Information, Production Networks and Optimal Taxation*

Siyu Chen†

November 10, 2024

Click here for the most recent version

Abstract

This paper studies optimal taxation in an economy with information frictions and a production network across industries. I show that when all industries share the same information structure, production efficiency holds and optimal policy features no taxes on intermediate goods. Deviating from this benchmark, the optimal policy generally features non-zero taxes on intermediate goods and unequal taxes on consumption goods when information structure is heterogeneous across industries: The government should impose higher revenue taxes on an industry in recession when (i) it has greater information rigidity, (ii) its upstream industries have smaller information rigidity, and (iii) its input goods are also used by less informed industries. I quantify information heterogeneity across industries with a standard text analysis method. Industries exhibit varying degrees of attention to economic outcomes correlated with their exposure to business cycle shocks. The calibrated model indicates that, in response to the COVID-19 shock, China should shift its tax burden to the utility, agriculture, and transport industries. The optimal taxation leads to a welfare increase of 0.7% for the U.S. and 1.23% for China in terms of consumption, compared to the equilibrium that assumes a homogeneous information structure.

Keywords: optimal policy, production networks, informational frictions, business cycles

*I am deeply indebted to Yongseok Shin, Yu-Ting Chiang, and Yili Chien for their invaluable advice, guidance, and support. I also thank Gaetano Antinolfi, Francisco Buera, Jennifer La'O, Fei Tan, Ping Wang, and Zhen Huo for comments and suggestions, and all participants of the Economic Graduate Student Conference and Macro Study Group at WashU for helpful discussions. All errors are mine.

†Department of Economics, Washington University in St. Louis. Email: siyu.c@wustl.edu.

1 Introduction

Recently, a growing literature has emphasized the importance of production networks, where shocks can propagate through the production chain and significantly affect the entire economy. In a multisector economy with input-output linkages, the optimal taxation policy under complete information satisfies the production efficiency result, which ensures that marginal rates of transformation are equalized across technologies in different industries. The government achieves this by setting zero tax rates on intermediate goods and equalizing tax rates on consumption goods (Diamond and Mirrlees, 1971; Chari and Kehoe, 1999). However, this result relies critically on the assumption that agents possess complete information about the future state or, if uncertainty exists, that agents have common knowledge of it[1]. In contrast, sticky information (Sims, 2003, 2010), rational inattention (Mankiw and Reis, 2002), and higher-order uncertainty (Angeletos and La'O, 2020) prevent agents from fully learning the true state, leading to dispersed and heterogeneous expectations about the future across households (Guerreiro, 2023) and across firms in different industries (Song and Stern, 2024; Flynn and Sastry, 2024). Given this complexity, how should a Ramsey planner design optimal taxation in a world with production networks and incomplete information? Would the production efficiency result still hold? If not, what would the structure of the optimal taxation look like, both in theory and in practice, to minimize welfare costs from shocks like the Covid-19 pandemic?

In this paper, I address these questions within a multisector framework featuring input-output linkages across industries and information frictions. The paper makes several contributions. Theoretically, I prove that the production efficiency result holds with homogeneous information structure—that is, when the distributions of signals are the same across industries. For heterogeneous information structure, I derive a closed-form solution for the optimal taxation using perturbation of small shocks for Gaussian information and identify two key matrices within production networks that determine the optimal tax rates: input reliance matrix and output allocation matrix. The optimal policy generally features non-zero taxes on intermediate goods and unequal taxes on consumption goods. This outcome stems from the principle that commodities with inelastic demand or supply should be taxed more heavily (Ramsey, 1927; Chari and Kehoe, 1999;

[1]I follow the definition of complete (incomplete) information as outlined by Angeletos and Lian (2016). By complete information, I mean that agents have common knowledge of the economy's information set, though there may still be uncertainty about both aggregate and industry-specific shocks.

Stiglitz and Ramsey, 2015). Here, the inelasticity arises not from traditional supply or demand curves but from information frictions: when individuals have limited information about future states, they also have limited knowledge of state-contingent tax rates, meaning that taxing these industries introduces less distortion. However, taxation in one industry can affect others through input-output linkages, influencing labor supply or intermediate goods allocation. For example, in a two-sector economy with a vertical structure, taxing the downstream sector influences labor supply in the upstream sector by distorting the demand for upstream goods. Similarly, if an industry supplies two downstream industries, imposing taxes on only one downstream sector[2] disrupts the allocation of intermediate goods across industries. Consequently, this interdependence may require the government to adjust the overall tax system. I find that production networks reshape optimal taxation in a way that can be fully captured by the input reliance and output allocation matrices.

Empirically, I find that attention is consistently heterogeneous across industries, regardless of business cycle fluctuation, and that this asymmetry is positively correlated with industry's exposure to business cycle shocks. Quantitatively, I apply the model to compute optimal tax structures for both the U.S. and China, evaluating the welfare loss if the government assumes homogeneous information structure. Additionally, I compare China's 2019 tax reforms, implemented during the pandemic, with the model's optimal taxation and discuss implications for China's industrial policy.

Theory I begin with a static model in which agents receive signals about the underlying state, with the distribution of these signals varying across industries. This setting allows that agents in different industries may possess varying degrees of precision in their information about shocks. The government selects a state-contingent tax schedule to maximize social welfare, ensuring that tax revenues are sufficient to cover government spending for each realization of the state. Two key findings emerge. First, the production efficiency result holds when the information structure is homogeneous[3]. This extends the result of Diamond and Mirrlees (1971); Chari and Kehoe (1999) from a complete information environment to an incomplete information economy, as complete information can be viewed as a special case of a homogeneous information structure. Second, by applying a first-order perturbation to the economy with respect to the small

[2] Assume this industry has less information regarding either a negative productivity shock or a positive government spending shock

[3] While the information structure is homogeneous, this does not eliminate dispersed beliefs and actions among agents.

variance of shocks, we can derive the optimal tax functions in closed form. We find that the Ramsey planner is confronted with a trade-off between labor distortion in the first stage and intermediate goods distortion in the second stage. I explore how the interaction between production networks and information frictions shapes this trade-off and redefines optimal taxation through specific examples: when the production network satisfies the condition that each industry has at most one downstream industry (a structure I define as a 'Tree' network), or all industries are self-contained, or when only the most upstream industry[4] has either more or less information, our theorem implies that there are correspondingly simple rules for the Ramsey planner to set optimal taxation. Finally, I extend the model to a dynamic setting, showing that the main result holds, except that the consumption tax is no longer constant over time, even with the same information structure. In the dynamic model, the information precision from the static model is replaced by a function of the sequence of Kalman gains.

Evidence: To measure information frictions, I follow the approach of Song and Stern (2024) and apply text analysis to construct an attention index. For the U.S., I use the Securities and Exchange Commission (SEC) 10-Q filings of public firms, and for China, I use the annual reports of listed firms. The results show that attention is consistently heterogeneous across industries for both countries and is positively correlated with exposure to business cycle shocks.

Quantitative: To translate the attention index into the information precision, I follow the approach of Bui et al. (2024) and incorporate the regression model developed by Goldstein (2023). I refer to the Survey of Professional Forecasters (SPF) to calibrate for information frictions. The calibrated model indicates the optimal revenue tax rates for the U.S. are modest. The tax rates on wholesale and retail trade, manufacturing, and services are close to zero, while the government should provide slight subsidies to the FIRE and construction sectors while shifting the tax burden onto the agriculture and mining industries. In contrast, the optimal revenue tax rates for China are higher, and the Chinese government should shift its tax burden onto the utility, agriculture, and transport sectors. The Chinese government has long implemented industrial policies that subsidize selected industries. This paper finds that, considering the heterogeneous attention across industries within the manufacturing sector, industrial policy should favor more modernized industries during the pandemic. Optimal taxation that accounts

[4]The most upstream industry refers to one that uses only labor as input without relying on any other intermediate goods.

for information frictions would lead to a welfare increase of 0.7% for the U.S. and 1.23% for China, compared to a policy with uniform consumption taxes and zero revenue taxes as implied by the homogeneous information case. By some counterfactual exercises, I find that the production networks can have a fundamental impact on the optimal tax rates.

Literature This paper belongs to a large literature analyzing the optimal taxation, the so-called Ramsey problem, particularly for those that examine optimal taxation for each industry by considering production networks (Diamond and Mirrlees, 1971; Atkinson and Stiglitz, 1976; Chari et al., 1994; Chari and Kehoe, 1999; Scheuer and Werning, 2016). These studies address both linear and nonlinear taxation in either representative or heterogeneous agent models and consistently find that the Ramsey allocation satisfies production efficiency, which implies that setting tax on intermediate goods to be zero is optimal. Diamond and Mirrlees (1971) supports the idea of uniform commodity taxes when production efficiency is prioritized. Atkinson and Stiglitz (1976) discusses the conditions under which uniform taxation of goods is optimal, particularly under separable preferences between goods and leisure. The contribution of this paper to this literature is to introduce information frictions into the model, which allows me to explore how information frictions influence optimal taxation, particularly in the framework with production networks. This extension is crucial as it provides a more realistic framework for understanding fiscal policy in the world with incomplete information.

A series of recent papers do consider the interaction of information frictions and production networks (Atolia and Chahrour, 2020; Chahrour et al., 2021; Bui et al., 2024; Lian, 2021; Pellet and Tahbaz-Salehi, 2023). For instance, Chahrour et al. (2021) shows that information shocks disseminated by the media can independently drive business cycle fluctuations, using this mechanism to explain the 2009 Great Recession. Similarly, Bui et al. (2024) examines the propagation of noisy shocks and productivity shocks through production chains, finding that noise shocks exhibit greater persistence across production networks compared to TFP shocks. Pellet and Tahbaz-Salehi (2023) investigate how firms optimally select intermediate goods, revealing that firms tend to favor less volatile supply chains under conditions of incomplete information, even at the cost of foregoing more efficient options. While these studies treat policy rates as exogenous, this paper investigates optimal fiscal policy within a production network framework that incorporates information frictions.

Finally, there is a strand of literature that directly addresses optimal policy design under informational frictions. For instance, Angeletos and La'O (2020) examines optimal monetary and fiscal policy in an environment where firms face both real and nominal rigidities, but they do not account for production networks. La'O and Tahbaz-Salehi (2022) study optimal monetary policy within a production network framework, showing that optimal policy is shaped by the interaction of an industry's position within the network and the degree of price stickiness. Wang et al. (2024) extend this framework to an open economy, finding that monetary policy should place large weights on inflation in sectors with small direct or indirect import shares through downstream sectors. Fang et al. (2024) considers endogenous information rigidities, demonstrating that the optimal price stabilization index and endogenous price rigidity are jointly determined and interact with one another.

This study complements the series of works initiated by La'O and Tahbaz-Salehi (2022). While their analysis focuses on monetary policy, this paper centers on fiscal policy. Nevertheless, I find that, much like monetary policy, optimal fiscal policy is significantly shaped by the interaction between production networks and informational frictions.

Outline The rest of the paper is organized as follows: Section 2 introduces the benchmark model. Section 3 formulates the Ramsey problem by characterizing the equilibrium conditions. In Section 4, I discuss the main theorem for both homogeneous and heterogeneous information cases, using examples to illustrate the role of production networks. Section 5 calibrates the model and presents the quantitative results. Section 6 concludes. Technical proofs are mostly delegated to the Appendix.

2 The Benchmark Model

In this section, I describe the benchmark model for a static economy. In the appendix, I extend this framework into a dynamic setting. The model features a representative household with N industries, each consisting of a continuum of firms with different information about the underlying states. Based on their information, firms use intermediate inputs and labor input to produce output, which is sold both as an intermediate good to other industries and as inputs for the final consumption good. A benevolent Ramsey planner sets the fiscal policies under full commitment to maximize the welfare. Lump-sum taxes and transfers are ruled out.

2.1 Production

There is a set of N industries, denoted by $i \in \{1, \ldots, N\}$. Each industry contains a continuum of islands indexed by $k \in [0, 1]$. On island j of industry i, there is a representative firm that produces a variety y_{ij} using inputs from other industries, as well as labor. Firms in each industry employ Cobb-Douglas production technologies to transform intermediate inputs and labor into final products:

$$y_{ij} = z_i l_{ij}^{\alpha_i} \Pi_{k=1}^{N} x_{ij,k}^{a_{ij}} \tag{1}$$

where l_{ij} is the amount of labor hired by firms on island j of industry i, $x_{ij,k}$ is the quantity of good k used for production of good i on island j, α_i represents the output elasticity with respect to labor in industry i's production technology, and a_{ij} denotes the output elasticity with respect to intermediate goods from industry j, and z_i captures the Hicks-neutral productivity shock. The assumption of constant returns to scale technology implies that $\alpha_i + \sum_{j=1}^{N} a_{ij} = 1$ for all i.

The nominal profit net of taxes is given by

$$\pi_{ij} = (1 - \tau_i^{Ind}) p_i y_{ik} - w_{ik} l_{ik} - \sum_{j=1}^{N} p_j x_{ik,j} \tag{2}$$

where w_{ij} denotes the nominal wage rate in industry i and island j, p_i denotes the nominal price of goods produced by industry i, and τ_i^{Ind} denotes the revenue tax imposed on industry i.

To aggregate the goods produced by different industries into a final consumption good, there exists a final consumption goods sector. The consumption from different industries is aggregated using Cobb-Douglas technology:

$$Y = \Pi_{i=1}^{N} (\frac{c_i}{\beta_i})^{\beta_i}$$

where β_i denotes the consumption share of goods from industry i. The constant returns to scale technology implies that $\sum_{i=1}^{N} \beta_i = 1$

2.2 Household

The household has CRRA preferences over consumption C and labor n_{ij}

$$U(C, \{n_{ij}\}) = \frac{C^{1-\sigma} - 1}{1-\sigma} - \sum_{i=1}^{N} \frac{1}{\varepsilon + 1} \int_{j\in[0,1]} n_{ij}^{\varepsilon+1} dj$$

and faces a budget constraint expressed in nominal terms as

$$PC = \sum_{i=1}^{N} \int_{j\in[0,1]} [w_{ij} n_{ij} + \pi_{ij}] dj$$

where n_{ij} is labor supply in industry i and island j, P is price of final consumption goods, and π_{ij} is firm's profit in industry i and island j. The household has income sources from both wage payment and profit income. All his income is used to pay for the consumption.

2.3 Government

The government's budget constraint, in nominal terms, is given by

$$PG = \sum_{i=1}^{N} \int_{j\in[0,1]} \tau_i^{Ind} p_i y_{ij} \, dj + \sum_{i=1}^{N} \tau_i^{C} p_i c_i \tag{3}$$

where G denotes the real government spending and τ_i^C is the tax rate on consumption goods from industry i. The government can finance its spending through both consumption taxes and industry revenue taxes (both are proportional tax rates), but it can not use lump-sum taxes or transfers.

2.4 Market clearing:

Market clearing in the goods market is given by

$$Y = C + G \tag{4}$$

which is the resource constraint for the economy: the final output is either consumed by households or utilized by the government. Market clearing for industry good i is given by

$$\sum_{k=1}^{N} \int_{j\in[0,1]} x_{kj,i}\, dj + c_i = \int_{j\in[0,1]} y_{ij}\, dj \tag{5}$$

where $\int_{j\in[0,1]} x_{kj,i}\, dj$ represent the overall use of goods i as intermediate goods in industry j and c_i represents the use goods i as input for final consumption goods. The output of industry i must satisfy the combined demand for both consumption goods and intermediate goods. Market clearing for labor requires

$$n_{ij} = l_{ij} \tag{6}$$

The labor demand equals the labor supply on every island and industry.

2.5 The Informational Structure

Nature first draws a random variable s from the set S, which contains all possible states for the economy. Its probability is denoted by $\Psi(s)$. The variable s contains not only the innovation of fundamentals (Beaudry and Portier, 2006; Jaimovich and Rebelo, 2009) like the productivity of each industry $z_i(s)$ and the real government spending $G(s)$ but also the information frictions of the economy (Lorenzoni, 2009; Angeletos and La'O, 2013). The economy proceeds in two stages.

Stage 1:

In this stage, the representative household assigns one worker to each island of each industry. Unlike the perfect information economy where everyone has common knowledge about the underlying state s, I assume that workers and firms on island j of the industry i receive a noisy and idiosyncratic signal ω_{jk} about s at stage 1. I don't specify ω_{jk} as it may be arbitrary information. It may contain information not only about fundamentals but also about the beliefs of other firms. Given this information ω_{jk}, agents form their beliefs about both the underlying shocks and the actions of other agents. I denote with $\phi_i(\omega_{ij}|s)$ the probability of receiving signal ω_{ij} for agents in industry i conditional on state s, with $\phi_i(s|\omega_{ij})$ the probability of s conditional on receiving signal ω_{ij} and with $\phi_i(\omega_{ij}, s)$ the joint probability of s and ω_{ij}. The probability functions ϕ_i are dependent on i, so I allow different industries to have different information precision regarding the state s.

Firms aim to maximize the expected utility-adjusted after-tax profit. The firm's problem $\mathcal{P}_{\text{Firm}}$ is formulated as:

$$\max_{l_{ij}} \mathbb{E}_{ij} \left[U_c \frac{1}{P}[(1-\tau_i^{Ind})p_i y_{ij} - w_{ij} l_{ij} - \sum_{k=1}^{N} p_k x_{ij,k}] \right]$$
$$s.t. \quad y_{ij} = z_i l_{ij}^{\alpha_i} \Pi_{k=1}^{N} x_{ij,k}^{a_{i,j}}$$

where U_c is the stochastic discount factor measured by the marginal utility and $\mathbb{E}_{ij}$ represents the firm's expectation, which is associated with the information ω_{ij} available to it. For workers, they aim to maximize their expected utility by choosing labor supply n_{ij} based on the signal ω_{ij}. Thus, the problem of workers $\mathcal{P}_{\text{Worker}}$ is formulated as

$$\max_{n_{ij}} \mathbb{E}_{ij} \left[\frac{C^{1-\sigma}-1}{1-\sigma} - \sum_{i=1}^{N} \frac{1}{\varepsilon+1} \int_{j\in[0,1]} n_{ij}^{\varepsilon+1} dj \right]$$
$$s.t. \quad PC = \sum_{i=1}^{N} \int_{j\in[0,1]} [w_{ij} n_{ij} + \pi_{ij}] dj$$

Stage 2:

In the second stage, goods markets open, market prices $\{p_1, \ldots, p_N\}$ and tax rates $\{\tau_1^{Ind}, \ldots, \tau_N^{Ind}\}$ and $\{\tau_1^C, \ldots, \tau_N^C\}$ are realized. The true state s becomes common knowledge across all agents. Firms on islands of different industries then decide on the quantity of intermediate goods $x_{ik,j}$ from other industries based on their initial decision of labor, the prices of intermediate inputs, and the revenue tax rate to maximize their profits π_{ik}. Frims on the final consumption goods sector make decisions for the industry's goods $\{c_i\}$ to maximize their profit U, given the realized prices of and taxes on consumption goods.

2.6 Equilibrium and Ramsey Problem

The aggregate quantities $C(s), Y(s), G(s), y_i(s), c_i(s)$ are determined as functions of the state s in accordance with the specified tax rates. The labor functions $\{n_{ij}(\omega_{ij}), l_{ij}(\omega_{ij})\}$ are measurable to the signal ω_{ij} since they are decided in stage 1, and intermediate input $x_{ij}(\omega_{ij}, s)$ are measurable to the tuple (ω_{ij}, s) as it depends on the labor input at first stage and prices at the second stage. In the aggregate level, let $L_i(s)$ denote the total labor demand in industry i and $X_{ij}(s)$ denote the total demand of intermediate goods from industry i in industry i:

$$L_i(s) \equiv \int_{k\in[0,1]} l_{ik}(\omega_{ik})dk; \quad X_{ij}(s) \equiv \int_{k\in[0,1]} x_{ik,j}(\omega_{ik}, s)dk \tag{7}$$

Then I define the equilibrium for the model:

Definition of Equilibrium

An equilibrium, based on the given tax rates, is a collection of allocations:

$$\xi \equiv \{Y(s), C(s), G(s), y_i(s), c_i(s), X_{ij}(s), L_i(s), y_{ij}(\omega_{ij}, s), x_{ij,k}(\omega_{ij}, s), n_{ij}(\omega_{ij}), l_{ij}(\omega_{ij})\}$$

such that (i) $n_{ij}(\omega_{ij})$ solves the worker's problem at stage 1; (ii) $C(s)$ solve the household's problem at stage 2; (iii) $l_{ij}(\omega_{ij})$, $x_{ij}(\omega_{ij}, s)$ solve the firm's problem at both stages; (iv) the resource constraint (4) is satisfied; (v) the government's budget constraint is satisfied; and (vi) all markets clear.

Now I define the Ramsey planner's problem. The Ramsey planner chooses state-contingent tax functions:

$$\boldsymbol{\tau^{Ind}}(s) \equiv \{\tau_1^{Ind}(s), \ldots, \tau_N^{Ind}(s)\},$$
$$\boldsymbol{\tau^{C}}(s) \equiv \{\tau_1^{C}(s), \ldots, \tau_N^{C}(s)\}$$

to maximize the expected utility of the representative household $E_s\left[U(C(s), \{n_{ij}(\omega_{ij})\}\right]$. By setting different state-contingent policy functions, the economy achieves different distribution of equilibrium, and these different equilibrium distributions lead to varying levels of welfare. Consequently, the benevolent Ramsey planner utilizes tax instruments (state-contingent proportional taxes) to select an equilibrium distribution from the set of all feasible distributions that maximizes household welfare, while ensuring that tax revenues are sufficient to cover government spending, $P(s)G(s)$, for every state s.

3 Solving the Ramsey Problem

I use the primal approach to solve the Ramsey problem. To ensure that a Ramsey outcome constitutes a competitive equilibrium, I must show first that all possible allocations in the choice set of the Ramsey planner constitute a competitive equilibrium. The following proposition states the conditions:

Proposition 1 (***Conditions to Support a Competitive Equilibrium***)**.** *Given the aggregate and industrial shocks $\{G(s), \{z_i(s)\}_{i=1}^N\}$, the allocation ξ and the price $\{P(s), p_i(s), w_{ij}(\omega_{ij})\}$ can be supported as a competitive equilibrium if and only if there exist functions $\mathcal{T}_i(s)$ and $P_i(s)$ measurable to the state s, and they satisfy the following conditions:*

1. *The aggregate resource constraint:*

$$\Pi_{i=1}^J \left(\frac{z_i(s) L_i^{\alpha_i}(s) \Pi_{k=1}^N X_{ik}^{a_{ik}}(s) - \sum_{k=1}^N X_{ki}(s)}{\beta_i} \right)^{\beta_i} - G(s) = C(s) \tag{8}$$

2. *the implementability constraint:*

$$C(s)^{1-\sigma} = \sum_{i=1}^N \mathcal{T}_i(s) L_i(s) \tag{9}$$

3. *the first-order conditions for intermediate goods:*

$$\frac{a_{ij}}{\alpha_i} \mathcal{T}_i(s) L_i(s) = P_j(s) X_{ij}(s) \tag{10}$$

4. *the first-order conditions for labor:*

$$\int \left[E_{s'|\omega_{ik}} \mathcal{T}_i(s') \right]^{\frac{1}{\varepsilon}} \phi_i(\omega_{ik}|s) d\omega_{ik} = L_i(s) \tag{11}$$

Proof. Please see the appendix A. □

Ramsey problem:

Based on proposition 1, the Ramsey planner is to maximize the utility function of the representative household:

$$\int_{s\in S} [\frac{C(s)^{1-\sigma} - 1}{1-\sigma} - \frac{1}{\varepsilon+1} \sum_{i=1}^N \int [E_{s'|\omega_{ik}} \mathcal{T}_i(s')]^{\frac{\varepsilon+1}{\varepsilon}} \phi_i(\omega_{ik}|s) d\omega_{ik}] \Psi(s) ds \tag{12}$$

subject to the constraints (8) - (11). Here the disutility function of labor is replaced by using $\mathcal{T}_i(s)$ as in equilibrium

$$n_{ik}(\omega_{ik}) = \left[E_{s'|\omega_{ik}} \mathcal{T}_i(s')\right]^{\frac{1}{\varepsilon}}$$

So, the optimization only involves functions that are measurable with s. The first-order conditions of the Ramsey problem are shown in the appendix A.

4 Optimal Taxation

Definition 1. *The information structure is homogeneous iff each industry i receive signals from the same distribution ω conditional on s: $\phi_i(\omega|s) \equiv \phi(\omega|s), \forall i$.*

Proposition 2 (***Homogeneous Information Structure***)**.** *Assume there are both government spending shocks and industrial productivity shocks $\{G(s), \{z_i(s)\}_{i=1}^N\}$. If the information structure is homogeneous across industries, the optimal taxation is to set*

$$\tau_i^{Ind}(s) = 0, \ \forall i, s$$
$$\tau_i^C(s) = \tau_j^C(s), \ \forall i, j, s$$

Proof. Please see the appendix A. □

When the information is the homogeneous, the optimal taxation is to set the industrial revenue tax to be zero and the consumption goods tax to be the same. This proposition extends the production efficiency results of Diamond and Mirrlees (1971) and Chari and Kehoe (1999) from a complete information environment into an information frictional economy. I can treat complete information as a special case of homogeneous information structure. Apart from the assumption of symmetry, there are no other constraints on the structure of the signals. The information friction can take any form, with signals potentially conveying both fundamental information and beliefs about others. People can still have heterogeneous beliefs about the economy at the first stage, both within and across industries. The proposition holds as long as the information friction remains the same, which gives us the benchmark result.

What happens if the information is heterogeneous? For example, firms in industry i may have more precise knowledge about a productivity shock within their own industry than those outside of it (Fang et al., 2024). When it comes to aggregate shocks, the impact varies across industries, leading firms to allocate differing levels of attention to the shock depending on their exposure to it.

In general, the answer to this question is not easy, as it relates to the higher-order belief of agents. The way I make the problem tractable is to use the perturbation approach. The perturbation approach has been employed by Bhandari et al. (2017, 2021) to solve the Ramsey problem for representative-agent (RA) and heterogeneous-agent (HA) models with complete information [5]. This paper extends the application of this approach to the case of incomplete information. I restrict the information structure to be Gaussian.

Assumption 1. *The state s and the signal ω_{ik} are normally distributed as follows:*

$$s \sim \mathcal{N}(0, \sigma_s^2), \quad \omega_{ik} = s + u_{ik,s}, \quad u_{ik,s} \sim \mathcal{N}(0, \sigma_{is}^2)$$

For this information structure, the asymmetry in information frictions is captured by the differing variances of the noise terms σ_{is}^2 across industries. To apply the perturbation approach, consider a sequence of economies indexed by a perturbation parameter δ that scales the size of the shocks and noises:

$$s(\delta) = \delta s; \quad u_{ik,s}(\delta) = u_{ik,s}\delta \tag{13}$$

The economy with $\delta = 1$ corresponds to the economy to be approximated. When δ converges to 0, the sequence of economies converges to a deterministic economy without shocks, which can be solved easily. Equilibrium objects are approximated through a Taylor expansion with respect to δ over the sequence of economies. In the expansion of the tax function with respect to δ, when δ is small, the first-order effect dominates. The first-order expansion of the policy functions consists of two parts: the derivative with respect to the shock δs and the derivative with respect to the scalar δ (the scale of variance of shocks & noises). Due to certainty equivalence, the second component is zero[6], so the tax functions are approximated as linear functions of the shock s:

[5] As previously discussed, the complete information here does not necessarily rule out the uncertainty of the fundamental, but it does rule out the uncertainty of the economy's information set.

