0468)	(0.00111)	(0.0967)	(0.000630)
Bottom 20% centrality × 2008 crisis	0.187	0.123***	-0.0705	-0.124***
	(0.158)	(0.0136)	(0.182)	(0.0257)
20-40% centrality × 2008 crisis	0.0541	0.0814***	-0.404	0.105***
	(0.128)	(0.00485)	(0.376)	(0.00446)
40-60% centrality × 2008 crisis	-0.0744	0.0364***	0.0729	0.0375
	(0.188)	(0.00156)	(0.157)	(0.0343)
60-80% centrality × 2008 crisis	0.0556	-0.335***	0.166	-0.308***
	(0.229)	(0.00709)	(0.250)	(0.00665)
Top 20% centrality × 2008 crisis	-0.148**	-0.328***	-0.249**	-0.268***
	(0.0669)	(0.00116)	(0.108)	(0.000713)
Observations	109	152,157	98	116,380
R-squared	0.078	0.657	0.078	0.816

Note: This table compare the R&D expense changes of sectors with different levels of eigenvector centrality in recent recessions in the 1991, 2001 recessions and the Great Recession. As is noted by the coefficients of Top 20% centrality × 2008 crisis, the most important sectors in the innovation network suffers significantly more than other sectors in the Great Recession, while this is not the case for the 2001 and 1991 recessions. Columns 1 and 3 are simple OLS, and Columns 2 and 4 are WLS weighted by sector R&D value. Robust standard errors in parentheses with *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5 Networks and Asset Pricing

In this section, we explore one application of the theory in asset pricing.

5.1 Networks and Pricing Kernel

In this section, we consider the implications of the theory on asset pricing. Intuitively, a shock to the arrival rate at period t affects $\boldsymbol{A}_t$ at first. Through the propagation via the innovation network, the effect will persist at periods $s = t+1, t+2, ...\infty$. Thus, to link the asset pricing with the network structure, the SDF $M_{t,t+1}$ must be linked to the consumption growth of periods $t+s,\ s \geq 2$. Epstein-Zin preference enables us to establish such linkages.

Here, we will focus on the Cobb-Douglas production technology to obtain a closed-form solution. For a more general case, we provide an approximation in Appendix A.2. To obtain the intuition, we first focus on the case where companies learn only from the recent innovations as 11. We discuss the general case in the next subsection. Under Cobb-Douglas, we have

$$\begin{aligned}
&\Delta c_{t+1} = \Delta y_{t+1} = \frac{1}{1-\eta}\boldsymbol{s}'\Delta\boldsymbol{a}_{t+1} \\
&\Delta\boldsymbol{a}_{t+1} = \boldsymbol{\mu}_t + \sigma_A \boldsymbol{z}^A_{t+1} \\
&\boldsymbol{\mu}_{t+1} = ((1-\rho)\boldsymbol{I} + \boldsymbol{W})\boldsymbol{\mu}_t + \boldsymbol{W}\sigma_A\boldsymbol{z}^A_{t+1} + \sigma_u\boldsymbol{z}^u_{t+1}.
\end{aligned} \tag{37}$$

Under the E-Z preference, the logarithm of the SDF is

$$m_{t+1} = \log(M_{t,t+1}) = \theta\log(\delta) - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1)r_{m,t+1}. \tag{38}$$

To solve for the logarithm of state price density, we need to pin down the equilibria market return $r_{m,t+1}$ as a function of the state variables. Denote the logarithm of price dividend ratio for the aggregate dividend process as z_t, the market portfolio return can be written as approximately [Campbell and Shiller, 1988] [19]

$$r_{m,t+1} = \kappa_0 + \kappa_1 z_{t+1} + \Delta c_{t+1} - z_t, \tag{39}$$

with κ_0 and κ_1 are constant of log-linearization. From equation 37, $\boldsymbol{\mu}_t$ are sufficient statistics for future arrival rate $\boldsymbol{\mu}_{t+1}$ and $\Delta \boldsymbol{a}_{t+1}$. Besides, the future growth of consumption or dividend linearly depends on $\boldsymbol{\mu}_t$. Thus, $\boldsymbol{\mu}_t$ are the state variables of the economy. Conjecture that z_t is linear in the state variables: $z_t = b_0 + \boldsymbol{b}_1' \boldsymbol{\mu}_t$. Substituting 39 and the log of SDF into the Euler equation:

$$E_t exp(m_{t+1} + r_{m,t+1}) = 1 \tag{40}$$

The solution to the coefficients can be obtained with the method of undetermined coefficients as below. The details of derivation are in Appendix **??**

$$\begin{aligned} (1-\kappa_1)b_0 &= \log(\delta) + \kappa_0 + \frac{\theta}{2} var_t(r_{m,t+1} - 1/\psi \Delta c_{t+1}) \\ \boldsymbol{b_1} &= -\frac{1-\psi}{\psi(1-\eta)}[\boldsymbol{I} - \kappa_1((1-\rho)\boldsymbol{I} + \boldsymbol{W'})]^{-1}\boldsymbol{s} \end{aligned} \tag{41}$$

Substituting the expressions for $r_{m,t+1}$ and the dynamics of Δc_{t+1} back to the log of the SDF in equation 38, we can express innovations in the pricing kernel in terms of underlying shocks (risks).

---

[19] Specifically, denote

$$z_t = \log(\frac{W_t - D_t}{D_t})$$

as the logarithm of the price-dividend ratio of the aggregate dividend process. Based on the definition of market return, we have

$$R_{m,t+1} = \frac{W_{t+1}}{W_t - D_t} = \frac{1+e^{z_{t+1}}}{e^{z_t}} \frac{D_{t+1}}{D_t} = \frac{1+e^{z_{t+1}}}{e^{z_t}} \frac{C_{t+1}}{C_t}$$

since, in equilibrium, we have $C_t = D_t$. Thus,

$$r_{m,t+1} = \log(1 + e^{z_{t+1}}) - z_t + \Delta c_{t+1}$$

We take a linear approximation around $\bar{z}$, the unconditional expectation of z_t,

$$r_{m,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + \Delta c_{t+1}$$

with $\kappa_0 = \log(1 + e^{\bar{z}}) - \bar{z}\kappa_1$ and $\kappa_1 = \frac{e^{\bar{z}}}{1+e^{\bar{z}}}$

Proposition 5.1 (No Lagged Learning) *Under the Cobb-Douglas,*

$$
\begin{aligned}
m_{t+1} - E_t m_{t+1} &= -f_{m,0}\frac{\boldsymbol{s}'}{1-\eta}\boldsymbol{\epsilon}^A_{t+1} - f_{m,1}\frac{\boldsymbol{s}'}{1-\eta}[\kappa_1((1-\rho)\boldsymbol{I}+\boldsymbol{W})-\boldsymbol{I}]^{-1}[\boldsymbol{W}\boldsymbol{\epsilon}^A_{t+1}+\boldsymbol{\epsilon}^u_{t+1}] \\
&= -f_{m,0}(E_{t+1}-E_t)\Delta c_{t+1} - f_{m,1}\frac{\boldsymbol{s}'}{1-\eta}[\boldsymbol{I}-\kappa_1((1-\rho)\boldsymbol{I}+\boldsymbol{W})]^{-1}[E_{t+1}-E_t]\boldsymbol{\mu}_{t+1}
\end{aligned} \tag{42}
$$

with $f_{m,0} = 1-\theta+\theta/\psi = \gamma$, and $f_{m,1} = \kappa_1(1-\theta)(1-1/\psi)$.The first term of equation 42 is the one related to CCAMP, and the second term comes from the innovation network.

When $\theta = 1$, the preference is time separable and the inter-temporal marginal rate of substitution (IMRS) only depends on the Δc_{t+1}, and the second part of 42 disappears since $f_{m,1} = 0$. Consequently, only risks associated with the growth of consumption at period $t+1(\Delta c_{t+1})$ are priced, while risks associated with the consumption growth in the future like Δc_{t+s}, with $s \geq 2$ are not priced. Thanks to the lagged knowledge diffusion effects - the shock $(E_{t+1}-E_t)\boldsymbol{\mu}_{t+1}$ only affects consumption growth in the future, i.e., $\Delta c_{t+s},\ s \geq 2$, the risks associated with the innovation network will not be priced.

When $\theta \neq 1$, the SDF $M_{t,t+1}$ depends not only on the current consumption growth but also on the continuation utility U_{t+1} which in turn depends on the continuation consumption growth $\Delta c_{t+s},\ s \geq 2$. Therefore, shocks to future consumption growth are priced. A shock to $\boldsymbol{\mu_{t+1}}$ will affect the realized $\boldsymbol{\Delta a_{t+2}}$, which in turn changes the arrival rate in the future. Through this propagation, the initial shock to the arrival rate will have a persistent effect on future consumption growth and its effect on the marginal utility will be amplified significantly.

5.1.1 General Cases with Lagged Learning

In this section, we provide results when firms can learn from historical innovations beyond the $\boldsymbol{\mu}_t$. Specifically, we model the arrival intensity as

$$
\Delta\mu_{it+1} = -\rho\mu_{it} + \sum_j W_{ij}\varphi(L)\Delta\log(A_{jt+1}) + \epsilon^u_{it+1}. \tag{43}
$$

Where $\varphi(L) = \sum_{s\geq 0}\varphi_s L^s$ and $\varphi_0 = 1$, with L the lag-operator. Under this case, the state variables at t include $(\boldsymbol{\mu}_t, \Delta\boldsymbol{a}_s, \forall s \leq t)$.

Proposition 5.2 (Lagged Learning Process) *Under the general case 43, the innova-*

tion to the SDF takes the form of

$$
\begin{aligned}
m_{t+1} - E_t m_{t+1} &= -\theta/\psi(c_{t+1} - E_t c_{t+1}) + (\theta - 1)(r_{m,t+1} - E_t r_{m,t+1}) \\
&= -f_{m,0}[\Delta c_{t+1} - E_t \Delta c_{t+1}] - f_{m,1}\frac{\boldsymbol{s'}}{1-\eta}\Big[\boldsymbol{I} - \kappa_1[(1-\rho)\boldsymbol{I} + (1+\psi_0)\boldsymbol{W}]\Big]^{-1} \\
&\times \Big[\boldsymbol{\epsilon}^u_{t+1} + (1+\psi_0)\boldsymbol{W}\boldsymbol{\epsilon}^A_{t+1}\Big]
\end{aligned}
\tag{44}
$$

with $f_{m,0} = 1 - \theta + \theta/\psi = \gamma$, $f_{m,1} = \kappa_1(1-\theta)(1-1/\psi)$, and $\psi_0 = \sum_{s\geq 1}\kappa_1^s\varphi_s$. When firms only learn from the most recent innovation, that is, $\varphi_s = 0, \forall s \geq 1$, then $\psi_0 = 0$. The results are reduced to the cases in equation 42. Allowing for more lagged learning does not change our results too much qualitatively but make the SDF more volatile. For the proof of this general result, see the appendix **??**.

5.2 Nexus of Recovery and Cross-Sectional Asset Pricing

This section shows cross sectional returns provide information on the prospect recovery of the economy. Denote the unexpected shock to arrival rate as $\boldsymbol{\epsilon}_{t+1} = [\mathbb{E}_{t+1} - \mathbb{E}_t]\boldsymbol{\mu}_{t+1} = \sum_{j\in[J]}\alpha_{j,t+1}\boldsymbol{v}_j$, with $\alpha_{j,t+1} = (\boldsymbol{v}_j, \boldsymbol{\epsilon}_{t+1})$ the loading of the shock on $\boldsymbol{v}_j, j \in [J]$.

From 42, we have

$$
\begin{aligned}
m_{t+1} - \mathbb{E}_t m_{t+1} = &- f_{m,0}(\mathbb{E}_{t+1} - \mathbb{E}_t)\Delta c_{t+1} \\
&- \frac{f_{m,1}}{1-\eta}\sum_{j\in[J]}[1 - \kappa_1 + \kappa_1(\rho - \lambda_j(\boldsymbol{W}))]^{-1}(\boldsymbol{s}, \boldsymbol{v}_j)\alpha_{jt+1}.
\end{aligned}
\tag{45}
$$

Note that $\kappa_1 \in (0,1)$ and very close to 1 empirically. When the shock $[\mathbb{E}_{t+1} - \mathbb{E}_t]\boldsymbol{\mu}_{t+1}$ is roughly parallel to the eigenvector centrality, the volatility of the log of the state price would increase sharply. Empirically, we find that $\lambda_j(\boldsymbol{M})$ and $(\boldsymbol{s}, \boldsymbol{v}_j)$ decline sharply as j increases. Thus, only the first several eigenvectors play non-negligible role in asset pricing empirically.

Proposition 5.3 *If the learning process satisfies 9, The expected return of any asset is*

$$
\begin{aligned}
E_t[r_{it+1} - r_{f,t+1}] + \frac{1}{2}Var_t(r_{it+1}) = f_{m,0}&Cov_t(r_{i,t+1}, (E_{t+1} - E_t)\Delta c_{t+1}) \\
&+ \frac{f_{m,1}}{1-\eta}\sum_j[1 - \kappa_1 + \kappa_1(\rho - \lambda_j(\boldsymbol{W}))]^{-1}(\boldsymbol{s}, \boldsymbol{v}_j)Cov_t(r_{i.t+1}, \alpha_{jt+1}),
\end{aligned}
\tag{46}
$$

with $f_{m,0} = 1 - \theta + \theta/\psi = \gamma$, *and* $f_{m,1} = \kappa_1(1-\theta)(1-1/\psi)$.

