Planning Against Disasters in Dynamic Production Networks

Vasco M. Carvalho, Matias Covarrubias and Galo Nuño*

March 20, 2025

Abstract

In dynamic multisector economies, the planner's optimal capital allocation can serve to minimize the aggregate impact of shocks cascading through nonlinear production networks. We show analytically that (*i*) optimal capital allocation under uncertainty involves deliberately over-investing in upstream sectors in order to mitigate severe economic downturns; (*ii*) this efficient strategy reduces the average level of consumption and gives rise to a high welfare cost of business cycles. Deploying novel deep-learning techniques in a general environment we show, quantitatively that: (*iii*) the ergodic distribution of the simulated nonlinear economy features higher mean capital levels in key upstream sectors, lower mean levels of macroeconomic aggregates, realistic aggregate volatility and a welfare cost of business cycles one order of magnitude larger than in standard linear models.

*Vasco M. Carvalho, University of Cambridge and CEPR, Matias Covarrubias, Bank of Spain, Galo Nuño, Bank of International Settlements and Bank of Spain. This manuscript supersedes a previous version entitled 'Nets on Nets: Amplification and Nonlinearities in Dynamic Production Networks.' The paper does not necessarily reflect the views of the Bank of Spain or Eurosystem. The authors are grateful to Isaac Baley, Luis Garicano, Basile Grassi, Ben Moll, Mathias Trabandt, Simon Scheidegger, Jaume Ventura, and Christian von Lehm for useful comments; and to Charles Parry, Jesus Villota, and Bowen Wang for excellent research assistance.

1 Introduction

Recent research on production networks has demonstrated that nonlinearities and complementarities between sectors can dramatically amplify the impact of shocks. When inputs are complementary — as is often the case with specialized components, rare materials, or energy inputs — disruptions in upstream sectors can cascade throughout the economy with outsized effects. This powerful intuition sheds light on classical issues in macroeconomics and holds the promise of better understanding, for example, the nonlinear aggregate effects of oil shocks or the origins of aggregate disaster risk.

In this paper, we confront the benevolent planner of textbook dynamic neoclassical economies with this novel logic of nonlinear production networks. We ask whether, in efficient economies, the planner's allocation of *dynamic* inputs, such as capital, can help mitigate the tail risk that has been shown to emerge in *static* networked environments. We do so both analytically, in a simplified environment, and quantitatively, in a general state-of-the-art dynamic multi-sector setup. We start by showing, analytically, that optimal capital allocation under uncertainty involves deliberately – and efficiently – over-investing in upstream sectors in order to minimize the risk of catastrophic downstream transmission and aggregate disruptions. While it significantly mitigates severe economic downturns, we also show that this optimal pre-allocation strategy of capital reduces the average level of consumption. The latter, in turn, gives rise to a large welfare cost of business cycles such that the optimal avoidance of disaster risk by a neoclassical planner leads to significant lifetime sacrifices in consumption levels. In the second part of the paper, we offer a quantitative benchmark. We show how to adapt and deploy frontier deep neural network tools that can capture the full nonlinear effects in a canonical dynamic production networks setup featuring both intermediate input and investment good linkages across sectors. We find that, relative to the deterministic steady state, the ergodic distribution of the simulated nonlinear economy features higher mean capital levels in key upstream sectors, lower mean levels of key macroeconomic aggregates, realistic aggregate volatility and, importantly, a high welfare cost of business cycles.

Our first contribution is to characterize, analytically, optimal capital allocation under uncertainty in a tractable two-period, two-sector networked economy, featuring an upstream intermediate input-supplying sector and a downstream sector, serving final demand consumption by the representative household and reliant on the upstream input to do so. Crucially, both sectors also need capital in order to produce. Unlike most of the extant literature – which has focused on characterizing static economies – we are interested in characterizing the planner's allocation of pre-determined dynamic inputs, such as capital, across sectors and how this de-

pends on nonlinearities and complementarities in the within-period problem that obtains once capital allocation choices are made. A key focus is on what we call "pre-allocation" of capital before productivity shocks are realized, i.e. how the allocation of a pre-determined input to a particular sector compares to what would be optimal in a *deterministic* environment.

Our key analytical result shows that whenever the downstream production function features complementarity across inputs and risk aversion is not too low, the planner's allocation results in excess pre-allocation of capital to the upstream, input-supplying sector (relative to the deterministic optimum). This is because productivity shocks upstream have a nonlinear effect on aggregate output rendering possible aggregate consumption disasters for sufficiently low realizations of shocks. This, in turn, renders insurance against negative upstream shocks more valuable to the planner, who deploys capital to the input supplying sector for this purpose. Our second analytical result shows that, indeed, the elasticity of aggregate consumption with respect to negative upstream shocks is strictly smaller when the planner pursues this excess pre-allocation strategy, relative to what obtains at deterministic optimum allocations.

Our third key analytical result shows that complementarities also imply that pre-allocating capital upstream is costly in terms of lower average consumption level at the stochastic steady state (again, relative to its deterministic counterpart).[1] This is because, holding the productivity of the downstream sector fixed, the planner is: (i) allocating excess capital upstream, to be used precisely in states of the world where that sector is particularly unproductive, and (ii) allocating too little capital downstream, in states of the world where the return to investing in downstream capacity would be relatively high. Thus, in mitigating nonlinear consumption disasters, the planner faces a fundamental tradeoff in terms of the level of aggregate consumption. This result implies that the welfare cost of business cycle does not come from observable volatility of consumption or consumption disasters. As discussed, the planner is able to avert these through manipulating input allocations. Instead, the welfare cost of business cycle is associated with a reduction in average consumption.

In the second part of the paper, we present a quantitative benchmark by deploying tools that can capture the full nonlinear effects of dynamic, nonlinear, production networks economies. Our quantitative environment is canonical, allowing for flexible substitution possibilities in consumption, investment and production and flexible adjustment cost specifications for both primary inputs, labor and capital. In detail, our baseline environment features a representative household with GHH preferences over consumption and labor. In turn, each of these

[1]The stochastic steady state is the steady state of an uncertain economy where all shocks happen to be realized at their (zero) mean. For nonlinear economies, such as ours, this will in general differ from the steady state that obtains in the corresponding deterministic economy.

is a CES aggregate of, respectively, disaggregated final consumption goods – allowing for flexible substitution possibilities in household consumption – and labor employed across the different sectors in the economy, allowing us to control the degree of flexibility of labor reallocation across sectors. Each of the sectors is populated by a representative firm that combines primary inputs – labor and capital – and intermediate inputs in order to produce sectoral gross output. The latter, in turn, can be sold as final consumption, as an intermediate input used in other sectors or as an investment good, also sold to other sectors. We again allow for successive CES nests, allowing for substitution/complementarity over the two primary inputs, across the intermediate inputs necessary for production and across primary and intermediate inputs. Sector-specific capital follows standard law of motions and allows for sector-specific adjustment costs. Sector-specific investment combines investment goods produced by other sectors, again in a CES fashion. All uncertainty derives from sector-specific persistent productivity processes and we assume throughout a perfectly competitive market structure across all markets. We render this complex disaggregated environment quantitative by calibrating it to data moments for 37 sectors in the US, using data spanning 1948 to 2018. We match steady-state expenditure shares in the model against time-averaged data counterparts, as well as feature empirically relevant input-output and investment linkages across sectors, while also taking into account empirical sector-specific depreciation rates for capital. Whenever possible, we set elasticities to consensus parameterizations already present in the literature.

Solving for the full non-linear stochastic equilibrium of this economy poses a major challenge: even for our relatively low levels of disaggregation, the full global solution implies that finding an equilibrium requires solving a stochastic dynamic programming problem with 74 state variables. Our second key contribution in this paper is to extend the recent 'deep equilibrium nets' method of Azinovic et al. (2022) to our particular quantitative environment. The main advantage of deploying deep neural networks as function approximators is that they are less prone to the curse of dimensionality than other approximators such as polynomials, or fourier series, thus making it feasible to resolve our high-dimensional dynamic nonlinear production network environment globally. In both the main body of the paper and the Appendix, we discuss how to adapt and implement in detail this deep neural network methodology in our environment alongside accuracy checks that provide reassurance that the solutions are accurate.

With this method in hand, our third main contribution is to document quantitative features of the global solution of our baseline economy. Specifically we can both compute the ergodic distribution by simulation, capturing how the (potentially) nonlinear propagation of Gaus-

sian shocks interacts with the anticipation effects of risk on agents and provide generalized impulse responses at the stochastic steady state. We report five key quantitative results. Our first main finding is that the mean of the ergodic distribution of key aggregates – consumption, labor, GDP, investment and total intermediates produced – is always below that of the deterministic steady state. Second, we find that the implied aggregate volatility is substantial – e.g. the standard deviation of annual GDP is 3.4%. Third, our preliminary results imply that, despite the above successes, nonlinear production network economies do not display quantitatively important higher-order moments: in the ergodic distribution, kurtosis is negligible, skewness is small throughout and, depending on the macro aggregate under consideration, can have the wrong sign relative to data. Fourth, there is substantial heterogeneity across sectors in the ergodic mean allocation, with key upstream sectors – like Mining, Oil and Gas – displaying relatively higher capital stocks whereas downstream sectors display relatively lower capital stocks. Fifth, we also show how, in counterfactual economies with higher sectoral TFP volatility, consumption and labor volatilities increase less than one-to-one. That is, as the risks of nonlinear disasters following large shocks become material, the planner spends more resources in keeping aggregate volatility from spiraling upwards.

Taken together, these results echo our analytical findings in the first part of the paper and suggest that the planner in this dynamic, fully nonlinear economy, is redistributing capital towards upstream sectors and away from downstream sectors in order to dampen large aggregate fluctuations originating in central upstream sectors. By doing this, the planner avoids the endogenous disasters stressed elsewhere in the production networks literature, at the cost of lower mean levels of aggregates. We show how the nonlinear impulse responses to sectoral TFP shocks are attenuated in comparison to the log-linear ones.[2]

Finally, we find that the welfare costs of business cycles are substantial relative to standard calculations in the literature: our global solution implies that the representative household would do away with 1% of lifetime consumption in order to live in a counterfactual economy without shocks. Combining the results above, this large welfare cost of business cycles is largely the direct result of the mean of the ergodic distribution being below that of the deterministic economy, rather than aggregate endogenous disaster risk, which is absent.

Related Literature Our paper relates to three different strands of literature. The first is the rapidly growing literature on production networks (see Acemoglu et al. (2012) for the seminar reference and Carvalho and Tahbaz-Salehi (2019) and Baqaee and Rubbo (2023) for

[2]This contrasts with the nonlinear response to an unanticipated zero-probability shock, which is amplified with respect to the log-linear case, in line with the findings of Baqaee and Farhi (2019).

recent overviews of this literature. In particular, our work relates to Baqaee and Farhi (2019) analysis of nonlinear, higher order effects in production networks. In particular, Baqaee and Farhi (2019) show how complementarity and intermediate input linkages together can generate endogenous skewness and aggregate disasters, depending on the size and location - in the production network - of micro-shocks. Relative to Baqaee and Farhi (2019) and the literature that followed – emphasizing nonlinearities in (static) production networks – our contribution is to show that in dynamic environments, the efficient allocation of predetermined inputs alters business cycle properties of these models.[3] On the one hand, we show how the pre-allocation of capital limits the extent to which aggregate nonlinearities – in the form of skewness and aggregate disasters – can emerge as important features in this class of models. On the other hand, our analysis shows that this aggregate disaster possibility and the ensuing optimal preallocation of capital yields a high welfare cost of business cycles. As a second contribution to this literature, this paper is the first to offer a quantitative framework that enables a systematic evaluation of nonlinear networks. This is because all extant quantitative characterizations either provide quantitative illustrations in *static* economies (e.g. Baqaee and Farhi (2019)) – where the absence of endogenous dynamics renders it difficult to benchmark results against competing dynamic general equilibrium environments – or rely on approximation techniques (e.g. Horvath (2000), Atalay (2017) and Vom Lehn and Winberry (2022) [4]) – such as log-linearized solutions – which, by definition, are ill-suited to fully capture the effects of the non-linear equilibrium dynamics that characterize these environments.

Second, our paper relates to the long-standing literature on the welfare cost of business cycles in dynamic equilibrium economies. As is well known, starting from the seminal contribution of Lucas (1987), simple textbook environments are typically thought to be unable to deliver high welfare costs of uncertainty. This realization, in turn, has motivated a long list of influential research on, for example, the role of preferences, market incompleteness, and various frictional environments or risk processes. Our main contribution is to overturn the baseline assertion in the literature: we show that a 'vanilla' real business cycle model – with no distortions, frictions, exotic preferences, or unconventional exogenous driving processes – can deliver a high welfare cost of business cycles. Further, we show that the mechanism through which the model attains this high welfare cost is related to - but distinct from - recent resolutions of

[3] Our emphasis on how predetermined inputs interact and alter other standard features of production networks is shared with Kopytov et al. (2024), who show that, with endogenous network formation but in an otherwise linear and static Cobb-Douglas environment, pre-commitment to risky suppliers may lead firms to optimally choose less risky but less productive suppliers. See also Pellet and Tahbaz-Salehi (2023) for a similar mechanism.

[4] Carvalho (2007), Foerster et al. (2011) and Liu and Tsyvinski (2024) also study dynamic environments but focus on Cobb-Douglas, linear economies.

the 'welfare cost puzzle' relying on disasters and higher order properties of consumption risk (e.g. Barro (2009), Martin (2008) or Jorda et al. (2024)). In particular, our novel mechanism relies on (efficient) capital reallocation across sectors – and the ensuing decline in average output and consumption *levels* – in order to *avoid* disaster risk, rather than on how large disaster realizations change welfare calculations.

Third, our paper relates to a recent literature on the use of deep learning to solve high-dimensional general equilibrium models (e.g., Maliar et al., 2021; Han et al., 2021; Gu et al., 2024).[5] Our paper extends the "deep equilibrium nets"methodology by Azinovic et al. (2022) to the case of a planning problem in a production network.

The rest of the paper proceeds as follows. In Section 2, we provide a simple two-sector model that allows us to analytically characterize capital allocation under uncertainty in a multi-sector networked economy. In Section 3, we exhibit the model and the system of equations that describe the solution. In Section 4, we discuss the calibration and the numerical method. Finally, in section 5 we detail the results for the global solution of the model.

2 A simple two-sector model

In this section, we analyze a stylized two-sector model that allows us to analytically characterize capital allocation under uncertainty in a multi-sector networked economy. A key focus is on what we call "preallocation" of capital, that is, the excess allocation of capital to a particular sector compared to what would be optimal in a *deterministic* environment. We use this term because capital must be allocated before productivity shocks are realized, and uncertainty leads to systematically higher capital shares for the benefited sector. This theoretical framework provides deeper insights into the main mechanisms behind three key results we observed in our quantitative model: preallocation of capital toward upstream sectors, mitigation of consumption disasters, and welfare costs of business cycles through lower average consumption. By focusing on a tractable two-sector setting, we can derive closed-form solutions illuminating the economic forces at play when the planner faces uncertainty about sectoral productivity. Appendix A presents all detailed algebraic derivations and proofs.

2.1 Setting

There are two sectors. Sector 1 ("upstream") produces the intermediate good, and sector 2 ("downstream") produces the final good. The intermediate good is produced with capital,

[5] See Fernandez-Villaverde et al. (2024) for a recent review.

and is subject to a productivity shock. The final good is produced using capital and the intermediate good and is not subject to a productivity shock. The planner has one unit of a factor K that needs to be allocated to the two sectors K_1 and K_2. The key aspect of this problem is that capital is allocated before the TFP shock is realized, creating a decision under uncertainty. After the shock is realized, there are no remaining decisions to be made since goods produced and consumed are determined by the preallocated capital.

Let Q_i, K_i represent gross-output and capital in sector $i = \{1, 2\}$, and C represents consumption of the final good. The problem is defined by the following equations:

$$K_1 + K_2 = 1, \qquad \text{(Resource constraint)} \qquad (1)$$

$$Q_1 = AK_1, \qquad \text{(Upstream production)} \qquad (2)$$

$$Q_2 = \left((1-\gamma_q) (Q_1)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q (K_2)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1}}, \qquad \text{(Downstream production)} \qquad (3)$$

$$C = Q_2, \qquad \text{(Consumption)} \qquad (4)$$

where A is a random variable, which takes a low value $A_L = 1 - \Delta_A < 1$ with probability p and a high value $A_H = 1 + \Delta_A > 1$ with probability $1 - p$ (notice that $\mathbb{E}[A] = 1$), and σ_q is the elasticity of substitution in production between the two goods. A representative household maximizes the expected utility:

$$\mathbb{E}[U(C)] = \mathbb{E}\left[\frac{C^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} \right] \qquad (5)$$

where the expectation is taken with respect to the random variable A and ϵ_c is the intertemporal elasticity of substitution (so ϵ_c^{-1} is risk aversion).

2.2 First-order conditions

We consider the first-best allocation produced by a benevolent social planner. This coincides with the allocation in the competitive equilibrium, assuming that firms in both sectors operate under perfect competition.

Using the resource constraint (1), we can substitute $K_2 = 1 - K_1$ to obtain an unconstrained optimization problem with a single control variable K_1. This allows us to write consumption in each state (low L and high H) as:

$$C_S = \left((1-\gamma_q) \left(A_s K_1\right)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q \left(1 - K_1\right)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1}}, \quad S \in \{L, H\} \tag{6}$$

Now we can write the objective function in terms of these state specific consumption levels:

$$\mathbb{E}[U(C)] = p \frac{C_L^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} + (1-p) \frac{C_H^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} \tag{7}$$

To find the optimal allocation, we first need the derivative of consumption with respect to upstream capital K_1. This shows how shifting capital between sectors affects consumption in each state:

$$\frac{\partial C_s}{\partial K_1} = \left((1-\gamma_q) \left(\frac{A_S K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} A_S - \gamma_q \left(\frac{1-K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} \right), \quad S \in \{L, H\} \tag{8}$$

The only control in this problem is the upstream capital k_1. The first-order condition (FOC) implies that the derivative of the expected utility with respect to K_1 is zero:

$$\begin{aligned} \frac{\partial \mathbb{E}[U(C)]}{\partial K_1} = p \frac{1}{C_L^{\epsilon_c^{-1}}} & \left((1-\gamma_q) \left(\frac{A_L K_1}{C_L} \right)^{\frac{-1}{\sigma_q}} A_L - \gamma_q \left(\frac{1-K_1}{C_L} \right)^{\frac{-1}{\sigma_q}} \right) \\ + (1-p) \frac{1}{C_H^{\epsilon_c^{-1}}} & \left((1-\gamma_q) \left(\frac{A_H K_1}{C_H} \right)^{\frac{-1}{\sigma_q}} A_H - \gamma_q \left(\frac{1-K_1}{C_H} \right)^{\frac{-1}{\sigma_q}} \right) = 0 \end{aligned} \tag{9}$$

This first-order condition is central to our analysis of preallocation. Each term represents the marginal impact of allocating capital to the upstream sector in different productivity states, weighted by the marginal utility in that state. We will analyze the economic forces behind

each component of this expression in Section 2.4.

2.3 Deterministic optimum

We analyze the deterministic problem, that is, the problem without aggregate risk. To this end, we set $A = 1$ (the expected value of the shock). Then, the first-order condition is:

$$\left((1-\gamma_q)\left(\frac{K_1}{C_S}\right)^{\frac{-1}{\sigma_q}} - \gamma_q\left(\frac{1-K_1}{C_S}\right)^{\frac{-1}{\sigma_q}}\right) = 0, \tag{10}$$

where we use the subindex S as both states coincide in this case.

Solving for K_1^{determ} (the optimal capital allocation in the deterministic case):

$$K_1^{determ} = \frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}} \tag{11}$$

This expression shows that the optimal deterministic allocation depends only on the production share parameters γ_q and the elasticity of substitution σ_q. Intuitively, when γ_q is higher, that is when downstream capital is more important in production, less capital is allocated to the upstream sector. The effect of σ_q depends on the relative magnitudes of γ_q and $(1-\gamma_q)$. In the special case where $\gamma_q = 1/2$, meaning both sectors are equally important in production, we have $K_1^{determ} = 1/2$ regardless of the value of σ_q.

2.4 Incentives to preallocate upstream

Although at the optimum $\partial\mathbb{E}[U(C)]/\partial K_1 = 0$, we can use the marginal expected utility to compare optimal capital allocation under uncertainty versus the deterministic case. To determine whether there is preallocation toward the upstream sector (i.e., whether the optimal capital under uncertainty K_1^* satisfies $K_1^* > K_1^{determ}$), we can evaluate the marginal expected utility at the deterministic solution. If this value is positive, it means the planner would benefit from increasing K_1 beyond the deterministic optimum when facing uncertainty. Figure 1 illustrates this approach by showing the marginal expected utility curve (solid blue line) as a function of K_1, calculated assuming $\sigma_q = 0.25$, $\gamma_q = 0.5$, risk aversion $\epsilon_c^{-1} = 1$, and $\Delta_A = 0.5$. The intersection of this curve with the horizontal axis determines the stochastic optimum K_1^*, while the deterministic optimum K_1^{determ} is shown by the vertical dotted-dashed red line. As evident from the figure, when the marginal expected utility is positive at K_1^{determ}, the stochastic optimum involves higher allocation to the upstream sector.

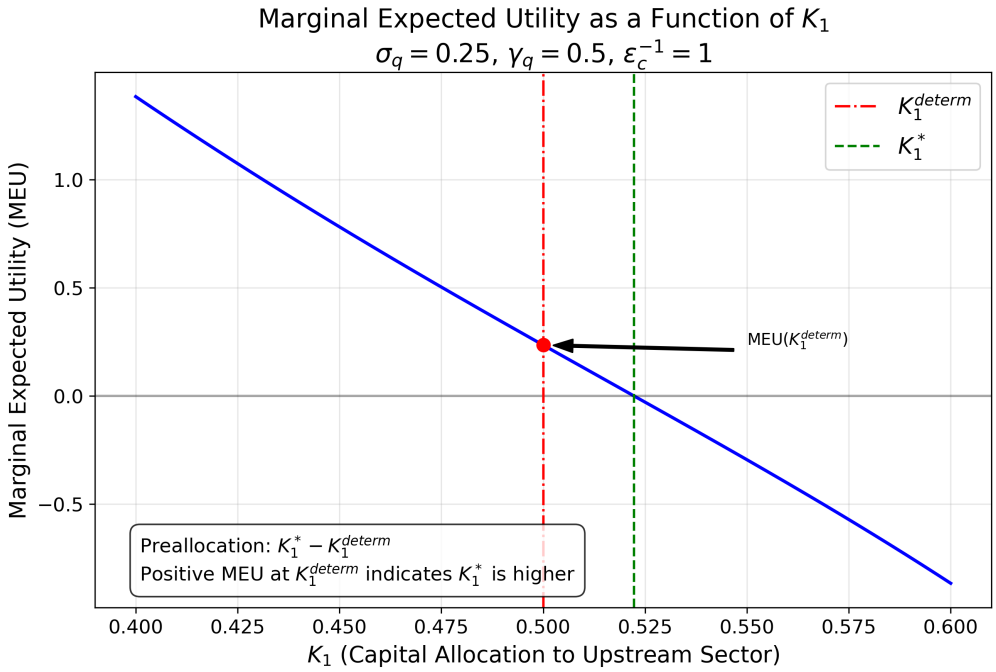


Figure 1: Marginal expected utility (MEU) as a function of upstream capital allocation. This figure illustrates our proof approach: when MEU evaluated at the deterministic optimum K_1^{determ} is positive, the stochastic optimum K_1^* must be higher, resulting in preallocation to the upstream sector. The parameters used are $\sigma_q = 0.5$, $\gamma_q = 0.5$, risk aversion $\epsilon_c^{-1} = 1$, and $\Delta_A = 0.5$. The intersection of the marginal expected utility curve with the horizontal axis determines the stochastic optimum.