[6] For higher-order approximations, the derivative with respect to δ is not zero

$$\tau_i^C(\delta s;\delta) \approx \bar{\tau}_i^C + \frac{\boldsymbol{\partial \tau_i^C}}{\boldsymbol{\partial s}}|_{\delta=\mathbf{0}}\delta s + \underbrace{\frac{\partial \tau_i^C}{\partial \delta}}_{=0\text{ (CE)}}\delta; \quad \tau_i^{Ind}(\delta s;\delta) = \underbrace{\bar{\tau}_i^{ind}}_{=0\text{ (PE)}} + \frac{\boldsymbol{\partial \tau_i^{Ind}}}{\boldsymbol{\partial s}}|_{\delta=\mathbf{0}}\delta s + \underbrace{\frac{\partial \tau_i^{Ind}}{\partial \delta}}_{=0\text{ (CE)}}\delta$$

where $\bar{\tau}_i^C$ and $\bar{\tau}_i^{Ind}$ are consumption goods and revenue tax rates in the no-shock economy. For the production efficiency result, we know $\bar{\tau}_i^{Ind} = 0$ and $\bar{\tau}_i^C$ are equalized. The expansion of government spending shock and industrial productivity shocks with respect to δ is given by:

$$\log G(\delta s) \approx \log \bar{G} + \underbrace{\frac{1}{\bar{G}}\frac{\partial G}{\partial s}|_{\delta=0}}_{\partial G}\delta s; \quad \log Z_i(\delta s) \approx \log \bar{Z}_i + \underbrace{\frac{1}{\bar{Z}_i}\frac{\partial Z_i}{\partial s}|_{\delta=0}}_{\partial Z}\delta s$$

where ∂G and ∂Z represent the percentage changes of the government spending shock and the industrial productivity shocks, respectively. I set up a list of useful notations before I go to the main theorem. For the information friction, the vector $\boldsymbol{\lambda}$ measures the precision of information across industries, where the i-th element λ_i is defined as $\frac{\sigma_{is}^{-2}}{\sigma_s^{-2}+\sigma_{is}^{-2}}$. This expression serves as the signal-to-noise ratio, reflecting the accuracy of the signal for each industry. The value of λ_i ranges from 0 to 1: 0 indicates no information ($\sigma_{is}^2 \to \infty$), while 1 indicates perfect information ($\sigma_{is}^2 = 0$). For the entire economy, I define the average precision $\bar{\lambda}$ as follows:

$$\bar{\lambda} \equiv \frac{\sum_{k=1}^N \alpha_k D_k \lambda_k}{\sum_{k=1}^N \alpha_k D_k}$$

which is a weighted average of precision across industries, where α_k is the labor share and D_k is the Domar weight of industry k. The weight $\frac{\alpha_i D_i}{\sum_{k=1}^N \alpha_k D_k}$ is associated with the ratio of the tax revenue from industry i (by using consumption good tax) in the no-shock equilibrium. I measure the difference in information frictions as $\hat{\boldsymbol{\lambda}} \equiv \boldsymbol{\lambda} - \bar{\lambda}$, which reflects the deviation of each industry's precision from the average precision. In the production network, $\boldsymbol{\alpha}$ is defined as:

$$\boldsymbol{\alpha} \equiv \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_N \end{pmatrix} \tag{14}$$

where the diagonal elements α_1 to α_N denote the output elasticity of labor for each industry. $\boldsymbol{A}$ is the matrix of a_{ij}, called the **input reliance matrix**, where each row represents the share of intermediate goods used in producing the output of each industry. $\boldsymbol{R}$ is the matrix of $R_{ij} \equiv \frac{\bar{X}_{ji}}{\sum_{k=1}^{N} \bar{X}_{ki}}$, referred to as the **output allocation matrix**, where each row shows the proportion of its output used as intermediate goods by other industries relative to its total use of intermediate goods. These fractions are computed in the deterministic economy when $\delta = 0$.

The vector $\boldsymbol{\beta}$ represents the shares of consumption goods from each industry. $\boldsymbol{D}$ is the vector of Domar weights $D_i \equiv \frac{p_i y_i}{PY}$, also evaluated in the steady state without shocks.

Theorem 2 (***Heterogeneous Information Structure***)**.** *(1) The optimal industry revenue tax $\boldsymbol{\tau^{Ind}}$ is given by*

$$\boldsymbol{\tau^{Ind}}(s) = -(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})(\boldsymbol{I} - \boldsymbol{R})(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}\hat{\boldsymbol{\lambda}}$$

(2) The optimal consumption goods tax $\boldsymbol{\tau^C}$ is given by:

$$\boldsymbol{\tau^C}(s) = \bar{\tau}^C \boldsymbol{e} - (\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})\boldsymbol{R}(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}\hat{\boldsymbol{\lambda}}.$$

where $\chi_i := \chi_i(\sigma, \varepsilon, \bar{\lambda}, \bar{G}, \bar{Y})$ and $\bar{\tau}^C := \bar{\tau}^C(\sigma, \varepsilon, \delta s, \bar{\lambda}, \bar{G}, \bar{Y}, \hat{\boldsymbol{\lambda}})$ are constants and satisfy: $\chi_Z > 0$ and $\chi_G < 0$, and $\boldsymbol{e}$ is a vector of ones.

Proof. Please see the appendix A. $\square$

I have a very simple formula for the tax function of industrial revenue and consumption goods. The last term reflects the variation in information precision. Under the same information structure, this vector becomes **0**, which directly confirms Proposition 2 within this information structure, where all revenue taxes are zero, and consumption taxes are uniform. The product of matrix $(\boldsymbol{I} - \boldsymbol{R})(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}$ and $\boldsymbol{R}(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}$ capture how the interaction of difference of precision and production networks affects the optimal taxation. This interaction is closely related to the input reliance matrix $\boldsymbol{A}$ and output allocation matrix $\boldsymbol{R}$ I formerly define. The roles of the matrices $(\boldsymbol{I}-\boldsymbol{R})$ for the revenue tax and $\boldsymbol{R}$ for the consumption goods tax, along with $\left(\boldsymbol{I}-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{R})\right)^{-1}$, will be discussed separately using examples later. $(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})$ is just a scalar which captures the response efficiency from the shocks to the tax rates. Here χ_Z is the response efficiency for productivity shock $\boldsymbol{D}^T\partial\boldsymbol{Z}$ which is positive and χ_G is the response efficiency for government spending shock $\partial\boldsymbol{G}$ which is negative. These response efficiency parameters are related to the no-shock ratio of government spending, the average information precision, and the preference parameters. As $\partial\boldsymbol{Z}$ is the vector of industrial productivity shock, the product $\boldsymbol{D}^T\partial\boldsymbol{Z}$ can be directly treated as the aggregate TFP shock. By using the Hulten theorem (Hulten, 1978; Baqaee and Farhi, 2019) , we know marginally how much industrial productivity affects the aggregate TFP is associated with its Domar weight.

Unlike the homogeneous information case, when information is heterogeneous, it's immediate that the tax rate on intermediate goods are non-zero, and the tax rate for final consumption goods are not equalized, and they are shaped by the interaction of production networks and information frictions. To explain this mechanism in detail, I refer to the examples below:

4.1 A Motivation Example:

I first study a two-sector vertical structure network (see graph *a* in Figure 1). I let industry 1 be the downstream industry and industry 2 be the upstream. From the theorem, the optimal taxes are [7]

$$\boldsymbol{\tau}^{\boldsymbol{Ind}} = (\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G}) \frac{\varepsilon}{\bar{\lambda}} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} \tag{15}$$

$$\boldsymbol{\tau}^{\boldsymbol{C}} = \bar{\tau}^C \boldsymbol{e} + (\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G}) \frac{\varepsilon}{\bar{\lambda}} \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} \tag{16}$$

where $\hat{\lambda}_1 = \lambda_1 - \bar{\lambda}$ and $\hat{\lambda}_2 = \lambda_2 - \bar{\lambda}$ and they take opposite signs. Focusing on the scenario of a positive government spending shock $\partial \boldsymbol{G} > 0$. Consider the first case where the downstream industry has less information about the government spending shock than the upstream. According to equations (15)(16), for the revenue tax, the planner should tax the downstream industry 1 and subsidize the upstream industry 2, and for the consumption good tax, the planner should comparatively increase the tax rates for the upstream industry as $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ and $\chi_G < 0$. The rationale is that when the upstream industry is less informed about government spending, it is also less informed about the future tax rate. As a result, an increase in revenue tax rates for the downstream leads to a relatively inelastic decrease in labor supply compared to the upstream industry. Therefore, the government primarily taxes the downstream industry to raise additional revenue as it distorts labor less. This rationale aligns with the literature, which emphasizes that factors that are either inelastically demanded or supplied should be taxed more heavily (Ramsey, 1927; Chari and Kehoe, 1999; Stiglitz and Ramsey, 2015). The inelasticity here doesn't originate from the supply or demand curves but instead results from information frictions in the market. However, increasing the tax rate on the downstream industry alone is not optimal as it would reduce demand for downstream goods at the second stage of production. Unlike the downstream firms, the upstream firms know more precisely that the government spending would go up, and in this sense, they know more precisely that the upstream would be taxed more heavily for their revenue, which reduces the demand for their product. Thus, the wedge created by the revenue tax is passed on to the upstream industry, and it would distort

[7] Mathematically, when $\sum_{k=1}^{N} \bar{X}_{ki} = 0$ for some i, $R_{ij} = \frac{\bar{X}_{ji}}{\sum_{k=1}^{N} \bar{X}_{ki}}$ is not good determined. This is the case when one or some industry goods are used only as consumption goods. In that case, the consumption goods tax and revenue tax are isomorphic. Since one tax instrument is redundant for those industries, there can be infinite ways of taxation to achieve the optimal equilibrium for the Ramsey planner. This pattern does not affect the validity of my theorem. $\boldsymbol{\tau}^{\boldsymbol{Ind}} + \boldsymbol{\tau}^{\boldsymbol{C}}$ perfectly cancels the term associated with matrix $\boldsymbol{R}$. So I can always set matrix $\boldsymbol{R}$ by letting R_{ij} to be 0 if it is associated with $\sum_{k=1}^{N} \bar{X}_{ki}$=0. The expression for my theorem still gives the optimal taxation even though the optimal taxation is not unique.

labor more heavily as they have more precise information. To offset this distortion, the government simultaneously subsidizes the upstream industry. In summary, the downstream taxation is justified by the inelasticity caused by information frictions, while the upstream subsidy arises from input-output linkages in the production network. [8]

In the second stage, when all tax rates are realized, if the upstream product is also used as input for consumption goods, positive revenue taxes for the downstream industry 1 would cause the allocation inefficiency as more upstream goods would be allocated to the final consumption good sector instead of the upstream industry. Thus, the Ramsey planner also increases the consumption tax for the upstream industry.

Conversely, when the upstream industry is less informed about government spending, I have $\hat{\lambda}_1 > 0 > \hat{\lambda}_2$. For the positive government spending shock, the government should tax the revenue of the upstream firms as their labor is more inelastic to the change of the tax rate. At the same time, the government subsidizes the downstream industry[9] and reduce the consumption tax for the downstream industry.

4.2 More General Cases:

Cases I: *Tree* Network:

Definition 2. *The production network is said to be a* ***Tree*** *network if every industry has at most one direct downstream industry.*

[8]The alternative way to understand this is directly looking at the first order conditions for the two sector vertical structure without using perturbation. Assuming that $\beta_1 = 1$ (only the downstream goods are used as consumption goods), I have

$$L_1(s) = E_{\omega_{1j}|s}\left[(E_{s'|\omega_{1j}}[\frac{1-\tau_1^{Ind}(s')}{1+\tau_1^C(s')} a_1 (\frac{L_2(s')}{L_1(s')})^{1-a_1}])^{\frac{1}{\varepsilon}} \right]$$

$$L_2(s) = E_{\omega_{2j}|s}\left[(E_{s'|\omega_{2j}}[\frac{(1-\tau_1^{Ind}(s'))(1-\tau_2^{Ind}(s'))}{(1+\tau_1^C(s'))} (1-a_1)(\frac{L_1(s')}{L_2(s')})^{a_1}])^{\frac{1}{\varepsilon}} \right]$$

The wedge on the downstream is only associated with the revenue tax rates on industry 1 while the wedge on the upstream is associated with revenue tax rates on both industries. Moreover, the wedge is not $\frac{1-\tau_1^{Ind}(s)}{1+\tau_1^C(s)}$ for upstream and $\frac{(1-\tau_1^{Ind}(s))(1-\tau_2^{Ind}(s))}{(1+\tau_1^C(s))}$ for downstream at state s. Instead, it is associated with $E_{\omega_{1j}|s}\left[E_{s'|\omega_{1j}}[\frac{1-\tau_1^{Ind}(s')}{1+\tau_1^C(s')}]\right]$ and $E_{\omega_{2j}|s}\left[E_{s'|\omega_{2j}}[\frac{(1-\tau_1^{Ind}(s'))(1-\tau_2^{Ind}(s'))}{(1+\tau_1^C(s'))}]\right]$ which shows how the wedges are affected by the information frictions for each industry.

[9]The government can also keep the revenue tax rate for the downstream industry to be the same as the upstream revenue tax does not distort the downstream labor supply. As discussed formerly, the revenue tax and consumption good tax for the downstream here is isomorphic, and there is an infinite way of taxation for the downstream to achieve the optimal equilibrium for the Ramsey planner.

The tree network applies for all examples in figure 1 including two sector vertical structure model as I discussed, and even an around-about production as graph ***c***:

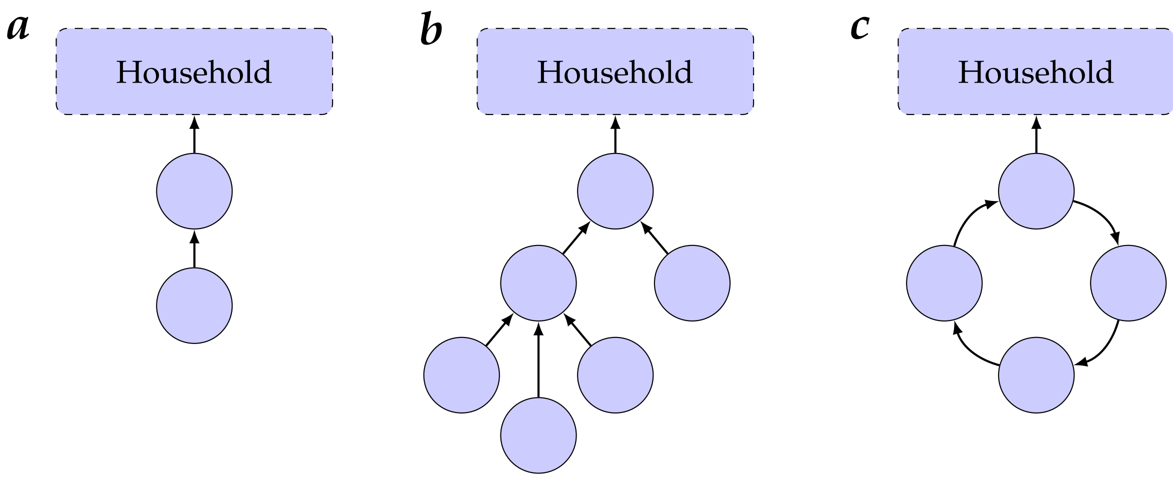

Figure 1: ***Tree*** Networks: examples

If the production network is a ***Tree*** network, I can simplify the bracket of the matrix in the theorem:

$$(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{R}))^{-1} = -\frac{\varepsilon}{\bar{\lambda}}\boldsymbol{I} \tag{17}$$

Therefore, the optimal tax rates become

$$\boldsymbol{\tau}^{\boldsymbol{Ind}} = \frac{\varepsilon}{\bar{\lambda}}(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})(\boldsymbol{I} - \boldsymbol{R})\hat{\boldsymbol{\lambda}} \tag{18}$$

$$\boldsymbol{\tau}^{\boldsymbol{C}} = \bar{\bar{\tau}}^C \boldsymbol{e} + \frac{\varepsilon}{\bar{\lambda}}(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})\boldsymbol{R}\hat{\boldsymbol{\lambda}} \tag{19}$$

Based on equations (18) and (19), the optimal taxation follows this pattern: given a positive government spending shock (or negative productivity shock), if industry i reduces its information precision, the government should increase the revenue tax on industry i. It should reduce the revenue tax on industry j if and only if j is the direct upstream of industry i. For the consumption goods tax, the government should increase the tax for these direct upstream industries of industry i.

The reasoning is similar to the two-sector model: when one industry becomes less informed, the revenue tax increases for that industry; at the same time, the revenue taxes are reduced for its upstream industries to eliminate further labor distortions in those industries, as well as in the industries upstream of these upstream industries, and this process continues along the production chain, affecting all connected industries. The consumption taxes of the direct upstream industries increase to remove the distortion of intermediate goods in the second stage. This pattern generalizes the findings from the two-sector vertical structure model into a more general production network, and it has been captured by the matrix $(\boldsymbol{I} - \boldsymbol{R})$ for the expression of the revenue tax and $\boldsymbol{R}$ for the expression of consumption good tax.

Cases II: Multiple Downstreams

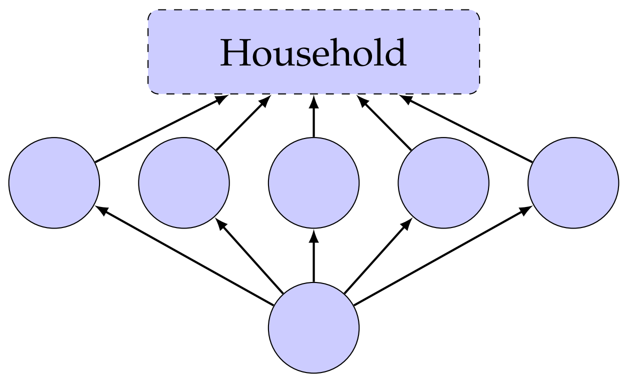

Figure 2: Multiple Downstream

To explain the role of the matrix $(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{R}))^{-1}$ in the theorem, I consider a simple production network with single upstream and multiple downstream industries shown in Figure 2. The input reliance matrix and output allocation matrix for this production network are given by:

$$A = \begin{pmatrix} 0 & \cdots & 0 & a_{1N} \\ 0 & \cdots & 0 & a_{2N} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{N-1,N} \\ 0 & \cdots & 0 & 0 \end{pmatrix}; \quad R = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ b_1 & \cdots & b_{N-1} & 0 \end{pmatrix} \tag{20}$$

where

$$b_i = \frac{\beta_i a_{iN}}{\sum_{k=1}^{N-1} \beta_k a_{kN}}$$

I have the following proposition:

Proposition 3. *For a positive government spending shock or a negative productivity shock s, if industry $i \in \{1, \cdots, N-1\}$ reduces its precision $\Delta\lambda_i < 0$ about s, then the change of the optimal taxation follows*
(1) $\Delta\tau_i^{Ind} > \Delta\tau_j^{Ind} > 0, \forall j \in \{1, \cdots, N-1\}$;
(2) $|\Delta\tau_j^{Ind}| \gtreqless |\Delta\tau_k^{Ind}|$ iff $a_{jN} \gtreqless a_{kN}$, $\forall j, k \neq i$;
(2) $\Delta\tau_N^{Ind} = -\sum_{k=1}^{N-1} b_k \Delta\tau_k^{Ind}$.

Unlike the previous case, the revenue tax increases for all downstream industries from 1 to $N-1$. If industry j relies more heavily on upstream industry N than industry k, the tax increase for industry j should be greater than that for industry k. The upstream industry N should be subsidized more (taxed less), with its tax reduction being the weighted average of the tax increases for all its downstream industries. The weight of industry i is proportional to its consumption share β_i and its input reliance a_{iN} on upstream industry N.

The reason is as follows: if industry i reduces its precision of information, it should be taxed more, and its upstream industry N should be subsidized more. However, simply replicating this tax strategy will no longer be optimal. Raising the tax on industry i alone distorts the allocation of intermediate goods for upstream industry N, which has multiple downstream paths to consumption goods. In simpler cases, when the upstream industry has at most one downstream industry, adjusting the consumption goods tax would be sufficient to remove distortions of products used for intermediate goods and used for the consumption goods. However, with multiple downstream industries, this would not possible. Therefore, the Ramsey planner must raise the revenue tax for all downstream industries.

Yet, when the revenue tax for industry j ($j \neq i$) increases, labor in industry j is further distorted because the information precision for that industry remains unchanged. This creates a trade-off between labor distortions in the first stage and intermediate goods allocation distortions in the second stage. Consequently, the tax increase for the other downstream industries should be smaller than for industry i as their precision doesn't change but is still greater than zero to reduce the distortion of allocation.

The degree to which an industry's revenue tax rate increases depends on its reliance on upstream inputs. If industry j relies more heavily on the upstream compared with another industry k, we want to tax it more because it is associated with a larger share of intermediate goods and a smaller share of labor input. Thus, increasing its tax rate greatly reduces the distortion of intermediate goods for its upstream industry N without intriguing a large distortion of labor supply in industry j. This rationale gives us the second result.

For the last property, the increase in the subsidy for the upstream industry serves to counteract the labor distortion caused by the tax increase on its downstream industries. The weight b_i represents the fraction of labor in the industry N used to produce intermediate goods i, relative to the total labor used for all intermediate goods production. So, this weighted average gives the optimal subsidy needed to fully correct the labor distortion in the upstream industry (given that there are multiple paths it goes to the final consumption good).

To understand how it relates to $(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}$, we can think about the "expansion" of the matrix as $\boldsymbol{I} + \boldsymbol{M} + \boldsymbol{M}^2 + ...$ where $\boldsymbol{M} \equiv \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR})$. The tax changes for other industries are primarily influenced by $\boldsymbol{\alpha}^{-1}\boldsymbol{AR}$. Since $\boldsymbol{\alpha}$ is diagonal, the critical component is $\boldsymbol{AR}$. When multiplied by the difference in information precision, $\boldsymbol{AR}\hat{\boldsymbol{\lambda}}$, the term $\boldsymbol{R}\hat{\boldsymbol{\lambda}}$ indicates that taxes are adjusted for downstream industries if their upstream industry's information precision differs from the average. Then $\boldsymbol{AR}\hat{\boldsymbol{\lambda}}$ suggests taxing the upstream industries of these downstream sectors to correct intermediate goods distortions. Thus, $\boldsymbol{I} + \boldsymbol{M}$ captures the process as the previous discussion: the government first taxes the industry that has less information, subsidizes the downstream, then realizes it is not optimal and adjusts the tax for the upstream of these downstream. The process does not end here, as once the taxes of those upstream go up, the government needs to adjust the subsidy for the new downstream industries of them and then revise the tax rate for the new upstream of these new downstream again. This iterative process continues as the expansion of the inverse matrix $\left(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR})\right)^{-1}$. In summary, the role of the second matrix $\left(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR})\right)^{-1}$ is to address intermediate goods distortions in the second stage, while the first matrix $(\boldsymbol{I} - \boldsymbol{R})$ or $\boldsymbol{R}$ is to address labor distortions in the first stage.

Cases III: Most Upstream Industry

Proposition 4. *If industry i only uses labor as input, and its precision increases by $\Delta\lambda_i > 0$,* *the optimal tax should change correspondingly as*

$$\Delta\tau_j^{Ind} = 0, \forall j \neq i; \quad \Delta\tau_j^C = \Delta\tau_i^C > 0, \forall i, j \tag{21}$$

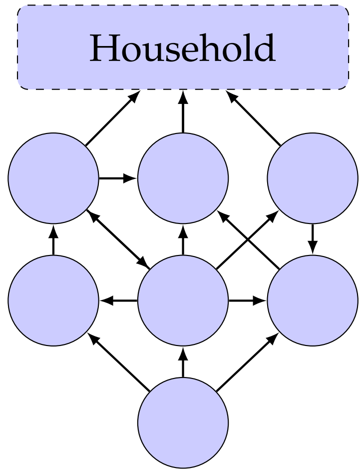

Figure 3

Proof. Without loss of generality, let's assume that industry N only uses labor as input. The input-output matrices are

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N-1,1} & a_{N-1,2} & \cdots & a_{N-1,N} \\ 0 & 0 & \cdots & 0 \end{pmatrix}; \quad R = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,N-1} & 0 \\ r_{2,1} & r_{2,2} & \cdots & r_{2,N-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ r_{N,1} & r_{N,2} & \cdots & r_{N,N-1} & 0 \end{pmatrix}$$

Then I can rewrite $(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))$ in block matrix as

$$\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}) = \begin{pmatrix} M_{(N-1)\times(N-1)} & \mathbf{0}_{(N-1)\times 1} \\ \mathbf{0}_{1\times(N-1)} & m_{1\times 1} \end{pmatrix}$$

The inverse of the matrix becomes

$$(\boldsymbol{I}-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}=\begin{pmatrix} M^{-1}_{(N-1)\times(N-1)} & \mathbf{0}_{(N-1)\times 1} \\ \mathbf{0}_{1\times(N-1)} & m^{-1}_{1\times 1}\end{pmatrix}$$

By substituting it into our theorem, I have

$$(\boldsymbol{I}-\boldsymbol{R})(\boldsymbol{I}-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{pmatrix}0\\ \vdots \\ \Delta\lambda_N\end{pmatrix}=\begin{pmatrix}0\\ \vdots \\ m^{-1}_{1\times 1}\Delta\lambda_N\end{pmatrix}$$

$$\boldsymbol{R}(\boldsymbol{I}-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{pmatrix}0\\ \vdots \\ \Delta\lambda_N\end{pmatrix}=\mathbf{0}$$

which establishes this result.

□

When the industry is the most upstream and its precision changes, only the revenue tax for that industry is affected. This is because the tax on the most upstream industry does not pass through to distort the downstream industries. Additionally, the consumption tax changes uniformly across all industries.

4.3 Discussion and Extension

According to the analysis for the static model, the optimal taxation for the economy can be summarized by the following three principles: the industry should be taxed for its revenue if (i) it has greater information rigidity, (ii) its upstream industries have smaller information rigidity, and (iii) its input goods are also used by less informed industries in recession (or when government expenditure goes up). The consumption tax changes the opposite way of the revenue tax to remove the distortion between consumption goods and intermediate goods. The above framework can be easily extended to allow for public signals and multiple shocks:

1. Public Signal: To incorporate a public signal, assume that at stage 1, each agent receives not only a private signal ω_{ij} about s but also a public signal $z \equiv s + u_z$, where u_z is normally distributed with mean zero and variance σ_z^2. The government employs tax instruments contingent on both the underlying state s and the noise shock z. For any aggregate variable, like $\mathcal{T}(\delta s, \delta z; \delta)$, a first-order perturbation yields the approximation

$$\mathcal{T}(\delta s, \delta z; \delta) \approx \bar{\mathcal{T}} + \frac{\partial \mathcal{T}}{\partial s}\delta s + \frac{\partial \mathcal{T}}{\partial z}\delta z.$$

Thus, I obtain a system of equations in δs and δz. In particular, the expectation term $\mathbb{E}_s\left[\mathbb{E}_{\omega_{ij},z}\left[s'\right]\right]$ is now given by

$$\mathbb{E}_s\left[\mathbb{E}_{\omega_{ij},z}\left[s'\right]\right] = \lambda_{i,\omega}s + \lambda_{i,z}z,$$

where

$$\lambda_{i,\omega} = \frac{\sigma_{is}^{-2}}{\sigma_s^{-2} + \sigma_{is}^{-2} + \sigma_z^{-2}}, \quad \lambda_{i,z} = \frac{\sigma_z^{-2}}{\sigma_s^{-2} + \sigma_{is}^{-2} + \sigma_z^{-2}}. \tag{22}$$

When examining the equations with respect to δs, they revert to the same form as those without a public signal, except that agents' information extracted from the private signal is now weighted by $\lambda_{i,\omega}$. The theorem 2 remains valid by using this definition of information precision from the private signals.

2. Multiple Shocks: I focus initially on a single shock s but can extend the analysis to multiple shocks $\boldsymbol{s} \equiv [s_1, s_2, \ldots, s_M]'$. Given that we are applying first-order perturbations of tax functions, the optimal taxes for multiple shocks are expressed as follows:

$$\boldsymbol{\tau}^{\text{Ind}}(\boldsymbol{s}) = \sum_{i=1}^{M} \boldsymbol{\tau}^{\text{Ind}}(s_i)$$
$$\boldsymbol{\tau}^{\text{C}}(\boldsymbol{s}) = \bar{\boldsymbol{\tau}}^{\text{C}} + \sum_{i=1}^{M}\left(\boldsymbol{\tau}^{\text{C}}(s_i) - \bar{\boldsymbol{\tau}}^{\text{C}}\right).$$

The change in taxation corresponding to shock s_i is linked to the precision $\boldsymbol{\lambda}_i$ of each industry with respect to that specific shock. [10]

3. Dynamic Model: I also extend the framework into a dynamic setting with an infinite horizon where the government uses government bonds and state-contingent assets to

[10] Fang et al. (2024) compute an attention matrix based on browsing intensity. This framework can be applied to determine optimal taxation under heterogeneous industrial productivity shocks, given a known attention matrix that accounts for the varying levels of attention each industry directs toward others.

smooth the tax revenue, shocks are persistent, and agents receive a history of signals to predict the underlying state. The main lesson holds. For proposition 1, when the information structure is homogeneous consumption tax rates are the same across industries, and the revenue tax rates are all zeros. The difference is that the consumption tax rates are not constant across time. The reason is that it is optimal to tax the agents when they collectively have less information as it causes less distortion of their labor input than when they understand the shocks. The theorem 1 also holds for the heterogeneous information structure, except that the precision parameter λ_i is not the simple signal-to-noise ratio. Instead, it is the function of the sequence of Kalman gains. See the appendix for the details.

