Distribution of Market Power, Endogenous Growth, and Monetary Policy *

Yumeng Gu[†] Sanjay R. Singh[‡]

February 2024

Abstract

We incorporate incumbent innovation in a Keynesian growth framework to generate an endogenous distribution of market power across firms. Existing firms increase markups over time through successful innovation. Entrant innovation disrupts the accumulation of market power by incumbents. Using this environment, we highlight a novel misallocation channel for monetary policy. A contractionary monetary policy shock causes an increase in markup dispersion across firms by discouraging entrant innovation relative to incumbent innovation. We characterize the circumstances when contractionary monetary policy may increase misallocation.

Keywords: Monetary policy, Markup dispersion, Allocative efficiency, Endogenous distribution of market power.

*We thank David Baqaee, Nicolas Caramp, James Cloyne, Pablo Cuba-Borda, Keshav Dogra, Romain Duval, Antonio Fatás, Robert Feenstra, John Fernald, Luca Fornaro, Òscar Jordà, Matthias Meier, Katheryn Russ, Michaela Schmöller, Alan Taylor, and seminar participants at UC Davis, the North American Summer, the Asia, the Australia, and the China Meetings of the the Econometric Society 2021 for helpful comments. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco, the Federal Reserve System, the International Monetary Fund, IMF Executive Board, or the IMF management. All errors are our own.

[†]International Monetary Fund (ygu@imf.org).

[‡]Federal Reserve Bank of San Francisco and Department of Economics, University of California, Davis (sjrsingh@ucdavis.edu).

1 Introduction

In this paper, we study the business cycle interaction of monetary policy with TFP growth and product market power. We characterize a novel channel through which monetary policy can affect allocative efficiency in the economy.

We incorporate incumbent innovation into a *Keynesian* growth model (Benigno and Fornaro, 2018). In the model, firms accumulate market power through incumbent innovation over time (Peters, 2020). Successful innovation by entrants results in the displacement of existing firms and disrupts the accumulation of market power by the incumbents. A cross-sectional distribution of markups endogenously responds to business cycle shocks through variations in entrant innovation. A contractionary monetary policy shock reduces the incentives for potential entrants to innovate. If existing firms continue to innovate unaffected by such shocks, they accumulate higher market power with successful innovation. Thus, a monetary policy-induced recession may reduce the incentives for potential entrants to innovate, lower the allocative efficiency across firms, and lead to an endogenous slowdown in TFP growth.

The interaction of market power and monetary policy is at the heart of the new Keynesian models. In a standard textbook new Keynesian model (Galí 2015, Chapter 3), the assumption of Calvo price rigidities introduces price dispersion across firms. When there is a contractionary monetary policy shock, some firms get to reset prices while others do not. The firms that reset prices maintain a constant markup, while the markups go up for firms that cannot adjust prices since marginal costs fall with reduced aggregate demand. In the textbook model, with monopolistic competition, homogeneous price rigidity across firms, and constant elasticity of demand, the markup dispersion goes up irrespective of whether the shock is contractionary or expansionary. The markup dispersion can decrease with expansionary monetary policy shocks in our setting.

Our misallocation channel is independent of the co-variability of pass-through with the level of markups. We show that even when the pass-through of marginal costs into markups is the same for all firms, monetary policy has misallocation effects through its impact on the cross-sectional distribution of market power across firms. The effect arises entirely from the extensive margin of firm entry. To our knowledge, this mechanism is distinct from other complementary analyses of misallocation effects of monetary policy. Baqaee, Farhi and Sangani (2024) and Meier and Reinelt (2022) study allocative efficiency effects of monetary policy and are closely related to the theme of our paper. A key to the misallocation channel highlighted in these papers is the negative covariance between the level of markups and the pass-through of marginal costs into prices. In Meier and Reinelt (2022), this negative covariance endogenously arises from heterogeneous price rigidity across firms in an environment with aggregate risk. Baqaee et al. (2024) consider general settings when the negative covariance could arise from variable elasticity of demand faced by the firms or heterogeneous price rigidities. Relatedly, Baqaee and Farhi (2019) provide a general treatment of misallocation in general equilibrium models (see also Basu and Fernald, 2002).

The quantitative importance of the mechanism we highlight is limited. We explain why the degree of misallocation is bounded above by two forces that counteract each other. Following a contractionary monetary policy shock, the entrant innovation rate declines relative to the incumbent innovation rate. A *dispersion* effect manifests through an increase in markup dispersion, giving rise to greater aggregate misallocation. Meanwhile, a *level* effect occurs as the average level of markups rises alongside the increase in markup dispersion. This conditionally counter-cyclical movement in the average markup level (i.e., the average markup level increases in response to a monetary contraction) generates an opposing force to the misallocation effect of monetary policy arising from the rise in markup dispersion. These two opposing forces preclude us from finding quantitatively large effects. In an oligopolistic competition environment with heterogeneous firms that feature an endogenous return to scale and customer acquisition, Gu (2022) remedies this dampening of misallocation through a mechanism that generates a pro-cyclical response of average markups to a monetary easing, potentially aligning these two forces to increase misallocation.

We further ask how the resulting misallocation depends on the initial condition of the economy. In particular, whether the misallocation is higher in a steady state with a high entrant innovation rate relative to incumbent innovation. We show that the state dependency is ambiguous and that a higher rate of steady state entrant innovation does not necessarily lead to more misallocation. A caveat of these exercises is that the initial conditions are functions of the deep parameter of the model. For our baseline parametrization, we numerically show how the degree of misallocation varies according to the steady-state level of the entrant's innovation. Specifically, we find that while the level effect weakens monotonically as the entrant's innovation increases, the dispersion effect is stronger at very low and very high levels of the entrant's innovation rate. The dispersion effect dominates in magnitude, and the misallocation worsens as the steady state level of entrant innovation rises, holding the incumbent's rate of innovation fixed.

In Section 2, we show the misallocation channel of monetary policy through a series of propositions. The baseline model assumes that the potential entrants are myopic and that the incumbent innovation is exogenous. These assumptions give us analytical tractability and help highlight our key mechanism. We generalize these results in Section 3 by first, relaxing the myopic assumption for the potential entrants, second, endogenizing incumbent innovation over the business cycles, and third, adopting a more general constant-elasticity-of-substitution (CES) demand structure. The misallocation channel of monetary policy we highlight is robust to these theoretical extensions.

We contribute to the literature on the interaction of monetary policy with the productive potential of the economy (Anzoategui, Comin, Gertler and Martinez, 2019; Bianchi, Kung and Morales, 2019; Garga and Singh, 2021; Moran and Queralto, 2018) by endogenizing allocative efficiency over the business cycle.[1] Relative to papers considering endogenous productivity growth in business cycle models, our model underscores the short-run variation in allocative efficiency across firms and the medium-run effects of monetary policy. Our setup is closer to that of Benigno and Fornaro (2018), who model a scenario where pessimistic expectations put the economy on a low growth trajectory. We build on their work and endogenize a distribution of market power across firms by allowing incumbent firms to accumulate market power over time.

Peters (2020) characterized the misallocation effects from changes in the distribution of market power in a model with both incumbent and entrant innovation.[2] Peters and Walsh

[1]Cerra, Fatás and Saxena (2022) provide a recent literature review. See Cerra and Saxena (2008); Jordà, Singh and Taylor (2021) for evidence on persistent effects of temporary shocks or Jordà, Schularick and Taylor (2013) for evidence on deep and protracted recovery/non-recovery from financial crises relative to normal recessions for advanced economies. For additional references, see, for example, Acharya, Bengui, Dogra and Wee (2022); Annicchiarico and Pelloni (2016); Ates and Saffie (2021); Cloyne, Martinez, Mumtaz and Surico (2022); Fatás and Mihov (2013); Fatás and Singh (2024); Fornaro and Wolf (2023); Furlanetto, Lepetit, Robstad, Rubio Ramírez and Ulvedal (2023); Guerron-Quintana and Jinnai (2019); Queralto (2020, 2022); Vinci and Licandro (2021).

[2]Acemoglu and Cao (2015) provide a textbook extension of the Schumpeterian growth framework incorporating simultaneous innovation by incumbents and entrants.

(2020) extend Peters (2020)'s framework to quantitatively assess the role of low population growth in explaining several secular trends in the US. We extend Peters (2020)'s work to study the misallocation effect of monetary policy through its impact on the distribution of markups across firms. Greaney and Walsh (2023) also uses a similar framework to study the amplification of debt deleveraging during the Great Recession.

These misallocation and business cycle analyses relate to an earlier literature studying firm dynamics. Jaimovich and Floetotto (2008) and Etro and Colciago (2010) propose models in which oligopolistic competition leads to an inverse relationship between the number of competitors and the level of price markup. Bilbiie, Ghironi and Melitz (2008, 2012) model a negative relation between markups and the number of firms via translog preferences on consumers (Feenstra, 2003). A contractionary business cycle shock by reducing firm entry, and thus the number of competitors, makes demand less elastic and allows firms to charge higher markups.[3] In our paper, the distribution of markups across firms and not just average markup is endogenous. Moreover, the endogeneity of average markups here arises from the interaction of entrant and incumbent innovation and not due to strategic decisions by firms with a time-varying number of competitors. Atkeson and Burstein (2008) and De Blas and Russ (2015) study models with heterogeneous firms and a distribution of markups that arises endogenously due to the strategic behavior of firms. As noted in Baqaee et al. (2024), the covariance between the pass-through of marginal costs into prices and the level of markups is key in these models to generate time-varying misallocation.

Our misallocation channel is also not covered in the Darwinian effect of firm entry isolated

[3]See also Bergin and Corsetti (2008) and Faia (2012) for additional analyses of monetary policy in oligopolistic environments.

by Baqaee, Farhi and Sangani (2023). The Darwinian effect of entry relies on non-CES demand structure – the entry of new firms causes the aggregate price index to fall, resulting in a reallocation towards high-markup (low price elasticity of demand) firms, holding markups constant. Not all incumbents are exposed to firm entry similarly, leading to a selection of the fittest/most productive. In our setup, the Darwinian effect is silenced since all incumbents are identically exposed to firm entry, and there exists no selection of the most productive.

In more recent work, Morlacco and Zeke (2021) connect low-interest rates to the increased market power of firms through a customer capital channel (Gourio and Rudanko, 2014) in a duopolistic industry setup. Afrouzi and Caloi (2023) connect cyclicality of output growth to cyclicality of markups in models with variable markups. Afrouzi, Drenik and Kim (2023) investigate the role of customer acquisition and average sales in determining a firm's market share and price-cost markup, respectively. Liu, Mian and Sufi (2022) emphasize a link between market power, real interest rates, and low productivity growth in an endogenous growth model with two firms (leader and follower) within an industry engaged in a gradual step-by-step innovation race (Aghion, Harris, Howitt and Vickers, 2001). Duval, Furceri, Lee and Tavares (2021) and Ferrando, McAdam, Petroulakis and Vives (2021) find low product-markup firms tend to be more responsive to monetary policy shocks than high markup firms. Duval, Furceri, Lee and Tavares (2023) explores this state dependence through a financial constraints channel (Aghion, Farhi and Kharroubi, 2015, 2019). Burya, Mano, Timmer and Weber (2022) investigate the interaction between monetary policy and labor market power.

Structure of the paper. The rest of the paper is organized as follows. Section 2 proposes the baseline theoretical model and discusses the main theoretical implications in a series of propositions. Section 3 presents various theoretical extensions that generalize our baseline results, and discusses consistent empirical evidence from the existing literature. Section 4 concludes.

2 Theory

This section builds heavily on earlier seminal work of Aghion and Howitt (1992) and more recent works of Benigno and Fornaro (2018) and Peters (2020). In an otherwise standard endogenous growth framework of vertical innovation, we follow Benigno and Fornaro (2018) in constructing a *Keynesian growth* model by incorporating nominal wage rigidities. Nominal wage rigidities introduce a mechanism for monetary policy to have real effects. The new element we introduce to the Benigno and Fornaro (2018) framework is that both the entrants and the incumbent firms can innovate, and these two forces of creative destruction (the traditional force) and own innovation (the new addition) interact with each other to collectively determine the growth rate of productivity and endogenously generate a cross-sectional distribution of markups. Monetary policy shocks can affect the cross-sectional distribution of markups by changing incentives for innovation by entrants.

2.1 The Environment

Households. The economy is populated with a measure one of infinitely lived households. Each household has the following preferences over a unique final consumption good. The lifetime utility of the representative household is given by

$$\mathbf{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} \right) \right],$$

where C_t denotes consumption at time t, $\beta \in (0,1)$ is the subjective discount factor, and σ is the inverse of the elasticity of intertemporal substitution.

Each household is endowed with one unit of labor, and labor supply is inelastic. There is a one-period risk-free bond B_t traded in the units of currency and pays a (net) nominal interest rate i_t. Households also own an equal share of all firms, and they receive Φ_t in dividends from the firms' profits each period.

The intertemporal problem of a representative household is to choose C_t and B_{t+1} to maximize its lifetime utility subject to a budget constraint and a non-Ponzi constraint:

$$P_t C_t + B_{t+1} = W_t L_t + B_t(1 + i_{t-1}) + \Phi_t,$$

where P_t is the price of the final good in units of currency,[4] and W_t is the nominal wage. As such, $W_t L_t$ is the household's labor income. Gross inflation rate, Π_t, is defined as the growth rate of P_t.

Final Good Production. The perfectly competitive final good, Y_t is a Cobb-Douglas aggregate of a continuum of differentiated intermediate goods, and each variety x_{it} is denoted with subscript i such that

$$\ln Y_t = \int_0^1 \ln x_{it} di. \tag{1}$$

We assume monopolistic competition across intermediate good varieties. The assumption of unitary elasticity of demand provides analytical tractability. Let P_t denote the aggregate price level. Then the demand for each variety i is $x_{it} = \frac{P_t Y_t}{p_{it}}$, where p_{it} is the price of

[4]Following Woodford (1998), our setup is that of a "cashless limit" environment of a monetary economy.

intermediate variety i. That is, all varieties have equal sales-share in final output. The aggregate price level is then given by: $\ln P_t = \int_0^1 \ln p_{it} di$.

Intermediate Goods Production. We assume that within each product market i, there are a number of firms S_{it}, each denoted with f, that supply a perfectly substitutable good x_{fit}. That is, $x_{it} = \sum_{f \in S_{it}} x_{fit}$. Each firm f operates a constant returns to scale technology: $x_{fit} = q_{fit}\, l_{fit}$, where q_{fit} is productivity of the firm, and l_{fit} is the labor hired by the firm. A firm f takes economy-wide wage rate as given. Firms within a given variety i compete with each other á la Bertrand competition. As a result, only the most efficient firm in market i produces the intermediate variety q_{it}. The presence of competing firms, however, implies a limit pricing strategy for the active producer, i.e., the most efficient firm, within a given variety i, sets price equal to the marginal cost of the immediate follower or the second most efficient firm within that variety. Denote the immediate follower with a superscript F. The equilibrium markup for variety i is thus given by

$$\mu_{it} \equiv \frac{p_{it}}{W_t/q_{it}} = \frac{W_t/q_{it}^F}{W_t/q_{it}} = \frac{q_{it}}{q_{it}^F}, \tag{2}$$

where W_t is the equilibrium wage that the firms take as given, and $\frac{W_t}{q_{it}^F}$ denotes the marginal cost of the immediate follower F. We suppress the subscript f since only the leader is active in production of variety i. Equilibrium profits for a producer of variety i are given as $\Phi_{it} = (1 - \mu_{it}^{-1}) P_t Y_t$.

Innovation and Creative Destruction. Growth in the model stems from innovations that improve the quality of the intermediate varieties. These innovations come from two sources: (i) creative destruction from a new firm that displaces the incumbent in production of variety i, and (ii) own innovation from the incumbent that increases its market power.[5] Following Aghion and Howitt (1992), productivity improvements, within each variety, are ranked on a quality ladder with each successive improvement $\lambda > 1$ step size higher than the previous innovation. Regardless of whether the innovation arises from incumbents or the entrants, we assume the same step-size improvement.

Creative destruction from an entrant reduces the quality gap to unity, whereas own innovation increases quality gap, thus empowering the incumbent to charge higher markups. This innovation structure implies that the equilibrium markups can be written as

$$\mu_{it} = \frac{q_{it}}{q_{it}^F} = \frac{\lambda^{s_{it}}}{\lambda^{s_{it}^F}} = \lambda^{s_{it}-s_{it}^F} \equiv \lambda^{\Delta_{it}}, \tag{3}$$

where s_{it} and s_{it}^F denote respective steps on the ladder, and $\Delta_{it} = s_{it} - s_{it}^F \geq 1$ represents an active firm's productivity advantage relative to its immediate follower. We also assume that there is a large exogenous upper limit in maximum quality gap, denoted by a finite positive integer $\bar{N}$. For the baseline model, we focus on the case with the maximum quality gap $\bar{N} \to \infty$. In Section 2.5 we show that in the log-linearized equilibrium, we do not need to account for the quality gap distribution, and aggregation of the economy ensures a finite state space.[6] In the appendix, we provide steady state equilibrium conditions and log-linearized competitive equilibria for both cases when $\bar{N}$ is a finite number and the limit

[5] Peters (2020) also allows two additional sources of innovations: (i) entrants innovating on new product varieties without displacing any incumbents, and (ii) incumbents expanding their product lines. We abstract from those two sources.

[6] Presenting the special case with a finite $\bar{N}$ is nevertheless useful, as we study theoretical extensions involving the CES preferences, in which firms with sufficiently large quality gaps would charge the unconstrained CES markup in place of the limit price, effectively imposing an upper limit to quality gap Δ.

case when $\bar{N} \to \infty$.

Consequently, equilibrium profits for intermediate good producer i can be written as function of the quality gap between the productivity of the leader and the immediate follower. That is, $\Phi_{it} = (1 - \lambda^{-\Delta_{it}})P_t Y_t \equiv \Phi_t(\Delta_i)$.

Each potential entrant, in variety i, invests labor resources into conducting research and development (R&D). With probability $0 < z_{it} < 1$, she is successful in making a process improvement, and gets the monopoly rights (patent) of the production of intermediate variety i in the following period. If she fails to innovate, the incumbent continues to produce with its productivity q_{it}. The entrepreneur in each variety i chooses probability z_{it} to maximize the expected discounted profits (from the patent). As in Benigno and Fornaro (2018), we assume that the monopolist gets the patent for only one period. That is,

$$\max_{z_{it}} \mathbf{E}_t \left[D_{t,t+1} z_{it} \Phi_{t+1}(1) \right] - \delta_z z_{it}^{\zeta_z} W_t,$$

where $\zeta_z \geq 1$ denotes the curvature of the cost function, $D_{t,t+1} = \beta \frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}} \frac{P_t}{P_{t+1}}$ is the stochastic discount factor, and $\Phi_t(1)$ is the equilibrium profits of a producer with unit quality gap.[7] Each potential entrant hires $\delta_z z_{it}^{\zeta_z}$ workers to conduct R&D. We focus on a symmetric equilibrium whereby entrants in all varieties i choose the same innovation success rate z_t.

Conditional on not being displaced by an innovating entrant, we assume that the firm's ownership is allocated to a random entrepreneur in each successive period.[8] With an ex-

[7]We make parameter assumptions to ensure that $z_t > 0$ in competitive equilibrium. We therefore do not explicitly incorporate the complementary slackness condition in the first order condition of the entrant's problem.

[8]This assumption of one period patents is made following Aghion and Howitt (2008) and Acemoglu, Aghion, Bursztyn and Hemous (2012). It considerably simplifies the analysis. In their papers, it is made when the existing firm does not undertake any new innovation. In our setup, we allow the surviving firm to innovate at an exogenous rate.

ogenous probability $0 < I_t < 1$, the non-displaced firm accumulates market power through successful innovation in process improvements. This assumption simplifies our analysis considerably. Furthermore note that while we assume an exogenous incumbent innovation rate I for our baseline model. In Section 3.4, we discuss evidence that R&D expenditures of U.S. listed firms do not exhibit a significant response to monetary policy shocks. Given our focus on short-run responses to temporary adverse shocks, we view this assumption as empirically relevant. We nevertheless relax this assumption in theoretical extensions.

Cross-Sectional Distribution of Markups. Let $\{\upsilon_t(\Delta)\}_{\Delta=1}^{\bar{N}}$ denote the measure of product varieties with quality gap Δ at time t. Given an initial distribution of markups, the next period distribution of markups evolves according to innovations by entrants and incumbents, with the following laws of motion:

$$\upsilon_{t+1}(\Delta) = z_t + (1 - z_t)(1 - I)\upsilon_t(\Delta) \quad \text{if } \Delta = 1, \tag{4}$$

$$\upsilon_{t+1}(\Delta) = (1 - z_t)(1 - I)\upsilon_t(\Delta) + (1 - z_t)I\upsilon_t(\Delta - 1) \quad \text{if } 2 \leq \Delta \leq \bar{N} - 1, \tag{5}$$

$$\upsilon_{t+1}(\Delta) = (1 - z_t)\upsilon_t(\Delta) + (1 - z_t)I\upsilon_t(\Delta - 1) \quad \text{if } \Delta = \bar{N}. \tag{6}$$

where equation 4 describes the evolution of the measure of firms at quality gap 1. On average z_t varieties will be creatively destroyed by entrants and will feature price-markups given by unit quality gap. In remaining sectors, that is measure $1 - z_t$, incumbents could move to higher step ladder and charge higher markups with probability I. Of the measure of product varieties that were at quality gap 1, $(1 - z_t)(1 - I)$ fraction will remain at quality gap 1. Likewise, in equation 5, for a given quality gap Δ less than $\bar{N}$ and larger than 1,

$(1 - z_t)(1 - I)\upsilon_t(\Delta)$ measure of varieties will survive displacement by entrants and not be able to move up to a higher step ladder; meanwhile, $(1 - z_t)I\upsilon_t(\Delta - 1)$ measure of varieties will survive displacement by entrants and move up the quality ladder and charge a higher markup based on the new quality gap Δ. Similarly, equation 6 summarizes the evolution of quality gap $\Delta = \bar{N}$ when further own innovation is no longer possible. As $\bar{N} \to \infty$, this equation is dropped in the baseline model.

2.2 Nominal Rigidities

We introduce nominal rigidities through a simple indexation rule following Benigno and Fornaro (2018). Nominal wages are assumed to follow the following process:

$$W_t = \pi^w W_{t-1}, \tag{7}$$

where $\pi^w > 0$ is a constant wage-inflation rate.