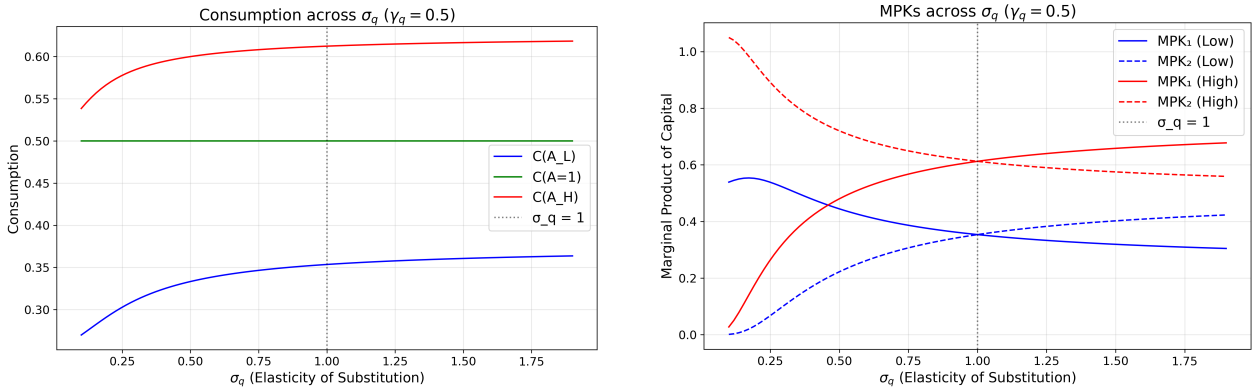
Before analyzing preallocation analytically, we first dissect the terms in the expression for marginal expected utility, to get a better intuition of the economic forces at play. The expression for marginal expected utility can be written as:

$$\left.\frac{\partial \mathbb{E}[U(C)]}{\partial K_1}\right|_{K_1^{determ}} = p \underbrace{\frac{1}{C_L^{\epsilon_c^{-1}}}}_{\text{Marginal Utility L}} \left(\underbrace{(1-\gamma_q)\left(\frac{A_L K_1^{determ}}{C_L}\right)^{\frac{-1}{\sigma_q}} A_L}_{\text{MPK}_{1,L}} - \underbrace{\gamma_q \left(\frac{1-K_1^{determ}}{C_L}\right)^{\frac{-1}{\sigma_q}}}_{\text{MPK}_{2,L}} \right)$$

$$+ (1-p) \underbrace{\frac{1}{C_H^{\epsilon_c^{-1}}}}_{\text{Marginal utility H}} \left(\underbrace{(1-\gamma_q)\left(\frac{A_H K_1^{determ}}{C_H}\right)^{\frac{-1}{\sigma_q}} A_H}_{\text{MPK}_{1,H}} - \underbrace{\gamma_q \left(\frac{1-K_1^{determ}}{C_H}\right)^{\frac{-1}{\sigma_q}}}_{\text{MPK}_{2,H}} \right)$$

This expression breaks down the marginal impact of allocating more capital to the upstream

sector. The first term in each state, $\text{MPK}_{1,S}$, represents the marginal product of capital in the upstream sector, which contributes positively to consumption. The second term, $\text{MPK}_{2,S}$, represents the opportunity cost in terms of foregone production in the downstream sector. The total marginal product in each state is the difference between these two effects. To understand how these components vary with the elasticity of substitution, we examine consumption and marginal products across different values of σ_q. Figure 2 presents these relationships.



(a) Consumption levels across different elasticities of substitution

(b) Marginal products of capital across different elasticities of substitution

Figure 2: Analysis of consumption and marginal products across elasticities of substitution (σ_q). Panel (a) shows consumption under low productivity ($A_L = 0.5$, blue line), normal productivity ($A = 1$, green line), and high productivity ($A_H = 1.5$, red line) states. Panel (b) shows the marginal product of capital in the upstream sector (MPK_1, solid lines) and downstream sector (MPK_2, dashed lines) under both low productivity (blue) and high productivity (red) states. All values are evaluated at the deterministic optimal capital allocation. The vertical dotted line marks $\sigma_q = 1$, the boundary between complementarity ($\sigma_q < 1$) and substitutability ($\sigma_q > 1$). Parameters: $\gamma_q = 0.5$ and $\Delta_A = 0.5$.

The analysis of consumption and marginal products across different elasticities of substitution reveals two fundamental mechanisms that shape capital allocation in networked production economies: an insurance motive driven by nonlinear productivity effects, and an efficiency cost arising from the reallocation of capital across sectors.

The insurance motive is clearly visible in Figure 2, panel (a). We observe a striking asymmetry in how productivity shocks affect consumption, particularly when inputs are complements ($\sigma_q < 1$). For example, at $\sigma_q = 0.25$, moving from the average productivity ($A = 1$) to low productivity (A_L) reduces consumption by approximately 39.4%, while moving from average to high productivity (A_H) increases consumption by only 15.6%. This nonlinearity means

that negative shocks to the upstream sector have disproportionately severe effects on overall consumption compared to positive shocks of equal magnitude. As σ_q increases toward and beyond 1 (perfect substitutes), this asymmetry diminishes, as the downstream sector can more effectively substitute its own capital for the intermediate input when upstream productivity is low. For a risk-averse agent facing strong complementarities, this asymmetry creates a powerful incentive to insure against the downside risk by strengthening the upstream sector.

The efficiency cost of this insurance is revealed in panel (b) of Figure 2, which shows the marginal products of capital across sectors and states. The behavior of marginal products is particularly illuminating as we approach the Leontief case ($\sigma_q \to 0$), where inputs become perfect complements. In this limit, we observe that MPK_2 in the low productivity state and MPK_1 in the high productivity state both approach zero. This pattern emerges because, with perfect complementarities, adding resources to the non-bottleneck sector yields no marginal benefit to production, since output is constrained by the least abundant input.

On the other hand, the asymmetry in marginal products in bottleneck sectors reveals the efficiency cost of providing insurance. In the low productivity state, sector 1 (upstream) becomes the bottleneck, while in the high productivity state, sector 2 (downstream) constrains production. However, a crucial asymmetry emerges: the marginal product of capital in sector 2 during the high productivity state substantially exceeds the marginal product in sector 1 during the low productivity state. This disparity arises because the productivity shock directly affects the marginal productivity of capital in the upstream sector. When upstream productivity is low, even though sector 1 is the bottleneck, the marginal benefit of allocating additional capital to this sector is dampened by the negative productivity shock itself. Thus, providing insurance through upstream preallocation requires moving capital from a highly productive use (downstream sector in good times) to a relatively unproductive one (upstream sector in bad times).

For a risk-neutral agent focused solely on expected returns, preallocation upstream would be suboptimal. The expected marginal product (averaging across states with equal probability) is negative when $\sigma_q < 1$, indicating that shifting capital downstream would increase expected consumption. However, for a sufficiently risk-averse agent, the calculus changes fundamentally. The high marginal utility of consumption in the low state, amplified by the nonlinear impact of productivity shocks, outweighs the efficiency cost in the high state.

This analysis illuminates the central economic trade-off driving preallocation. When inputs are complements, strengthening the upstream sector provides valuable insurance against negative productivity shocks, but it has a cost in terms of efficiency. Whether this cost outweighs

the insurance value depends on the risk aversion of the agent and the elasticity of substitution between inputs. Next, we will provide a formal characterization of the conditions under which preallocation occurs.

2.5 Analytical conditions for preallocation upstream

Now, we will use the same strategy of evaluating the marginal expected utility at the deterministic solution to characterize the analytical conditions under which preallocation occurs. First, we evaluate the consumption level at the deterministic allocation:

$$C_S|_{K_1^{determ}} = \left(\frac{1}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}}\right)\left(A_S^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{\sigma_q}{\sigma_q - 1}}$$

Next, we evaluate the derivative of consumption with respect to K_1 at the deterministic solution in each of the two states S:

$$\left.\frac{\partial C_S}{\partial K_1}\right|_{K_1^{determ}} = \frac{\left(A_S^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_S^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}}$$

The FOC evaluated at the deterministic optimum is:

$$\left.\frac{\partial \mathbb{E}[U(C)]}{\partial K_1}\right|_{K_1^{determ}} \propto p\frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} + (1-p)\frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}},$$

where we have removed the positive constant $\left[\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}\right]^{\epsilon_c^{-1}}$.

To show that we get preallocation, that is $K_1^* > K_1^{determ}$, we need to prove that the marginal expected utility with respect to K_1 is positive when evaluated at the deterministic solution, $\left.\frac{\partial \mathbb{E}[U(C)]}{\partial K_1}\right|_{K_1^{determ}} > 0$. This would indicate that the planner would increase K_1 beyond the deterministic optimum when facing uncertainty. Thus, we are trying to find parametric

conditions such that:

$$(1-p)\frac{A_H^{1-\sigma_q^{-1}}-1}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}} > p\frac{1-A_L^{1-\sigma_q^{-1}}}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}}$$

For the following results, we consider the case where $p = 1/2$, which gives us symmetric positive and negative productivity shocks. The following proposition characterizes the conditions under which preallocation exists.

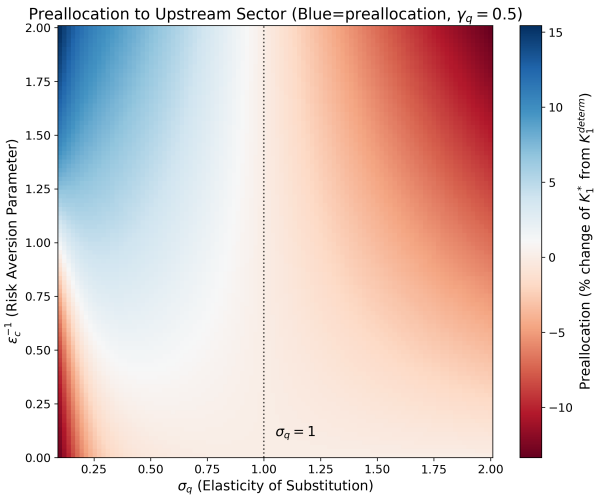
Proposition 1. *If inputs are complements ($\sigma_q < 1$), the planner preallocates capital to the upstream sector (that is, $K_1^* > K_1^{determ}$) under any of the following conditions, up to the first-order approximation:*

1. *If risk aversion is high enough, $\epsilon_c^{-1} \geq 1$.*
2. *If risk aversion is moderate $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q) < \epsilon_c^{-1} < 1$.*
3. *If risk aversion is low, $0 < \epsilon_c^{-1} < \bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q)$, and the downstream sector's importance in production is sufficiently high, $\gamma_q > \bar{\gamma}_q(\triangle_A)$.*

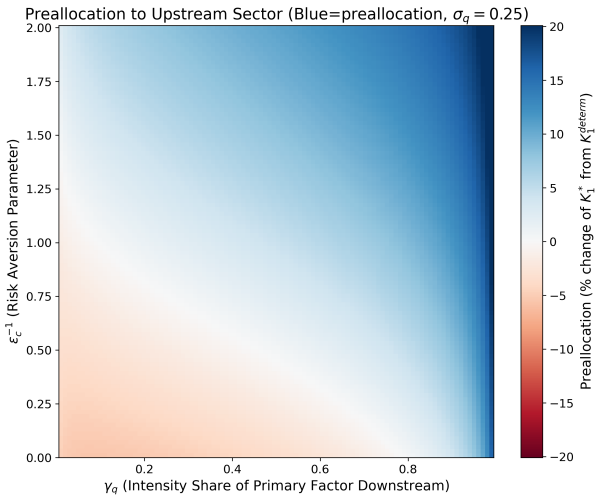
The thresholds $\bar{\gamma}_q(\triangle_A)$ and $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q)$ are determined by the size of productivity shocks and production parameters.

The proof of this proposition is presented in Appendix A.3.

Figure 3 provides computational verification of this theoretical result across a wide range of parameter values. Panel (a) illustrates how preallocation depends on the elasticity of substitution between inputs and on risk aversion: with substitutable inputs ($\sigma_q > 1$), capital is preallocated *downstream*, while with complementary inputs ($\sigma_q < 1$), capital is generally preallocated upstream except when both risk aversion and the elasticity of substitution are low. Panel (b) examines how preallocation varies with the downstream sector's intensity share (γ_q) when inputs are complements ($\sigma_q = 0.5$). The results show that preallocation to the upstream sector occurs across most parameter combinations, except when both risk aversion and the downstream intensity share are low. The blue regions in both panels represent parameter combinations where preallocation to the upstream sector occurs, while the dashed black line in panel (b) shows the boundary case of zero preallocation.



(a) Preallocation as a function of risk aversion and elasticity of substitution



(b) Preallocation as a function of risk aversion and downstream intensity share

Figure 3: Preallocation to the upstream sector. Panel (a) shows the percentage difference between stochastic optimum K_1^* and deterministic optimum K_1^{determ} across elasticity of substitution σ_q and risk aversion ϵ_c^{-1}, with $\gamma_q = 0.5$. The vertical line at $\sigma_q = 1$ marks the boundary between complementarity and substitutability. Panel (b) shows preallocation across downstream intensity share γ_q and risk aversion ϵ_c^{-1}, assuming $\sigma_q = 0.25$. In both panels, blue regions indicate positive preallocation upstream, while red regions indicate the opposite. The dashed black line in panel (b) marks zero preallocation. We use $\Delta_A = 0.5$ throughout.

2.6 Consequences of preallocation upstream

Having established the conditions under which preallocation occurs, we explore its consequences for the economy. Specifically, we analyze how preallocation affects the impact of productivity shocks on consumption and the distribution of consumption across states of the world. These results will provide further insights into how preallocation serves as an insurance mechanism against negative productivity shocks in networked production economies.

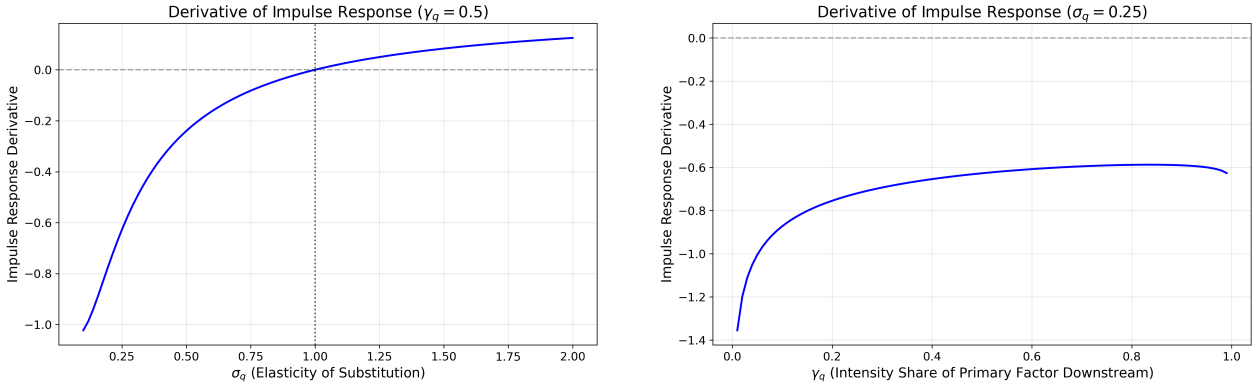
Proposition 2. *If inputs are complements ($\sigma_q < 1$), then increasing capital allocation to the upstream sector beyond the deterministic optimum decreases the impulse response to negative productivity shocks:*

$$\left.\frac{\partial IR(K_1, A_L)}{\partial K_1}\right|_{K_1 = K_1^{determ}} < 0$$

where $IR(K_1, A_L)$ measures the percentage change in consumption resulting from a negative productivity shock.

The proof of this proposition is presented in Appendix A.4.

This proposition formalizes the insurance value of preallocation described in subsection 2.4. When upstream and downstream sectors are complements, allocating more capital to the upstream sector than would be optimal in a deterministic environment dampens the effect of negative productivity shocks on consumption. The negative derivative indicates that each additional unit of capital allocated to the upstream sector reduces the consumption drop caused by adverse shocks. Figure 4 demonstrates this insurance benefit by showing the derivative of the absolute value of the impulse response with respect to capital upstream evaluated at the deterministic optimum, across the parameter space. A negative value for the derivative of the impulse response implies that the impulse response gets smaller in absolute value. As we can see, the derivative is negative when we have complementarities ($\sigma_q < 1$), and positive otherwise. Also, the negative impact increases nonlinearly as the elasticity of substitution decreases. This computational analysis confirms our theoretical prediction that strengthening the upstream sector serves as a form of insurance against productivity disasters.



(a) Impact of preallocation upstream on negative IR as a function of elasticity of substitution

(b) Impact of preallocation upstream on negative IR as a function of downstream share

Figure 4: Impact of preallocation upstream on the negative impulse response of consumption. The plot shows the derivative of the absolute value of the impulse response with respect to capital upstream evaluated at the deterministic optimum, across the parameter space. A negative value for the y axis means that the negative impulse response gets smaller in absolute value. Panel (a) varies the elasticity of substitution σ_q with a fixed $\gamma_q = 0.5$, while panel (b) varies the downstream intensity share γ_q with a fixed $\sigma_q = 0.25$. We use $\Delta_A = 0.5$ throughout.

This insurance mechanism operates precisely because negative shocks to the upstream sector have disproportionate effects on downstream production when inputs are complements. By preallocating more capital upstream, the planner reduces the severity of these disruptions. Importantly, this dampening effect only occurs when inputs are complements ($\sigma_q < 1$). With substitutable inputs, preallocation toward the downstream sector would provide better

insurance.

Having established the insurance benefits of preallocation, we now turn to analyzing its costs. Specifically, we show that allocating more capital to the upstream sector comes at the expense of lower expected consumption, highlighting the fundamental trade-off that the planner faces between insurance and efficiency.

Proposition 3. *Up to the first-order approximation, if inputs are complements ($\sigma_q < 1$), the effect of upstream preallocation on expected consumption depends on the relative strength of the upstream sector:*

1. *When the upstream sector is large enough ($\gamma_q < \bar{\gamma}_q(\triangle_A)$), increasing upstream capital allocation beyond the deterministic optimum decreases expected consumption:*

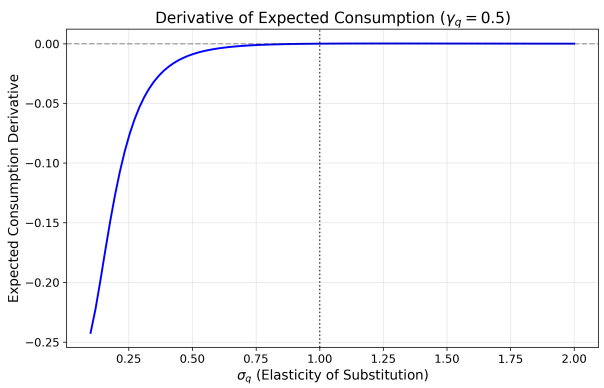
$$\left.\frac{\partial \mathbb{E}\{C(A, K_1)\}}{\partial K_1}\right|_{K_1 = K_1^{determ}} < 0.$$

2. *When the upstream sector's role is more limited ($\gamma_q > \bar{\gamma}_q(\triangle_A)$), increasing upstream capital allocation increases expected consumption.*

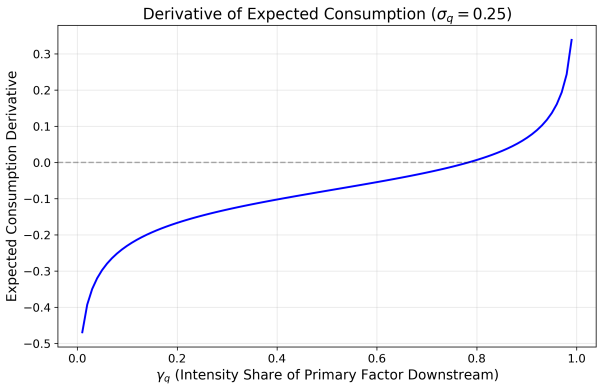
The proof of this proposition is presented in Appendix A.5.

This proposition reveals the fundamental cost-benefit trade-off of strengthening the upstream sector through preallocation. When inputs are complements, under the condition that the upstream sector is large enough ($\gamma_q < \bar{\gamma}_q(\triangle_A)$), allocating more capital upstream reduces expected consumption. The planner accepts this efficiency loss because a robust upstream sector provides crucial insurance benefits, as described in the previous proposition.

Figure 5 illustrates this trade-off by showing how preallocation affects expected consumption across the parameter space. The computational results confirm our theoretical prediction: when the upstream sector's relative importance is high (low γ_q), the insurance benefits of preallocation come at the cost of lower average consumption.



(a) Impact of preallocation upstream on expected consumption as a function of elasticity of substitution



(b) Impact of preallocation upstream on expected consumption as a function of downstream intensity share

Figure 5: Effect of preallocation on expected consumption. The plot shows the derivative of expected consumption with respect to capital upstream evaluated at the deterministic optimum, across the parameter space. Panel (a) varies the elasticity of substitution σ_q with fixed $\gamma_q = 0.5$, while panel (b) varies the downstream intensity share γ_q with fixed $\sigma_q = 0.25$. We use $\Delta_A = 0.5$ throughout.

Notably, this finding reveals a key aspect of the welfare cost of business cycles in production network economies. Rather than manifesting primarily through consumption disasters, which are largely mitigated by the optimal strengthening of the upstream sector, the welfare cost appears through permanently lower average consumption. The planner sacrifices mean consumption to build resilience in the critical upstream portion of the production chain.

This trade-off is particularly stark when upstream-downstream complementarities are strong. In such cases, the planner accepts a larger reduction in average consumption to fortify the upstream sector against severe disruptions that would otherwise cascade through the production network.

3 Quantitative Model

We introduce here the full quantitative model. There are a finite number of perfectly competitive sectors indexed by $j = 1, ..., N$. A representative household consumes goods and supplies labor to firms in each sector. Time is discrete and infinite.

3.1 Households

The representative household has the following preferences over consumption of each good j, which we denote C_{jt}, and labor on industry j, which we denote L_{jt}:

$$U = \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1-\epsilon_c^{-1}} \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}} \right] \quad \text{where}$$

$$C_t = \left(\sum_{j=1}^{N} \xi_j^{\frac{1}{\sigma_c}} (C_{jt})^{1-\sigma_c^{-1}} \right)^{\frac{1}{1-\sigma_c^{-1}}}, \quad \sum_{j=1}^{N} \xi_j = 1 \text{ and } L_t = \left(\sum_{j=1}^{N} (L_{jt})^{1+\sigma_l^{-1}} \right)^{\frac{1}{1+\sigma_l^{-1}}}$$

where β is the discount factor, ϵ_c is the intertemporal elasticity of substitution (or the inverse of the relative risk aversion), ϵ_l is the Frisch elasticity of labor, ξ_j captures the time-invariant preference for good j, σ_c is the elasticity of substitution across goods, σ_l controls the degree of labor reallocation between sectors, and θ is a normalization constant.

3.2 Firms

The representative firm in sector j produces gross output Q_{jt} using capital K_{jt}, labor L_{jt}, and intermediate inputs M_{jt}.[6] The production function is:

$$Q_{jt} = \left[(\mu_j)^{\sigma_q^{-1}} (Y_{jt})^{1-\sigma_q^{-1}} + (1-\mu_j)^{\sigma_q^{-1}} (M_{jt})^{1-\sigma_q^{-1}} \right]^{\frac{1}{1-\sigma_q^{-1}}}, \text{ where}$$

$$Y_{jt} = A_{jt} \left[(\alpha_j)^{\sigma_y^{-1}} (K_{jt})^{1-\sigma_y^{-1}} + (1-\alpha_j)^{\sigma_y^{-1}} (L_{jt})^{1-\sigma_y^{-1}} \right]^{\frac{1}{1-\sigma_y^{-1}}}.$$

Variable Y_{jt} denotes value-added production, α_j captures the share of capital in value-added, μ_j parametrizes the share of materials in gross output, σ_q is the elasticity of substitution between primary outputs (e.g. capital and labor) and materials, and A_{jt} is an industry-specific shock to value added productivity that follows the process

$$\log A_{jt+1} = \rho_j \log A_{jt} + \varepsilon_{jt+1}^A$$

where ρ_j represents industry-specific persistence and the shocks ε_{jt+1}^A are distributed multivariate normal with mean 0 and variance-covariance matrix Σ^A. The industry-specific productivity shocks may be correlated, that is, the variance-covariance matrix may not be

[6] This variable is also labeled as material in some papers (e.g., Rotemberg and Woodford, 1993).

diagonal.

Firms can accumulate capital by producing an industry-specific investment good I_{jt} facing capital adjustment costs denoted by Φ_{jt}:

$$K_{jt+1} = (1-\delta_j)K_{jt} + I_{jt} - \Phi_{jt},$$
$$\Phi_{jt} = \frac{\phi}{2}\left(\frac{I_{jt}}{K_{jt}} - \delta_j\right)^2 K_{jt}$$

where δ_j is the industry-specific depreciation rate, and ϕ parametrize the adjustment cost function.

The investment good is produced by bundling goods produced by other industries:

$$I_{jt} = \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right)^{\sigma_I^{-1}} \left(I_{ijt}\right)^{1-\sigma_I^{-1}}\right)^{\frac{1}{1-\sigma_I^{-1}}}, \quad \text{where} \quad \sum_{i=1}^{N} \gamma_{ij}^{I} = 1$$

where γ_{ij}^{I} represents the importance of good i in the production of the investment good for sector j, and σ_I is the elasticity of substitution between inputs of the investment bundle. In the same vein, the intermediate input is produced using the following bundle:

$$M_{jt} = \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{m}\right)^{\sigma_m^{-1}} \left(M_{ijt}\right)^{1-\sigma_m^{-1}}\right)^{\frac{1}{1-\sigma_m^{-1}}}, \quad \text{where} \quad \sum_{i=1}^{N} \gamma_{ij}^{m} = 1.$$

Parameters γ_{ij}^{m} and σ_m are analogous to the parameters γ_{ij}^{I} and σ_I discussed for the investment bundle.

3.3 Market Clearing and the Planner's First Order Conditions

The market clearing conditions for each good is:

$$Q_{jt} = C_{jt} + \sum_{i=1}^{N} \left(M_{jit} + I_{jit}\right),$$

which implies that gross output equals final consumption, intermediate inputs, and investment goods.

In order to obtain the first order conditions, we will use the fact that the model satisfies the first welfare theorem, so we can formulate the problem as a planning problem. The planner's Lagragian is given by:

$$\begin{aligned}\mathcal{L} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ & \frac{1}{1-\epsilon_c^{-1}} \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}} \\ & + \sum_{j=1}^{N} P_{jt}^{\mathrm{k}} \left[I_{jt} + (1-\delta_j) K_{jt} - \Phi_{jt} - K_{jt+1} \right] \\ & + \sum_{j=1}^{N} P_{jt} \left[Q_{jt} - C_{jt} - \sum_{i=1}^{N} \left[M_{jit} + I_{jit} \right] \right] \Bigg\}\end{aligned}$$

where P_{jt}^k is the Lagrange multiplier associated to the capital accumulation constraint, P_{jt} is the Lagrange multiplier associated to the market clearing condition.