5 Quantitative

To quantitatively apply the theorem, I have to calibrate the information rigidity for different industries. This is completed by using two steps. In the first step, I use the text analysis to construct the attention index for different industries following Song and Stern (2024). In step 2, I construct the mapping from the attention index to information frictions.

5.1 Text-Based Measure

To analyze how different industries pay attention to various economic topics, I use a dictionary-based approach that counts the frequency of keywords associated with each topic. These keywords, which are detailed in the appendix, are primarily selected based on their frequency in Econoday. Econoday is a well-known service that provides notifications on major economic events and is also the source for the Bloomberg Economic Calendar. According to the model's first-order perturbation, aggregate output is a linear function of productivity. Thus, to capture the attention directed towards TFP shocks, I compute an industry's attention to output. The output-related topic is defined using six keywords: GDP, economic growth, macroeconomic conditions, construction spending, national activity, and recession.

Data for US: I use electronically available 10-Q filings from publicly listed U.S. companies, as required by the Securities and Exchange Commission (SEC), covering the period from 1994 to 2023. These quarterly filings, mandated under Regulation S-K, include audited financial statements and descriptions of business conditions. As illustrated in Figure A1, the content of Apple's 10-Q filings provides an example of such disclosures. For each firm j in industry i at time period t, the firm is marked as attentive to a specific topic s if any keywords associated with that topic are mentioned in its 10-Q filing. In such cases, the dummy variable d_{ijst} is assigned a value of 1; otherwise, it is set to 0. This can be represented as follows:

$$d_{ijst} = \mathbf{1}(\text{Total topic-}s\text{ words} > 0) \tag{23}$$

The attention index for industry i at time t for topic s is then calculated as the average of d_{ijst} values for all firms in that industry:

$$\text{Attention}_{ist} = \frac{\sum_{j=1}^{N_{it}} d_{ijst}}{N_{it}} \tag{24}$$

where N_{it} denotes the total number of firms in industry i at time t. Consequently, the attention index for industry i represents the proportion of firms that are attentive to topic s during period t.

Data for China: In contrast, for Chinese industries, an equivalent database of quarterly filings does not exist. Therefore, I rely on the annual reports of firms listed on the Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) between 2001 and 2022. The attention index for these firms is constructed using the same approach as that for the U.S., based on the frequency of keywords appearing in these reports.

Figure 1 shows the attention index for various topics, revealing heterogeneity across industries. Attention to output differs across industries in both China and the U.S. Both countries show little attention to government spending, while firms in China pay significantly more attention to fiscal policy compared to those in the U.S. Fiscal policy is associated with government grants and subsidies, which is consistent with the fact that Chinese firms often rely on government support for development. Additionally, Chinese firms show greater attention to input-output linkages, though this attention remains heterogeneous across industries [11].

To examine whether heterogeneous attention holds over time, I calculate the average attention index over 5-year epochs, except for the shorter periods of 2020–2023 for the U.S. and 2021–2022 for China. Figure 2 displays the attention to output across different epochs for both countries. The top graph shows the U.S. data, while the bottom graph shows the data for China.

The results indicate that certain industries consistently exhibit higher attention to output. This pattern reveals a persistent asymmetry in attention. In the U.S., sectors such as FIRE, construction, manufacturing, and services exhibit greater attention on output, while industries like agriculture, mining, and wholesale trade exhibit lower levels of attention. Similarly, in China, finance and construction industries are generally more attentive to output than industries such as mining and agriculture.

[11] Attention to production refers to whether a firm is concerned with its intermediate inputs from upstream or demand from downstream, without specifying the industries involved. Fang et al. (2024) analyze firms' attention allocation across all industries within the production networks.

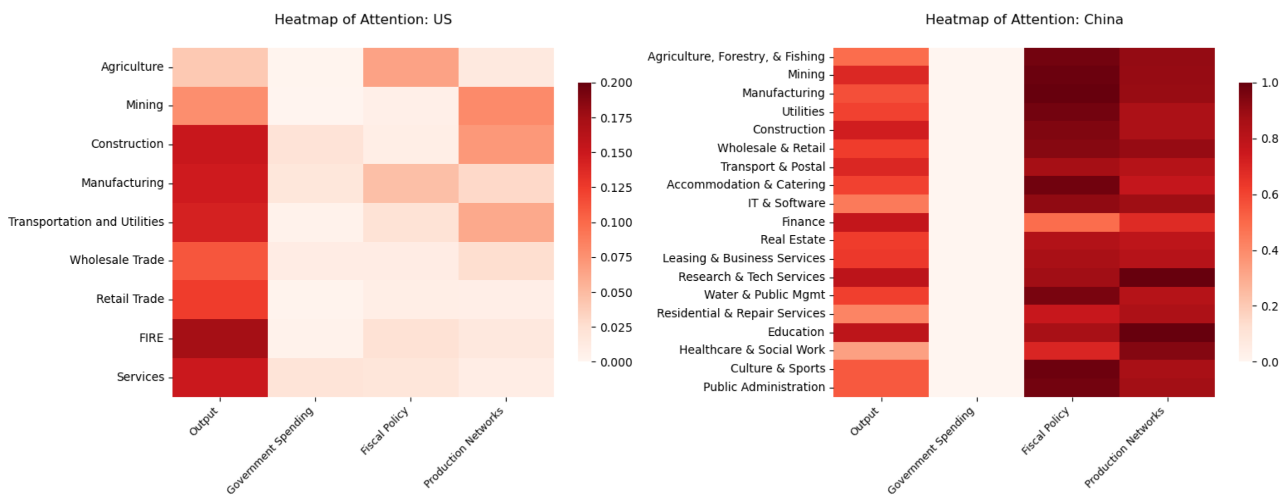

Figure 1: The heatmap of attention

Notes: The attention index is constructed based on the 10-Q filings of all publicly listed companies in the U.S. from 1994Q1 to 2023Q4, and the annual reports of all listed firms on the Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) in China from 2001 to 2022. U.S. industries are classified using the two-digit NAICS system, and Chinese industries follow the standards defined by the National Bureau of Statistics of China.

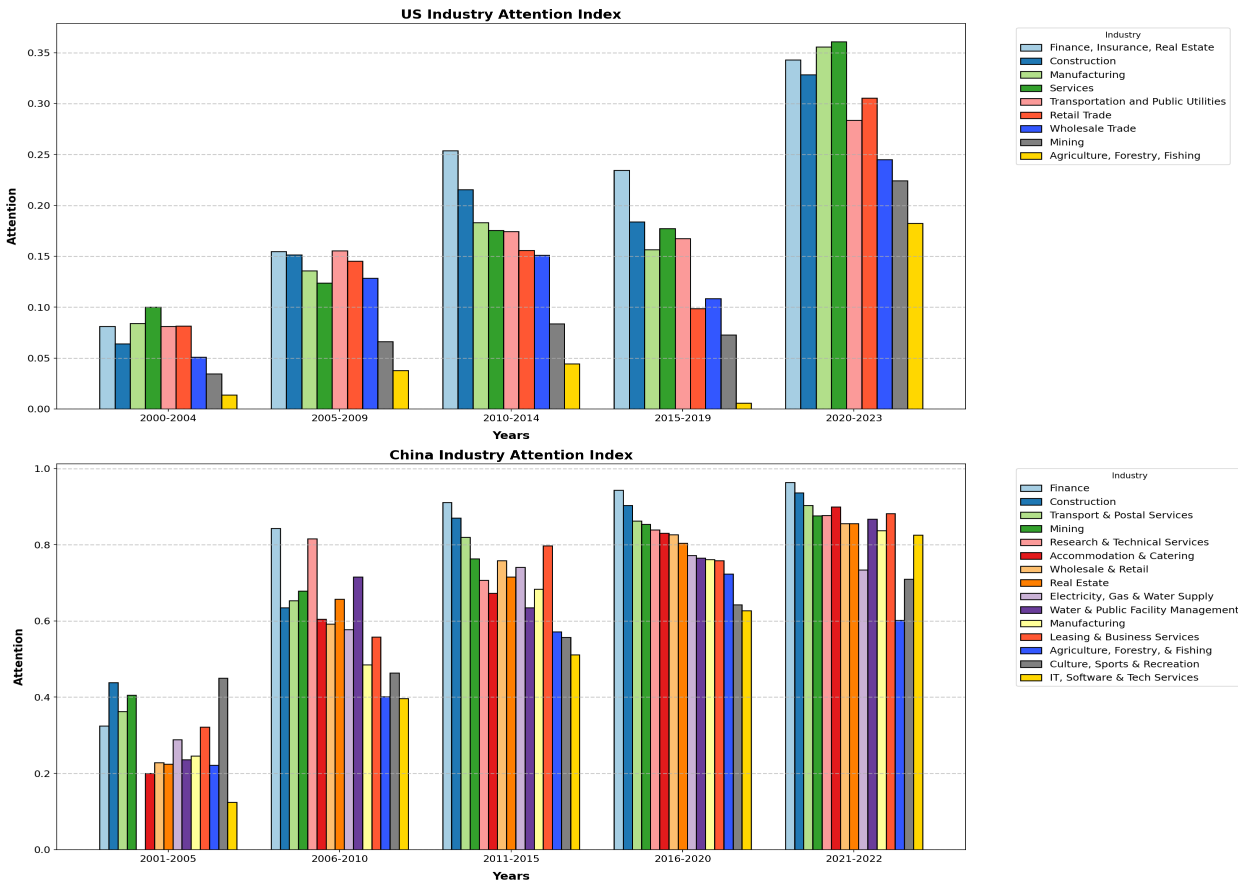

Figure 2: Industry Attention Over Time

Notes: The attention index is constructed by averaging quarterly attention values over specified ranges of year.

To check whether the constructed attention index is a good proxy for information frictions, I conduct regressions on both forecast error and forecast dispersion. I use data from the Survey of Professional Forecasters (SPF). In particular, I use the dataset of individual forecasters, focusing exclusively on those from the finance sector [12]. The forecast error is defined as the difference between the forecasted and actual real GDP values: forecast $\text{error}_{i,t} = \mathbb{E}_{i,t}[rGDP_t] - rGDP_t$. I use the standard deviation of the

[12] In the SPF, the variable industry takes a value of 1, 2, or 3, indicating the forecaster is from the finance sector, non-financial sector, or unknown, respectively. I only use forecasters who are identified as being in the finance industry.

forecast errors for all forecasters in each period to capture the information dispersion.

$$|\text{forecast error}_{i,t}| = \gamma_0 + \gamma_1 \text{Attention}_t + X_t + \varepsilon_{i,t}$$
$$\text{SD}(|\text{forecast error}_{i,t}|)_t = \gamma_0 + \gamma_1 \text{Attention}_t + X_t + \varepsilon_{i,t}$$

For the regression, I include the control variable NBER recession to account for potential increases in forecast error and dispersion during recessions. NBER recession is set to 1 during recession periods and 0 otherwise. [13].

	|Forecast Error|		SD(|Forecast Error|)	
	(1)	(2)	(1)	(2)
Attention	-0.0182*	-0.0209**	-0.0432***	-0.0457***
	(0.00949)	(0.00957)	(0.0165)	(0.0166)
NBER Recession		0.00625**		0.00632
		(0.00311)		(0.00580)
Constant	0.0184***	0.0180***	0.0210***	0.0206***
	(0.00191)	(0.00191)	(0.00333)	(0.00335)

Table 1: How attention affects the forecast error and dispersion

Table 1 shows that increased attention reduces both forecast error and dispersion. This pattern aligns with the model and validates the attention index I constructed as a proxy for information uncertainty within the industry. The negative coefficients for the NBER shock, both for forecast error and dispersion, suggest that uncertainty does increase in recession if attention remains unchanged. This outcome may result from larger volatility in aggregate TFP or increased subjective uncertainty during economic downturns (Chiang, 2023; Flynn and Sastry, 2024).

What drives heterogeneous attention across industries? Table 3 shows that an industry's attention to output is positively associated with its exposure to business cycle shocks. Industry exposure to these shocks is measured by the correlation between the growth rate of industry output and the growth rate of GDP, using data from the Bureau

[13] According to the SPF classification, financial service providers encompass institutions in sectors such as insurance, investment and commercial banking, payment services, hedge funds, mutual funds, associations within financial services, and asset management. To align with the attention index constructed based on the two-digit NAICS system, I set the attention index to reflect the Finance, Insurance, and Real Estate (FIRE) industry.

of Economic Analysis from 1997 to 2023.[14] [15]

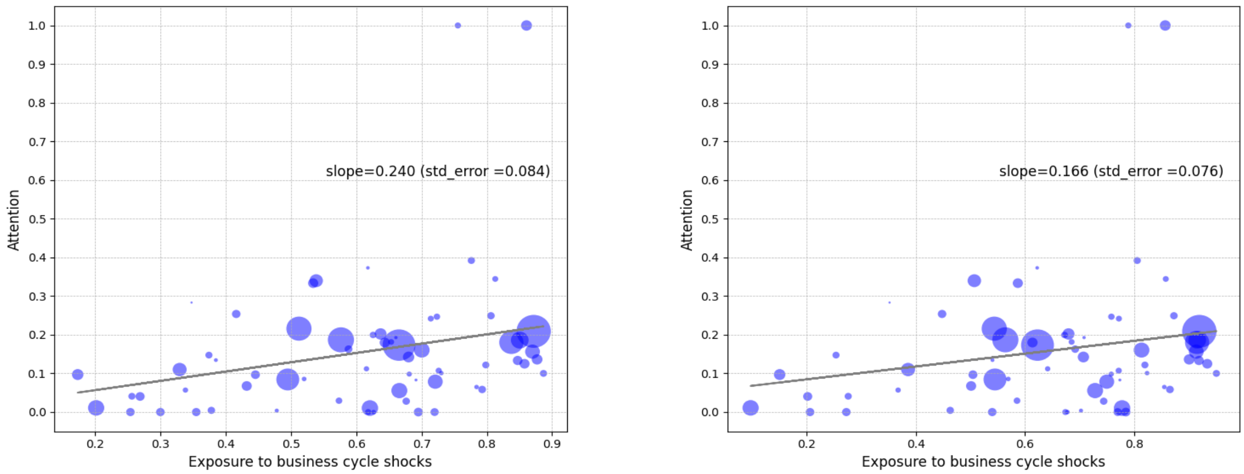

Figure 3: Exposure and attention

Note: The exposures to business shocks are computed using the correlation of the detrended growth rate of industry output and GDP growth. The left figure uses the HP filter, and the right applies a 3-year moving average. The attention is computed by taking the average of the attention index from 1997 to 2023. The industry is specified by the 3-digit NAICS system.

5.2 Regression: from attention to precision

In this section, I map the attention index into information precision. To eliminate the scale effect, I use the growth rate of real GDP instead of the level of real GDP. The forecasted growth rate of real GDP for period $t+h$ by individual i in industry j at period t is defined as:

$$E_{i,j,t}[g_{t+h}] = \frac{E_{i,j,t}[rGDP_{t+h}]}{E_{i,j,t}[rGDP_{t+h-1}]} - 1 \tag{25}$$

where forecast data is obtained from the SPF. To restrict the forecasters to be in a single industry, I only include those who are in the financial sector. When the information structure is Gaussian, forecaster i in industry j updates their forecast based on the

[14]Growth rates are detrended using either an HP filter or a 3-year moving average.

[15]In the appendix, I use an alternative measure of exposure to shocks by examining the correlation between an industry's TFP and labor productivity with the aggregate TFP and labor productivity.

history of signals. This update in the static model follows the Kalman filter:

$$E_{i,j,t}[g_t] = \left(1 - \lambda_{j,t}^{\text{private}}\right) E_{i,j,t-1}[g_t] + \lambda_{j,t}^{\text{private}} \hat{x}_{i,j,t}^{\text{private}} \tag{26}$$

where $\lambda_{j,t}^{\text{private}}$ is the Kalman gain of the private signal and $\hat{x}_{i,j,t}^{\text{private}}$ is the innovation term based on the signal received at period t. The Kalman gain $\lambda_{j,t}^{\text{private}}$ is specific to both the industry j and the time period t. Here, I restrict attention to the private signal and set the forecast horizon to the current period. For a more general information structure that incorporates public signals, see the appendix, where the model is extended to a dynamic setting. For any forecaster i, the innovation term $\hat{x}_{i,j,t}^{\text{private}}$ is uncorrelated with his previous forecast $E_{i,j,t-1}[g_t]$: he extracts only the component of the new signal that is orthogonal to all previously received ones.

I impose the key assumption: the Kalman gain of the private signal $\lambda_{j,t}^{\text{private}}$ has an affine relationship to the attention of industry j at period t:

$$\lambda_{j,t}^{\text{private}} = \beta_0 + \beta_1 * Attention_{j,t} \tag{27}$$

The assumption of this affine relationship is employed by Bui et al. (2024) to determine the precision of public signals from the intensity of news coverage. I use this assumption to map attention into precision. Then the remaining step is to get parameters β_0 and β_1.

Equation (26) implies

$$\bar{E}_{j,t}[g_t] = (1 - \lambda_{j,t}^{\text{private}})\bar{E}_{j,t-1}[g_t] + \lambda_{j,t}^{\text{private}} \bar{x}_{j,t}^{\text{private}} \tag{28}$$

where the average forecast of the industry is updated in the same way. Combining with (26) and (28), I have

$$\underbrace{E_{i,j,t}[g_t] - \bar{E}_{j,t}[g_t]}_{\text{Forecast Difference at Period } t} = (1 - \lambda_{j,t}^{\text{private}})(\underbrace{E_{i,j,t-1}[g_t] - \bar{E}_{j,t-1}[g_t]}_{\text{Forecast Difference at Period } t-1}) + \text{error}_{i,j,t} \tag{29}$$

where the $\text{error}_{i,j,t}$ is given by the difference of innovation $\hat{x}_{i,j,t}^{\text{private}} - \bar{x}_{j,t}^{\text{private}}$, and it is uncorrelated with the past forecast $E_{i,j,t-1}[g_t]$ and $\bar{E}_{j,t-1}[g_t]$.

Substituting (27) into (29), I have

$$
\begin{aligned}
\underbrace{E_{i,j,t}[g_t] - \bar{E}_{j,t}[g_t]}_{\text{Forecast Difference at Period } t} &= (1-\beta_0)(\underbrace{E_{i,j,t-1}[g_t] - \bar{E}_{j,t-1}[g_t]}_{\text{Forecast Difference at Period } t-1}) \\
&+\beta_1(\underbrace{\bar{E}_{j,t-1}[g_t] - E_{i,j,t-1}[g_t]}_{\text{Forecast Difference at Period } t-1}) * Attention_{j,t} + \text{error}_{i,j,t}
\end{aligned} \tag{30}
$$

The regression equation (30) uses the forecast difference between individuals and the mean as both the dependent and independent variables. This approach follows Goldstein (2023), with the distinction that the Kalman gain is not constant but varies over time for different levels of attention. This approach allows me to use the individual data to get the estimation of β_0 and β_1.

Alternatively by substituting (27) into (26), I have

$$
\begin{aligned}
&\underbrace{E_{i,j,t}[g_t] - E_{i,j,t-1}[g_t]}_{\text{Forecast Revision at Period t}} = \\
&-\beta_0 E_{i,j,t-1}[g_t] - \beta_1 E_{i,j,t-1}[g_t] * Attention_{j,t} + \text{error}_{i,j,t}
\end{aligned} \tag{31}
$$

This gives me an alternative regression model to estimate the key parameters β_0 and β_1. Once I have their values, I can back out the precision of any other industry j at period t by using affine relationship (27).

Table 2 shows the regression results. The results indicate that greater attention is associated with higher precision. I refer to the first regression model of forecast difference for the baseline calibration. The quantitative analysis focuses on the COVID-19 shock, and I verify that the precision values for all industries lie within the range of 0 to 1 after 2019.

	(1) Forecast Difference	(2) Forecast Revision
β_1	0.361***	0.444**
	(0.142)	(0.204)
β_0		0.0751
		(0.0535)
$1-\beta_0$	0.289***	
	(0.0763)	

Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 2: The estimation of β_0 and β_1

Note: The first column presents the results for regression (30), using the difference between the mean forecast and individual forecasts as the dependent and independent variable. The second column shows the results for regression (31), with forecast revisions as the dependent variable.

5.3 Calibration:

To calibrate input-output linkages in the model, I use the input-out data from the Asian Development Bank (ABD) for China and the U.S. Bureau of Economic Analysis (BEA) for the United States. The calibration for both countries is shown in the table below:

	U.S.			**China**		
Param.	**Value**	**Source**	**Related to**	**Value**	**Source**	**Related to**
σ^{-1}	2	—	IES	2	—	IES
ε	1.0	—	Frisch	1.0	—	Frisch
β_i		BEA	consumption share		ADB	consumption share
α_i		BEA	labor share		ADB	labor share
a_{ij}		BEA	input-output matrix		ADB	input-output matrix
R_{ij}		BEA	input-output matrix		ADB	input-output matrix
$\frac{\bar{G}}{\bar{Y}}$	0.365	IMF	Spending-to-GDP	0.354	IMF	Spending-to-GDP
Attention$_i$	—	10-Q	attention index	—	annual report	attention index
β_0	0.361	regression	info precision	0.667	regression	info precision
β_1	0.611	regression	info precision	0.410	regression	info precision

Table 3: Calibrated Parameters

For preference, I set $\sigma^{-1} = 2$ and $\varepsilon = 1.0$. I treat the year 2019 before the COVID-19 shock as the steady state for the model. Thus I refer to the input-out table data for both countries at 2019. The steady-state government spending ratio $\frac{G}{Y}$ is set at 0.3679 for the U.S. and 0.3679 for China using data from the IMF's Public Finances in Modern

History Database at 2019. a_{ij} is computed by the share of intermediate goods to produce the output. Since the model uses only labor, α_i is computed by $1-\sum_{j=1}^{N} a_{ij}$ to ensure that the production function is a constant return to scale. R_{ij} is computed by using the nominal cost of each industry to buy intermediate goods i [16]. β_i is computed by using both household consumption and government consumption. I first calculate its consumption by households and the government and then compute the share of each industry relative to the total consumption across all industries. The attention index Attention_i is from the text analysis. For China, it is constructed by an annual report of all SSE and SZSE listed firms at 2020. For the United States, it is constructed by taking an average of the quarterly attention index at year 2020 by using the 10-Q fillings. β_0 and β_1 are from the regression[17].

The input reliance matrix A, the output allocation matrix R, and the diagonal matrix of labor share $\boldsymbol{\alpha}$ can be directly computed from the calibration. The information provision can be computed by using equation (27). For the Covid-19 shock, I set a 5 % negative TFP shock for the quantitative exercises for both countries. This level is modest. Bloom et al. (2023) find total factor productivity (TFP) fell by up to 6% during 2020-21 for U.S. The NBS of China estimates that China's GDP contracted by 6.8% year-on-year in the first quarter of 2020 due to the impact of the pandemic. This exercise applies when there is a uniform 5% decline in industrial productivity. It also holds when the decline is not uniform across industries, as long as $D^T \boldsymbol{z} = 5\%$, in accordance with the theorem.

5.4 Quantitative results:

I compute the optimal tax rates in this section. To decompose the effect of production networks, I consider two counterfactual exercises: (i) the self-contained production where I assume each industry produces without using intermediate goods from other industries (α_i is kept the same but it could use its own goods as input); (ii) the symmetric input-output structure where each industry depend equally on the rest industries (α_i is kept the same but a_{ij} is set to be equalized across j).

Figure 4 - Figure 7 shows the optimal taxation of both industrial revenue tax and consumption goods for both countries. Firstly, we know when information is homo-

[16]For the model, R_{ij} is for the real good allocation. I assume that all industries pay the same price p_i to purchase industry good i, so I can directly use the nominal cost of each industry to buy intermediate good i to compute R_{ij}.

[17]see the appendix for the calibration of β_0 and β_1 for China

geneous, all the revenue tax lies on the x-axis (the production efficiency result). The red bars show optimal tax rates by using the 2019 input-output table as calibration for the network. The green and blue bars show the optimal tax rates for the counterfactual analysis. The consumption goods tax changes dramatically when the production network is self-contained. At the same time, the revenue tax is all zeros. [18]. The reason is as follows. When production networks are self-contained, there is no interconnection between industries. Then, the optimal taxation seems to tax the industry with less information about the shock as its labor is more inelastic to the taxation. However, the government should strictly prefer the consumption goods tax to the revenue tax because it does not distort the allocation of output for use as intermediate goods for its own industry or input for consumption goods. However, revenue tax does distort this allocation: if one firm supplies its goods to another firm in the same industry, if the downstream firm is taxed on its revenue, it reduces its demand for the supplier's goods and leads to a less-than-optimal allocation of products to the downstream firm. In this counterfactual scenario, production networks fundamentally alter the optimal taxation: compared to the actual input-output relationships, if each industry were to rely solely on its product as input, the government would shift from taxing industrial revenue to taxing consumption goods instead [19].

[18] The self-contained network is exactly a special case for the 'Tree' networks I specify in the examples. The industry has, at most, one upstream industry, which turns out to be itself. I prove that the optimal taxation for the tree network is to subsidize (tax) the industry, which has different precision, and at the same time tax (subsidize) its direct downstream to the same level. Here, both the industry and the downstream of the industry is the industry itself. The tax and subsidy cancel each other. So the revenue remains the same. But for the consumption good tax in the ***Tree*** network, for the downstream of that industry (this is again the industry itself), it changes to remove the distortion between its used as intermediate good and as consumption input.

[19] To the extreme case, if all goods are used for consumption and there are no intermediate goods, the self-contained scenario goes back the standard horizontal structure where consumption good tax and industry revenue tax are isomorphic: one of them is redundant.

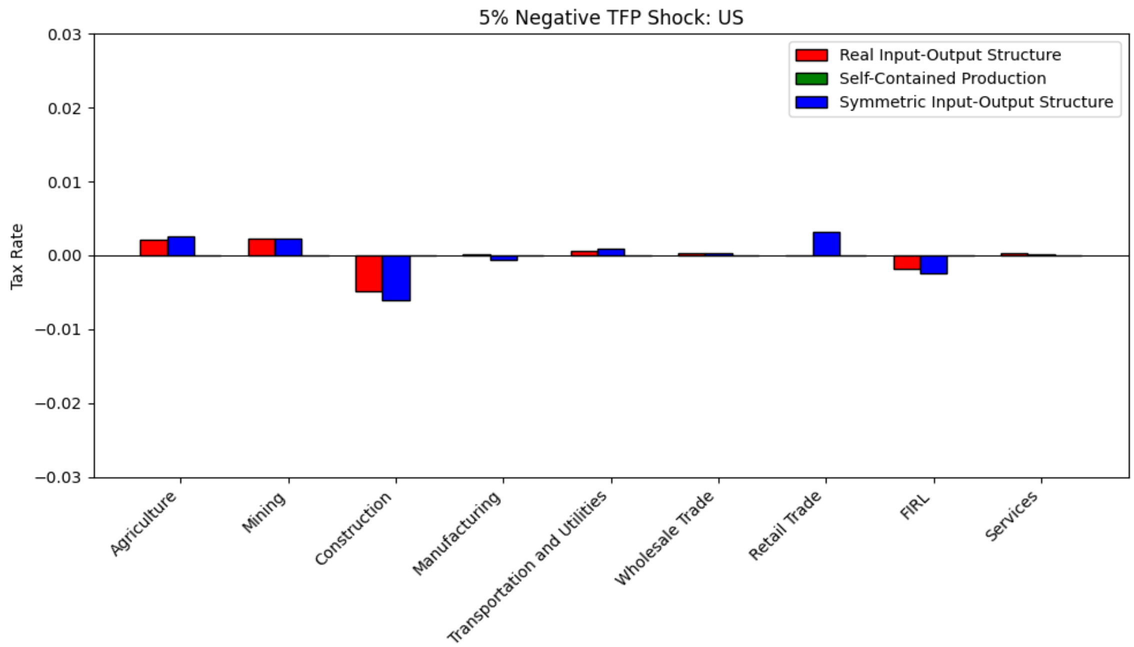

Figure 4: Optimal Industrial Revenue Tax: U.S.