2.3 Government Policy

As in Benigno and Fornaro (2018), the central bank sets the nominal interest rate on short-term government bond according to the following rule:

$$\frac{1 + i_t}{1 + i_{ss}} = \left(\frac{L_t - \delta_z z_t^{\zeta_z}}{L_{ss} - \delta_z z_{ss}^{\zeta_z}} \right)^{\phi} \epsilon_t^m = \left(\frac{L_{pt}}{L_p} \right)^{\phi} \epsilon_t^m, \tag{8}$$

where $\phi > 0$ and ϵ_t^m is an exogenous AR(1) process with persistence $\rho \in (0, 1)$. The central bank sets nominal interest rate in order to target deviation of employment in the production sector from its steady state level by cutting the interest rates in response to falling employment levels. In addition, the short-term government bonds are in net zero supply, and the government balances it budget every period.

2.4 Aggregation and Market Clearing

We can now discuss aggregation of the economy. In equilibrium, labor hired by each firm i is given by

$$l_{it} = \frac{x_{it}}{q_{it}} = \frac{P_t Y_t}{p_{it} q_{it}} = \frac{P_t Y_t}{\mu_{it} W_t} = \frac{P_t Y_t}{W_t} \mu_{it}^{-1}, \tag{9}$$

and the total amount of labor hired by intermediate goods producers, denoted with L_{pt}, is given by

$$L_{pt} = \int_0^1 l_{it} di = \frac{P_t Y_t}{W_t} \int_0^1 \mu_{it}^{-1} di. \tag{10}$$

Equilibrium nominal wage W_t is given by

$$W_t = exp(\ln P_t + \int_0^1 \ln q_{it}^F di) = P_t \times exp(\int_0^1 \ln \frac{q_{it}}{\mu_{it}} di) = P_t Q_t \times exp(\int_0^1 \ln \mu_{it}^{-1} di), \tag{11}$$

where $\ln Q_t = \int_0^1 \ln q_{it} di$ denotes technical efficiency of the economy. Long-run growth in this economy is captured by increases in Q_t. Entrant and incumbent innovation increases quality of intermediate goods by a step size of λ. A fraction z_t of products are improved through entrant innovation each period. After the entrant innovation is realized, I fraction of the remaining sectors see an improvement through incumbent innovation. Firms produce with improved quality in the following period. Consequently, the growth rate of technical efficiency, denoted by g_{t+1}, is given by

$$g_{t+1} = \frac{Q_{t+1}}{Q_t} = (z_t + I - z_t I)(\lambda - 1) + 1$$

Using equations (10) and (11), Aggregate output can be written as

$$Y_t = Q_t L_{pt} \mathcal{M}_t, \text{ where } \mathcal{M}_t \equiv \frac{\exp\left(\int_0^1 \ln \mu_{it}^{-1} di\right)}{\int_0^1 \mu_{it}^{-1} di} = \frac{\exp\{-\mathbf{E}_i[\ln \mu_{i,t}]\}}{\mathbf{E}_i[\mu_{i,t}^{-1}]}, \tag{12}$$

and $\mathcal{M}_t$ is a measure of allocative efficiency in the economy. Higher markup dispersion, leads to a lower value of $\mathcal{M}_t$. From equation (12), notice that $\mathcal{M}_t = 1$ if and only if markups are all equalized across producers. In the absence of markup dispersion, there are no allocative efficiency losses. Furthermore, $\mathcal{M}_t \leq 1$, and $\mathcal{M}_t$ decreases as the dispersion of markups increases, i.e., higher markup dispersion reduces allocative efficiency.

Labor market clears, with labor used in intermediate goods' production and R&D adds up to the total labor employed, L_t, in the economy:

$$L_{pt} + \delta_z z_t^{\zeta_z} = L_t \leq 1$$

Market clearing for final goods implies that total consumption expenditure equals total output produced:

$$Y_t = C_t$$

2.5 Equilibrium

We formally define the competitive equilibrium of the economy in Appendix A.1. In order to obtain a stationary system of equations, we normalize the equilibrium equations by dividing the non-stationary level of output by the level of technical efficiency Q_t, and dividing nominal wage level with the product of aggregate price level, P_t, and level of technical efficiency Q_t. We define $y_t = \frac{Y_t}{Q_t}$ as the normalized output, $w_t = \frac{W_t}{P_t Q_t}$ as the normalized real wage. Likewise, we formally define the normalized competitive equilibrium in Appendix A.1.

We find the balanced growth path (BGP) of the economy by imposing restrictions on parameters such that deterministic steady state equilibrium satisfies (i) there is full employment, $L = 1$, (ii) entrant innovation rate is bounded, $z \in \left(0, \min\left(1, \frac{1}{\delta_z^{\zeta_z - 1}}\right)\right)$, and (iii) net nominal interest rate is positive, $i > 0$. The upper bound on z ensures that the (normalized) consumption is positive in the steady state. In numerical simulations, we verify that $z_t \in (0, 1)$, and $c_t > 0$ for all t.[9]

Steady State. As the maximum quality gap $\bar{N} \to \infty$, the steady state equilibrium is characterized by a set of constant values $\{g, y, z, L, i, \Pi\}$ that satisfies six equations provided in Appendix A.3. Proposition 1 states that in steady state, the difference equations from Equations 4 - 6 can be used to derive an analytical expression for the cross-sectional distribution of markups, which in turn can be used to derive analytical expression of the aggregate misallocation measure $\mathcal{M}$. This case of $\bar{N} \to \infty$ is a restatement of the result derived by Peters (2020).

[9] If there is no incumbent innovation, that is the maximum quality gap $\bar{N} = 1$, the model simplifies to the setup in Benigno and Fornaro (2018).

Proposition 1. *Let z be the steady state probability of successful entry for outside entrepreneurs in a stationary competitive equilibrium, and let $\vartheta_I \equiv \frac{z}{(1-z)I}$ and $\phi_I \equiv \frac{1}{1+\vartheta_I}$. As the maximum quality gap $\bar{N} \to \infty$, then following results are obtained along the balanced growth path:*

1. The distribution of quality gaps Δ is given by the following probability mass function $\upsilon(\Delta)$:

$$\upsilon(\Delta) = \left(\frac{1}{1+\vartheta_I}\right)^{\Delta} \vartheta_I \ \forall \Delta \tag{13}$$

2. The distribution of markups is given by the following cumulative distribution function:

$$F(\mu, \Delta) = 1 - \left(\frac{1}{\mu}\right)^{\frac{\ln(1+\vartheta_I)}{\ln \lambda}} \quad \forall \Delta \tag{14}$$

3. The measure of aggregate misallocation is given by:

$$\mathcal{M} = \left(\lambda^{-\frac{1+\vartheta_I}{\vartheta_I}}\right)\left(\frac{\lambda(1+\vartheta_I)-1}{\vartheta_I}\right) \tag{15}$$

Proof. See Appendix A.2. □

Proposition 1 states that the probability of successful entry z relative to the probability for incumbents to survive displacement (i.e., $(1-z)I$) determines the distribution of markups and the measure of aggregate misallocation along the BGP. Furthermore, the distribution of markups approximates a Pareto distribution with scale parameter equals to 1 and tail parameter equals to $\frac{\ln(1+\vartheta_I)}{\ln \lambda}$ for all Δ. The tail parameter is increasing in z, such that if the steady state rate of successful entry declines, the distribution of markups would have a thicker tail, and the dispersion as summarized by $\mathcal{M}$ would go up.

When $\bar{N} \to \infty$, our model collapses to the one studied by Peters (2020). It is nevertheless useful to consider a case with a finite $\bar{N}$, which we later rely on in extensions of the baseline model. We summarize the special case for a finite maximum quality gap $\bar{N} < \infty$ as the following corollary:

Corollary 1. *For a finite maximum quality gap $\bar{N} < \infty$, the following results hold alone the BGP respectively:*

1. $\upsilon(\Delta) = \begin{cases} \left(\frac{1}{1+\vartheta_I}\right)^{\Delta} \vartheta_I, & \text{if } \Delta = 1, 2, ..., \bar{N}-1 \\ \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} & \text{if } \Delta = \bar{N} \end{cases}$

2. $F(\mu, \Delta) = \mathbf{P}(\lambda^{\Delta} \leq \mu) = \begin{cases} 1 - \left(\frac{1}{\mu}\right)^{\frac{\ln(1+\vartheta_I)}{\ln \lambda}}, & \text{if } \Delta = 1, 2, ..., \bar{N}-1 \\ 1, & \text{if } \Delta = \bar{N} \end{cases}$

3. $\mathcal{M}_{\bar{N}} = \frac{\lambda^{-\Xi_{\bar{N}}}}{\Lambda_{\bar{N}}}$,

where $\Lambda_{\bar{N}} = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}\left[1 - \left(\frac{1}{\lambda(1+\vartheta_I)}\right)^{\bar{N}-1}\right] + \lambda^{-\bar{N}}\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}$, *and* $\Xi_{\bar{N}} = \vartheta_I\left[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I}\right] + \bar{N}\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}$.

Given that the measure of quality gaps $\upsilon(\Delta)$ is probability mass function (pmf) in the BGP, it follows that $\upsilon_t(\Delta)$ is also a pmf. We summarize this result in Corollary 2.

Corollary 2. *Since $\upsilon(\Delta)$ is a probability mass function for discrete variable all quality gap Δ in steady state, $\upsilon_t(\Delta)$ would also be a probability mass function for all quality gap Δ.*

Log-linearized Equilibrium. We log-linearize the normalized competitive equilibrium around the steady state and define the following log-linearized equilibrium:

Definition. As $\bar{N} \to \infty$, the log-linearized competitive equilibrium of the economy is defined as a sequence of variables $\{d\log y_t, d\log g_{t+1}, dz_t, di_t, dL_t, d\log \Pi_t, dm_{t+1}, d\log \mathcal{M}_{t+1}\}$

that satisfy the following 8 equations, for a given sequence of exogenous shocks $\{d\log\epsilon_t^m\}$ and given initial values of state variables $\{dm_0, d\log\mathcal{M}_0\}$.

Consumption Euler equation

$$-\sigma d\log y_t = -\sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1} + d\log(1+i_t) - d\mathbf{E}_t \log\Pi_{t+1} \tag{16}$$

Entrant's innovation decision

$$[\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1}] + \mathbf{E}_t d\log y_{t+1} + d\log g_{t+1} = dm_t + \frac{\zeta_z - 1}{z} dz_t \tag{17}$$

Wage rigidity equation

$$d\log\Pi_t = -d\log g_t + dm_{t-1} - dm_t \tag{18}$$

Technical efficiency growth equation

$$d\log g_{t+1} = \frac{(\lambda-1)(1-I)}{(z+I(1-z))(\lambda-1)+1} dz_t \tag{19}$$

Measure of aggregate misallocation

$$dm_{t+1} = \alpha_m dz_t + (1-z) dm_t \tag{20}$$

$$d\log\mathcal{M}_{t+1} = (\alpha_m - \alpha_\lambda)dz_t + (1-z)\left[1 - I\left(1-\frac{1}{\lambda}\right)\right]d\log\mathcal{M}_t + (1-z)I\left(1-\frac{1}{\lambda}\right)dm_t, \text{ where} \tag{21}$$

$$\begin{aligned}
&\kappa_A \equiv 1-(1-I)\frac{\vartheta_I}{1+\vartheta_I},\ \kappa_B \equiv \frac{1}{1-z}, \vartheta_I \equiv \frac{z}{(1-z)I} \\
&\mathcal{S}_I^\star \equiv \sum_{\Delta=2}^{\infty}\Delta\upsilon(\Delta) = \frac{1+\vartheta_I}{\vartheta_I} - \frac{\vartheta_I}{1+\vartheta_I}, \\
&\mathcal{S}_\lambda^\star \equiv \sum_{\Delta=2}^{\infty}\lambda^{-\Delta}\upsilon(\Delta) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1} - \lambda^{-1}\left(\frac{\vartheta_I}{1+\vartheta_I}\right), \\
&\alpha_m \equiv -(\ln\lambda)[\kappa_A - \kappa_B\mathcal{S}_I^\star], \\
&\alpha_\Lambda \equiv \frac{1}{\Lambda}[\lambda^{-1}\kappa_A - \kappa_B\mathcal{S}_\lambda^\star], \Lambda \equiv \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}.
\end{aligned}$$

Resource constraint

$$d\log y_t = d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log\mathcal{M}_t \tag{22}$$

Monetary policy rule

$$d\log(1+i_t) = \phi d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log\epsilon_t^m, \phi > 0, i_t > 0 \tag{23}$$

The consumption euler equation is the log-linearized version of household's inter-temporal utility maximization condition; the entrant's innovation decision is the log-linearized transformation of outside entrepreneur's maximization condition; the wage rigidity equation is the log-linearized version of price inflation derived from combining equations (7) and (11);

technical efficiency growth equation summarizes how does percentage changes in rate of entrant's innovation translate into growth rate of technical efficiency. Resource constraint and monetary policy rule are obtained from log-linearizing their counterparts in the normalized competitive equilibrium, respectively.

While the consumption Euler equation, wage rigidity equation, resource constraint, and monetary policy rule take familiar forms as in the business cycle literature, and the technical efficiency growth equation as well as the entrant's innovation decision are familiar to scholars of business cycle model augmented with endogenous growth, the new ingredient is the misallocation block with the aggregate misallocation measure (eqns 20 - 21), which is the log-linearized transformation of the misallocation measure around the steady state. In particular, Equation 21 shows how a percentage change in rate of entrant's innovation could translate into changes in measure of aggregate misallocation through the evolution of the distribution of varieties with different quality gaps. Notably, relative to the log-linearized equilibrium defined in Appendix A.4 for a finite $\bar{N}$, the limiting case when $\bar{N} \to \infty$ does not need to keep track of the evolution of the quality gaps, and the misallocation block can be written recursively.

2.6 Calibration

In order to illustrate the dynamics for the baseline model, we calibrate the model with parameters provided in Table 1. A model period corresponds to a year. We target the real rate of interest to be 1.5% ($r_{\text{target}} = 1.015$) as in Benigno and Fornaro (2018). The full employment steady state growth rate to be 2% ($g^* = 1.02$). Step size of successful innovation is set at $\lambda = 1.069$ which is equal to average improvements from innovation quoted by Akcigit and Kerr (2018). We calibrate the own-variety incumbent innovation rate, I, and entrant innovation rate, z, using the fact that the fraction of internal patents is equal to 21.5% (Akcigit and Kerr, 2018). That is, $\frac{I}{z+I} = 0.215$. This statistic, along with steady state growth rate and innovation step size, implies an incumbent innovation rate of approximately 6.57%, and a creative destruction rate of 24%.[10]

Table 1: Parameters

	Value	Source/Target
Elas. intertemporal substitution	$1/\sigma = 0.5$	Standard value
Discount factor	$\beta = 0.96$	Standard value
Wage inflation at steady state	$\pi^w = 1.02$	
Innovation step size	$\lambda = 1.0698$	Akcigit and Kerr (2018).
Own variety improvement	$I = 0.0657$	Fraction of internal patents
Parameter for entrant's R&D cost	$\delta_z = 0.0701$	First order condition for entrant innovation
Curvature of entrant's R&D cost	$\zeta_z = 1.5$	
Persistence of monetary shock	$\rho^m = 0.9$	Standard value

Notes: Model period corresponds to a year.

2.7 Allocative Efficiency

Aggregate TFP, in the normalized economy, depends on the *dispersion* of markups. Given our assumption of discrete quality gaps that a firm can be on, $\mathcal{M}_t$ can be expressed as function of quality gaps, Δ, and measure of firms on those quality gaps, $\{\upsilon_t(\Delta)\}_{\Delta=1}^{\infty}$, as follows:

$$\mathcal{M}_t = \frac{\exp\left(\int_0^1 \ln \mu_{it}^{-1} di\right)}{\int_0^1 \mu_{it}^{-1} di} = \frac{\exp\left(-(\ln\lambda)\sum_{\Delta=1}^{\infty} \Delta \upsilon_t(\Delta)\right)}{\sum_{\Delta=1}^{\infty} \lambda^{-\Delta}\upsilon_t(\Delta)}. \tag{24}$$

Changes in normalized output can be decomposed into changes in labor employed in the

[10] Note that $\lambda = \frac{g-1}{z+(1-I)z} + 1$.

intermediate goods' production, and changes in measure of aggregate misallocation. That is, $d \log y_t = d \log L_{pt} + d \log \mathcal{M}_t$. Changes in misallocation measure can be further decomposed into a weighted changes in quality gap distribution. We summarize the result in Proposition 2 that changes in allocative efficiency are denoted with changes in $\mathcal{M}_t$.

Proposition 2. *Changes in allocative efficiency are given by changes in $\mathcal{M}_t$, which are summarized by a weighted average of changes in measure of varieties at each step*

$$d \log \mathcal{M}_{t+1} = \alpha_1 d\upsilon_{t+1}(1) + \sum_{j=2}^{\infty} \alpha_j d\upsilon_{t+1}(j), \tag{25}$$

where weights $\alpha_j = -(\ln \lambda)(j) - \frac{\lambda^{-j}}{\Lambda} < 0 \ \forall j = 1, 2, ...,$ and are all negative.

Proof. It follows from equation 24, which in turn follows from aggregation equation 12. □

Proposition 2 states that change in aggregate misallocation is determined by a weighted average of changes in mass of firms at each quality gap. A reduction in entrant innovation rate, z_t, has two effects. First, it implies that on average less number of products will face a successful entrant in the following period. Second, of the firms that are at unit quality gap, they have a higher likelihood of moving up the quality ladder. On net, the effect of a reduction in z_t is to reduce the mass of firms at unit quality gap, whereas at all quality gaps larger than one, a reduction in z_t increases the mass of firms at those quality gaps.

The mass of varieties at unit quality gap is increasing in the rate of entrant's innovation z_t, whereas the masses at all other quality gaps greater than 1 are decreasing in z_t. Furthermore, given that $\upsilon_t(\Delta)$ is a pmf at each quality gap. Then, it follows that an increase in mass at unit quality gap is equal to the sum of the reduction in masses at all other quality gaps.[11] As such, following a reduction in z_t, next period's measure of misallocation would increase if the magnitudes of weighted change in mass at unit quality gap would be smaller than the magnitude of the weighted sum of masses at the other quality gaps. Proposition 3 provides the sufficient condition that guarantees that the misallocation is an increasing function of entrant innovation rate.

Proposition 3. *Consider an economy along the balanced growth path. At time 0, there is an unanticipated negative shock to entrant innovation rate, z_t, that is, $dz_t < 0$. Aggregate allocative efficiency at time 1 would fall, if condition 26 is satisfied, that is, $\frac{d \log \mathcal{M}_1}{dz_0} > 0$ if:*

$$\ln \lambda - \frac{(\lambda - 1)}{\lambda^2 \Lambda} > 0, \tag{26}$$

where $\Lambda = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}$ is the production labor's share of income along the BGP.

Proof. See Appendix B.2. □

Proposition 3 summarizes the key insights of this paper. It states that if the condition in Equation 26 is satisfied, the measure of aggregate misallocation goes up next period after an unanticipated reduction in entrant innovation rate. The condition ensures that the weights α_j at each quality gap would increase in magnitudes such that the changes in mass of firms at non-unitary quality gaps are weighted more. Condition 26 would hold if the BGP features a high production labor's share of income (and correspondingly, a low level of

[11] Follows from Corollary 2, $\upsilon_t(\Delta)$ is a pmf at each quality gap, then for all t it must be the case that:

$$-d\upsilon_{t+1}(1) = \sum_{j=2}^{\infty} d\upsilon_{t+1}(j).$$

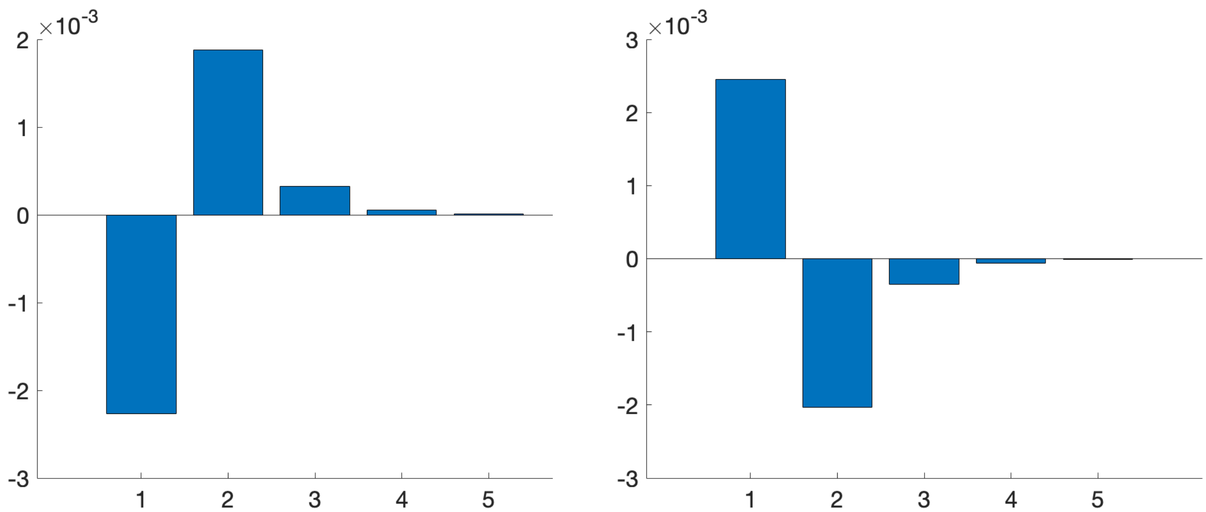
average markups). This implies that the rate of entrant innovation needs to be sufficiently high relative to the rate of incumbent innovation along the BGP, a condition we find to be empirically relevant (Akcigit and Kerr, 2018). It follows that the magnitude of the weighted reduction in mass of varieties with unit quality gap would be smaller than the magnitude of the weighted sum of all the other varieties with quality gaps larger than one such that misallocation measure at time 1 would go up.