In Appendix B, we provide a detailed derivation of all first-order conditions, as well as additional .

4 Calibration and solution method

We calibrate the model in two steps: first, by directly setting parameters from data or existing literature, and second, by internally calibrating remaining parameters to match key empirical moments. Table 1 summarizes all parameter values and their sources.

For our dataset, we rely heavily on the empirical work of Vom Lehn and Winberry (2022)[7]. The dataset spans 1948-2018 and draws primarily from BEA Tables. The BEA Fixed Assets data is used to construct the investment network and depreciation parameters; Input-Output data provides the intermediates network parameters; and GDP-by-Industry data yields value added and employment observations. The aggregation level is $N = 37$ private non-farm sectors, with non-manufacturing sectors at the 2-digit NAICS level and manufacturing sectors at the 3-digit level.

[7]Data can be downloaded https://sites.google.com/site/cvomlehn/research?authuser=0. Our model generalizes some assumptions and functional forms used by Vom Lehn and Winberry (2022), so our calibration strategy differs. In particular, we have imperfect labor reallocation, and general CES for all aggregators. This mostly modify how to calibrate TFP shocks, and we need to add additional empirical targets, as explained below.

Table 1: Model Calibration

Parameter	Symbol	Value	Source
Panel A: External Parameters			
Discount factor	β	0.96	Standard
IES	ϵ_c	0.5	Standard
Frisch elasticity	ϵ_l	0.5	Standard
Intermediate elasticity	σ_m	0.1	Atalay (2017)
K-L elasticity	σ_y	0.8	Oberfield and Raval (2021)
VA-Int. elasticity	σ_q	0.5	Intermediate Value
Investment elasticity	σ_I	0.5	Intermediate Value
Consumption elasticity	σ_c	0.5	Intermediate Value
Depreciation	δ_j	By sector	BEA (1947-2018)
TFP persistence	ρ_j	By sector	Solow residuals
TFP covariance	Σ_A	By sector	Solow residuals
Panel B: Internal Parameters			
HH preferences	ξ_j	By sector	Exp. shares
VA intensity	μ_j	By sector	Exp. shares
Capital intensity	α_j	By sector	Exp. shares
Investment network	γ_{ij}^I	By sector-pair	Exp. shares
I-O network	γ_{ij}^M	By sector-pair	Exp. shares
Capital adj. cost	ϕ	Calibrated	Inv. volatility
Labor realloc.	σ_l	Calibrated	Labor volatility

Notes: Panel A shows parameters from data or literature. Panel B shows parameters matched to empirical moments. Model calibrated annually with $N = 37$ sectors (BEA 1948-2018). TFP processes detrended using fourth-order log-polynomial.

4.1 Externally-calibrated parameters

Household preferences. At annual frequency, we set the discount factor (β) to 0.96, implying a 4% real interest rate in the deterministic steady state. The intertemporal elasticity of substitution (ϵ_c) is 0.5, corresponding to a risk aversion of 2. The Frisch Elasticity of Labor Supply (ϵ_l) is 0.5.

Elasticities of substitution. Following Atalay (2017) and Vom Lehn and Winberry (2022), we set the elasticity of substitution between intermediate goods (σ_m) to 0.1. Based on Oberfield and Raval (2021), the elasticity between capital and labor (σ_y) is 0.8. We set the elasticities between value-added and intermediates (σ_q), between investment goods (σ_I), and

between consumption goods (σ_c) to 0.5.[8]

Depreciation. Capital depreciation rates (δ_j) reflect each sector's implied depreciation rates derived from BEA Fixed Assets data (1947-2018). Each δ_j is calculated by averaging the annual depreciation rates, weighting by the quantity of each capital good type used in sector j.

TFP. Parameters governing the TFP process (ρ_j, Σ_A) are calibrated using sector-level Solow residuals. Given our CES specification for value added, TFP variation is computed as:

$$\Delta \log A_{jt} = \log\left(\frac{Y_{jt}}{Y_{j,t-1}}\right) - \left(\frac{1}{1-\sigma_y^{-1}}\right) \log\left[\frac{\alpha_{jt}^{\sigma_y^{-1}} K_{jt}^{1-\sigma_y^{-1}} + (1-\alpha_{jt})^{\sigma_y^{-1}} L_{jt}^{1-\sigma_y^{-1}}}{\alpha_{jt}^{\sigma_y^{-1}} K_{j,t-1}^{1-\sigma_y^{-1}} + (1-\alpha_{jt})^{\sigma_y^{-1}} L_{j,t-1}^{1-\sigma_y^{-1}}}\right],$$

where α_{jt} denotes the capital share of sector j at time t. We allow factor shares to vary over time to account for technological changes, though results remain robust with constant shares. Capital input (K_{jt}) is constructed using the perpetual inventory method, with initial values from BEA nominal year-end capital stock data for 1948. We detrend sector-level TFP using a fourth-order log-polynomial, balancing nonlinearity capture against overfitting.[9]

4.2 Internally-calibrated parameters

We move now to the internally-calibrated parameters.

Expenditure shares, IO network and investment networks. Given that we have general CES aggregators at all levels, expenditure shares do not match exactly intensity shares of the corresponding CES aggregator. For example, $1-\alpha$ is not exactly the labor share, thought it is proportional to it. This also applies to the input-output network and the investment network, which also use CES aggregators. In Appendix B.10, we describe how we can obtain an analytic expression for expenditure shares for all aggregators in the model. Using this expression, we iterate on the intensity shares of the model until the expenditure shares in the deterministic steady state matches the data. Specifically, this procedure is used to calculate the intensity parameters associated with household preferences over goods (ξ_j), value-added intensity share (μ_j), capital share(α_j), the investment network (γ_{ij}^I), and the input-output network (γ_{ij}^M).

[8] Limited empirical evidence exists for these elasticities as most studies assume Cobb-Douglas. Our robustness analyses confirm that results remain qualitatively unchanged with this parameterization.

[9] Results remain stable with lower-order polynomial detrending.

Parameters set to match volatilities. The capital adjustment cost parameter (ϕ) matches the weighted average volatility of sectoral investment, while the labor reallocation parameter (σ_l) matches the weighted average volatility of sectoral labor.

4.3 Solution method

Our analysis of preallocation and nonlinear shock propagation requires a global solution method. With this solution in hand, we can simulate the economy to compute the ergodic distribution of sectoral and aggregate variables, capturing both nonlinear shock propagation and the anticipation effects of risk on agents' behavior. This ergodic distribution centers around the stochastic steady state (SSS)—defined as the equilibrium state where agents fully internalize the impact of future shocks on equilibrium outcomes, but current shock realizations are zero. This approach differs fundamentally from standard first- or second-order perturbation methods, which approximate the solution locally around the deterministic steady state (DSS)—a steady state where shocks are not just zero but impossible, precluding any preallocation behavior.

The global solution of our model presents a significant computational challenge. With 37 sectors and two state variables per sector (capital and TFP), we must solve a stochastic dynamic programming problem with 74 state variables—pushing the boundaries of current computational capabilities. We overcome this challenge by extending the 'deep equilibrium nets' method developed by Azinovic et al. (2022) to allow for high dimensional shocks, through the use montecarlo simulation to build the expectation terms that appear in the euler equations.

While we provide detailed technical documentation in Appendix C, the core approach can be summarized as follows: We approximate all equilibrium objects (policies and prices) as functions of the state variables using deep neural networks. The solution process then reduces to training these networks by minimizing violations of the equilibrium conditions, which consists of euler equations and technological constraints. This is accomplished through an iterative procedure that alternates between simulation and optimization steps. In each iteration, we first simulate the economy using current network parameters, then use the simulated data to update these parameters, minimizing equilibrium condition errors. This process continues until we achieve the desired numerical precision.

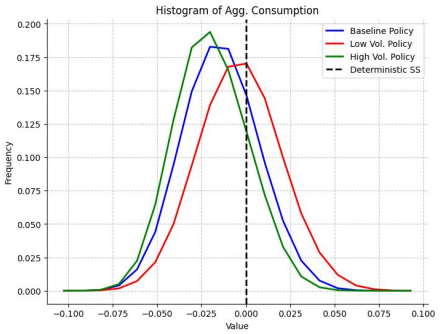
5 Quantitative results

5.1 Ergodic distribution

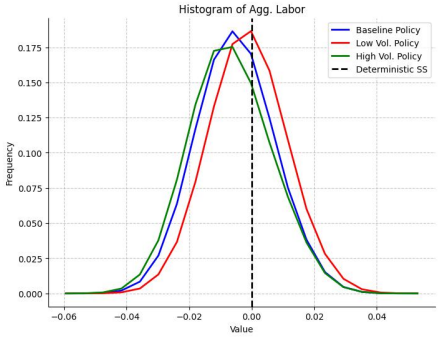
We start by analyzing how risk can lead to preallocation in the full model. Armed with our new solution method, we consider three scenarios differing on the shock volatility. First, the baseline scenario, discussed in Section 4. Second, a *low volatility* scenario, in which the volatility of all sectoral TFP shocks is divided by two. This scenario gets us closer to the linear model. Finally, a *high-volatility* scenario in which shocks are twice as volatile as in the baseline.

We solve the model nonlinearly for each of these scenarios. Then we compute the ergodic distribution by simulating 500,000 periods using in the three cases the volatility of the baseline scenario. That implies that the allocations in the high- and low-volatility scenarios are suboptimal, as the planner forecasts volatilities that do not materialize in the simulation. Results are displayed in Figure 6. Each panel displays the ergodic distribution of some aggregate variable, with a dashed vertical line marking the position of the DSS. Table 2 summarizes this information.

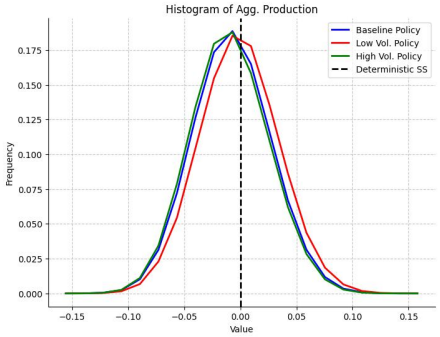
Figure 6: Ergodic distribution of key aggregate variables



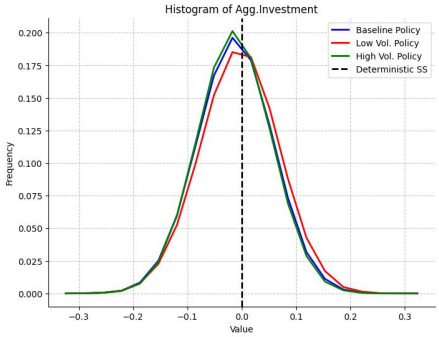
(a) Consumption



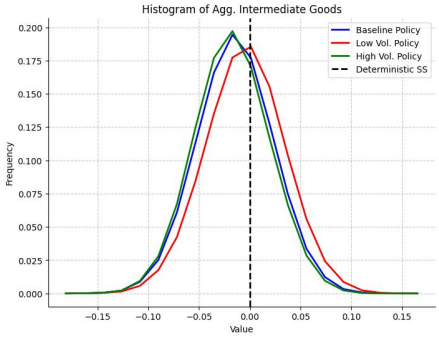
(b) Labor



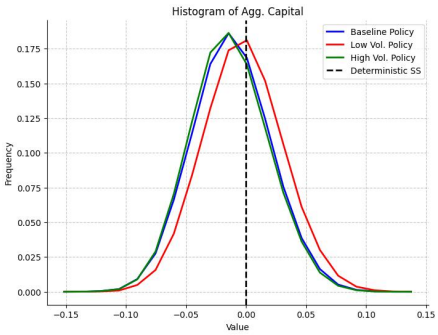
(c) GDP



(d) Investment



(e) Intermediates



(f) Capital

Note: The figure shows the ergodic distribution of specific aggregate variables, using the 500,000 periods simulation of the global solution. For each variable, we show the case for three policy functions: baseline policy, the policy trained under low volatility (shocks scaled by 0.5), and the policy trained under high volatility (shocks scaled by 1.5). The figures are calculated using the same shocks for the four policies, that is, the only thing that varies is the policy function. The baseline volatility was used to simulate the shocks.

Variable	Policy	Mean (%)	Sd (%)	Skewness	Kurtosis
Consumption	Benchmark	-1.4	2.1	0.089	-0.102
	High volatility	-2.0	2.0	0.145	-0.193
	Low volatility	-0.3	2.3	0.067	-0.059
	Log-linear	0.0	3.1	-0.001	-0.028
Labor	Benchmark	-0.5	1.2	0.076	-0.076
	High volatility	-0.7	1.3	0.131	-0.138
	Low volatility	-0.1	1.2	0.059	-0.047
	Log-linear	0.0	1.3	-0.002	-0.019
GDP	Benchmark	-0.8	3.4	0.073	-0.062
	High volatility	-1.0	3.4	0.086	-0.102
	Low volatility	-0.1	3.5	0.064	-0.038
	Log-linear	0.0	3.8	0.000	-0.016
Investment	Benchmark	-1.3	6.9	-0.020	-0.001
	High volatility	-1.4	6.7	-0.019	-0.014
	Low volatility	-0.5	7.1	-0.035	0.001
	Log-linear	0.0	7.9	0.001	0.012
Intermediates	Benchmark	-1.5	3.7	-0.002	-0.025
	High volatility	-1.8	3.7	0.019	-0.067
	Low volatility	-0.5	3.9	-0.024	-0.005
	Log-linear	0.0	4.2	-0.002	-0.009
Capital	Benchmark	-1.2	3.3	0.056	-0.027
	High volatility	-1.4	3.2	0.083	-0.064
	Low volatility	-0.3	3.3	0.045	-0.016
	Log-linear	0.0	3.7	0.011	-0.016

Table 2: Descriptive statistics of key aggregate variables

Note: This table shows descriptive statistics of aggregate variables, using the 500,000 periods simulation of the global solution. For each variable, we show the case for the four policy functions: baseline policy, the policy trained under low volatility (shocks scaled by 0.5), the policy trained under high volatility (shocks scaled by 1.5), and the policy trained under log-linear approximation (no expected shock as it is a perturbation around the DSS). The statistics are calculated using the same shocks for the four policies, that is, the only thing that varies is the policy function. The benchmark volatility, calculated from data, was used to simulate the shocks. Mean and standard deviation are expressed in percentage deviations from the DSS for easier interpretation (for example, -1.4% means the variable is on average 1.4% below its DSS value). Skewness and kurtosis are calculated over variables expressed in log deviations from the DSS.

Several important results emerge. First, all aggregate variables display a mean below the DSS. That implies that once aggregate risk is factored in, agents find it optimal to reduce their investment and labor supply in a more volatile environment. This translates into lower capital and labor, and hence lower consumption and output. This mechanism is amplified as shock volatility increases. That showcases how the model nonlinearity becomes more prevalent as shocks are more volatile. Notice how the impact is heterogeneous across variables. While a fourfold increase in volatility between the low- and high-volatility scenarios implies a sevenfold decrease in consumption, from 0.3 percent to 2.0 percent, it is only threefold in the case of

investment, from 0.5 percent to 1.3 percent.

Second, the standard deviation of all aggregate variables decreases with expected shock volatility. In other words, in counterfactual economies with higher sectoral TFP volatility, the standard deviation of key macro variables increases less than one-to-one The planner is thus trading a lower mean by a reduction in standard deviation as shock variance increases. This is in line with the theoretical results presented in Section 2. Furthermore, the implied aggregate volatility is substantial – e.g. the standard deviation of annual GDP is 3.4%.

Third, the economy displays negative excess kurtosis and a negligible degree of skewness. The negative kurtosis reflects the fact that the economy reduces large excursions from the mean, that is, the economy displays 'thinner tails' than the (log-)normal. This reflects the fact that policy functions bend for large shock values in the global solution, whereas they are (log-)linear in the linear solution (we will return to this point in subsection 5.3).

5.2 Stochastic steady state

We next analyze the SSS. A comparison between Table 2 and 3 clearly shows how the SSS is very close to the mean of the ergodic distribution. The advantage of the SSS is that it allows us to analyze in detail the sectoral characteristics of the equilibrium under the global solution.

Policy	Consumption (%)	Labor (%)	GDP (%)	Investment (%)	Intermediates (%)	Capital (%)
Low Volatility	-0.43	-0.16	-0.28	-0.37	-0.43	-0.45
Baseline	-1.56	-0.58	-1.02	-1.32	-1.50	-1.48
High Volatility	-1.89	-0.62	-1.07	-1.14	-1.47	-1.21

Table 3: Aggregate Variables in Stochastic Steady State

Note: This table shows aggregate variables in the SSS. We show the case for the baseline model, the model with low volatility (shocks scaled by 0.5), and the model with high volatility (shocks scaled by 1.5). Values are expressed as percentage deviations from the DSS (for example, -1.56% means consumption is 1.56% below its dDSS value). The SSS is calculated by sampling 1,000 points from the full simulation, and simulating forward but setting shocks to zero. Then, we take the resulting 1,000 ending points, verify they are the same within 0.01% tolerance, and average them.

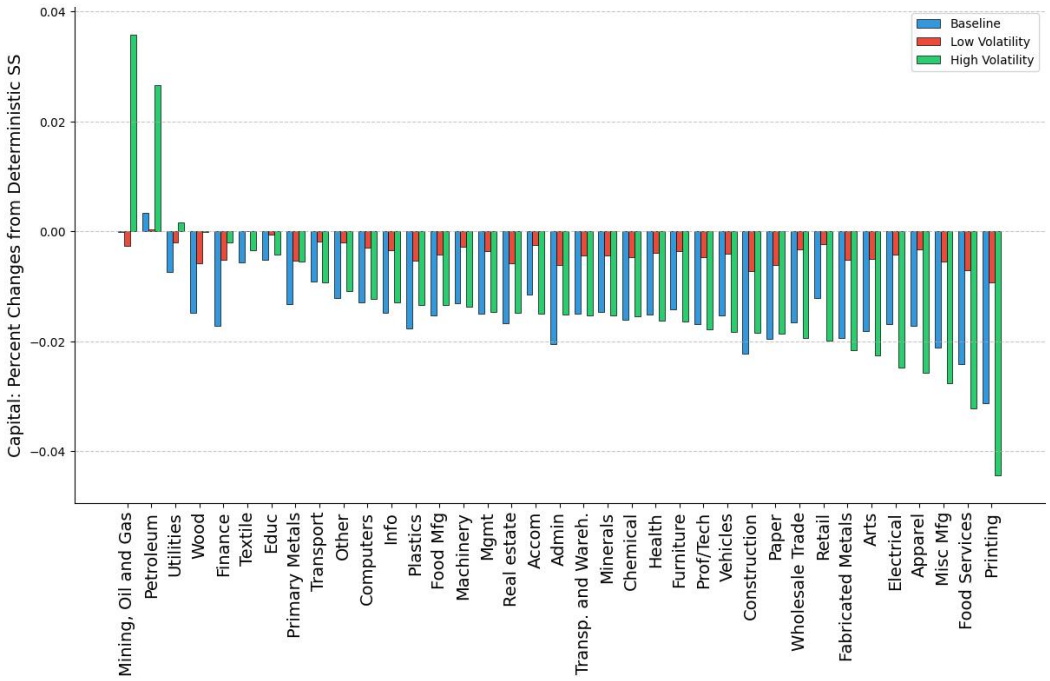
Figure 7 displays the capital allocation across sectors as a difference with respect to the DSS. This reflects how the anticipation of aggregate risk affects agents' decisions regarding investment.

From the aggregate results above, we know that aggregate capital is lower in the SSS than in the DSS. This result, however, masks a considerable degree of sectoral heterogeneity. In

the baseline scenario, aggregate capital is 1.5 percent lower in the SSS than in the DSS. This translates into a decline in capital across most sectors of a magnitude between roughly 1-2 percent (blue). Capital is particularly penalized in Printing and Food Services. There is, however, an increase in the stock of capital in one upstream sector: Mining, Oil, and Gas.

This heterogeneous capital allocation gets much amplified in the high-volatility scenario. While most sectors experience a decline of a similar magnitude as in the baseline, the striking exceptions are Petroleum and Mining, Oil, and Gas, which get an increase larger than 2 percent compared to the DSS. The SSS does exhibit preallocation towards upstream sectors, in line with the theoretical prescriptions of Section 2.

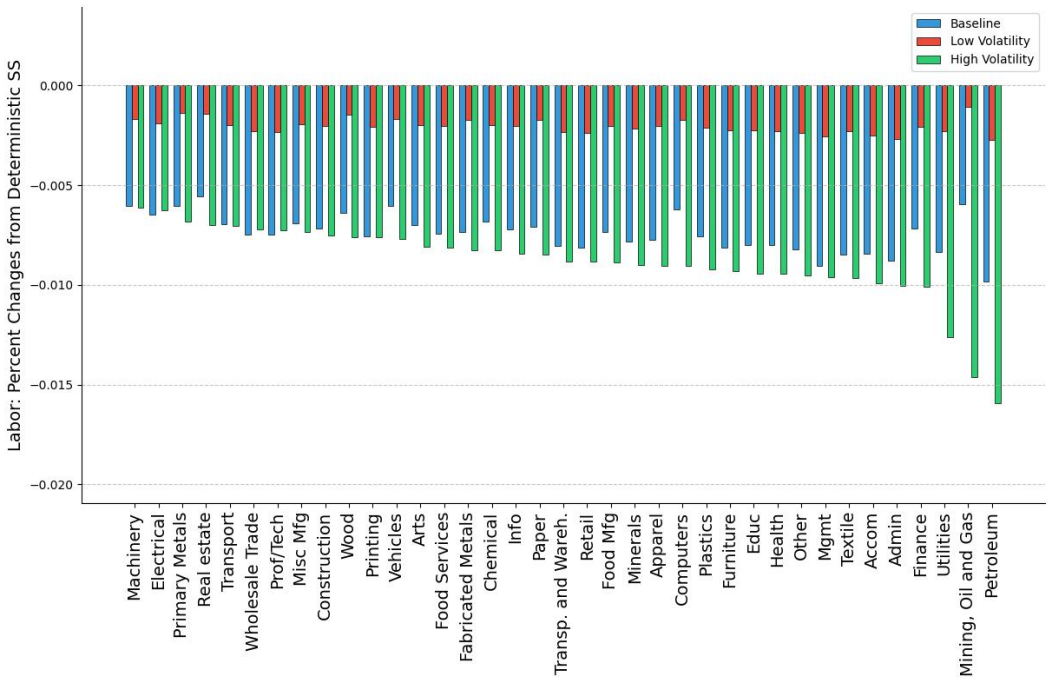
Figure 7: Distribution of Capital in the Stochastic Steady State



Note: This figure shows the capital of each sector in the stochastic steady state. We show the case for the baseline model, the model with low volatility (shocks scaled by 0,5), and the model with high volatility (shocks scaled by 1.5). Sectors are ordered using the value for the high volatility policy. Variables are in log deviations from the deterministic steady state, so 0.05 can be interpreted as 5% above the deterministic steady state. The stochastic steady state is calculated by sampling 1000 points from the full simulation and simulating forward trajectories with the shocks set to zero. Then, we take the resulting 1000 ending points, verify they are the same within 0.01% tolerance, and average them.

Figure 8 shows the distribution of labor in the stochastic SSS, compared to the DSS. We observe the opposite pattern compared to the distribution of capital. In the high volatility case, labor is lower in Petroleum and Mining, Oil and Gas. Notice, however, how SSS of labor is not exactly the specular image of that of capital, as Machinery and Electrical are the two sectors less penalized in terms of labor.

Figure 8: Distribution of Labor in the Stochastic Steady State



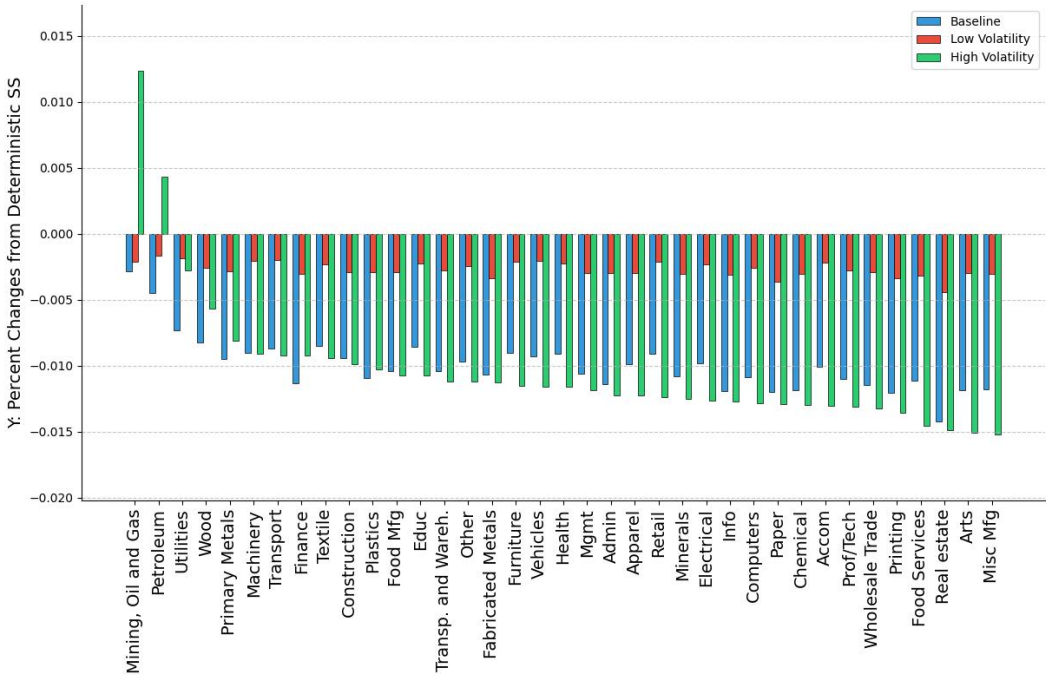
Note: This figure shows labor of each sector in the stochastic steady state. We show the case for the baseline model, the model with low volatility (shocks scaled by 0,5), and the model with high volatility (shocks scaled by 1.5). Sectors are ordered using the value for the high volatility policy. Variables are in log deviations from the deterministic steady state, so 0.05 can be interpreted as 5% above the deterministic steady state. The stochastic steady state is calculated by sampling 1000 points from the full simulation and simulating forward trajectories with the shocks set to zero. Then, we take the resulting 1000 ending points, verify they are the same within 0.01% tolerance, and average them.