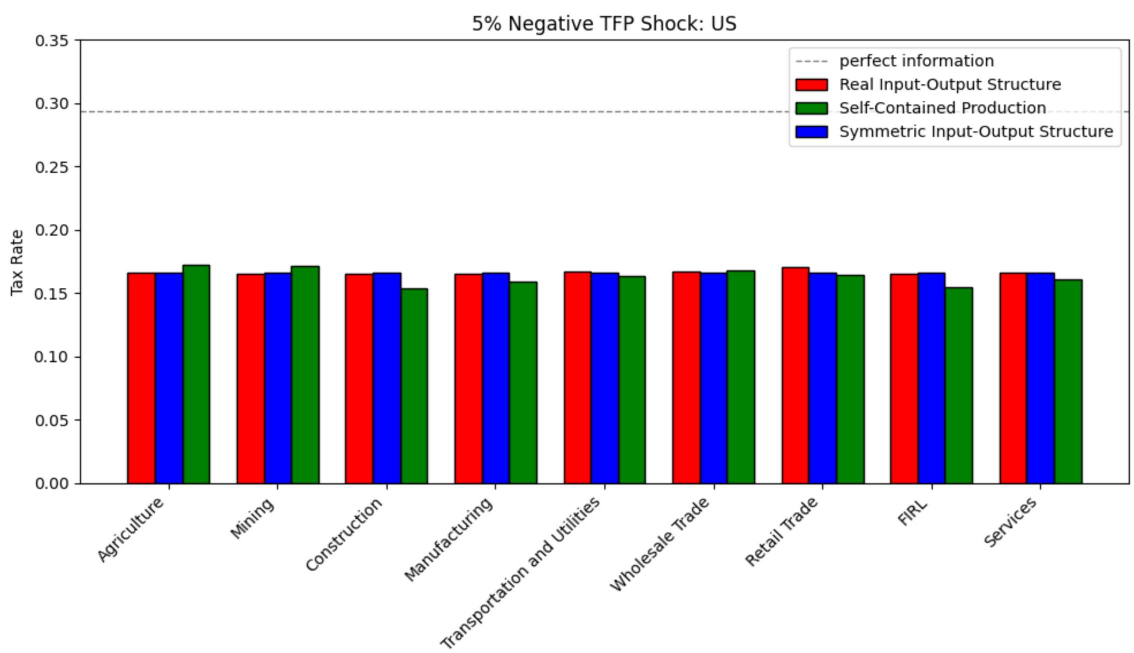

Figure 5: Optimal Consumption Good Tax: U.S.

The optimal taxation for the U.S. (red bar) is non-zero for the Covid-19 shock, but their values are modest. The tax rates of wholesale and retail trade, manufacturing and services are all close to zeros. The government should slightly subsidize FIRE and construction while shifting its tax burden on the agriculture and mining industries. [20] For the second counterfactual, when networks are symmetric across industries, the absolute tax rates mostly go up, but it does not change the sign of tax rates. The reason is that in accurate input-output linkage, the agriculture and mining industries rely greatly on intermediate goods from the manufacturing industry. Those two industries pay less attention to the output than construction, which also relies on the intermediate goods from manufacturing to production. Thus, more significant tax differences between those industries will cause an extraordinary misallocation of intermediate goods from manufacturing in the accurate input-output linkage. In the end, the optimal taxation takes smaller absolute values.

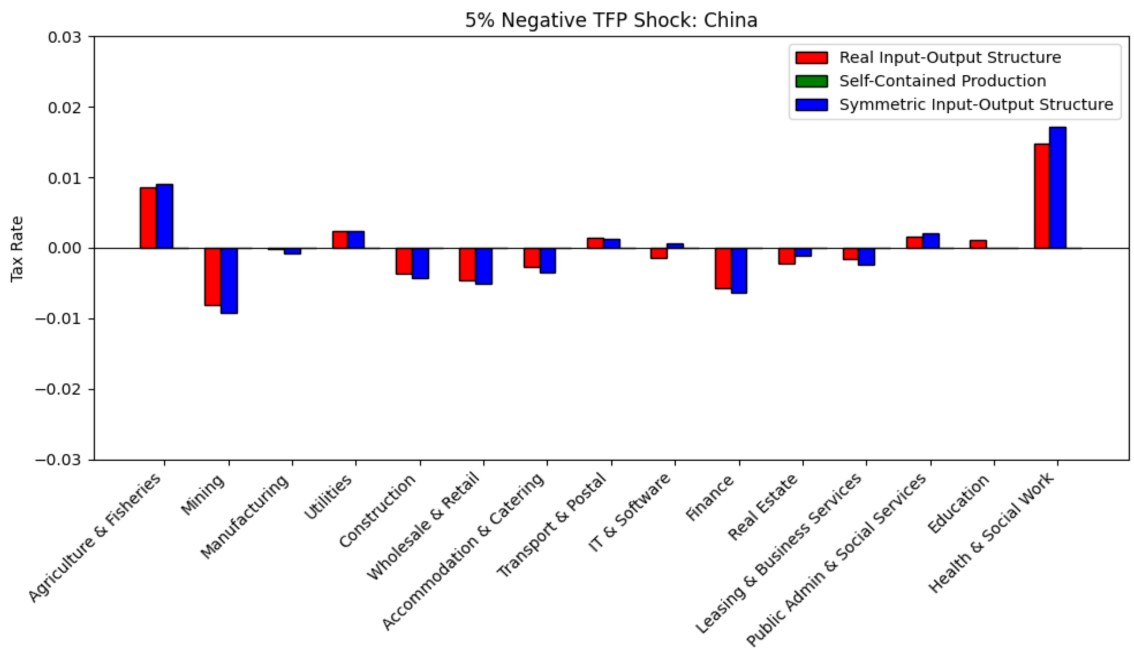

Figure 6: Optimal Industrial Revenue Tax: China

[20] I don't consider inequality here but it remains an interesting extension of this model.

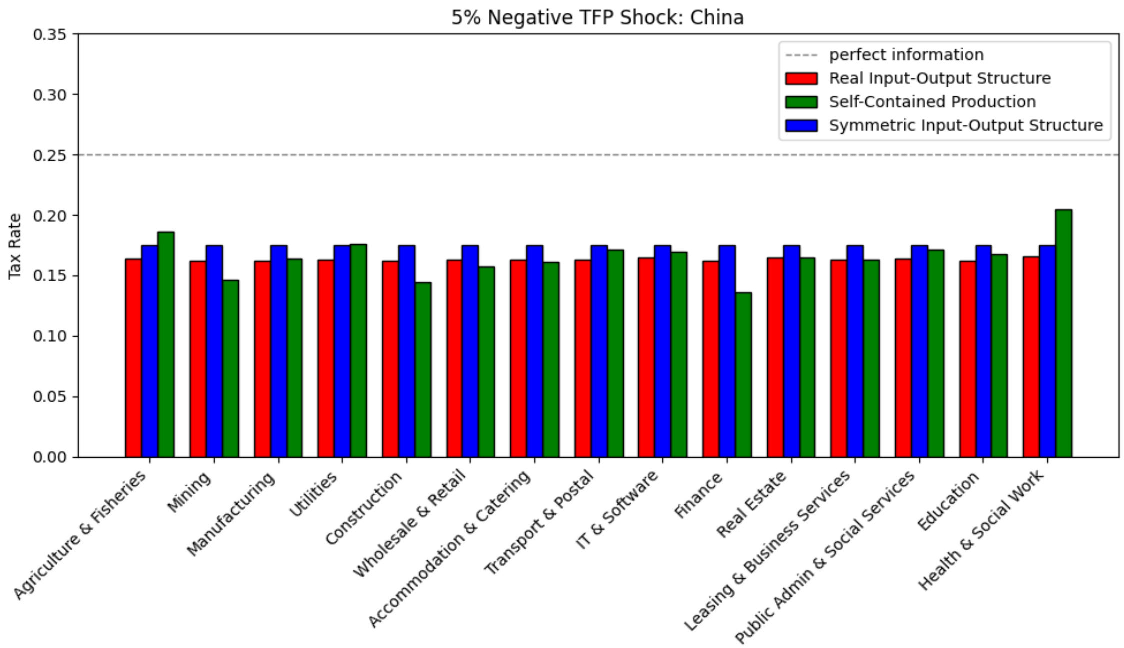

Figure 7: Optimal Consumption Good Tax: China

The optimal taxation for China is also non-zero for the Covid-19 shock, and the tax burden should be set on Agriculture, Utilities, Transport and Postal, Public Administration and Social Services, Health, and Social Work. Industries like Mining, Finance, Wholesale and Retail, Construction, and Real Estate should be subsidized the most during the pandemic. China has announced a reduced value-added tax for different industries by 2019. The tax reform reduces the tax rates for the current 16 % rate for industries including manufacturing to 13%, reduces the current 10% rate for industries such as transportation, postal services, construction, real estate, and agricultural products to 9%, and keeps the 6% rate unchanged mainly covering sectors such as I.T. and software, health care, finance, social services, and telecommunications services. [21] The quantitative results suggest that this tax reform implemented during the pandemic leads to some welfare loss as the government should subsidize the industry like finance instead of industries like agriculture and postal.

5.5 Industrial Policy in China:

To take a closer look at the case of China, we know the Chinese government has implemented so-called industrial policies to provide financial support and subsidies to

[21] For small companies in China, their value added tax base is their total revenue.

selected industries. These 'implicit' subsidies are not reflected in the uniform tax rates applied across the manufacturing sector. To investigate this, I break down the manufacturing sector into smaller industries, construct its input-output table, and compute the optimal taxation for those industries in the manufacturing sector, assuming the government can set varying tax rates on these industries through its industrial policy.

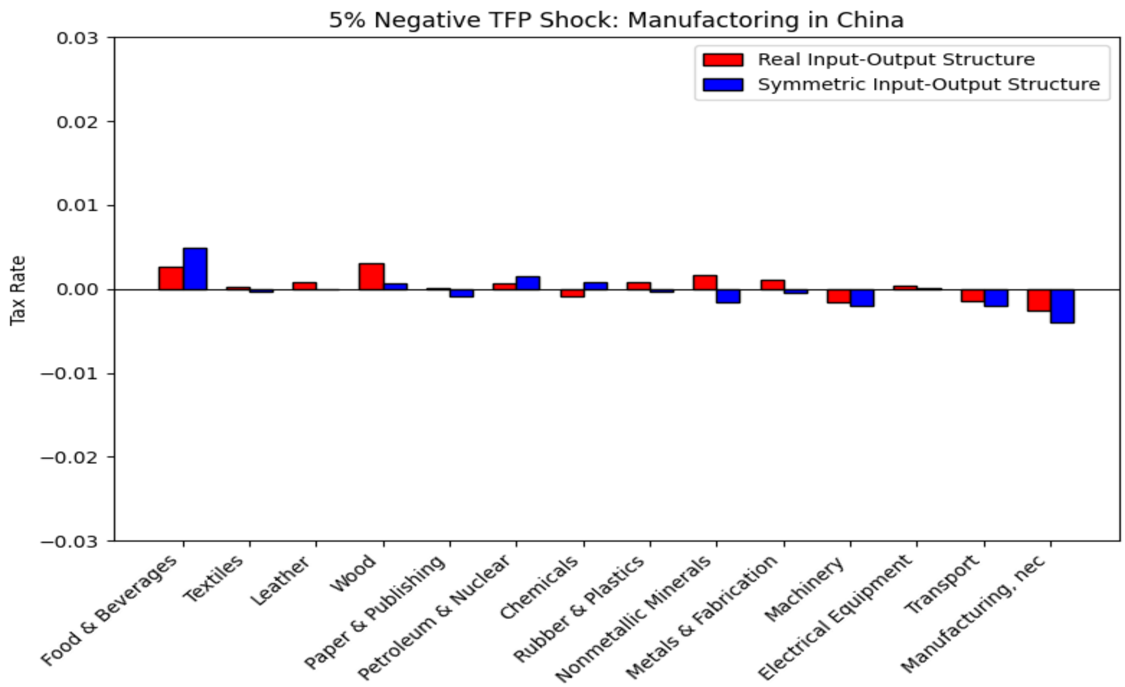

Figure 8: Optimal Revenue Tax Within Manufacturing Industry in China

Figure 8 shows the result of optimal taxes within the manufacturing sector. The tax burden primarily falls on simpler industries, such as Food and Beverage, Wood Products, and Nonmetallic Minerals within the manufacturing sector, while government subsidies are directed towards more modernized industries like Transport Equipment and Machinery. The production networks plays a crucial role in determining the sign of the optimal taxation for some industries: in a counterfactual scenario with a symmetric input-output structure, the sign of optimal taxation shifts for industries such as Metal, Nonmetallic Minerals, Chemicals, and Rubber & Plastics.

5.6 Welfare Loss:

I compute the welfare gains for both countries. Compared to the scenario where the government uses a zero revenue tax and equalized consumption tax, I calculate the percentage welfare gain when tax rates are determined by considering the varying precision of industries. For China, the welfare gain is 1.23%, and for the U.S., it is 0.7%. To check robustness, I test different values for β_0 and β_1, and the welfare gains remain non-negligible.

Welfare Gain: China		
	$\beta_1 = 0.445$	$\beta_1 = 0.667$
$\beta_0 = 0.273$	0.32%	0.71%
$\beta_0 = 0.410$	0.62%	1.23%
Welfare Gain: US		
	$\beta_1 = 0.221$	$\beta_1 = 0.331$
$\beta_0 = 0.407$	0.53%	0.56%
$\beta_0 = 0.611$	0.64%	0.70%

Table 4: Welfare gain under optimal taxation

Note: The welfare gain is computed as the percentage change in utility between the scenario where optimal taxes are imposed and the scenario where zero revenue tax and constant consumption tax are applied.

6 Conclusion

This study has explored the optimal taxation within a framework that integrates both production networks and informational frictions. By developing a model where industries are interconnected through input-output linkages and where firms possess different levels of information precision about shocks, I find how these factors jointly influence optimal tax policy:

The theoretical analysis shows that the production efficiency result holds in a homogeneous information environment. However, with heterogeneous information structure, these results must be adjusted. The Ramsey planner imposes non-zero tax rates on intermediate goods and differentiated taxes on final goods, with the tax rates determined by the precision of information available to different industries and two key matrices: the input reliance matrix and the output allocation matrix within the production network.

In the quantitative exercises, the calibrated model suggests that the optimal revenue tax rates for the U.S. are modest, while the Chinese government should shift its tax burden onto the utility, agriculture, and technology sectors. Counterfactual analysis reveals that production networks can significantly influence the determination of optimal tax rates.

References

Angeletos, George-Marios and Chen Lian, "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," in "Handbook of Macroeconomics," Vol. 2 2016, chapter 14, pp. 1065–1240.

— **and Jennifer La'O**, "Sentiments," *Econometrica*, 2013, *81* (2), 739–779.

— **and** — , "Optimal Monetary and Fiscal Policy with Incomplete Information," *Journal of Political Economy*, 2020, *128* (3), 945–999.

Atkinson, Anthony B. and Joseph E. Stiglitz, "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 1976, *6* (1-2), 55–75.

Atolia, Manoj and Ryan Chahrour, "Intersectoral linkages, diverse information, and aggregate dynamics," *Review of Economic Dynamics*, 2020, *36*, 270–292.

Baqaee, David Rezza, "Cascading Failures in Production Networks," *Econometrica*, 2018, *86* (5), 1819–1838.

— **and Emmanuel Farhi**, "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Beaudry, Paul and Franck Portier, "Stock Prices, News, and Economic Fluctuations," *American Economic Review*, 2006, *96* (4), 1293–1307.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris, "Information and Interaction," *Yale University, Cowles Foundation Discussion Paper*, 2017, *2088.*

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J. Sargent, "Fiscal Policy and Debt Management with Incomplete Markets," *The Quarterly Journal of Economics*, 2017, *132* (2), 617–663.

— **,** — **,** — **, and** — , "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, November 2021, *89* (6), 2559–2599.

— **, Jaroslav Borovička, and Paul Ho**, "Survey Data and Subjective Beliefs in Business Cycle Models," *Review of Economic Studies*, 2024. Forthcoming.

Bloom, Nicholas, Philip Bunn, Paul Mizen, Pawel Smietanka, and Gregory Thwaites, "The Impact of COVID-19 on Productivity," *The Review of Economics and Statistics*, 2023, pp. 1–45.

Bourany, Thomas, David Evans, and Mikhail Golosov, "A Perturbational Approach for Approximating Heterogeneous-Agent Models," 2023. Working Paper.

Bui, Huong, Zhen Huo, Andrei Levchenko, and Nitya Pandalai-Nayar, "Noisy Global Value Chains," *Working Paper*, 2024.

Chahrour, Ryan, Philipp J. Nimark, and Jens Pitschner, "News and Business Cycles: Revisiting the Evidence," *Review of Economic Studies*, 2021, *88* (3), 1186–1216.

Chamley, Christophe, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 1986, *54* (3), 607–622.

Chari, V. V. and Patrick J. Kehoe, "Optimal Fiscal and Monetary Policy," *Handbook of Macroeconomics*, 1999, *1C*, 1671–1745.

_, Lawrence J. Christiano, and Patrick J. Kehoe, "Optimal Fiscal Policy in a Business Cycle Model," *Journal of Political Economy*, 1994, *102* (4), 617–652.

Chiang, Yu-Ting, "Attention and fluctuations in macroeconomic uncertainty," 2023. Mimeo, Federal Reserve Bank of St. Louis.

Chien, YiLi and Yi Wen, "The Ramsey steady-state conundrum in heterogeneous-agent economies," *Journal of Economic Theory*, September 2024, *220*, 105873.

Diamond, Peter A. and James A. Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 1971, *61* (1), 8–27.

Fang, Yue, Zhen Huo, Minghao Li, and Jieran Wu, "Inattentive Network: Evidence and Theory," 2024. Working Paper.

Flynn, Joel P. and Karthik Sastry, "Attention Cycles," *Working Paper*, June 2024, (32553).

Goldstein, Nathan, "Tracking Inattention," *Journal of the European Economic Association*, December 2023, *21* (6), 2682–2725.

Guerreiro, Joao, "Belief Disagreement and Business Cycles," 2023. Working Paper.

Hulten, Charles R., "Growth Accounting with Intermediate Inputs," *The Review of Economic Studies*, 1978, *45* (3), 511–518.

Jaimovich, Nir and Sergio Rebelo, "Can News about the Future Drive the Business Cycle?," *American Economic Review*, 2009, *99* (4), 1097–1118.

Konishi, Hideki, "Final and intermediate goods taxation in an oligopolistic economy with free entry," *Journal of Public Economics*, August 1990, *42* (3), 371–386.

La'O, Jennifer and Alireza Tahbaz-Salehi, "Optimal Monetary Policy in Production Networks," *Econometrica*, May 2022, *90*, 1295–1336.

Lian, Chen, "A theory of narrow thinking," *The Review of Economic Studies*, 2021, *88* (5), 2344–2374.

Lorenzoni, Guido, "A Theory of Demand Shocks," *American Economic Review*, 2009, *99* (5), 2050–2084.

Lucas, Robert E. Jr. and Nancy L. Stokey, "Optimal fiscal and monetary policy in an economy without capital," *Journal of Monetary Economics*, January 1983, *12* (1), 55–93.

Mankiw, N. Gregory and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 2002, *117* (4), 1295–1328.

Mirrlees, James A., "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, 1976, *6* (4), 327–358.

Morris, Stephen and Hyun Song Shin, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 1998, *88* (3), 587–597.

_ **and** _ , "Social Value of Public Information," *American Economic Review*, 2002, *92* (5), 1521–1534.

Pellet, Thomas and Alireza Tahbaz-Salehi, "Rigid production networks," *Journal of Monetary Economics*, 2023, *137*, 86–102.

Ramsey, F. P., "A Contribution to the Theory of Taxation," *Economic Journal*, 1927, *37* (145), 47–61.

Scheuer, Florian and Iván Werning, "Mirrlees Meets Diamond-Mirrlees," Technical Report 5830, CESifo Working Paper Series April 2016. 35 Pages.

Shin, Yongseok, "Ramsey Meets Bewley: Optimal Government Financing with Incomplete Markets." Working Paper.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

__, "Rational Inattention and Monetary Economics," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Amsterdam: North-Holland, 2010, pp. 155–181.

Song, Wenting and Samuel Stern, "Firm Inattention and the Efficacy of Monetary Policy: A Text-Based Approach," *Review of Economic Studies*, 2024.

Stiglitz, Joseph E. and F. P. Ramsey, "In Praise of Frank Ramsey's Contribution to the Theory of Taxation," *The Economic Journal*, March 2015, *125* (583), 235–268. 125th Anniversary Issue.

Wang, Yicheng, Zhesheng Qiu, Le Xu, and Francesco Zanetti, "Monetary Policy in Open Economies with Production Networks," 2024. Available at SSRN 4955035.

Werning, Iván, "Optimal Fiscal Policy with Redistribution," *Quarterly Journal of Economics*, 2007, *122* (3), 925–967.

Online Appendix of "Information, Production Networks and Optimal Taxation"

Appendix A Theory

Proof of proposition 1:

1.The 'Only If' Part:

I solve the equilibrium conditions backwardly: at stage 2, the firm maximizes the after-tax profit (2) with complete information. The first-order conditions of the intermediate goods are thus given by

$$a_{ij}\frac{(1-\tau_i^{Ind}(s))p_i(s)z_i(s)(l_{ij}(\omega_{ij}))^{\alpha_i}\Pi_{k=1}^{N}x_{ij,k}^{a_{ik}}(\omega_{ij},s)}{x_{ij,k}(\omega_{ij},s)}=p_k(s) \tag{32}$$

where the left-hand is the marginal profit of intermediate goods, and the right-hand side is the marginal cost. The equation (32) holds for any island j, industry i and all intermediate goods k. Substituting (32) into the production function (1), I have

$$y_{ij}(\omega_{ij},s)=\underbrace{\left[z_i(s)\frac{\prod_{k=1}^{N}a_{ik}^{a_{ik}}}{\prod_{k=1}^{N}p_k^{a_{ik}}(s)}\Big((1-\tau_i^{Ind}(s))p_i(s)\Big)^{1-\alpha_i}\right]^{\frac{1}{\alpha_i}}}_{\Psi_i^y(s)}\ell_{ij}(\omega_{ij}) \tag{33}$$

and

$$x_{ij,k}(\omega_{ij},s)=\underbrace{a_{ij}(1-\tau_i^{Ind}(s))\frac{p_i(s)}{p_k(s)}\Psi_i^y(s)}_{\Psi_{ik}^l(s)}l_{ij}(\omega_{ij}) \tag{34}$$

By using (7), I have that $X_{ik}(s)=\Psi_{ik}^l(s)L_i(s)$ and thus I can transform the first order condition (32) by using the aggregate values:

$$a_{ik}\frac{(1-\tau_i^{Ind}(s))p_i(s)z_i(s)(L_i(s))^{\alpha_i}\Pi_{k=1}^{N}X_{ij}^{a_{ik}}(s)}{X_{ij}(s)}=p_j(s) \tag{35}$$

The first order conditions of industry's good j used in the final consumption goods sector is given by

$$\beta_j \frac{P(s)\Pi_{i=1}^{N}(\frac{c_i(s)}{\beta_i})^{\beta_i}}{c_j(s)} = p_j(s)(1+\tau_j^C(s)) \tag{36}$$

where the left-hand side is a marginal benefit, and the right-hand side is the marginal cost. The marginal cost of good j includes both its price and the associated consumption taxes.

At stage 1, for the firms' problem $\mathcal{P}_{\text{Firm}}$, combining with (33), the first order condition of labor demand $l_{ij}(\omega_{ij})$ is given by:

$$E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{(1-\tau_i^{Ind}(s'))p_i(s')}{P(s')} \Psi_i^y(s') = E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{w_{ik}(\omega_{ij})}{P(s')}$$

which can be transformed into the following expression by using (34):

$$\begin{aligned} E_{s'|\omega_{ij}} C(s')^{-\sigma} & \frac{\alpha_i(1-\tau_i^{Ind}(s'))p_i(s')z_i(s')(l_{ij}(\omega_{ij}))^{\alpha_i-1}\Pi_{k=1}^{N} x_{ij,k}^{a_{ij}}(\omega_{ij},s')}{P(s')} \\ &= E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{w_{ij}(\omega_{ij})}{P(s')} \end{aligned}$$

Similarly, I can rewrite the above equation by using the aggregate input:

$$\begin{aligned} E_{s'|\omega_{ij}} C(s')^{-\sigma} & \frac{\alpha_i(1-\tau_i^{Ind}(s'))p_i(s')z_i(s')(L_i(s'))^{\alpha_i-1}\Pi_{j=1}^{N} X_{ij}^{a_{ij}}(s')}{P(s')} \\ &= E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{w_{ij}(\omega_{ij})}{P(s')} \end{aligned} \tag{37}$$

as in the second stage, the marginal product of labor must be the same across all firms within the same industry, given the Cobb-Douglass technology. For the worker's optimization problem $\mathcal{P}_{\text{Worker}}$ at stage 1, the first order condition of labor supply n_{ij} on island j of industry i is given by

$$E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{w_{ij}(\omega_{ij})}{P(s')} = l_{ij}^{\varepsilon}(\omega_{ij})$$

where the left-hand side is the marginal benefit of labor, and the right-hand side is its marginal disutility. Combining the above two equations, I have

$$E_{s'|\omega_{ij}} C(s')^{-\sigma} \frac{\alpha_i(1-\tau_i^{Ind}(s'))p_i(s')z_i(s')(L_i(s'))^{\alpha_i-1}\Pi_{j=1}^{N} X_{ij}^{a_{ij}}(s')}{P(s')} = l_{ij}^{\varepsilon}(\omega_{ij}) \tag{38}$$

The equation (38) has a direct interpretation. The total revenue in industry j minus its total cost of intermediate goods is $\alpha_i p_i(s) z_i(s) (L_i(s'))^{\alpha_i} \Pi_{k=1}^{N} X_{ik}^{a_{ik}}(s')$[22]. The left-hand side is the marginal revenue of employment in terms of social welfare, and the right-hand side is its marginal cost in terms of labor disutility. In contrast to stage 2, where the first-order conditions are determined by current state variables s (since it is common knowledge), stage 1 incorporates expectations about future states s'. The future prices $p_i(s)$ are endogenously determined by labor input $\{L_j(s)\}$ for all industries. The choice of labor at stage 1, therefore, introduces a sophisticated layer of strategic interaction among industries. I set $P_i(s)$ and $\mathcal{T}_i(s)$ as follows:

$$P_i(s) \equiv C(s)^{-\sigma} \frac{p_i(s)}{P(s)} \tag{39}$$

$$\mathcal{T}_i(s) \equiv C(s)^{-\sigma} \frac{p_i(s)}{P(s)} (1-\tau_i^{Ind}(s)) z_i(s) \alpha_i L_i^{\alpha_i-1}(s) \Pi_{k=1}^{N} X_{ik}^{a_{ik}}(s) \tag{40}$$

where $P_i(s)$ denote the value of good j and $\mathcal{T}_i(s)$ denote the marginal benefit of labor, both in terms of the utility. Using (38), the labor $l_{ij}(\omega_{ij})$ is given by

$$l_{ij}(\omega_{ij}) = (E_{s'|\omega_{ij}} \mathcal{T}_i(s'))^{\frac{1}{\varepsilon}} \tag{41}$$

which gives (11) by aggregating all labor within one industry. For the representative household, the budget constraint can be rewritten as

[22] In the complete information model, it is also the share of labor cost. When information is incomplete, firms will have positive or negative profits, and the expenditure of labor $w_{ij}l_{ij}$ is not necessarily the fixed share a_i of its total revenue.

$$P(s)C(s) = \sum_{i=1}^{N} \int_{k\in[0,1]} [w_{ik}n_{ik} + \pi_{ik}] = \sum_{i=1}^{N} (1-\tau_i^{Ind}(s))p_i(s)z_i(s)\alpha_i L_i^{\alpha_i}(s)\Pi_{k=1}^{N} X_{ik}^{a_{ik}}(s) \tag{42}$$

where the second equality holds by using the F.O.C.s of firms at stage 2. Multiplying both sides by $\frac{C(s)^{-\sigma}}{P(s)}$ and by using the new notation of $P_i(s)$ and $\mathcal{T}_i(s)$, I have

$$C(s)^{1-\sigma} = \sum_{i=1}^{N} \mathcal{T}_i(s)L_i(s) \tag{43}$$

, which is the implementability constraint for the Ramsey problem.

2.The 'If' Part:

I construct the allocations, prices, and tax functions. The output y_i and c_i can be computed using the production function, the labor input L_i, and the intermediate input X_{ij}. The labor supply l_{ij} and demand n_{ij} on each island is determined by equation (41). Normalizing P to be 1. p_i can be derived by using equation (39). Then τ_i^{Ind} can be derived by using (40). The island output y_{ij} and intermediate input $x_{ij,k}$ can be derived by using equations (33)(34). The consumption good tax τ_i^C can be derived by using equation (36). The household budget constraint is satisfied for the way of my construction when the implementability condition holds. The government budget constraint is automatically satisfied in equilibrium when the household budget constraint and resource constraints are met. Finally, it's straightforward to check that all the f.o.cs of equilibrium hold.