Graphically, as shown in Figure 1a, following a one time adverse shock to rate of entrant's innovation at time 0, mass of firms at unit quality gap decreases, while masses of firms at all quality gaps greater than one increase at time 1. Figure 1b shows the graphical representation of change in misallocation measure the next period following the one time shock to entrant's innovation as weighted changes in mass of varieties. Recall that all weights are negative, then at unit quality gap, the positive weighted change in mass of varieties implies that $\mathcal{M}_1$ would increase (and that misallocation measure would decrease), whereas at all other quality gaps, the negative weighted changes in mass implies that $\mathcal{M}_1$ would decrease (and that misallocation measure would go up). On net, misallocation measure would increase if condition (26) holds, and that the sum of the areas is negative, i.e., aggregate allocative efficiency would fall. The greater is the absolute sum, the higher would be the fall in aggregate allocative efficiency following a percentage point reduction in z_0.

Figure 1: Changes in distribution from time 0 to time 1 after an unanticipated negative shock to entrant innovation rate

(a) Changes in distribution of varieties

(b) Change in allocative efficiency



Proposition 4 provides the necessary and sufficient condition for allocative efficiency to fall following a one-time percentage point reduction in rate of entrant's innovation.

Proposition 4. *Consider an economy along the balanced growth path. At time 0, there is an unanticipated negative shock to entrant innovation rate, z_t, that is, $dz_t < 0$. Aggregate allocative efficiency at time 1 would fall, if and only if condition 27 is satisfied, that is,* $\frac{d \log \mathcal{M}_1}{dz_0} > 0$ *if and only if:*

$$\left(\ln \lambda + \frac{\lambda^{-1}}{\Lambda}\right) \kappa_1 < (\ln \lambda) \kappa_B \mathcal{S}_I^\star + \frac{\kappa_B}{\Lambda} \mathcal{S}_\lambda^\star \tag{27}$$

where κ_1, κ_B, Λ, $\mathcal{S}_I^\star$, and $\mathcal{S}_\lambda^\star$ are all positive and defined in Equation 96.

Proof. See Appendix B.2. □

The necessary and sufficient condition is derived by rewriting the changes in masses of firms at each quality gap $dv_{t+1}(j)$ as the product of coefficient κ_j and the unanticipated shock to entrant innovation rate dz_t. Then as long as the weighted changes in masses of firms at each quality gap satisfies $|\alpha_1\kappa_1| < |\sum_{j=2}^{\infty}\alpha_j\kappa_j|$, then this would imply that the changes in mass of firms at non-unitary quality gaps are weighted more than those of firms at unit quality gap. Hence, allocative efficiency at time 1 would fall following a reduction in entrant innovation rate at time 0.

2.8 Misallocation effects of Monetary Policy

We study a special case of the model to analytically present the misallocation effects of monetary policy. Consider an economy along the balanced growth path. We assume that central bank can perfectly stabilize production labor L_p at the long-run level $\bar{L}_p$ period 1 onwards. Furthermore, nominal wages evolve according to following wage rigidity equation, as in Benigno and Fornaro (2018):

$$W_t = \pi^w W_{t-1},$$

where $\pi^w > 0$ is a constant.

There is a one-time unanticipated increase in nominal interest rate i_0, after which the nominal interest rate returns back to the level consistent with satisfying full employment in the production sector. We can then show that entrant innovation falls on impact, and consequently there are misallocation effects in the following period.

Proposition 5. *Consider an economy along the balanced growth path. At time 0, there is* *an unanticipated increase in nominal interest rate. $di_0 > 0$. Suppose further that the central bank restores full employment in production sector from $t \geq 1$, such that $d\log(L_t - \delta_z z_t^{\zeta_z}) = 0 \;\forall t \geq 1$. Entrant innovation rate falls if (1) $\sigma > 1$, (2) condition in proposition 4 is satisfied. That is $\frac{dz_0}{di_0} < 0$.*

Proof. See Appendix B.3. □

Proposition 5 constructs the misallocation effects by enabling the contractionary monetary shock to lower entrant innovation with $\sigma > 1$ assumption, i.e., low inter-temporal elasticity of substitution. The second requirement that proposition 4 be satisfied guarantees the connection to increased misallocation from lower entry rate.

However, note that the quantitative import of this misallocation channel may be limited as discussed in the following corollary.

Corollary 3. *It follows that there exits two counteracting forces - a level effect and a dispersion effect for the distribution of markups - such that the misallocation effect of monetary policy is muted. That is:*

$$d\log \mathcal{M}_1 = (\alpha_m - \alpha_\lambda) dz_0,$$

where $dz_0 = -\left(\frac{1}{\alpha_\lambda + \frac{\sigma(\zeta_z - 1)}{z}}\right) d\log(1 + i_0)$. $\alpha_m > 0$ measures the dispersion effect, and $\alpha_\lambda > 0$ measures the level effect. They are defined respectively in equation 109.

Proof. See Appendix B.3. □

Misallocation effect of a monetary policy shock depends on the conditional movements of real wage and production labor's share of income as shown in Equation 28.

$$\mathcal{M}_{t+1} = \frac{\exp(m_{t+1})}{\Lambda_{t+1}} = \frac{\exp(-(\ln\lambda)\sum_{\Delta=1}^{\infty}\Delta\upsilon_{t+1}(\Delta))}{\sum_{\Delta=1}^{\infty}\lambda^{-\Delta}\upsilon_{t+1}(\Delta)} = \frac{w_{t+1}}{\Lambda_{t+1}}, \tag{28}$$

where w_{t+1} is the real wage, and $\Lambda_{t+1} = \frac{\left(L_t - \delta_z z_{t+1}^{\zeta_z}\right)w_{t+1}}{y_{t+1}}$ is the production labor's share of income. Notice that market power not only leads to allocative inefficiency through an increase in markup dispersion, it also distorts factor prices, in this case wages, relative to their social marginal product through a reduction in labor demand. Unlike $\mathcal{M}_t$, Λ_t only depends on the level of markups, and can be interpreted as a labor wedge.

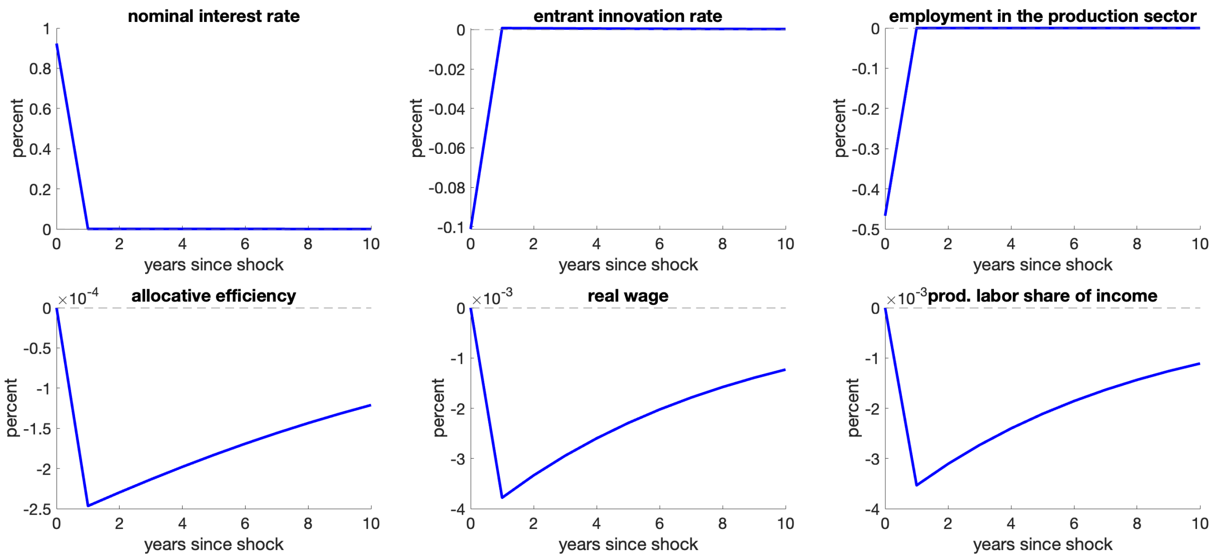
Our baseline model generates conditionally pro-cyclical movements in both the real wage and the production labor's share of income, and the counteracting forces attenuate the misallocation effects. Specifically, following a contractionary monetary policy shock, entrant's innovation rate z_t declines, it follows that the measure of firms at unit quality gap $\upsilon_{t+1}(1)$ falls, whereas the measures of firms at all other quality gaps rise. The distribution of markups expands, giving rise to two effects simultaneously. A *dispersion* effect occurs as markup dispersion rises, leading to more misallocation ($\mathcal{M}_t$ decreases). A *level* effect occurs as the production labor's share of income decreases, since the average level of markups inevitably rises and the labor wedge widens (Λ_t decreases, pushing $\mathcal{M}_t$ up).

In other words, a conditionally counter-cyclical movement in the average markup (pro-cyclical movement in the production labor's share of income) acts as an opposing force to the misallocation effect of monetary policy that arises from the increase in markup dispersion. As a result of the counteracting forces, the quantitative importance of the mechanism we highlight is limited. That said, our results suggest that the strength of the misallocation effects of monetary policy hinges on the conditional movements of both the first moment (level) and the second moment (dispersion) of markup distribution.[12]

Graphically, Figure 2 shows the modelled impulse responses of the nominal interest rate, the entrant innovation rate, the production labor, the measure of aggregate misallocation, the real wage, and the production labor's share of income following a one-time contractionary monetary policy shock in period 0. The central bank fully restores the employment level of the production labor from period 1 onwards. Entrant's innovation rate decreases following the monetary contraction, it follows that the allocative efficiency falls as the markup distribution becomes more dispersed. The response is persistent yet muted. Since as the markup distribution becomes more dispersed, the average level of markups also rises, and the production labor's share of income falls. The level effect of the conditionally counter-cyclical markups strongly attenuates the misallocation effect arising from the increase in markup dispersion.

[12]There is an active debate on whether or not price-cost markups are indeed counter-cyclical conditional on a monetary shock. In a textbook new Keynesian model with sticky prices (Galí, 2015), level of markups increases in response to a contractionary monetary policy shock, i.e., markups are counter-cyclical conditional on a monetary policy shock. Nekarda and Ramey (2020) provide empirical evidence that suggests a pro-cyclical movement of markups. Likewise, Cantore, Ferroni and León-Ledesma (2021) find that the labor share of income increases in response to a monetary tightening, thus a counter-cyclical labor share of income.

Figure 2: Model-based impulse response to a one-time monetary policy shock



Notes: The figure shows the impulse responses of the nominal interest rate, entrant innovation rate, production labor, the measure of misallocation, real wage, and the production labor's share of income to a one-time contractionary monetary policy shock at time 0. The central bank fully stabilizes the production labor from time 1 onwards.

We noted earlier that the sufficient condition for the misallocation effects of monetary policy is that the economy has a sufficiently high entrant's innovation rate relative to incumbent's innovation rate. This implies a steady state with a high density of firms at the unitary quality gap and lower steady state level of average markup. In Proposition 6, we summarize the state-dependency of the misallocation effect. We find that the misallocation effect appears stronger when there is a low density of firms at the unitary quality gap in the steady state. The intuition is: when there are very few firms at the unitary quality gap, a contractionary monetary policy shock deters firm entry, which in turn would cause the markup distribution to shift further to the right. The increase in dispersion, and hence the misallocation effect, appears stronger when the initial distribution has a fatter right tail. At the same time, the misallocation effect appears stronger when the economy starts from a low level of average markups.

Proposition 6. *Consider an economy along the balanced growth path. At time 0, there is an unanticipated increase in nominal interest rate $di_0 > 0$. Suppose further that the central bank restores full employment in production sector from $t \geq 1$. The elasticity of the misallocation measure with respect to the density of firms with unitary quality gap is negative, whereas the elasticity with respect to the production labor's share of income is positive. That is, $\frac{\partial(\alpha_m - \alpha_\lambda)}{\partial \upsilon(1)} < 0$, and $\frac{\partial(\alpha_m - \alpha_\lambda)}{\partial \Lambda} > 0$.*

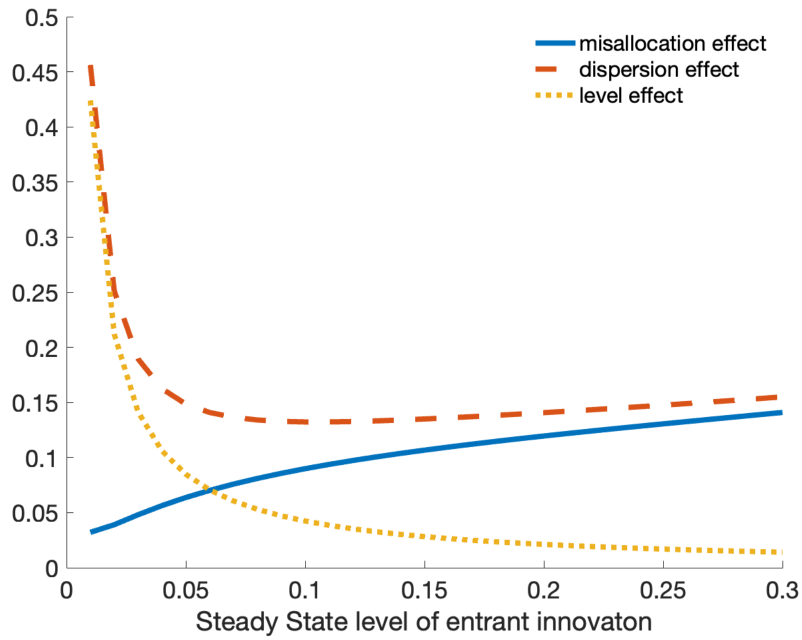
Proof. See Appendix B.4. □

In steady state, it can be easily shown that both Λ and $\upsilon(1)$ are increasing in entrant's innovation rate z. Yet the increases in both Λ (hence a reduction in average markup) and $\upsilon(1)$ following an increase in steady sate entrant's innovation exert counteracting forces on the elasticity of the misallocation measure. As such, whether the misallocation effect is stronger with a higher rate of steady state entrant's innovation is ambiguous. Note that a caveat to these exercises is that the initial conditions are functions of the deep parameter of the model.

To understand whether a higher entrant innovation rate in the steady state is associated with a stronger misallocation effect requires a breakdown of the dispersion and the level effect at different rates of steady state entrant's innovation. At a very low level of entrant innovation rate with few firms sitting on the unitary quality gap, the dispersion effect, as noted earlier, is strong in response to a fall in firm entry. However, the overall misallocation effect also depends on the change in the level of markups. It turns out that if the level effect is also strong as the distribution becomes increasingly dispersed, average markup would increase fast. These two effects counteract each other and give rise to a muted misallocation response. In contrast, at a high level of steady state entrant's innovation, there are a high density firms at the unitary quality gap, a fall in firm entry would not shift the distribution much to the right, resulting in a weakened dispersion effect. If the level effect is in turn weakened in response to the slight right-shift in the distribution, then the misallocation effect would strengthen if the level effect is weakened to a greater extent.

To further investigate which effect dominates, in Figure 3 we graphically show how the misallocation effect varies with respect to the steady state level of entrant's innovation, conditioning on incumbent innovation rate. In particular, we show that while the level effect α_λ weakens monotonically as entrant's innovation rises, the dispersion effect α_m is stronger at both very low level of entrant's innovation and very high level; even though the dispersion effect is strongest at the lower end. That said, the dispersion effect nevertheless dominates in magnitudes and the misallocation effect intensifies as we increase the steady state level of entrant innovation. In a steady state with higher rate of creative destruction relative to incumbent innovation, the misallocation effect of monetary policy is greater,

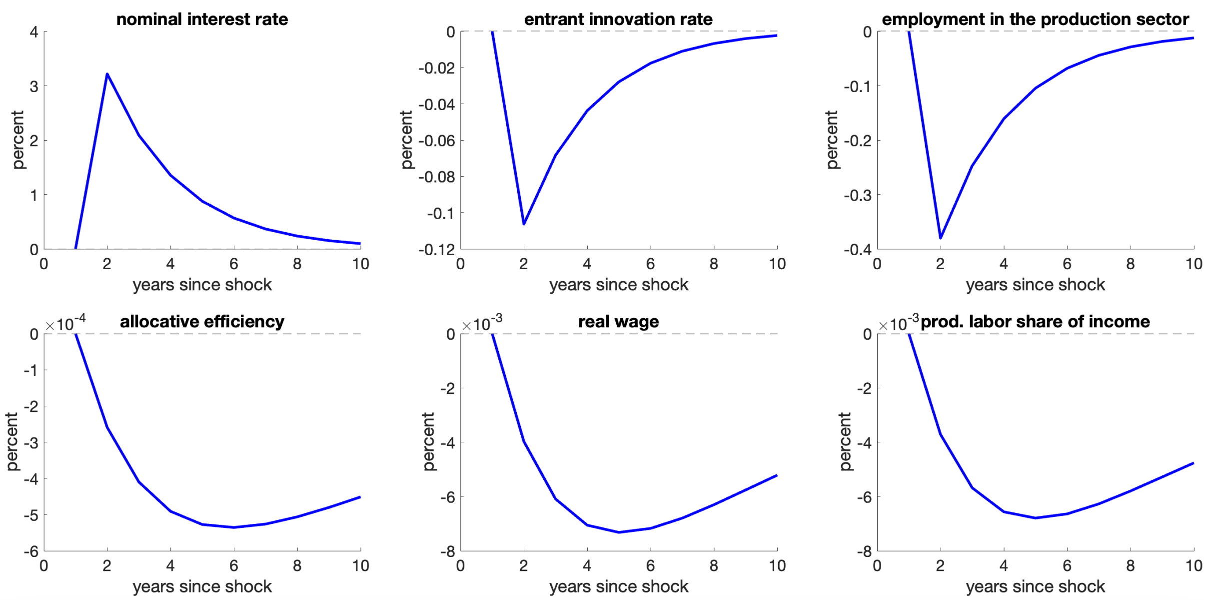
Figure 3: State-dependent misallocation effect of monetary policy



Notes: The figure shows the state-dependent misallocation effect of monetary policy with respect to the steady state rates of entrant's innovation. It breaks down the misallocation effect into a dispersion effect and a level effect. Calibration is discussed in Table 1.

To summarize, at the core of the mechanism is that a contractionary monetary policy shock reduces allocative efficiency across firms through reducing entrant's innovation rate relative to that of the incumbent. In particular, this misallocation effect induced by the asymmetrical responses is stronger if we start with a steady state with high rate of entrant innovation relative to incumbent innovation. A reduction in entrant innovation would push up markup dispersion if there is incumbent innovation at the same time. In our calibration, even though the reduction in entrant innovation also pushes up the average level of markups, exerting a counteracting force to the dispersion effect, the dispersion effect dominates, and the overall misallocation effect is stronger when the economy starts from a high rate of entrant innovation.

Figure 4: Model-based impulse response to a one-time monetary policy shock



Notes: The figure shows the impulse responses of the nominal interest rate, entrant innovation rate, production labor, the measure of misallocation, real wage, and the production labor's share of income to a one-time contractionary monetary policy shock.

2.8.1 Modelled Dynamic Responses

Proposition 5 is a special case that can be analytically solved, to further illustrate the dynamics of the baseline model, we calibrate the model with parameters provided in Table 1. Following a one-time contractionary monetary policy shock, Figure 4 graphically illustrates the dynamic responses of nominal interest rate, entrant innovation rate, employment level in the production sector, and allocative efficiency. Following a contractionary monetary policy shock, real interest rate increases, depressing nominal wages as a result. Because of the sticky nominal wages, aggregate demand would fall. The increase in real interest rate combined with a reduced expectation of future aggregate demand leads to a reduction in rate of entrant's innovation, which translates into a lowered growth rate as incumbent's rate of innovation is kept unchanged. Facing a reduced probability of being displaced, incumbents could further move up in quality ladder and charge a higher markup with probability I, resulting in reduced allocative efficiency.

3 Extensions and Empirical Discussion

The baseline model provides us with analytical tractability at the expense of specific assumptions, such that outside entrepreneurs are only allowed to survive for one period upon successful entry; incumbent's innovation probability is exogenously given and is thus invariant over the business cycle. Finally, the unitary demand elasticity structure assumed in the Cobb-Douglas aggregation in the final good sector is a special case of the more general CES demand structure. In this section, we relax each assumption and study how does each of them affects the model.

3.1 Incorporating value functions

In the baseline model, we assume that successful entrants could only earn profits for one period, we relax the assumption and consider instead that outside entrepreneurs' maximization problems takes the following form in a setting in which successful entrant can survive to the next period as an incumbent firm:

$$\max_{z_t}\{D_{t,t+1}z_t V_{t+1}(1) - \delta_z z_t^{\zeta_z} W_t\},$$

where $V_{t+1}(1)$ is incumbent's value function with unitary quality gap, and

$$V_t(1) = (1-\lambda^{-1})P_t Y_t + \mathbf{E}_t\{D_{t,t+1}(1-z_t)[IV_{t+1}(2) + (1-I)V_{t+1}(1)]\}.$$

The value function of outside entrepreneur now consists of two parts: (i) the profits generated from unitary quality gap $\Delta = 1$, (ii) the expected profits for the next period, conditioning on surviving creative destruction this period and increasing the existing quality gap by one step, this happens at the exogeneous probability of own innovation I.

It turns out that incorporating value functions into entrant's optimization problem does not necessitate keeping track of all value functions and difference equations for each quality gap Δ. In fact, we show in Appendix C.1 that to define the full log-linearized competitive equilibrium we only need the value function at the unitary quality gap $V_t(1)$ and the difference in value functions at $\Delta = 2$ and the unitary quality gap, $dV_t(1) \equiv V_t(2) - V_t(1)$. Figure 5 graphically illustrates the dynamic responses of the key variables in the model to a contractionary monetary policy shock. As expected, the dynamic response of entrant's innovation rate is slightly muted relative to the baseline model, as now outside entrepreneurs are able to retain their patent and continue to innovate upon successful entry. Otherwise, incorporating value functions into entrant's decision does not materially change the properties of the baseline model. Allocative efficiency still falls as a result of the decline in entrant's innovation when the incumbents continue to innovate.