PROPOSITION 5.4 *For the more general learning process 10, the innovation of logarithm of the SDF*

$$\begin{aligned} m_{t+1} - E_t m_{t+1} = & - f_{m,0}[\Delta c_{t+1} - E_t \Delta c_{t+1}] \\ & - \frac{f_{m,1}}{1-\eta} \sum_j [1 - \kappa_1 + \kappa_1(\rho - \psi_0 \lambda_j(\boldsymbol{W}))]^{-1} (\tilde{\boldsymbol{\epsilon}}_{t+1}, \boldsymbol{v}_j)(\boldsymbol{s}, \boldsymbol{v}_j) \end{aligned} \tag{47}$$

with $\tilde{\boldsymbol{\epsilon}}_{t+1} = \boldsymbol{\epsilon}^u_{t+1} + \psi_0 \boldsymbol{W} \boldsymbol{\epsilon}^A_{t+1}$ *and* $\psi_0 = \sum_{s \geq 0} \varphi_s$. *Furthermore,*

$$\begin{aligned} E_t[r_{it+1} - r_{f,t+1}] + \frac{1}{2} Var_t(r_{it+1}) = & f_{m,0} Cov_t(r_{i,t+1}, (E_{t+1} - E_t)\Delta c_{t+1}) \\ & + \frac{f_{m,1}}{1-\eta} \sum_j [1 - \kappa_1 + \kappa_1(\rho - \psi_0 \lambda_j(\boldsymbol{W}))]^{-1} (\boldsymbol{s}, \boldsymbol{v}_j) Cov_t(r_{i.t+1}, \tilde{\alpha}_{jt+1}) \end{aligned} \tag{48}$$

with $\tilde{\alpha}_{jt+1} = (\boldsymbol{v}_j, \tilde{\boldsymbol{\epsilon}}_{t+1})$, $f_{m,0} = 1 - \theta + \theta/\psi = \gamma$, *and* $f_{m,1} = \kappa_1(1-\theta)(1-1/\psi)$.

There are several things worth mentioning:

1. The persistent impact of the initial shock is closely related to behaviors of asset prices by comparing Equation 23 and Propositions 5.3 or 5.4. When the initial shock highly correlates with the eigenvector centrality of the innovation network, its impact on future innovation will become very persistent, while the SDF's conditional volatility will increase sharply.

2. Under the decomposition of Proposition 5.3, there are J factors, the number of sectors in the market. The risk premium associated with factor j is

$$RP_j = \frac{f_{m,1}}{1-\eta} \Big[1 - \kappa_1 + \kappa_1[\rho - \psi_0 \lambda_j(\boldsymbol{W})] \Big]^{-1} (\boldsymbol{s}, \boldsymbol{v}_j) \tag{49}$$

   $\kappa_1 = \frac{e^{\bar{z}}}{1+e^{\bar{z}}}$, with $\bar{z}$ the unconditional expected log of aggregate price-dividend ratio. In real data, the parameter κ_1 is close to but slightly smaller than 1, $\kappa_1 \approx 0.997$ [Campbell and Shiller, 1988; Bansal and Yaron, 2004]. Empirically, we document that $\lambda_j(\boldsymbol{W})$ drops significantly, thus

   2.1 If ρ is non-negligible, only factors such that $\rho \approx \psi_0 \lambda_j(\boldsymbol{W})$ exhibit very high risk-premium. Combining the fact that $\lambda_j(\boldsymbol{W})$ drop sharply, we conclude that only the first several factors are important in practice.

2.2 If ρ is close to zero, then $\rho - \psi_0\lambda_j(\boldsymbol{W}) \approx 0$. Under this case, only factors with large $(\boldsymbol{s}, \boldsymbol{v}_j)$ are important in practice.

3. The theory rationalize the long-run risk in a networked economy. When $\rho \approx \psi_0\lambda_1(\boldsymbol{W})$, and $(\boldsymbol{s}, \boldsymbol{v}_1)$ and $(\delta\boldsymbol{\mu}_{t+1}, \boldsymbol{v}_1)$ are non-negligible, the impacts of the initial shock on future growth are very persistent and significantly amplified. The pricing kernel becomes very volatile and the expected market return rises sharply to compensate for the risk.

4. Besides the common preference parameter $f_{m,1}$ shared by all network factors, the magnitude of risk premium associated with the j^{th} factor depends on $[1 - \kappa_1 + \kappa_1(\rho - \psi_0\lambda_j(\boldsymbol{W}))]^{-1}$ and $(\boldsymbol{s}, \boldsymbol{v}_j)$. The first term depends on the eigenvalue distribution of the innovation network while the second captures the interaction between the production and innovation networks.

Empirically, remember $\psi_0 = 1$, $\lambda_1(\boldsymbol{W}) \approx 1$, and $\rho \approx \lambda_1(\boldsymbol{W})$, thus the volatility of the pricing kernel is proportional to $(1 - \kappa)^{-1} \approx 330$ when the cross-sectional shock follows the eigenvector centrality of the innovation network. On the contrary, if there is no technology spillover, the pricing kernel is proportional to $(1 - \kappa + \rho)^{-1} \approx 3$ since $\rho \approx 0.3$. As a result, due to the technology spillover and the low rank structure of the innovation network, the volatility of the pricing kernel is amplified to 100 folders when the shocks follows the eigenvector centrality of the innovation network. This amplified volatility can well explain the risk-premium and risk-free rate puzzles even for a modest risk-averse coefficient [Mehra and Prescott, 1985; Bansal and Yaron, 2004].

6 Conclusion

In this paper, we propose a production economy incorporating both the innovation network and production network. We examine the dynamic interactions among the cross-sectional shock to technology progress, innovation network, and production network. We emphasize the crucial role of the network's structure and shock's sectoral distribution in the amplification and persistence of the shock's impact.

We first show that the technology spillover effect depends on the direction of the shock to technology progress. The economy exhibits a stronger technology spillover effect when the shock parallels the sectoral importance (i.e., eigenvector centrality) in the innovation network. On the contrary, the spillover effect becomes weakest when the shock is parallel to the eigenvector associated with the smallest eigenvalue of the innovation network.

Several facts on the structure of the innovation network and the cross-sectional shocks are documented in this paper. The empirical evidence suggests that the US's innovation network has a low-rank structure so that the leading eigenvalue dominates in magnitude the remaining ones of the innovation network. Furthermore, the most potent spillover-effect roughly cancels out the depreciation effect of the technology shock. The unique structure implies that the shock's sectoral distribution reveals information on the economy's recovery path when an adverse shock hits the economy. Specifically, conditional on the shock's initial aggregate effect, the economy will experience a prolonged recovery process only if important sectors in the innovation network suffer more than their counterparts. However, the economy will recover quickly from the recession if the shock follows other directions. This channel indicates that the sectors' heterogeneous exposure to the technology shock plays a vital role in shaping the prolonged slow recovery. Thus understanding why some crucial sectors in the innovation network suffer more during some recessions while suffering much less during other episodes is an important research topic in the future.

Besides the persistence, we show that two sufficient statistics fully capture the shock's amplification effect. The first is the inner product between the cross-sectional shock and the sectoral importance in the innovation network. This coefficient captures how shocks are propagated through the technology innovation-network. The second coefficient is the inner product between sectors' importance in the innovation network and their importance in the production network. This coefficient captures how the initial technology shocks are propagated from the innovation network to the production network.

We document a stable and robust interaction between the innovation network and production network. The shock will be amplified significantly when the shock parallels the sectoral importance in the innovation network (i.e., the eigenvector centrality). Empirically, we document that the pivotal sectors in the innovation network suffer much more than their periphery counterparts during the great recessions of 2008. In contrast, this pattern reverses during other recessions like the recession of 1998 or 2001.

References

Daron Acemoglu, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Daron Acemoglu, Ufuk Akcigit, and William Kerr. Networks and the macroeconomy: An empirical exploration. *NBER Macroeconomics Annual*, 30(1):273–335, 2016a.

Daron Acemoglu, Ufuk Akcigit, and William R Kerr. Innovation network. *Proceedings of the National Academy of Sciences*, 113(41):11483–11488, 2016b.

P Aghion and P Howitt. A model of growth through creative destruction. *Econometrica*, 60 (2), 1992.

Mohammad Ahmadpoor and Benjamin F Jones. The dual frontier: Patented inventions and prior scientific advance. *Science*, 357(6351):583–587, 2017.

Franklin Allen, Junhui Cai, Xian Gu, Jun"QJ" Qian, Linda Zhao, and Wu Zhu. Ownership networks and firm growth: What do five million companies tell about chinese economy. *Working Paper*, 2019.

Brian DO Anderson and John B Moore. *Optimal filtering*. Courier Corporation, 2012.

Diego Anzoategui, Diego Comin, Mark Gertler, and Joseba Martinez. Endogenous technology adoption and r&d as sources of business cycle persistence. *American Economic Journal: Macroeconomics*, 11(3):67–110, 2019.

Enghin Atalay. How important are sectoral shocks? *American Economic Journal: Macroeconomics*, 9(4):254–80, 2017.

Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509, 2004.

David Rezza Baqaee. Cascading failures in production networks. *Econometrica*, 86(5):1819–1838, 2018.

David Rezza Baqaee and Emmanuel Farhi. The macroeconomic impact of microeconomic shocks: beyond hulten's theorem. *Econometrica*, 87(4):1155–1203, 2019.

Jean-Noël Barrot and Julien Sauvagnat. Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592, 2016.

Marianne Baxter and Robert G King. Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics*, 81(4):575–593, 1999.

Ben Bernanke and Mark Gertler. Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1):14–31, 1989.

Francesco Bianchi, Howard Kung, and Gonzalo Morales. Growth, slowdowns, and recoveries. *Journal of Monetary Economics*, 101:47–63, 2019.

Saki Bigio and Jennifer La'o. Financial frictions in production networks. *NBER working paper*, (w22212), 2016.

Nicholas Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013.

Nicholas Bloom, Charles I Jones, John Van Reenen, and Michael Webb. Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44, 2020.

Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182, 1987.

Phillip Bonacich. Some unique properties of eigenvector centrality. *Social networks*, 29(4): 555–564, 2007.

Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201, 2001.

Markus K Brunnermeier, Thomas M Eisenbach, and Yuliy Sannikov. Macroeconomics with financial frictions: A survey. Technical report, National Bureau of Economic Research, 2012.

John Y Campbell and Robert J Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1988.

Fabio Canova. *Methods for applied macroeconomic research*. Princeton university press, 2011.

Diego Comin and Mark Gertler. Medium-term business cycles. *American Economic Review*, 96(3):523–551, 2006.

Maarten De Ridder and Coen Teulings. Endogenous growth and lack of recovery from the global crisis. *Voxeu article, Available, July*, 2017.

Romain Duval, Gee Hee Hong, and Yannick Timmer. Financial frictions and the great productivity slowdown. *The Review of Financial Studies*, 33(2):475–503, 2019.

LG Epstein and SE Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica*, 57(4):937–969, 1989.

John G Fernald, Robert E Hall, James H Stock, and Mark W Watson. The disappointing recovery of output after 2009. Technical report, National Bureau of Economic Research, 2017.

Andrew T Foerster, Pierre-Daniel G Sarte, and Mark W Watson. Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1):1–38, 2011.

A Galeotti, B Golub, and S Goyal. Targeting interventions in networks. *Econometrica*, 2020.

Nicolae Gârleanu, Leonid Kogan, and Stavros Panageas. Displacement risk and asset returns. *Journal of Financial Economics*, 105(3):491–510, 2012.

Nicolae Garleanu, Stavros Panageas, and Jianfeng Yu. Technological growth and asset pricing. *The Journal of Finance*, 67(4):1265–1292, 2012.

Zvi Griliches. Distributed lags: A survey. *Econometrica: journal of the Econometric Society*, pages 16–49, 1967.

Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools. Technical report, National Bureau of Economic Research, 2001.

Bernard Herskovic. Networks in production: Asset pricing implications. *The Journal of Finance*, 73(4):1785–1818, 2018.

Robert J Hodrick and Edward C Prescott. Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16, 1997.

Michael Horvath. Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, 45(1):69–106, 2000.

Michael Horvath et al. Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks. *Review of Economic Dynamics*, 1(4):781–808, 1998.

Charles R Hulten. Growth accounting with intermediate inputs. *The Review of Economic Studies*, 45(3):511–518, 1978.

Adam B Jaffe. Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value. *American Economic Review*, 76(5):984–1001, 1986.

Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598, 1993.

Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1): 39–43, 1953.

Robert G King and Mark W Watson. Money, prices, interest rates and the business cycle. *The Review of Economics and statistics*, pages 35–53, 1996.

Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of political economy*, 105(2): 211–248, 1997.

Leonid Kogan, Dimitris Papanikolaou, and Noah Stoffman. *Technological innovation: Winners and losers*. Number w18671. National Bureau of Economic Research, 2013.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2): 665–712, 2017.