Still, as we can see in Figure 9, value-added production is higher in those upstream sectors such as Mining, Oil and Gas, and Petroleum, so the higher level of capital dominates the lower level of labor.

As we discuss in Section 5.3 next, this distributional pattern in the SSS increases resilience

to negative shocks, since capital takes time to accumulate and cannot be readily adjusted as a response to a negative shock, while labor can be more easily reallocated.

Figure 9: Distribution of Value Added in the Stochastic Steady State



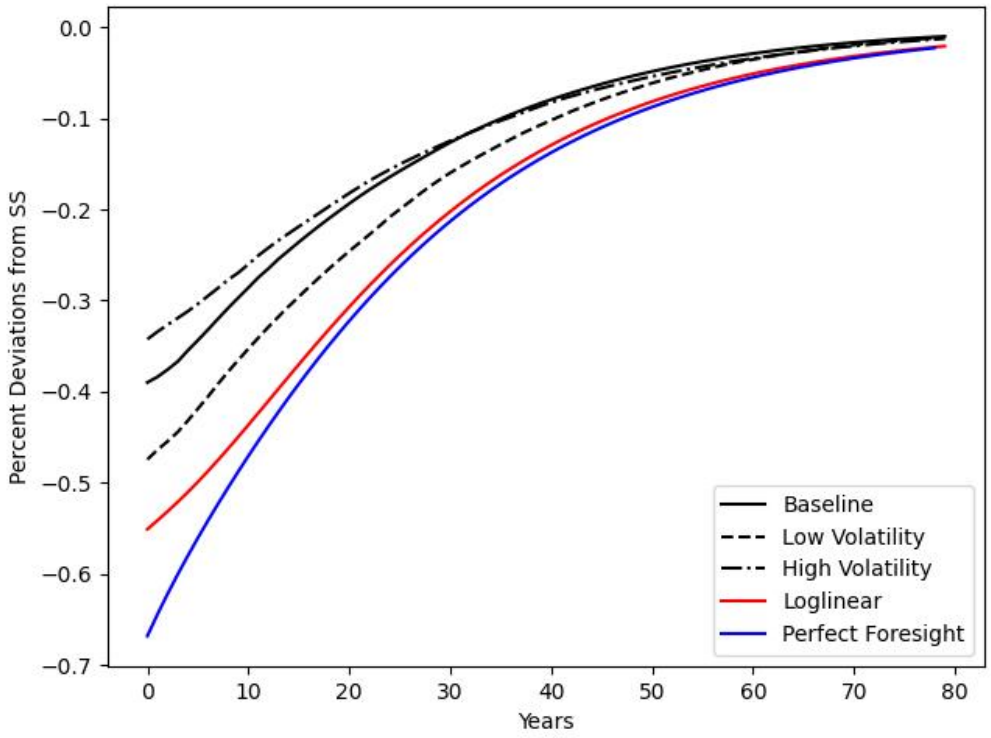
Note: This figure shows value added production of each sector in the stochastic steady state. We show the case for the baseline model, the model with low volatility (shocks scaled by 0,5), and the model with high volatility (shocks scaled by 1.5). Sectors are ordered using the value for the high volatility policy. Variables are in log deviations from the deterministic steady state, so 0.05 can be interpreted as 5% above the deterministic steady state. The stochastic steady state is calculated by sampling 1000 points from the full simulation and simulating forward trajectories with the shocks set to zero. Then, we take the resulting 1000 ending points, verify they are the same within 0.01% tolerance, and average them.

5.3 Impulse Responses

To examine the dynamic properties of the global solution, we analyze impulse response functions following a large sectoral productivity shock. Figure 10 shows how aggregate consumption responds to a negative 20% TFP shock in Mining, Oil, and Gas, which is a key upstream sector characterized by substantial capital preallocation, as previously discussed. The figure compares three solution methods: our global solution (black lines), a log-linear approximation

(red line), and a perfect foresight solution (blue line).[10] The global solution results include the baseline (solid line), as well as the low- (dashed) and high- (dotted-dashed) volatility scenarios.

Figure 10: Impulse Response of Aggregate Consumption to a Mining, Oil, and Gas Shock



Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Mining, Oil, and Gas (an upstream sector). The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.

The perfect-foresight solution, often denoted as an 'MIT shock', coincides with the log-linear approximation for small shocks (Boppart et al., 2018), but in the case of a large shock, such as the one considered here, they diverge as the perfect-foresight solution captures the nonlinear elements of the response. We see that nonlinearities amplify the impact on aggregate consumption by a 15%, in line with the results by Baqaee and Farhi (2019).

We also compare the perfect foresight response with that under the baseline global solution. This comparison captures the effect of input preallocation, since the baseline solution deviates from the SSS due to ex-ante uncertainty, while the perfect-foresight solution deviates from

[10] Given the nonlinearity of the model, these are *generalized* impulse responses.

the DSS. Figure 10 shows that the baseline policy is attenuated by 50% compared to the perfect foresight solution. This result is similar for shocks in other sectors, both upstream and downstream, as displayed in Appendix D.[11]. The fact that the planner dampens the consumption response through capital preallocation explains why consumption volatility is lower under the global solution, as shown in Table 2 above.

5.4 Welfare Cost of Business Cycles

Finally, we conclude this section by looking at the welfare implications of business cycles under the light of our model. Table 4 reports different measures of welfare costs, always defined in consumption-equivalent terms.[12]

Policy	Full Nonlinear (%)	Loglinear (%)	C fixed at SS (%)	L fixed at SS (%)	Mean at SS (%)
Low Volatility	-0.46	-0.05	0.16	-0.63	-0.03
Baseline	-1.05	-0.11	0.41	-1.51	-0.06
High Volatility	-1.50	-0.19	0.57	-2.18	-0.10

Table 4: Welfare Cost of Business Cycle

Note: This table shows the welfare cost of business cycles under different scenarios. All values are expressed as percentage changes in consumption equivalent terms (for example, -1.05% means households would need to decrease their consumption by 1.05% in the deterministic steady state to be indifferent between that and living in the stochastic economy). We show results for three cases: baseline model, model with low volatility (shocks scaled by 0.5), and model with high volatility (shocks scaled by 1.5). The welfare cost is calculated as described in subsection B.7, using 1,000 different trajectories from the full simulation, each of length 2,000 periods. We then calculate the mean welfare at the first period of each slice. The "Full Nonlinear" and "Loglinear" columns show the welfare cost using the global solution and loglinear approximation respectively. "C fixed at SS" shows the cost when labor follows the observed paths but consumption is fixed at the deterministic steady state. "L fixed at SS" fixes labor at the deterministic steady state. "Mean at SS" shows the pure effect of volatility by renormalizing the observed trajectories so their mean matches the deterministic steady state.

First, we display the losses in the global solution under the baseline calibration (1.05 percent), which are one order of magnitude larger than if the model is solved using loglinear methods (0.11 percent). This result makes evident how linear approximations may be greatly downplaying the true costs of aggregate fluctuations. This result is a direct consequence of the fact that mean consumption is lower in the SSS than in the DSS, and thus the mean consumption along the business cycle falls in the global solution.

Naturally, there are two opposing effects. The SSS displays lower consumption, and lower

[11] There we show impulse responses for shocks to construction, machinery, petroleum, retail, and real estate.

[12] They are defined as the consumption amount that the representative household would be willing to pay in order to live in a counterfactual economy without shocks.

labor supply, where the latter is welfare-improving. To capture exclusively the impact of consumption on welfare, in the second column we recompute welfare in the global solution keeping labor constant at its DSS value. Welfare now increases, reflecting the lower labor in the SSS. Conversely, if we keep labor fixed at its DSS value in the third column, welfare losses are amplified as the only effect is the fall in consumption.

The last column captures exclusively the impact of volatility on the nonlinear model, abstracting from the difference in means. It shows how welfare losses are higher under the loglinear solution, reflecting the higher volatility in that case.

Finally, we compare the low- and high- volatility scenarios. The welfare loss in the high-volatility scenario is 1.5%. This reflects both the lower mean in this case and the higher volatility.

Our results invite us to revisit the traditional view of low welfare costs of business cycles once we abandon the unrealistic assumptions of one single sector and local dynamcis around the DSS.

References

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

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

Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger, "Deep equilibrium nets," *International Economic Review*, 2022.

Baqaee, David and Elisa Rubbo, "Micro Propagation and Macro Aggregation," *Annual Review of Economics*, 2023, *15* (Volume 15, 2023), 91–123.

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

Barro, Robert J., "Rare Disasters, Asset Prices, and Welfare Costs," *American Economic Review*, March 2009, *99* (1), 243–64.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 2018, *89* (C), 68–92.

Carvalho, Vasco, "Aggregate fluctuations and the network structure of intersectoral trade," Economics Working Papers 1206, Department of Economics and Business, Universitat Pompeu Fabra November 2007.

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

Fernandez-Villaverde, Jesus, Galo Nuno, and Jesse Perla, "Taming the Curse of Dimensionality: Quantitative Economics with Machine Learning," Papers October 2024.

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

Gu, Zhouzhou, Mathieu Lauri‘ere, Sebastian Merkel, and Jonathan Payne, "Global Solutions to Master Equations for Continuous Time Heterogeneous Agent Macroeconomic Models," Papers 2406.13726, arXiv.org June 2024.

Han, Jiequn, Yucheng Yang, and Weinan E, "DeepHAM: A global solution method for heterogeneous agent models with aggregate shocks," *arXiv preprint arXiv:2112.14377*, 2021.

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

Jorda, Oscar, Moritz Schularick, and Alan M. Taylor, "Disasters Everywhere: The Costs of Business Cycles Reconsidered," *IMF Economic Review*, March 2024, *72* (1), 116–151.

Kopytov, Alexandr, Bineet Mishra, Kristoffer Nimark, and Mathieu Taschereau-Dumouchel, "Endogenous Production Networks Under Supply Chain Uncertainty," *Econometrica*, 2024, *92* (5), 1621–1659.

Lehn, Christian Vom and Thomas Winberry, "The investment network, sectoral comovement, and the changing US business cycle," *The Quarterly Journal of Economics*, 2022, *137* (1), 387–433.

Liu, Ernest and Aleh Tsyvinski, "A Dynamic Model of Input–Output Networks," *The Review of Economic Studies*, 2024, *91* (6), 3608–3644.

Lucas, Robert E., *Models of Business Cycles*, Basil Blackwell, 1987.

Maliar, Lilia, Serguei Maliar, and Pablo Winant, "Deep learning for solving dynamic economic models.," *Journal of Monetary Economics*, 2021, *122*, 76–101.

Martin, Ian W. R., "Disasters and the Welfare Cost of Uncertainty," *American Economic Review*, May 2008, *98* (2), 74–78.

Oberfield, Ezra and Devesh Raval, "Micro data and macro technology," *Econometrica*, 2021, *89* (2), 703–732.

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

Rotemberg, Julio J and Michael Woodford, "Dynamic general equilibrium models with imperfectly competitive product markets," 1993.

A Two sector model appendix

A.1 Setting

The full description of the model is presented in Section 2. Here we briefly summarize the key equations needed for the proofs.

This vertical network economy consists of two sectors. Sector 1 ("upstream") produces the intermediate good by using capital and is subject to a productivity shock. In contrast, sector 2 ("downstream") produces the final good using both capital and the intermediate good but is not subject to any productivity shocks. The planner is endowed with one unit of capital K which must be allocated between the two sectors K_1 and K_2. A key feature of the problem is that the capital allocation decision must be made before the realization of the total factor productivity shock, so we have an allocation decision under uncertainty. Once the shock is realized, no further decisions are required, as the quantities of goods produced and consumed are fully determined by the preallocated capital. Let Q_i, K_i represents gross-output and capital in sector $i = \{1, 2\}$, and C represents consumption of the final good. The problem is defined by the following equations:

$$K_1 + K_2 = 1, \quad Q_1 = AK_1, \quad Q_2 = \left((1-\gamma_q) (Q_1)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q (K_2)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1}}, \quad C = Q_2$$

where A is a stochastic variable, which takes a low value $A_L = 1 - \Delta_A < 1$ with probability $p = 1/2$ and a high value $A_H = 1 + \Delta_A > 1$ with probability $1 - p = 1/2$ ($\mathbb{E}[A] = 1$), and σ_q is the elasticity of substitution in production between the two goods. The planner maximizes:

$$\mathbb{E}[U(C)] = \mathbb{E}\left[\frac{C^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} \right]$$

where ϵ_c is the intertemporal elasticity of substitution (so ϵ_c^{-1} is risk aversion).

A.2 First-order Conditions

If we replace $K_2 = 1 - K_1$, we have an unconstrained problem with one control variable, K_1. Specifically, consumption in each scenario can be written as

$$C_S = \left((1-\gamma_q) (A_s K_1)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q (1 - K_1)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1}}, \quad S \in \{L, H\}$$

The objective function is:

$$\begin{aligned} \mathbb{E}[U(C)] &= E\left[\frac{C^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} \right] \\ &= p \frac{C_L^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} + (1-p) \frac{C_H^{1-\epsilon_c^{-1}}}{1-\epsilon_c^{-1}} \end{aligned}$$

Before we calculate the first order condition (F.O.C.), we will calculate the derivative of C with respect to K_1:

$$\begin{aligned} \frac{\partial C_s}{\partial K_1} &= \frac{\sigma_q}{\sigma_q - 1} \left((1-\gamma_q) (A_S K_1)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q (1 - K_1)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1} - 1} \\ &\quad \times \left(\frac{\sigma_q - 1}{\sigma_q} \right) \left((1-\gamma_q) (A_S K_1)^{\frac{-1}{\sigma_q}} A_S - \gamma_q (1 - K_1)^{\frac{-1}{\sigma_q}} \right) \qquad , \quad S \in \{L, H\} \\ &= C_S^{\frac{1}{\sigma_q}} \left((1-\gamma_q) (A_S K_1)^{\frac{-1}{\sigma_q}} A_S - \gamma_q (1 - K_1)^{\frac{-1}{\sigma_q}} \right) \qquad , \quad S \in \{L, H\} \\ &= \left((1-\gamma_q) \left(\frac{A_S K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} A_S - \gamma_q \left(\frac{1 - K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} \right) \qquad , \quad S \in \{L, H\} \end{aligned}$$

Then, the F.O.C. is:

$$\begin{aligned} \frac{\partial \mathbb{E}[U(C)]}{\partial K_1} &= p \frac{1}{C_L^{\epsilon_c^{-1}}} \left((1-\gamma_q) \left(\frac{A_L K_1}{C_L} \right)^{\frac{-1}{\sigma_q}} A_L - \gamma_q \left(\frac{1 - K_1}{C_L} \right)^{\frac{-1}{\sigma_q}} \right) \\ &\quad + (1-p) \frac{1}{C_H^{\epsilon_c^{-1}}} \left((1-\gamma_q) \left(\frac{A_H K_1}{C_H} \right)^{\frac{-1}{\sigma_q}} A_S - \gamma_q \left(\frac{1 - K_1}{C_H} \right)^{\frac{-1}{\sigma_q}} \right) \\ &= 0 \end{aligned}$$

A.3 Proof for Proposition 1: Preallocation toward the upstream sector

While in the optimum we have that $\partial \mathbb{E}[U(C)]/\partial K_1 = 0$, we can use the expression of the marginal expected utility to evaluate how optimal capital allocation in the uncertainty case compares to the optimum for the deterministic case. To get the deterministic solution, we set $A = 1$. Then, the FOC is:

$$\left((1-\gamma_q) \left(\frac{K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} - \gamma_q \left(\frac{1-K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} \right) = 0$$

or

$$\begin{aligned}
(1-\gamma_q) \left(\frac{K_1}{C_S} \right)^{\frac{-1}{\sigma_q}} &= \gamma_q \left(\frac{1-K_1}{C_S} \right)^{\frac{-1}{\sigma_q}}, \\
\left(\frac{K_1}{1-K_1} \right)^{\frac{-1}{\sigma_q}} &= \frac{\gamma_q}{1-\gamma_q}, \\
\frac{K_1}{1-K_1} &= \left(\frac{1-\gamma_q}{\gamma_q} \right)^{\sigma_q}, \\
K_1 &= \left(\frac{1-\gamma_q}{\gamma_q} \right)^{\sigma_q} - K_1 \left(\frac{1-\gamma_q}{\gamma_q} \right)^{\sigma_q}, \\
K_1 &= \frac{\left(\frac{1-\gamma_q}{\gamma_q} \right)^{\sigma_q}}{\left(\frac{1-\gamma_q}{\gamma_q} \right)^{\sigma_q} + 1}, \\
K_1 &= \frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}}.
\end{aligned}$$

Now we evaluate C_S at the deterministic capital allocation

$$\begin{aligned}
C_S|_{K_1^{determ}} &= \left((1-\gamma_q) \left(A_S \frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}} \right)^{\frac{\sigma_q - 1}{\sigma_q}} + \gamma_q \left(1 - \frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}} \right)^{\frac{\sigma_q - 1}{\sigma_q}} \right)^{\frac{\sigma_q}{\sigma_q - 1}} \\
&= \left(\frac{1}{\gamma_q^{\sigma_q} + (1-\gamma_q)^{\sigma_q}} \right) \left(A_S^{1-\frac{1}{\sigma_q}} (1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q} \right)^{\frac{\sigma_q}{\sigma_q - 1}}
\end{aligned}$$

Next, we evaluate the derivative:

$$\begin{aligned}
\left.\frac{\partial C}{\partial K_1}\right|_{K_1^{determ}} &= \left((1-\gamma_q)\left(\frac{A_S\frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}}}{C_S}\right)^{\frac{-1}{\sigma_q}} A_S - \gamma_q\left(\frac{1-\frac{(1-\gamma_q)^{\sigma_q}}{\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}}}{C_S}\right)^{\frac{-1}{\sigma_q}}\right) \\
&= \left[\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}\right]^{\frac{1}{\sigma_q}} \frac{\left(A_S^{1-\frac{1}{\sigma_q}}-1\right)}{C_S^{-\frac{1}{\sigma_q}}} \\
&= \left[\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}\right]^{\frac{1}{\sigma_q}} \frac{\left(A_S^{1-\frac{1}{\sigma_q}}-1\right)}{\left(\left(\frac{1}{\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}}\right)\left(A_S^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\sigma_q}{\sigma_q-1}}\right)^{-\frac{1}{\sigma_q}}} \\
&= \frac{\left(A_S^{1-\frac{1}{\sigma_q}}-1\right)}{\left(A_S^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}}
\end{aligned}$$

The FOC evaluated at the deterministic optimum is

$$\begin{aligned}
\left.\frac{\partial \mathbb{E}[U(C)]}{\partial K_1}\right|_{K_1^{determ}} &= p\frac{1}{\left.C_L^{\epsilon_c^{-1}}\right|_{K_1^{determ}}}\left.\frac{\partial C_L}{\partial K_1}\right|_{K_1^{determ}} + (1-p)\frac{1}{\left.C_H^{\epsilon_c^{-1}}\right|_{K_1^{determ}}}\left.\frac{\partial C_H}{\partial K_1}\right|_{K_1^{determ}} \\
&= p\left[\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}\right]^{\epsilon_c^{-1}}\frac{\left(A_L^{1-\frac{1}{\sigma_q}}-1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}} \\
&\quad +(1-p)\left[\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}\right]^{\epsilon_c^{-1}}\frac{\left(A_H^{1-\frac{1}{\sigma_q}}-1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}}, \\
&\propto p\frac{\left(A_L^{1-\frac{1}{\sigma_q}}-1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}} + (1-p)\frac{\left(A_H^{1-\frac{1}{\sigma_q}}-1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}},
\end{aligned}$$

where in the last line we take out the positive constant $\left[\gamma_q^{\sigma_q}+(1-\gamma_q)^{\sigma_q}\right]^{\epsilon_c^{-1}}$.

If the derivative of upstream capital on expected utility, evaluated at the deterministic allocation, is positive, then it is optimal for the social planner to allocate more capital to the upstream sector. Thus, by setting $p = 1 - p = \frac{1}{2}$, the necessary and sufficient condition for allocating extra capital to sector 1 is

$$\frac{(1+\Delta_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\Delta_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} > \frac{1 - (1-\Delta_A)^{1-\sigma_q^{-1}}}{\left[(1-\Delta_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}}. \quad (12)$$

The following proposition 1 fully characterizes the conditions for the above inequality to hold when $\sigma_q < 1$.

Proposition. *If inputs are complements ($\sigma_q < 1$), the planner preallocates capital to the upstream sector (that is, $K_1^* > K_1^{determ}$) under any of the following conditions, up to the first-order approximation:*

1. *If risk aversion is high enough, $\epsilon_c^{-1} \geq 1$.*
2. *If risk aversion is moderate $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q) < \epsilon_c^{-1} < 1$.*
3. *If risk aversion is low, $0 < \epsilon_c^{-1} < \bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q)$, and the downstream sector's importance in production is sufficiently high, $\gamma_q > \bar{\gamma}_q(\triangle_A)$.*

The thresholds $\bar{\gamma}_q(\triangle_A)$ and $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q)$ are determined by the size of productivity shocks and production parameters.

Proof. Our target is to discuss the conditions for the following inequality to hold when $\sigma_q < 1$.

$$F \equiv \frac{\left((1-\triangle_A)^{1-\frac{1}{\sigma_q}} - 1\right)}{\left((1-\triangle_A)^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} + \frac{\left((1+\triangle_A)^{1-\frac{1}{\sigma_q}} - 1\right)}{\left((1+\triangle_A)^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} > 0$$

We apply first-order approximation around $\sigma_q = 1$. As the appendix A.6.1 shows,

$$\lim_{\sigma_q \to 1^+} F = 0 = \lim_{\sigma_q \to 1^-} F \quad \forall \epsilon_c^{-1} \geq 0$$

and

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q = 1} = (1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \frac{\log(1+\triangle_A)}{D(\epsilon_c^{-1}, \gamma_q)} + (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \frac{\log(1-\triangle_A)}{D(\epsilon_c^{-1}, \gamma_q)}$$

where $D(\epsilon_c^{-1}, \gamma_q)$ is strictly positive for any $(\epsilon_c^{-1}, \gamma_q) \in [0,1] \times [0,1]$. Therefore, the sign of the first-order derivative depends on the values of $\triangle_A$, γ_q, and ϵ_c^{-1}.

When $\epsilon_c^{-1} = 1$, due to the strict concavity of log function, given $\triangle_A \in (0,1)$

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=0} = \log(1+\triangle_A) + \log(1-\triangle_A) < 1$$

When $\epsilon_c^{-1} > 1$,

$$\begin{aligned}
\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} &< (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \frac{\log(1+\triangle_A)}{D} + (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \frac{\log(1-\triangle_A)}{D} \\
&= (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \Big[\frac{\log(1+\triangle_A)}{D} + \frac{\log(1-\triangle_A)}{D} \Big] \\
&< (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} \Big[\frac{\log(1)}{D} + \frac{\log(1)}{D} \Big] = 0
\end{aligned}$$

Since $\log(1+\triangle_A) > 0 > \log(1-\triangle_A)$ and $(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)} > (1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}$, by adding more positive numbers, we get the first inequality. The second strict inequality is again due to the strict concavity of log function.

Lastly, when $\epsilon_c^{-1} < 1$, we first note that

$$\lim_{\gamma_q \to 0} \lim_{\epsilon_c^{-1} \to 0^+} \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} > 0, \quad \lim_{\gamma_q \to 0} \lim_{\epsilon_c^{-1} \to 1^-} \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0 \quad \text{(Opposite Signs on } \epsilon_c^{-1}\text{'s boundary)}$$

$$\lim_{\gamma_q \to 0} \lim_{\epsilon_c^{-1} \to 0^+} \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} > 0, \quad \lim_{\gamma_q \to 1} \lim_{\epsilon_c^{-1} \to 0^+} \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0 \quad \text{(Opposite Signs on } \gamma_q\text{'s boundary)}$$

These two pairs of opposite signs indicate the existence of threshold values for γ_q and ϵ_c^{-1}. Notice that $\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1}$ is continuously differentiable and strictly monotone in ϵ_c^{-1} and γ_q when $\triangle_A \in (0,1)$. Therefore, we can apply Intermediate Value Theorem (IVT) to show that given any $\triangle_A \in (0,1)$, there exists a unique threshold value $\bar{\gamma}_q(\triangle_A)$ such that

1. if $\gamma_q > \bar{\gamma}_q(\triangle_A)$, $\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0$ when $\epsilon_c^{-1} < 1$. Together with previous discussion, $\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0 \ \forall \epsilon_c^{-1} \geq 0$.