3.2 Endogenizing incumbent's innovation

We assumed that rate of incumbent innovation I was exogenous and thus unaffected by monetary policy shocks in the baseline model. Consider instead that incumbents are allowed to optimally choose their rates of own innovation $I_t(\Delta)$ by hiring R&D labor to continue increasing their relative quality. Incumbents' maximization problems can be written as:

$$\max_{I_t(\Delta)} \left\{(1-\lambda^{-\Delta})P_tY_t + \mathbf{E}_t D_{t,t+1}(1-z_t)[I_tV_{t+1}(\Delta+1) + (1-I_t)V_{t+1}(\Delta)] - c(I_t,\Delta)W_t\right\},$$

where $c(I_t, \Delta)$ denotes costs of own innovation in units of R&D labor hired. To make analytical progress in solving the more general case, we follow Peters (2020) to assume a particular functional form for the cost function such that

$$c(I_t, \Delta) = \lambda^{-\Delta}\delta_I I_t(\Delta)^{\zeta_I}, \zeta_I > 1,$$

where δ_I determines the efficiency of own innovation, and $\zeta_I > 1$ ensures convexity of the cost function such that there exists a unique solution. Solving incumbent's maximization problems give the optimal rate of own innovation:

$$I_t^*(\Delta) = \left[\frac{\beta \mathbf{E}_t \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}[\tilde{V}_{t+1}(\Delta+1) - \tilde{V}_{t+1}(\Delta)]}{\zeta_I \lambda^{-\Delta}\delta_I w_t}\right]^{\frac{1}{\zeta_I - 1}}, \tag{29}$$

where $\tilde{V}_{t+1}(\Delta)$ is the normalized value function for each Δ. As $\bar{N} \to \infty$, we show in Appendix C.2 that the optimal rate of own innovation is independent of quality gap such that the aggregate own innovation is $I_t^* = I_t^*(\Delta)\forall\Delta$. However, for a finite $\bar{N}$, the optimal own innovation depends on quality gap and the aggregate own innovation I_t^* is the expected value of $I_t^*(\Delta)$ across the distribution of quality gaps. Endogenizing incumbent's innovation alters difference equations 4 - 6 by introducing a time subscript to the rate of own innovation I, and now the optimal own innovation rate depends on existing quality gaps. The evolution of the distribution of quality gaps is now determined by the counteracting forces of entrant's innovation and incumbents' continuing innovation. These two counteracting forces - entrant innovation compressing the distribution, whereas incumbents' own innovation expanding it - shape the evolution of the cross-sectional distribution of markups, such that it endogenously responds to business cycle shocks.

In Figure 6 of Appendix C.2, we graphically illustrate the dynamics of the model subject to a contractionary monetary policy shock. Both entrant and incumbent innovation rates fall as the monetary contraction leads to a reduction in the expected future aggregate demand; notably the former falls by a greater extent. This muted responsiveness of incumbent innovation is at the heart of the mechanism we propose: a contractionary monetary policy shock reduces allocative efficiency across firms by reducing entrant innovation rate more relative to the reduction in incumbent innovation it causes.

3.3 CES demand function

We had assumed that household's preferences take a particular form of unitary elasticity in the baseline model, now we relax the assumption to study the case with a general CES demand structure:

$$Y_t = \left(\int_0^1 x_{it}^{\frac{\eta-1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}}, \eta > 1. \tag{30}$$

With CES demand, firms with quality gaps that are sufficiently large will find it optimal to charge the unconstrained markup $\frac{\eta}{\eta-1}$ instead of the limit price. The distribution of markups is thus truncated at a finite quality gap. It follows that

$$\mu_{it}(\Delta) = \min\left\{\frac{\eta}{\eta-1}, \lambda^{\Delta}\right\}. \tag{31}$$

Aggregate output can thus be rewritten as

$$Y_t = Q_t L_{pt} \mathcal{M}_t, \text{ where } \mathcal{M}_t \equiv \frac{\left[\int_0^1 \mu_{it}^{1-\eta}\left(\frac{q_{it}}{Q_t}\right)^{\eta-1} di\right]^{\frac{\eta}{\eta-1}}}{\int_0^1 \mu_{it}^{-\eta}\left(\frac{q_{it}}{Q_t}\right)^{\eta-1} di}, \tag{32}$$

where $Q_t \equiv (\int q_{it}^{\eta-1} di)^{\frac{1}{\eta-1}}$ is a measure of technical efficiency, while $\mathcal{M}_t$ is a measure of allocative efficiency in the economy.

The entrant's optimization problem implies:

$$\beta \mathbf{E}_t\left[\frac{Y_{t+1}^{-\sigma}}{Y_t^{-\sigma}} \frac{P_t}{P_{t+1}} \Phi_{i,t+1}(1, q_{i,t+1})\right] = \zeta_z \delta_z z_t^{\zeta_z - 1} W_t, \tag{33}$$

where $\Phi_{i,t+1}(1, q_{i,t+1}) = \left(1 - \frac{1}{\lambda}\right) \lambda^{1-\eta} \mathbf{E}_i \left[\frac{q_{i,t+1}}{\tilde{W}_{t+1}}\right]^{\eta-1} P_{t+1} Y_{t+1}$ is the per periods profit function for a successful entrant that produces a variety with unitary markup and quality $q_{i,t+1}$, and real wages is given by $\tilde{W}_t \equiv \frac{W_t}{P_t} = \left[\int_0^1 \mu_{it}^{1-\eta} q_{it}^{\eta-1} di\right]^{\frac{1}{\eta-1}}$.

Similar to the baseline case, we assume that incumbent's own innovation occurs at an exogenous probability I, this simplifies the analysis as now the marginal distributions of quality gaps Δ and quality q are *independent* such that: $F_t(\Delta, q) = F_{\Delta,t}(\Delta) F_{q,t}(q)$. Since markups only depend on Δ, it follows that the distribution of markups is independent of the distribution of q, and that the evolution of quality gaps are governed by the same difference equations 4 - 6. We show in Appendix C.3 that the optimal rate of entry z_t is independent of q, and in the special case $\eta = 1$, the model collapses to the baseline case. Since z_t alone determines the evolution of quality gaps, it follows that the measure of aggregate misallocation is pinned down by the distribution of quality gaps:

$$\mathcal{M}_t \equiv \frac{\left[\int \mu_{it}^{1-\eta} di\right]^{\frac{\eta}{\eta-1}}}{\int \mu_{it}^{-\eta} di} = \frac{\left[\sum_{\Delta=1}^{\bar{N}} \mu(\Delta)^{1-\eta} \upsilon_t(\Delta)\right]^{\frac{\eta}{\eta-1}}}{\sum_{\Delta=1}^{\bar{N}} \mu(\Delta)^{-\eta} \upsilon_t(\Delta)}. \tag{34}$$

The main insight from the baseline case is preserved in the more general setup, as the distribution of quality gaps alone determines allocative efficiency.

In Figure 7 of Appendix C.3, we show the dynamic responses of key variables in the extended model following a contractionary monetary policy shock. Allocative efficiency falls as a result of the decline in entrant's innovate rate as in the baseline model, and the impulse responses are somewhat amplified with the introduction of the CES demand.

3.4 Empirical Discussion

Finally, we briefly discuss the empirical evidence from the literature that is consistent with the mechanism we presented, namely that monetary policy shocks can increase misallocation. Meier and Reinelt (2022) document that contractionary U.S. monetary policy shocks increase inter-sectoral markup dispersion across firms in the Compustat data. Garga and Singh (2021, Online Appendix) find that contractionary monetary policy shocks reduce firm entry, but do not significantly affect R&D expenditures of the U.S. listed firms, which we take as a proxy for incumbent innovation. Similarly, numerous studies document that aggregate productivity declines in response to contractionary monetary shocks (see, for example, Moran and Queralto, 2018).

Taken together, these previous empirical findings suggest that a tightening of monetary policy does not affect own innovation of the established incumbents as much as it affects the innovation efforts of the potential entrants. As a result, a monetary tightening lowers firm entry, and at the aggregate, lowers productivity. We view the empirical findings as providing suggestive evidence in support of our proposed mechanism, in particular that a contractionaty monetary policy shock reduces allocative efficiency across firms through a greater reduction of entrant innovation relative to incumbent innovation.

4 Conclusion

In this paper, we propose a novel misallocation channel of monetary policy that operates on the extensive margin of firm entry and incumbent's own innovation. We add to a standard business-cycle framework that incorporates endogenous growth a simple modification: in addition to allowing entrants to undertake R&D investment to displace incumbents via the traditional channel of creative destruction, we allow incumbent firms to engage in own innovation of their existing products to increase their markups as long as they remain active (Peters, 2020).

We show theoretically that a contractionary monetary policy shock leads to a reduction of outside entrepreneurs' R&D efforts such that the traditional force of creative destruction is weakened. Importantly, we find that as long as the incumbent R&D is affected to a lesser extent by a monetary tightening, such that the average incumbent innovation remains largely unchanged over the business cycles, this reduction in entrant's innovation would push up both the average level and the dispersion of the markup distribution. Sectors where incumbents continue innovate, see their product market power goes up. This implies that the propagation of an adverse business shock shock over time generates a more dispersed cross-sectional distribution of markups, resulting in an endogenous short-term misallocation of inputs across firms.

References

Acemoglu, Daron, and Dan Cao. 2015. "Innovation by entrants and incumbents." *Journal of Economic Theory*, 157: 255–294.

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The environment and directed technical change." *American Economic Review*, 102(1): 131–66.

Acharya, Sushant, Julien Bengui, Keshav Dogra, and Shu Lin Wee. 2022. "Slow Recoveries and Unemployment Traps: Monetary Policy in a Time of Hysteresis." *The Economic Journal*.

Afrouzi, Hassan, and Luigi Caloi. 2023. "Endogenous Firm Competition and the Cyclicality of Markups." *The Review of Economics and Statistics*, 1–45.

Afrouzi, Hassan, Andres Drenik, and Ryan Kim. 2023. "Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition." National Bureau of Economic Research.

Aghion, P, and P Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica*, 60(2).

Aghion, Philippe, and Peter W Howitt. 2008. *The Economics of Growth.* MIT Press.

Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers. 2001. "Competition, imitation and growth with step-by-step innovation." *The Review of Economic Studies*, 68(3): 467–492.

Aghion, Philippe, Emmanuel Farhi, and Enisse Kharroubi. 2015. "Liquidity and growth: the role of counter-cyclical interest rates."

Aghion, Philippe, Emmanuel Farhi, and Enisse Kharroubi. 2019. "Monetary policy, product market competition and growth." *Economica*, 86(343): 431–470.

Akcigit, Ufuk, and William R Kerr. 2018. "Growth through heterogeneous innovations." *Journal of Political Economy*, 126(4): 1374–1443.

Annicchiarico, Barbara, and Alessandra Pelloni. 2016. "Innovation, Growth and Optimal Monetary Policy." *CEIS Working Paper*.

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

Ates, Sina T., and Felipe E. Saffie. 2021. "Fewer but better: Sudden stops, firm entry, and financial selection." *American Economic Journal: Macroeconomics*, 13(3): 304–56.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-market, trade costs, and international relative prices." *American Economic Review*, 98(5): 1998–2031.

Baqaee, David Rezza, and Emmanuel Farhi. 2019. "Productivity and Misallocation in General Equilibrium*." *The Quarterly Journal of Economics*, 135(1): 105–163.

Baqaee, David Rezza, Emmanuel Farhi, and Kunal Sangani. 2023. "The Darwinian Returns to Scale." *Review of Economic Studies*, rdad061.

Baqaee, David Rezza, Emmanuel Farhi, and Kunal Sangani. 2024. "The Supply-Side Effects of Monetary Policy." *Journal of Political Economy.*

Basu, Susanto, and John G Fernald. 2002. "Aggregate productivity and aggregate technology." *European Economic Review*, 46(6): 963–991.

Benigno, Gianluca, and Luca Fornaro. 2018. "Stagnation Traps." *The Review of Economic Studies*, 85(3): 1425–1470.

Bergin, Paul R, and Giancarlo Corsetti. 2008. "The extensive margin and monetary policy." *Journal of Monetary Economics*, 55(7): 1222–1237.

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

Bilbiie, Florin O, Fabio Ghironi, and Marc J Melitz. 2008. "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety." In *NBER Macroeconomics Annual 2007, Volume 22*. 299–353. University of Chicago Press.

Bilbiie, Florin O, Fabio Ghironi, and Marc J Melitz. 2012. "Endogenous entry, product variety, and business cycles." *Journal of Political Economy*, 120(2): 304–345.

Burya, Anastasia, Rui Mano, Yannick Timmer, and Anke Weber. 2022. "Monetary Policy Under Labor Market Power." International Monetary Fund.

Cantore, Cristiano, Filippo Ferroni, and Miguel León-Ledesma. 2021. "The missing link: monetary policy and the labor share." *Journal of the European Economic Association*, 19(3): 1592–1620.

Cerra, Valerie, and Sweta Chaman Saxena. 2008. "Growth dynamics: the myth of economic recovery." *American Economic Review*, 98(1): 439–457.

Cerra, Valerie, Antonio Fatás, and Sweta Chaman Saxena. 2022. "Hysteresis and Business Cycles." *Journal of Economic Perspectives*, forthcoming.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico. 2022. "Short-term tax cuts, long-term stimulus." National Bureau of Economic Research.

De Blas, Beatriz, and Katheryn N Russ. 2015. "Understanding markups in the open economy." *American Economic Journal: Macroeconomics*, 7(2): 157–80.

Duval, Romain A, Davide Furceri, Raphael Lee, and Marina M Tavares. 2021. "Market power and monetary policy transmission." International Monetary Fund.

Duval, Romain A, Davide Furceri, Raphael Lee, and Marina M Tavares. 2023. "Market Power, Financial Constraints, and Monetary Transmission." *AEA Papers and Proceedings*, 113: 99–104.

Etro, Federico, and Andrea Colciago. 2010. "Endogenous market structures and the business cycle." *The Economic Journal*, 120(549): 1201–1233.

Faia, Ester. 2012. "Oligopolistic competition and optimal monetary policy." *Journal of Economic Dynamics and Control*, 36(11): 1760–1774.

Fatás, Antonio, and Ilian Mihov. 2013. "Recoveries." C.E.P.R. Discussion Papers CEPR Discussion Papers 9551.

Fatás, Antonio, and Sanjay R. Singh. 2024. "Supply or demand? Policy makersâ confusion in the presence of hysteresis." *European Economic Review*, 161: 104617.

Feenstra, Robert C. 2003. "A homothetic utility function for monopolistic competition models, without constant price elasticity." *Economics Letters*, 78(1): 79–86.

Ferrando, Annalisa, Peter McAdam, Filippos Petroulakis, and Xavier Vives. 2021. "Product market structure and monetary policy: evidence from the euro area."

Fornaro, Luca, and Martin Wolf. 2023. "The scars of supply shocks: Implications for monetary policy." *Journal of Monetary Economics.*

Furlanetto, Francesco, Antoine Lepetit, Ørjan Robstad, Juan Rubio Ramírez, and Pål Ulvedal. 2023. "Estimating hysteresis effects." *AEJ Macroeconomics, forthcoming.*

Galí, Jordi. 2015. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications.* Princeton University Press.

Garga, Vaishali, and Sanjay R Singh. 2021. "Output hysteresis and optimal monetary policy." *Journal of Monetary Economics*, 117: 871–886.

Gourio, Francois, and Leena Rudanko. 2014. "Customer capital." *Review of Economic Studies*, 81(3): 1102–1136.

Greaney, Brian, and Conor Walsh. 2023. "Demand, growth, and deleveraging." *Review of Economic Dynamics*, 51: 795–812.

Guerron-Quintana, Pablo A., and Ryo Jinnai. 2019. "Financial frictions, trends, and the great recession." *Quantitative Economics*, 10(2): 735–773.

Gu, Yumeng. 2022. "Monetary Policy, Intangibles, and the Cyclicality of Markups."

Jaimovich, Nir, and Max Floetotto. 2008. "Firm dynamics, markup variations, and the business cycle." *Journal of Monetary Economics*, 55(7): 1238–1252.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2013. "When credit bites back." *Journal of Money, Credit and Banking*, 45(2): 3–28.

Jordà, Òscar, Sanjay R Singh, and Alan M Taylor. 2021. "The long-run effects of monetary policy." National Bureau of Economic Research Working Paper 26666.

Liu, Ernest, Atif Mian, and Amir Sufi. 2022. "Low interest rates, market power, and productivity growth." *Econometrica*, 90(1): 193–221.

Meier, Matthias, and Timo Reinelt. 2022. "Monetary policy, markup dispersion, and aggregate tfp." *Review of Economics and Statistics*, 1–45.

Moran, Patrick, and Albert Queralto. 2018. "Innovation, productivity, and monetary policy." *Journal of Monetary Economics*, 93: 24–41.

Morlacco, Monica, and David Zeke. 2021. "Monetary policy, customer capital, and market power." *Journal of Monetary Economics*, 121: 116–134.

Nekarda, Christopher J, and Valerie A Ramey. 2020. "The cyclical behavior of the price-cost markup." *Journal of Money, Credit and Banking*, 52(S2): 319–353.

Peters, Michael. 2020. "Heterogeneous markups, growth, and endogenous misallocation." *Econometrica*, 88(5): 2037–2073.

Peters, Michael, and Conor Walsh. 2020. "Population Growth and Firm Dynamics." Columbia University Manuscript.

Queralto, Albert. 2020. "A model of slow recoveries from financial crises." *Journal of Monetary Economics*, 114: 1–25.

Queralto, Albert. 2022. "Monetary Policy in a Model of Growth." Federal Reserve Board of Governors International Finance Discussion Papers 1340.

Vinci, Francesca, and Omar Licandro. 2021. "Switching-track after the Great Recession." CESifo Working Paper 9107.

Woodford, Michael. 1998. "Doing without money: controlling inflation in a post-monetary world." *Review of Economic Dynamics*, 1(1): 173–219.

A Competitive equilibrium

A.1 Definition

Definition A.1 (Competitive Equilibrium). For a given finite $\bar{N}$, the competitive equilibrium is defined as a sequence of $(7+\bar{N})$ quantities $\{g_{t+1}, z_t, L_t, Y_t, \{\upsilon_{t+1}(i)\}_{i=1}^{\bar{N}}, m_{t+1}, \Lambda_{t+1}, \mathcal{M}_{t+1}\}_{t=0}^{\infty}$ and 4 prices $\{P_t, i_t, W_t, D_{t,t+1}\}$ satisfying the following $(11+\bar{N})$ equations, for a given sequence of exogenous shocks $\{\epsilon_t^m\}$, and initial conditions are given by $\{Q_0, P_0, \{\upsilon_0(i)\}_{i=1}^{\bar{N}}\}$, and $L_t \leq 1$ for all $t \geq 0$.

1. Consumption Euler Equation and Stochastic Discount Factor

$$1 = \beta \mathbf{E}_t \left[\frac{Y_{t+1}^{-\sigma}}{Y_t^{-\sigma}} \frac{P_t}{P_{t+1}} (1 + i_t) \right] \tag{35}$$

$$D_{t,t+1} = \beta \mathbf{E}_t \left[\frac{Y_{t+1}^{-\sigma}}{Y_t^{-\sigma}} \frac{P_t}{P_{t+1}} \right] \tag{36}$$

2. Outside Entrepreneur's Innovation Decision

$$D_{t,t+1}(1 - \lambda^{-1}) \mathbf{E}_t [P_{t+1} Y_{t+1}] = \zeta_z \delta_z z_t^{\zeta_z - 1} W_t \tag{37}$$

3. Wage Rigidity

$$W_t = \pi^w W_{t-1} \tag{38}$$

$$P_t = \frac{W_t}{Q_t \exp\{m_t\}} \tag{39}$$

4. Technical Efficiency Growth Rate

$$g_{t+1} = \frac{Q_{t+1}}{Q_t} = (z_t + I - z_t I)(\lambda - 1) + 1 \tag{40}$$

5. Measure of Aggregate Misallocation

$$m_{t+1} = -\mathbf{E}_i[\ln \mu_{i,t+1}] = -\int_0^1 \ln \mu_{i,t+1} di = -(\ln \lambda) \sum_{\Delta=1}^{\bar{N}} \Delta \upsilon_{t+1}(\Delta) \tag{41}$$

$$\Lambda_{t+1} = \mathbf{E}_i[\mu_{i,t+1}^{-1}] = \int_0^1 \mu_{i,t+1}^{-1} di = \sum_{\Delta=1}^{\bar{N}} \lambda^{-\Delta} \upsilon_{t+1}(\Delta) \tag{42}$$

$$\mathcal{M}_{t+1} = \frac{\exp(m_{t+1})}{\Lambda_{t+1}} = \frac{\exp(-(\ln \lambda) \sum_{\Delta=1}^{\bar{N}} \Delta \upsilon_{t+1}(\Delta))}{\sum_{\Delta=1}^{\bar{N}} \lambda^{-\Delta} \upsilon_{t+1}(\Delta)} \tag{43}$$

6. Market Clearing Conditions and Aggregate Production Function

$$Y_t = Q_t L_{pt} \mathcal{M}_t = Q_t (L_t - \delta_z z_t^{\zeta_z}) \mathcal{M}_t \tag{44}$$

7. Monetary Policy Rule

$$1 + i_t = (1 + i_{ss}) \left(\frac{L_t - \delta_z z_t^{\zeta_z}}{L_{ss} - \delta_z z_{ss}^{\zeta_z}} \right)^{\phi} \epsilon_t^m, \phi > 0 \tag{45}$$

8. Evolution of Quality Gaps

$$\upsilon_{t+1}(\Delta) = z_t + (1 - z_t)(1 - I)\upsilon_t(\Delta) \quad \text{if } \Delta = 1 \tag{46}$$

$$\upsilon_{t+1}(\Delta) = (1 - z_t)(1 - I)\upsilon_t(\Delta) + (1 - z_t) I \upsilon_t(\Delta - 1) \quad \text{if } 2 \leq \Delta \leq \bar{N} - 1 \tag{47}$$

$$\upsilon_{t+1}(\Delta) = (1 - z_t)\upsilon_t(\Delta) + (1 - z_t) I \upsilon_t(\Delta - 1) \quad \text{if } \Delta = \bar{N} \tag{48}$$

where $\{\upsilon_{t+1}(\Delta)\}_{\Delta=1}^{\bar{N}}$ denote the measure of varieties with quality gap Δ at time t. As $\bar{N} \to \infty$, equation 48 will be dropped.

Stationarization of the System. The competitive equilibrium defined above is non-stationary: output and nominal wage are co-integrated with technical efficiency Q_t. We normalize them as follows:

$$y_t \equiv \frac{Y_t}{Q_t}; w_t \equiv \frac{W_t}{P_t Q_t}.$$

Definition A.2 (Normalized Competitive Equilibrium). The normalized competitive equilibrium is defined as a set of $(7+\bar{N})$ stationary quantities $\{g_{t+1}, z_t, L_t, y_t, \{\upsilon_{t+1}(i)\}_{i=1}^{\bar{N}}, m_{t+1}, \Lambda_{t+1}, \mathcal{M}_{t+1}\}_{t=0}^{\infty}$ and 3 stationary prices $\{\Pi_t, i_t, w_t\}$ satisfying the following $(10+\bar{N})$ equations, for a given sequence of exogenous shocks $\{\epsilon_t^m\}$, and initial conditions are given by $\{Q_0, P_0, \{\upsilon_0(i)\}_{i=1}^{\bar{N}}\}$, and $L_t \leq 1$ for all $t \geq 0$.