Howard Kung and Lukas Schmid. Innovation, growth, and asset prices. *The Journal of Finance*, 70(3):1001–1037, 2015.

Ernest Liu. Industrial policies in production networks. *The Quarterly Journal of Economics*, 134(4):1883–1948, 2019.

John B Long and Charles I Plosser. Real business cycles. *Journal of Political Economy*, 91 (1):39–69, 1983.

Rajnish Mehra and Edward C Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1985.

Ulrich K Müller and Mark W Watson. Long-run covariability. *Econometrica*, 86(3):775–804, 2018.

Alexei Onatski and Francisco Ruge-Murcia. Factor analysis of a large dsge model. *Journal of Applied Econometrics*, 28(6):903–928, 2013.

Albert Queralto. A model of slow recoveries from financial crises. *Journal of Monetary Economics*, 2019.

Yucheng Yang and Wu Zhu. Resource reallocation, non-linear effect, and asset pricing in general nested production networks. 2020.

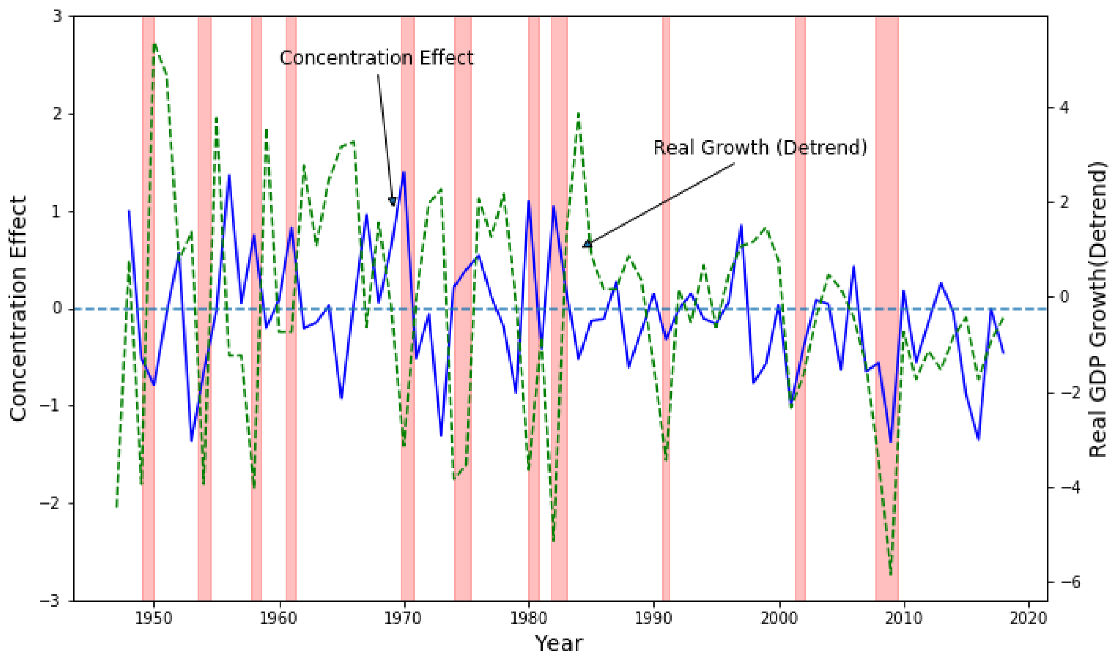
Wu Zhu. Networks, linking complexity, and cross predictability. 2020.

Table 7: Summary Statistics for Firms in Different Centrality Quantiles

Centrality Quantile	Average R&D Expense ($million)	Average Market Value ($million)	Average Book-to-Market Value
Bottom 20%	4.57	2300.10	0.931
20-40%	13.21	2920.24	0.828
40-60%	14.36	3071.22	0.824
60-80%	27.75	2893.71	0.736
Top 20%	76.26	2867.46	0.832

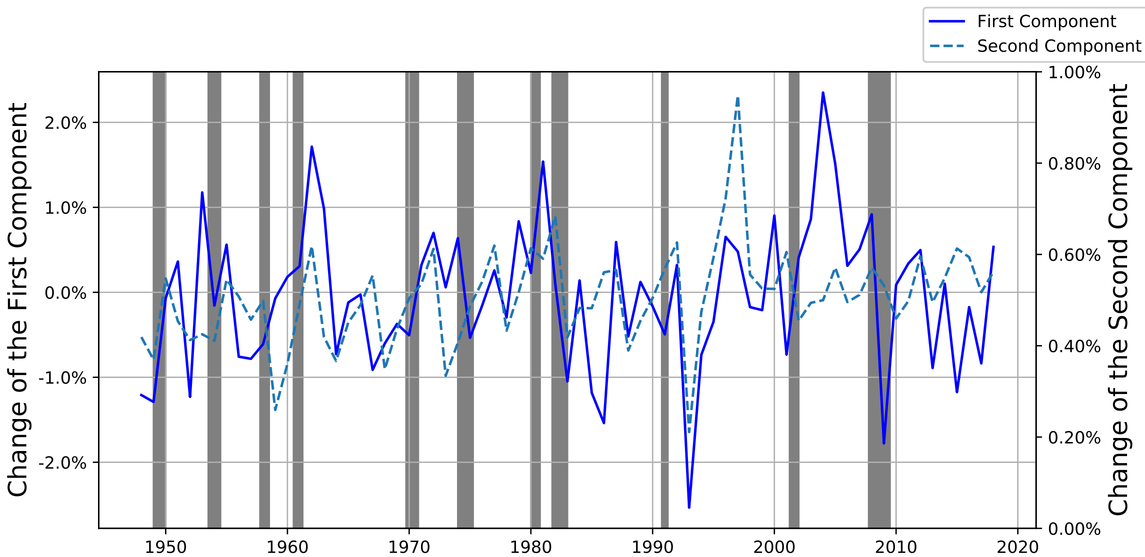
Note: This table reports average R&D expenses, average market value (both in million of dollars), average book-to-market value of Compustat firms in different centrality quantiles. To construct the table, each year, we first sort all sectors (at three-digit NAICS level) into five groups based on the eigenvector centrality of the last year. Within each group, we pool all the firm-year observations together and report their aggregate R&D expense, the average market value, and the average Book-to-Market ratio. The sample period is from 1952 to 2014 at an annual frequency.

Figure 10: Non-linear Effect from Change in Concentration



Note: This figure illustrates the non-linear effect from the change in concentration on growth. The real blue line represent the growth from the concentration effect(indexed by the left y-axis). The dash green line represents the detrend real growth(indexed by the right y-axis). At year t, we calculate the concentration as $N^c(t) = -\sum_j s_{jt} \log(s_{jt})$ with s_{jt} is the output share of sector j at the three-digit NAICS level at the end of calendar year t. We construct the output share of three-digit NAICS sector from the output table by sectors provided by the BEA and BLS. The non-linear growth effect from change in the concentration at year is defined $N^c(t) - N^c(t-1)$. This concentration effect reflects the contribution of resource reallocation across sectors to the growth.

Figure 11: Non-linear Effect from Change in Sparsity

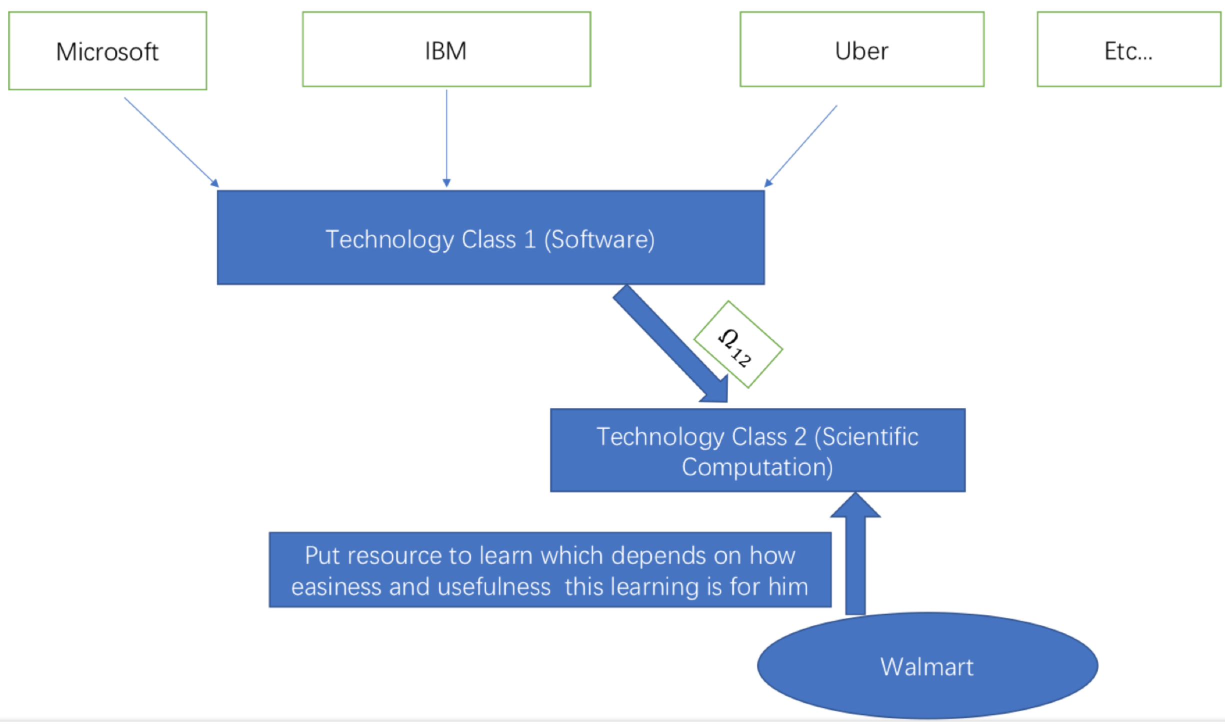


Note: This figure illustrates the non-linear effect from the change in the adjusted sparsity on growth, the resource reallocation within the sector. We decompose the adjusted sparsity N_{it}^{s} into two components $N_{it}^{s,1}$ and $N_{it}^{s,2}$ with $N_{jt}^{s,1} = \sum_j \theta_{ijt} \log(\theta_{ijt})$ and $N_{jt}^{s,2} = \sum_j \theta_{ij} \log(\frac{\theta_{ijt}}{\theta ij})$. Thus,

$$N_{it}^{s} = N_{it}^{s,1} + \frac{v_i}{1-v_i} N_{it}^{s,2}$$

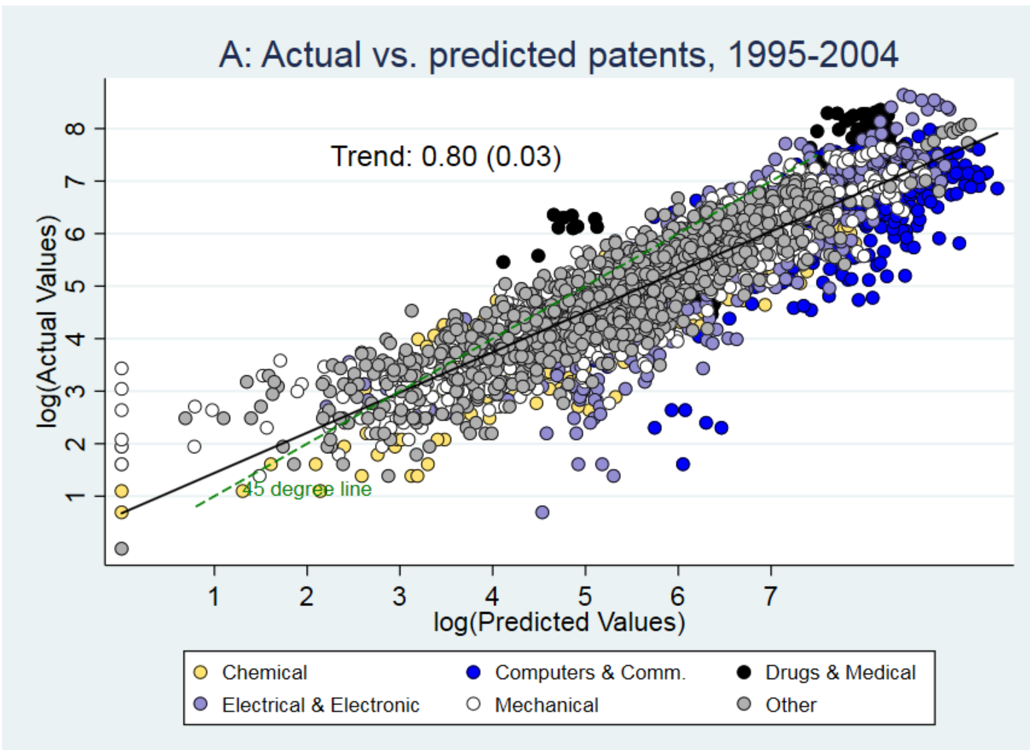
The real blue line represents the growth from the change in first component(indexed by the left y-axis), $\frac{\eta}{1-\eta}\Delta \boldsymbol{s}_t \boldsymbol{N}_t^{s,1}$.The dash gray line represents the the second component (indexed by the right y-axis), $\frac{\eta}{1-\eta}\Delta \boldsymbol{s}_t \boldsymbol{N}_t^{s,2}$.At year t, we construct the sparsity measure is the based on the input-output table at the three-digit NAICS level at the end of calendar year t. We follow Herskovic [2018] to choose $\eta = 0.35$.