2. if $\gamma_q \leq \bar{\gamma}_q(\triangle_A)$, by applying IVT again, there exists a unique threshold value $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q) \in$

$(0,1)$ such that

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} \begin{cases} > 0 & \forall \epsilon_c^{-1} < \bar{\epsilon}_c^{-1}(\gamma_q, \triangle_A) \\ = 0 & \forall \epsilon_c^{-1} = \bar{\epsilon}_c^{-1}(\gamma_q, \triangle_A) \\ < 0 & \forall \epsilon_c^{-1} \in (\bar{\epsilon}_c^{-1}(\gamma_q, \triangle_A), 1) \end{cases}$$

Finally,

$$F \approx 0 + \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} \times (\sigma_q - 1)$$

When either $\gamma_q > \bar{\gamma}_q(\triangle_A)$ or $[\gamma_q \leq \bar{\gamma}_q(\triangle_A)] \wedge [\epsilon_c^{-1} \in (\bar{\epsilon}_c^{-1}(\gamma_q, \triangle_A), 1)]$, $\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0$, and thus F is bigger than 0 if and only if $\sigma_q < 1$ up to the first-order approximation. □

A.4 Proof for Proposition 2: Preallocation upstream reduces the impact of negative shocks

Next, we will analyze the implications of preallocating capital upstream for the impact of a productivity shock. We define a productivity shock as moving from $A = \mathbb{E}[A] = 1$ to A_L. Furthermore, we define the absolute value of the impulse response of consumption, as a function of original capital stock in sector 1, as:

$$IR(K_1, A_L) \equiv |C(K_1, A_L) - C(K_1, \mathbb{E}[A])| = C(K_1, \mathbb{E}[A]) - C(K_1, A_L)$$

Following this definition, we will prove in our next proposition that the absolute value of the impulse response of consumption to a negative shock is smaller when we set capital to the optimum under uncertainty (K_1^*) versus the deterministic optimum (K_1^{determ}). To do that, we will follow the same strategy as in our main proposition, which is to evaluate the derivate of $IR(K_1, A_L)$ with respect to K_1 at the deterministic optimum K_1^{determ}, and show that when $\sigma_q < 1$, we have that $\frac{\partial IR(K_1, A_L)}{\partial K}\Big|_{K_1 = K_1^{determ}} < 0$. The interpretation of this result is that preallocation towards the upstream sector reduces the impact of negative shocks in consumption.

Proposition. *If inputs are complements ($\sigma_q < 1$), then increasing capital allocation to the upstream sector beyond the deterministic optimum decreases the impulse response to negative productivity shocks:*

$$\frac{\partial IR(K_1, A_L)}{\partial K_1}\Big|_{K_1 = K_1^{determ}} < 0$$

where $IR(K_1, A_L)$ measures the percentage change in consumption resulting from a negative

productivity shock.

Proof. We start by building $IR(K_1^{determ})$ using our expression for $C_S|_{K_1=K_1^{determ}}$:

$$\begin{aligned}
\left.\frac{\partial IR(K_1, A_L)}{\partial K_1}\right|_{K_1=K_1^{determ}} &= \left.\frac{\partial C(K_1, \mathbb{E}[A])}{\partial K_1}\right|_{K_1=K_1^{determ}} - \left.\frac{\partial C(K_1, A_L)}{\partial K_1}\right|_{K_1=K_1^{determ}} \\
&= \frac{\left(1^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(1^{1-\frac{1}{\sigma_q}} (1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} - \frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}} (1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} \\
&= -\frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}} (1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}}.
\end{aligned}$$

Since $A_L \in (0,1)$, we have that

$$\sigma_q < 1 \Leftrightarrow \left(A_L^{1-\sigma_q^{-1}} - 1\right) > 0 \Leftrightarrow \left.\frac{\partial IR(K_1, A_L)}{\partial K_1}\right|_{K_1=K_1^{determ}} < 0.$$

□

In the case of positive shocks

$$\begin{aligned}
\left.\frac{\partial IR(K_1, A_H)}{\partial K_1}\right|_{K_1=K_1^{determ}} &= \left.\frac{\partial C(K_1, A_H)}{\partial K_1}\right|_{K_1=K_1^{determ}} - \left.\frac{\partial C(K_1, \mathbb{E}[A])}{\partial K_1}\right|_{K_1=K_1^{determ}} \\
&= \frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}} (1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}},
\end{aligned}$$

and using a similar reasoning as above, we get

$$\sigma_q < 1 \Leftrightarrow \left(A_H^{1-\sigma_q^{-1}} - 1\right) < 0 \Leftrightarrow \left.\frac{\partial IR(K_1)}{\partial K_1}\right|_{K_1=K_1^{determ}} < 0.$$

Therefore preallocating capital upstream also dampens the effect of positive shocks.

A.5 Proof for Proposition 3: Preallocation upstream reduces the expected consumption

Moreover, we can explore how preallocation towards upstream sector affects the mean of aggregate consumption under different conditions by following the same strategy of proof.

Proposition. *Up to the first-order approximation, if inputs are complements ($\sigma_q < 1$), the effect of upstream preallocation on expected consumption depends on the relative strength of the upstream sector:*

1. *When the upstream sector is large enough ($\gamma_q < \bar{\gamma}_q(\triangle_A)$), increasing upstream capital allocation beyond the deterministic optimum decreases expected consumption:*

$$\frac{\partial \mathbb{E}\{C(A, K_1)\}}{\partial K_1}\Big|_{K_1 = K_1^{determ}} < 0.$$

2. *When the upstream sector's role is more limited ($\gamma_q > \bar{\gamma}_q(\triangle_A)$), increasing upstream capital allocation increases expected consumption.*

Proof. We start by showing that

$$\begin{aligned}
\mathbf{E}\{C(A, K_1)\} &= pC(A_L, K_1) + (1-p)C(A_H, K_1) \\
\frac{\partial \mathbf{E}\{C(A, K_1)\}}{\partial K_1}|_{K_1 = K_1^{determ}} &= p\frac{\partial C(A_L, K_1)}{\partial K_1}|_{K_1 = K_1^{determ}} + (1-p)\frac{\partial C(A_H, K_1)}{\partial K_1}|_{K_1 = K_1^{determ}} \\
&= p\frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} + (1-p)\frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}}
\end{aligned}$$

Following our previous assumption, we set $p = 1/2$, thus our target is to analyze the sign of

$$F \equiv \frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} + \frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}}$$

Notice that this is a special case of proposition 1 where $\epsilon_c^{-1} = 0$. By taking a second-order

local approximation for F around $\sigma_q = 1$, as the appendix A.6.1 shows, we have

$$\lim_{\sigma_q \to 1} F = 0$$

and

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} = (1 + \triangle_A)^{1-\gamma_q} \log(1 + \triangle_A) + (1 - \triangle_A)^{1-\gamma_q} \log(1 - \triangle_A)$$

By Intermediate Value Theorem, there exists a unique threshold value $\bar{\gamma}_q(\triangle_A)$ such that

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} \begin{cases} > 0 & \text{if } \gamma_q < \bar{\gamma}_q(\triangle_A) \\ = 0 & \text{if } \gamma_q = \bar{\gamma}_q(\triangle_A) \\ < 0 & \text{if } \gamma_q > \bar{\gamma}_q(\triangle_A) \end{cases}$$

Since the values on the boundary are strictly bounded away from 0, $\bar{\gamma}_q(\triangle_A) \in (0, 1)$. Finally, when $\gamma_q > \bar{\gamma}_q(\triangle_A)$,

$$F \approx 0 + \underbrace{\frac{\partial}{\partial \sigma_q}|_{\sigma_q=1} F}_{<0} \times (\sigma_q - 1)$$

F is bigger than 0 if and only if $\sigma_q < 1$ up to the first-order approximation. □

A.6 Full proofs

A.6.1 Proof for Proposition 1

Recall from A.3 that our target is to discuss the sign of the following objective function when $\sigma_q < 1$,

$$F \equiv \frac{(1+\Delta_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\Delta_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} + \frac{(1-\Delta_A)^{1-\sigma_q^{-1}} - 1}{\left[(1-\Delta_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}}$$

By applying first-order Taylor approximation around $\sigma_q = 1$, we proceed with the whole proof in two general steps.

1. Firstly, we show that in a Cobb-Douglass economy,

$$\lim_{\sigma_q \to 1} F = 0$$

   Extra allocation of capital to certain sector does not exist.

2. Secondly, given the first-order approximation,

$$F \approx 0 + \frac{\partial}{\partial \sigma_q} F|_{\sigma_q = 1}(\sigma_q - 1)$$

   it suffices to discuss the conditions for the following to hold

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q = 1} < 0$$

Step 1 for Proposition 1

When $\epsilon_c^{-1} = 1$

$$\lim_{\sigma_q \to 1} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]} = 0 = \lim_{\sigma_q \to 1} \frac{1 - (1-\triangle_A)^{1-\sigma_q^{-1}}}{\left[(1-\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]}$$

When $\epsilon_c^{-1} > 1$ First note that

$$\lim_{\sigma_q \to 1} \frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1} = DNE$$

since

$$\lim_{\sigma_q \to 1^+} \frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1} = \infty \quad \lim_{\sigma_q \to 1^-} \frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1} = -\infty$$

We need to check the limit for limits from both sides.

1. If we approach 1 from the right hand-side, we have the indeterminate form $(\frac{0}{1^\infty})$. We first find the limit for the denominator (1^∞)

$$\begin{aligned}
\lim_{\sigma_q \to 1^+} f &= \lim_{\sigma_q \to 1^+} \Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]^{\frac{\epsilon^{-1}\sigma_q - 1}{\sigma_q - 1}} \\
\lim_{\sigma_q \to 1^+} \log(f) &= \lim_{\sigma_q \to 1^+} \frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1} \log\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big] \\
&= \lim_{\sigma_q \to 1^+} \frac{\overbrace{\log\Big[(1+\triangle_A)^{1-\sigma^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma}\Big]}^{0}}{\underbrace{\frac{\sigma_q - 1}{\epsilon_c^{-1}\sigma_q - 1}}_{0}} \\
&= \lim_{\sigma_q \to 1^+} \frac{\frac{\Big\{\log(\gamma_q)\gamma_q^{\sigma_q} + \Big[\log(1-\gamma_q)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}} + \log(1+\triangle_A)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2})\Big]\Big\}}{\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]}}{\frac{\epsilon_c^{-1}-1}{(\epsilon_c^{-1}\sigma_q - 1)^2}} \\
&= \frac{\log(\gamma_q)\gamma_q + \Big[\log(1-\gamma_q)(1-\gamma_q) + \log(1+\triangle_A)(1-\gamma_q)\Big]}{(\epsilon_c^{-1}-1)^{-1}}
\end{aligned}$$

   where the last second equality is due to L'hospital rule, and

$$\begin{aligned}
&\frac{\partial}{\partial \sigma_q}\Big[(1+\triangle_A)^{1-\sigma_c^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big] \\
&= \log(\gamma_q)\gamma_q^{\sigma_q} + \Big[\log(1-\gamma_q)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}} + \log(1+\triangle_A)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2})\Big]
\end{aligned}$$

   Therefore,

$$\lim_{\sigma_q \to 1^+} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} = 0$$

   The same applies for the RHS when $A = 1 - \triangle_A$.

2. If we approach 1 from the left hand-side, we have the indeterminate form (01^∞). Based

on the above calculation,

$$\lim_{\sigma_q \to 1^-} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} = 0$$

So, when $\epsilon_c^{-1} > 1$,

$$\lim_{\sigma_q \to 1^-} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} = 0 = \lim_{\sigma_q \to 1^+} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q{}^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}}$$

and thus

$$\lim_{\sigma_q \to 1} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma^{\sigma}\right]^{\frac{\epsilon^{-1}\sigma_q - 1}{\sigma_q - 1}}} = \lim_{\sigma_q \to 1} \frac{1-(1-\triangle_A)^{1-\sigma^{-1}}}{\left[(1-\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} \tag{13}$$

When $\epsilon_c^{-1} < 1$ We can carry out the same calculation for limits from both sides as the above showed.

Therefore, given any $\epsilon_c^{-1} \geq 0$,

$$\lim_{\sigma_q \to 1^-} F = 0 = \lim_{\sigma_q \to 1^+} F$$

Step 2 for Proposition 1

In this step, we discuss the sign of the limit of the following sum of two derivatives by rewriting it as the sum of the limits of two derivatives.

$$\lim_{\sigma_q \to 1} \left\{ \frac{\partial}{\partial \sigma_q} \frac{(1+\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1+\triangle_A)^{1-\sigma_q{}^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right]^{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1}}} + \frac{\partial}{\partial \sigma_q} \frac{(1-\triangle_A)^{1-\sigma_q^{-1}} - 1}{\left[(1-\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma)^{\sigma_q} + \gamma^{\sigma_q}\right]^{\frac{\epsilon^{-1}\sigma_q - 1}{\sigma_q - 1}}} \right\}$$
$$= \lim_{\sigma_q \to 1} \frac{\Omega_1^1 - \Omega_2^1}{D_1^2} + \lim_{\sigma_q \to 1} \frac{\Omega_1^2 - \Omega_2^2}{D_2^2} \tag{14}$$

where we define

$$D_1 \equiv \Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\Big]^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}} \quad , \quad D_2 \equiv \Big[(1-\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q}+\gamma_q^{\sigma_q}\Big]^{\frac{\epsilon_c^{-1}\sigma_q-1}{\sigma_q-1}}$$

$$\Omega_1^1 \equiv [\frac{\partial}{\partial\sigma_q}((1+\triangle_A)^{1-\sigma_q^{-1}}-1)]D_1 \quad , \quad \Omega_1^2 \equiv [\frac{\partial}{\partial\sigma_q}((1-\triangle_A)^{1-\sigma_q^{-1}}-1)]D_2$$

$$\Omega_2^1 \equiv ((1+\triangle_A)^{1-\sigma_q^{-1}}-1)(\frac{\partial}{\partial\sigma_q}D_1) \quad , \quad \Omega_2^2 \equiv ((1-\triangle_A)^{1-\sigma_q^{-1}}-1)(\frac{\partial}{\partial\sigma_q}D_2)$$

The following proof of step 2 is proceeded by firstly calculating the two limits separately and then discussing the conditions for their sum to be positive. WLOG, we offer a detailed explanation for finding the limits of $D_1, \Omega_1^1, \Omega_2^1$ as $\sigma_q \to 1$. The symmetric result then follows for $D_2, \Omega_1^2, \Omega_2^2$.

We start by calculating the limit for D_1. Due to L'hospital rule,

$$\lim_{\sigma_q\to 1} D_1 = \exp\Big\{\frac{\log(\gamma_q)\gamma_q+\Big[\log(1-\gamma_q)(1-\gamma_q)+\log(1+\triangle_A)(1-\gamma_q)\Big]}{(\epsilon_c^{-1}-1)^{-1}}\Big\}$$

We therefore have

$$\lim_{\sigma_q\to 1}\Omega_1^1 = \lim_{\sigma_q\to 1}\log(1+\triangle_A)(1+\triangle_A)^{1-\sigma_q^{-1}}\sigma_q^{-2}D_1 = \log(1+\triangle_A)\lim_{\sigma_q\to 1} D_1$$

For the limit of Ω_2^1, since

$$\begin{aligned}\frac{\partial}{\partial x}f(x)^{g(x)} &= \frac{\partial}{\partial x}e^{g(x)\log(f(x))} \\ &= f(x)^{g(x)}\Big[g'(x)\log(f(x))+g(x)\frac{f'(x)}{f(x)}\Big]\end{aligned}$$

we have

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma_q} D_1 \\
&= \underbrace{\exp\Big\{ \frac{\epsilon_c^{-1}\sigma - 1}{\sigma_q - 1} \log\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]\Big\}}_{g_1^1(\sigma_q)=D_1} \\
&\times \Bigg\{ \underbrace{\frac{\epsilon_c^{-1}(\sigma-1) - (\epsilon_c^{-1}\sigma_q - 1)}{(\sigma_q-1)^2} \log\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]}_{g_2^1(\sigma_q)} \\
&+ \underbrace{\frac{\epsilon_c^{-1}\sigma_q - 1}{\sigma_q - 1} \frac{\Big[\log(\gamma_q)\gamma_q^{\sigma_q} + \log(1-\gamma_q)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}} + \log(1+\triangle_A)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2})\Big]}{(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}}}_{g_3^1(\sigma_q)} \Bigg\}
\end{aligned}
$$

So we can express the limit of Ω_2^1 as

$$
\lim_{\sigma_q \to 1} \Omega_2^1 = \lim_{\sigma \to 1} D_1 \lim_{\sigma_q \to 1} \Big[\underbrace{((1+\triangle_A)^{1-\sigma_q^{-1}} - 1) g_2^1(\sigma_q)}_{(1)} + \underbrace{((1+\triangle_A)^{1-\sigma_q^{-1}} - 1) g_3^1(\sigma_q)}_{(2)} \Big]
$$

We first discuss the limit of component (1) from both sides. By applying L'hospital rule,

$$
\begin{aligned}
&(1-\epsilon_c^{-1}) \lim_{\sigma_q \to 1^+} \frac{\overbrace{\Big[(1+\triangle_A)^{1-\sigma_q^{-1}} - 1\Big]}^{0^+} \overbrace{\log\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma}\Big]}^{0^-}}{\underbrace{(\sigma_q - 1)^2}_{0^+}} \\
&= (1-\epsilon_c^{-1}) \log(1+\triangle_A) \lim_{\sigma_q \to 1} \frac{\log\Big[(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\Big]}{2(\sigma_q - 1)} + (1-\epsilon_c^{-1}) \lim_{\sigma_q \to 1} g_4^1 \frac{\Big[(1+\triangle_A)^{1-\sigma_q^{-1}} - 1\Big]}{2(\sigma_q - 1)} \\
&= \frac{1}{2}(1-\epsilon_c^{-1}) \log(1+\triangle_A) \lim_{\sigma_q \to 1} g_4^1 + \frac{1}{2}(1-\epsilon^{-1}) \lim_{\sigma_q \to 1} g_4^1 \lim_{\sigma_q \to 1} \log(1+\triangle_A)(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2}) \\
&= (1-\epsilon_c^{-1}) \log(1+\triangle_A) \lim_{\sigma_q \to 1} g_4^1
\end{aligned}
$$

where

$$
g_4^1 = \frac{1}{(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma^{\sigma}} \Bigg\{ \log(\gamma_q)\gamma_q + \Big[\log(1-\gamma_q)(1-\gamma_q) + \log(1+\triangle_A)(1-\gamma_q)\Big] \Bigg\}
$$

The same applies for $\sigma_q \to 1^-$.

We then discuss the limit of component (2) from both sides. Here we apply L'hospital rule twice.

$$
\begin{aligned}
&\lim_{\sigma_q \to 1^+} \Big[(1+\triangle_A)^{1-\sigma_q^{-1}} - 1\Big] g_3^1(\sigma_q) \\
&= \lim_{\sigma_q \to 1^+} \frac{\Big\{\log(\gamma_q)\gamma_q^{\sigma_q} + \Big[\log(1-\gamma_q)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}} + \log(1+\triangle_A)(1-\gamma_q)^{\sigma_q}(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2})\Big]\Big\}}{(1+\triangle_A)^{1-\sigma_q^{-1}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}} \\
&\times \lim_{\sigma_q \to 1^+} \frac{\overbrace{\Big[(1+\triangle_A)^{1-\sigma_q^{-1}} - 1\Big]}^{0^+}}{\underbrace{\frac{\sigma_q - 1}{\epsilon_c^{-1}\sigma_q - 1}}_{0^+}} \\
&= \left\{\log(\gamma_q)\gamma_q + \Big[\log(1-\gamma_q)(1-\gamma_q) + \log(1+\triangle_A)(1-\gamma_q)\Big]\right\} \frac{\log(1+\triangle_A)(1+\triangle_A)^{1-\sigma_q^{-1}}(\sigma_q^{-2})}{\frac{\epsilon_c^{-1}-1}{(\epsilon_c^{-1}\sigma_q - 1)^2}} \\
&= -\lim_{\sigma_q \to 1} g_4^1 \frac{\log(1+\triangle_A)}{\frac{1}{1-\epsilon_c^{-1}}}
\end{aligned}
$$

The same applies for $\sigma_q \to 1^-$.

And, the calculations for (limits of) $D_1, \Omega_1^1, \Omega_2^1$ can also apply to $D_2, \Omega_1^2, \Omega_2^2$. Therefore, by combining the above results,

$$
\lim_{\sigma_q \to 1} \Omega_2^1 = \lim_{\sigma_q \to 1} g_1^1(\sigma_q) \Big[\log(1+\triangle_A) \frac{\lim_{\sigma_q \to 1} g_4^1}{\frac{1}{1-\epsilon_c^{-1}}} - \log(1+\triangle_A) \frac{\lim_{\sigma_q \to 1} g_4^1}{\frac{1}{1-\epsilon_c^{-1}}}\Big] = 0
$$

$$
\lim_{\sigma_q \to 1} \Omega_2^2 = \lim_{\sigma_q \to 1} g_1^2(\sigma_q) \Big[\log(1-\triangle_A) \frac{\lim_{\sigma_q \to 1} g_4^2}{\frac{1}{1-\epsilon_c^{-1}}} - \log(1-\triangle_A) \frac{\lim_{\sigma_q \to 1} g_4^2}{\frac{1}{1-\epsilon_c^{-1}}}\Big] = 0
$$

All of the results above hold for all $\epsilon_c^{-1} \geq 0$

Based on the previous calculation, our original target equation 14 is reduced to

$$
\begin{aligned}
&\lim_{\sigma_q \to 1} \frac{\Omega_1^1 - \Omega_2^1}{D_1^2} + \lim_{\sigma_q \to 1} \frac{\Omega_1^2 - \Omega_2^2}{D_2^2} \\
&= \lim_{\sigma_q \to 1} \frac{\Omega_1^1}{D_1^2} + \lim_{\sigma_q \to 1} \frac{\Omega_1^2}{D_2^2}
\end{aligned}
$$

We can now begin establishing our ideal results by splitting the discussion into three cases according to the partitions of risk aversion.

Case 1: $\epsilon_c^{-1} = 1$

$$\lim_{\sigma_q \to 1} D_1 = \lim_{\sigma_q \to 1} D_2 = 1$$

From previous calculation

$$\lim_{\sigma_q \to 1} \Omega_1^1 = \log(1 + \triangle_A) \lim_{\sigma_q \to 1} D_1$$

$$\lim_{\sigma_q \to 1} \Omega_1^2 = \log(1 - \triangle_A) \lim_{\sigma_q \to 1} D_2$$

We thus have the objective function as

$$\begin{aligned} \log(1 + \triangle_A) + \log(1 - \triangle_A) &= \frac{1}{2}(\log(1 + \triangle_A) + \log(1 - \triangle_A)) + \frac{1}{2}(\log(1 + \triangle_A) + \log(1 - \triangle_A)) \\ &< \log(1) + \log(1) = 0 \end{aligned}$$

due to the strict concavity of $\log(\cdot)$, establishing the result.

Case 2: $\epsilon_c^{-1} > 1$

$$\begin{aligned} \lim_{\sigma_q \to 1} D_1 &= \exp\Big(\frac{\log(\gamma_q)\gamma_q + \Big[\log(1 - \gamma_q)(1 - \gamma_q) + \log(1 + \triangle_A)(1 - \gamma_q)\Big]}{(\epsilon_c^{-1} - 1)^{-1}}\Big) \\ &= \exp\Big(\frac{\log(\gamma_q)\gamma_q + \log(1 - \gamma_q)(1 - \gamma_q)}{(\epsilon_c^{-1} - 1)^{-1}}\Big) \exp\Big(\frac{\log(1 + \triangle_A)(1 - \gamma_q)}{(\epsilon_c^{-1} - 1)^{-1}}\Big) \end{aligned}$$

$$\begin{aligned} \lim_{\sigma_q \to 1} D_2 &= \exp\Big(\frac{\log(\gamma_q)\gamma_q + \Big[\log(1 - \gamma_q)(1 - \gamma_q) + \log(1 - \triangle_A)(1 - \gamma_q)\Big]}{(\epsilon_c^{-1} - 1)^{-1}}\Big) \\ &= D \exp\Big(\frac{\log(1 - \triangle_A)(1 - \gamma_q)}{(\epsilon_c^{-1} - 1)^{-1}}\Big) \end{aligned}$$

For expositional convenience, lets define

$$D \equiv \exp\Big(\frac{\log(\gamma_q)\gamma_q + \log(1 - \gamma_q)(1 - \gamma_q)}{(\epsilon_c^{-1} - 1)^{-1}}\Big)$$

The objective function thus becomes

$$
\begin{aligned}
&\frac{\log(1+\triangle_A)}{D\exp\Big(\frac{\log(1+\triangle_A)(1-\gamma_q)}{(\epsilon_c^{-1}-1)^{-1}}\Big)}+\frac{\log(1-\triangle_A)}{D\exp\Big(\frac{\log(1-\triangle_A)(1-\gamma_q)}{(\epsilon_c^{-1}-1)^{-1}}\Big)}\\
&=(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}\\
&<(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}\\
&=(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\Big[\frac{\log(1+\triangle_A)}{D}+\frac{\log(1-\triangle_A)}{D}\Big]\\
&<(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\Big[\frac{\log(1)}{D}+\frac{\log(1)}{D}\Big]=0
\end{aligned}
$$

where the first inequality comes from

$$
\begin{aligned}
&(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}\\
&=(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}\\
&\underbrace{-\big[(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}-(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\big]\frac{\log(1+\triangle_A)}{D}}_{>0}\\
&<(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}
\end{aligned}
$$

and the second inequality is again due to the strict concavity of log function.