1. Consumption Euler Equation

$$y_t^{-\sigma} = \beta \mathbf{E}_t \left[y_{t+1}^{-\sigma} g_{t+1}^{-\sigma} \frac{(1+i_t)}{\Pi_{t+1}} \right] \tag{49}$$

2. Outside Entrepreneur's Innovation Decision

$$\beta \mathbf{E}_t \left[\frac{c_{t+1}^{1-\sigma} g_{t+1}^{1-\sigma}}{c_t^{-\sigma}} \right] (1-\lambda^{-1}) = \zeta_z \delta_z z_t^{\zeta_z - 1} w_t \tag{50}$$

3. Wage Rigidity

$$w_t = \exp\left\{ \int_0^1 \ln \mu_{it}^{-1} di \right\} = exp(m_t) \tag{51}$$

$$\Pi_t = \frac{P_t}{P_{t-1}} = \pi^w g_t^{-1} \frac{w_{t-1}}{w_t} \tag{52}$$

4. Technical Efficiency Growth Rate

$$g_{t+1} = \frac{Q_{t+1}}{Q_t} = (z_t + I - z_t I)(\lambda - 1) + 1 \tag{53}$$

5. Measure of Aggregate Misallocation

The three equations that govern $\{m_{t+1}, \Lambda_{t+1}, \mathcal{M}_{t+1}\}$ are give by Equations 41 - 43.

6. Market Clearing Conditions and Aggregate Production Function

$$y_t = L_{pt}\mathcal{M}_t = (L_t - \delta_z z_t^{\zeta_z})\mathcal{M}_t \tag{54}$$

7. Monetary Policy Rule

$$1 + i_t = (1 + i_{ss})\left(\frac{L_t - \delta_z z_t^{\zeta_z}}{L_{ss} - \delta_z z_{ss}^{\zeta_z}}\right)^{\phi} \epsilon_t^m, \phi > 0 \tag{55}$$

8. Evolution of Quality Gaps

The remaining $\bar{N}$ equations that govern state variables $\{\upsilon_{t+1}(i)\}_{i=1}^{\bar{N}}$ are give by Equations 46 - 48.

A.2 Derivation of markup distribution

Limiting case. We first consider the limiting case with $\bar{N} \to \infty$ for distribution of quality gaps $\{\upsilon_t(\Delta)\}_{\Delta=1}^{\infty}$, in steady state, we have $\upsilon_{t+1}(\Delta) = \upsilon_t(\Delta)\forall\Delta$, Equations 46 and 47 imply that

$$\upsilon(\Delta) = \left[\frac{(1-z)I}{z + (1-z)I}\right]^{\Delta} \frac{z}{(1-z)I} = \left[\frac{1}{1 + \frac{z}{(1-z)I}}\right]^{\Delta} \frac{z}{(1-z)I} = \left(\frac{1}{1+\vartheta_I}\right)^{\Delta} \vartheta_I,$$

where $\vartheta_I = \frac{z}{(1-z)I}$. It can then be shown that $\mathbf{P}(\Delta \leq d) = 1-(\frac{1}{1+\vartheta_I})^d = 1-e^{-\ln(1+\vartheta_I)\times d}$. This in turn implies that $\ln\mu = \Delta\ln\lambda$ is exponentially distributed with parameter $\theta = \frac{\ln(1+\vartheta_I)}{\ln\lambda}$. That is $\mathbf{P}(\lambda^{\Delta} \leq \mu) = 1 - \mu^{-\theta}$. Given the distribution for μ, we can show that

$$\Lambda = \sum_{i=1}^{\infty} \lambda^{-i}\upsilon(i) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)} \sum_{i=0}^{\infty}\left(\frac{1}{\lambda(1+\vartheta_I)}\right)^i = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1},$$

$$\mathcal{M} = \frac{e^{-\ln\lambda(\sum_{i=1}^{\infty} i\upsilon(i))}}{\Lambda} = \left(\lambda^{-\frac{1+\vartheta_I}{\vartheta_I}}\right)\left(\frac{\lambda(1+\vartheta_I)-1}{\vartheta_I}\right),$$

since $\sum_{i=1}^{\infty} i\upsilon(i) = \vartheta_I \sum_{i=1}^{\infty} i \left(\frac{1}{1+\vartheta_I}\right)^i = \frac{1+\vartheta_I}{\vartheta_I}$.

Finite markup case. We then consider the special case with finite upper limit $\bar{N}$ such that $\Delta = 1, 2, ..., \bar{N}$, in steady state, we have $\upsilon_{t+1}(\Delta) = \upsilon_t(\Delta) \forall \Delta$, Equations 46 and 47 likewise imply that $\upsilon(\Delta) = \left(\frac{1}{1+\vartheta_I}\right)^{\Delta} \vartheta_I$ for $\Delta \in \{1, 2, ..., \bar{N}-1\}$, where $\vartheta_I = \frac{z}{(1-z)I}$. From Equation 48, it can be shown that in steady state $\upsilon(\bar{N}) = \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}$.

Probability mass function. In steady state, $\{\upsilon(\Delta)\}_{\Delta=1}^{\bar{N}}$ is a probability mass function of the quality gaps $\Delta \in \{1, 2, ..., \bar{N}\}$. This is because $1 - \sum_{i=1}^{\bar{N}-1} \upsilon(i) = 1 - \frac{\vartheta_I}{1+\vartheta_I} \sum_{i=0}^{\bar{N}-2} \left(\frac{1}{1+\vartheta_I}\right)^i = \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} = \upsilon(\bar{N})$ such that $\sum_{\Delta=1}^{\bar{N}} \upsilon(\Delta) = 1$, and $\forall \Delta \in \{1, 2, ..., \bar{N}\}$, $\upsilon(\Delta) > 0$.

Distribution of markups for finite case. Analogous to the infinite case, It can then be shown that for $1 \leq \Delta \leq \bar{N} - 1$, $\mathbf{P}(\Delta \leq d) = 1 - (\frac{1}{1+\vartheta_I})^d = 1 - e^{-\ln(1+\vartheta_I) \times d}$. This in turn implies that $\ln \mu = \Delta \ln \lambda$ is exponentially distributed with parameter $\theta = \frac{\ln(1+\vartheta_I)}{\ln \lambda}$. That is:

$$F(\mu, \Delta) = \mathbf{P}(\lambda^{\Delta} \leq \mu) = \begin{cases} 1 - \mu^{-\frac{\ln(1+\vartheta_I)}{\ln \lambda}}, & \text{if } \Delta = 1, 2, ..., \bar{N} - 1 \\ 1, & \text{if } \Delta = \bar{N} \end{cases}$$

Steady state misallocation measure. We can thus derive steady state misallocation measure $\mathcal{M}_{\bar{N}}$ as follows:

$$\begin{aligned} \Lambda_{\bar{N}} &= \sum_{i=1}^{\bar{N}-1} \lambda^{-i} \upsilon(i) + \lambda^{-\bar{N}} \upsilon(\bar{N}) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)} \sum_{i=0}^{\bar{N}-2} \left(\frac{1}{\lambda(1+\vartheta_I)}\right)^i + \lambda^{-\bar{N}} \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} \\ &= \frac{\vartheta_I}{\lambda(1+\vartheta_I) - 1} \left(1 - \left(\frac{1}{\lambda(1+\vartheta_I)}\right)^{\bar{N}-1}\right) + \lambda^{-\bar{N}} \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} \end{aligned}$$

Similarly, $\sum_{i=1}^{\bar{N}} i\upsilon(i) = \vartheta_I \sum_{i=1}^{\bar{N}-1} i \left(\frac{1}{1+\vartheta_I}\right)^i + \bar{N}\upsilon(\bar{N}) = \vartheta_I \left[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I}\right] + \bar{N} \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}$, where $\phi_I \equiv \frac{1}{1+\vartheta_I}$, let $\Xi_{\bar{N}} \equiv \sum_{i=1}^{\bar{N}} iv(i)$, such that $m_t = -\Xi(\ln \lambda)$. Finally, we have

$$\mathcal{M}_{\bar{N}} = \frac{\lambda^{-\Xi_{\bar{N}}}}{\Lambda_{\bar{N}}}, \text{ where } \Xi_{\bar{N}} = \vartheta_I \left[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I} \right] + \bar{N}(\frac{1}{1+\vartheta_I})^{\bar{N}-1}.$$

It can be easily shown that $\mathcal{M}_{\bar{N}} \to \mathcal{M}$ as $\bar{N} \to \infty$.

A.3 Steady States

Limiting case. As $\bar{N} \to \infty$, the balanced growth path (BGP) for the economy are characterized by constant values for technical efficiency growth rate g, probability of successful entry z, employment L, normalized output y, and price inflation Π satisfying:

$$g^\sigma = \frac{\beta(1+i)}{\Pi} \tag{56}$$

$$\beta g^{1-\sigma}(1-\lambda^{-1})(L-\delta_z z) = \zeta_z \delta_z z^{\zeta_z - 1}\Lambda, \text{ where } \Lambda = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}, \vartheta_I = \frac{z}{(1-z)I} \tag{57}$$

$$\Pi = \frac{\pi^w}{g} \tag{58}$$

$$g = (z + I - zI)(\lambda - 1) + 1 \tag{59}$$

$$y = (L - \delta_z z^{\zeta_z})\mathcal{M}, \text{ where } \mathcal{M} = \left(\lambda^{-\frac{1+\vartheta_I}{\vartheta_I}}\right)\left(\frac{\lambda(1+\vartheta_I)-1}{\vartheta_I}\right) \tag{60}$$

The five equations are (i) household's Euler equation, (ii) outside entrepreneur's innovation decision, (iii) wage rigidity equation, (iv) growth rate of technical efficiency, and (v) resource constraint.

Finite markup case. For any given $\bar{N}$, the balanced growth path (BGP) for the economy are characterized by constant values for technical efficiency growth rate g, probability of successful entry z, employment L, normalized output y, and price inflation Π satisfying:

$$g^{\sigma} = \frac{\beta(1+i)}{\Pi} \tag{61}$$

$$\beta g^{1-\sigma}(1-\lambda^{-1})(L-\delta_z z) = \zeta_z \delta_z z^{\zeta_z - 1}\Lambda_{\bar{N}}, \text{ where} \tag{62}$$

$$\Lambda_{\bar{N}} = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}(1-(\frac{1}{\lambda(1+\vartheta_I)})^{\bar{N}-1}) + \lambda^{-\bar{N}}(\frac{1}{1+\vartheta_I})^{\bar{N}-1}, \vartheta_I = \frac{z}{(1-z)I}$$

$$\Pi = \frac{\pi^w}{g} \tag{63}$$

$$g = (z + I - zI)(\lambda - 1) + 1 \tag{64}$$

$$y = (L - \delta_z z^{\zeta_z})\mathcal{M}_{\bar{N}}, \text{ where} \tag{65}$$

$$\mathcal{M}_{\bar{N}} = \frac{\lambda^{-\Xi_{\bar{N}}}}{\Lambda_{\bar{N}}}, \text{ where } \Xi_{\bar{N}} = \vartheta_I[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I}] + \bar{N}(\frac{1}{1+\vartheta_I})^{\bar{N}-1}, \phi_I = \frac{1}{1+\vartheta_I}$$

The five equations are (i) household's Euler equation, (ii) outside entrepreneur's innovation decision, (iii) wage rigidity equation, (iv) growth rate of technical efficiency, and (v) resource constraint.

A.4 Log-linearized equilibrium

Finite markup case. The limiting case is summarized in Section 2.5 of the main text, for the special case with $\bar{N} < \infty$, we define the log-linearized competitive equilibrium as follows:

Definition. For a given finite $\bar{N}$, the log-linearized competitive equilibrium of the economy is defined as a sequence of variables $\{d\log y_t, d\log g_{t+1}, dz_t, di_t, dL_t, d\log \Pi_t, \{dv_{t+1}(\Delta)\}_{\Delta=1}^{\bar{N}}, dm_{t+1}, d\log \mathcal{M}_{t+1}\}$ that satisfy the following $8+\bar{N}$ equations, for a given sequence of exogenous shocks $\{d\log \epsilon_t^m\}$ and given initial values of state variables $\{\{dv_0(\Delta)\}_{\Delta=1}^{\bar{N}}, dm_0, d\log \mathcal{M}_0\}$.

Consumption Euler equation

$$-\sigma d\log y_t = -\sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1} + d\log(1+i_t) - d\mathbf{E}_t \log \Pi_{t+1} \tag{66}$$

Entrant's innovation decision

$$[\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1}] + \mathbf{E}_t d\log y_{t+1} + d\log g_{t+1} = dm_t + \frac{\zeta_z - 1}{z} dz_t \quad (67)$$

Wage rigidity equation

$$d\log \Pi_t = -d\log g_t + dm_{t-1} - dm_t \quad (68)$$

Technical efficiency growth equation

$$d\log g_{t+1} = \frac{(\lambda-1)(1-I)}{(z+I(1-z))(\lambda-1)+1} dz_t \quad (69)$$

Measure of aggregate misallocation

$$dm_{t+1} = -(\ln\lambda)\sum_{\Delta=1}^{\bar{N}} \Delta d\upsilon_{t+1}(\Delta) \quad (70)$$

$$\begin{aligned} d\log \mathcal{M}_{t+1} = &\left[\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)\kappa_A + (\ln\lambda)\kappa_B \mathcal{S}_I + \frac{\kappa_B}{\Lambda}\mathcal{S}_\lambda - \left(-(\ln\lambda)\bar{N} - \frac{\lambda^{-\bar{N}}}{\Lambda}\right)\kappa_C\right] dz_t \\ &+ f\left(\sum_{\Delta=1}^{\bar{N}} d\upsilon_t(\Delta)\right), \text{ where} \end{aligned} \quad (71)$$

$$\begin{aligned} &\kappa_A \equiv 1-(1-I)\upsilon(1),\ \kappa_B \equiv \frac{1}{1-z},\ \kappa_C \equiv \kappa_B \upsilon(\bar{N}), \\ &\mathcal{S}_I \equiv \sum_{\Delta=2}^{\bar{N}-1} \Delta \upsilon(\Delta) = \vartheta_I\left[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I}\right] - \frac{\vartheta_I}{1+\vartheta_I}, \vartheta_I \equiv \frac{z}{(1-z)I}, \phi_I \equiv \frac{1}{1+\vartheta_I}, \\ &\mathcal{S}_\lambda \equiv \sum_{\Delta=2}^{\bar{N}-1} \lambda^{-\Delta}\upsilon(\Delta) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}\left[1-\left(\frac{1}{\lambda(1+\vartheta_I)}\right)^{\bar{N}-1}\right] - \lambda^{-1}\frac{\vartheta_I}{1+\vartheta_I}. \end{aligned}$$

Resource constraint

$$d\log y_t = d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log \mathcal{M}_t \tag{72}$$

Monetary policy rule

$$d\log(1+i_t) = \phi d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log \epsilon_t^m, \phi > 0, i_t > 0 \tag{73}$$

Evolution of quality gaps ($\bar{N}$ equations)

$$d\upsilon_{t+1}(\Delta) = \begin{cases} [1-(1-I)\upsilon(\Delta)]dz_t + (1-z)(1-I)d\upsilon_t(\Delta) & \text{if } \Delta = 1, \\ [-(1-I)\upsilon(\Delta) - I\upsilon(\Delta-1)]dz_t + & \\ \quad (1-z)(1-I)d\upsilon_t(\Delta) + (1-z)Id\upsilon_t(\Delta-1) & \text{if } 2 \leq \Delta \leq \bar{N}-1, \\ [-\upsilon(\Delta) - I\upsilon(\Delta-1)]dz_t + (1-z)d\upsilon_t(\Delta) + (1-z)Id\upsilon_t(\Delta-1) & \text{if } \Delta = \bar{N}, \end{cases} \tag{74}$$

where $\upsilon(\Delta) = (\frac{1}{1+\vartheta_I})^{\Delta}\vartheta_I$, and $\vartheta_I = \frac{z}{(1-z)I}$ for $\Delta \leq \bar{N}-1$, and $\upsilon(\bar{N}) = (\frac{1}{1+\vartheta_I})^{\bar{N}-1}$.

Relative to the limiting case, the finite markup case requires keeping track of the evolution of the quality gaps. The consumption euler equation is the log-linearized version of household's inter-temporal utility maximization condition; the entrant's innovation decision is the log-linearized transformation of outside entrepreneur's maximization condition; the wage rigidity equation is the log-linearized version of price inflation; technical efficiency growth equation summarizes how does percentage changes in rate of entrant's innovation translate into growth rate of technical efficiency. Resource constraint and monetary policy rule are obtained from log-linearizing their counterparts in the normalized competitive equilibrium, respectively.

Finally, the misallocation block with the aggregate misallocation measure and the evolution of quality gaps (eqns 70 - 71 and eqn 74, respectively), both of which are log-linearized transformation of the misallocation measure and evolution of masses at each quality gap around the steady state. In particular, Equation 71 shows how a percentage change in rate of entrant's innovation could translate into changes in measure of aggregate misallocation through the evolution of the distribution of varieties with different quality gaps as governed by Equation 74.

B Theoretical Derivation

B.1 Derivation of the log-linearized equilibrium

Log-linearization of the finite case. Endogenous variables: $y_t, L_t, z_t, g_{t+1}, i_t, \Pi_t, m_{t+1}, \mathcal{M}_{t+1}, \{\upsilon_{t+1}(\Delta)\}_{\Delta=1}^{\bar{N}}$

Consumption Euler equation

$$-\sigma d\log y_t = -\sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1} + d\log(1+i_t) - d\mathbf{E}_t \log \Pi_{t+1} \tag{75}$$

Outside entrepreneur's innovation decision

$$[\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1}] + \mathbf{E}_t d\log y_{t+1} + d\log g_{t+1} = dm_t + \frac{\zeta_z - 1}{z} dz_t \tag{76}$$

Wage rigidity equation

$$d\log \Pi_t = -d\log g_t + dm_{t-1} - dm_t, \text{ where} \tag{77}$$

Growth rate of technical efficiency

$$d\log g_{t+1} = \frac{(\lambda-1)[(1-I)dz_t + (1-z)dI_t]}{(z+I(1-z))(\lambda-1)+1} \tag{78}$$

Resource constraint

$$d\log y_t = d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log \mathcal{M}_t,\ \text{where } d\log(L_t - \delta_z z_t^{\zeta_z}) = \frac{dL_t - \zeta_z \delta_z z^{\zeta_z - 1} dz_t}{L - \delta_z z^{\zeta_z}} \tag{79}$$

Monetary policy rule

$$d\log(1+i_t) = \phi d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log \epsilon_t^m, \phi > 0, i_t > 0 \tag{80}$$

Evolution of quality gaps ($\bar{N}$ equations)

$$dv_{t+1}(\Delta) = [1-(1-I)v(\Delta)]dz_t + [-(1-z)v(\Delta)]dI_t + (1-z)(1-I)dv_t(\Delta),\ \text{if } \Delta = 1 \tag{81}$$

$$\begin{aligned} dv_{t+1}(\Delta) &= [-(1-I)v(\Delta) - Iv(\Delta-1)]dz_t + [(1-z)(v(\Delta-1) - v(\Delta))]dI_t \\ &\quad + (1-z)(1-I)dv_t(\Delta) + (1-z)Idv_t(\Delta-1),\ \text{if } 2 \le \Delta \le \bar{N}-1 \\ &\equiv \mathcal{C}_A(\Delta)dz_t + \mathcal{C}_B(\Delta)dI_t + \mathcal{C}_C dv_t(\Delta) + \mathcal{C}_D dv_t(\Delta-1) \end{aligned} \tag{82}$$

$$dv_{t+1}(\Delta) = [-v(\Delta) - Iv(\Delta-1)]dz_t + [(1-z)v(\Delta-1)]dI_t \tag{83}$$

$$\quad + (1-z)dv_t(\Delta) + (1-z)Idv_t(\Delta-1),\ \text{if } \Delta = \bar{N}, \tag{84}$$

where $v(\Delta) = (\frac{1}{1+\vartheta_I})^{\Delta}\vartheta_I$, and $\vartheta_I = \frac{z}{(1-z)I}$ for $\Delta \le \bar{N}-1$, and $v(\bar{N}) = (\frac{1}{1+\vartheta_I})^{\bar{N}-1}$.