Figure 12: Knowledge Diffusion across Firms



Note: This figure illustrates how a representative firm i learns from others in our empirical construction of knowledge diffusion matrix and our model of micro foundation. Specifically, firm i (Walmart) put resources to establish several research groups, each of them focusing on various technology fields. Consider the research group in technology class 2 (scientific computation), they learn knowledge and insights not only from the patents in technology class 2, but also from patents in other fields, say class 1 (Software). The knowledge flow from class 1 to class 2 is captured by Ω_{21}. On the other hands, there are many companies (like Microsoft, IBM, and Uber etc) contributing to knowledge and insights in the class 1.The more the knowledge firm j contributes to the technology class 1, the easier the researchers in scientific computation of firm i learn from class 2, and the more resources firm i (Walmart) puts in class 2, then the more the knowledge and insights firm i will obtain from j through this channel.

Figure 13: Patent Predictability Based on the Innovation Network



Note: This figure shows the predictability of patent issuance in the downstream sectors using the historical patent issuance in the upstream sectors in the innovation network.

Appendix

A Derivations of the Network Model

A.1 The Knowledge Diffusion Matrix

In this section, we provide a heuristic micro foundation to the innovation network that extends the static version by Bloom et al. [2013] to the general dynamic settings.

A.1.1 Arrival of Innovations

Suppose there are $[\boldsymbol{J}] = \{1, 2, ..., J\}$ sectors (i.e., representative firm), and $[\boldsymbol{T}] = \{1, 2, ..., T\}$ as the set of technology classes in technology space. Denote

$$\boldsymbol{A}(t) = [A_1(t), ..., A_j(t), ..., A_J(t)]' \tag{50}$$

as the joint technology stock at time t, with $A_j(t)$ the technology stock of sector $j \in [\boldsymbol{J}]$. Within the time interval $[t, t+1]$, denote $N_j(t, t+1)$ as the new arrival of technology inventions. The technology will be lifted up due to new innovation, we write as

$$\log(\frac{A_j(t+1)}{A_j(t)}) = \sum_{i \leq N_j(t,t+1)} [z_{ji}(t+1) + z_j(t+1) + z(t+1)] \tag{51}$$

Here, we decompose the lift-up effect of each invention into three components - innovation specific, firm specific and time specific, i.e., we allow for inner product of shocks between sectors. Without loss of generality, we set $\mathbb{E}_t z_{ji}(t+1) = 0$, $g_{A,j} = \mathbb{E}_t z_j(t+1)$ and $g_A = \mathbb{E}_j z(t+1)$. If we assume $z_{ji}(t+1), z_j(t+1)$, and $z(t+1)$ are independent of $N_j(t, t+1)$, then

$$E_t \log[\frac{A_j(t+1)}{A_j(t)}] = E_t N_j(t, t+1) E_t[z_{ji}(t+1) + z_j(t+1) + z(t+1)]$$

and

$$E_t \log[\frac{A_j(t+1)}{A_j(t)}] = \mu_j(t)$$

with $\mu_j(t)$ is the arrival intensity of new technology between t and $t+1$.

A.1.2 Endogenous Learning

In this subsection, we model the process of arrival intensity as a learning process in a dynamic setting. In reality, one firm in a specific industry (e.g. Microsoft) may issued many patents, spanning several technology classes in technology space $[\boldsymbol{T}]$. We denote $F_{q\leftarrow j}(t)$ as the fraction of sector j's technology in technology class q, denote

$$\boldsymbol{F_j(t)} = (F_{1\leftarrow j}, ..., F_{q\leftarrow j}, ..., F_{T\leftarrow j})' \tag{52}$$

as the technology distribution of sector j in technology space $[\boldsymbol{T}]$. Empirically, the distribution $\boldsymbol{F}_j$ is very persistent over time if we interpret it as the patent distribution in technology space [Acemoglu et al., 2016b]. Denote

$$TP_q(t+1) = \sum_{k\in[J]} F_{q\leftarrow k} log[\frac{A_k(t+1)}{A_k(t)}] \tag{53}$$

the technology insights in field q contributed by all sectors between t and $t+1$, which is the available technology can be learned by all sectors (companies).

Figure 12 shows how firms learn from each other. Firm j (Walmart) put efforts and resources to establish research groups or laboratories in various technology fields - scientific computation, software, and forecasting etc. Consider researchers in scientific computation, they not only obtain insights from the technology progress in scientific computation, but also from the technology progress in other fields like software contributed by other companies like *Microsoft, IBM, and Uber etc.* Specifically, we denote the effort or resource vector as

$$\boldsymbol{e}_j(t) = (e_{j,1}, ..., e_{j,\tau}, ..., e_{j,T})'(t)$$

with $e_{j\tau}(t)$ the resources put by j in technology field τ at t. Given the efforts $\boldsymbol{e}_j(t)$, we model the arrival intensity at $t+1$ as

$$\mu_j(t) = (1-\theta_0)\mu_j(t-1) + \theta_1 \sum_{\tau\in[\boldsymbol{T}]} f(e_{j\tau}(t), l_{j\tau} \sum_q \Omega_{\tau\leftarrow q} TP_q(t)) + \epsilon^u_{jt} \tag{54}$$

We explain this setup one by one as follows

1. If there is no learning or R&D, the arrival process is mean-reverse (i.e., $\theta_0 > 0$)

$$\mu_j(t) = (1 - \theta_0)\mu_j(t-1) + \epsilon_t^u$$

2. $TP_q(t)$ is the technology knowledge in field q that researchers can learn. $\Omega_{\tau\leftarrow q}$ measures the easiness of researchers with expertise in field τ learning from field q. We use $\Omega_{\tau\leftarrow q}$ to capture the technology closeness between fields τ and q and allow for heterogeneity in the easiness. This is quite intuitive, it is easier for researchers in *Algorithm* to learn from technology class *Optimization* than *Hardware Designers.* The higher $\Omega_{\tau\leftarrow q}$ is, the easier it is that knowledge flows from q to τ.Overall, we use

$$\Omega_{\tau\leftarrow q} TP_q(t))$$

to capture the amount of knowledge in field q available to field τ adjusted by learning easiness.

3. In real world, sometimes it is more difficult for some sectors (firms) than others to obtain or apply the useful insights from a given technology field. To capture this effect, we use $l_{j,\tau}$ to measure *learning ability* of firm (industry) j from τ, thus, the term measures

$$l_{j\tau} \sum_q \Omega_{\tau\leftarrow q} TP_q(t))$$

the easiness of researchers in technology field τ of firm (industry) j from others.

4. $f(e_{j\tau}(t), l_{j\tau} \sum_q \Omega_{\tau\leftarrow q} TP_q(t))$ is the *search-match* function in technology field τ for firm (industry) j.With more efforts put in searching for ideas, it is more likely to get new ideas and promote the arrival rate of invention. For simplicity, we further assume the *search-match* function take the form of

$$f(x_1, x_2) = \alpha_0 x_2 + x_1^{\alpha} x_2^{1-\alpha}$$

$\alpha_0 x_2$ captures the externality due to technology diffusion, and $\alpha \in (0,1)$ capture the diminishing marginal effect in learning.

5. The promotion effect due to learning is additive across research groups

$$\sum_{\tau \in [\boldsymbol{T}]} f(e_{j\tau}(t), l_{i\tau} \sum_q \Omega_{\tau \leftarrow q} TP_q(t))$$

with $TP_q(t)$ the pool of knowledge in technology class q from which firms can learn new insights.

A.1.3 Optimal Searching Efforts

Firms choose efforts $\boldsymbol{e}_j(t) = (e_{j,1}, ..., e_{j,\tau}, ..., e_{j,T})'(t)$ to improve the arrival rate $\mu(t)$ which will in turn accelerate the arrive of the new invention in the future. At period t, given Based on the search and match function in equations 54 and **??**, the optimal level of resources put in field τ is

$$\tilde{V}_j(t)\alpha\theta_1[\frac{1}{\tilde{V}_j(t)}]^\alpha \Big[\frac{l_{j\tau}\sum_q \Omega_{\tau\leftarrow q} TP_q(t)}{e_{j\tau}(t)}\Big]^{1-\alpha} = 1 \tag{55}$$

Thus, the optimal amount of resources put into searching for ideas or insights would be

$$e^*_{j\tau}(t) = \tilde{V}_j(t)(\alpha\theta_1)^{\frac{1}{1-\alpha}} l_{i\tau} \sum_q \Omega_{\tau\leftarrow q} TP_q(t) = \tilde{V}_j(t)(\alpha\theta_1)^{\frac{1}{1-\alpha}} l_{j\tau} \sum_{q,k} \Omega_{\tau\leftarrow q} F_{q\leftarrow k} \Delta \log(A_k(t)) \tag{56}$$

The dynamics of arrival intensity follows

$$\mu_j(t) = (1-\theta_0)\mu_j(t-1) + (\alpha_0 + \tilde{\theta}_1) \sum_{\tau q,k} l_{j\tau} \Omega_{\tau\leftarrow q} F_{q\leftarrow k} \Delta \log(A_k(t)) \tag{57}$$

with $\tilde{\theta}_1 = \theta_1(\alpha\theta_1)^{\frac{1}{1-\alpha}}$. Let us denote $\rho = \theta_0$, and

$$W_{jk} = (\alpha_0 + \tilde{\theta}_1) \sum_{\tau,q} l_{j\tau} \Omega_{\tau\leftarrow q} F_{q\leftarrow k} \tag{58}$$

We have

$$\boldsymbol{\mu}(t) = (1-\rho)\boldsymbol{\mu}(t-1) + \boldsymbol{W} \Delta \log(\boldsymbol{A}(t)) \tag{59}$$

in matrix form. Also note that it is very easy for us to incorporate shocks $\sigma_u \boldsymbol{z}^u(t)$ to the arrival intensity in equation 54. Thus, the reduced diffusion process can be written as

$$\boldsymbol{\mu}(t) = (1-\rho)\boldsymbol{\mu}(t-1) + \boldsymbol{W}\Delta\log(\boldsymbol{A}(t)) + \sigma_u \boldsymbol{z}^u(t) \tag{60}$$

A.2 General Cases

In this section, we address the general cases of the nested input-output networks. We start with the basic equations to characterize the equilibrium.

A.2.1 The General Equilibrium

Note that, in equilibrium, all dividends from companies will be consumed

$$C_t = \sum_j D_{jt} = (1-\eta)\sum_j P_{jt}Y_{jt} = (1-\eta)Y_t \Longrightarrow \Delta c_{t+1} = \Delta y_{t+1} \tag{61}$$

with $c_t = \log(C_t)$, $y_t = \log(Y_t)$, $\Delta c_{t+1} = c_{t+1} - c_t$, and $\Delta y_{t+1} = y_{t+1} - y_t$.

Using the market clear condition for firm i, we have

$$\begin{aligned}
&c_{it} + \sum_{j\in[J]} X_{jit} = Y_{it} \Longrightarrow P_{it}c_{it} + \sum_{j\in[J]} P_{it}X_{jit} = P_{it}Y_{it} \\
&\Longrightarrow \alpha_i C_t + \sum_j \frac{P_{it}X_{jit}}{I_{jt}\lambda_{jt}}\frac{I_{jt}\lambda_{jt}}{P_{jt}Y_{jt}}P_{jt}Y_{jt} = \alpha_i C_t + \eta\sum_j \tilde{\theta}_{jit}P_{jt}Y_{jt} = P_{it}Y_{it} \\
&\Longrightarrow \alpha_i(1-\eta)Y_t + \sum_j \tilde{\theta}_{jit}P_{jt}Y_{jt} = P_{it}Y_{it}
\end{aligned} \tag{62}$$

The above is just the accounting identity of input-output table. $\tilde{\theta}_{jit} = \frac{P_{it}X_{jit}}{I_{jt}\lambda_{jt}}$ is the reliance of sector or firm j on firm i - the share of expenditure of firm j on firm i. From the optimization of intermediate firms, we have $\tilde{\theta}_{jit} = \theta_{ji}^{v_j}P_{it}^{1-v_j}\lambda_{jt}^{v_j-1}$. Using equation 5, we have

$$\tilde{\theta}_{jit} = \frac{P_{it}^{1-v_j}\theta_{ji}^{v_j}}{\sum_k P_{kt}^{1-v_j}\theta_{jk}^{v_j}} \tag{63}$$

$\tilde{\theta}_{jit}$ is the price adjusted input dependence of sector j on i. The intuition for equation 62 is that - the final product of sector i will be used either as consumption, $\alpha_i(1-\eta)Y_t$, or the input to other sectors, $\tilde{\theta}_{jit}P_{jt}Y_{jt}$. We also can write equation 62 using share of each sector

$$\alpha_i(1-\eta) + \eta \sum_j \tilde{\theta}_{jit} s_{jt} = s_{it} \Longrightarrow \boldsymbol{s} = (1-\eta)[I - \eta \tilde{\boldsymbol{\Theta}}_t']^{-1} \boldsymbol{\alpha} \tag{64}$$

with $s_{jt} = P_{jt}Y_{jt}/Y_t, \forall j \in [J]$, $\boldsymbol{s} = (s_{1t}, ..., s_{Jt})'$, $\boldsymbol{\alpha} = (\alpha_{1t}, ..., \alpha_{Jt})'$ and $\tilde{\boldsymbol{\Theta}}_t = (\tilde{\theta}_{ijt})_{|J|\times|J|}$. There are several things worth mentioning.