The Ramsey problem:

I write out the Langrange for the Ramsey problem:

$$\mathcal{L} = \int [\frac{C(s)^{1-\sigma} - 1}{1-\sigma} - \frac{1}{\varepsilon+1} \sum_{i=1}^{N} \int [E_{s'|\omega_{ik}} \mathcal{T}_i(s')]^{\frac{\varepsilon+1}{\varepsilon}} \phi_i(\omega_{ik}|s) d\omega_{ik}$$

$$+\mu_R(s)[\Pi_{i=1}^{J}(\frac{z_i(s) L_i^{\alpha_i}(s) \Pi_{k=1}^{N} X_{ik}^{a_{ik}}(s) - \sum_{k=1}^{N} X_{ki}(s)}{\beta_i})^{\beta_i} - G(s) - C(s)]$$

$$+\mu_G(s)[C(s)^{1-\sigma} - \sum_{i=1}^{N} \mathcal{T}_i(s) L_i(s)]$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N} \mu_{ij}(s)[\frac{a_{ij}}{\alpha_i}\mathcal{T}_i(s) L_i(s) - P_j(s) X_{ij}(s)]$$

$$+\sum_{i}^{N} \mu_i^L(s)(\int \left[E_{s'|\omega_{ik}} \mathcal{T}_i(s')\right]^{\frac{1}{\varepsilon}} \phi_i(\omega_{ik}|s) d\omega_{ik} - L_i(s))]\mu(s) ds$$

F.O.Cs are:

$$\mathbf{L_i}(\mathbf{s}): \quad \mu_R(s)\frac{\partial Y(s)}{\partial L_i(s)} - \mu_i^L(s) - \mu_G(s)\mathcal{T}_i(s) + \sum_{j=1}^{N} \mu_{ij}(s)\frac{a_{ij}}{\alpha_i}\mathcal{T}_i(s) = 0 \tag{44}$$

$$\mathbf{P_i}(\mathbf{s}): \quad \sum_{i=1}^{N} \mu_{ij}(s) X_{ij}(s) = 0 \tag{45}$$

$$\begin{aligned}\mathcal{T}_{\mathbf{i}}(\mathbf{s}): \quad & -\frac{1}{\varepsilon}\int \tilde{\varphi}_i(\omega_{ik}|s)[E_{s'|\omega_{ik}}\mathcal{T}_i(s')]^{\frac{1}{\varepsilon}} d\omega_{ik} - \mu_G(s) L_i(s) + \sum_{j=1}^{N}\frac{\mu_{ij}(s) a_{ij}}{\alpha_i} L_i(s) \\ & + \frac{1}{\varepsilon}\int \mu_i^L(\tilde{s})\frac{\varphi(\tilde{s})}{\varphi(s)}[\int([E_{s'|\omega_{ik}}\mathcal{T}_i(s')]^{\frac{1}{\varepsilon}-1}\varphi_i(s|\omega_{ik})\varphi_i(\omega_{ik}|\tilde{s}) d\omega_{ik}] d\tilde{s} = 0\end{aligned} \tag{46}$$

$$\mathbf{X_{ij}}(\mathbf{s}): \quad \mu_R(s)\frac{\partial Y(s)}{\partial X_{ij}(s)} - \mu_{ij}(s) P_j(s) = 0, \tag{47}$$

$$\mathbf{C}(\mathbf{s}): \quad ((1-\sigma)\mu_G(s) + 1) C(s)^{-\sigma} = \mu_R(s) \tag{48}$$

where $\tilde{\varphi}_i(\omega_{ij}|s) = \int \frac{\varphi(\tilde{s})}{\varphi(s)} \varphi_i(s|\omega_{ik})\varphi_i(\omega_{ik}|\tilde{s}) d\tilde{s}$ is the probability of receiving signal ω_{ik} given that his belief of state is s.

Proof of proposition 2:

I use the guess and verify strategy:

Step 1:

Guess $\mu_{ij}(s) = 0$, $\mathcal{T}_i(s) = k_i \mathcal{T}(s)$, $\mu_i^L(s) = k_i \mu^L(s)$. The coefficient k_i and functional forms $\mathcal{T}(s)$ and $\mu^L(s)$ are undetermined at this moment, and they are going to be solved later.

Given the guess, from (11)

$$L_i(s) = k_i^{\frac{1}{\varepsilon}} E_{\omega_{ik}|s}[E_{s'|\omega_{ik}} \mathcal{T}(s')]^{\frac{1}{\varepsilon}} = k_i^{\frac{1}{\varepsilon}} L(s; \mathcal{T}) \tag{49}$$

where $L(s; \mathcal{T}) \equiv E_{\omega_{ik}|s}[E_{s'|\omega_{ik}} \mathcal{T}(s')]^{\frac{1}{\varepsilon}}$ [23]. This equation finds $L_i(s)$ once I determine $\mathcal{T}$. As $\mu_{ij}(s) = 0$, from (47) it implies $\frac{\partial Y(s)}{\partial X_{ij}(s)} = 0$. Using these $N \times N$ equations and the given set $\{L_i(s)\}_{i=1}^N$, I can first solve for $X_{ij}(s)$ (which includes $N \times N$ unknowns). Once these are determined, obtaining $Y_i(s)$, $C_i(s)$, and the aggregate output $Y(s)$ becomes straightforward by applying the relevant production functions. They satisfy

$$a_{ij} \frac{\beta_i Y_i(s)}{C_i(s)} / \frac{\beta_j Y_j(s)}{C_j(s)} = \frac{X_{ij}(s)}{Y_j(s)} \tag{50}$$

Aggregating both sides over i, I have

$$\sum_i a_{ij} \frac{\beta_i Y_i(s)}{C_i(s)} / \frac{\beta_j Y_j(s)}{C_j(s)} = 1 - \frac{C_j(s)}{Y_j(s)}$$

Thus

$$\beta_j = \frac{\beta_j Y_j(s)}{C_j(s)} - \sum_i a_{ij} \frac{\beta_i Y_i(s)}{C_i(s)}$$

[23]I can have a uniform functional form $L(s; \mathcal{T})$ for all industries as the information structure is homogeneous.

which in matrix form is

$$\boldsymbol{\beta} = (\boldsymbol{I} - \boldsymbol{A}^T) \begin{bmatrix} \frac{\beta_1 Y_1(s)}{C_1(s)} \\ \vdots \\ \frac{\beta_N Y_N(s)}{C_N(s)} \end{bmatrix} \tag{51}$$

It implies $[\frac{\beta_1 Y_1(s)}{C_1(s)}, \cdots, \frac{\beta_N Y_N(s)}{C_N(s)}]' = (I - A^T)^{-1}\beta$. Their values are constant regardless of the state s. Let $\frac{\beta_i Y_i(s)}{C_i(s)} = [(I - A^T)^{-1}\beta]_i \equiv \hat{D}_i$. Substituting it back into (50), I have

$$\frac{X_{ij}(s)}{Y_j(s)} = a_{ij}\frac{\hat{D}_i}{\hat{D}_j} \tag{52}$$

Multiplying both sides of (44) with $L_i(s)$ and substituting (48)(49), I have

$$\alpha_i \frac{\beta_i Y_i(s)}{C_i(s)}((1-\sigma)\mu_G(s)+1)Y(s)C(s)^{-\sigma} = k_i^{\frac{\varepsilon+1}{\varepsilon}}(\mu^L(s)L(s;\mathcal{T}) + \mu_G(s)\mathcal{T}(s)L(s;\mathcal{T}))$$

To ensure the above equation holds, I set $k_i = (\hat{D}_i\alpha_i)^{\frac{\varepsilon}{\varepsilon+1}}$, as it follows from $\frac{\beta_i Y_i(s)}{C_i(s)} = \hat{D}_i$. Dividing both sides by $\hat{D}_i\alpha_i$, I get a single equation

$$((1-\sigma)\mu_G(s)+1)Y(s)C(s)^{-\sigma} = \mu^L(s)L(s;\mathcal{T}) + \mu_G(s)\mathcal{T}(s)L(s;\mathcal{T}) \tag{53}$$

In order to satisfy the equilibrium condition (35), $P_j(s)$ is defined as

$$P_j(s) = \frac{\frac{a_{ij}}{\alpha_i}\mathcal{T}_i(s)L_i(s)}{X_{ij}(s)} = \frac{\hat{D}_j\mathcal{T}(s)L(s;\mathcal{T})}{Y_j(s)} \tag{54}$$

For the implementability constraint to hold,

$$C(s)^{1-\sigma} = (\sum_{i=1}^{N} k_i^{\frac{\varepsilon+1}{\varepsilon}})\mathcal{T}(s)L(s;\mathcal{T}) = \mathcal{T}(s)L(s;\mathcal{T}) \tag{55}$$

Combining with the resource constraint, I have

$$(Y(s) - G(s))^{1-\sigma} = \mathcal{T}(s)L(s;\mathcal{T}) \tag{56}$$

From the previous discussion, I know $Y(s) - G(s)$ is a function of $\mathcal{T}$(s). Thus, I determine $\mathcal{T}(s)$ as the function that solves the above equation.

From (46), given the guess, I need

$$k_i^{\frac{1}{\varepsilon}}(\varepsilon)\mu_G(s)L(s;\mathcal{T}) + k_i^{\frac{1}{\varepsilon}}L(s;\mathcal{T})$$
$$= k_i^{\frac{1}{\varepsilon}} \int \mu^L(\tilde{s})\frac{\varphi(\tilde{s})}{\varphi(s)}[\int([E_{s'|\omega_{ik}}\boldsymbol{\mathcal{T}}(s')]^{\frac{1}{\varepsilon}-1}\varphi_i(s|\omega_{ik})\varphi_i(\omega_{ik}|\tilde{s})d\omega_{ik}]d\tilde{s} \tag{57}$$

Dividing both sides by $k_i^{\frac{1}{\varepsilon}}$, I have a single equation. Combining with (53), I have

$$((\varepsilon+1)\frac{\mu^L(s)L(s;\mathcal{T}) - Y(s)C(s)^{-\sigma}}{(1-\sigma)Y(s)C(s)^{-\sigma} - \mathcal{T}(s)L(s;\mathcal{T})} + 1)L(s;\mathcal{T})$$
$$= \int \mu^L(\tilde{s})\frac{\varphi(\tilde{s})}{\varphi(s)}[\int([E_{s'|\omega_{ik}}\boldsymbol{\mathcal{T}}(s')]^{\frac{1}{\varepsilon}-1}\varphi_i(s|\omega_{ik})\varphi_i(\omega_{ik}|\tilde{s})d\omega_{ik}]d\tilde{s} \tag{58}$$

Once I find $\mathcal{T}(s)$ using (56), I can also compute $C(s), Y(s), L(s;\mathcal{T})$ following the way I discussed. Then I solve $\mu^L(s)$ by using (58) and I define $\mu_G(s)$ ensuring (53) holds and $\mu_R(s)$ ensuring (48) holds.

Now I verify the guess is correct. Given the guess, conditions (8)(9)(11)(45)(47)(48) hold automatically. Given the values of k_i, (44) and (46) are reduced to conditions (53) and (58) and those conditions are satisfied by the way I define $\mu^L(s)$ and $\mu_G(s)$. Besides, (35) is satisfied by the way I construct $P_j(s)$.

Step 2:

The industry revenue tax satisfies: $\mathcal{T}_j(s)L_j(s) = \frac{(1-\tau_j^{Ind})\alpha_j p_j(s)Y_j(s)}{P(s)}C(s)^{-\sigma}$. Then

$$(1-\tau_j^{Ind}(s)) = \frac{\mathcal{T}_j(s)L_j(s)}{\alpha_j P_j(s)Y_j(s)} = 1 \tag{59}$$

which implies that industry revenue tax is zero for all industries.

The taxation on the consumption goods satisfies: $p_j(s)C_j(s)(1+\tau_j^C(s)) = \beta_j P(s)Y(s)$. Transforming it, I have

$$1+\tau_j^C(s) = \frac{\beta_j Y(s)C^{-\sigma}(s)}{P_j(s)C_j(s)} = \frac{1}{P_j(s)Y_j(s)}\frac{\beta_j Y_j(s)}{C_j(s)}C^{1-\sigma}(s)\frac{Y(s)}{C(s)} = \frac{Y(s)}{C(s)}$$

which implies that the taxation on the consumption goods are equal across different industries.

Proof of Theorem 1:

I use the perturbation approach to solve equations (44)-(48):

Zeroth Order Perturbation:

For the budget constraint:

$$\bar{Y} - \bar{G} = \bar{C}, \quad \bar{C}^{1-\sigma} = \sum_{i=1}^{N} \bar{L}_i^{\varepsilon+1}, \quad \frac{a_{ij}}{a_i}\bar{L}_i^{\varepsilon+1} = \bar{P}_j \bar{X}_{ij}, \quad \bar{T}_i = \bar{L}_i^{\varepsilon} \tag{60}$$

For the F.O.Cs:

$$\sum_{i=1}^{N} \bar{\mu}_{ij}\bar{X}_{ij} = 0, \quad \bar{\mu}_R \frac{\partial \bar{Y}}{\partial X_{ij}} = \bar{\mu}_{ij}\bar{P}_j, \quad ((1-\sigma)\bar{\mu}_G + 1)\bar{C}^{-\sigma} = \bar{\mu}_R$$

$$\bar{\mu}_i^L \bar{L}_i^{1-\varepsilon+1} = \bar{\mu}_R \frac{\partial \bar{Y}}{\partial L_i}\bar{L}_i^{1-\varepsilon+1} - \bar{\mu}_G + \sum_{j=1}^{N} \bar{\mu}_{ij}\frac{a_{ij}}{\alpha_i} = (\varepsilon)(\bar{\mu}_G + \sum_{j=1}^{N} \bar{\mu}_{ij}\frac{a_{ij}}{\alpha_i}) + 1$$

From proposition 2, I know that the costates $\bar{\mu}_{ij} = 0$ as I can let $\varphi(\omega|s)$ to be Dirac distribution $\delta(\omega - s)$ when agents in all industries have perfect information. Then I have

$$\frac{\partial \bar{Y}}{\partial L_i}\bar{L}_i^{-\varepsilon} = \frac{\varepsilon + 1\bar{\mu}_G + 1}{\bar{\mu}_R} = \frac{\varepsilon + 1\bar{\mu}_G + 1}{((1-\sigma)\bar{\mu}_G + 1)}\bar{C}^{\sigma} \tag{61}$$

where $\frac{\varepsilon+1\bar{\mu}_G+1}{((1-\sigma)\bar{\mu}_G+1)}$ gives the wedge of labor supply for the steady state economy. In steady state, the industry revenue taxes are all zeros $\bar{\tau}_j^{Ind} = 0$.

First Order Perturbation:

To get the first-order perturbations, I derive the following useful lemma:

Lemma 1. *Let $\varphi_i(s|\tilde{s}) \equiv \int \varphi_i(s|\omega_{ik})\varphi_i(\omega_{ik}|\tilde{s})d\omega_{ik}$ represent the probability that the state is perceived to be s, given that the actual state is $\tilde{s}$. Then I have:*

$$\int \tilde{s}\frac{\varphi_i(s|\tilde{s})\varphi(\tilde{s})}{\varphi(s)}d\tilde{s} = \lambda_i s, \quad \int \omega_{ij}\tilde{\varphi}_i(\omega_{ij}|s)d\omega_{ij} = s$$

where $\lambda_i = \frac{\sigma_{is}^2}{\sigma_{is}^2+\sigma_{ie}^2}$ *denote the precision of industry i about the true state.*

Proof. Given Gaussian shocks and signals,

$$\frac{\varphi_i(s|\tilde{s})\varphi(\tilde{s})}{\varphi(s)} = \int e^{\frac{s^2-\tilde{s}^2}{2\sigma_{is}^2}} \frac{1}{\sqrt{2\pi}\sigma_{ie}} e^{-\frac{(\omega_{ik}-\tilde{s})^2}{2\sigma_{ie}^2}} \frac{1}{\sqrt{2\pi\frac{\sigma_{is}^2\sigma_{ie}^2}{\sigma_{is}^2+\sigma_{ie}^2}}} e^{-\frac{(s-\lambda_i\omega_{ik})^2}{2(\sigma_{is}^{-2}+\sigma_{ie}^{-2})^{-1}}} d\omega_{ik}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{ie}} \frac{1}{\sqrt{2\pi\frac{\sigma_{is}^2\sigma_{ie}^2}{\sigma_{is}^2+\sigma_{ie}^2}}} \int e^{-\frac{1}{2}(\tilde{s}^2(\sigma_{is}^{-2}+\sigma_{ie}^{-2})+\sigma_{ie}^{-2}s^2-2\sigma_{ie}^{-2}\omega_{ik}(s+\tilde{s})+(\frac{\sigma_{is}^2}{\sigma_{is}^2+\sigma_{ie}^2}+1)\sigma_{ie}^{-2}\omega_{ik}^2)} d\omega_{ik}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{ie}} \frac{1}{\sqrt{2\pi\frac{\sigma_{is}^2\sigma_{ie}^2}{\sigma_{is}^2+\sigma_{ie}^2}}} e^{-\frac{1}{2}(\tilde{s}^2(\sigma_{is}^{-2}+\sigma_{ie}^{-2})+\sigma_{ie}^{-2}s^2-\frac{(s+\tilde{s})^2}{(\frac{\sigma_{is}^2}{\sigma_{is}^2+\sigma_{ie}^2}+1)\sigma_{ie}^2})}$$

$$\times \int e^{-\frac{1}{2}(\frac{\sigma_{is}^2}{\sigma_{is}^2+\sigma_{ie}^2}+1)\sigma_{ie}^{-2}(\omega_{ik}-(\frac{\sigma_{is}^2}{\sigma_{is}^2+\sigma_{ie}^2}+1)^{-1}(s+\tilde{s}))^2} d\omega_{ik}$$

$$= \frac{1}{\sqrt{2\pi\frac{\sigma_{is}^2\sigma_{ie}^2}{2\sigma_{is}^2+\sigma_{ie}^2}}} e^{-\frac{(\tilde{s}-\lambda_i s)^2}{2\frac{\sigma_{is}^2\sigma_{ie}^2}{2\sigma_{is}^2+\sigma_{ie}^2}}}$$

As $\frac{\varphi_i(s|\tilde{s})\varphi(\tilde{s})}{\varphi(s)}$ is the density of $\mathcal{N}(\lambda_i s, \frac{\sigma_{is}^2\sigma_{ie}^2}{2\sigma_{is}^2+\sigma_{ie}^2})$, I prove the first part. For the second one, I have

$$\tilde{\varphi}_i(\omega_{ij}|s) = \frac{\varphi_i(s|\omega_{ik})}{\varphi(s)} \int \varphi(\tilde{s})\varphi_i(\omega_{ik}|\tilde{s})d\tilde{s} = \frac{\varphi_i(s|\omega_{ik})\varphi(\omega_{ik})}{\varphi(s)} = \varphi_i(\omega_{ij}|s)$$

by using the fact that $\varphi_i(s|\omega_{ik})\varphi(\omega_{ik}) = \varphi(s)\varphi_i(\omega_{ij}|s)$. $\tilde{\varphi}_i(\omega_{ij}|s)$ and $\varphi(\omega_{ik})$ have different meanings, but they are intrinsically the same as we assume people have rational expectations. Thus, I prove the second part. □

Budget Constraint:

To simplify the expression, I mean $\frac{\partial X(s)}{\partial s}|_{s=0}$ by using $\frac{\partial X(s)}{\partial s}$. Specifically, $\frac{\partial X(s)}{\partial s}|_{s=0}$ denotes evaluating the partial derivative of $X(s)$ with respect to s at $s=0$. For The first order perturbation of the implementability constraint, I have

$$
\begin{aligned}
(1-\sigma)\frac{\partial \log(C(s))}{\partial s} &= \sum_{i=1}^{N}\frac{\bar{T}_i\bar{L}_i}{\sum_{i=1}^{N}\bar{T}_i\bar{L}_i}(\frac{\partial \log(L_i(s))}{\partial s}+\frac{\partial \log(\mathcal{T}_i(s))}{\partial s}) \\
&= (1+\bar{\tau}^C)\sum_{i=1}^{N}D_i a_i(\frac{\partial \log(L_i(s))}{\partial s}+\frac{\partial \log(\mathcal{T}_i(s))}{\partial s}) \qquad (62)
\end{aligned}
$$

where D_i is the Domar Weight of industry i. I show the last equality step by step. Using (60), I have $\bar{T}_i\bar{L}_i = \bar{L}_i^{\varepsilon+1}$. Combining (61) and the following equation

$$
\frac{\partial \bar{Y}}{\partial L_i}\bar{L}_i^{-\varepsilon} = \alpha_i\beta_i\frac{\bar{Y}}{\bar{C}_i}\frac{\bar{Y}_i}{\bar{L}_i}\bar{L}_i^{-\varepsilon} = \alpha_i\bar{Y}\frac{\beta_i\bar{Y}}{P_i\bar{C}_i}\frac{P_i\bar{Y}_i}{\bar{Y}}\bar{L}_i^{-\varepsilon-1} = \alpha_i\bar{Y}(1+\bar{\tau}^C)D_i\bar{L}_i^{-\varepsilon-1}
$$

We know that $\bar{L}_i^{\varepsilon+1} \sim \alpha_i D_i$. Thus, I show the last equality holds. For the Domar Weight, let $\mathbf{D} \equiv \begin{pmatrix} D_1 \\ \vdots \\ D_N \end{pmatrix}$ by using f.o.cs of steady state, we know it satisfies:

$$
\mathbf{D} = \frac{1}{1+\bar{\tau}^C}(I-A^T)^{-1}\beta \qquad (63)
$$

For the resource constraint, I have

$$
\frac{\partial \log(C(s))}{\partial s} = \frac{\bar{Y}}{\bar{C}}(\sum_{i=1}^{N}\beta_i\frac{\partial \log(C_i(s))}{\partial s}+\frac{\partial \log(A(s))}{\partial s}) - \frac{\bar{G}}{\bar{C}}\frac{\partial \log(G(s))}{\partial s} \qquad (64)
$$

For the equilibrium condition constraints, by using

$$
\begin{aligned}
E_{\omega_{ik}|s}[E_{s'|\omega_{ik}}\mathcal{T}_i(s')]^{\frac{1}{\varepsilon}} &\approx \frac{1}{\varepsilon}E_{\omega_{ik}|s}[\bar{\mathcal{T}}_i^{\frac{1}{\varepsilon}-1}E_{s'|\omega_{ik}}(ds'\frac{\partial \mathcal{T}_i(s)}{\partial s}|_{s=0})] \\
= \frac{1}{\varepsilon}\frac{\partial \mathcal{T}_i(s)}{\partial s}|_{s=0}\bar{\mathcal{T}}_i^{\frac{1}{\varepsilon}-1}E_{\omega_{ik}|s}[E_{s'|\omega_{ik}}ds'] &= \frac{1}{\varepsilon}\lambda_i ds\frac{\partial \mathcal{T}_i(s)}{\partial s}|_{s=0}\bar{\mathcal{T}}_i^{\frac{1}{\varepsilon}-1}
\end{aligned}
$$

I have

$$
\frac{\partial \log(\mathcal{T}_i(s))}{\partial s} = \lambda_i^{-1}(\varepsilon)\frac{\partial \log(L_i(s))}{\partial s} \qquad (65)
$$

$$
\frac{\partial \log(\mathcal{T}_i(s))}{\partial s}+\frac{\partial \log(\mathcal{L}_i(s))}{\partial s}-\frac{\partial \log(\mathcal{P}_j(s))}{\partial s} = \frac{\partial \log(\mathcal{X}_{ij}(s))}{\partial s} \qquad (66)
$$

F.O.Cs:

For $L_i(s)$:

$$\frac{\partial\mu_R(s)}{\partial s}\frac{\partial\bar{Y}}{\partial L_i}+\frac{\partial\frac{\partial Y(s)}{\partial L_i(s)}}{\partial s}\bar{\mu}_R-\frac{\partial\mu_i^L(s)}{\partial s}-\frac{\partial\mu_G(s)}{\partial s}\bar{\mathcal{T}}_i+\frac{\partial\mathcal{T}_i(s)}{\partial s}\bar{\mu}_G+\sum_{j=1}^{N}\frac{a_{ij}}{\alpha_i}\bar{\mathcal{T}}_i\frac{\partial\mu_{ij}(s)}{\partial s}=0$$

Dividing both sides by $\mathcal{T}_i$ and using (60)(61) and (65), I have

$$\begin{aligned}\frac{\partial\mu_i^L(s)}{\partial s}\bar{L}_i^{-\varepsilon}=\frac{\partial\mu_R(s)}{\partial s}\frac{\varepsilon+1\bar{\mu}_G+1}{((1-\sigma)\bar{\mu}_G+1)}\bar{C}^{\sigma}+\frac{\partial log(\frac{\partial Y(s)}{\partial L_i(s)})}{\partial s}(\varepsilon+1\bar{\mu}_G+1)-\frac{\partial\mu_G(s)}{\partial s}\\+\lambda_i^{-1}(\varepsilon)\frac{\partial\log(L_i(s))}{\partial s}\bar{\mu}_G+\sum_{j=1}^{N}\frac{a_{ij}}{\alpha_i}\frac{\partial\mu_{ij}(s)}{\partial s}\end{aligned}\tag{67}$$

For $\mathcal{T}_i(s)$, using (65) and lemma 1, I have

$$\frac{\partial\mu_G(s)}{\partial s}(\varepsilon)=(\varepsilon)\sum_{j=1}^{N}\frac{a_{ij}}{\alpha_i}\frac{\partial\mu_{ij}(s)}{\partial s}-(\varepsilon)((\varepsilon)\bar{\mu}_G+1)\frac{\partial log(L_i(s))}{\partial s}+\lambda_i\frac{\partial\mu_i^L(s)}{\partial s}\bar{L}_i^{-\varepsilon}\tag{68}$$

Substituting (67) into (68), I have

$$\begin{aligned}\frac{\partial\mu_G(s)}{\partial s}[\lambda_i^{-1}\varepsilon+1]=[\lambda_i^{-1}\varepsilon+1]\sum_{j=1}^{N}\frac{a_{ij}}{\alpha_i}\frac{\partial\mu_{ij}(s)}{\partial s}-\lambda_i^{-1}\varepsilon((\varepsilon+1)\bar{\mu}_G+1)\frac{\partial log(L_i(s))}{\partial s}\\+\frac{\partial\mu_R(s)}{\partial s}\frac{(\varepsilon+1)\bar{\mu}_G+1}{((1-\sigma)\bar{\mu}_G+1)}\bar{C}^{\sigma}+\frac{\partial log(\frac{\partial Y(s)}{\partial L_i(s)})}{\partial s}((\varepsilon+1)\bar{\mu}_G+1)\end{aligned}$$

Substuting $\frac{\partial Y(s)}{\partial L_i(s)}=\alpha_i\beta_i\frac{Y(s)}{C_i(s)}\frac{Y_i(s)}{L_i(s)}$, it can be simplified as

$$\begin{aligned}\frac{\partial\mu_G(s)}{\partial s}[\lambda_i^{-1}\varepsilon+1]-\frac{\partial\mu_R(s)}{\partial s}\frac{(\varepsilon+1)\bar{\mu}_G+1}{((1-\sigma)\bar{\mu}_G+1)}\bar{C}^{\sigma}=[\lambda_i^{-1}\varepsilon+1]\sum_{j=1}^{N}\frac{a_{ij}}{\alpha_i}\frac{\partial\mu_{ij}(s)}{\partial s}\\+((\varepsilon+1)\bar{\mu}_G+1)(\frac{\partial\log(Y(s))}{\partial s}+\frac{\partial\log(Y_i(s))}{\partial s}-\frac{\partial\log(C_i(s))}{\partial s}-[\lambda_i^{-1}\varepsilon+1]\frac{\partial\log(L_i(s))}{\partial s})\end{aligned}\tag{69}$$

For the f.o.c of X_{ij}, I have

$$\frac{\partial \mu_{ij}(s)}{\partial s} = \frac{\partial \frac{\partial Y(s)}{\partial X_{ij}(s)}}{\partial s} \frac{\bar{\mu}_R}{\bar{P}_j}$$

Since

$$\frac{\partial Y(s)}{\partial X_{ij}(s)} = \beta_i a_{ij} \frac{Y(s)}{C_i(s)} \frac{Y_i(s)}{X_{ij}(s)} - \beta_j \frac{Y(s)}{C_j(s)}$$

and the steady state value $\frac{\partial \bar{Y}}{\partial X_{ij}}$ is zero , I can further get

$$\frac{\partial \mu_{ij}(s)}{\partial s} = \frac{\beta_j \bar{Y}}{\bar{C}_j} \frac{\bar{\mu}_R}{\bar{P}_j} (\frac{\partial \log(Y_i(s))}{\partial s} - \frac{\partial \log(C_i(s))}{\partial s} + \frac{\partial \log(C_j(s))}{\partial s} - \frac{\partial \log(X_{ij}(s))}{\partial s})$$