Block of misallocation measure

$$d\log \mathcal{M}_{t+1} = dm_{t+1} - d\log \Lambda_{t+1},$$

where

$$dm_{t+1} = -(\ln\lambda)\sum_{\Delta=1}^{\bar{N}} \Delta dv_{t+1}(\Delta) \equiv d\log w_{t+1}, \tag{85}$$

where w_t is normalized real wage, the equation holds because $w_t = \exp(m_t)$, and

$$d \log \Lambda_{t+1} = \frac{d\Lambda_{t+1}}{\Lambda} = \frac{1}{\Lambda} \sum_{\Delta=1}^{\bar{N}} \lambda^{-\Delta} dv_{t+1}(\Delta),$$

where $\Lambda = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1} \left(1 - \left(\frac{1}{\lambda(1+\vartheta_I)}\right)^{\bar{N}-1}\right) + \lambda^{-\bar{N}} \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}$,

$$\begin{aligned}
dm_{t+1} &= -(\ln \lambda) dv_{t+1}(1) - (\ln \lambda) \sum_{\Delta=2}^{\bar{N}-1} \Delta [\mathcal{C}_A(\Delta) dz_t + \mathcal{C}_B(\Delta) dI_t + \mathcal{C}_C dv_t(\Delta) + \mathcal{C}_D dv_t(\Delta - 1)] \\
&\quad - (\ln \lambda) \bar{N} dv_{t+1}(\bar{N}) \\
&= -(\ln \lambda) dv_{t+1}(1) - (\ln \lambda) \{[-(1-I)\mathcal{S}_I - I(1+\vartheta_I)\mathcal{S}_I]\, dz_t + [(1-z)\vartheta_I \mathcal{S}_I]\, dI_t\} - (\ln \lambda) \bar{N} dv_{t+1}(\bar{N}) \\
&\quad - (\ln \lambda) \sum_{\Delta=2}^{\bar{N}-1} \Delta [\mathcal{C}_C dv_t(\Delta) + \mathcal{C}_D dv_t(\Delta - 1)],
\end{aligned}$$

$$\begin{aligned}
d\Lambda_{t+1} &= \sum_{\Delta=1}^{\bar{N}} \lambda^{-\Delta} dv_{t+1}(\Delta) \\
&= \lambda^{-1} dv_{t+1}(1) + \sum_{\Delta=2}^{\bar{N}-1} \lambda^{-\Delta} [\mathcal{C}_A(\Delta) dz_t + \mathcal{C}_B(\Delta) dI_t + \mathcal{C}_C dv_t(\Delta) + \mathcal{C}_D dv_t(\Delta - 1)] + \lambda^{-\bar{N}} dv_{t+1}(\bar{N}) \\
&= \lambda^{-1} dv_{t+1}(1) + \{[-(1-I)\mathcal{S}_\lambda - I(1+\vartheta_I)\mathcal{S}_\lambda]\, dz_t + [(1-z)\vartheta_I \mathcal{S}_\lambda]\, dI_t\} + \lambda^{-\bar{N}} dv_{t+1}(\bar{N}) \\
&\quad + \sum_{\Delta=2}^{\bar{N}-1} \lambda^{-\Delta} [\mathcal{C}_C dv_t(\Delta) + \mathcal{C}_D dv_t(\Delta - 1)],
\end{aligned}$$

where

$$
\begin{aligned}
d\upsilon_{t+1}(1) &= [1-(1-I)\frac{\vartheta_I}{1+\vartheta_I}]dz_t + [-(1-z)\frac{\vartheta_I}{1+\vartheta_I}]dI_t + (1-z)(1-I)d\upsilon_t(1), \\
d\upsilon_{t+1}(\bar{N}) &= \left[-\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} - I\vartheta_I\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}\right]dz_t + \left[(1-z)\vartheta_I\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}\right]dI_t \\
&\quad + (1-z)d\upsilon_t(\Delta) + (1-z)Id\upsilon_t(\Delta-1), \\
\mathcal{C}_C &= (1-z)(1-I), \\
\mathcal{C}_D &= (1-z)I, \\
\mathcal{S}_I &= \sum_{\Delta=2}^{\bar{N}-1}\Delta\upsilon(\Delta) = \vartheta_I\left[\frac{\phi_I(1-\phi_I^{\bar{N}-1})}{(1-\phi_I)^2} - \frac{(\bar{N}-1)\phi_I^{\bar{N}}}{1-\phi_I}\right] - \frac{\vartheta_I}{1+\vartheta_I}, \text{ where } \phi_I = \frac{1}{1+\vartheta_I}, \\
\mathcal{S}_\lambda &= \sum_{\Delta=2}^{\bar{N}-1}\lambda^{-\Delta}\upsilon(\Delta) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}\left[1-\left(\frac{1}{\lambda(1+\vartheta_I)}\right)^{\bar{N}-1}\right] - \lambda^{-1}\frac{\vartheta_I}{1+\vartheta_I}.
\end{aligned}
$$

Rewrite $dm_{t+1}, d\Lambda_{t+1}$ in terms of $dz_t, dI_t, \{d\upsilon_t(\Delta)\}_{\Delta=1}^{\bar{N}}$:

$$
\begin{aligned}
dm_{t+1} = &\left\{-(\ln\lambda)\underbrace{\left[1-(1-I)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{\kappa_A} - (\ln\lambda)\underbrace{\left(-1-\frac{z}{1-z}\right)}_{-\kappa_B}\mathcal{S}_I\right. \\
&\left. -(\ln\lambda)\bar{N}\underbrace{\left(-1-\frac{z}{1-z}\right)\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}}_{-\kappa_C}\right\}dz_t \\
&+\left\{-(\ln\lambda)\underbrace{\left[-(1-z)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{-\kappa_D} - (\ln\lambda)\underbrace{(1-z)\vartheta_I}_{\kappa_E}\mathcal{S}_I - (\ln\lambda)\bar{N}\underbrace{(1-z)\vartheta_I\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}}_{\kappa_F}\right\}dI_t \\
&-(\ln\lambda)(1-z)\sum_{\Delta=1}^{\bar{N}-1}\Delta d\upsilon_t(\Delta) - (\ln\lambda)(1-z)d\upsilon_t(\bar{N})(\bar{N}-I) \\
= &-(\ln\lambda)[\kappa_A - \kappa_B\mathcal{S}_I - \kappa_C\bar{N}]dz_t - (\ln\lambda)[-\kappa_D + \kappa_E\mathcal{S}_I + \kappa_F\bar{N}]dI_t + (1-z)dm_t \\
&+(\ln\lambda)(1-z)Id\upsilon_t(\bar{N})
\end{aligned}
$$

$$
\begin{aligned}
d\Lambda_{t+1} =& \left\{ \lambda^{-1} \underbrace{\left[1-(1-I)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{\kappa_A} + \underbrace{\left(-1-\frac{z}{1-z}\right)}_{-\kappa_B} \mathcal{S}_\lambda + \lambda^{-\bar{N}} \underbrace{\left(-1-\frac{z}{1-z}\right)\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}}_{-\kappa_C} \right\} dz_t \\
&+ \left\{ \lambda^{-1} \underbrace{\left[-(1-z)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{-\kappa_D} + \underbrace{(1-z)\vartheta_I}_{\kappa_E} \mathcal{S}_\lambda + \lambda^{-\bar{N}} \underbrace{(1-z)\vartheta_I \left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1}}_{\kappa_F} \right\} dI_t \\
&+ (1-z)\sum_{\Delta=1}^{\bar{N}-1} \left[1-I+I\lambda^{-1}\right] \lambda^{-\Delta} d\upsilon_t(\Delta) + \lambda^{-\bar{N}}(1-z)d\upsilon_t(\bar{N}) \\
=&[\lambda^{-1}\kappa_A - \kappa_B\mathcal{S}_\lambda - \lambda^{-\bar{N}}\kappa_C]dz_t + [-\lambda^{-1}\kappa_D + \kappa_E\mathcal{S}_\lambda + \lambda^{-\bar{N}}\kappa_F]dI_t \\
&+ (1-z)\left(1-I+\frac{I}{\lambda}\right)d\Lambda_t + (1-z)\left(I-\frac{I}{\lambda}\right)\lambda^{-\bar{N}} d\upsilon_t(\bar{N})
\end{aligned}
$$

Since $d\log\mathcal{M}_{t+1} = dm_{t+1} - d\log\Lambda_{t+1}$,, we can rewrite $d\log\mathcal{M}_{t+1}$ as:

$$
\begin{aligned}
d\log\mathcal{M}_{t+1} =& \left[\underbrace{\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)\kappa_A}_{\ominus} + \underbrace{(\ln\lambda)\kappa_B\mathcal{S}_I}_{\oplus} + \underbrace{\frac{\kappa_B}{\Lambda}\mathcal{S}_\lambda}_{\oplus} - \underbrace{\left(-(\ln\lambda)\bar{N} - \frac{\lambda^{-\bar{N}}}{\Lambda}\right)\kappa_C}_{\ominus} \right] dz_t \\
&+ \left[\underbrace{\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)(-\kappa_D)}_{\oplus} - \underbrace{(\ln\lambda)\kappa_E\mathcal{S}_I}_{\oplus} - \underbrace{\frac{\kappa_E}{\Lambda}\mathcal{S}_\lambda}_{\oplus} + \underbrace{\left(-(\ln\lambda)\bar{N} - \frac{\lambda^{-\bar{N}}}{\Lambda}\right)\kappa_F}_{\ominus} \right] dI_t \\
&+ (1-z)dm_t - \frac{(1-z)\left(1-I+\frac{I}{\lambda}\right)}{\Lambda}d\Lambda_t + \left[(\ln\lambda)(1-z)I - \frac{(1-z)\left(I-\frac{I}{\lambda}\right)\lambda^{-\bar{N}}}{\Lambda}\right]d\upsilon_t(\bar{N})
\end{aligned} \tag{86}
$$

where

$$\kappa_A \equiv 1-(1-I)\frac{\vartheta_I}{1+\vartheta_I} = 1-(1-I)\upsilon(1)$$
$$\kappa_B \equiv \frac{1}{1-z}$$
$$\kappa_C \equiv \left(\frac{1}{1-z}\right)\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} = \left(\frac{1}{1-z}\right)\upsilon(\bar{N})$$
$$\kappa_D \equiv (1-z)\frac{\vartheta_I}{1+\vartheta_I} = (1-z)\upsilon(1)$$
$$\kappa_E \equiv (1-z)\vartheta_I$$
$$\kappa_F \equiv (1-z)\vartheta_I\left(\frac{1}{1+\vartheta_I}\right)^{\bar{N}-1} = (1-z)\vartheta_I\upsilon(\bar{N})$$

Limiting case. As $\bar{N}\to\infty$, we can rewrite the system as:

$$\mathcal{S}_I \to \mathcal{S}_I^\star \equiv \sum_{\Delta=2}^{\infty}\Delta\upsilon(\Delta) = \frac{1+\vartheta_I}{\vartheta_I} - \frac{\vartheta_I}{1+\vartheta_I}$$

$$\mathcal{S}_\lambda \to \mathcal{S}_\lambda^\star \equiv \sum_{\Delta=2}^{\infty}\lambda^{-\Delta}\upsilon(\Delta) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1} - \lambda^{-1}\left(\frac{\vartheta_I}{1+\vartheta_I}\right)$$

Hence,

$$\begin{aligned} dm_{t+1} &= \left\{-(\ln\lambda)\underbrace{\left[1-(1-I)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{\kappa_A} - (\ln\lambda)\underbrace{\left(-1-\frac{z}{1-z}\right)}_{-\kappa_B}\mathcal{S}_I^\star\right\}dz_t \\ &\quad + \left\{-(\ln\lambda)\underbrace{\left[-(1-z)\frac{\vartheta_I}{1+\vartheta_I}\right]}_{-\kappa_D} - (\ln\lambda)\underbrace{(1-z)\vartheta_I}_{\kappa_E}\mathcal{S}_I^\star\right\}dI_t + (1-z)dm_t \\ &= \underbrace{-(\ln\lambda)[\kappa_A - \kappa_B\mathcal{S}_I^\star]}_{\equiv\alpha_m}dz_t - (\ln\lambda)[-\kappa_D + \kappa_E\mathcal{S}_I^\star]dI_t + (1-z)dm_t \end{aligned} \tag{87}$$

$$
\begin{aligned}
d\Lambda_{t+1} =& \left\{ \lambda^{-1} \underbrace{\left[1-(1-I)\frac{\vartheta_I}{1+\vartheta_I} \right]}_{\kappa_A} + \underbrace{\left(-1-\frac{z}{1-z} \right)}_{-\kappa_B} \mathcal{S}^\star_\lambda \right\} dz_t \\
&+ \left\{ \lambda^{-1} \underbrace{\left[-(1-z)\frac{\vartheta_I}{1+\vartheta_I} \right]}_{-\kappa_D} + \underbrace{(1-z)\vartheta_I}_{\kappa_E} \mathcal{S}^\star_\lambda \right\} dI_t + (1-z)\left(1-I+\frac{I}{\lambda} \right) d\Lambda_t \\
=& [\lambda^{-1}\kappa_A - \kappa_B \mathcal{S}^\star_\lambda] dz_t + [-\lambda^{-1}\kappa_D + \kappa_E \mathcal{S}^\star_\lambda] dI_t + (1-z)\left(1-I+\frac{I}{\lambda} \right) d\Lambda_t \qquad (88)
\end{aligned}
$$

Hence,

$$
d\log\Lambda_{t+1} = \underbrace{\frac{1}{\Lambda}[\lambda^{-1}\kappa_A - \kappa_B \mathcal{S}^\star_\lambda]}_{\equiv \alpha_\Lambda} dz_t + \frac{1}{\Lambda}[-\lambda^{-1}\kappa_D + \kappa_E \mathcal{S}^\star_\lambda] dI_t + \frac{1}{\Lambda}(1-z)\left(1-I+\frac{I}{\lambda} \right) d\Lambda_t \quad (89)
$$

$$
\begin{aligned}
d\log\mathcal{M}_{t+1} =& \left[\underbrace{\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda} \right)\kappa_A}_{\ominus} + \underbrace{(\ln\lambda)\kappa_B \mathcal{S}^\star_I}_{\oplus} + \underbrace{\frac{\kappa_B}{\Lambda}}_{\oplus} \mathcal{S}^\star_\lambda \right] dz_t \\
&+ \left[\underbrace{\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda} \right)(-\kappa_D)}_{\oplus} - \underbrace{(\ln\lambda)\kappa_E \mathcal{S}^\star_I}_{\oplus} - \underbrace{\frac{\kappa_E}{\Lambda}}_{\oplus} \mathcal{S}^\star_\lambda \right] dI_t \\
&+ (1-z) dm_t - \frac{(1-z)\left(1-I+\frac{I}{\lambda} \right)}{\Lambda} d\Lambda_t \qquad (90) \\
=& (\alpha_m - \alpha_\Lambda) dz_t + \mathcal{C}_I dI_t + (1-z) d\log\mathcal{M}_t + \frac{(1-z)I\left(1-\frac{1}{\lambda} \right)}{\Lambda} d\Lambda_t \\
=& (\alpha_m - \alpha_\Lambda) dz_t + \mathcal{C}_I dI_t + (1-z) d\log\mathcal{M}_t + (1-z)I\left(1-\frac{1}{\lambda} \right)(dm_t - d\log\mathcal{M}_t) \\
=& (\alpha_m - \alpha_\Lambda) dz_t + \mathcal{C}_I dI_t + (1-z)\left[1-I\left(1-\frac{1}{\lambda} \right) \right] d\log\mathcal{M}_t + (1-z)I\left(1-\frac{1}{\lambda} \right) dm_t
\end{aligned}
$$

B.2 Proofs of Propositions 3 and 4

To prove Proposition 3, change in measure of aggregate misallocation $d\log\mathcal{M}_{t+1}$ can be written as

$$d\log\mathcal{M}_{t+1} = \alpha_1 d\upsilon_{t+1}(1) + \sum_{j=2}^{\infty} \alpha_j d\upsilon_{t+1}(j), \tag{91}$$

where

$$\alpha_j = -(\ln\lambda)(j) - \frac{\lambda^{-j}}{\Lambda} < 0 \ \forall j. \tag{92}$$

α are all negative and α is decreasing in j for $j \geq 2$, since $\frac{d\alpha_j}{dj} = -\ln\lambda + (\ln\lambda)\frac{\lambda^{-j}}{\Lambda} = (\ln\lambda)(-1 + \frac{1}{\lambda^j\Lambda}) < 0$ for $j \geq 2$, such that $|\alpha_2| < |\alpha_3| < ... < |\alpha_{\bar{N}}| < ...$ For $|\alpha_1| < |\alpha_2|$ to holds, it must be that:

$$|\alpha_2| - |\alpha_1| = \ln\lambda + \frac{1}{\lambda^2\Lambda} - \frac{1}{\lambda\Lambda} > 0. \tag{93}$$

If the condition in Equation 93 is satisfied, then the weights are decreasing in j and that $|\alpha_j|$ is increasing in $j\forall j$.

It can be easily shown that $d\upsilon_{t+1}(1)$ is increasing in z_t, whereas $d\upsilon_{t+1}(j)\ \forall j \neq 1$ are decreasing in z_t, and that:

$$-d\upsilon_{t+1}(1) = \sum_{j=2}^{\infty} d\upsilon_{t+1}(j) \tag{94}$$

As such, whether or not misallocation increase would depend on the relative magnitudes of weighted $\alpha_1 d\upsilon_{t+1}(1)$ versus the weighted sum of all the other $\alpha_j d\upsilon_{t+1}(j)$. Since $|\alpha_j|$ is increasing in j for $j \geq 2$, it follows that if the sufficient condition in Equation 93 holds, measure of misallocation is increasing in z_t, since:

$$|\alpha_1 d\upsilon_{t+1}(1)| < |\sum_{j=2}^{\infty} \alpha_j d\upsilon_{t+1}(j)|. \tag{95}$$

For the finite markup case with $\bar{N} < \infty$, the above equation would still hold.

To prove Proposition 4, for the limiting case when $\bar{N} \to \infty$, we rewrite $d\upsilon_{t+1}(j)$ as $\kappa_j dz_t$.

It follows that: as long as $|\alpha_1\kappa_1| < |\sum_{j=2}^{\infty}\alpha_j\kappa_j|$ is satisfied, misallocation measure at time 1 would fall in response to a reduction in entry rate at time 0. That is:

$$\left(\ln\lambda + \frac{\lambda^{-1}}{\Lambda}\right)\kappa_1 < (\ln\lambda)\kappa_B\mathcal{S}_I^\star + \frac{\kappa_B}{\Lambda}\mathcal{S}_\lambda^\star \tag{96}$$

where

$$\begin{aligned}
\kappa_1 &\equiv 1-(1-I)\frac{\vartheta_I}{1+\vartheta_I}, \kappa_B \equiv \frac{1}{1-z}\\
\Lambda &\equiv \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}\\
\mathcal{S}_I^\star &\equiv \frac{1+\vartheta_I}{\vartheta_I} - \frac{\vartheta_I}{1+\vartheta_I}\\
\mathcal{S}_\lambda^\star &\equiv \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1} - \lambda^{-1}\left(\frac{\vartheta_I}{1+\vartheta_I}\right)
\end{aligned}$$

To prove the special case with a finite markup, we again rewrite $d\upsilon_{t+1}(j)$ as $\kappa_j dz_t$, where

$$\kappa_1 \equiv 1-(1-I)\upsilon(1) = \kappa_A \tag{97}$$

$$\kappa_j \equiv -[(1-I)\upsilon(j) + I\upsilon(j-1)], \text{ for } j = 2, ..., \bar{N}-1 \tag{98}$$

$$\kappa_N \equiv -[\upsilon(\bar{N}) + I\upsilon(\bar{N}-1)] = -\kappa_C, \tag{99}$$

change in aggregate misallocation measure $d\log\mathcal{M}_{t+1}$ can be rewritten as

$$d\log\mathcal{M}_{t+1} = \alpha_1\kappa_1 dz_t + \sum_{j=2}^{\bar{N}}\alpha_j\kappa_j dz_t \tag{100}$$

$$= \left[\underbrace{\left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)\kappa_1}_{\ominus} + \underbrace{(\ln\lambda)\kappa_B\mathcal{S}_I}_{\oplus} + \underbrace{\frac{\kappa_B}{\Lambda}\mathcal{S}_\lambda}_{\oplus} + \underbrace{\left(-(\ln\lambda)\bar{N} - \frac{\lambda^{-\bar{N}}}{\Lambda}\right)\kappa_N}_{\oplus}\right]dz_t. \tag{101}$$

It follows that if and only if the coefficient in front of z_t is positive, $d\log\mathcal{M}_{t+1}$ is increasing in dz_t. That is:

$$|\alpha_1\kappa_1| < |\sum_{j=2}^{\bar{N}} \alpha_j\kappa_j|. \tag{102}$$

And that $\sum_{j=1}^{\bar{N}} \alpha_j\kappa_j$ determines how much does aggregate measure of misallocation next period change following one percentage point change in z_t. That is

$$\underbrace{\left(\ln\lambda + \frac{\lambda^{-1}}{\Lambda}\right)\kappa_A}_{\oplus} < \underbrace{(\ln\lambda)\kappa_B\mathcal{S}_I}_{\oplus} + \underbrace{\frac{\kappa_B}{\Lambda}\mathcal{S}_\lambda}_{\oplus} + \underbrace{\left((\ln\lambda)\bar{N} + \frac{\lambda^{\bar{N}}}{\Lambda}\right)\kappa_C}_{\oplus}. \tag{103}$$

B.3 Proof of Proposition 5

Consider an economy along the balanced growth path. At time 0, there is an unanticipated increase in nominal interest rate. $di_0 > 0$. Suppose further that the central bank restores full employment in production sector from $t \geq 1$, such that $d\log(L_t - \delta_z z_t^{\zeta_z}) = 0 \forall t \geq 1$.

From Equation 69, we have

$$d\log g_{t+1} = \underbrace{\frac{(\lambda-1)(1-I))}{(z+I(1-z))(\lambda-1)+1}}_{\equiv\alpha_g>0} dz_t.$$

From Equation 68, we have

$$d\log\Pi_{t+1} = -\log g_{t+1} + dm_t - dm_{t+1}, \quad \forall t \geq 1,$$
$$d\log\Pi_1 = -\log g_1 - dm_t, \text{ for } t = 0.$$

From Equation 70, we have

$$dm_{t+1} = -(\ln\lambda)\left[d\upsilon_{t+1}(1) + \sum_{\Delta=2}^{\bar{N}} \Delta d\upsilon_{t+1}(\Delta)\right].$$

Since $d\upsilon_1(1)$ is increasing in dz_0, while $d\upsilon_1(\Delta)\forall\Delta \geq 2$ is decreasing in dz_0, and $\sum_{\Delta=1}^{\bar{N}} d\upsilon_1(\Delta) =$

0. Then $\sum_{\Delta=1}^{\bar{N}} \Delta dv_1(\Delta)$ is decreasing in dz_0. Hence

$$dm_1 = -(\ln\lambda)(\kappa_A - \kappa_B \mathcal{S}_I - \bar{N}\kappa_C)dz_0 = \underbrace{(\ln\lambda)(\kappa_B \mathcal{S}_I + \bar{N}\kappa_C - \kappa_A)}_{\equiv \alpha_m > 0} dz_0.$$

Similarly, $d\Lambda_1 = \sum_{\Delta=1}^{\bar{N}} \lambda^{-\Delta} dv_1(\Delta)$ is an increasing function of dz_0. Hence

$$d\log\Lambda_1 = \frac{d\Lambda_1}{\Lambda} = \underbrace{\frac{1}{\Lambda}(\lambda^{-1}\kappa_A - \kappa_B \mathcal{S}_\lambda - \lambda^{-\bar{N}}\kappa_C)}_{\equiv \alpha_\lambda > 0} dz_0.$$

Therefore, we have

$$d\log\Pi_1 = -(\alpha_g + \alpha_m)dz_0,$$

and from Equation 71, suppose that Proposition 4 holds, then

$$d\log\mathcal{M}_1 = (\alpha_m - \alpha_\lambda)dz_0, \text{ where } \alpha_m - \alpha_\lambda > 0.$$

From Equation 72, and since the central bank would stabilize the economy to full production employment in period 1 onward, we have

$$d\log y_1 = d\log\mathcal{M}_1 = (\alpha_m - \alpha_\lambda)dz_0,$$
$$d\log y_0 = d\log(L_0 - \delta_z z_0^{\zeta_z}).$$

From consumption euler equation 66, we have

$$-\sigma d\log y_0 - [\sigma\alpha_\lambda + (1-\sigma)\alpha_m + (1-\sigma)\alpha_g]dz_0 = d\log(1+i_0), \tag{104}$$

and from entrant's innovation condition 67

$$d\log y_0 = \frac{\sigma-1}{\sigma}(\alpha_m - \alpha_\lambda + \alpha_g)dz_0 + \frac{\zeta_z - 1}{z}dz_0. \tag{105}$$

We focus on the case: $\sigma > 1$, then from Equation 104, we have:[13]

$$dz_0 = -\frac{1}{\alpha_\lambda + \frac{\sigma(\zeta_z - 1)}{z}} d\log(1+i_0),$$

i.e., dz_0 decreases on impact on a contractionary monetary policy shock, and

$$d\log y_0 = -\underbrace{\left(\frac{\sigma-1}{\sigma}\right)\left(\frac{\alpha_m - \alpha_\lambda + \alpha_g}{\alpha_\lambda + \frac{\sigma(\zeta_z-1)}{z}}\right)}_{\oplus} d\log(1+i_0) - \frac{\zeta_z - 1}{z}\left(\frac{1}{\alpha_\lambda + \frac{\sigma(\zeta_z-1)}{z}}\right) d\log(1+i_0), \tag{106}$$

i.e., output decreases on impact, and

$$d\log \mathcal{M}_1 = -\left(\frac{\alpha_m - \alpha_\lambda}{\alpha_\lambda + \frac{\sigma(\zeta_z-1)}{z}}\right) d\log(1+i_0), \tag{107}$$

i.e., allocative efficiency decreases the following period.