1. The reliance of sector j on sector i, $\tilde{\theta}_{jit}$, is endogeneous, depending on $\boldsymbol{A}_t$ and the structure parameters in our model. For the Cobb-Douglas case, $\tilde{\theta}_{jit} = \theta_{ji}$ is constant over time.

2. The intuition of equation 64 is very clear - the products of firm or sector i will be used either in consumption, i.e. $\alpha_i(1-\eta)$, or used as input for other sectors, i.e., $\eta \sum_j \tilde{\theta}_{jit} s_{jt}$. The importance of firm or sector i,s_{it} is defined recursively, depending on the importance of its downstream firms and their reliance on i.

3. How technology shocks affect the reallocation of resources across sectors can be fully captured by $\tilde{\boldsymbol{\Theta}}_t$, which in turn is observable from the input-output table. Thus, $\tilde{\boldsymbol{\Theta}}_t$ is the sufficient statistics capturing the reallocation effects.

4. When $v_j > 1$, the substitute effect dominates the income effect. When the price P_{it} increases, firm j can substitute away X_{it}, the effective reliance of firm j on the firm i, $\tilde{\theta}_{jit}$, declines. In partial equilibrium, if there is bad shock to firm or sector i, leading to an increase in price of firm i, we would expect firm or sector j replace away from the product i. Conditional on $s_{jt}, j \neq i$, we expect the share of firm i to decline as shown in 64.

5. When $v < 1$, the substitute effects are dominated, the firm or sector j can not fully substitute away from i to offset the price shock exposure to i. Under this case, the effective reliance on i increases.

From the first order condition of firm optimization, we have

$$
\begin{aligned}
&[s_{it}Y_t]^{1-\eta} = \lambda_{it}^{-\eta} P_{it} A_{it} \eta^{\eta}, i \in [J] \\
&\implies (1-\eta)[\log(s_{it}) + \log(Y_t)] = -\eta \log(\lambda_{it}) + \log(P_{it}) + \log(A_{it}) + \eta \log(\eta) \\
&\implies (1-\eta)[\log(\boldsymbol{s_t}) + \mathbf{1} \log(Y_t)] = -\eta \log(\boldsymbol{\lambda_t}) + \log(\boldsymbol{P_t}) + \log(\boldsymbol{A_t}) + \eta \log(\eta)\mathbf{1}
\end{aligned}
\tag{65}
$$

where $\mathbf{1} = (1, 1, ..., 1)$. Besides, from $\tilde{\theta}_{jit} = \theta_{ji}^{v_j} P_{it}^{1-v_j} \lambda_{jt}^{v_j - 1}$, we have

$$
\begin{aligned}
&(1-v_j)\log(\lambda_{jt}) = (1-v_j)\log(P_{it}) + v_j \log(\theta_{ji}) - \log(\tilde{\theta}_{jit}) \\
&\implies (1-v_j)\tilde{\theta}_{jit}\log(\lambda_{jt}) = \tilde{\theta}_{jit}[(1-v_j)\log(P_{it}) + v_j \log(\theta_{ji}) - \log(\tilde{\theta}_{jit})] \\
&\implies (1-v_j)\log(\lambda_{jt}) = (1-v_j)\sum_{i\in[J]} \tilde{\theta}_{jit}\log(P_{it}) + v_j \sum_{i\in[J]} \tilde{\theta}_{jit}\log(\theta_{ji}) - \sum_{i\in[J]} \tilde{\theta}_{jit}\log(\tilde{\theta}_{jit})
\end{aligned}
\tag{66}
$$

The last equation is obtained by summing across all i's and use $\sum_i \tilde{\theta}_{jit} = 1$.

Definition A.1 *We define*

$$
N_{jt}^{\theta} = \sum_{i\in[J]} \tilde{\theta}_{jit}\log(\tilde{\theta}_{jit}) + \frac{v_j}{1-v_j}\sum_{i\in[J]} \tilde{\theta}_{jit}\log\left(\frac{\tilde{\theta}_{jit}}{\theta_{ji}}\right), \forall v_j \neq 1
$$

with $N_{jt}^{\theta} = \sum_{i\in[J]} \theta_{ji}\log(\theta_{ji})\ \forall v_j = 1$.

where N_{jt}^{θ} is the sparsity of input-output matrix for firm j similar to Herskovic [2018]. There are two things worth mentioning. First, Herskovic [2018] obtains the measure of sparsity for Cobb-Douglas case, which is $\sum_i \tilde{\theta}_{jit}\log(\tilde{\theta}_{jit})$ and assume it is exogenous. Second, he takes the log-linear approximation around Cobb-Douglas case to approximate for the general case, we extend the sparsity definition to the general case accurately, but adjusted by an additional term $\frac{v_j}{1-v_j}\sum_{i\in[J]} \tilde{\theta}_{jit}\log\left(\frac{\tilde{\theta}_{jit}}{\theta_{ji}}\right)$.

From equation 66, we have $\log(\lambda_{jt}) = \sum_{i\in J}\tilde{\theta}_{jit}\log(P_{it}) - N_{it}^{\theta}$ [20]

$$\log(\boldsymbol{\lambda_t}) = \boldsymbol{\Theta}_t \log(\boldsymbol{P}_t) - \boldsymbol{N}_t^{\theta} \tag{67}$$

Besides, we also have normalization condition,

$$\prod[\frac{P_{it}}{\alpha_i}]^{\alpha_i} = 1 \Longrightarrow \boldsymbol{\alpha}' \log(\boldsymbol{P}_t) = \boldsymbol{\alpha}' \log(\boldsymbol{\alpha}) \tag{68}$$

A.2.2 Solution to the General Nested Networks

There are $3|J|+1$ unknown variables in equations $64, 65, 67$, and 68 - $\boldsymbol{P_t} = (P_{1t}, ..., P_{Nt})$, $\boldsymbol{s_t} = (s_{1t}, ..., s_{Nt}), \boldsymbol{\lambda_t} = (\lambda_{1t}, ..., \lambda_{Nt})$ and the aggregate output Y_t, and $3|J|+1$ of independent equations. In equilibrium, $\tilde{\theta}_{jit}$ can be solved as a function of $\boldsymbol{A}_t$. Most importantly, $\tilde{\theta}_{jit}$ can be recovered from the input-output table which enables us to obtain a closed-form solution to the general cases as a function of $\boldsymbol{A}_t$ and the observable $\tilde{\boldsymbol{\Theta}}_t$. Substitute equation 67 into 65, we obtain

$$\begin{aligned} &(1-\eta)[\log(\boldsymbol{s_t}) + \boldsymbol{1}\log(Y_t)] = \eta\boldsymbol{N}_t^{\theta} + [\boldsymbol{I} - \eta\tilde{\boldsymbol{\Theta}}_t]\log(\boldsymbol{P_t}) + \log(\boldsymbol{A_t}) + \eta\log(\eta)\boldsymbol{1} \\ &\Longrightarrow (1-\eta)[\boldsymbol{s}_t'\log(\boldsymbol{s_t}) + \log(Y_t)] = \eta\boldsymbol{s}_t'\boldsymbol{N}_t^{\theta} + (1-\eta)\boldsymbol{\alpha}'\log(\boldsymbol{\alpha}) + \boldsymbol{s}_t'\log(\boldsymbol{A}_t) + \eta\log(\eta) \end{aligned} \tag{69}$$

Here, we have used the equations 64, 68,and $\boldsymbol{s}_t'\boldsymbol{1} = 1$. Thus, the output would takes the form of

$$\log(Y_t) = \boldsymbol{s}_t'\Big[-\log(\boldsymbol{s}_t) + \frac{\eta}{1-\eta}\boldsymbol{N_t^{\theta}} + \frac{1}{1-\eta}\log(\boldsymbol{A}_t)\Big] + \frac{\eta}{1-\eta}log(\eta) + \boldsymbol{\alpha}'\boldsymbol{\alpha} \tag{70}$$

[20] In literature, researchers usually notice that

$$\lambda_{jt}^{1-v} = \sum_k \theta_{ji}^{v} P_{it}^{1-v},\ v \neq 1,\ \text{and}\ \lambda_{it} = \prod_j (\frac{P_{jt}}{\theta_{ij}})^{\theta_{ij}}\ \text{if}\ v = 1$$

For $v = 1$, we have

$$\log(\boldsymbol{\lambda}_t) = \boldsymbol{\Theta}\log(\boldsymbol{P}_t) - \boldsymbol{N^{\theta}},\ \text{if}\ v = 1$$

where $\boldsymbol{N^{\theta}} = (\sum_j \theta_{1j}\log(\theta_{1j}), ..., \sum_j \theta_{Jj}\log(\theta_{Jj}))'$ is the sparsity of the input-output networks. When $v \neq 1$, there is neither linear nor log-linear relationship between $\boldsymbol{\lambda_t}$ and $\boldsymbol{P}_t$ based on the above equation, they conclude that there is no-closed form solution to the general case. However, if we note that $\tilde{\theta}_{jit} = \theta_{ji}^{v_j} P_{it}^{1-v_j}\lambda_{jt}^{v_j-1}$, we still have a log-linear relationship between $\boldsymbol{\lambda}_t$ and $\boldsymbol{P}_t$.

Furthermore, the price in equilibrium can be determined as

$$\log(\boldsymbol{P}_t) = (\boldsymbol{I} - \eta\tilde{\boldsymbol{\Theta}}_t')^{-1}\Big[(1-\eta)\log(\boldsymbol{s_t}) + (1-\eta)\mathbf{1}\log(Y_t) - \eta\boldsymbol{N}_t^{\theta} - \log(\boldsymbol{A}_t) - \eta\log(\eta)\mathbf{1}\Big] \quad (71)$$

In the general case, the system goes beyond the Hulten's theorem [Hulten, 1978; Baqaee and Farhi, 2019] by including additional terms $\boldsymbol{s}_t'\log(\boldsymbol{s}_t)$, and $-\boldsymbol{s}_t'\boldsymbol{N_t^{\theta}}$, both of which can be estimated from the data.

B State Space Model Estimate

B.1 Estimation with TFP Data

So far we measure the innovation process $\Delta\boldsymbol{a}_t$ with patent data. One caveat of doing so is that patent issuance or filing can only measure R&D activities up to a time lag, as they are the results of the R&D several months or even years ago. In this section, we measure the innovation process with sector-level TFP data in the US, and investigate the technology shock and innovation network.

B.1.1 Model Setup

Different from patent-based measure of innovation, TFP growth can be decomposed into two parts: technological progress and improvement in management efficiency. As a result, we modify our model and divide the log TFP process $\tilde{\boldsymbol{a}}_{\boldsymbol{t}}$ into technology and efficiency components as

$$\tilde{\boldsymbol{a}}_{\boldsymbol{t}} = \boldsymbol{a_t} + \boldsymbol{m_t} \quad (72)$$

Similar to the previous model, the technological growth is knowledge diffusion plus noise:

$$\boldsymbol{\Delta a_t} = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t^a \text{ with } \boldsymbol{\epsilon}_t^a \sim \mathcal{N}(0, \sigma_a^2\boldsymbol{I})$$

here the arrival intensity of innovation $\boldsymbol{\mu}_t$ follows

$$\boldsymbol{\mu}_{t+1} = (1-\rho)\boldsymbol{\mu}_t + \boldsymbol{W}\boldsymbol{\Delta a}_t + \boldsymbol{\epsilon}_{t+1}^u = [(1-\rho)\boldsymbol{I} + \boldsymbol{W}]\boldsymbol{\mu}_t + \boldsymbol{W}\boldsymbol{\epsilon}_t^a + \boldsymbol{\epsilon}_{t+1}^u, \text{ with } \boldsymbol{\epsilon}_t^u \sim \mathcal{N}(0, \sigma_u^2\boldsymbol{I}) \quad (73)$$

where $\boldsymbol{\epsilon}_t^u$ is the technology shock process, and

$$\boldsymbol{W} = \Xi \tilde{\boldsymbol{W}}, \Xi = \text{diag}\{\xi_1, ..., \xi_J\}, \tilde{\boldsymbol{W}} \text{ is from the citation matrix data.} \tag{74}$$

The growth of management efficiency $\boldsymbol{m}_t$ follows AR(1) process

$$\boldsymbol{\Delta m}_{t+1} = \rho_m \boldsymbol{\Delta m}_t + \boldsymbol{\epsilon}_{t+1}^m, \text{ with } \rho_m \text{ constant, } \boldsymbol{\epsilon}_t^m \sim \mathcal{N}(0, \sigma_m^2 \boldsymbol{I}) \text{ independent with } \boldsymbol{\epsilon}_t^u \tag{75}$$

The model can be written as a state space form with the measurement equation as:

$$\boldsymbol{\Delta \tilde{a}_t} = \boldsymbol{\mu_t} + \boldsymbol{\Delta m_t} + \boldsymbol{\epsilon}_t^a, \text{ with } \boldsymbol{\epsilon}_t^a \sim \mathcal{N}(0, \sigma_a^2 \boldsymbol{I}) \tag{76}$$

The state equation is

$$\begin{pmatrix} \boldsymbol{\mu}_{t+1} \\ \boldsymbol{\Delta m}_{t+1} \end{pmatrix} = \begin{pmatrix} (1-\rho)\boldsymbol{I} + \Xi \tilde{\boldsymbol{W}} & 0 \\ 0 & \rho_m \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\Delta m}_t \end{pmatrix} + \begin{pmatrix} \Xi \tilde{\boldsymbol{W}} \boldsymbol{\epsilon}_t^a + \boldsymbol{\epsilon}_{t+1}^u \\ \boldsymbol{\epsilon}_{t+1}^m \end{pmatrix} \tag{77}$$

In this problem, we hope to estimate $\rho, \rho_m, \Xi, \sigma_u^2, \sigma_a^2, \sigma_m^2$ and look into the interaction of the technology shock process $\boldsymbol{\epsilon}_t^u$ and the knowledge diffusion matrices.