Case 3: $\epsilon_c^{-1}<1$ We still have the objective function

$$
(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D}+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D}
$$

However, now the sign of this equation depends on the choices of ϵ_c^{-1} and γ_q.

We first note that since D is finite and strictly positive over $\epsilon_c^{-1}\in[0,1)$, it suffices to discuss the sign of

$$
(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1+\triangle_A)+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1-\triangle_A)
$$

Secondly, given $\gamma_q\in(0,1)$, the above equation is continuously differentiable and strictly monotone in ϵ_c^{-1}. Moreover, although it is negative when ϵ_c^{-1}, its left boundary value depends

on the choice of γ_q.

$$(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1+\triangle_A)+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1-\triangle_A)|_{\epsilon_c^{-1}=1}<0$$

$$\begin{aligned}&(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1+\triangle_A)+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1-\triangle_A)|_{\epsilon_c^{-1}=0}\\&=(1+\triangle_A)^{1-\gamma_q}\log(1+\triangle_A)+(1-\triangle_A)^{1-\gamma_q}\log(1-\triangle_A)\gtreqless 0\end{aligned}\tag{15}$$

Again, given $\triangle_A\in(0,1)$, the above function is continuously differentiable and strictly monotone in γ_q. As $\gamma_q\to 1$,

$$\log(1+\triangle_A)+\log(1-\triangle_A)<0$$

As $\gamma_q\to 0$

$$(1+\triangle_A)\log(1+\triangle_A)+(1-\triangle_A)\log(1-\triangle_A)>0\quad\triangle_A\in(0,1)$$

The above equation is due to the fact that it equals to 0 at $\triangle_A=0$ and its first-order derivative is uniformly bounded from below by 0 for any $\triangle_A\in(0,1)$

$$\begin{aligned}&\frac{\partial}{\partial\triangle_A}(1+\triangle_A)\log(1+\triangle_A)+(1-\triangle_A)\log(1-\triangle_A)\\&=\frac{1}{1+\triangle_A}+\log(1+\triangle_A)+\frac{\triangle_A}{1+\triangle_A}-\frac{1}{1-\triangle_A}-\log(1-\triangle_A)+\frac{\triangle_A}{1-\triangle_A}\\&=\log(1+\triangle_A)-\log(1-\triangle_A)>0\quad\forall\ \triangle_A>0\end{aligned}$$

Since the values on the boundary of the support exhibit opposite signs and the function is continuously differentiable and monotone in γ_q, because of intermediate value theorem (IVT), the following set is an non-empty and singleton set.

$$\left|\{\gamma_q:(1+\triangle_A)^{1-\gamma_q}\log(1+\triangle_A)+(1-\triangle_A)^{1-\gamma_q}\log(1-\triangle_A)=0,0<\gamma_q<1\}\right|=1$$

Denote this value by $\bar{\gamma}_q$. When $\gamma_q<\bar{\gamma}_q(\triangle_A)$, in the linear utility case, social planner does not allocate extra capital to the upstream sectors. Since the values on the boundary are strictly bounded away from 0, $\bar{\gamma}_q(\triangle_A)\in(0,1)$. We therefore find the critical value for the boundary value of the original function 15 to switch its sign.

After pinning down the sign of boundary value 15 when $\epsilon_c^{-1}=0$, we can characterize the threshold value for ϵ_c^{-1}: for any $\gamma_q\in(0,1)$,

1. If $\gamma_q > \bar{\gamma}_q(\triangle_A)$,

$$(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1+\triangle_A)}{D} + (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\frac{\log(1-\triangle_A)}{D} < 0 \quad \forall \epsilon_c^{-1} \in [0,1)$$

2. If $\gamma_q < \bar{\gamma}_q(\triangle_A)$, then by the same logic, due to IVT, the following set is an non-empty and singleton set.

$$\left|\{\epsilon_c^{-1} : (1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_c)}\log(1+\triangle_A)+(1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1-\triangle_A) = 0, 0 \le \epsilon_c^{-1} \le 1\}\right| = 1$$

Denote this value by $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q)$. Then, in this case

$$\frac{\left[(1+\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1+\triangle_A) + (1-\triangle_A)^{(1-\epsilon_c^{-1})(1-\gamma_q)}\log(1-\triangle_A)\right]}{D} \begin{cases} > 0 & \forall \epsilon_c^{-1} \in [0, \bar{\epsilon}^{-1}(\bar{\gamma}_q(\triangle_A))) \\ = 0 & \forall \epsilon_c^{-1} = \bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q) \\ < 0 & \forall \epsilon_c^{-1} > (\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q), 1) \end{cases}$$

3. When $\gamma_q = \bar{\gamma}_q(\triangle_A)$, extra allocation of capital does not exist if and only if $\epsilon_c^{-1} = 0$ since $\bar{\epsilon}_c^{-1}(\triangle_A, \gamma_q) = 0$.

Finally,

$$F \approx 0 + \frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} \times (\sigma_q - 1)$$

Combining the results from cases 1, 2, and 3, when either $\gamma_q > \bar{\gamma}_q(\triangle_A)$ or $[\gamma_q \le \bar{\gamma}_q(\triangle_A)] \wedge [\epsilon_c^{-1} \in (\bar{\epsilon}_c^{-1}(\gamma_q, \triangle_A), 1)]$, $\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} < 0$, and thus F is bigger than 0 if and only if $\sigma_q < 1$ up to the first-order approximation.

A.6.2 Proof for Proposition 2

Let's start by checking the expression for the IRF

$$\begin{aligned}
\left.\frac{\partial IR(K_1, A_L)}{\partial K_1}\right|_{K_1=K_1^{determ}} &= \left.\frac{\partial C(K_1, \mathbb{E}[A])}{\partial K_1}\right|_{K_1=K_1^{determ}} - \left.\frac{\partial C(K_1, A_L)}{\partial K_1}\right|_{K_1=K_1^{determ}} \\
&= \frac{\left(\mathbb{E}[A]^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(\mathbb{E}[A]^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} - \frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} \\
&= -\frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}}
\end{aligned}$$

Since

$$\begin{aligned}
\sigma_q &< 1 \\
\sigma_q^{-1} &> 1 \\
-\sigma_q^{-1} &< -1 \\
1 - \sigma_q^{-1} &< 0 \\
(A_L)^{1-\sigma_q^{-1}} &> 1
\end{aligned}$$

We have

$$\frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q - 1}}} > 0 \quad \text{iff} \quad \sigma_q < 1$$

thus establishing the result.

We then move to prove that

$$IR(K_1, A_H) \equiv |C(K_1, A_H) - C(K_1, \mathbb{E}[A])| = C(K_1, A_H) - C(K_1, \mathbb{E}[A])$$

$$\left.\frac{\partial IR(K_1, A_H)}{\partial K_1}\right|_{K_1=K_1^{determ}} < 0$$

Note that

$$\frac{\partial IR(K_1, A_H)}{\partial K_1}\Big|_{K_1=K_1^{determ}} = \frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}} - \frac{\left(\mathbb{E}[A]^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(\mathbb{E}[A]^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}}$$
$$= \frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}}$$

By the same logic,

$$\frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}} < 0 \quad \text{iff} \quad \sigma_q < 1$$

We thus establish the second proposition.

A.6.3 Proof for Proposition 3

By definition, it is equivalent to show the sign of

$$F \equiv \frac{\left(A_L^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_L^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}} + \frac{\left(A_H^{1-\frac{1}{\sigma_q}} - 1\right)}{\left(A_H^{1-\frac{1}{\sigma_q}}(1-\gamma_q)^{\sigma_q} + \gamma_q^{\sigma_q}\right)^{\frac{-1}{\sigma_q-1}}}$$

which is a special case of our proof for proposition 1. By taking first-order approximation around $\sigma_q = 1$, we have

$$\lim_{\sigma_q \to 1} F = 0$$

and

$$\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1} = \frac{1}{D}\Big[(1+\triangle_A)^{(1-\gamma_q)}\log(1+\triangle_A) + (1-\triangle_A)^{(1-\gamma_q)}\log(1-\triangle_A)\Big]$$

Following the discussion for case 3 in the proof for the Step 2 of Proposition 1, we know that by intermediate value theorem, there exists a threshold value $\bar{\gamma}_q(\triangle_A)$ such that if $\gamma_q > \bar{\gamma}_q(\triangle_A)$,

$\frac{\partial}{\partial \sigma_q}|_{\sigma_q=1} F$ is smaller than 0. Therefore, we can conclude that when $\gamma_q > \bar{\gamma}_q(\triangle_A)$,

$$F \approx 0 + \underbrace{\frac{\partial}{\partial \sigma_q} F|_{\sigma_q=1}}_{<0} \times (\sigma_q - 1)$$

F is bigger than 0 if and only if $\sigma_q < 1$ up to the first-order approximation.

B First-order conditions and auxiliary results regarding the full model

In this appendix, we derive the first-order conditions of the full model. We also provide some additional results

The Lagrangian of the planner is:

$$\begin{aligned}\mathcal{L} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ & \frac{1}{1-\epsilon_c^{-1}} \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}} \\ & + \sum_{j=1}^{N} P_{jt}^{\mathrm{k}} \left[I_{jt} + (1-\delta_j) K_{jt} - \Phi_{jt} - K_{jt+1} \right] \\ & + \sum_{j=1}^{N} P_{jt} \left[Q_{jt} - C_{jt} - \sum_{i=1}^{N} \left[M_{jit} + I_{jit} \right] \right] \Bigg\},\end{aligned}$$

where

$$
\begin{aligned}
C_t &= \left(\sum_{j=1}^{N} \xi_j^{\sigma_c^{-1}} (C_{jt})^{1-\sigma_c^{-1}} \right)^{\frac{1}{1-\sigma_c^{-1}}}, \\
L_t &= \left(\sum_{j=1}^{N} (L_{jt})^{1+\sigma_l^{-1}} \right)^{\frac{1}{1+\sigma_l^{-1}}}, \\
Q_{jt} &= \left[(\mu_j)^{\sigma_q^{-1}} (Y_{jt})^{1-\sigma_q^{-1}} + (1-\mu_j)^{\sigma_q^{-1}} (M_{jt})^{1-\sigma_q^{-1}} \right]^{\frac{1}{1-\sigma_q^{-1}}}, \\
Y_{jt} &= A_{jt} \left[(\alpha_j)^{\sigma_q^{-1}} (K_{jt})^{1-\sigma_y^{-1}} + (1-\alpha_j)^{\sigma_y^{-1}} (L_{jt})^{1-\sigma_y^{-1}} \right]^{\frac{1}{1-\sigma_y^{-1}}}, \\
I_{jt} &= \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right)^{\sigma_I^{-1}} (I_{ijt})^{1-\sigma_I^{-1}} \right)^{\frac{1}{1-\sigma_I^{-1}}}, \\
M_{jt} &= \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{m}\right)^{\sigma_m^{-1}} (M_{ijt})^{1-\sigma_m^{-1}} \right)^{\frac{1}{1-\sigma_m^{-1}}}, \\
\Phi_{jt} &= \frac{\phi}{2} \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right)^2 K_{jt}.
\end{aligned}
$$

We will start by writing down all the derivatives of the CES aggregators and the adjustment cost function. See sub-appendix B.11 to see the details of the CES algebra involved:

$$
\begin{aligned}
\frac{\partial C_t}{\partial C_{jt}} &= \left(\xi_j \frac{C_t}{C_{jt}} \right)^{\sigma_c^{-1}} & \frac{\partial L_t}{\partial L_{jt}} &= \left(\frac{L_{jt}}{L_t} \right)^{\sigma_l^{-1}} \\
\frac{\partial Q_{jt}}{\partial Y_{jt}} &= \left(\mu_j \frac{Q_{jt}}{Y_{jt}} \right)^{\sigma_q^{-1}} & \frac{\partial Q_{jt}}{\partial M_{jt}} &= \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\sigma_q^{-1}} \\
\frac{\partial Y_{jt}}{\partial K_{jt}} &= A_{jt}^{1-\sigma_y^{-1}} \left(\alpha_j \frac{Y_{jt}}{K_{jt}} \right)^{\sigma_y^{-1}} & \frac{\partial Y_{jt}}{\partial L_{jt}} &= A_{jt}^{1-\sigma_y^{-1}} \left((1-\alpha_j) \frac{Y_{jt}}{L_{jt}} \right)^{\sigma_y^{-1}} \\
\frac{\partial M_{jt}}{\partial M_{ijt}} &= \left(\gamma_{ij}^{M} \frac{M_{jt}}{M_{ijt}} \right)^{\sigma_m^{-1}} & \frac{\partial I_{jt}}{\partial I_{ijt}} &= \left(\gamma_{ij}^{I} \frac{I_{jt}}{I_{ijt}} \right)^{\sigma_I^{-1}} \\
\frac{\partial \Phi_{jt}}{\partial I_{jt}} &= \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) & \frac{\partial \Phi_{jt}}{\partial K_{jt}} &= -\frac{\phi}{2} \left(\frac{I_{jt}^2}{K_{jt}^2} - \delta_j^2 \right).
\end{aligned}
$$

B.1 FOC for consumption

We start with

$$\frac{\partial \mathcal{L}}{\partial C_{jt}} = \beta^t \left(\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \frac{\partial C_t}{\partial C_{jt}} - P_{jt} \right) = 0.$$

Next, we replace the derivatives:

$$\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \frac{\partial C_t}{\partial C_{jt}} = P_{jt},$$
$$\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \left(\xi_j \frac{C_t}{C_{jt}} \right)^{\sigma_c^{-1}} = P_{jt}.$$

B.2 FOC for labor

We start with

$$\frac{\partial \mathcal{L}}{\partial L_{jt}} = \beta^t \left(- \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \theta \left(L_t \right)^{\epsilon_l^{-1}} \frac{\partial L_t}{\partial L_{jt}} + P_{jt} \frac{\partial Q_{jt}}{\partial L_{jt}} \right) = 0.$$

Next, we replace the derivatives, using the chain rule when needed:

$$\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \theta \left(L_t \right)^{\epsilon_l^{-1}} \frac{\partial L_t}{\partial L_{jt}} = P_{jt} \frac{\partial Q_{jt}}{\partial Y_{jt}} \frac{\partial Y_{jt}}{\partial L_{jt}},$$
$$\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{-\epsilon_c^{-1}} \theta \left(L_t \right)^{\epsilon_l^{-1}} \left(\frac{L_{jt}}{L_t} \right)^{\sigma_l^{-1}} = P_{jt} A_{jt}^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt}}{Y_{jt}} \right)^{\sigma_q^{-1}} \left(\left(1 - \alpha_j \right) \frac{Y_{jt}}{L_{jt}} \right)^{\sigma_y^{-1}}.$$

B.3 FOC with respect to capital in next period

We start with

$$\frac{\partial \mathcal{L}}{\partial K_{jt+1}} = \beta^t \left(-P_{jt}^k \right) + \beta^{t+1} \mathbb{E}_t \left(P_{jt+1}^k \left(\left(1 - \delta_j \right) - \frac{\partial \Phi_{jt+1}}{\partial K_{jt+1}} \right) + P_{jt+1} \frac{\partial Q_{jt+1}}{\partial K_{jt+1}} \right) = 0.$$

Next, we replace the derivatives, using the chain rule when needed:

$$
\begin{aligned}
P_{jt}^{k} &= \beta \mathbb{E}_t \left(P_{jt+1}^{k} \left((1-\delta_j) - \frac{\partial \Phi_{jt+1}}{\partial K_{jt+1}} \right) + P_{jt+1} \frac{\partial Q_{jt+1}}{\partial Y_{jt+1}} \frac{\partial Y_{jt+1}}{\partial K_{jt+1}} \right), \\
P_{jt}^{k} &= \beta \mathbb{E}_t \left[P_{jt+1}^{k} \left((1-\delta_j) + \frac{\phi}{2} \left(\frac{I_{jt+1}^2}{K_{jt+1}^2} - \delta_j^2 \right) \right) \right. \\
&\quad \left. + P_{jt+1} A_{jt+1}^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt+1}}{Y_{jt+1}} \right)^{\sigma_q^{-1}} \left(\alpha_j \frac{Y_{jt+1}}{K_{jt+1}} \right)^{\sigma_y^{-1}} \right].
\end{aligned}
$$

B.4 FOC for intermediates and system reduction

We start with

$$
\frac{\partial \mathcal{L}}{\partial M_{ijt}} = \beta^t \left(P_{jt} \frac{\partial Q_{jt}}{\partial M_{ijt}} - P_{it} \right) = 0.
$$

Next, we replace the derivatives, using the chain rule when needed:

$$
\begin{aligned}
P_{jt} \frac{\partial Q_{jt}}{\partial M_{jt}} \frac{\partial M_{jt}}{\partial M_{ijt}} &= P_{it}, \\
P_{jt} \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\sigma_q^{-1}} \left(\gamma_{ij}^{M} \frac{M_{jt}}{M_{ijt}} \right)^{\sigma_m^{-1}} &= P_{it}
\end{aligned}
$$

We want to use this FOC to get rid of M_{ijt}. First, we solve for M_{ijt}:

$$
M_{ijt} = \left(\frac{P_{jt}}{,} P_{it} \right)^{\sigma_m} \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\frac{\sigma_m}{\sigma_q}} \gamma_{ij}^{m} M_{jt}
$$

We solve for M_{jt} by aggregating from the solution for M_{ijt}. First, we construct the term $\left(\gamma_{ij}^{m}\right)^{\frac{1}{\sigma_m}} M_{ijt}^{1-\sigma_m^{-1}}$ that is inside the aggregator:

$$
\begin{aligned}
M_{ijt} &= \left(\frac{P_{jt}}{P_{it}} \right)^{\sigma_m} \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\frac{\sigma_m}{\sigma_q}} \gamma_{ij}^{m} M_{jt}, \\
\left(\gamma_{ij}^{m}\right)^{\frac{1}{\sigma_m}} M_{ijt}^{1-\sigma_m^{-1}} &= \left(\gamma_{ij}^{m}\right)^{\frac{1}{\sigma_m}} \left(\frac{P_{it}}{P_{jt}} \right)^{1-\sigma_m} \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\frac{\sigma_m - 1}{\sigma_q}} \left(\gamma_{ij}^{m} M_{jt} \right)^{1-\sigma_m^{-1}}, \\
\left(\gamma_{ij}^{m}\right)^{\frac{1}{\sigma_m}} M_{ijt}^{1-\sigma_m^{-1}} &= \gamma_{ij}^{m} \left(\frac{P_{it}}{P_{jt}} \right)^{1-\sigma_m} \left((1-\mu_j) \frac{Q_{jt}}{M_{jt}} \right)^{\frac{\sigma_m - 1}{\sigma_q}} \left(M_{jt} \right)^{1-\sigma_m^{-1}}.
\end{aligned}
$$

Next, we sum over all the goods in the aggregator:

$$\begin{aligned}
\sum_{i=1}^{N}\left(\gamma_{ij}^{m}\right)^{\frac{1}{\sigma_m}} M_{ijt}^{1-\sigma_m^{-1}} &= \sum_{i=1}^{N}\gamma_{ij}^{m}\left(\frac{P_{it}}{P_{jt}}\right)^{1-\sigma_m}\left(\left(1-\mu_j\right)\frac{Q_{jt}}{M_{jt}}\right)^{\frac{\sigma_m-1}{\sigma_q}} M_{jt}^{1-\sigma_m^{-1}}, \\
M_{jt}^{1-\sigma_m^{-1}} &= \left(P_{jt}\right)^{\sigma_m-1}\left(\left(1-\mu_j\right)\frac{Q_{jt}}{M_{jt}}\right)^{\frac{\sigma_m-1}{\sigma_q}} M_{jt}^{1-\sigma_m^{-1}}\sum_{i=1}^{N}\gamma_{ij}^{m}\left(P_{it}\right)^{1-\sigma_m}, \\
M_{jt}^{\frac{\sigma_m-1}{\sigma_q}} &= \left(\left(1-\mu_j\right)Q_{jt}\right)^{\frac{\sigma_m-1}{\sigma_q}}\left(P_{jt}\right)^{\sigma_m-1}\sum_{i=1}^{N}\left(\gamma_{ij}^{m}\right)\left(P_{it}\right)^{1-\sigma_m}, \\
M_{jt} &= \left(\left(1-\mu_j\right)Q_{jt}\right)P_{jt}^{\sigma_q}\left(\sum_{i=1}^{N}\left(\gamma_{ij}^{m}\right)\left(P_{it}\right)^{1-\sigma_m}\right)^{\frac{\sigma_q}{\sigma_{m-1}}}.
\end{aligned}$$

Following the CES algebra in sub-appendix B.11, we define the price index for the M_{jt} bundle as:

$$P_{jt}^{m} = \left(\sum_{i=1}^{N}\left(\gamma_{ij}^{m}\right)\left(P_{it}\right)^{1-\sigma_m}\right)^{\frac{1}{1-\sigma_m}}.$$

so we can write the FOC for M_{jt} as:

$$M_{jt} = \left(1-\mu_j\right)\left(\frac{P_{jt}^{m}}{P_{jt}}\right)^{-\sigma_q} Q_{jt}.$$

We can use this expression for $M_{jt}$to simplify our solution for M_{ijt}:

$$\begin{aligned}
M_{ijt} &= \left(\frac{P_{it}}{P_{jt}}\right)^{-\sigma_m}\left(\left(1-\mu_j\right)\frac{Q_{jt}}{M_{jt}}\right)^{\frac{\sigma_m}{\sigma_q}}\gamma_{ij}^{m}M_{jt}, \\
&\left(\frac{P_{it}}{P_{jt}}\right)^{-\sigma_m}\left(\left(1-\mu_j\right)\frac{Q_{jt}}{\left(1-\mu_j\right)\left(\frac{P_{jt}^{m}}{P_{jt}}\right)^{-\sigma_q}Q_{jt}}\right)^{\frac{\sigma_m}{\sigma_q}}\gamma_{ij}^{m}M_{jt}, \\
&= \gamma_{ij}^{m}\left(\frac{P_{it}}{P_{jt}^{m}}\right)^{-\sigma_m}M_{jt}.
\end{aligned}$$

Finally, using this solution for M_{ijt}, we can calculate the supply of intermediate goods of each

sector, which we denote $M_{jt}^{out} = \sum_{i=1}^{N} M_{jit}$:

$$M_{jt}^{out} = \sum_{i=1}^{N} M_{jit} = \sum_{i=1}^{N} \gamma_{ji}^{m} \left(\frac{P_{jt}}{P_{it}^{m}} \right)^{-\sigma_m} M_{it}.$$

B.5 FOC for investment and system reduction

We start with

$$\frac{\partial \mathcal{L}}{\partial I_{ijt}} = \beta^t \left(P_{jt}^{k} \left(\frac{\partial I_{jt}}{\partial I_{ijt}} - \frac{\partial \Phi_{jt}}{\partial I_{ijt}} \right) - P_{it} \right) = 0.$$

Next, we replace the derivatives, using the chain rule when needed:

$$\begin{aligned} P_{jt}^{k} \left(\frac{\partial I_{jt}}{\partial I_{ijt}} - \frac{\partial \Phi_{jt}}{\partial I_{jt}} \frac{\partial I_{jt}}{\partial I_{ijt}} \right) &= P_{it}, \\ P_{jt}^{k} \frac{\partial I_{jt}}{\partial I_{ijt}} \left(1 - \frac{\partial \Phi_{jt}}{\partial I_{jt}} \right) &= P_{it}, \\ P_{jt}^{k} \left(\gamma_{ij}^{I} \frac{I_{jt}}{I_{ijt}} \right)^{\sigma_I^{-1}} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right) &= P_{it}. \end{aligned}$$

We want to use this FOC to eliminate I_{ijt}. First, we solve for I_{ijt}:

$$I_{ijt} = \gamma_{ij}^{I} \left(\frac{P_{it}}{P_{jt}^{k}} \right)^{-\sigma_I} I_{jt} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{\sigma_I}.$$

