Redistribution and Reallocation: Monetary Policy with Heterogeneous Entrepreneurs*

Nicholas Hoffman
Job Market Paper
Tepper School of Business, Carnegie Mellon University
nchoffma@andrew.cmu.edu

This version: October 8, 2024

Abstract

I study the aggregate and distributional impacts of monetary policy in an economy with entrepreneurs who earn persistent, heterogeneous returns on investments in their private businesses. Using an analytically tractable model in which entrepreneurs face credit constraints, I demonstrate that a decrease in the monetary policy rate redistributes capital from less productive entrepreneurs, to more productive. As a result, acommodative monetary policy shocks increase both aggregate productivity and wealth inequality, as they do in the data. In an economy calibrated to match US data, redistribution amplifies the effect of expansionary policy on output by about 20%, and leads to 30% higher investment at the peak of the business cycle. In comparative statics exercises, I demonstrate that the aggregate effect of monetary expansions is hump-shaped in the degree of wealth inequality at the time of the policy change. When inequality is low, an increase in top wealth concentration leads to larger effects on output, whereas when wealth is highly concentrated at the top, further increases in inequality reduce the effectiveness of monetary policy.

Keywords: Macroeconomics, Monetary Economics, Entrepreneurs, Wealth Distribution

*This is my job market paper, and the first chapter of my PhD Thesis at Carnegie Mellon University. I am deeply indebted to Ali Shourideh, Laurence Ales, and Ariel Zetlin-Jones for invaluable guidance. I also extend my sincere gratitude to Andre Sztutman, Liyan Shi, Daniel Carroll, Eric Young, Kevin Mott, David Childers, and participants at the 2024 Midwest Macroeconomics Meetings and in the Carnegie Mellon Macroeconomics Reading Group for helpful discussions and feedback.

1 Introduction

This paper studies the role of heterogeneous entrepreneurs determining the effect of monetary policy on the aggregate economy. I embed nominal rigidities into model in which agents accumulate wealth which they may invest into a private business, which they operate with idiosyncratic and stochastic productivity. These entrepreneurs face financial constraints, in that they may issue nominal debt to finance their operations, up to a collateral limit. The cost of this debt, in nominal terms, is set by a central bank. Financial frictions give rise to a steady state that features a misallocation of capital. I characterize analytically the response of my economy to an unanticipated change in monetary policy, showing that a reduction in the nominal interest rate redistributes assets from entrepreneurs with lower skill, to those with higher skill. This redistributive channel *amplifies* the response of aggregate employment, output, and investment to the change in monetary policy by increasing aggregate (total factor) productivity. I further show how the extent of this amplification, and thus the effect of monetary policy, depends on the distribution of assets among entrepreneurs—i.e., the wealth distribution—at the time of the rate change. When inequality is low, and assets are evenly held across the distribution of skills, the economy acts as one with a representative producer, and increasing wealth concentration in the steady state increases the impact of a monetary expansion on aggregates, by allowing for larger increases in productivity and investment following the expansionary shock. By contrast, if steady state wealth inequality is high, then general equilibrium forces dampen the scope for redistribution, and further increases in inequality *reduce* the efficacy of monetary policy in stimulating output.

Empirical evidence points to a few effects of monetary expansions, both on aggregates and distributions, that are unaccounted for by both standard representative agent New Keynesian models (e.g. Galí, 2015) and recent heterogeneous-agent models (Kaplan et al., 2018, Auclert et al., 2020, Kekre and Lenel, 2022). First, accommodative monetary policy changes increase aggregate total factor productivity, as documented in Christiano et al. (2005) and Baqaee et al. (2021). Additionally, these expansionary changes in interest rate policy increase wealth inequality over the business cycle, as measured by top wealth shares or the Gini coefficient—see Feilich (2021) and Medlin (2023), and a review in Colciago et al. (2019). Finally, the data suggest that monetary policy shocks have larger effects on output in regions and time periods where wealth inequality is higher (Matusche and Wacks, 2023). These facts suggest a few features of the economy that have been largely, if not entirely, overlooked by the literature on monetary policy. First, they suggest that rather than being exogeneously given, productivity is a function of *allocations*, which are endogeneous to monetary policy. The positive effect of expansionary shocks on productivity also implies long-run distortions: if the allocation of capital were efficient, then there would be no scope for reallocations resulting from monetary easings to increase productivity. Furthermore, the data suggest a tight interplay between the weatlh distribution, and the effects of monetary policy, where each depends on the other. I argue that these observations from the data are all related: in an economy wher production is undertaken in large part by private businesses, and the operation of these businesses is constrained by the net worth of their owners, then aggregate output and productivity are functions of the wealth distribution. Furthermore, heterogeneity in entrepreneurs' wealth and returns imply that interest rate policy has differential effects across households. These differential effects give rise to a tight relationship between monetary policy and the wealth distribution: changes in interest rates affect the allocation of assets, and the overall effect on aggregates depends on how assets are distributed to begin with.