B.1.2 Data and Estimation Results

For estimation, we use the sector level TFP data from the BEA-BLS database[21]. We use the the official multifactor productivity (MFP) estimates as the measure of $\boldsymbol{\tilde{a}_t}$ in Equation (72). The data is available at the sector level annually from 1987 to 2018, and we further construct 3-digit NAICS level TFP each year to investigate its interaction with the knowledge diffusion matrix we construct in Section **??**.

Since the TFP data only cover a shorter time period than patent data, here we further assume that $\Xi = \xi \boldsymbol{I}$. Thus we only need to estimate $\rho, \rho_m, \xi, \sigma_u^2, \sigma_a^2, \sigma_m^2$. Using the state space estimation algorithm in Appendix D, the estimation results are in Table 5. Similar to the estimation results with patent data, here $1 - \rho + \xi = 0.9763$, which is very close to 1.

[21] Data and relevant documents are available at this BLS webpage.

With the parameter estimates, we can recover the technology shock process $\boldsymbol{\epsilon}_t^u$, and investigate its interaction with the knowledge diffusion matrices over time. In Figure 9, we plot the inner product of the technology shocks and the leading eigenvector of knowledge diffusion matrices over time. The grey shadow area plots the NBER recessions. We can see that during the Great Recession, the negative TFP shock get mostly amplified through the knowledge diffusion network.

C Data Appedix

C.1 Patent Data and Construction of Diffusion Matrix

In this appendix, we describe more details on the construction the standardized innovation diffusion matrix $\tilde{\boldsymbol{W}}$. We mainly use the patent datasets constructed by Zhu [2020] who makes two significant improvements in patent citations and assignees, compared to the data by Hall, Jaffe, and Trajtenberg [2001] and Kogan, Papanikolaou, Seru, and Stoffman [2017]. First, Zhu [2020] constructs a more complete patent citation datasets, which is roughly 170% of that in Kogan, Papanikolaou, Seru, and Stoffman [2017]. Second, Zhu [2020] develop new algorithm to match the patent assignees and companies in CRSP/CompuStat where 45% of patents are matched to CRSP/CompuStat, much higher than that by Hall, Jaffe, and Trajtenberg [2001] and Kogan, Papanikolaou, Seru, and Stoffman [2017]. These improvements are important for us to construct a more complete innovation network.

Briefly speaking, the patent datasets include three sub-datasets - patent issuance, patent assignee, and patent citation. The patent issuance and assignee datasets can be traced back to 1920 and updated to 2014, while the patent citation dataset can only be traced back to 1947. At each year t, we use the patent citation in the past five years to construct the innovation diffusion matrix. Thus, our dynamic diffusion matrix $\tilde{\boldsymbol{W}}$ is available between 1952 and 2015. For the sample before 1952 or after 2015, we implicitly assume the knowledge diffusion matrices to be stable [Acemoglu, Akcigit, and Kerr, 2016a], and use the latest one to make approximation.

C.1.1 Static Diffusion Matrix

To illustrate the intuition of the construction on $\tilde{\boldsymbol{W}}$, we first consider the construction in a static context. In the next subsection, we describe in more detail about the construction of dynamic diffusion matrix.

Figure 12 shows how firms learn knowledge from others. Firm i (Walmart) put efforts and resources to establish research groups or liboratories in various technology fields. For example, the company could have several research groups - scientific computation, software, and forecasting etc. Consider the researchers in the scientific computation (technology class τ), they not only obtain insights from the new knowledge of the field scientific computation, but also from the new knowledge of other technology fields (say, like software) created by other companies like *Microsoft, IBM, and Uber etc.*. On the other hand, firm j (e.g. Microsoft) may issued many patents, spanning several technology classes in technology space $[\boldsymbol{T}]$.

Empirically, we interpret the technology fields as the classes based on patent classification maintained by US Patent Office (USPTO). Based on the usage and property of the patent, US Patent Office (USPTO) assign each patent into one or several technology classes and subclasses called USPC. Specifically, the patent is first assigned into three-digit main class, within the main class, the patent will be further assigned into more detailed 6-digit subclass. In our paper, we focus on the three-digit main classes.

To measure the knowledge flow intensity from the representative firm in industry j to i, we denote $\boldsymbol{l} = (l_{i,1}, ..., l_{i,T})$ as the patent or reference distribution of industry i over the technology space, denote $\boldsymbol{F}_q = (F_{q,1}, ..., F_{q,J}), q \in [\boldsymbol{T}]$ as the distribution of patents in technology field q over industries, and $\Omega_{\tau,q}$ as the easiness of researchers with expertise in fields τ learning insights from q. We proxy $\tilde{\boldsymbol{W}}$ for

$$\tilde{W}_{i,j} = \sum_{q,\tau} l_{i\tau}\Omega_{\tau,q}F_{q,j} \tag{78}$$

The intuition here is straightforward. The knowledge that the representative firm in industry i learns from j depends on the three parts - 1) the distribution of the resources of the representative firm in industry i in various technology fields (τ), captured by $l_{i\tau}, \tau \in [\boldsymbol{T}]$, 2) the easiness of researchers in technology class τ learning from q, captured by $\Omega_{\tau,q}$, 3) the amount of knowledge in technology field q that the representative firm in j contributes to, captured by $F_{q,j}$.

To estimate the easiness of the knowledge flow from q to τ, we construct a technology class-to-class citation matrix based on the patent-to-patent citations. Specifically, denote $Cit_{\tau,q}$ as the total number of references made by patents in technology class τ to the patents in technology class q [22]. We proxy the easiness measure as

$$\Omega_{\tau,q} = \frac{Cit_{\tau,q}}{\sum_q' Cit_{\tau,q'}}$$

C.1.2 Construction of Dynamic $\tilde{\boldsymbol{W}}$

In the previous subsection, we mainly talk about the construction of $\tilde{\boldsymbol{W}}$ in a static and abstract setting which provide useful insights to us. In this subsection, we describe how to construct it from our real patent datasets. We use the patent datasets constructed by Zhu [2020]. We construct the time-varying $\tilde{\boldsymbol{W}}_{ij}(t)$[23] as $\sum_{\tau,q\in[\boldsymbol{T}]} l_{i,\tau}\Omega_{\tau,q}F_{q,j}$.

At year t, we estimate the $\boldsymbol{l}_i(t) = (l_{i,1}(t), ..., l_{i,T}(t))$ based on five-year rolling window. Specifically, at year t, we calculate $l_{i,\tau}(t)$ as the fraction of patents, granted to firms in industry i in the past five years, that belong to technology class τ. If one patent is assigned to several technology classes (say, k classes) by US Patent Office (USPTO), we split the patent into

[22] If the patent belongs to each technology class, we equally split the patents into these classes
[23] Here, we denote as $\tilde{\boldsymbol{W}}(t)$ to indicate time-varying

the k classes equally, and accrue $1/k$ patent to each class. Obvisouly, $\sum_{\tau\in[\boldsymbol{T}]} l_{i,\tau}(t) = 1$.

To estimate the time-varying $\Omega_{\tau,q}(t)$. At year, we first find all patents, issued between year t and $t-5$, that belong to the technology class τ. If the issued patent is in several technology classes, we split it equally among these classes. For the patent in class τ issued in year $t-s$ with $0 \leq s < 5$, we count its references to patents in class q that issued between $t-s$ and $t-s-10$. Denote $Cit_{\tau,q}(t)$ as the total number of references by patents in class τ to patents in class q, we estimate $\Omega_{\tau,q}(t)$ as $\frac{Cit_{\tau,q}(t)}{\sum_{q'} Cit_{\tau,q'}}$.

Figure **??** shows the sparsity of $\Omega_{\tau,q}(t), \tau, q \in [\boldsymbol{T}]$ at year 2014. For class τ, we define its in-degree as the number of three-digit classes to which patents in τ refer in the past ten years, that is, $\#\{q \in [\boldsymbol{T}], \Omega_{\tau,q}(t) > 0\}$. Similarly, we define the out-degree as the number of classes citing the class τ, $\#\{q \in [\boldsymbol{T}], \Omega_{q,\tau}(t) > 0\}$. We can see that the both in-degree and out-degree distribution are highly left-skewed, most of the technology classes only cite or are cited by few technology classes. However, there are non-negligible technology classes citing or being cited by hundreds of other technology classes.

Finally, to estimate the $\boldsymbol{F}_q(t) = (F_{q,1}(t), ..., F_{q,J}(t)), q \in [\boldsymbol{T}]$. At year t, we first count the number of patents issued between t and $t-5$ in the technology class q, and then use the fraction attributed to the firms in industry j to estimate $F_{q,j}(t), j \in [J]$.

C.2 Sector Output Data

In this section, we describe how to construct US sector gross output $\{\mathbf{s}_j\}$ at three-digit NAICS code level from 1947 to 2018. We use data from both the Bureau of Economic Analysis (BEA) and the Bureau of Labor Statistics (BLS). Though BEA is the most authorized data provider of US industry output data, they did not provide gross output data for detailed manufacturing industries from 1977 to 1997. Fortunately, BLS construct a detailed industry output measure for 205 sectors each year from 1972 to 2018[24]. For the year from 1947 to

[24]See "Data for Researchers" on this BLS webpage, last accessed on January 6, 2020. According to BLS report which can also be downloaded from this webpage, they did a very careful job to construct the industry

1972, we use the historical gross output by industry data from BEA[25].

The 1972-2018 BLS industry output data are coded in the BLS 205 Order Industry Sectoring Plan, and BLS provide an official guidance to map their industry classification to the NAICS 2017 classification. The 1947-1971 data are coded in the BEA IO code, which could be mapped to the NAICS 2017 classification by hand[26]. After the mapping, we have industry output data from 1947 to 2018 coded in NAICS 2017 classification.

Since the original industry division was not in three-digit NAICS level, each entry of the data after the mapping may correspond to the total output of more than one three-digit NAICS industry, or may correspond to the output of some 4-digit NAICS industries. To construct $\{\mathbf{s}_j\}$ for each single three-digit NAICS industry, we first follow Acemoglu, Akcigit, and Kerr [2016a] to split the output equally to each industry if one data entry is the total output of several industries. Then we merge the data to three-digit NAICS level and calculate the sector output shares $\{\mathbf{s}_j\}$ for each three-digit NAICS industry every year from 1947 to 2018.

C.3 Sector Input-Output Data

In this section, we describe how to construct US sector input-output $\{\tilde{\mathbf{\Theta}}_t\}$ at three-digit NAICS code level for non-government sectors[27] every year between 1947 and 2018. The data source is the Bureau of Economic Analysis (BEA), including a 71-industry input-output table for each year in 1997-2018, and a 65-industry table for each year in 1963-1996, and a 46-industry table for each year in 1947-1962[28]. Since BEA release various versions of input-

output measures from 1972 to 2018. We also compare the BLS data with BEA data for 1997-2017 when BEA also provide detailed industry output for all the industries, and find that they are highly consistent with each other.

[25]The data can be found in the gross output link on this BEA webpage, last accessed on January 6, 2020.

[26]BEA only provides a verbal description for its historical industry classification, while we find the BEA historical classification matches almost perfectly to 2-digit or three-digit level NAICS 2017 classification with few exceptions.

[27]Though the BEA IO table covers all sectors including government sectors, the map between those government sectors and NAICS sectors is missing. So we follow the literature like Bigio and La'o [2016] to only focus the non-government sectors in this paper. We would also drop the government sectors when calculating the sector gross output shares.

[28]The data for 1997-2018 can be found in the "Make-Use Tables" section, and data for 1947-1962 and 1963-1996 in the "Historical Make-Use Tables" section, both on this BEA webpage, last accessed on January

output tables, we follow the literature like Bigio and La'o [2016] to use the "Use Table after Redefinitions" for 1997-2018, and use the "Use Table before Redefinitions" for 1947-1996 as the redefined data are not available then.

Similar to the sector output data, we need to map the original data in the BEA IO code to NAICS code. The 1997-2018 input-output data are provided together with a concordance between its industry classification and NAICS 2007, while the 1947-1996 data are provided with a concordance with NAICS 2012 classification. So we first map the BEA IO code to the NAICS 2007/2012, then map the NAICS 2007/2012 to NAICS 2017 to get consistent industry classification[29].