Combining with (61) (65) and (66), I finally have

$$\begin{aligned}\frac{\partial \mu_{ij}(s)}{\partial s} = ((\varepsilon+1)\bar{\mu}_G + 1)(&\frac{\partial \log(Y_i(s))}{\partial s} - \frac{\partial \log(C_i(s))}{\partial s} + \frac{\partial \log(C_j(s))}{\partial s} \\ &-[\lambda_i^{-1}\varepsilon + 1]\frac{\partial \log(L_i(s))}{\partial s} + \frac{\partial \log(P_j(s))}{\partial s})\end{aligned} \tag{70}$$

Then

$$\begin{aligned}\sum_{j=1}^{N} \frac{a_{ij}}{\alpha_i} \frac{\partial \mu_{ij}(s)}{\partial s} = (\varepsilon\bar{\mu}_G + 1)(&\frac{1-\alpha_i}{\alpha_i}\frac{\partial \log(Y_i(s))}{\partial s} - \frac{1-\alpha_i}{\alpha_i}\frac{\partial \log(C_i(s))}{\partial s} + \sum_{j=1}^{N} \frac{a_{ij}}{\alpha_i}\frac{\partial \log(C_j(s))}{\partial s} \\ &-[\lambda_i^{-1}(\varepsilon - 1) + 1]\frac{1-\alpha_i}{\alpha_i}\frac{\partial \log(L_i(s))}{\partial s} + \sum_{j=1}^{N} \frac{a_{ij}}{\alpha_i}\frac{\partial \log(P_j(s))}{\partial s})\end{aligned} \tag{71}$$

And

$$\begin{aligned}\sum_{i=1}^{N} \bar{X}_{ij}\frac{\partial \mu_{ij}(s)}{\partial s} = ((\varepsilon+1)\bar{\mu}_G + 1)(&\sum_{i=1}^{N} \bar{X}_{ij}\frac{\partial \log(Y_i(s))}{\partial s} - \sum_{i=1}^{N} \bar{X}_{ij}\frac{\partial \log(C_i(s))}{\partial s} + X_j^{Used}\frac{\partial \log(C_j(s))}{\partial s} \\ &-\sum_{i=1}^{N} \bar{X}_{ij}[\lambda_i^{-1}\varepsilon + 1]\frac{\partial \log(L_i(s))}{\partial s} + X_j^{Used}\frac{\partial \log(P_j(s))}{\partial s})\end{aligned} \tag{72}$$

where $X_j^{Used} \equiv \sum_{i=1}^{N} \bar{X}_{ij}$ which represents the total use of industry goods j as intermediate goods. The perturbation of the f.o.c of P_j implies

$$\sum_{i=1}^{N} \bar{X}_{ij} \frac{\partial \mu_{ij}(s)}{\partial s} = 0 \tag{73}$$

And the perturbation of the f.o.c of $C_i(s)$ gives

$$\bar{C}^{\sigma} \frac{\partial \mu_R(s)}{\partial s} = (1-\sigma)\frac{\partial \mu_G(s)}{\partial s} - \sigma((1-\sigma)\bar{\mu}_G + 1)\frac{\partial logC(s)}{\partial s} \tag{74}$$

I use the matrix notations to solve the above equations:

$$\partial \mathbf{L} \equiv \begin{bmatrix} \frac{\partial \log(L_1(s))}{\partial s} \\ \vdots \\ \frac{\partial \log(L_N(s))}{\partial s} \end{bmatrix}, \quad \partial \mathbf{C} \equiv \begin{bmatrix} \frac{\partial \log(C_1(s))}{\partial s} \\ \vdots \\ \frac{\partial \log(C_N(s))}{\partial s} \end{bmatrix}, \quad \partial \mathbf{Y} \equiv \begin{bmatrix} \frac{\partial \log(Y_1(s))}{\partial s} \\ \vdots \\ \frac{\partial \log(Y_N(s))}{\partial s} \end{bmatrix} \quad \partial \mathbf{P} \equiv \begin{bmatrix} \frac{\partial \log(P_1(s))}{\partial s} \\ \vdots \\ \frac{\partial \log(P_N(s))}{\partial s} \end{bmatrix}$$

$$\boldsymbol{\alpha} \equiv \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_N \end{bmatrix}, \quad \bar{\boldsymbol{X}} \equiv [\bar{X}_{ij}], \quad \boldsymbol{X^{Used}} \equiv \begin{bmatrix} X_1^{Used} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & X_N^{Used} \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_N \end{bmatrix}$$

Combining (72) and (73), I have

$$(\boldsymbol{X^{Used}} - \boldsymbol{X^T})\partial\mathbf{C} + \boldsymbol{X^T}\partial\mathbf{Y} - \boldsymbol{X^T}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} + \boldsymbol{X^{Used}}\partial\mathbf{P} = \mathbf{0} \tag{75}$$

Lemma 2.

$$\partial\mathbf{C} = \boldsymbol{C}^{-1}[((\boldsymbol{Y}(\boldsymbol{I}-\boldsymbol{\alpha}) - \boldsymbol{X^T})(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}) + \boldsymbol{\alpha}\boldsymbol{Y})\partial\mathbf{L} + (\boldsymbol{X^{Used}} - \boldsymbol{YA})\boldsymbol{\partial P} + \boldsymbol{Y}\partial\mathbf{Z}] \tag{76}$$

$$\partial\mathbf{Y} = [(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I}-\boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial\mathbf{L} - \boldsymbol{A}\partial\mathbf{P} + \partial\mathbf{Z} \tag{77}$$

Proof. For the Cobb-Douglas production function, $Y_i = z_i(s)L_i^{\alpha_i}(s)\Pi_{k=1}^{N} X_{ik}^{a_{ik}}(s)$. From (65) and (66), we know that

$$\sum_{k=1}^{N} \frac{\partial log(X_{ik}^{a_{ik}}(s))}{\partial s} = \sum_{k=1}^{N} a_{ik}\frac{\partial log(X_{ik}(s))}{\partial s} = \sum_{k=1}^{N} a_{ik}((\lambda_i^{-1}\varepsilon + 1)\frac{\partial \log(\mathcal{L}_i(s))}{\partial s} - \frac{\partial \log(\mathcal{P}_k(s))}{\partial s})$$

which implies

$$\begin{bmatrix} \sum_{k=1}^{N} \frac{\partial log(X_{1k}^{a_{1k}}(s))}{\partial s} \\ \vdots \\ \sum_{k=1}^{N} \frac{\partial log(X_{Nk}^{a_{Nk}}(s))}{\partial s} \end{bmatrix} = (\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha})\partial \mathbf{L} - \boldsymbol{A}\partial \mathbf{P}$$

Thus I show the expression for $\partial \mathbf{Y}$. As $C_i(s) = Y_i(s) - \sum_{k=1}^{N} X_{ki}(s)$, I have

$$\frac{\partial \log(C_i(s))}{\partial s} = \frac{Y_i(s)}{C_i(s)} \frac{\partial \log(Y_i(s))}{\partial s} - \frac{1}{C_i(s)} \sum_{k=1}^{N} X_{ki}(s) \frac{\partial \log(X_{ki}(s))}{\partial s}$$

Again by using (65) and (66),

$$\sum_{k=1}^{N} X_{ki}(s) \frac{\partial \log(X_{ki}(s))}{\partial s} = \sum_{k=1}^{N} X_{ki}(s)((\lambda_k^{-1}\varepsilon + 1)\frac{\partial \log(\mathcal{L}_k(s))}{\partial s} - \frac{\partial \log(\mathcal{P}_i(s))}{\partial s})$$

which implies

$$\begin{bmatrix} \sum_{k=1}^{N} X_{ki}(s) \frac{\partial \log(X_{k1}(s))}{\partial s} \\ \vdots \\ \sum_{k=1}^{N} X_{ki}(s) \frac{\partial \log(X_{kN}(s))}{\partial s} \end{bmatrix} = \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})\partial \mathbf{L} - \boldsymbol{X}^{\boldsymbol{Used}}\boldsymbol{\partial P}$$

Then it's straightforward to have

$$\partial \mathbf{C} = \boldsymbol{C}^{-1}[\boldsymbol{Y}\partial \mathbf{Y} - \boldsymbol{X}^{T}(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})\partial \mathbf{L} + \boldsymbol{X}^{\boldsymbol{Used}}\boldsymbol{\partial P}] \tag{78}$$

Substituting the expression of $\partial \mathbf{Y}$, I prove the expression for $\partial \mathbf{C}$.

□

Next, I prove the following key lemma 3, which helps us to largely simplify the equations I need to solve:

Lemma 3. *The solution $\partial \mathbf{C}$ and $\partial \mathbf{Y}$ satisfy:*

$$\partial \mathbf{Y} - \partial \mathbf{C} = \mathbf{0} \tag{79}$$

Proof. By using lemma 2:

$$
\begin{aligned}
&\boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{Y} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon\boldsymbol{I}]\partial\mathbf{L} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{P} \\
&= \boldsymbol{X}^{\boldsymbol{Used}}([(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial\mathbf{L} - A\partial\mathbf{P} + \partial\mathbf{Z}) - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{P} \\
&= \boldsymbol{X}^{\boldsymbol{Used}}[(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial\mathbf{L} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} \\
&\quad + \boldsymbol{X}^{\boldsymbol{Used}}(\boldsymbol{I} - A)\partial\mathbf{P} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{Z}
\end{aligned}
$$

And

$$
\begin{aligned}
\partial\mathbf{C} - \partial\mathbf{Y} &= \boldsymbol{C}^{-1}[((\boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{\alpha}) - \boldsymbol{X}^{\boldsymbol{T}})(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}) + \boldsymbol{\alpha}\boldsymbol{Y})\partial\mathbf{L} + (\boldsymbol{X}^{\boldsymbol{Used}} - \boldsymbol{Y}A)\partial\boldsymbol{P} + \boldsymbol{Y}\partial\mathbf{Z}] \\
&\quad - [(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial\mathbf{L} - A\partial\mathbf{P} + \partial\mathbf{Z} \\
&= \boldsymbol{C}^{-1}[((\boldsymbol{Y} - \boldsymbol{C})((\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha})\partial\mathbf{L} - \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})\partial\mathbf{L} \\
&\quad + (\boldsymbol{X}^{\boldsymbol{Used}} - (\boldsymbol{Y} - \boldsymbol{C})A)\partial\mathbf{P} + (\boldsymbol{Y} - \boldsymbol{C})\partial\mathbf{Z}]
\end{aligned}
$$

Market clearing conditions of each industry imply $\boldsymbol{Y} - \boldsymbol{C} = \boldsymbol{X}^{\boldsymbol{Used}}$. Thus

$$
\begin{aligned}
\partial\mathbf{C} - \partial\mathbf{Y} = \boldsymbol{C}^{-1}[&\boldsymbol{X}^{\boldsymbol{Used}}[(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial\mathbf{L} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} \\
&+ \boldsymbol{X}^{\boldsymbol{Used}}(\boldsymbol{I} - A)\partial\mathbf{P} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{Z}]
\end{aligned}
$$

Comparing two expressions, $\boldsymbol{C}(\partial\mathbf{C} - \partial\mathbf{Y}) = \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{Y} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{P}$. The equation (75) can be rewritten as

$$
\begin{aligned}
(\boldsymbol{X}^{\boldsymbol{Used}} - \boldsymbol{X}^{\boldsymbol{T}})(\partial\mathbf{C} - \partial\mathbf{Y}) + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{Y} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial\mathbf{L} + \boldsymbol{X}^{\boldsymbol{Used}}\partial\mathbf{P} = \mathbf{0} \\
\Rightarrow (\boldsymbol{Y} - \boldsymbol{X}^{\boldsymbol{T}})(\partial\mathbf{C} - \partial\mathbf{Y}) = \mathbf{0}
\end{aligned}
$$

which proves the lemma. □

By using lemma 2 and lemma 3 , I have

$$\begin{bmatrix} \sum_{j=1}^{N} \frac{a_{1j}}{\alpha_1} \frac{\partial \mu_{1j}(s)}{\partial s} \\ \vdots \\ \sum_{j=1}^{N} \frac{a_{Nj}}{\alpha_N} \frac{\partial \mu_{Nj}(s)}{\partial s} \end{bmatrix} = ((\varepsilon+1)\bar{\mu}_G+1)\boldsymbol{\alpha}^{-\mathbf{1}}[(\boldsymbol{I}-\boldsymbol{\alpha})(\partial\mathbf{Y}-(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\mathbf{L})$$

$$+(\boldsymbol{A}-(\boldsymbol{I}-\boldsymbol{\alpha}))\partial\mathbf{C}+\boldsymbol{A}\partial\mathbf{P})]$$

$$=((\varepsilon+1)\bar{\mu}_G+1)\boldsymbol{\alpha}^{-\mathbf{1}}[-(\boldsymbol{I}-\boldsymbol{\alpha})(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\mathbf{L}+\boldsymbol{A}\partial\mathbf{P}+\boldsymbol{A}\partial\mathbf{C}]$$

$$=((\varepsilon+1)\bar{\mu}_G+1)\boldsymbol{\alpha}^{-\mathbf{1}}(-\partial\mathbf{Y}+\partial\mathbf{Z}+\boldsymbol{\alpha}\partial\mathbf{L}+\boldsymbol{A}\partial\mathbf{C})$$

$$=((\varepsilon+1)\bar{\mu}_G+1)[\partial\mathbf{L}-\boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{A})\partial\mathbf{Y}+\boldsymbol{\alpha}^{-\mathbf{1}}\partial\mathbf{Z}]$$

Substituting the above equation into (69), and using lemma 3:

$$\boldsymbol{\lambda}^{-1}\varepsilon\frac{\partial\mu_G(s)}{\partial s}+\Lambda\boldsymbol{e}=((\varepsilon+1)\bar{\mu}_G+1)(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})[-\boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{A})\partial\mathbf{Y}+\boldsymbol{\alpha}^{-\mathbf{1}}\partial\mathbf{Z}] \tag{80}$$

where $\boldsymbol{e}$ is the vector of ones and Λ is a constant given by

$$\Lambda=\frac{\partial\mu_G(s)}{\partial s}-\frac{\partial\mu_R(s)}{\partial s}\frac{(\varepsilon+1)\bar{\mu}_G+1}{((1-\sigma)\bar{\mu}_G+1)}\bar{C}^{\sigma}-((\varepsilon+1)\bar{\mu}_G+1)\frac{\partial\log(Y(s))}{\partial s}$$

Combining with (74), I have

$$\Lambda=((\varepsilon+1)\bar{\mu}_G+1)(\sigma\frac{\partial\log(C(s))}{\partial s}-\frac{\partial\log(Y(s))}{\partial s})+(1-\frac{(\varepsilon+1)\bar{\mu}_G+1}{\bar{\mu}_G+(1-\sigma)^{-1}})\frac{\partial\mu_G(s)}{\partial s}$$

For the shocks, I define

$$\partial\boldsymbol{A}\equiv\frac{\partial log(A(s))}{\partial s},\quad \partial\boldsymbol{G}\equiv\frac{\partial log(G(s))}{\partial s},\quad \partial\boldsymbol{Z}\equiv\frac{\partial log(Z(s))}{\partial s}$$

Using (62)(64), I have

$$\Lambda=((\varepsilon+1)\bar{\mu}_G+1)[\frac{\sigma}{(1-\sigma)}(1+\bar{\tau}^C)\boldsymbol{D}^T\boldsymbol{\alpha}(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\mathbf{L}-\beta^T\partial\mathbf{C}-\partial\mathbf{A}]$$
$$+(1-\frac{(\varepsilon+1)\bar{\mu}_G+1}{\bar{\mu}_G+(1-\sigma)^{-1}})\frac{\partial\mu_G(s)}{\partial s} \tag{81}$$

$$\frac{(1+\bar{\tau}^C)}{(1-\sigma)}\boldsymbol{D}^T\boldsymbol{\alpha}(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\mathbf{L}=\frac{\bar{Y}}{\bar{C}}(\beta^T\partial\mathbf{C}+\partial\mathbf{A})-\frac{\bar{G}}{\bar{C}}\partial\boldsymbol{G} \tag{82}$$

To summarize, given shocks $\{\partial\boldsymbol{A},\partial\boldsymbol{G},\partial\boldsymbol{Z}\}$, I have equations (75)(76)(77)(80)(82) to solve for $\{\partial\mathbf{L},\partial\mathbf{P},\partial\mathbf{C},\partial\mathbf{Y},\frac{\partial\mu_G(s)}{\partial s}\}$.

Taxation

By construction, as $P_j=\frac{p_j(s)}{P(s)}C(s)^{-\sigma}$ and $\mathcal{T}_j(s)L_j(s)=\frac{(1-\tau_j^{Ind})\alpha_j p_j(s)Y_j(s)}{P(s)}C(s)^{-\sigma}$, I have

$$(1-\tau_j^{Ind}(s))=\frac{\mathcal{T}_j(s)L_j(s)}{\alpha_j P_j(s)Y_j(s)} \tag{83}$$

In the steady state, $\bar{\tau}_j^{Ind}=0$, I have

$$\partial\boldsymbol{\tau}^{\boldsymbol{Ind}}\equiv\begin{bmatrix}\frac{\partial\tau_1^{Ind}(s)}{\partial s}\\ \vdots\\ \frac{\partial\tau_N^{Ind}(s)}{\partial s}\end{bmatrix}=\partial\boldsymbol{Y}+\partial\boldsymbol{P}-\partial\boldsymbol{\mathcal{T}}-\partial\boldsymbol{\mathcal{L}}=\partial\boldsymbol{Y}+\partial\boldsymbol{P}-(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\boldsymbol{\mathcal{L}} \tag{84}$$

As $p_j(s)C_j(s)(1+\tau_j^C(s))=\beta_j P(s)Y(s)$, I have

$$\partial\boldsymbol{\tau}^{\boldsymbol{C}}\equiv\begin{bmatrix}\frac{\partial\tau_1^{C}(s)}{\partial s}\\ \vdots\\ \frac{\partial\tau_N^{C}(s)}{\partial s}\end{bmatrix}=(1+\bar{\tau}^C)([\beta^T\partial\mathbf{C}+\partial\mathbf{A}-\frac{\sigma}{(1-\sigma)}(1+\bar{\tau}^C)\boldsymbol{D}^T\boldsymbol{\alpha}(\boldsymbol{\lambda}^{-1}\varepsilon+\boldsymbol{I})\partial\mathbf{L}]\boldsymbol{e}-\partial\mathbf{P}-\partial\mathbf{C}) \tag{85}$$

When $\lambda^i=\bar{\lambda},\forall i$, it's easy to check that $\partial\boldsymbol{\mathcal{L}}=l\boldsymbol{e},\partial\boldsymbol{\mathcal{P}}=\bar{\lambda}^{-1}\varepsilon l\boldsymbol{e}-(I-A)^{-1}\partial\boldsymbol{Z},\partial\boldsymbol{C}=\partial\boldsymbol{Y}=l\boldsymbol{e}+(I-A)^{-1}\partial\boldsymbol{Z}$. l is a constant determined by the steady state values and shocks. By using the fact that $\beta_i\frac{\bar{Y}_i}{\bar{C}_i}=(1+\bar{\tau}^C)D_i$, I can compute l:

$$l=\frac{\frac{\bar{Y}}{\bar{C}}((1+\bar{\tau}^C)\boldsymbol{D}^T\partial\boldsymbol{Z}+\partial\boldsymbol{A})-\frac{\bar{G}}{\bar{C}}\partial\boldsymbol{G}}{\frac{1}{(1-\sigma)}(\bar{\lambda}^{-1}\varepsilon+1)-\frac{\bar{Y}}{\bar{C}}} \tag{86}$$

Immediately, I have $\partial\boldsymbol{\tau}^{\boldsymbol{Ind}}=\mathbf{0}$ for this homogeneous information case. This result is what I have shown in the proposition 2. For different λ^i, let $\bar{\lambda} \equiv \frac{\sum_{i=1}^N \alpha_i D_i \lambda_i}{\sum \alpha_i D_i}$. The equations are non-linear in λ_i, so I do the linear approximation of these equations for λ_i around $\bar{\lambda}$.

Let $\tilde{\lambda_i} = \lambda_i^{-1} - \bar{\lambda}^{-1}$. Heterogeneous information structure leads to changes in the solutions, which I define as the following new set of solutions:

$$\partial \boldsymbol{\mathcal{L}} = l\boldsymbol{e} + \partial \tilde{\boldsymbol{\mathcal{L}}}, \quad \partial \boldsymbol{\mathcal{P}} = \bar{\lambda}^{-1} \varepsilon l \boldsymbol{e} - (I - A)^{-1} \partial \boldsymbol{Z} + \partial \tilde{\boldsymbol{\mathcal{P}}}, \quad \partial \boldsymbol{Y} = l\boldsymbol{e} + (I - A)^{-1} \partial \boldsymbol{Z} + \partial \tilde{\boldsymbol{Y}}$$

The linear approximation of (77) for λ_i is

$$\partial \tilde{\boldsymbol{Y}} = \varepsilon l (\boldsymbol{I} - \boldsymbol{\alpha}) \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} + [(\varepsilon \bar{\lambda}^{-1} + 1)(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}] \partial \tilde{\boldsymbol{\mathcal{L}}} - A \partial \tilde{\boldsymbol{\mathcal{P}}} \tag{87}$$

The equation (75) is equivalent to

$$\begin{aligned} \boldsymbol{X}^{\boldsymbol{Used}}[(\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial \mathbf{L} - \boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}^{-1}\varepsilon + \boldsymbol{I}]\partial \mathbf{L} \\ + \boldsymbol{X}^{\boldsymbol{Used}}(\boldsymbol{I} - \boldsymbol{A})\partial \mathbf{P} + \boldsymbol{X}^{\boldsymbol{Used}}\partial \mathbf{Z} = 0 \end{aligned}$$

Then I have

$$\begin{aligned} \boldsymbol{X}^{\boldsymbol{Used}} \varepsilon l (\boldsymbol{I} - \boldsymbol{\alpha}) \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} - \boldsymbol{X}^{\boldsymbol{T}} \varepsilon l \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} + \boldsymbol{X}^{\boldsymbol{Used}}[(\bar{\lambda}^{-1}\varepsilon + 1)(\boldsymbol{I} - \boldsymbol{\alpha}) + \boldsymbol{\alpha}]\partial \tilde{\boldsymbol{\mathcal{L}}} \\ - \boldsymbol{X}^{\boldsymbol{T}}(\bar{\lambda}^{-1}\varepsilon + 1)\partial \tilde{\boldsymbol{\mathcal{L}}} + \boldsymbol{X}^{\boldsymbol{Used}}(\boldsymbol{I} - \boldsymbol{A})\partial \tilde{\boldsymbol{\mathcal{P}}} = 0 \end{aligned}$$

Let's define $\boldsymbol{R} \equiv (\boldsymbol{X}^{\boldsymbol{Used}})^{-1}\boldsymbol{X}^{\boldsymbol{T}}$ and $\boldsymbol{\mathcal{L}} \equiv (\boldsymbol{I} - \boldsymbol{A})^{-1}$.The (i, j) element of $\boldsymbol{R}$ is $\frac{X_{ji}}{\sum_{k=1}^N X_{ki}}$, and it is the proportion of industry i's output used as intermediate goods by industry j, relative to the total output of industry i consumed as intermediate goods across all industries. So $\boldsymbol{R}$ is the matrix for input-output linkage, and $\boldsymbol{\mathcal{L}}$ is Leontief inverse. Transforming the above matrix, I have

$$\partial\tilde{\boldsymbol{\mathcal{P}}} = \boldsymbol{\mathcal{L}}(-\varepsilon l(\boldsymbol{I}-\boldsymbol{\alpha})\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} + \varepsilon l\boldsymbol{R}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}$$
$$-((\bar{\lambda}^{-1}\varepsilon+1)(\boldsymbol{I}-\boldsymbol{\alpha})+\boldsymbol{\alpha})\partial\tilde{\boldsymbol{\mathcal{L}}} + \boldsymbol{R}(\bar{\lambda}^{-1}\varepsilon+1)\partial\tilde{\boldsymbol{\mathcal{L}}} \tag{88}$$

For (80), I have

$$(\bar{\lambda}^{-1}\varepsilon\frac{\partial\tilde{\mu}_G(s)}{\partial s}+\tilde{\Lambda})\boldsymbol{e}+\varepsilon\frac{\partial\mu_G(s)}{\partial s}|_{\bar{\lambda}}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}$$
$$= -((\varepsilon+1)\bar{\mu}_G+1)\varepsilon l\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} - ((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{A})\partial\tilde{\mathbf{Y}} \tag{89}$$

Substituting (87) and (88) into (89) and noticing that $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{A}\boldsymbol{\mathcal{L}} = \boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\mathcal{L}} = \boldsymbol{A}$, I have

$$-\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial\tilde{\mu}_G(s)}{\partial s}+\tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)}\boldsymbol{e} - \frac{\varepsilon(\frac{\partial\mu_G(s)}{\partial s}|_{\bar{\lambda}}+((\varepsilon+1)\bar{\mu}_G+1)l)}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}$$
$$= \boldsymbol{\alpha}^{-1}[\varepsilon l(\boldsymbol{I}-\boldsymbol{\alpha}-\boldsymbol{A}\boldsymbol{R})\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} + ((\bar{\lambda}^{-1}\varepsilon+1)(\boldsymbol{I}-\boldsymbol{\alpha}-\boldsymbol{A}\boldsymbol{R})+\boldsymbol{\alpha})\partial\tilde{\boldsymbol{\mathcal{L}}}] \tag{90}$$

For the industry revenue tax

$$\partial\boldsymbol{\tau}^{\boldsymbol{Ind}} = -\varepsilon l\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} + \partial\tilde{\boldsymbol{Y}} + \partial\tilde{\boldsymbol{P}} - (\bar{\lambda}^{-1}\varepsilon+1)\partial\tilde{\boldsymbol{\mathcal{L}}}$$

Combining with (87) and (88),

$$\partial \boldsymbol{\tau}^{\boldsymbol{Ind}} = -(\boldsymbol{I}-\boldsymbol{R})[\varepsilon l \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} + (\bar{\lambda}^{-1}\varepsilon + 1)\partial \tilde{\boldsymbol{\mathcal{L}}}] \tag{91}$$

Let $\partial \hat{\boldsymbol{\mathcal{L}}} = \varepsilon l \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} + (\bar{\lambda}^{-1}\varepsilon + 1)\partial \tilde{\boldsymbol{\mathcal{L}}}$. Then from (90) I have

$$\begin{aligned} &\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial \bar{\mu}_G(s)}{\partial s} + \tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)}\boldsymbol{e} + \frac{\varepsilon(\frac{\partial \mu_G(s)}{\partial s}|_{\bar{\lambda}} + ((\varepsilon+1)\bar{\mu}_G+1)l)}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)} \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix} \\ &= (\boldsymbol{I} - \boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))\partial \hat{\boldsymbol{\mathcal{L}}} - \partial \tilde{\boldsymbol{\mathcal{L}}} \end{aligned}$$

which implies

$$\begin{aligned} \partial \hat{\boldsymbol{\mathcal{L}}} &= (\frac{\varepsilon}{\varepsilon+\bar{\lambda}}\boldsymbol{I} - \boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))^{-1} \\ &\cdot(\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial \bar{\mu}_G(s)}{\partial s} + \tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)}\boldsymbol{e} + \frac{\varepsilon)\frac{\partial \mu_G(s)}{\partial s}|_{\bar{\lambda}}}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)} \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix}) \end{aligned}$$

By using $(\frac{\varepsilon}{\varepsilon+\bar{\lambda}}\boldsymbol{I} - \boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))\boldsymbol{e} = \frac{\varepsilon}{\varepsilon+\bar{\lambda}}\boldsymbol{e} - \boldsymbol{e} = \frac{-\lambda}{\varepsilon+\bar{\lambda}}\boldsymbol{e}$ and $(\boldsymbol{I}-\boldsymbol{R})\boldsymbol{e} = \mathbf{0}$, it shows that

$$(\boldsymbol{I}-\boldsymbol{R})(\frac{\varepsilon}{\varepsilon+\bar{\lambda}}\boldsymbol{I} - \boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial \bar{\mu}_G(s)}{\partial s} + \tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)(\bar{\lambda}^{-1}\varepsilon+1)}\boldsymbol{e} = \mathbf{0}$$