We solve for I_{jt} by aggregating up from the solution for I_{ijt}

$$
\begin{aligned}
I_{ijt} &= \gamma_{ij}^{I} \left(\frac{P_{it}}{P_{jt}^{k}} \right)^{-\sigma_I} I_{jt} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{\sigma_I}, \\
\left(\gamma_{ij}^{I}\right)^{\sigma_I^{-1}} I_{ijt}^{1-\sigma_I^{-1}} &= \left(\gamma_{ij}^{I}\right)^{\sigma_I^{-1}} \left(\frac{P_{it}}{P_{jt}^{k}} \right)^{1-\sigma_I} \left(\gamma_{ij}^{I} I_{jt}\right)^{1-\sigma_I^{-1}} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{\sigma_I - 1}, \\
I_{jt}^{1-\sigma_I^{-1}} &= \sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right) \left(\frac{P_{it}}{P_{jt}^{k}} \right)^{1-\sigma_I} \left(I_{jt}\right)^{1-\sigma_I^{-1}} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{\sigma_I - 1}. \\
1 &= \left(P_{jt}^{k}\right)^{\sigma_I - 1} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{\sigma_I - 1} \sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right) \left(P_{it}\right)^{1-\sigma_I}, \\
P_{jt}^{k} &= \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{-1} \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right) \left(P_{it}\right)^{1-\sigma_I} \right)^{\frac{1}{1-\sigma_I}}.
\end{aligned}
$$

We define the frictionless price index of capital goods as

$$
\tilde{P}_{jt}^{k} = \left(\sum_{i=1}^{N} \left(\gamma_{ij}^{I}\right) \left(P_{it}\right)^{1-\sigma_I} \right)^{\frac{1}{1-\sigma_I}}.
$$

Then, we can write the FOC for I_{jt} as

$$
P_{jt}^{k} = \tilde{P}_{jt}^{k} \left(1 - \phi \left(\frac{I_{jt}}{K_{jt}} - \delta_j \right) \right)^{-1}.
$$

Next, we calculate the amount of goods of a sector that goes to other sectors as investment goods. We define

$$
I_{jt}^{out} = \sum_{i=1}^{N} I_{jit}.
$$

Using the FOC for I_{ijt}, we have

$$
I_{jt}^{out} = \sum_{i=1}^{N} \gamma_{ji}^{I} \left(\frac{P_{jt}}{P_{it}^{k}} \right)^{-\sigma_I} I_{it} \left(1 - \phi \left(\frac{I_{it}}{K_{it}} - \delta_i \right) \right)^{\sigma_I}.
$$

B.6 Full system of equations

The full system we get is:

$$
\begin{aligned}
\log A_{jt+1} &= \rho_j \log A_{jt} + \varepsilon^A_{jt+1}, \\
K_{jt+1} &= (1-\delta_j)K_{jt} + I_{jt} - \frac{\phi}{2}\left(\frac{I_{jt}}{K_{jt}} - \delta_j\right)^2 K_{jt}, \\
P_{jt} &= \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}}\right)^{-\epsilon_c^{-1}} \left(\xi_j \frac{C_t}{C_{jt}}\right)^{\sigma_c^{-1}}, \\
\frac{\theta \left(L_t\right)^{\epsilon_l^{-1}} \left(\frac{L_{jt}}{L_t}\right)^{\sigma_l^{-1}}}{\left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}}\right)^{\epsilon_c^{-1}}} &= P_{jt} A_{jt}^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt}}{Y_{jt}}\right)^{\sigma_q^{-1}} \left((1-\alpha_j)\frac{Y_{jt}}{L_{jt}}\right)^{\sigma_y^{-1}}, \\
P^k_{jt} &= \beta \mathbb{E}_t \left[P_{jt+1} A_{jt+1}^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt+1}}{Y_{jt+1}}\right)^{\sigma_q^{-1}} \left(\alpha_j \frac{Y_{jt+1}}{K_{jt+1}}\right)^{\sigma_y^{-1}} \right], \\
&\quad + P^k_{jt+1} \left((1-\delta_j) + \frac{\phi}{2}\left(\frac{I^2_{jt+1}}{K^2_{jt+1}} - \delta_j^2\right)\right)\Bigg], \\
P^m_{jt} &= \left(\sum_{i=1}^N \left(\gamma^m_{ij}\right)\left(P_{it}\right)^{1-\sigma_m}\right)^{\frac{1}{1-\sigma_m}}, \\
M_{jt} &= (1-\mu_j)\left(\frac{P^m_{jt}}{P_{jt}}\right)^{-\sigma_q} Q_{jt}, \\
M^{out}_{jt} &= \sum_{i=1}^N \gamma^m_{ji} \left(\frac{P_{jt}}{P^m_{it}}\right)^{-\sigma_m} M_{it}, \\
P^k_{jt} &= \left(\sum_{i=1}^N \left(\gamma^I_{ij}\right)\left(P_{it}\right)^{1-\sigma_I}\right)^{\frac{1}{1-\sigma_I}} \left(1 - \phi\left(\frac{I_{jt}}{K_{jt}} - \delta_j\right)\right)^{-1}, \\
I^{out}_{jt} &= \sum_{i=1}^N \gamma^I_{ji} \left(\frac{P_{jt}}{P^k_{it}}\right)^{-\sigma_I} I_{it} \left(1 - \phi\left(\frac{I_{it}}{K_{it}} - \delta_i\right)\right)^{\sigma_I}, \\
Q_{jt} &= C_{jt} + M^{out}_{jt} + I^{out}_{jt}, \\
Q_{jt} &= \left[(\mu_j)^{\sigma_q^{-1}} \left(Y_{jt}\right)^{1-\sigma_q^{-1}} + (1-\mu_j)^{\sigma_q^{-1}} \left(M_{jt}\right)^{1-\sigma_q^{-1}}\right]^{\frac{1}{1-\sigma_q^{-1}}}, \\
Y_{jt} &= A_{jt}\left[(\alpha_j)^{\sigma_y^{-1}} \left(K_{jt}\right)^{1-\sigma_y^{-1}} + (1-\alpha_j)^{\sigma_y^{-1}} \left(L_{jt}\right)^{1-\sigma_y^{-1}}\right]^{\frac{1}{1-\sigma_y^{-1}}}, \\
C_t &= \left(\sum_{j=1}^N \xi_j^{\frac{1}{\sigma_c}} \left(C_{jt}\right)^{1-\sigma_c^{-1}}\right)^{\frac{1}{1-\sigma_c^{-1}}}, \\
L_t &= \left(\sum_{j=1}^N \left(L_{jt}\right)^{1+\sigma_l^{-1}}\right)^{\frac{1}{1+\sigma_l^{-1}}}.
\end{aligned}
$$

B.7 Welfare

In order to calculate welfare, we can write the intertemporal utility of the representative household as:

$$V_t = \frac{1}{1-\epsilon_c^{-1}} \left(C_t - \theta \frac{L_t^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}} + \beta E_t V_{t+1}.$$

Steady state welfare is:

$$\bar{V} = \frac{1}{1-\beta} \frac{1}{1-\epsilon_c^{-1}} \left(\bar{C} - \theta \frac{\bar{L}^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}}.$$

Then, in a given period, we can get an interpretable measure of welfare by calculating the fraction of steady-state consumption that you would need to give up to achieve that level of welfare in the steady state. We denote such consumption-equivalent welfare as $\hat{V}_t^c$:

$$V_t = \frac{1}{1-\beta} \frac{1}{1-\epsilon_c^{-1}} \left(\bar{C}(1+\hat{V}_t^c) - \theta \frac{\bar{L}^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right)^{1-\epsilon_c^{-1}}.$$

We can solve analytically for consumption-equivalent welfare $\hat{V}_t^c$:

$$\hat{V}_t^c = \frac{1}{\bar{C}} \left[\left(V_t(1-\beta)(1-\epsilon_c^{-1}) \right)^{\frac{1}{1-\epsilon_c^{-1}}} + \theta \frac{\bar{L}^{1+\epsilon_l^{-1}}}{1+\epsilon_l^{-1}} \right] - 1.$$

We will analyze how $\hat{V}_t^c$ is affected by productivity shocks.

B.8 Vectorizations

In terms of programming, it will be useful to vectorize the equations that involve sums over sectoral variables. We will use variables in bold to denote vectors where each element represents the corresponding sectoral value, and use $*$ to denote element by element multiplication. When we raise a vector to the power of a parameter, we mean element-to-element

exponentiation. Then, we get the following equations:

$$\begin{aligned}
\boldsymbol{P_t^m} &= \left(\Gamma_M^{'} \boldsymbol{P_t}^{1-\sigma_m}\right)^{\frac{1}{1-\sigma_m}}, \\
\boldsymbol{M_t^{out}} &= \boldsymbol{P_t}^{-\sigma_m} * \Gamma_M \left(\boldsymbol{P_t^m}\right)^{\sigma_m} * \boldsymbol{M_t} \\
\boldsymbol{\tilde{P}_t^k} &= \left(\Gamma_I^{'} \boldsymbol{P_t}^{1-\sigma_I}\right)^{\frac{1}{1-\sigma_I}}, \\
\boldsymbol{\tilde{I}_t^{out}} &= \boldsymbol{P_t}^{-\sigma_I} * \Gamma_I \left(\boldsymbol{P_t^k}\right)^{\sigma_I} * \boldsymbol{I_t}.
\end{aligned}$$

B.9 Steady state

There are three dynamic equations in the model:

$$\begin{aligned}
K_{jt+1} &= (1-\delta_j)K_{jt} + I_{jt} - \frac{\phi}{2}\left(\frac{I_{jt}}{K_{jt}} - \delta_j\right)^2 K_{jt}, \\
a_{jt+1} &= \rho_j a_{jt} + \epsilon_{jt}, \\
P_{jt}^k &= \beta \mathbb{E}_t \left[P_{jt+1} A_{jt+1}^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt+1}}{Y_{jt+1}}\right)^{\sigma_q^{-1}} \left(\alpha_j \frac{Y_{jt+1}}{K_{jt+1}}\right)^{\sigma_y^{-1}} \right. \\
&\quad \left. + P_{jt+1}^k \left((1-\delta_j) + \frac{\phi}{2}\left(\frac{I_{jt+1}^2}{K_{jt+1}^2} - \delta_j^2\right)\right)\right].
\end{aligned}$$

First, notice that $I_{jt} = \delta_j K_{jt}$ implies $K_{jt+1} = K_{jt}$ and makes the adjustment costs equal to zero. Then, in the steady state we have:

$$\begin{aligned}
\bar{K} &= \bar{I}_j/\delta_j, \\
\bar{P}_j^k &= \frac{\beta}{1-\beta(1-\delta)} \bar{P}_j \left(\mu_j \frac{\bar{Q}_j}{\bar{Y}_j}\right)^{\sigma_q^{-1}} \left(\alpha \frac{\bar{Y}_j}{\bar{K}_j}\right)^{\sigma_y^{-1}}.
\end{aligned}$$

B.10 Intensity shares mapping to expenditure shares

In sub-appendix B.11, we show that for standard CES aggregators, the intensity shares (e.g., ξ_j for consumption bundle or α_j for value-added function) do not correspond to expenditure shares. In the same sub-appendix, we show that how to map expenditures shares to intensity shares once you know the steady state equilibrium variables. We use tilde notation to refer to expenditures, so $\tilde{\xi}_j$ is the consumption share of good j. Then, for a given steady state equilibrium the relation between the expenditure share in steady state and the intensity share

used to calculate the steady state is:

$$\tilde{\xi}_j = \xi_j^{\sigma_c^{-1}} \left(\frac{\bar{C}_j}{\bar{C}} \right)^{1-\sigma_c^{-1}},$$
$$\tilde{\mu}_j = \mu_j^{\sigma_q^{-1}} \left(\frac{\bar{Y}_j}{\bar{Q}_j} \right)^{1-\sigma_q^{-1}},$$
$$\tilde{\alpha}_j = \alpha_j^{\sigma_y^{-1}} \left(\frac{\bar{K}_j}{\bar{Y}_j} \right)^{1-\sigma_y^{-1}},$$
$$\tilde{\gamma}_{ij}^m = \left(\gamma_{ij}^m\right)^{\sigma_m^{-1}} \left(\frac{\bar{M}_{ij}}{\bar{M}_j} \right)^{1-\sigma_m^{-1}},$$
$$\tilde{\gamma}_{ij}^I = \left(\gamma_{ij}^I\right)^{\sigma_I^{-1}} \left(\frac{\bar{I}_{ij}}{\bar{I}_j} \right)^{1-\sigma_I^{-1}}.$$

Since we do not solve explicitly for $\bar{M}_{ij}$ and $\bar{I}_{ij}$ we are going to use the first order conditions to solve for $\bar{M}_{ij}/\bar{M}_j$ and $\bar{I}_{ij}/\bar{I}_j$. We get

$$\tilde{\gamma}_{ij}^m = \gamma_{ij}^m \left(\frac{P_{it}}{P_{jt}^m} \right)^{1-\sigma_m},$$
$$\tilde{\gamma}_{ij}^I = \gamma_{ij}^I \left(\frac{P_{it}}{P_{jt}^k} \right)^{1-\sigma_I}.$$

Given this mapping, a naive approach would be simply to replace the intensity shares with the equations that map the empirical shares with the model shares. Nevertheless, if we use this mapping equations as endogenous equations in the steady state system of equations, the output of each aggregator become indeterminate. To see why, we can replace the mapping in the consumption aggregator and rearrange to obtain:

$$C_t = \bar{C} \left(\sum_{j=1}^{N} \tilde{\xi}_j \left(\frac{C_{jt}}{\bar{C}_j} \right)^{1-\sigma_c^{-1}} \right)^{\frac{1}{1-\sigma_c^{-1}}}.$$

This equation for aggregate C cannot be used in steady state, since if we replace the time t values for steady-state values we get a tautology ($\bar{C} = \bar{C}$). Given this, the correct approach is to include in the steady state system the difference between the model-implied expenditure shares as additional expressions to minimize.

B.11 CES algebra

The objective of this appendix is to obtain a mapping between the intensity shares (the primitive parameters that appear in the CES aggregator) and the expenditure shares (input expenditure/aggregate expenditure). We start with the common CES aggregator:

$$X = A\left(\sum_{j=1}^{N} \xi_j^{\sigma_x^{-1}} {X_j}^{1-\sigma_x^{-1}}\right)^{\frac{1}{1-\sigma_x^{-1}}}.$$

where A is a constant (e.g., TFP in the value added aggregator). The derivative is:

$$\begin{aligned}
\frac{\partial X}{\partial X_j} &= \frac{A}{1-\sigma_x^{-1}}\left(\sum_{j=1}^{N} \xi_j^{\sigma_x^{-1}} (X_j)^{1-\sigma_x^{-1}}\right)^{\frac{\sigma_x^{-1}}{1-\sigma_x^{-1}}} \left(\xi_j^{\sigma_x^{-1}}\left(1-\sigma_x^{-1}\right)(X_j)^{-\sigma_x^{-1}}\right), \\
&= \frac{A}{A^{\sigma_x^{-1}}} A^{\sigma_x^{-1}}\left(\sum_{j=1}^{N} \xi_j^{\sigma_q^{-1}} (X_j)^{1-\sigma_x^{-1}}\right)^{\frac{\sigma_x^{-1}}{1-\sigma_x^{-1}}} \xi_j^{\sigma_x^{-1}} (X_j)^{-\sigma_x^{-1}}, \\
&= A^{1-\sigma_x^{-1}}\left(A\left(\sum_{j=1}^{N} \xi_j^{\sigma_x^{-1}} (X_j)^{1-\sigma_x^{-1}}\right)^{\frac{1}{1-\sigma_x^{-1}}}\right)^{\sigma_x^{-1}} \xi_j^{\sigma_x^{-1}} (X_j)^{-\sigma_x^{-1}}.
\end{aligned}$$

Notice that we can now plug the definition of the aggregator. We get:

$$\begin{aligned}
\frac{\partial X_t}{\partial X_{jt}} &= A^{1-\sigma_x^{-1}} X^{\sigma_x^{-1}} \xi_j^{\sigma_x^{-1}} (X_j)^{-\sigma_x^{-1}}, \\
&= A^{1-\sigma_x^{-1}}\left(\xi_j \frac{X}{X_j}\right)^{\sigma_x^{-1}}
\end{aligned}$$

From this first-order condition, we are going to get a price aggregator. We start from the budget constraint:

$$\sum_{j=1}^{N} P_j X_j = Y_j.$$

where Y_{jt} represents income. The optimization problem is

$$\max_{\{X_j\}_{j=1}^N} X \quad \text{s.t.} \quad \sum_{j=1}^N P_j X_j = Y_j,$$

The first-order conditions with respect to X_{jt} are:

$$\frac{\partial X}{\partial X_j} - \lambda P_j = 0.$$

So, for al goods $i \neq j$, we have:

$$\frac{\partial X}{\partial X_j}\frac{1}{P_j} = \frac{\partial X}{\partial X_i}\frac{1}{P_i},$$
$$A^{1-\sigma_x^{-1}}\left(\xi_j \frac{X}{X_j}\right)^{\sigma_x^{-1}} \frac{1}{P_j} = A^{1-\sigma_x^{-1}}\left(\xi_i \frac{X}{X_i}\right)^{\sigma_x^{-1}} \frac{1}{P_i},$$

and we get

$$X_j = \left(\frac{P_i}{P_j}\right)^{\sigma_x} \frac{\xi_j}{\xi_i} X_i.$$

Let's replace that into the aggregator:

$$\begin{aligned} X &= A\left(\sum_{j=1}^N \xi_j^{\sigma_x^{-1}} \left(\left(\frac{P_i}{P_j}\right)^{\sigma_x} \frac{\xi_j}{\xi_i} X_i\right)^{1-\sigma_x^{-1}}\right)^{\frac{1}{1-\sigma_x^{-1}}}, \\ &= \frac{1}{\xi_i}\left(P_i\right)^{\sigma_x} X_i A\left(\sum_{j=1}^N \xi_j \left(P_j\right)^{1-\sigma_x}\right)^{\frac{1}{1-\sigma_x^{-1}}}. \end{aligned}$$

We get an expression for X_i in terms of aggregates:

$$X_i = \xi_i \left(\frac{1}{P_i}\right)^{\sigma_x} X A^{-1}\left(\sum_{j=1}^N \xi_j^{\sigma_x} \left(P_j\right)^{1-\sigma_x}\right)^{\frac{-1}{1-\sigma_x^{-1}}}.$$

We calculate total expenditure:

$$P_i X_i = P_i^{1-\sigma_x} \xi_i X A^{-1} \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right)^{\frac{-1}{1-\sigma_x^{-1}}}.$$

Adding up all the goods:

$$\begin{aligned} Y &= X A^{-1} \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right) \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right)^{\frac{-1}{1-\sigma_x^{-1}}}, \\ &= X A^{-1} \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right)^{\frac{1}{1-\sigma_x}}. \end{aligned}$$

We can define the price aggregator

$$P = A^{-1} \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right)^{\frac{1}{1-\sigma_x}}.$$

such that $Y_t = X_t P_t$. Also, we can plug in the price aggregator in our expression for X_{it} to obtain:

$$\begin{aligned} X_i &= \xi_i \left(\frac{1}{P_i} \right)^{\sigma_x} X A^{-1} \left(\sum_{j=1}^{N} \xi_j \left(P_j \right)^{1-\sigma_x} \right)^{\frac{-1}{1-\sigma_x^{-1}}}, \\ &= \xi_i \left(\frac{P}{P_i} \right)^{\sigma_x} X. \end{aligned}$$

The problem with the standard formulation of the CES is that the taste parameters ξ_i do not correspond to expenditure shares unless $\sigma_x = 1$:

$$\xi_i = \left(\frac{X_i}{X} \right) \left(\frac{P_i}{P} \right)^{\sigma_x}$$

Given this, we propose a mapping between intensity shares ξ_i and expenditure shares $\tilde{\xi}_i$:

$$
\begin{aligned}
\xi_i &= \left(\frac{X_i}{X}\right)\left(\frac{P_i}{P}\right)^{\sigma_x}, \\
&= \left(\frac{X_i}{X}\right)\left(\frac{X}{X_i}\frac{X_i}{X}\frac{P_i}{P}\right)^{\sigma_x}, \\
&= \left(\frac{X_i}{X}\right)^{1-\sigma_x}\left(\tilde{\xi}_i\right)^{\sigma_x}.
\end{aligned}
$$

This implies the following mapping from intensity shares to expenditure shares:

$$
\tilde{\xi}_i = \xi_i^{\sigma_x^{-1}}\left(\frac{X_i}{X}\right)^{1-\sigma_x^{-1}}.
$$

C Solution Method

To find the global solution, we adapt the Deep Equilibrium Nets method (Azinovic et al., 2022) to allow for Monte-Carlo based expectations, since our model has a large number of shocks. This method consists in using neural networks as function approximators for the policy functions that map the state variables $\{K_{jt}, Z_{jt}\}_j$ to the endogenous variables (policies and prices), which we denote as $\{X_{jt}\}_j$, and train the neural net to reduce the error in the system of equations describing the equilibrium, presented in subsection B.6. One of the key elements of this methodology is that we use the neural network to simulate the model and obtain points of the state space over which we will minimize the loss function. Also, we use the neural net to estimate the expectation terms that appear in the system of equations that describe the solution.

More specifically, the system of equations that we are trying to solve contains state and endogenous variables at time t, and expectations that depend on endogenous variables at time $t+1$. In order to evaluate the average error in the system of equations associated to some neural net parameters, we first sample points of the state space by simulating episodes using the neural net to step the model forward. Then, we use again the neural network to get the policies at those points of the state space. Finally, to calculate the expectation terms at each point of the state space, we simulate many one-period transitions, use the neural net to recover the terms inside the expectation for each one-period ahead simulation, and then averaging them up to get the expectation. The only non-trivial part of this evaluation is to evaluate the expectation terms. For models in which we have an explicit and deterministic equation

linking the endogenous state today (K_{jt}) with the endogenous state tomorrow (K_{jt+1}), such as real business cycle models, we can simulate many one-period transitions by sampling only the exogenous state, since we already know K_{jt+1}.

Before we explain formally how to calculate the expectation terms, we are going to introduce the neural net approximator. Let $S_t \in \mathbb{R}^{2N}$ be the state vector $S = [K_t, a_t]$. The neural net approximator, with L layers indexed by i, is:

$$\mathcal{F}(\{W, b_l\}_l; S_t) = \sigma_l(W_l * [K_t, a_t] + b_l)$$

Thus, the neural net is a function $\mathcal{F}(\theta_t; K_t, a_t)$ where θ_t correspond to the internal parameters of the neural net. The output of the neural net is a vector of non-negative real values. In particular, we define the policy variables as

$$X_t = \mathcal{F}(\theta_t; K_t, a_t)$$

Next, we show how to calculate the expectation function given that we are at particular point of the state space and our current policy has parameters θ_t.

C.1 The expectation function

In the system of equations presented in subsection B.6, the only expectation term appears in the F.O.C. for K_{jt+1}:

$$P_{jt}^k = \beta \mathbb{E}_t \left[P_{jt+1} \left(A_{jt+1}\right)^{1-\sigma_y^{-1}} \left(\mu_j \frac{Q_{jt+1}}{Y_{jt+1}}\right)^{\sigma_q^{-1}} \left(\alpha \frac{Y_{jt+1}}{K_{jt+1}}\right)^{\sigma_y^{-1}} \right.$$
$$\left. + P_{jt+1}^k \left((1-\delta_k) + \frac{\phi}{2} \left(\frac{I_{jt+1}^2}{K_{jt+1}^2} - \delta_j^2 \right) \right) | K_t, a_t \right]$$

Notice that given policy parameters θ_t, K_{jt+1} is already determined, so the expectation is taken over realizations of $a_{t+1} \equiv \log(A_{t+1})$, whose conditional probability distribution is

$$p(a_{t+1}|a_t) = \mathcal{N}(\rho a_t, \Sigma^a)$$

In order to calculate our loss function, we need to calculate that expectation term at each

point of the state space we visit. Next, we break down how that expectation would be constructed using a montecarlo simulation. First, given K_t and the policy parameters θ_t, we calculate next period capital:

$$K_{t+1} = (1 - \delta_j)K_t + 1_{I_t}^{'}\mathcal{F}(\theta_t; K_t, a_t)$$

where $1_{I_t}^{'}$ is a vector that selects, among all the policies, the vector of sectoral investments. Second, we sample 128 realization of the shock vector using the conditional distribution $(a_{t+1}|a_t) = \mathcal{N}(\rho a_t, \Sigma^a)$. Third, we we are going to use our policy function to obtain the policies in $t+1$ for a each draw of $a_{t+1} \sim \mathcal{N}(\rho a_t, \Sigma^a)$ and the predetermined K_{t+1}:

$$X_{t+1} = \mathcal{F}(\theta_t; K_{t+1}, a_{t+1})$$

Finally, we use K_{t+1} and the policies X_{t+1} to calculate the term inside the expectation for each draw of the shock vector, and we get the expectation by taking the average across all draws.

C.2 The loss function

Now that we have solved for the expectation function $\Phi_t(K_t, a_t, \theta_t)$, we can calculate the loss function in period t. First, we specify the endogenous variables in period t given K_t, a_t and the parameters of the policy θ_t:$X_t = \mathcal{F}(\theta_t; K_t, a_t)$. Then, we calculate the period error, given as the quadratic loss in the system of equations that describe the solution. The error is written as right hand side divided by left hand side minus. For example, the loss for the market clearing equation of sector j is:

$$\mathcal{L}_{Q_j} \equiv \left(\frac{\left[(mu_j)^{\frac{1}{\sigma_q}} (Y_{jt})^{\frac{\sigma_q - 1}{\sigma_q}} + (1 - \mu_j)^{\frac{1}{\sigma_q}} (M_{jt})^{\frac{\sigma_q - 1}{\sigma_q}} \right]^{\frac{\sigma_q}{\sigma_q - 1}}}{Q_{jt}} \mathbf{1} \right)^2$$

then, we take this loss and square it.