To construct $\{\tilde{\mathbf{\Theta}}_t\}$ between all the pairs of single three-digit NAICS industries, we need to split data entries that correspond to multiple NAICS industries. Similar to the work on sector output shares, we split the input-output value equally to each industry if one data entry is the total input or output of multiple NAICS industries. Then we merge the data to three-digit NAICS level to get the input-output table $\{\tilde{\mathbf{\Theta}}_t\}$ between all the three-digit NAICS industry pairs every year from 1947 to 2018.

C.4 NAICS Classification for Firms since 1925

In this paper, we classify industries following the 2017 version of the North American Industry Classification System (NAICS). However, NAICS code is generally not available in the Compustat data until 1997, the year when NAICS was introduced[30]. Also, there have been five versions of NAICS classification: 1997, 2002, 2007, 2012 and 2017, among which the same industry code could means different industries. To make our analysis more consistent, we map all public firms in the US since 1925 into the 2017 NAICS classification. As some firms may operate in different industries over the year, the mapping is constructed for every

6, 2020. Note that the BEA used to only release the input-output tables for every five years before 1997, while they recently released a new dataset including annual input-output table since 1947, which is used in this paper.

[29]For example, there are sectors 513 and 514 in NAICS 2012, which are merged into 515 in NAICS 2017.

[30]NAICS is not available in the CRSP data provided by WRDS until 2004.

pair of {firm, year}. With a CRSP dataset on the universe of US public firms since 1925 with their (Standard Industrial Classification) SIC and NAICS code each year whenever possible, the construction procedure proceed as follows:

1. Map all NAICS code to the 2017 NAICS version. For each {firm, year} pair with NAICS information available, take that NAICS code as the most recent version of NAICS in that year, and map that NAICS code into the 2017 version of NAICS classification with the official concordances provided by the Census Bureau[31].

   Among all 115,154 data entries in CRSP with NAICS information available, 110,757 (96.2%) can be mapped to the 2017 NAICS in this way, with the remaining 4,397 entries' NAICS cannot be found in the most updated version of NAICS then. There are 297 unique 6-digit NAICS codes for those 4,397 data entries, and we search for them in other versions of NAICS. 132 of them can be uniquely spotted in only one of the historical NAICS version. The remaining 140 NAICS codes can be found in more than one historical version of NAICS. For those, we take them as the latest version. This is totally acceptable, as we manually check that most of those NAICS codes are exactly concordance with each other in different versions of NAICS. There are also 25 NAICS code cannot be found in any historical versions of NAICS, so we would not make use of those NAICS information, and would leave those firms to further steps.

2. Fill in NAICS code with SIC code. After the first step, we could get roughly 26.4% of the CRSP data mapped to the 2017 NAICS, and we need to map the remaining data using their SIC information. We take all SIC codes in CRSP as the most recent 1987 version, and map them to the 2017 NAICS codes using the official concordances provided by the Census Bureau. Similar to the last step, we also search in historical SIC classifications for those SIC codes that cannot be found in the 1987 SIC, and map them into the 1987 SIC thus 2017 NAICS.

[31] The official concordances can be found on this webpage, last accessed on March 30, 2020. In practice, we merge the concordances between 1987 SIC and 2002 NAICS, 1997 NAICS and 2002 NAICS, 2002 NAICS and 2007 NAICS, 2007 NAICS and 2012 NAICS, 2012 NAICS and 2017 NAICS, so that we can get the concordances between all historical NAICS and the 2017 NAICS, and also between 1987 SIC and 2017 NAICS.

After the procedures detailed as above, we could find the 2017 NAICS industry classifications for all the US public firms since 1925. To note, one firm could belong to one or more 6-digit NAICS industries in a given year, and we would split firm's variable value into each NAICS industry when we calculate measures in the industry level. How we do the split would not matter much since most of our analysis would be done in the three-digit NAICS level (so called "subsector").

C.5 Compustat Data and Industry R&D Expenditure

The second proxy for the innovation shocks $\delta\boldsymbol{\mu}_t = (\delta\mu_{1t}, ..., \delta\mu_{Jt})$ is the realized drop in R&D expenditure at industry level. We use the Compustat North America data to calculate YoY change of quarterly[32] R&D expenditure at the three-digit NAICS level. The quarterly data on R&D expenditure begins in 1988Q3, so our measures for YoY expenditure change begins in 1989Q3 and ends in 2018Q4. For each firm, we drop their missing values of R&D expenditure before their first positive report, and set missing values after their first report as 0. We also exclude all the negative report values of R&D expenditure.

To calculate YoY change of quarterly R&D expenditure at the three-digit NAICS level from 1988Q3 to 2018Q4, we proceed as below. First, using the NAICS classification for firms constructed in Section C.4, for those firms that belong to multiple three-digit NAICS industry in a given year, we equally split their R&D expenditure value to each of the NAICS industry. Second, for each three-digit NAICS industry in each quarter, we first select all the firms that report R&D expenditures both in the current quarter and in the quarter one year ago, and then we calculate the industry value of R&D expenditure in the current quarter and in the quarter one year ago as the sum of the R&D expenditure of those firms (or subfirms) in that quarters. Finally, we drop industries with 0 total R&D expenditure one year ago, and calculate the YoY expenditure change for remaining industries.

[32] To note, the calendar quarters in the Compustat data are different from the calendar quarters we used to understand with one month difference. For example, 2007Q1 in Compustat means Feb 2007 to Apr 2007, while 2007Q4 means Oct 2007 to Jan 2008. We will take this difference into account for later calculation.

Finally, we also winsorize change values outside the 95% percentile and 5% percentile of the YoY expenditure change.

C.6 TFP Data at the Industry Level

To model the innovation process, we also use the sector level TFP data from the BEA-BLS database[33]. We use the the official multifactor productivity (MFP) estimates. The data is available at the sector level annually from 1987 to 2018, and we follow similar procedures as above to construct 3-digit NAICS level TFP each year to investigate its interaction with the knowledge diffusion matrix we construct in Section **??**.

D State Space Estimation of Innovation Network Dynamics

D.1 Derivation of Maximum Likelihood Estimators

The likelihood function

$$p(\Delta \boldsymbol{a}_{0:T}, \boldsymbol{\mu}_{0:T}|\boldsymbol{\Theta}) = p(\Delta \boldsymbol{a}_0, \boldsymbol{\mu}_0) \prod_{t=0}^{T-1} p(\Delta \boldsymbol{a}_{t+1}, \boldsymbol{\mu}_{t+1}|\Delta \boldsymbol{a}_{0:t}, \Delta \boldsymbol{\mu}_{0:t}) \tag{79}$$

Note that $p(\Delta \boldsymbol{a}_{t+1}, \boldsymbol{\mu}_{t+1}|\Delta \boldsymbol{a}_{0:t}, \Delta \boldsymbol{\mu}_{0:t}) = p(\Delta \boldsymbol{a}_{t+1}|\boldsymbol{\mu}_{t+1})p(\boldsymbol{\mu}_{t+1}|\boldsymbol{\mu}_t, \Delta \boldsymbol{a}_t)$ from our model. Thus,

$$L(\boldsymbol{\Theta}|\Delta \boldsymbol{a}_{0:T}, \boldsymbol{\mu}_{0:T}) = \log[p(\Delta \boldsymbol{a}_0, \boldsymbol{\mu}_0)] + \sum_{t=0}^{T-1}[\log(p(\Delta \boldsymbol{a}_{t+1}|\boldsymbol{\mu}_{t+1})) + \log(p(\boldsymbol{\mu}_{t+1}|\boldsymbol{\mu}_t, \Delta \boldsymbol{a}_t))] \tag{80}$$

Note that

$$\log(p(\Delta \boldsymbol{a}_{t+1}|\boldsymbol{\mu}_{t+1})) = -0.5\log(2\pi) + 0.5\log(|\boldsymbol{\Sigma}_A^{-1}|) - \frac{1}{2}Tr(\boldsymbol{\Sigma}_A^{-1}(\Delta \boldsymbol{a}_{t+1} - \boldsymbol{\mu}_{t+1})(\Delta \boldsymbol{a}_{t+1} - \boldsymbol{\mu}_{t+1})')$$

[33] Data and relevant documents are available at this BLS webpage.

and

$$\log(p(\boldsymbol{\mu}_{t+1}|\boldsymbol{\mu}_t, \Delta \boldsymbol{a}_t)) = -0.5\log(2\pi) + 0.5\log(|\boldsymbol{\Sigma}_u^{-1}|) - \frac{1}{2}Tr(\boldsymbol{\Sigma}_u^{-1}((\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t - \boldsymbol{A}\boldsymbol{\mu}_t))(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t - \boldsymbol{A}\boldsymbol{\mu}_t)') \tag{81}$$

Thus, we have

$$\begin{aligned} &L(\boldsymbol{\Theta}|\Delta\boldsymbol{a}_{0:T}, \boldsymbol{\mu}_{0:T}) = \\ &\frac{T+1}{2}\log(|\boldsymbol{\Sigma}_A^{-1}|) - \frac{1}{2}Tr(\boldsymbol{\Sigma}_A^{-1}\sum_{t=0}^{T}(\Delta\boldsymbol{a}_t\Delta\boldsymbol{a}_t' - \Delta\boldsymbol{a}_t\boldsymbol{\mu}_t' - \boldsymbol{\mu}_t\Delta\boldsymbol{a}_t' + \boldsymbol{\mu}_t\boldsymbol{\mu}_t')) \\ &+ \frac{T}{2}\log(|\boldsymbol{\Sigma}_u^{-1}|) - \frac{1}{2}Tr(\sum_{t=0}^{T-1}\boldsymbol{\Sigma}_u^{-1}((\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)' - (\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)\boldsymbol{\mu}_t'\boldsymbol{A}' \\ &- \boldsymbol{A}\boldsymbol{\mu}_t(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)' + \boldsymbol{A}\boldsymbol{\mu}_t\boldsymbol{\mu}_t'\boldsymbol{A}') + \text{constant} \end{aligned} \tag{82}$$

We take expectation of the log of likelihood conditional on the observable variables, and replace

1. $\mathbb{E}[\boldsymbol{\mu}_t|\Delta\boldsymbol{a}_{0:T}] = \boldsymbol{\mu}_{t|T}$
2. $\mathbb{E}[\boldsymbol{\mu}_t\boldsymbol{\mu}_t'|\Delta\boldsymbol{a}_{0:T}] = \boldsymbol{\mu}_{t|T}\boldsymbol{\mu}_{t|T}' + \boldsymbol{P}_{t|T}$
3. $\mathbb{E}[\boldsymbol{\mu}_t\boldsymbol{\mu}_{t+1}'|\Delta\boldsymbol{a}_{0:T}] = \boldsymbol{\mu}_{t|T}\boldsymbol{\mu}_{t+1|T}' + \boldsymbol{L}_t[\boldsymbol{P}_{t+1|T} + (\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t})\boldsymbol{\mu}_{t+1|T}']$

In our simplified case, we have restriction that $\boldsymbol{\Sigma}_A = \sigma_A^2\boldsymbol{I}, \boldsymbol{\Sigma}_u = \sigma_u^2\boldsymbol{I}, \boldsymbol{W} = \boldsymbol{\Lambda}\tilde{\boldsymbol{W}}, \boldsymbol{A} = (1-\rho)\boldsymbol{I}$. We have

$$\frac{\partial L}{\partial \sigma_A^{-2}} = 0 \Rightarrow \sigma_A^2 = \frac{1}{J(T+1)}Tr(\mathbb{E}(\sum_{t=0}^{T}(\Delta\boldsymbol{a}_t\Delta\boldsymbol{a}_t' - \Delta\boldsymbol{a}_t\boldsymbol{\mu}_t' - \boldsymbol{\mu}_t\Delta\boldsymbol{a}_t' + \boldsymbol{\mu}_t\boldsymbol{\mu}_t'))|\Delta\boldsymbol{a}_{0:T}) \tag{83}$$

$$\begin{aligned} \frac{\partial L}{\partial \sigma_u^{-2}} = 0 \Rightarrow \sigma_u^2 = &\frac{1}{JT}\sum_{t=0}^{T-1}Tr(\mathbb{E}((\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)' - (\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)\boldsymbol{\mu}_t'\boldsymbol{A}' \\ &- \boldsymbol{A}\boldsymbol{\mu}_t(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)' + \boldsymbol{A}\boldsymbol{\mu}_t\boldsymbol{\mu}_t'\boldsymbol{A}'|\Delta\boldsymbol{a}_{0:T})) \end{aligned} \tag{84}$$

$$\frac{\partial L}{\partial(1-\rho)} = 0 \Rightarrow 2(1-\rho) = \frac{Tr(\sum_{t=0}^{T-1}\mathbb{E}[(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)\boldsymbol{\mu}_t' + \boldsymbol{\mu}_t(\boldsymbol{\mu}_{t+1} - \boldsymbol{W}\Delta\boldsymbol{a}_t)'|\Delta\boldsymbol{a}_{0:T}]}{\sum_{t=0}^{T-1} Tr(\mathbb{E}[\boldsymbol{\mu}_t\boldsymbol{\mu}_t'|\Delta\boldsymbol{a}_{0:T}])}) \tag{85}$$

$$\frac{\partial L}{\partial(\lambda_i)} = 0 \Rightarrow \lambda_i = \frac{\sum_{t=0}^{T-1}\mathbb{E}[(\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t)_i(\boldsymbol{\mu_{t+1}} - \boldsymbol{A\mu_t})_{\boldsymbol{i}}|\boldsymbol{\Delta a_{0:T}}]}{\sum_{t=0}^{T-1}\mathbb{E}[((\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t)(\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t)')_{ii}|\Delta\boldsymbol{a}_{0:T}]} \tag{86}$$

D.2 Estimation with Gradient methods

Since Equations 84, 85, 86 are nonlinear functions of parameters of interest, we follow Canova [2011] (Section 6.3) to use gradient methods for estimation.