Equilibrium is guaranteed to exist if the following conditions are satisfied: (1) $\sigma > 1$, and (2) $\alpha_m > \frac{\sigma}{\sigma-1}\alpha_\lambda$. This can be shown from entrant's innovation equation 67:

$$-\left[(\sigma-1)\alpha_m - \sigma\alpha_\lambda\right]\frac{1}{\alpha_\lambda + \frac{\sigma(\zeta_z-1)}{z}} d\log(1+i_0) = (\sigma-1)d\log\mathcal{M}_2 + (\sigma-1)\alpha_g dz_1 + \frac{\zeta_z - 1}{z}dz_1. \tag{108}$$

Since production labor $L_{pt} = L - \delta_z z^{\zeta_z}$ is fixed from period 1 onward, dz_1 can at most be zero. Since $\sigma > 1$, it follows that $d\log\mathcal{M}_2 < 0$, such that the RHS is strictly negative. This implies equilibrium condition $\alpha_m > \frac{\sigma}{\sigma-1}\alpha_\lambda$.

Proof of Corollary 3. Following a reduction in entry rate, there exits two counteracting forces - a level effect due to the widening of the labor wedge (production labor's share of income decreases as average level of markups increases) and a dispersion effect for the distribution of markups - such that the misallocation effect of monetary policy is muted. In

[13] It can be shown that if $\sigma = 1$, then $d\log y_0 = 0$, and if $\sigma < 1$, then $d\log y_0 > 0$. That is, aggregate output is unchanged or goes up on impact of a contractionary monetary policy shock.

this special case we study, at time 1:

$$\begin{aligned} d\log \mathcal{M}_1 &= dm_1 - d\log \Lambda_1 \\ &= (\alpha_m - \alpha_\lambda) dz_0, \end{aligned}$$

where $dz_0 = -\left(\frac{1}{\alpha_\lambda + \frac{\sigma(\zeta z - 1)}{z}}\right) d\log(1+i_0)$, and that α_m measures the dispersion effect and α_λ corresponds to the strength of the level effect:

$$\begin{aligned} \alpha_m &= -(\ln \lambda)\left\{[1-(1-I)v(1)] - \frac{1}{1-z}\left[\frac{1+\vartheta_I}{\vartheta_I} - v(1)\right]\right\}, \\ \alpha_\lambda &= \frac{1}{\Lambda}\left\{\lambda^{-1}[(1-(1-I)v(1)] - \frac{1}{1-z}[\Lambda - \lambda^{-1}v(1)]\right\}. \end{aligned} \tag{109}$$

It can be shown that $\alpha_m > 0, \alpha_\lambda > 0$, thus, both real wage, dm_1, and the production labor's share of income, $d\log \Lambda_1$, are increasing in dz_0, such that on net, the misallocation effect is attenuated.

More generally, at time $t+1$, a contractionary monetary policy shock leads to a reduction in entry rate, $dz_t < 0$. Both dm_{t+1} and $d\log \Lambda_{t+1}$ would fall, and the net effect on misallocation is determined by the difference between the two coefficients, α_m and α_λ:

$$d\log \mathcal{M}_{t+1} = dm_{t+1} - d\log \Lambda_{t+1},$$

where $dm_{t+1} = \alpha_m dz_t + (1-z)dm_t$, and $d\log \Lambda_{t+1} = \alpha_\lambda dz_t + (1-z)\left(1 - I + \frac{I}{\lambda}\right) d\log \Lambda_t$.

B.4 Proof of Proposition 6

The state-dependent misallocation effect of monetary policy is such that: in an economy with low steady state level of average markups (high production labor's share of income), the misallocation effect of monetary policy is greater; in an economy with low steady state density of firms with unitary markups, the misallocation effect of monetary policy is greater.

The elasticity of the misallocation measure with respect to entrant's innovation rate is:

$$\begin{aligned}\alpha_m - \alpha_\lambda &= \left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)[1-(1-I)\upsilon(1)] + \left((\ln\lambda)\mathcal{S}_I^\star + \frac{\mathcal{S}_\lambda^\star}{\Lambda}\right)\left(\frac{1}{1-z}\right)\\ &= \left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)[1-(1-I)\upsilon(1)] + \left\{\ln\lambda\left[\frac{1+\vartheta_I}{\vartheta_I} - \upsilon(1)\right] + 1 - \frac{\lambda^{-1}}{\Lambda}\upsilon(1)\right\}\left(\frac{1}{1-z}\right)\end{aligned}$$

It follows that the elasticity is increasing in the production labor's share of income in the steady state:

$$\frac{\partial(\alpha_m-\alpha_\lambda)}{\partial\Lambda} = \frac{\lambda^{-1}}{\Lambda^2}[1-(1-I)\upsilon(1)] + \frac{\lambda^{-1}}{\Lambda^2}\upsilon(1)\left(\frac{1}{1-z}\right) > 0,$$

and decreasing in the steady state density of firms with unitary markups:

$$\frac{\partial(\alpha_m-\alpha_\lambda)}{\partial\upsilon(1)} = \left(-\ln\lambda - \frac{\lambda^{-1}}{\Lambda}\right)\left(\frac{z}{1-z} + I\right) < 0$$

In steady state, it can be easily shown by taking partial derivatives that both Λ and $\upsilon(1)$ are increasing in entrant's innovation rate z and decreasing in incumbent's innovation rate I. The increases in Λ and $\upsilon(1)$ (corresponding to a higher rate of entrant's innovation rate) have counteracting effects on the elasticity of the misallocation measure. In other words, whether the misallocation effect following a one-time shock to entrant's innovation is stronger in an economy with higher entrant's innovation rate is ambiguous.

C Theoretical Extensions

C.1 Incorporating value functions

In the baseline model, we assume that entrants could only earn profits for one period, we relax the assumption and consider instead that outside entrepreneurs' maximization problems takes the following form:

$$\max_{z_t}\{D_{t,t+1}z_tV_{t+1}(1)-\delta_z z_t^{\zeta_z}W_t\},\zeta_z>1$$

where $V_{t+1}(1)$ is incumbent's value function with unitary quality gap, and outside entrepreneur's first order condition is:

$$\mathbf{E}_t[D_{t,t+1}V_{t+1}(1)]=\delta_z\zeta_z z_t^{\zeta_z-1}W_t. \tag{110}$$

As for the incumbents, their value function can be written as

$$V_t(\Delta)=(1-\lambda^{-\Delta})P_tY_t+\mathbf{E}_t\{D_{t,t+1}(1-z_t)[IV_{t+1}(\Delta+1)+(1-I)V_{t+1}(\Delta)]\},$$

where $D_{t,t+1}$ is the stochastic discount factor, and I is the exogenous probability of successful own innovation. The value function of the firm consists of two parts: (i) the profits generated from existing quality gap Δ, (ii) the expected profits for the next period, conditioning on surviving creative destruction this period and increasing the existing quality gap by one step, this happens at the exogenous probability of own innovation I.

Stationarization of value function. Notice that the value functions are non-stationary, we normalize them as follows:

$$\tilde{V}_t(\Delta)\equiv\frac{V_t(\Delta)}{P_tQ_t}=(1-\lambda^{-\Delta})y_t+\mathbf{E}_t\left\{\beta\frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}}g_{t+1}^{1-\sigma}(1-z_t)[I\tilde{V}_{t+1}(\Delta+1)+(1-I)\tilde{V}_{t+1}(\Delta)]\right\},$$

and outside entrepreneur's first order condition can be normalized as:

$$\beta\mathbf{E}_t\left[\frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}}g_{t+1}^{1-\sigma}\tilde{V}_{t+1}(1)\right]=\delta_z\zeta_z w_t z_t^{\zeta_z-1}, \tag{111}$$

where

$$\tilde{V}_t(1) = (1-\lambda^{-1})y_t + \mathbf{E}_t\left\{\beta\frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}}g_{t+1}^{1-\sigma}(1-z_t)[Id\tilde{V}_{t+1}(1) + \tilde{V}_{t+1}(1)]\right\},$$

$$d\tilde{V}_t(1) \equiv \tilde{V}_t(2) - \tilde{V}_t(1) = \lambda^{-1}(1-\lambda^{-1})y_t + \mathbf{E}_t\left\{\beta\frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}}g_{t+1}^{1-\sigma}(1-z_t)\left[I\left(\frac{\lambda^{-1}-\lambda^{-2}}{1-\lambda^{-1}}\right) + (1-I)\right]d\tilde{V}_{t+1}(1)\right\}.$$

Steady states. In steady states, $\tilde{V}_t(\Delta) = \tilde{V}_{t+1}(\Delta) = \tilde{V}(\Delta)\forall\Delta$.

Conjecture that the value function takes the following form in the steady state

$$\tilde{V}(\Delta) = \alpha - \kappa\lambda^{-\Delta},$$

hence, it follows that $d\tilde{V}(\Delta) \equiv \tilde{V}(\Delta+1) - \tilde{V}(\Delta) = (1-\lambda^{-1})\,\kappa\lambda^{-\Delta}$.

Collecting terms, it can be shown at

$$\alpha = \frac{y}{1-\beta g^{1-\sigma}(1-z)},$$
$$\kappa = \frac{y}{1-\beta g^{1-\sigma}(1-z)[1-I(1-\lambda^{-1})]}.$$

Hence, the value function in balanced growth path (BGP) is given by

$$\tilde{V}(\Delta) = \frac{y}{1-\beta g^{1-\sigma}(1-z)} - \frac{y}{1-\beta g^{1-\sigma}(1-z)[1-I(1-\lambda^{-1})]}\lambda^{-\Delta},$$

and outside entrepreneur's innovation decision in BGP is

$$\beta g^{1-\sigma}\left[\frac{1}{1-\beta g^{1-\sigma}(1-z)} - \frac{\lambda^{-1}}{1-\beta g^{1-\sigma}(1-z)[1-I(1-\lambda^{-1})]}\right]y = \delta_z\zeta_z\exp(m)z^{\zeta_z-1}. \quad (112)$$

As $\bar{N} \to \infty$, the balanced growth path (BGP) for the economy are characterized by constant values for technical efficiency growth rate g, probability of successful entry z, normalized output y, price inflation Π, employment L, value functions $\tilde{V}(1)$ and $d\tilde{V}(1)$ satisfying:

$$g^{\sigma} = \frac{\beta(1+i_{ss})}{\Pi} \tag{113}$$

$$\beta g^{1-\sigma}\tilde{V}(1) = \delta_z \zeta_z \left(\lambda^{-\frac{1+\vartheta_I}{\vartheta_I}}\right) z^{\zeta_z - 1} \tag{114}$$

$$\tilde{V}(1) = \left[\frac{1}{1-\beta g^{1-\sigma}(1-z)} - \frac{\lambda^{-1}}{1-\beta g^{1-\sigma}(1-z)[1-I(1-\lambda^{-1})]}\right] y \tag{115}$$

$$d\tilde{V}(1) = \frac{(1-\lambda^{-1})\,\lambda^{-1}}{1-\beta g^{1-\sigma}(1-z)[1-I(1-\lambda^{-1})]} y \tag{116}$$

$$\Pi = \frac{\pi^w}{g} \tag{117}$$

$$g = (z + I - zI)(\lambda - 1) + 1 \tag{118}$$

$$y = (L - \delta_z z^{\zeta_z})\mathcal{M}, \text{ where} \tag{119}$$

$$\mathcal{M} = \frac{\exp(m)}{\Lambda} = \left(\lambda^{-\frac{1+\vartheta_I}{\vartheta_I}}\right)\left(\frac{\lambda(1+\vartheta_I)-1}{\vartheta_I}\right), \vartheta_I = \frac{z}{(1-z)I}$$

where i_{ss} represents the nominal interest rate alone the BGP. The eight equations are (i) household's Euler equation, (ii) outside entrepreneur's innovation decision, (iii) and (iv) value functions for unitary quality gap, (v) wage rigidity equation, (vi) growth rate of technical efficiency, and (vii) resource constraint.

Log-linearization. We log-linearize the normalized competitive equilibrium around its steady state and define the log-linearized equilibrium of the economy as follows:

Definition. As $\bar{N} \to \infty$, the log-linearized competitive equilibrium of the economy is defined as a sequence of variables $\{d\log y_t, d\log g_{t+1}, dz_t, di_t, dL_t, d\log \tilde{V}_t(1), d\log d\tilde{V}_t(1), d\log \Pi_t, dm_{t+1}, d\log \mathcal{M}_{t+1}\}$ that satisfy the following 10 equations, for a given sequence of exogenous shocks $\{d\log \epsilon_t^m\}$ and given initial values of state variables $\{dm_0, d\log \mathcal{M}_0\}$.

Consumption Euler equation

$$-\sigma d\log y_t = -\sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1} + d\log(1+i_t) - d\mathbf{E}_t \log \Pi_{t+1} \tag{120}$$

Entrant's innovation decision

$$[\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} - \sigma d\log g_{t+1}] + \mathbf{E}_t d\log \tilde{V}_{t+1}(1) + d\log g_{t+1} = dm_t + (\zeta_z - 1)d\log z_t \tag{121}$$

Value functions

$$d\log[\tilde{V}_t(1) - (1-\lambda^{-1})y_t] = [\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} + (1-\sigma)d\log g_{t+1}] + d\log(1-z_t) \tag{122}$$

$$+ \mathbf{E}_t d\log[Id\tilde{V}_{t+1}(1) + \tilde{V}_{t+1}(1)], \text{ where}$$

$$d\log[\tilde{V}_t(1) - (1-\lambda^{-1})y_t] = \frac{\tilde{V}(1)d\log\tilde{V}_t(1) - (1-\lambda^{-1})yd\log y_t}{\tilde{V}(1) - (1-\lambda^{-1})y}$$

$$\mathbf{E}_t d\log[Id\tilde{V}_{t+1}(1) + \tilde{V}_{t+1}(1)] = \frac{Id\tilde{V}(1)d\log d\tilde{V}_{t+1}(1) + \tilde{V}(1)d\log\tilde{V}_{t+1}(1)}{Id\tilde{V}(1) + \tilde{V}(1)}$$

$$d\log[d\tilde{V}_t(1) - \lambda^{-1}(1-\lambda^{-1})y_t] = [\sigma d\log y_t - \sigma \mathbf{E}_t d\log y_{t+1} + (1-\sigma)d\log g_{t+1}] + d\log(1-z_t) \tag{123}$$

$$+ \mathbf{E}_t \left[I\left(\frac{\lambda^{-1} - \lambda^{-2}}{1-\lambda^{-1}}\right) + (1-I)\right] d\log d\tilde{V}_{t+1}(1), \text{ where}$$

$$d\log[d\tilde{V}_t(1) - \lambda^{-1}(1-\lambda^{-1})y_t] = \frac{d\tilde{V}(1)d\log\tilde{V}_t(1) - \lambda^{-1}(1-\lambda^{-1})yd\log y_t}{d\tilde{V}(1) - \lambda^{-1}(1-\lambda^{-1})y}$$

Wage rigidity equation

$$d\log\Pi_t = -d\log g_t + dm_{t-1} - dm_t \tag{124}$$

Technical efficiency growth equation

$$d\log g_{t+1} = \underbrace{\frac{(\lambda-1)(1-I)}{(z+I(1-z))(\lambda-1)+1}}_{\equiv\alpha_g} dz_t = \alpha_g dz_t \tag{125}$$

Measure of aggregate misallocation

$$dm_{t+1} = \alpha_m dz_t + (1-z)dm_t \tag{126}$$

$$d\log\mathcal{M}_{t+1} = (\alpha_m - \alpha_\Lambda)dz_t + (1-z)\left[1 - I\left(1-\frac{1}{\lambda}\right)\right]d\log\mathcal{M}_t + (1-z)I\left(1-\frac{1}{\lambda}\right)dm_t, \text{ where} \tag{127}$$

$$\begin{aligned}
&\kappa_A \equiv 1-(1-I)\frac{\vartheta_I}{1+\vartheta_I},\ \kappa_B \equiv \frac{1}{1-z}, \vartheta_I \equiv \frac{z}{(1-z)I} \\
&\mathcal{S}_I^\star \equiv \sum_{\Delta=2}^{\infty}\Delta\upsilon(\Delta) = \frac{1+\vartheta_I}{\vartheta_I} - \frac{\vartheta_I}{1+\vartheta_I}, \\
&\mathcal{S}_\lambda^\star \equiv \sum_{\Delta=2}^{\infty}\lambda^{-\Delta}\upsilon(\Delta) = \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1} - \lambda^{-1}\left(\frac{\vartheta_I}{1+\vartheta_I}\right), \\
&\alpha_m \equiv -(\ln\lambda)[\kappa_A - \kappa_B\mathcal{S}_I^\star], \\
&\alpha_\Lambda \equiv \frac{1}{\Lambda}[\lambda^{-1}\kappa_A - \kappa_B\mathcal{S}_\lambda^\star], \Lambda \equiv \frac{\vartheta_I}{\lambda(1+\vartheta_I)-1}.
\end{aligned}$$

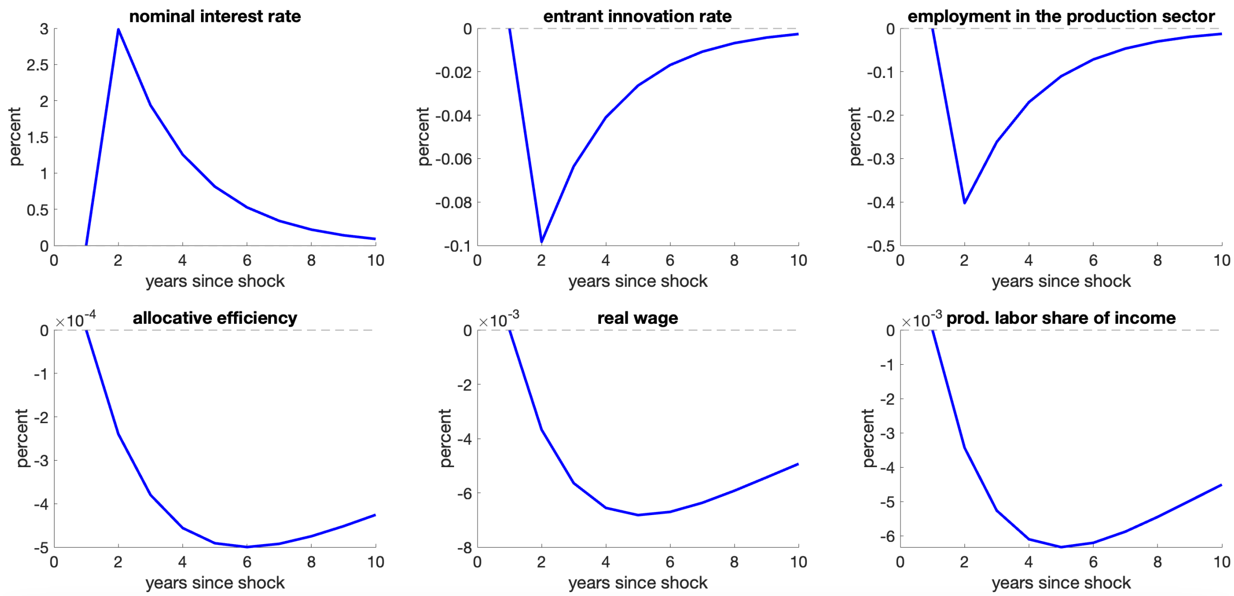
Resource constraint

$$d\log y_t = d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log\mathcal{M}_t \tag{128}$$

Monetary policy rule

$$d\log(1+i_t) = \phi d\log(L_t - \delta_z z_t^{\zeta_z}) + d\log\epsilon_t^m, \phi > 0, i_t > 0 \tag{129}$$

Figure 5: Model-based impulse response to a one-time monetary policy shock



Notes: The figure shows the impulse responses of the nominal interest rate, entrant innovation rate, production labor, the measure of misallocation, real wage, and the production labor's share of income to a one-time contractionary monetary policy shock.

Modelled dynamic responses. Figure 5 graphically illustrates the dynamic responses of nominal interest rate, entrant innovation rate, employment level in the production sector, allocative efficiency, real wage, and the production labor's share of income to a contractionary monetary policy shock.