Thus I find

$$\partial \boldsymbol{\tau}^{\boldsymbol{Ind}} = -(\boldsymbol{I}-\boldsymbol{R})\partial \hat{\boldsymbol{\mathcal{L}}} = \frac{-\frac{\partial \mu_G(s)}{\partial s}|_{\bar{\lambda}}}{(\varepsilon+1)\bar{\mu}_G+1}(\boldsymbol{I}-\boldsymbol{R})(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))^{-1} \begin{bmatrix} \tilde{\lambda_1} \\ \vdots \\ \tilde{\lambda_N} \end{bmatrix}$$

Combining (81) and (82), I have

$$\frac{-\frac{\partial \mu_G(s)}{\partial s}|_{\bar{\lambda}}}{(\varepsilon+1)\bar{\mu}_G+1} = \frac{(\frac{\bar{\lambda}^{-1}\varepsilon+1}{1-\sigma}-1)l-(1+\bar{\tau}^C)\boldsymbol{D}^T\partial\boldsymbol{Z}-\partial\boldsymbol{A}}{(\bar{\lambda}^{-1}\varepsilon+1)-\frac{(\varepsilon+1)\bar{\mu}_G+1}{\bar{\mu}_G+(1-\sigma)^{-1}}}$$

Substituting the expression of l and transforming the equation, I have

$$\frac{-\frac{\partial \mu_G(s)}{\partial s}|_{\bar{\lambda}}}{(\varepsilon+1)\bar{\mu}_G+1} = \chi_Z\boldsymbol{D}^T\partial\boldsymbol{Z}+\chi_A\partial\boldsymbol{A}+\chi_G\partial\boldsymbol{G}$$

$$\chi_Z = \frac{\frac{\bar{Y}}{\bar{C}}\frac{\bar{G}}{\bar{C}}(\bar{\lambda}^{-1}\varepsilon)+1)}{(\bar{\lambda}^{-1}\varepsilon+1-\frac{\bar{Y}}{\bar{C}}(1-\sigma))^2} > 0$$

$$\chi_A = \frac{\frac{\bar{G}}{\bar{C}}(\bar{\lambda}^{-1}\varepsilon+1)}{(\bar{\lambda}^{-1}\varepsilon+1-\frac{\bar{Y}}{\bar{C}}(1-\sigma))^2} > 0$$

$$\chi_G = -\frac{\frac{\bar{G}}{\bar{C}}(\bar{\lambda}^{-1}\varepsilon+1-(1-\sigma))}{(\bar{\lambda}^{-1}\varepsilon+1-\frac{\bar{Y}}{\bar{C}}(1-\sigma))^2} < 0$$

And it's easy to check that for χ_Z, χ_A and χ_G, their absolute values increase when $\frac{\bar{G}}{\bar{Y}}$ or $\bar{\lambda}$ increase if $\sigma \in [0,1]$. At the same time, I can get

$$\partial\hat{\boldsymbol{\mathcal{L}}} = -\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial\tilde{\mu}_G(s)}{\partial s}+\tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)}\boldsymbol{e}-$$
$$(\chi_Z\boldsymbol{D}^T\partial\boldsymbol{Z}+\chi_A\partial\boldsymbol{A}+\chi_G\partial\boldsymbol{G})(\boldsymbol{I}-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-\mathbf{1}}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} \tag{92}$$

For future use, I let $k \equiv -\frac{\bar{\lambda}^{-1}\varepsilon\frac{\partial\tilde{\mu}_G(s)}{\partial s}+\tilde{\Lambda}}{((\varepsilon+1)\bar{\mu}_G+1)}$ to simplify the equations. To solve for k, (82) implies

$$\boldsymbol{D}^T\boldsymbol{\alpha}\partial\hat{\mathbf{L}} = (1-\sigma)\beta^T\partial\tilde{\mathbf{Y}} \tag{93}$$

From (85), I have

$$\partial\tilde{\boldsymbol{\tau}}^{C} = (1+\bar{\tau}^{C})([\beta^{T}\partial\tilde{\mathbf{Y}} - \frac{\sigma}{(1-\sigma)}(1+\bar{\tau}^{C})\boldsymbol{D}^{T}\boldsymbol{\alpha}\partial\hat{\mathbf{L}}]\boldsymbol{e} - \partial\tilde{\mathbf{P}} - \partial\tilde{\mathbf{Y}}) \tag{94}$$

To replace $\partial\tilde{\boldsymbol{Y}}$ by $\partial\hat{\mathbf{L}}$, I use

$$\partial\tilde{\mathbf{Y}} = \mathcal{L}(\boldsymbol{I} - \boldsymbol{AR} - \boldsymbol{\alpha})\partial\hat{\mathbf{L}} + \frac{1}{(\bar{\lambda}^{-1}\varepsilon+1)}\mathcal{L}\boldsymbol{\alpha}(\partial\hat{\mathbf{L}} - \begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}) \tag{95}$$

$$\partial\tilde{\mathbf{Y}} + \partial\tilde{\mathbf{P}} = \boldsymbol{R}\partial\hat{\mathbf{L}} \tag{96}$$

Combining (93) (94) and (96),

$$\partial\tilde{\boldsymbol{\tau}}^{C} = \frac{\bar{Y}}{\bar{C}}((1 - \frac{\sigma}{(1-\sigma)}\bar{\tau}^{C})\boldsymbol{D}^{T}\boldsymbol{\alpha}\partial\hat{\mathbf{L}}\boldsymbol{e} - \boldsymbol{R}\partial\hat{\mathbf{L}}) \tag{97}$$

Substituting (92) into (97), I have

$$\begin{aligned}\partial\tilde{\boldsymbol{\tau}}^{C} = & -\frac{\bar{G}}{\bar{C}}k\boldsymbol{e} - \frac{\bar{Y}}{\bar{C}}(\chi_Z\boldsymbol{D}^{T}\partial\boldsymbol{Z} + \chi_A\partial\boldsymbol{A} + \chi_G\partial\boldsymbol{G})\boldsymbol{D}^{T}\boldsymbol{\alpha}(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}\boldsymbol{e}\\ & +(\chi_Z\boldsymbol{D}^{T}\partial\boldsymbol{Z} + \chi_A\partial\boldsymbol{A} + \chi_G\partial\boldsymbol{G})\boldsymbol{R}(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}\end{aligned} \tag{98}$$

Using (93) and (95) and fact that $\beta^{T}\mathcal{L} = \boldsymbol{D}^{T}(1+\bar{\tau}^{C})$ and $\boldsymbol{D}^{T}\boldsymbol{\alpha}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} = 0$, I have

$$k = \frac{\frac{\bar{Y}}{\bar{C}}(\bar{\lambda}^{-1}\varepsilon+1)(\chi_Z\boldsymbol{D}^{T}\partial\boldsymbol{Z} + \chi_A\partial\boldsymbol{A} + \chi_G\partial\boldsymbol{G})}{(\bar{\lambda}^{-1}\varepsilon + 1 - \frac{\bar{Y}}{\bar{C}}(1-\sigma))}\boldsymbol{D}^{T}\boldsymbol{\alpha}(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I}-\boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}\boldsymbol{e}$$

Substituting the expression of k into (98), I finally get

$$\partial\tilde{\boldsymbol{\tau}}^{\boldsymbol{C}} = -\frac{(\frac{\bar{Y}}{\bar{C}})^2(\bar{\lambda}^{-1}\varepsilon + \sigma\frac{\bar{Y}}{\bar{C}})}{(\bar{\lambda}^{-1}\varepsilon + 1 - \frac{\bar{Y}}{\bar{C}}(1-\sigma))}(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_A \partial \boldsymbol{A} + \chi_G \partial \boldsymbol{G})\boldsymbol{D}^T \boldsymbol{\alpha}(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix}\boldsymbol{e}$$

$$+(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_A \partial \boldsymbol{A} + \chi_G \partial \boldsymbol{G})\boldsymbol{R}(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1}\begin{bmatrix}\tilde{\lambda_1}\\ \vdots \\ \tilde{\lambda_N}\end{bmatrix} \tag{99}$$

From (85), I have

$$\partial\bar{\boldsymbol{\tau}}^{\boldsymbol{C}} = \frac{\bar{Y}}{\bar{C}}\frac{\bar{G}}{\bar{C}}\frac{(\bar{\lambda}^{-1}\varepsilon + \sigma)\partial \boldsymbol{G} - (\bar{\lambda}^{-1}\varepsilon + 1)(\frac{\bar{Y}}{\bar{C}}\boldsymbol{D}^T \partial \boldsymbol{Z} + \partial \boldsymbol{A})}{(\varepsilon\bar{\lambda}^{-1} + 1 - \frac{\bar{Y}}{\bar{C}}(1-\sigma))}\boldsymbol{e} \tag{100}$$

I know $\boldsymbol{\tau}^{\boldsymbol{C}} = \frac{\bar{G}}{\bar{C}}\boldsymbol{e} + \partial\bar{\boldsymbol{\tau}}^{\boldsymbol{C}} + \partial\tilde{\boldsymbol{\tau}}^{\boldsymbol{C}}$. Combining (99), (100) and the expression of χ_Z, χ_A and χ_G, I prove the theorem.

Proof for the examples

1. 'Tree' Networks:

Given this structure of production networks, I have R_{im} to be either 0 or 1. Whenever a_{im} is nonzero, it indicates that industry i requires input from industry m, so I must have $R_{mi} = 1$ and $R_{mj} = 1$ for any $j \neq i$ for industry m. Thus

$$(AR)_{ij} = \sum_{m=1}^{N} a_{im} R_{mj} = \begin{cases} \sum_{m=1}^{N} a_{im}, & \text{if } j = i, \\ 0, & \text{if } j \neq i. \end{cases}$$

which gives

$$(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon} \boldsymbol{\alpha}^{-1} (\boldsymbol{I} - \boldsymbol{AR}))^{-1} = -\frac{\varepsilon}{\bar{\lambda}} \boldsymbol{I}$$

The special case for this is a two sector model with vertical structure. Since $R = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$[24], I have the expression (15)(16) in the main text.

2. Multiple Downstreams:

Given the expressions of A and R in (20), I have

$$AR = \begin{pmatrix} a_{1N} b_1 & a_{1N} b_2 & \cdots & a_{1N} b_{N-1} & 0 \\ a_{2N} b_1 & a_{2N} b_2 & \cdots & a_{2N} b_{N-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{(N-1)N} b_1 & a_{(N-1)N} b_2 & \cdots & a_{(N-1)N} b_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

Thus,

[24] The validity of assigning a zero value to R follows from the discussion in the main text.

$$
I - AR = \begin{pmatrix} 1 - a_{1N}b_1 & -a_{1N}b_2 & \cdots & -a_{1N}b_{N-1} & 0 \\ -a_{2N}b_1 & 1 - a_{2N}b_2 & \cdots & -a_{2N}b_{N-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{(N-1)N}b_1 & -a_{(N-1)N}b_2 & \cdots & 1 - a_{(N-1)N}b_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}
$$

As $\boldsymbol{\alpha}^{-1} = \mathrm{diag}\left(\frac{1}{1-a_{1N}}, \frac{1}{1-a_{2N}}, \dots, \frac{1}{1-a_{(N-1)N}}, 1\right)$, I have

$$
I - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(I - AR) = \begin{pmatrix} 1 - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{1-a_{1N}b_1}{1-a_{1N}} & -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{1N}b_2}{1-a_{1N}} & \cdots & -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{1N}b_{N-1}}{1-a_{1N}} & 0 \\ -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{2N}b_1}{1-a_{2N}} & 1 - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{1-a_{2N}b_2}{1-a_{2N}} & \cdots & -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{2N}b_{N-1}}{1-a_{2N}} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{(N-1)N}b_1}{1-a_{(N-1)N}} & -\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{a_{(N-1)N}b_2}{1-a_{(N-1)N}} & \cdots & 1 - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\cdot\frac{1-a_{(N-1)N}b_{N-1}}{1-a_{(N-1)N}} & 0 \\ 0 & 0 & \cdots & 0 & 1 - \frac{\varepsilon+\bar{\lambda}}{\varepsilon} \end{pmatrix}
$$

The change in the industry's revenue tax satisfies

$$
\Delta\boldsymbol{\tau}^{\boldsymbol{Ind}} = -(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})(\boldsymbol{I} - \boldsymbol{R})(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{AR}))^{-1} \begin{pmatrix} 0 \\ \vdots \\ \Delta\lambda_i \\ \vdots \\ 0 \end{pmatrix}
$$

It gives us

$$
\Delta\tau_j^{Ind} = \begin{cases} -(\boldsymbol{\chi}_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \boldsymbol{\chi}_G \partial \boldsymbol{G})\frac{\varepsilon+\bar{\lambda}}{\varepsilon} a_{jN} b_i \frac{\left(1-\frac{\varepsilon+\bar{\lambda}}{\varepsilon}\right)\cdot\prod_{m\neq i,j}\left(\frac{\varepsilon+\bar{\lambda}}{\varepsilon}-1+a_{mN}\right)}{\mathcal{M}\prod_{m\neq i}(1-a_{mN})}\Delta\lambda_i & \text{if } j \neq i \text{ and } j < N \\ -\sum_{k=1}^{N-1} b_k \Delta\tau_k^{Ind} & \text{if } j = N. \end{cases}
$$

where $\mathcal{M} = \det\left(\boldsymbol{I} - \frac{\varepsilon+\bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}\left(\boldsymbol{I} - \boldsymbol{AR}\right)\right)$. $\mathcal{M}$ always takes the negative values if $\sum_{i=1}^{N} b_i = 1$ and $0 < a_{iN} < 1$. Thus I show that $\Delta\tau_j^{Ind} < 0$ when $j < N$ and $j \neq i$.

Finally, I have

$$\frac{\Delta\tau_j^{Ind}}{\Delta\tau_k^{Ind}} = \frac{\frac{\bar{\lambda}}{\varepsilon} + a_{kN}}{\frac{\bar{\lambda}}{\varepsilon} + a_{jN}}$$

which implies that $|\Delta\tau_j^{Ind}| \gtreqless |\Delta\tau_k^{Ind}|$ iff $a_{jN} \gtreqless a_{kN}, \forall j,k \neq i$;

Appendix B Empirical Evidence

B.1 Texture analysis of attention:

I employ dictionary-based frequency counts that identify the attention of different industries toward different macroeconomic topics. The methodology is based on the approach of Song and Stern (2023). For each topic, I match each topic with a keyword dictionary composed of terms and phrases frequently found in Econoday, which provides updates on significant economic events and is the service behind the Bloomberg economic calendar.

Topic	**Keywords**
Output	GDP, economic growth, macroeconomic condition, construction spending, national activity, recession
Government Spending	government spending
Fiscal Policy	fiscal deficit, fiscal policy, tax rebate, government subsidy, government support
Production Networks	intermediate input, intermediate goods, upstream, downstream

Table A1: Macroeconomic topics and keywords

For the U.S., I use all electronically available 10-K and 10-Q filings by publicly listed US companies between 1994 and 2023. For China, I use the annual report of all listed firms in Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) between 2001 and 2022. The keywords for China are the same set as for the U.S., except that they are translated into Chinese.

I classify U.S. industries using the 2-digit NAICS system. For China, the industry classification follows the standards in the ”Industrial Classification for National Economic Activities,” as defined by the National Bureau of Statistics of China.

Apple Inc.

Form 10-Q

For the Fiscal Quarter Ended December 30, 2023

TABLE OF CONTENTS

		Page
	Part I	
Item 1.	Financial Statements	1
Item 2.	Management's Discussion and Analysis of Financial Condition and Results of Operations	13
Item 3.	Quantitative and Qualitative Disclosures About Market Risk	18
Item 4.	Controls and Procedures	18
	Part II	
Item 1.	Legal Proceedings	19
Item 1A.	Risk Factors	19
Item 2.	Unregistered Sales of Equity Securities and Use of Proceeds	20
Item 3.	Defaults Upon Senior Securities	21
Item 4.	Mine Safety Disclosures	21
Item 5.	Other Information	21
Item 6.	Exhibits	21

Figure A1: Example: FORM 10-Q of Apple

Notes: The content of the 10-Q report for Apple for the fiscal quarter ended December 30.

The input-output data of China comes from the Asian Development Bank (ADB). To compile the industry's output data with the coding system following the NBS of China, the table below shows the mapping for the industry names:

Industry: NBS of China	**Industry: ADB**
Mining	Mining and quarrying
Manufacturing	Food, beverages, and tobacco Textiles and textile products Leather, leather products, and footwear Wood and products of wood and cork Pulp, paper, paper products, printing, and publishing Coke, refined petroleum, and nuclear fuel Chemicals and chemical products Rubber and plastics Other nonmetallic minerals Basic metals and fabricated metal Machinery, nec Electrical and optical equipment Transport equipment Manufacturing, nec; recycling
Utilities	Electricity, gas, and water supply
Construction	Construction
Wholesale & Retail	Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of fuel Wholesale trade and commission trade, except of motor vehicles and motorcycles Retail trade, except of motor vehicles and motorcycles; repair of household goods
Accommodation & Catering	Hotels and restaurants
Transport & Postal	Inland transport Water transport Air transport Other supporting and auxiliary transport activities; activities of travel agencies
IT & Software	Post and telecommunications
Finance	Financial intermediation
Real Estate	Real estate activities
Leasing & Business Services	Renting of M&Eq and other business activities
Public Admin & Social Services	Public administration and defense; compulsory social security Other community, social, and personal services
Education	Education
Health & Social Work	Health and social work

Table A2: Industry Classification

I construct an alternative measure of exposure to business shocks using the Integrated Industry-Level Production Account (KLEMS). This dataset provides estimates of total factor productivity (TFP) and labor productivity across various industries. An industry's exposure is measured by its correlation with either the aggregate TFP (using Total Factor Productivity at Constant National Prices for the United States) or labor productivity (using Constant GDP per capita for the United States) from the Federal Reserve Bank of St. Louis (FRED). The results confirm that an industry's attention remains positively associated with its exposure to business cycle shocks.

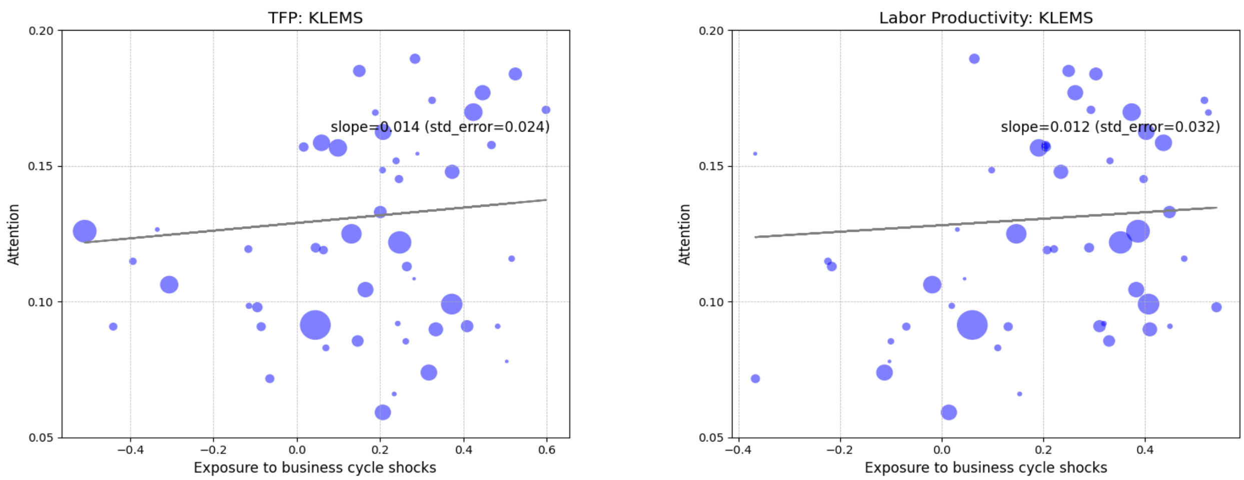

Figure A2: Exposure and attention: alternative construction

Note: The exposures to business shocks are computed using the correlation of industry TFP (left) or industry labor productivity (right) with the aggregate TFP or aggregate labor productivity. The attention is computed by taking the average attention index from 1997 to 2023. The three digit NAICS system specifies the industry.

B.2 Regression:

To check the robustness of the calibration for β_0 and β_1, I refer to other ways to construct the attention index. I use the attention index constructed by 10-K filling or the annualized 10-Q attention index (averaging them within the same year). The assumption is the same: the affine relationship exists between information precision and attention. Table A3 shows the regression outcome for these two attention indexes. The value of β_1 ranges from 0.342 to 0.691, with the baseline calibration for the U.S. set at 0.361, aligning closely with estimates from other approaches. Since I do not have firms' forecasts of overall economic output (GDP) for China, I use the 10-K attention result as the baseline calibration for China. Although the actual value of β_1 and β_1 may differ, in terms of the focus for the industrial policy by using the industrial revenue tax, the theory suggests that changes in optimal industrial revenue taxation are proportional to changes in β_1 and independent of changes in β_0.

	10-K Attention		Average 10-Q Attention	
	(1) Forecast Difference	(2) Forecast Revision	(1) Forecast Difference	(2) Forecast Revision
β_1	0.667**	0.691***	0.342	0.499**
	(0.332)	(0.206)	(0.235)	(0.206)
β_0		0.0141		0.0568
		(0.0555)		(0.0558)
$1-\beta_0$	0.590***		0.270***	
	(0.0993)		(0.0642)	

Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A3: The estimation of β_0 and β_1 for 10-K and average 10-Q attention

Note: The first column presents the results for regression (30), using the difference between the mean forecast and individual forecasts as the dependent and independent variable. The second column shows the results for regression (31), with forecast revisions as the dependent variable.

B.3 Quantitative:

To compute the optimal tax for the U.S. and China, I use the input-output table data from BEA for the US and ADB for China. The theorem shows that optimal taxation relies on two key matrices: the input reliance matrix A and the output allocation matrix R. Figure A3 shows the heatmap for these two matrices for both countries:

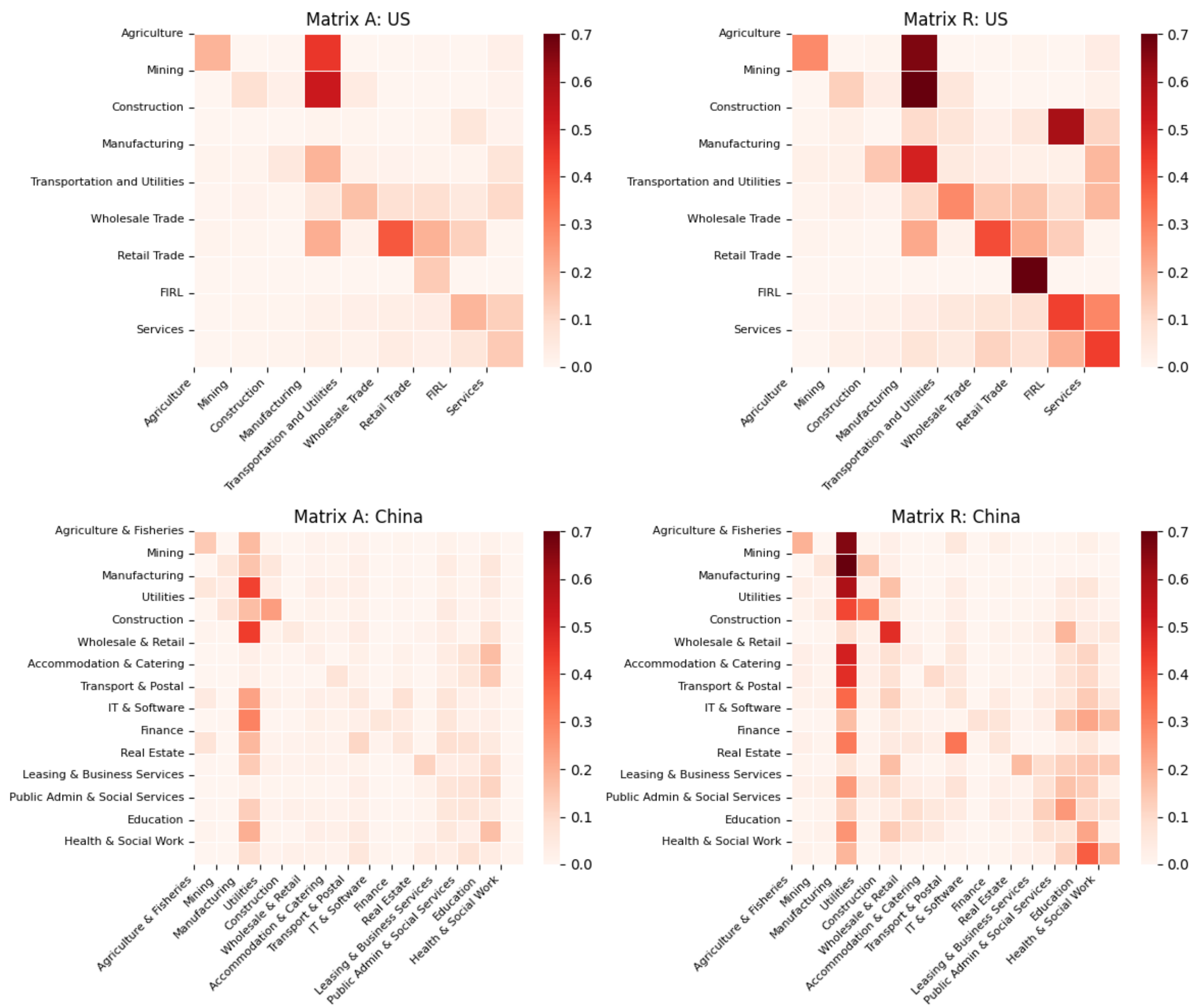

Figure A3: Heatmap of production network matrices

To compute the optimal taxation for the industries within the manufacturing sector, I recompute the input-output table for China, where the manufacturing sector is disaggregated into sub-industries to produce different types of goods. Figure A4 shows the heatmap:

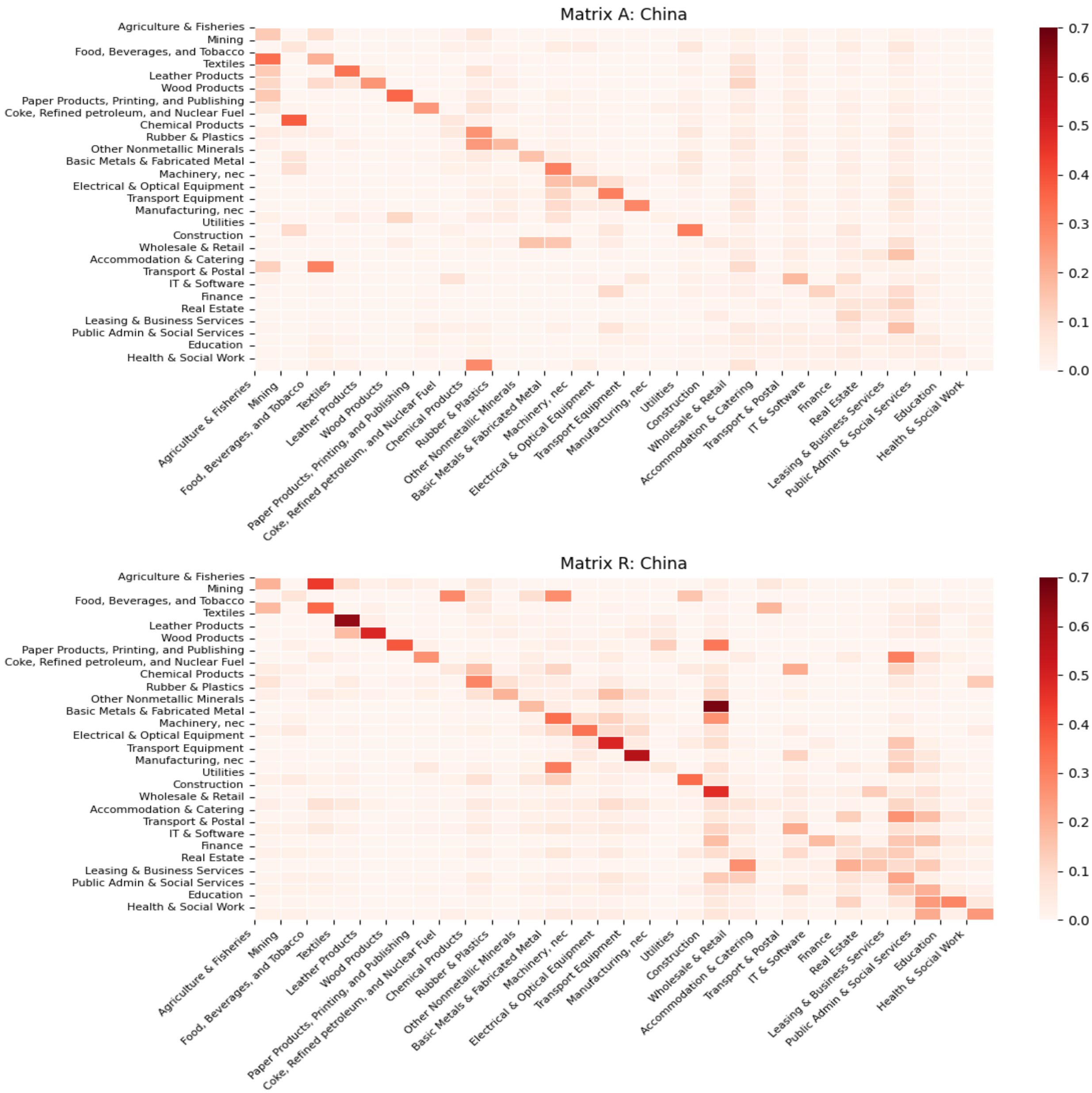

Figure A4: Heatmap include industries within manufacturing sector

Appendix C Dynamic Model

I extend the framework into a dynamic model:

Model

Preference and Technology:

The preference of a representative household over consumption and labor is given by:

$$\mathbb{E}_{HH} \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \sum_{i=1}^{N} \frac{1}{\varepsilon + 1} \int_{k \in [0,1]} n_{ik,t}^{\varepsilon+1} dj \right)$$

The household's period-t budget constraint can be expressed in nominal terms, as follows:

$$\begin{aligned} &(1 + \tau_t^C) P_t C_t + B_{t+1} + \int_{s \in S^{t+1}} Q_{t+1,s} D_{t+1,s} ds = \\ &\sum_{i=1}^{N} \int_{j \in [0,1]} [w_{ij,t} n_{ij,t} + \pi_{ij,t}] dj + (1 + R_t) B_t + D_{t,s^t} \end{aligned}$$

where B_{t+1} denote non-contingent debt instrument, R_t denotes the nominal interest rate between t and $t+1$. $D_{t+1,s}$ denote the quantities of state-contingent assets (or Arrow securities), $Q_{t+1,s}$ denote the cost of the state-contingent assets, and τ_t^C denote proportional tax on consumption. The production side is the same as the static model. The key assumption is that D_{0,s^0} equals zero for any realization of s_0. In other words, the time 0 shock is not insured by the state-contingent asset, meaning period 0 shock is treated as unexpected. For the Cobb-Douglas production function, the effective tax rate on the final consumption goods is given by

$$(1 + \tau_t^C) = \Pi_{m=1}^{N} (1 + \tau_{m,t}^C) \tag{101}$$

Government:

The government's budget constraint for period t, in nominal terms, is given by

$$B_t(1+R_t) + D_{t,s^t} + P_t G_t = B_{t+1} + \int_{s \in S^{t+1}} Q_{t+1,s} D_{t+1,s} ds + T_t \tag{102}$$

where G_t is exogenous real level of government spending and T_t is nominal level of tax revenue:

$$T_t = \sum_{i=1}^{N} \int_{j \in [0,1]} \tau_{i,t}^{Ind} p_{i,t} y_{ij,t} dj + \sum_{i=1}^{N} \tau_{i,t}^{C} p_{i,t} c_{i,t} \tag{103}$$

Information Friction:

Nature draws a random variable s_t from a set S_t in each period t. The state at t is represented by the history of shocks, denoted as $s^t \equiv \{s_0, s_1, \ldots, s_t\}$. The agents on island j of industry i receive a random variable ω_{ij}^t from a set Ω_i^t, following a probability distribution $\Phi_{i,t}(\omega_{ij}^t | s^t)$. This variable captures all the information that the firm and worker on island i of industry i has in period t. The probability distribution $\Phi_{i,t}$ is industry-specific, meaning that firms in different industries may receive signals from different distributions. There are two stages. In stage 1, the firm decides labor demand $l_{ij}(\omega_{ij}^t)$ to maximize the utility-adjusted after-tax profit, given his information $\omega_{ij,t}$. The worker decides labor supply $n_{ij}(\omega_{ij}^t)$ to maximize the expected utility. In stage 2, when prices and tax rates are realized, the firm chooses intermediate input $x_{ij}(\omega_{ij}^t, s^t)$ to maximize its profit.