C.3 Implementation

In this section, we present a step by step breakdown of the implementation. The first part of the implementation, which consists on calculating a loglinear policy using dynare, is done

in Matlab. The second and third part of the implementation, which consist on pretraining the neural net to fit the log linear policy and then use that solve the full nonlinear version of the model, are coded in Python Colab Notebooks. The notebooks are publicly available at RbcProdNet_pretrain and RbcProdNet_train, and all the codes, including the notebooks, can also be found in the github page github econjax.

Before we start our three steps, we start in Matlab calculating the **steady state**, described in section B.9

C.3.1 Solve Log-linear Model

We solve a version of the model with log-linearization around the steady state using dynare. We are going to use these policies to "pre-train" the neural net (more details in next subsection). . Next, we recover the state space representation used by dynare internally. Let $X_t = \{I_t, Q_t, P_t, L_t\}$ be the vector of endogenous variables, and denote the vector with logs as x_t .s_t contains the log of the states, and the sub-index ss represents steady state values. The representation is:

$$
\begin{aligned}
s_{t+1} - s_{ss} &= A(s_t - s_{ss}) + Be_t \\
x_t - x_{ss} &= C(s_{t_{-1}} - s_{ss}) + De_t
\end{aligned}
$$

A relevant detail of this stage is that timing conventions in dynare, and in particular the implementation of the log-linearized systems, introduce a subtle informational assumption. They write the evolution of log TFP as $a_t = \rho_t a_{t-1} + \epsilon_t$. Policies depend on $\{k_t, a_{t1}, \epsilon_t\}$, and the terms a_{t-1} and ϵ_t do not load according to the evolution of a_t. Trying to fit the neural net to these policies, conditioning only on $\{k_t, a_t\}$ leads to less than complete learning (up to 95% accuracy). Then, we will use$\{k_t, a_{t-1}, \epsilon_t\}$as an input for the neural net. We will also normalize those three vectors by subtracting the steady state and dividing by the standard deviation over dynare simulations. Thus, we denote observations as as $o_t = \{\tilde{k}_t, \tilde{a}_{t-1}, \tilde{e}_t\}$, where tildes are used to represent normalized variables. [13] As an output of this stage, which we pass to Python, we save the matrices $\{A, B, C, D\}$ and the standard deviation of all state and policy variables in the simulation. We are going to import in Python.

[13] Throughout the code, we will distinguish between state and observations, and normalized variables or not.

C.3.2 Pretrain Neural Net

Next, we "pretrain" the neural net to approximate the dynare policies. First, we simulate the model using the state space representation we extracted from dynare, and we store the states we visited and loglinear policy at each period. Second, we evaluate the neural net at the visited states, so we get the neural net policies. Third, we calculate the loss at each step, using the following loss function:

$$\mathcal{L}_t = \sum_{i=1}^{4N} (X_{it}^{neuralnet} / X_{it}^{loglinear} - 1)^2 / (4N)$$

We can also calculate the average accuracy in one period as

$$Acc_t = \sum_{i=1}^{4N} \left| \frac{X_{it}^{neuralnet} - X_{it}^{loglinear}}{X_{it}^{dynare}} \right| / (4N)$$

so 100% represents perfect fitting.

Table 5: Pretrain experiment hyperparameters

Hyperparameter	Value
Network Architecture	
NN hidden layers	[1024,1024]
Output layer	Softplus
Training Workflow	
episodes per step	1024
periods per episode	128
steps per epoch	100
number of epoches	1000
shock size scaling	1.5
Optimization	
Optimizer	Adam
Learning rate	5e-4 to 1e-5
batch size	16
Opt. Momentum	0.9
Mean gradients exp. decay	0.9
S.D. gradients exp. decay	0.999
Solution threshold	mean error $<$ 5e-6, max error $<$ 5e-5

In table 5, we show the configuration of the pre-train experiment. We will use vanilla Fully Connected Layer, also called Multi-layer Perceptron, with 2 hidden layers of 1024 nodes each. The input to the neural net is the normalized observations we just described in subsection C.3.1. We activate the output layer with a Softplus activation, which guarantees that the outputs are positive. The input to the neural net is the normalized observations we just described in subsection C.3.1. The targets that the neural net needs to hit are normalized in such a way that if the neural net outputs 1 for a variable, it is outputting the deterministic steady state of the variable. If it outputs 1.1, it is outputting a value that is 10% higher than the deterministic steady state.

In order to explain the training workflow, we need to introduce some notation. Whenever we update the neural net, we call it a step. But we let the computer to run for several steps before it gives us feedback on how the neural net is doing. We call this set of steps an epoch, so we have a parameter called steps per epoch, set to 100 in this case. Then, in a step, we are going to sample a number of episodes in parallel. An episode is a simulation of the model for a number of periods (in this case years). Hence, we have a parameter called episodes per step, set at 1024, and another parameter called periods per episode , set at 128. For the training, we scale up the shocks by 1.5 so we have more volatility.

For the optimization, we use Adam as an optimizer with default parameters. The learning rate follows a cosine decaying schedule starting at 5e-4 and ending at 1e-5. We split randomly all the periods collected in a step in minibatches of step 16, we calculate gradients of the loss function for each minibatch, and then we average the gradients across minibatches. While we are training for a fixed number of epochs, our criteria for the problem being solved is that the mean loss is below 5e-6 and the max loss is below 5e-5.

C.3.3 Solve the Nonlinear Model

FOllowing the same notation explained in the pretraining step, the configuration of the experiment is

Table 6: Train experiment hyperparameters

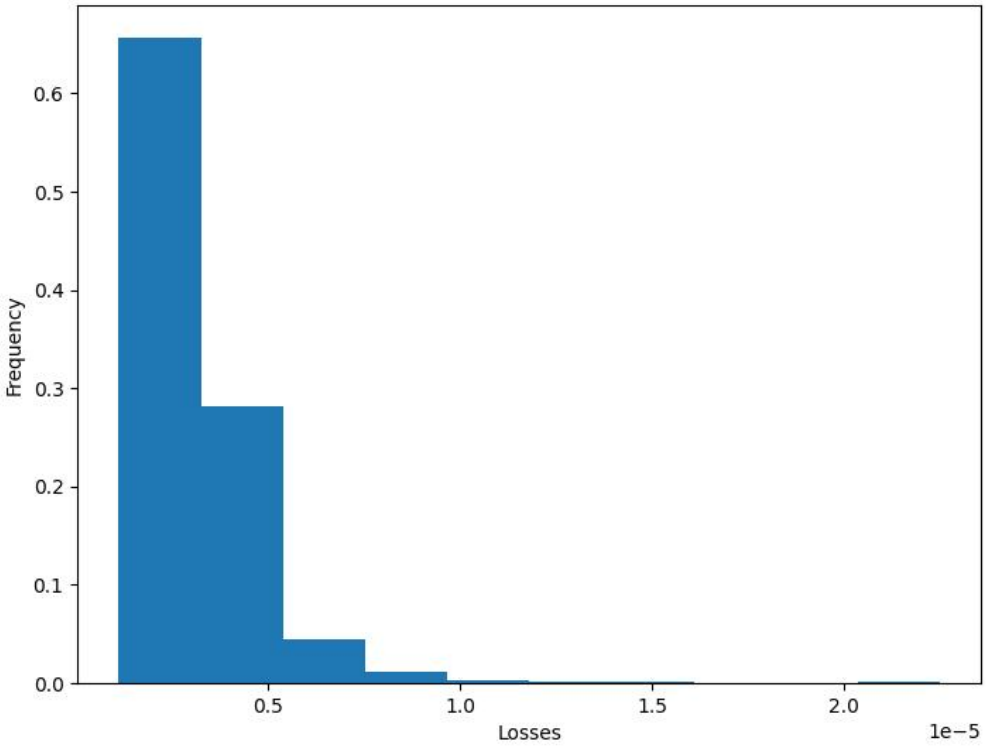
Hyperparameter	Value
Network Architecture	
NN hidden layers	[1024,1024]
Output layer	Softplus
Training workflow	
episodes per step	64
periods per episode	32
steps per epoch	100
number of epoches	1000
shock size scaling	1.5
Montecarlo draws	128
Optimization	
Optimizer	Adam
Learning rate	2e-5
batch size	16
Opt. Momentum	0.9
Mean gradients exp. decay	0.9
S.D. gradients exp. decay	0.999
Solution threshold	mean error $<$ 5e-6, max error $<$ 5e-5

Given that in the training step we need to estimate the expectations at each point of the state space that we visit, the training step is much more computationally intensive, so we have lower values for periods per episode and episodes per step. For the montecarlo estimation of expectations, we draw 128 samples of the shock vector in parallel (each shock vector has dimension 37 by 1). Since the pretraining leaves us close to the optimum, we fix the learning rate at a low value of 2e-5.

C.4 Accuracy Checks

In figure 11, we observe the histogram of period losses that result from a 10000 period simulation of the solved neural net policies. We see that most of the mass is concentrated below 5e-6.

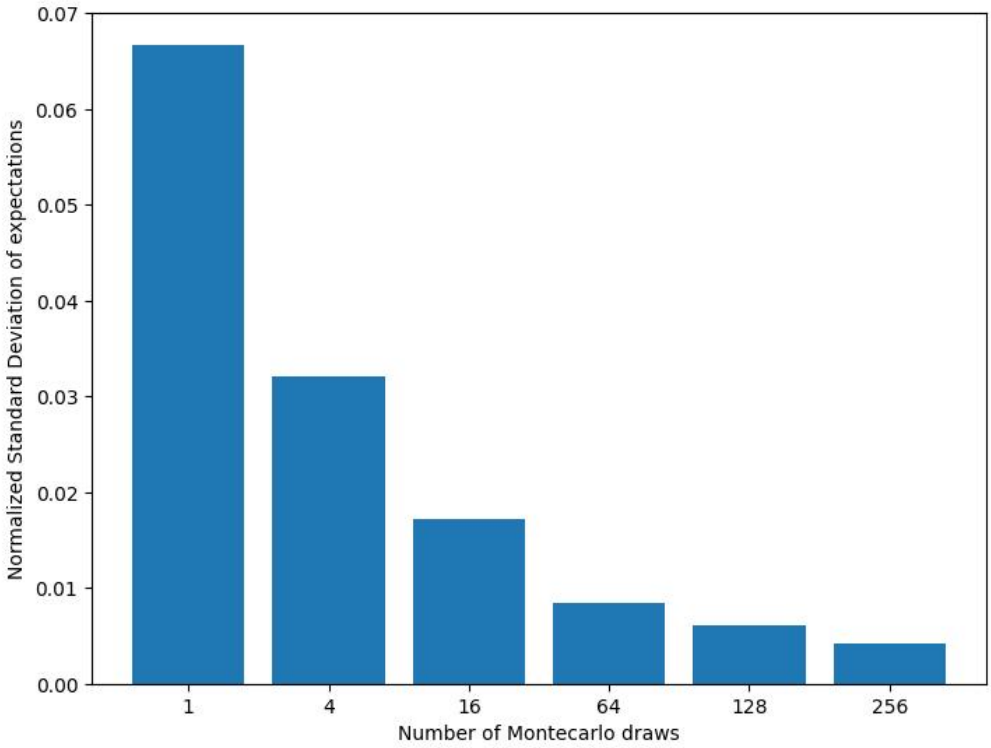
Figure 11: Distribution of Losses over Ergodic Distribution



Note: This figure shows the distribution of losses over the ergodic distribution. Ir order to calculate, we simulate the global solution for 10000 periods, calculate the loss in each period, and plot the histogram.

Another accuracy metric that is important to keep track of is the error introduced by the expectation estimation via the montecarlo method. In order to do this, we use the solved neural net and we calculate the expectation terms in random points of the state space. For each point of the state space, we repeat the montecarlo simulation 1000 times with different shock realizations. In figure 12, we see the accuracy of the montecarlo estimation of the expectation terms, measured as the standard deviation divided by the mean of the error. As we can see, the error diminishes considerably as we increase the number of random draws, but at 128, the number of draws we use, the error is already below 1%.

Figure 12: Error in Montecarlo Estimation of Expectations

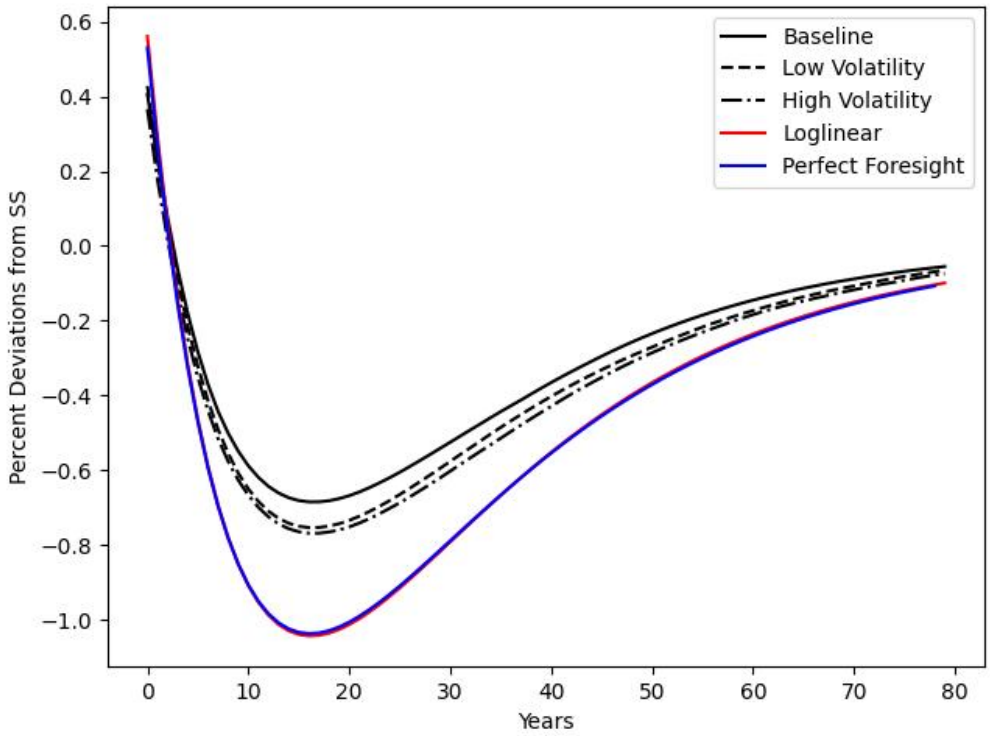


Note: This figure shows the accuracy of the montecarlo simulation method. In order to calculate it, we perform the simulation 1000 times for different draws of the shocks, and calculate accuracy statistics. In the y axis we present our measure of accuracy, that corresponds to the normalized standard deviation of the expectation estimation (standard deviation divided by the mean). The number of draws actually used in the experiments is 128.

D Auxiliary Figures for Section 5

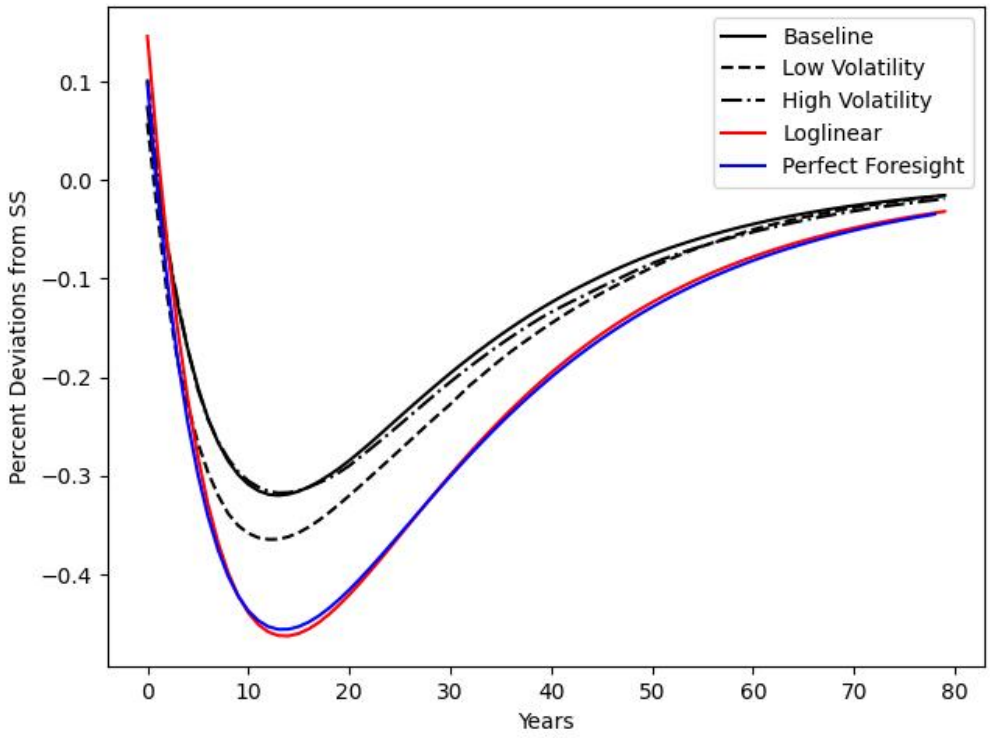
In this appendix, we present impulse responses for additional sectors.

Figure 13: Impulse Response of Aggregate Consumption to a Construction TFP Shock



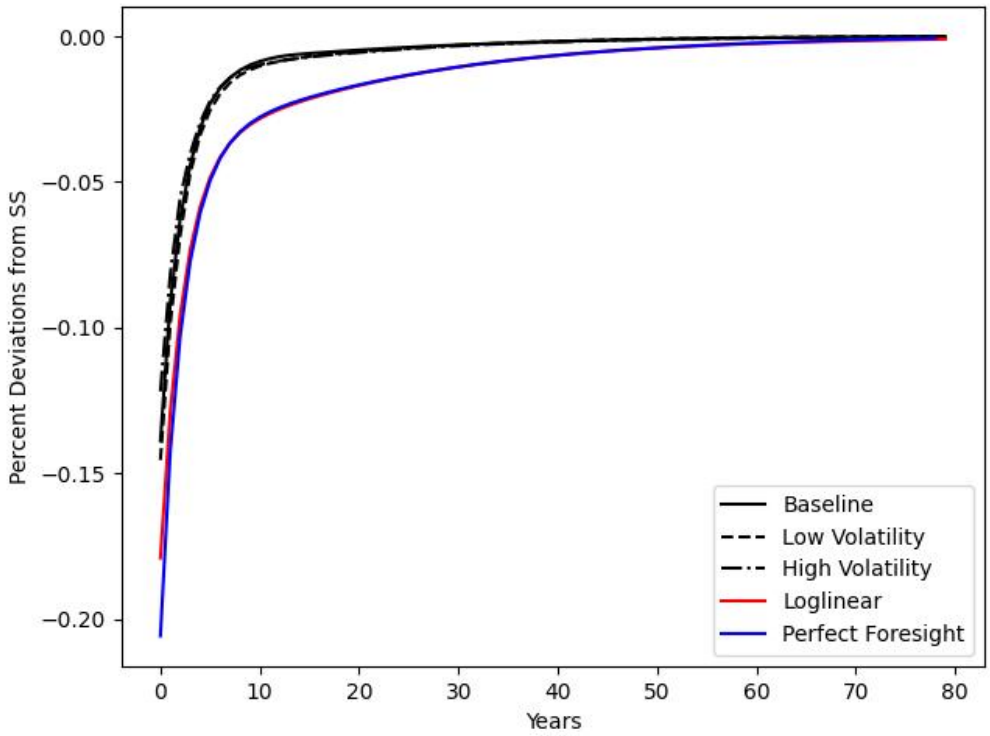
Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Construction. The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.

Figure 14: Impulse Response of Aggregate Consumption to a Machinery TFP Shock



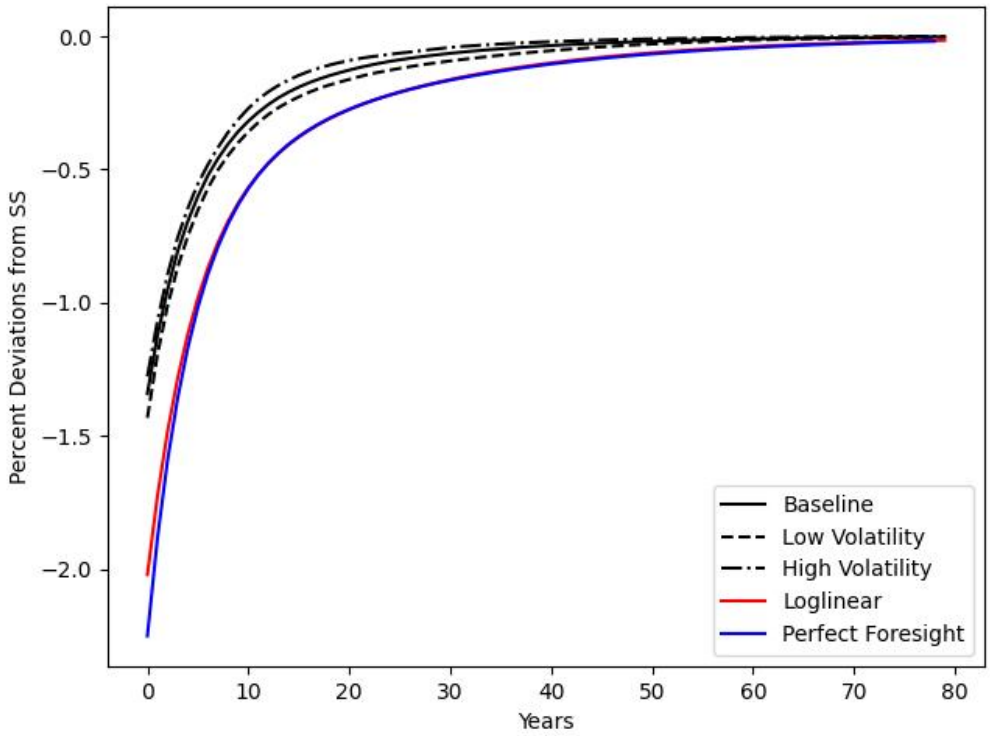
Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Machinery. The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.

Figure 15: Impulse Response of Aggregate Consumption to a Petroleum TFP Shock



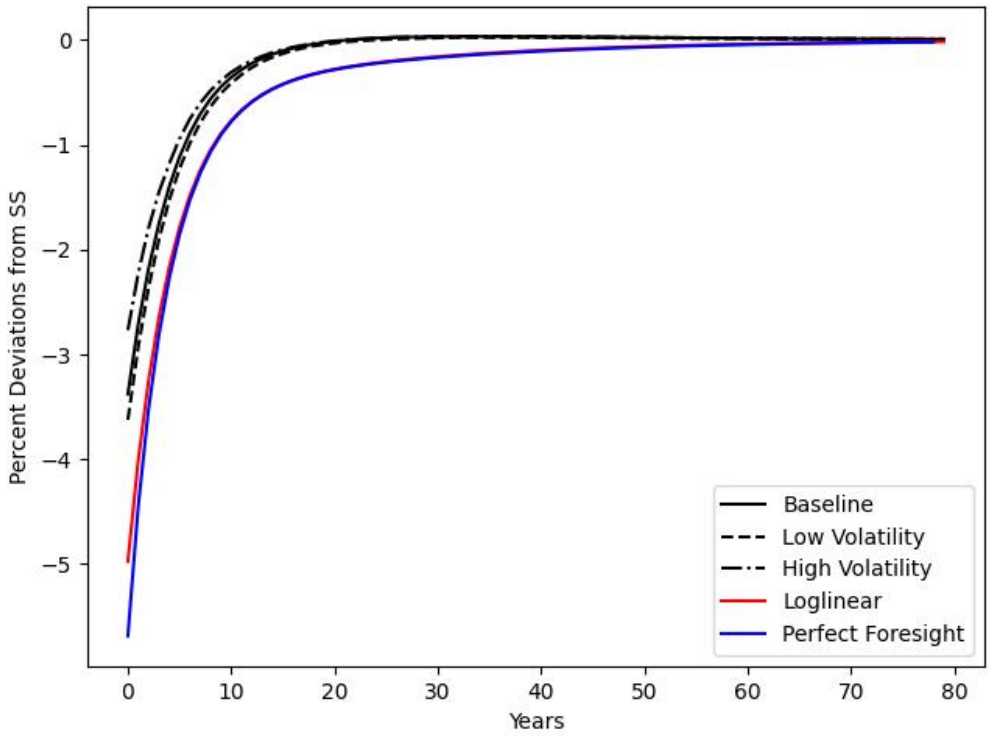
Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Petroleum. The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.

Figure 16: Impulse Response of Aggregate Consumption to a Retail TFP Shock



Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Retail. The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.

Figure 17: Impulse Response of Aggregate Consumption to a Real Estate TFP Shock



Note: This figure shows the impulse response of aggregate consumption to a 20% negative TFP shock in Real Estate. The black line represents the response using the global solution method, the red line shows the response using a log-linear approximation, and the blue line shows the perfect foresight solution. For the global solutions, the vertical axis shows log deviations relative to the stochastic steady state, while for the log-linear and perfect foresight solutions we show percentage deviations from the deterministic steady state.