C.2 Endogenizing incumbent innovation

In the baseline model, we assume that rates of incumbent innovation I are exogenous and thus unaffected by business cycle fluctuations. We now relax the assumption and allow incumbents to choose their rates of own innovation $I_t(\Delta)$. Incumbents' maximization problems can be written as:

$$\max_{I_t(\Delta)} \left\{ (1-\lambda^{-\Delta})P_tY_t + \mathbf{E}_t D_{t,t+1}(1-z_t)[I_tV_{t+1}(\Delta+1) + (1-I_t)V_{t+1}(\Delta)] - c(I_t,\Delta)W_t \right\},$$

where $c(I_t, \Delta)$ denotes costs of own innovation in units of R&D labor hired. To make progress in solving the more general case, we follow Peters (2020) to assume a particular functional form for the cost function such that

$$c(I_t, \Delta) = \lambda^{-\Delta}\delta_I I_t(\Delta)^{\zeta_I}, \zeta_I > 1$$

where δ_I determines the efficiency of own innovation, and $\zeta_I > 1$ ensures convexity of the cost function such that there exists a unique solution.

First order condition implies that:

$$\mathbf{E}_t D_{t,t+1}(1 - z_t)\left[V_{t+1}(\Delta + 1) - V_{t+1}(\Delta)\right] = \zeta_I \lambda^{-\Delta}\delta_I I_t^*(\Delta)^{\zeta_I - 1} W_t \tag{130}$$

Labor market clearing. The labor market clearing condition can be rewritten as:

$$L_t = L_{pt} + L_{rt} = L_{pt} + \delta_z z_t^{\zeta_z} + \int_0^1 \lambda^{-\Delta_i}\delta_I I_{it}^{\zeta_I} di = L_{pt} + \delta_z z_t^{\zeta_z} + \sum_{\Delta=1}^{\infty} \lambda^{-\Delta}\delta_I I_t(\Delta)^{\zeta_I} v_t(\Delta). \tag{131}$$

Stationarization of value function. Likewise, we rewrite the normalized value function $\tilde{V}_t(\Delta) \equiv \frac{V_t(\Delta)}{P_t Q_t}$ as:

$$\tilde{V}_t(\Delta) = (1-\lambda^{-\Delta})y_t - \lambda^{-\Delta}\delta_I I_t^*(\Delta)^{\zeta_I} w_t + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1 - z_t)[I_t^*(\Delta)\tilde{V}_{t+1}(\Delta + 1) + (1 - I_t^*(\Delta))\tilde{V}_{t+1}(\Delta)]\right\},$$

where first order condition implies that the optimal rate of own innovation is:

$$I_t^*(\Delta) = \left[\frac{\beta \mathbf{E}_t \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1 - z_t)[\tilde{V}_{t+1}(\Delta + 1) - \tilde{V}_{t+1}(\Delta)]}{\zeta_I \lambda^{-\Delta}\delta_I w_t}\right]^{\frac{1}{\zeta_I - 1}}. \tag{132}$$

Guess and verify that $I_t^*(\Delta) = I_t^* \forall \Delta$: let $d\tilde{V}_t(\Delta) \equiv \tilde{V}_{t+1}(\Delta + 1) - \tilde{V}_{t+1}(\Delta)$, it follows that:

$$d\tilde{V}_t(\Delta) = \lambda^{-\Delta}(1-\lambda^{-1})(y_t + \delta_I I_t^{*\zeta_I} w_t) + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1 - z_t)[I_t^* d\tilde{V}_{t+1}(\Delta + 1) + (1 - I_t^*)d\tilde{V}_{t+1}(\Delta)]\right\},$$

similarly,

$$d\tilde{V}_t(\Delta+1) = \lambda^{-\Delta}(\lambda^{-1}-\lambda^{-2})(y_t+\delta_I I_t^{*\zeta_I} w_t) + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1-z_t)[I_t^* d\tilde{V}_{t+1}(\Delta+2) \right.$$
$$\left. +(1-I_t^*)d\tilde{V}_{t+1}(\Delta+1)]\right\}.$$

Guess that:

$$d\tilde{V}_t(\Delta+1) = \frac{\lambda^{-1}-\lambda^{-2}}{1-\lambda^{-1}} d\tilde{V}_t(\Delta),$$

It can be easily verified that the equality holds such that:

$$d\tilde{V}_t(\Delta) = \lambda^{-\Delta}(1-\lambda^{-1})(y_t+\delta_I I_t^{*\zeta_I} w_t) + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1-z_t)\left[I_t^*\left(\frac{\lambda^{-1}-\lambda^{-2}}{1-\lambda^{-1}}\right)+(1-I_t^*)\right] d\tilde{V}_{t+1}(\Delta)\right\}. \tag{133}$$

Since optimal own innovation $I_t^*(\Delta)$ can be written as:

$$I_t^*(\Delta) = \left[\frac{\beta \mathbf{E}_t \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1-z_t) d\tilde{V}_{t+1}(\Delta)}{\zeta_I \lambda^{-\Delta}\delta_I w_t}\right]^{\frac{1}{\zeta_I - 1}}, \tag{134}$$

and $\lambda^{-\Delta}$ in both numerator and denominator should cancel out each other, we have $I_t^*(\Delta) = I_t^* \forall \Delta$.

We can thus rewrite first order condition of entrant's optimization problem:

$$\beta \mathbf{E}_t\left[\frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}\tilde{V}_{t+1}(1)\right] = \zeta_z \delta_z z_t^{\zeta_z - 1} w_t, \tag{135}$$

where

$$\tilde{V}_t(1) = (1-\lambda^{-1})y_t - \lambda^{-1}\delta_I I_t^{*\zeta_I} w_t + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1-z_t)[I_t^* d\tilde{V}_{t+1}(1) + \tilde{V}_{t+1}(1)]\right\},$$
$$d\tilde{V}_t(1) = \lambda^{-1}(1-\lambda^{-1})(y_t+\delta_I I_t^{*\zeta_I} w_t) + \mathbf{E}_t\left\{\beta \frac{y_{t+1}^{-\sigma}}{y_t^{-\sigma}} g_{t+1}^{1-\sigma}(1-z_t)\left[I_t^*\left(\frac{\lambda^{-1}-\lambda^{-2}}{1-\lambda^{-1}}\right)+(1-I_t^*)\right] d\tilde{V}_{t+1}(1)\right\}.$$

Steady states. In steady states, again conjecture that $\tilde{V}(\Delta) = \alpha - \kappa\lambda^{-\Delta}$, then the first order condition is:

$$I^*(\Delta) = \left[\frac{\beta g^{1-\sigma}(1-z)(1-\lambda^{-1})\kappa}{\zeta_I \delta_I w}\right]^{\frac{1}{\zeta_I - 1}}.$$

Suppose that $\frac{\kappa}{w}$ = constant (which we verify later), then we have $I^*(\Delta) = I^* \forall \Delta$.

From the value function, we can show that:

$$\alpha = \frac{y}{1-\beta g^{1-\sigma}(1-z)},$$
$$\kappa = \frac{y + \delta_I I^{*\zeta_I} w}{1-\beta g^{1-\sigma}(1-z)[1-I^*(1-\lambda^{-1})]}.$$

Since $\frac{\kappa}{w} = \frac{y/w + \delta_I I^{*\zeta_I}}{1-\beta g^{1-\sigma}(1-z)[1-I^*(1-\lambda^{-1})]}$, and $\frac{y}{w} = \frac{L_p}{\Lambda^*}$ in BGP, then $\frac{\kappa}{w}$ = constant.

Outside entrepreneur's innovation decision in BGP can be rewritten as

$$\beta g^{1-\sigma}\left[\frac{1}{1-\beta g^{1-\sigma}(1-z)} - \frac{(1+\delta_I I^{*\zeta_I}\frac{\Lambda^*}{L_p})\lambda^{-1}}{1-\beta g^{1-\sigma}(1-z)[1-I^*(1-\lambda^{-1})]}\right] y = \delta_z \exp\{m\}, \tag{136}$$

where $L_p = L - L_r = L - \delta_z z^{\zeta_z} - \delta_I I^{*\zeta_I}\Lambda^*$.

Calibration. We calibrate the extended models with parameters provided in Table 2. Incumbent's own innovation in steady state is $I^* = 0.0898$, while we keep the steady state rate of entrant's innovation at $z^* = 0.24$ as in the baseline model. Parameters for the productivity and curvature of incumbent's and entrant's R&D efforts are calibrated to match these moments.

Table 2: Parameters

	Value	Source/Target
Elas. intertemporal substitution	$1/\sigma = 0.5$	Standard value
Discount factor	$\beta = 0.96$	Standard value
Wage inflation at steady state	$\pi^w = 1.02$	
Innovation step size	$\lambda = 1.0649$	Akcigit and Kerr (2018)
Parameter for entrant's R&D cost	$\delta_z = 0.0832$	FOC for entrant innovation
Curvature of entrant's R&D cost	$\zeta_z = 1.5$	
Persistence of monetary shock	$\rho^m = 0.9$	Standard value
Parameter for incumbent's R&D cost	$\delta_I = 0.9$	
Curvature of incumbent's R&D cost	$\zeta_I = 2$	

Notes: Model period corresponds to a year.

Modelled dynamic responses. Figure 6 graphically illustrates the dynamic responses of nominal interest rate, entrant innovation rate, allocative efficiencyïŒ incumbent's own innovation rate, growth rate, employment level in the production sector, real wage and production labor's share of income to a contractionary monetary policy shock.

C.3 CES demand function

In the baseline model, we assume that household's preferences take a particular form of unitary elasticity, now we relax the assumption to study the case with a general CES demand structure that takes the form[14]

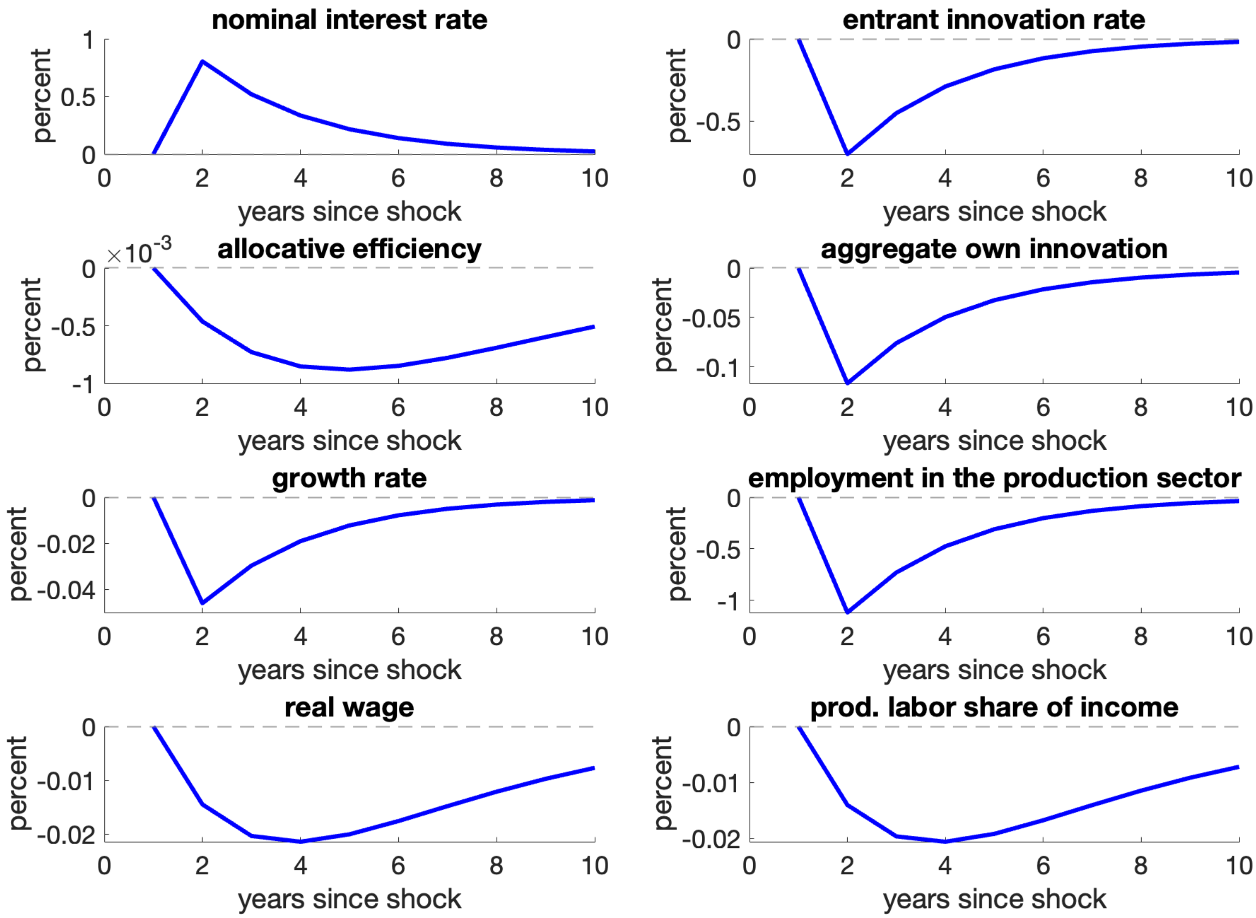
$$Y_t = \left(\int_0^1 x_{it}^{\frac{\eta-1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}}, \eta > 1. \tag{137}$$

With CES demand, firms with quality gaps that are sufficiently large will find it optimal to charge the unconstrained markup $\frac{\eta}{\eta-1}$ in stead of the limit price. It follows that

$$\mu_{it}(\Delta) = \min \left\{ \frac{\eta}{\eta - 1}, \lambda^{\Delta} \right\}. \tag{138}$$

[14]In calibration, we set the elasticity of substitution $\eta = 6$, following convention.

Figure 6: Model-based impulse response to a one-time monetary policy shock



Notes: The figure shows the impulse responses of the nominal interest rate, entrant innovation rate, incumbent's own innovation rate, production labor, TFP growth rate, the measure of misallocation, real wage, and the production labor's share of income to a one-time contractionary monetary policy shock.

Then we derive an expression for real wages $\tilde{W}_t$ to be used later:

$$\tilde{W}_t \equiv \frac{W_t}{P_t} = \left[\int_0^1 \mu_{it}^{1-\eta} q_{it}^{\eta-1} di \right]^{\frac{1}{\eta-1}}, \tag{139}$$

where in any period t, the joint distribution of quality gaps (which determine markups) and quality, $F_t(\Delta, q)$ would determine real wages at t.

Likewise, aggregate output can be rewritten as

$$Y_t = Q_t L_{pt} \mathcal{M}_t, \text{ where } \mathcal{M}_t \equiv \frac{\left[\int_0^1 \mu_{it}^{1-\eta} \left(\frac{q_{it}}{Q_t}\right)^{\eta-1} di\right]^{\frac{\eta}{\eta-1}}}{\int_0^1 \mu_{it}^{-\eta} \left(\frac{q_{it}}{Q_t}\right)^{\eta-1} di}, \tag{140}$$

where $Q_t \equiv (\int q_{it}^{\eta-1} di)^{\frac{1}{\eta-1}}$ is a measure of technical efficiency, while $\mathcal{M}_t$ is a measure of allocative efficiency in the economy.

It remains to show that the equilibrium profit function of a variety with markup μ_{it} and quality q_{it} is given as

$$\Phi_{it}(\Delta, q_{it}) = (1 - \mu_{it}^{-1}) P_{it} x_{it} = (1 - \mu_{it}^{-1}) \mu_{it}^{1-\eta} q_{it}^{\eta-1} \tilde{W}_t^{1-\eta} P_t Y_t. \tag{141}$$

The entrant's optimization problem implies:

$$\beta \mathbf{E}_t \left[\frac{Y_{t+1}^{-\sigma}}{Y_t^{-\sigma}} \frac{P_t}{P_{t+1}} \Phi_{i,t+1}(1, q_{i,t+1})\right] = \zeta_z \delta_z z_t^{\zeta_z - 1} W_t, \tag{142}$$

where $\Phi_{i,t+1}(1, q_{i,t+1}) = \left(1 - \frac{1}{\lambda}\right) \lambda^{1-\eta} \mathbf{E}_i \left[\frac{q_{i,t+1}}{\tilde{W}_{t+1}}\right]^{\eta-1} P_{t+1} Y_{t+1}$.

Similar to baseline model, we assume that incumbent's own innovation occurs at an exogenous probability I, this simplifies the analysis as now the marginal distributions of quality gaps Δ and quality q are *independent* such that: $F_t(\Delta, q) = F_{\Delta,t}(\Delta) F_{q,t}(q)$. Since markups only depend on Δ, it follows that the distribution of markups is independent of the distribution of q, and that the evolution of quality gaps are governed by the same equations 46 - 48. It remains to show that z_t is independent of q.

Equation 142 can be normalized to:

$$\beta \mathbf{E}_t \left[\frac{c_{t+1}^{1-\sigma} g_{t+1}^{1-\sigma}}{c_t^{-\sigma}} \left(1 - \frac{1}{\lambda}\right) \lambda^{1-\eta} \mathbf{E}_i \left[q_{i,t+1}\right]^{\eta-1} \tilde{W}_{t+1}^{1-\eta}\right] = \delta_z w_t, \tag{143}$$

where $w_t \equiv \frac{\tilde{W}_t}{Q_t} = \frac{\left[\int_0^1 \mu_{it}^{1-\eta} q_{it}^{\eta-1} di\right]^{\frac{1}{\eta-1}}}{(\int q_{it}^{\eta-1} di)^{\frac{1}{\eta-1}}}$ is the normalized real wage.

Making use of the fact that $\int_q \left(\frac{q}{Q_t}\right)^{\eta-1} dF_t(q) = 1$, entrant's optimal innovation decision can be rearranged as

$$\beta \mathbf{E}_t \left[\frac{c_{t+1}^{1-\sigma} g_{t+1}^{1-\sigma}}{c_t^{-\sigma}} \left(1 - \frac{1}{\lambda}\right) \lambda^{1-\eta} w_{t+1}^{1-\eta} \right] = \zeta_z \delta_z z_t^{\zeta_z - 1} w_t, \tag{144}$$

and that z_t is independent of quality q and in the special case $\eta = 1$, it collapses to the baseline case we analyze in the main text. Since z_t alone determines the evolution of quality gaps, it follows that the measure of aggregate misallocation can be rewritten as:

$$\mathcal{M}_t \equiv \frac{\left[\int \mu_{it}^{1-\eta} di\right]^{\frac{\eta}{\eta-1}}}{\int \mu_{it}^{-\eta} di} = \frac{\left[\sum_{\Delta=1}^{\bar{N}} \mu(\Delta)^{1-\eta} \upsilon_t(\Delta)\right]^{\frac{\eta}{\eta-1}}}{\sum_{\Delta=1}^{\bar{N}} \mu(\Delta)^{-\eta} \upsilon_t(\Delta)}, \tag{145}$$

similarly, normalized real wage $w_t = \left[\sum_{\Delta=1}^{\bar{N}} \mu(\Delta)^{1-\eta} \upsilon_t(\Delta)\right]^{\frac{1}{\eta-1}}$.

Finally, there exits a threshold value for quality gap Δ such that:

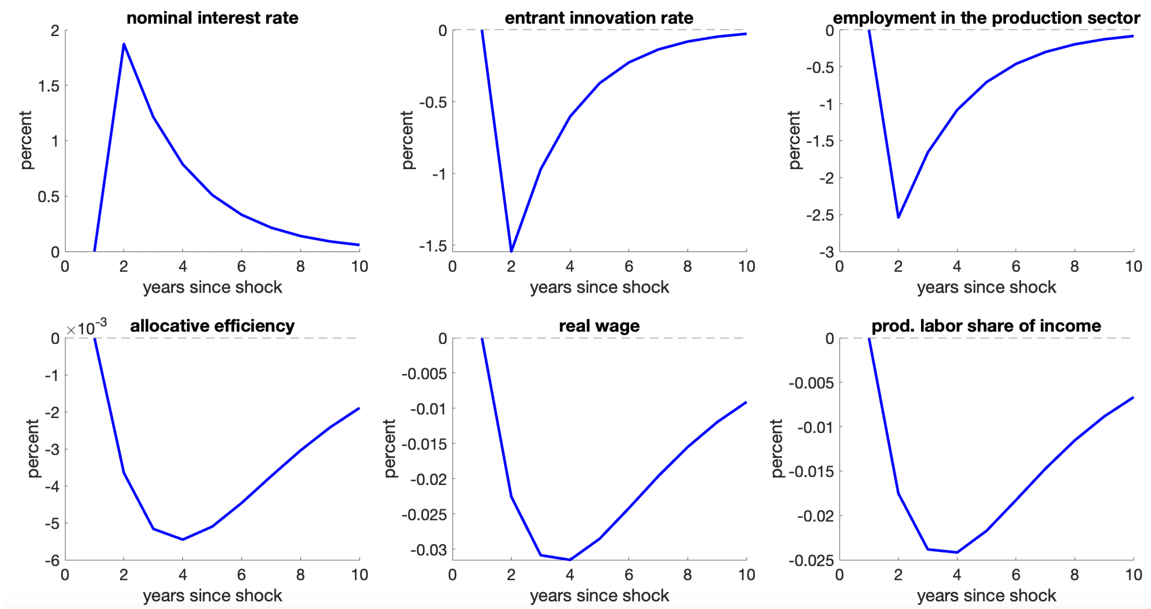
$$\mu(\Delta) = \begin{cases} \lambda^{\Delta}, & \text{if } \Delta < \frac{\ln\left(\frac{\eta}{\eta-1}\right)}{\ln \lambda} \\ \frac{\eta}{\eta-1} & \text{if } \Delta \geq \frac{\ln\left(\frac{\eta}{\eta-1}\right)}{\ln \lambda} \end{cases}. \tag{146}$$

Steady states. In steady state, $\upsilon(\Delta)$ would be exactly the same as in the baseline case, and they pin down steady state measure of aggregate misallocation $\mathcal{M}$ and normalized real wage w. The entrant's innovation decision is given by

$$\beta g^{1-\sigma}(1 - \lambda^{-1})\lambda^{1-\eta}(L - \delta_z z)\mathcal{M} = \zeta_z \delta_z z^{\zeta_z - 1} w^{\eta}. \tag{147}$$

Modelled dynamic responses. Figure 7 graphically illustrates the dynamic responses of nominal interest rate, entrant innovation rate, employment level in the production sector, and allocative efficiency to a contractionary monetary policy shock.

Figure 7: Model-based impulse response to a one-time monetary policy shock



Notes: The figure shows the impulse responses of the nominal interest rate, entrant innovation rate, production labor, the measure of misallocation, real wage, and the production labor's share of income to a one-time contractionary monetary policy shock.