I am restricting the information frictions to the production side where the representative household and the government have complete information as Angeletos and La'O (2020). The household chooses consumption $C_t(s^t)$, the non-contingent debt $B_{t+1}(s^t)$ and the state-contingent asset $D_{t+1,s^{t+1}}(s^{t+1})$ to maximize his expected utility function given the prices and policy rates subject to its budget constraint.

Ramsey Problem

The prices of different industry goods, consumption goods, government bonds, and state-contingent assets are functions of state s^t:

$$\{p_{i,t}(s^t), P_t(s^t), R_t(s^t), Q_{t,s^{t+1}}(s^{t+1}))\}$$

The Ramsey planner sets the policy rates that are state-contingent on the whole history s^t:

$$\{\tau_{i,t}^{Ind}(s^t), \tau_{i,t}^{C}(s^t)\}$$

The labor functions $n_{ij,t}(\omega_{ij}^t)$, $l_{ij,t}(\omega_{ij}^t)$ are measurable to the signal ω_{ij}^t, and intermediate input $x_{ij,t}(\omega_{ij}^t, s^t)$ are measurable to the tuple (ω_{ij}^t, s^t). Given the tax rates, an equilibrium is triplet of allocations, prices and policy rates that satisfy (i) $C_t(\cdot)$, $B_t(\cdot)$ and $D_{t,s}(\cdot)$ solve the household's problem; (ii) $l_{ij,t}(\cdot)$ and $x_{ij,t}(\cdot)$ solve the firm's problem; (iii) $n_{ij,t}(\cdot)$ solves the worker's problem (iv) the resource constraint is satisfied; (v) the government's budget constraint is satisfied; and (vi) all markets clear. The Ramsey planner chooses the tax functions to maximize welfare.

Optimal Taxation

The following proposition characterizes the feasible set of equilibrium:

Proposition 5. *A feasible allocation, $\xi \in \mathcal{X}$, is part of an equilibrium if and only if the following two properties hold: (i) the allocation satisfies the implementability condition:*

$$\sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t|s^0} C_t^{1-\sigma}(s^t) - \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t|s^0} \sum_{i=1}^{N} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t) = 0 \quad \forall s_0 \tag{104}$$

(ii) for any t and s^t, there exist functions $\psi_{i,t}^C(s^t)$ and $\psi_{i,t}^{Ind}(s^t)$ such that the equilibrium conditions from f.o.s of households and firms are satisfied:

$$\prod_{i=1}^{N} \left(\frac{z_{i,t}(s^t) L_{i,t}^{\alpha_i}(s^t) \prod_{k=1}^{N} X_{ik,t}^{a_{ik}}(s^t) - \sum_{k=1}^{N} X_{ki,t}(s^t)}{\beta_i} \right)^{\beta_i} - G_t(s^t) - C_t(s^t) = 0 \tag{105}$$

$$\frac{a_{ij}}{\alpha_i} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t) = P_{j,t}(s^t) X_{ij,t}(s^t) \quad \forall i, j \tag{106}$$

$$\mathbb{E}_{s'|\omega_{ik}} \mathcal{T}_L(s') = n_{ij,t}^{\varepsilon}(\omega_{ij}^t) \quad \forall i \tag{107}$$

$$L_{i,t}(s^t) = \int_{j \in [0,1]} n_{ij,t}(\omega_{ij}^t) dj \tag{108}$$

where

$$P_{i,t}(s^t) = \psi^C_{i,t}(s^t)\beta_i C_t(s^t)^{-\sigma}\frac{C_t(s^t)}{c_{i,t}(s^t)}$$

$$\mathcal{T}_{i,t}(s) = \psi^{Ind}_{i,t}(s^t)\beta_i C_t(s^t)^{-\sigma}\frac{C_t(s^t)}{c_{i,t}(s^t)} z_{i,t}(s^t)\alpha_i L^{\alpha_i-1}_{i,t}(s)\prod_{k=1}^{N} X^{a_{ik}}_{ik,t}(s^t)$$

Definition 3. *The information structure is homogeneous iff each industry i receive signals from the same distribution ω^t conditional on s^t: $\Phi_i(\omega^t|s^t) \equiv \Phi(\omega^t|s^t)$.*

Theorem 3 (***Generalizing Proposition* 2**)**.** *If the information structure is heterogeneous, the optimal taxation is to set*

$$\tau^{Ind}_{i,t}(s^t) = 0 \quad \forall i, t, s^t$$
$$\tau^C_{i,t}(s^t) = \bar{\tau}^C_t(s^t) \quad \forall i, t, s^t$$

We still have equalized tax rates on consumption goods and zero revenue tax. The difference from the complete information case is that the consumption goods tax is not constant. Instead, it depends on the realization of s^t. The reason is that the government should tax agents when they collectively have less information, as it distorts labor less. The information precision depends on the history of shocks. For heterogeneous information, assume that the shocks and information structure are Gaussian. I have the following generalizations:

Theorem 4 (***Generalizing Theorem* 2**)**.** *The optimal industry revenue tax at state s^t is given by*

$$\boldsymbol{\tau}^{\boldsymbol{Ind}}_{\boldsymbol{t}} = -(\chi_Z \boldsymbol{D}^T \partial \boldsymbol{Z} + \chi_G \partial \boldsymbol{G})(\boldsymbol{I} - \boldsymbol{R})(\boldsymbol{I} - \frac{\varepsilon + \bar{\lambda}}{\varepsilon}\boldsymbol{\alpha}^{-1}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{R}))^{-1}\hat{\boldsymbol{\lambda}}_{\boldsymbol{t}}$$

where $\hat{\lambda}_{j,t} = \lambda_{j,t} - \bar{\lambda}_t$, $\bar{\lambda}_t = \frac{\sum_{j=1}^N \alpha_j D_j \lambda_{j,t}}{\sum_{j=1}^N \alpha_j D_j}$, and

$$\lambda_{j,t} = 1 - \prod_{m=0}^{t} \mathcal{K}_{j,m}$$

where $\mathcal{K}_{j,m}$ is the Kalman weight on the past forecast for forecaster in industry j at period m.

Proof for the dynamic model

Proof of proposition 5

Proof. **Necessity.** I characterize the equilibrium conditions to prove the necessity: I first solve the optimal behavior of the representative households, firms, and workers on different islands in different industries. The representative household chooses the $\{C_t(s^t), B_t(s^t), D_t(t, s^t)\}$ to maximize his expected utility. The Lagrangian for the household's problem is given by

$$
\begin{aligned}
L^{HH} = \sum_{t=0}^{\infty} \beta^t \int \Bigg\{ & \frac{C_t(s^t)^{1-\sigma}-1}{1-\sigma} - \sum_{i=1}^{N} \frac{1}{\varepsilon+1} \int_{j\in[0,1]} n_{ij,t}^{\varepsilon+1}(\omega_{ij}^t)dj \\
& - \mu^{HH}(s^t)\left[(1+\tau_t^c(s^t))P_t(s^t)C_t(s^t) + B_{t+1}(s^t) + \int_{s\in S^{t+1}} Q_{t+1,s}(s)D_{t+1,s}(s)\right] \\
& + \mu^{HH}(s^t)\Bigg[\int_{j\in[0,1]}[w_{ij}(\omega_{ij}^t)n_{ij}(\omega_{ij}^t) + \pi_{ij}(\omega_{ij}^t, s^t)]dj \\
& + (1+R_t(s^{t-1}))B_t(s^{t-1}) + D_{t,s^t}(s^t)\Bigg]\Bigg\} d\Psi(s^t)
\end{aligned}
$$

The F.O.Cs are given by

$$
U_{c,t}(s^t) - \mu^{HH}(s^t)\left(1+\tau_t^c(s^t)\right)P_t(s^t) = 0, \quad \forall s^t \tag{109}
$$

$$
-\mu^{HH}(s^t) + \beta\mathbb{E}_{s^{t+1}|s^t}\mu^{HH}(s^{t+1})\left(1+R_{t+1}(s^t)\right) = 0, \quad \forall s^t \tag{110}
$$

$$
-\Pr(s^{t+1}|s^t)Q_{t+1,s^{t+1}}(s^{t+1})\mu^{HH}(s^t) + \beta\mu^{HH}(s^{t+1}) = 0, \quad \forall s^t \tag{111}
$$

Combining (109) and (110) I derive the household's Euler equation

$$
\frac{U_c(s^t)}{(1+\tau_t^c(s^t))P_t(s^t)} = \beta\mathbb{E}\left[\frac{U_c(s^{t+1})}{(1+\tau_t^c(s^{t+1}))P_{t+1}(s^{t+1})}(1+R_{t+1}(s^t))\bigg| s^t\right] \tag{112}
$$

From (109) and (111), I find that the state-contingent price satisfies

$$
Q_{t+1,s^{t+1}} = \beta\Pr(s^{t+1}\big|s^t)\frac{U_c(s^{t+1})}{U_c(s^t)}\frac{(1+\tau_t^c(s^t))P_t(s^t)}{(1+\tau_{t+1}^c(s^{t+1}))P_{t+1}(s^{t+1})} \tag{113}
$$

Multiplying the household budget constraint at s^t by $\beta^t\frac{U_c(s^t)}{(1+\tau_t^c(s^t))P_t(s^t)}\Pr(s^t)$, and then integrating over s^t conditional on s^0:

$$\beta^t \int \Bigg[U_{c,t}(s^t)C_t(s^t) + \frac{U_{c,t}(s^t)}{(1+\tau_t^c(s^t))}\frac{B_{t+1}(s^t)}{P_t(s^t)}$$
$$+ \int_{s\in S^{t+1}} \frac{U_{c,t}(s^t)}{(1+\tau_t^c(s^t))}\frac{Q_{t+1,s}(s)D_{t+1,s}(s)}{P_t(s^t)}ds\Bigg]\, d\Psi(s^t|S^0)$$
$$= \beta^t \int \Bigg[\frac{U_{c,t}(s^t)}{(1+\tau_t^c(s^t))}\frac{1}{P_t(s^t)}\sum_{i=1}^{N}\int_{j\in[0,1]}[w_{ij}(\omega_{ij}^t)n_{ij}(\omega_{ij}^t) + \pi_{ij}(\omega_{ij}^t, s^t)]dj$$
$$+\frac{U_{c,t}(s^t)}{(1+\tau_t^c(s^t))}\frac{(1+R_t(s^{t-1}))B_t(s^{t-1})}{P_t(s^t)} + \frac{U_{c,t}(s^t)}{(1+\tau_t^c(s^t))}\frac{D_{t,s^t}(s^t)}{P_t(s^t)}\Bigg]\, d\Psi(s^t|s^0) \tag{114}$$

Summing the above equation over t and combining with (112)(113), I have

$$\sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0} C_t^{1-\sigma}(s^t) - \sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0}\sum_{i=1}^{N}\int_{j\in[0,1]}[w_{ij}(\omega_{ij}^t)n_{ij}(\omega_{ij}^t) + \pi_{ij}(\omega_{ij}^t, s^t)]dj$$
$$= \sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0} C_t^{1-\sigma}(s^t) - \sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0}\sum_{i=1}^{N}(1-\tau_{i,t}^{\text{Ind}}(s^t))p_{i,t}(s^t)z_{i,t}(s^t)\alpha_i L_{i,t}^{\alpha_i}(s^t)\prod_{k=1}^{N}X_{ik,t}^{a_{ik}}(s^t)$$
$$= \sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0} C_t^{1-\sigma}(s^t) - \sum_{t=0}^{\infty}\beta^t \mathbb{E}_{s^t|s^0}\sum_{i=1}^{N}\mathcal{T}_{i,t}(s^t)L_{i,t}(s^t)$$
$$= \frac{U_{c,0}(s^0)}{(1+\tau_0^c(s^0))}\frac{(1+R_0(s^{-1}))B_0(s^{-1})}{P_0(s^0)} = 0 \tag{115}$$

where I assume $B_0 = 0$ to ensure that period 0 is not treated as special. $\mathcal{T}_{i,t}(s^t)$ and $P_{i,t}(s^t)$ are defined by in our static model:

$$P_{i,t}(s^t) \equiv C_t(s^t)^{-\sigma}\frac{p_{i,t}(s)}{P_t(s)} \tag{116}$$

$$\mathcal{T}_{i,t}(s) \equiv C_t(s^t)^{-\sigma}\frac{p_{i,t}(s^t)}{P_t(s^t)}(1-\tau_{i,t}^{Ind}(s^t))z_{i,t}(s^t)\alpha_i L_{i,t}^{\alpha_i-1}(s)\prod_{k=1}^{N}X_{ik,t}^{a_{ik}}(s^t) \tag{117}$$

I have the implementabiliy constraint (115). For the production side, I solve the equilibrium conditions backwardly. The equations are the similar to those in the static model. Then I prove proposition 5: .

Sufficiency. Take any allocation $\bar{\xi}$ that satisfies (115)(105)(106)(107). I now prove that there exists a set of tax rates

$$\left\{\tau_{it}^{C}(s^t), \tau_{it}^{Ind}(s^t)\right\},$$

a local wage $w_{ij,t}(\omega_{ij}^t)$, relative prices $\{p_{i,t}(s^t)\}_{i\in\{1,...,N\}}$, an interest rate function $R_{t+1}(s^t)$, and a path for nominal debt holdings $B_{t+1}(s^t)$ and assets $D_{t,s^{t+1}}(s^{t+1})$ that implement this allocation as an equilibrium. I construct the equilibrium prices and policies as follows.

By normalizing the price of final goods $P_t(s^t)$ to be 1, the relative prices are given by

$$\frac{p_{i,t}(s^t)}{P_t(s^t)} = C_t(s^t)^{\sigma} P_{i,t}(s^t)$$

The local wage rate is

$$w_{ij,t}(\omega_{ij}^t) = \frac{n_{ij,t}^{\varepsilon}(\omega_{ij}^t)}{E_{s'|\omega_{ij}^t} C_t(s')^{-\sigma} \frac{1}{P_t(s')}}$$

The state-contingent taxes are derived by using (116) (117):

$$(1+\tau_{i,t}^{C}(s^t)) = \frac{1}{\psi_{i,t}^{C}(s^t)}; \quad (1+\tau_{i,t}^{Ind}(s^t)) = \frac{\psi_{i,t}^{Ind}(s^t)}{\psi_{i,t}^{C}(s^t)}$$

The f.o.c of intermediate goods are satisfied by using (106). The f.o.c of workers are satisfied by the way I determine the wage. The f.o.c of the firms for labor are satisfied by using (107). The resource constraint is ensured by (105).

The interest rate functions are determined by using the Euler condition:

$$(1+R_{t+1}(s^t)) = \frac{\frac{U_c(s^t)}{(1+\tau_t^c(s^t))P_t(s^t)}}{\beta \mathbb{E}\left[\frac{U_c(s^{t+1})}{(1+\tau_t^c(s^{t+1}))P_{t+1}(s^{t+1})}\middle| s^t\right]}$$

The price of Arrow security is given by (113). By these prices, I know the f.o.cs of the household are satisfied. The last thing to check is the budget constraints of the representative household. Multiplying the household budget constraint by $\beta^t \frac{U_c(s^t)}{(1+\tau_t^c(s^t))P_t(s^t)} \Pr(s^t|s^m)$, integrating over s^t and summing the above equation over t. Combining with the expressions of asset prices, the government bond holdings are given by

$$B_m(s^{m-1}) = \frac{U_{c,m-1}(s^{m-1})}{(1+\tau^C_{m-1}(s^{m-1}))} \frac{1}{P_{m-1}(s^{m-1})} \times$$
$$\left[\sum_{t=m}^{\infty} \beta^t \mathbb{E}_{s^t|s^{m-1}} C_t^{1-\sigma}(s^t) - \sum_{t=m}^{\infty} \beta^t \mathbb{E}_{s^t|s^{m-1}} \sum_{i=1}^{N} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t)\right] \quad \forall m \geq 1, \forall s^m$$

The holdings of state-contingent assets are given by

$$D_{m,s^m}(s^m) = \frac{P_m(s^m)}{U_{c,m}(s^m)(1+\tau^C_m(s^m))} \left[\sum_{t=m}^{\infty} \beta^t \mathbb{E}_{s^t|s^m} C_t^{1-\sigma}(s^t) - \sum_{t=m}^{\infty} \beta^t \mathbb{E}_{s^t|s^m} \sum_{i=1}^{N} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t)\right]$$
$$-(1+R^C_m(s^{m-1}))B_m(s^{m-1}) \quad \forall m \geq 1, \forall s^m$$

Then, it's straightforward to check that the household's budget constraint is satisfied for $t \geq 1$. And using (104), the budget constraint also holds at time 0. Thus, I have completed the proof of this proposition.

□

Proof of theorem 3 and theorem 4

Proof. For the Ramsey problem, by combining (107)(108), I have

$$\mathbb{E}_{\omega^t_{ik}|s^t} \left[\mathbb{E}_{s'|\omega_{ik}} \mathcal{T}_{i,t}(s')\right]^{\frac{1}{\varepsilon}} = L_{i,t}(s^t) \tag{118}$$

Using the primal approach, the Ramsey planner chooses

$$\{C_{i,t}(\omega^t_{ij}), L_{i,t}(\omega^t_{ij}), T_{i,t}(\omega^t_{ij}), T_{i,t}(\omega^t_{ij}), X_{ij,t}(\omega^t_{ij})\} \tag{119}$$

which are measurable to s^t to maximize the expected utility

$$\sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t} \left\{ \frac{C_t^{1-\sigma}(s^t) - 1}{1-\sigma} - \frac{1}{1+\varepsilon} \sum_{i=1}^{N} \mathbb{E}_{\omega^t_{ik}|s^t} \left[\mathbb{E}_{s'|\omega_{ik}} \mathcal{T}_{i,t}(s')\right]^{\frac{1+\varepsilon}{\varepsilon}} \right\} \tag{120}$$

subject to the constraints (104)(105)(106)(118). The Lagrange for the Ramsey problem is

$$\mathcal{L}^{\text{Dynamic}} = \sum_{t=0}^{\infty} \beta^t \int \left\{ \frac{C_t^{1-\sigma}(s^t) - 1}{1-\sigma} - \frac{1}{1+\varepsilon} \sum_{i=1}^{N} \mathbb{E}_{\omega_{ik}^t | s^t} \left[\mathbb{E}_{s' | \omega_{ik}} \mathcal{T}_{i,t}(s') \right]^{\frac{1+\varepsilon}{\varepsilon}} \right.$$

$$+\mu_{R,t}(s^t)[\prod_{i=1}^{N} \left(\frac{z_{i,t}(s^t) L_{i,t}^{\alpha_i}(s^t) \prod_{k=1}^{N} X_{ik,t}^{a_{ik}}(s^t) - \sum_{k=1}^{N} X_{ki,t}(s^t)}{\beta_i} \right)^{\beta_i} - G_t(s^t) - C_t(s^t)]$$

$$+ \sum_{i=1}^{N} \mu_{L,t}(s^t) [\mathbb{E}_{\omega_{ik}^t | s^t} \left[\mathbb{E}_{s' | \omega_{ik}^t} \mathcal{T}_{i,t}(s') \right]^{\frac{1}{\varepsilon}} - L_{i,t}(s^t)]$$

$$\left. + \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{ij,t}(s^t) [\frac{a_{ij}}{\alpha_i} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t) - P_{j,t}(s^t) X_{ij,t}(s^t)] \right\} d\Psi(s^t)$$

$$+ \int \mu_{G,0}(s^0) \left\{ \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t | s^0} C_t^{1-\sigma}(s^t) - \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t | s^0} \sum_{i=1}^{N} \mathcal{T}_{i,t}(s^t) L_{i,t}(s^t) \right\} d\Psi(s^0)$$

I first solve this relaxed Ramsey problem and then verify this solution is within the feasible set of equilibrium for the original Ramsey problem. The f.o.cs of the problem are similar to those for the static model if I replace the state s in those functions by state s^t, except that for the implementability constraint, the multiplier is not state-contingent on s^t, but it is also related to the period 0 shock s^0. The effects of shocks after period 0 can be smoothed by the state-contingent assets $D_{t,s}$. I assume that $\mu_{ij}(s^t) = 0$, $\mathcal{T}_{i,t}(s^t) = (\bar{D}_i \alpha_i)^{\frac{\varepsilon}{\varepsilon+1}} \mathcal{T}_t(s^t)$, $\mu_{i,t}^L(s) = (\bar{D}_i \alpha_i)^{\frac{\varepsilon}{\varepsilon+1}} \mu_t^L(s^t)$. The proof strategy is similar to the static model. The implementability condition is different from the static model:

$$\sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t | s^0} (Y(s^t) - G(s^t))^{1-\sigma} - \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s^t | s^0} \mathcal{T}_t(s^t) L(s^t; \mathcal{T}_t) = 0$$

The equation (58) hols if I replace s by s^t and ω_{ij} by ω_{ij}^t. And (53) becomes

$$((1-\sigma)\mu_G(s_0) + 1) Y_t(s^t) C_t(s^t)^{-\sigma} = \mu_{L,t}(s^t) L_t(s^t; \mathcal{T}_t) + \mu_{G,t}(s^t) \mathcal{T}_t(s^t) L_t(s^t; \mathcal{T}_t)$$

Again, I can compute $L_{i,t}(s^t), X_{ij,t}(s^t), Y_t(s^t), C_t(s^t)$ from the guess and equilibrium conditions once I know $\mathcal{T}_t(s^t)$. Thus, I have three sets of equations shown above and three sets of unknows $\{\mathcal{T}_t(s^t), \mu_{L,t}(s^t), \mu_{G,t}(s_0)\}$. I got the equilibrium by solving the functional equations and combining the guess and how I constructed the solution. In this equilibrium, it's easy to check that the optimal revenue taxes $\tau_{j,t}^{Ind}(s^t))$ are zeros and the optimal consumption taxes $\tau_{j,t}^{C}(s^t)$ are equalized as in the static model.

When the information structure is heterogeneous, I refer to the perturbation approach. The difference is that, the counterpart for λ_j from the static model to the dynamic setting is

$$\lambda_j \equiv \mathbb{E}_{s^t}\left[\mathbb{E}_{\omega_{ij}^t}[s_0]\right]$$

With the state-contingent asset, the effect of shocks after 0 on the government budget constraint will be perfectly smoothed. Thus, what matters is only the period 0 shock. But the distribution beliefs of s_0 update and is associated with the underlying state s^t.

The forecasters i in industry j update his forecast of the s_0 rate using a Kalman filter, as follows [25]:

$$\begin{aligned} E_{i,j,t}[s_0] = &\left(1 - \sum_{n=1}^{M} \lambda_{n,j,t}^{\text{pub}} - \sum_{n=1}^{K} \lambda_{n,j,t}^{\text{private}}\right) E_{i,j,t-1}[s_0] \\ &+ \sum_{n=1}^{M} \lambda_{n,j,t}^{\text{pub}} \hat{x}_{n,t}^{\text{pub}} + \sum_{n=1}^{K} \lambda_{n,j,t}^{\text{private}} \hat{x}_{i,j,n,t}^{\text{private}} \end{aligned}$$

where $\lambda_{n,j,t}^{\text{pub}}$ and $\lambda_{n,j,t}^{\text{private}}$ are the Kalman gains for the n-th public signal $\hat{x}_{n,t}^{\text{pub}}$ and n-th private signal $\hat{x}_{i,j,n,t}^{\text{private}}$. Taking the expectation over ω_{ij}^t:

$$\begin{aligned} \mathbb{E}_{s^t}\left[\mathbb{E}_{i,t}[s_0]\right] = &\left(1 - \sum_{n=1}^{M} \lambda_{n,j,t}^{\text{pub}} - \sum_{n=1}^{K} \lambda_{n,j,t}^{\text{private}}\right) \mathbb{E}_{s^{t-1}}\left[\mathbb{E}_{i,t-1}[s_0]\right] \\ &+ \sum_{n=1}^{M} \lambda_{n,j,t}^{\text{pub}} s_0 + \sum_{n=1}^{K} \lambda_{n,j,t}^{\text{private}} s_0 \\ = &\,\mathcal{K}_{j,t}\mathbb{E}_{s^{t-1}}\left[\mathbb{E}_{i,t-1}[s_0]\right] + (1 - \mathcal{K}_{j,t})s_0 \\ = &\,\mathcal{K}_{j,t}(1 - \prod_{m=0}^{t-1} \mathcal{K}_{j,m})s_0 + (1 - \mathcal{K}_{j,t})s_0 = (1 - \prod_{m=0}^{t} \mathcal{K}_{j,m})s_0 \end{aligned}$$

The theorem can be proved using mathematical induction. Regardless of the number of shocks or whether the types of signals are known, this theorem can be easily applied by running a regression on the forecast data to obtain the Kalman weight on the previous forecast, $\mathcal{K}_{j,t}$.

□

[25] The signal received before $t = 0$ has no useful information about the unexpected shock s_0