Denote $\varphi_a = \sigma_A^{-2}, \varphi_u = \sigma_u^{-2}, \varphi_\rho = 1-\rho$, the first and second order derivatives are

$$\frac{dL}{d\varphi_a} = \frac{J(T+1)}{2\varphi_a} - \frac{1}{2}\mathbb{E}_T Tr(\sum_{t=0}^{T}(\Delta\boldsymbol{a}_t\Delta\boldsymbol{a}_t' - \Delta\boldsymbol{a}_t\boldsymbol{\mu}_t' - \boldsymbol{\mu}_t\Delta\boldsymbol{a}_t' + \boldsymbol{\mu}_t\boldsymbol{\mu}_t')) \tag{87}$$

$$\frac{dL}{d\varphi_u} = \frac{JT}{2\varphi_u} - \frac{1}{2}\mathbb{E}_T Tr(\sum_{t=0}^{T-1}(\boldsymbol{\mu}_{t+1} - \Lambda\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t - (1-\rho)\boldsymbol{\mu}_t)(\boldsymbol{\mu}_{t+1} - \Lambda\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t - (1-\rho)\boldsymbol{\mu}_t) \tag{88}$$

$$\frac{d^2L}{d\varphi_u^2} = -\frac{J(T)}{2\varphi_u^2} \tag{89}$$

$$\frac{dL}{d\varphi_\rho} = -\varphi_u\mathbb{E}_T Tr(\sum_{t=0}^{T-1}(1-\rho)\boldsymbol{\mu}_t\boldsymbol{\mu}_t' - (\boldsymbol{\mu}_{t+1} - \Lambda\tilde{W}\boldsymbol{\Delta a}_t)\boldsymbol{\mu}_t) \tag{90}$$

$$\frac{d^2L}{d\varphi_\rho^2} = -\varphi_u\mathbb{E}_T Tr(\sum_{t=0}^{T-1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t') \tag{91}$$

$$\frac{dL}{d\lambda_i} = \varphi_u\mathbb{E}_T\sum_{t=0}^{T-1}[(\mu_{t+1} - (1-\rho)\mu_t)_i(\tilde{W}a_t)_i - \lambda_i(\tilde{W}a_t)_i(\tilde{W}a_t)_i] \tag{92}$$

$$\frac{d^2L}{d{\lambda_i}^2} = -\varphi_u\mathbb{E}_T\sum_{t=0}^{T-1}[(\tilde{W}a_t)_i(\tilde{W}a_t)_i] \tag{93}$$

We use the second order method - Newton Raposon Method as in Canova [2011] to find the optimum,the step size can be obtain by $-f''/f'$, we have:

The step size of update for φ_a

$$\Delta\varphi_a = \varphi_a - \frac{\varphi_a^2}{J(T+1)}\mathbb{E}_T Tr(\sum_{t=0}^{T}(\Delta\boldsymbol{a}_t\Delta\boldsymbol{a}_t' - \Delta\boldsymbol{a}_t\boldsymbol{\mu}_t' - \boldsymbol{\mu}_t\Delta\boldsymbol{a}_t' + \boldsymbol{\mu}_t\boldsymbol{\mu}_t')) \tag{94}$$

The step size of update for φ_u

$$\Delta\varphi_u = \varphi_u - \frac{\varphi_u^2}{J(T)}\mathbb{E}_T Tr(\sum_{t=0}^{T-1}[(\boldsymbol{\mu}_{t+1} - \Lambda\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t - (1-\rho)\boldsymbol{\mu}_t)(\boldsymbol{\mu}_{t+1} - \Lambda\tilde{\boldsymbol{W}}\Delta\boldsymbol{a}_t - (1-\rho)\boldsymbol{\mu}_t)']) \tag{95}$$

The step size of update for φ_ρ

$$\Delta\varphi_\rho = \frac{\mathbb{E}_T Tr(\sum_{t=0}^{T-1}(\boldsymbol{\mu}_{t+1} - \Lambda\tilde{W}\boldsymbol{\Delta a}_t)\boldsymbol{\mu}_t)}{\mathbb{E}_T Tr(\sum_{t=0}^{T-1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t')} - \varphi_\rho \tag{96}$$

The step size of update for λ_i

$$\Delta\lambda_i = \frac{\mathbb{E}_T \sum_{t=0}^{T-1}[(\mu_{t+1} - (1-\rho)\mu_t)_i(\tilde{W}a_t)_i - \lambda_i(\tilde{W}a_t)_i(\tilde{W}a_t)_i]}{\mathbb{E}_T \sum_{t=0}^{T-1}[(\tilde{W}a_t)_i(\tilde{W}a_t)_i]} \tag{97}$$

D.3 Proof of Convergence of the EM Algorithm

Proof. a. $\boldsymbol{\mu}_{t+1|t} = \mathbb{E}[\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t}] = \boldsymbol{A}\boldsymbol{\mu}_{t|t} + \boldsymbol{W}\Delta\boldsymbol{a}_t$ since $\mathbb{E}[\boldsymbol{\epsilon}_t^u|\Delta\boldsymbol{a}_{0:t}] = 0$.

b. Note that $\boldsymbol{P}_{t+1|t} = \mathbb{V}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t}) = \mathbb{V}(\boldsymbol{A}\boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t^u|\Delta\boldsymbol{a}_{0:t})$. Thus, we have

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}' + \boldsymbol{\Sigma}_u$$

c. $\Delta\boldsymbol{a}_{t+1|t} = \mathbb{E}[\Delta\boldsymbol{a}_{t+1}|\Delta\boldsymbol{a}_{0:t}] = \mathbb{E}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t}) = \boldsymbol{\mu}_{t+1|t}$ since $\mathbb{E}[\boldsymbol{\epsilon}_{t+1}^A|\Delta\boldsymbol{a}_{0:t}] = 0$.

d. $\boldsymbol{F}_{t+1|t} = \mathbb{V}(\Delta\boldsymbol{a}_{t+1}|\Delta\boldsymbol{a}_{0:t}) = \mathbb{V}(\Delta\boldsymbol{u}_{t+1}|\Delta\boldsymbol{a}_{0:t}) + \boldsymbol{\Sigma}_A = \boldsymbol{P}_{t+1|t} + \boldsymbol{\Sigma}_A$

e. $\boldsymbol{P}_{t+1|t+1} = \mathbb{V}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t+1}) = \mathbb{V}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t}, \boldsymbol{v}_{t+1})$, with $\boldsymbol{v}_{t+1} = \boldsymbol{\Delta a}_{t+1} - \Delta\boldsymbol{a}_{t+1|t}$, using the property of joint normal of $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{v}_{t+1}$, we have

$$\mathbb{V}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t},\boldsymbol{v}_{t+1}) = \mathbb{V}(\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t}) - Cov(\boldsymbol{\mu}_{t+1},\boldsymbol{v}_{t+1}|\Delta\boldsymbol{a}_{0:t})\boldsymbol{F}_{t+1|t}^{-1}Cov(\boldsymbol{\mu}_{t+1},\boldsymbol{v}_{t+1}|\Delta\boldsymbol{a}_{0:t})'$$

Note that

$$Cov(\boldsymbol{\mu}_{t+1},\boldsymbol{v}_{t+1}|\Delta\boldsymbol{a}_{0:t}) = Cov(\boldsymbol{\mu}_{t+1},\boldsymbol{\Delta a}_{t+1}|\Delta\boldsymbol{a}_{0:t}) = \boldsymbol{P}_{t+1|t}$$

Thus, we have $\boldsymbol{P}_{t+1|t+1} = \boldsymbol{P}_{t+1|t} - \boldsymbol{P}_{t+1|t}\boldsymbol{F}_{t+1|t}^{-1}\boldsymbol{P}'_{t+1|t}$

f. $\boldsymbol{\mu}_{t+1|t+1} = \mathbb{E}[\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t+1}] = \mathbb{E}[\boldsymbol{\mu}_{t+1}|\Delta\boldsymbol{a}_{0:t},\boldsymbol{v}_{t+1}] = \boldsymbol{\mu}_{t+1|t} + Cov(\boldsymbol{\mu}_{t+1},\boldsymbol{v}_{t+1}|\Delta\boldsymbol{a}_{0:t})\boldsymbol{F}_{t+1|t}^{-1}\boldsymbol{v}_{t+1}$, therefore,

$$\boldsymbol{\mu}_{t+1|t+1} = \boldsymbol{\mu}_{t+1|t} + \boldsymbol{P}_{t+1|t}\boldsymbol{F}_{t+1|t}^{-1}\boldsymbol{v}_{t+1}$$

g. $\mu_{t|T} = \mathbb{E}[\mu_t|\Delta\boldsymbol{a}_{0:T}] = \mathbb{E}[\mu_t|\Delta\boldsymbol{a}_{0:t},\boldsymbol{v}_{t+1:T}]$ since $\boldsymbol{v}_{t+1:T}$ are independent of each and independent of $\Delta\boldsymbol{a}_{0:t}$, therefore, we have

$$\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t-1} + \sum_{j\geq t} Cov(\boldsymbol{\mu_t},\boldsymbol{v}_j|\Delta\boldsymbol{a}_{0:t-1})\mathbb{V}(\boldsymbol{v}_j|\Delta\boldsymbol{a}_{0:t-1})^{-1}\boldsymbol{v}_j$$

We further analyze the results for $Cov(\boldsymbol{\epsilon_t^u},\boldsymbol{v}_j|\Delta\boldsymbol{a}_{0:t-1})$. We first write the one-step forecast error

$$\boldsymbol{v}_{t+1} = \Delta\boldsymbol{a}_{t+1} - \Delta\boldsymbol{a}_{t+1|t} = \boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_{t+1|t} + \boldsymbol{\epsilon}_{t+1}^{A} \tag{98}$$

and

$$\begin{aligned}\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_{t+1|t} &= \boldsymbol{A}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t|t}) + \boldsymbol{\epsilon}_t^u \\ &= \boldsymbol{A}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t|t-1}) - \boldsymbol{A}\boldsymbol{K}_t\boldsymbol{v}_t + \boldsymbol{\epsilon}_t^u \\ &= \boldsymbol{L}_t(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t|t-1}) - \boldsymbol{A}\boldsymbol{K}_t\boldsymbol{\epsilon}_t^A + \boldsymbol{\epsilon}_t^u\end{aligned} \tag{99}$$

with $\boldsymbol{L}_t = \boldsymbol{A} - \boldsymbol{A}\boldsymbol{K}_t$. Now we derive a formula for

$$\begin{aligned}Cov(\boldsymbol{\mu_t},\boldsymbol{v}_j|\Delta\boldsymbol{a}_{0:t-1}) &= \mathbb{E}[\boldsymbol{\mu}_t(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j|j-1} + \boldsymbol{\epsilon}_j^A)'|\Delta\boldsymbol{a}_{0:t-1}] \\ &= \mathbb{E}[\boldsymbol{\mu}_t(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j|j-1})'|\Delta\boldsymbol{a}_{0:t-1}]\end{aligned} \tag{100}$$

Moreover,

$$
\begin{aligned}
Cov(\boldsymbol{\mu_t}, \boldsymbol{v}_t | \Delta \boldsymbol{a}_{0:t-1}) &= \boldsymbol{P}_{t|t-1} \\
Cov(\boldsymbol{\mu_t}, \boldsymbol{v}_j | \Delta \boldsymbol{a}_{0:t-1}) &= \boldsymbol{P}_{t|t-1} \boldsymbol{L}_t ... \boldsymbol{L}_{j-1}, \forall j > t
\end{aligned} \tag{101}
$$

Thus, we have

$$
\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t-1} + \sum_{j \geq t} \boldsymbol{P}_{t|t-1} \boldsymbol{L}_t ... \boldsymbol{L}_{j-1} \boldsymbol{F}_j^{-1} \boldsymbol{v}_j \tag{102}
$$

From this, we can easily get a recursive formula

$$
\boldsymbol{\mu}_{t|T} - \boldsymbol{\mu}_{t|t} = \boldsymbol{P}_{t|t} \boldsymbol{A}' \boldsymbol{P}_{t+1|t}^{-1} (\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}) \tag{103}
$$

These results are the same as the standard state-space model with small modification [Anderson and Moore, 2012]. Also, note that $\boldsymbol{L}_t = \boldsymbol{A} - \boldsymbol{A}\boldsymbol{K}_t = \boldsymbol{A} - \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{t|t-1}} \boldsymbol{F}_t^{-1} = \boldsymbol{P}_{t|t} \boldsymbol{A}' \boldsymbol{P}_{t+1|t}^{-1}$

■