In the United States, private firms are central to economic activity and constrained by capital market frictions, and their owners are disproportionately wealthy. Private firms account for half of all capital investment and about two-thirds of private employment in the US (Asker et al., 2015), and earn more than half of aggregate net business income (DeBacker et al., 2023). The owners of these firms also constitute between 65 and 76 percent of wealthiest one percent of US households (Cagetti and De Nardi, 2006). Finally, there is evidence in the data that income from operating a business is persistent over time (DeBacker et al., 2023). This persistence has important implications for the evolution of entrepreneurial wealth: if returns are persistent then, over time, we can expect differences in returns to manifest in differences in wealth, with entrepreneurs of higher returns accumulating more wealth than those with lower returns.

In addition to their prominent role in driving aggregate economic activity, the actions of private firms are tightly linked to credit market conditions and the balance sheets of those who operate them. The investment, hiring, and production decisions of these firms depends on the ability of their owners to either self-finance their operations, or obtain external financing through financial markets. Private firms in the US finance about 80% of their investment using external funds, as compared to about 20% for publicly-owned firms (Shourideh and Zetlin-Jones, 2012). De Nardi et al. (2007) document that the median firm operated by an entrepreneur has a leverage ratio of 12-25%; its debts are equal to between one-eighth and one-fourth of its equity value. Furthermore, they report that although entrepreneurs are on average wealthier than worker households, they are equally likely to report being credit constrained, unable to take on as much borrowing as they would like. These facts suggest that both collateral constraints and the cost of credit are crucial determinants of entrepreneurial activity.

Intuitively, there are a number of channels, in both partial and general equilibrium, by which heterogeneous exposure to monetary shocks can redistribute wealth between entrepreneurs and creditors, and within entrepreneurs. The proclivity of entrepreneurs to take on debt is one channel: to the extent that these liabilities are denoted in *nominal* terms and not inflation-protected, the net worth of entrepreneurs is subject to inflation risk, and unanticipated changes in the price level can potentially redistribute wealth between entrepreneurs and their creditors. This channel There is a second channel which redistributes towards wealthy, high-productivity entrepreneurs, which I term the *aggregate demand* channel. Entrepreneurs' returns are a function not only of their idiosyncratic productivities, but aggregate prices as well: the price that they earn for their output, and the wage rate that they pay to their workers. In expansions, the former rises by more than the latter, thereby increasing the return that an entrepreneur earns on one *effective* (productivity-adjusted) unit of investment capital. Because higher-productivity entrepreneurs are more effective in using capital, this shock increases their returns by more, thus increasing the wealth held by high-productivity entrepreneurs relative to lower-productivity.

A crucial determinant of both the strength of these channels, and steady-state wealth inequality, is the persistence of entrepreneurs' return shocks. When shocks are weakly autocorrelated, there is little scope for redistribution to amplify the effects of monetary policy: in this scenario, high-productivity entrepreneurs today are unlikely to be high-productivity tomorrow. Thus, although these entrepreneurs will benefit from redistribution via the two channels above, their new wealth does not translate to increased productivity tomorrow. On the other end, when shocks are very persistent, steady state wealth inequality will be high, as high-productivity entrepreneurs accumulate wealth over time to expand their business and loosen their collateral constraint. However, at very high levels of inequality, there is similarly little scope for amplification of monetary expansions through redistribution. The subtle reason behind this is the key theoretical contribution of my paper. For an entrepreneur to increase her wealth share from one period to the next, she must earn an idiosyncratic return on her wealth not just in excess of the risk-free rate, but in excess of the aggregate return to capital. If she does, then her wealth grows faster than the capital stock, and so her share of total capital increases. At high levels of wealth concentration, entrepreneurs who actively invest in their business are those with high returns, and thus the aggregate return on capital is very closely aligned with their private returns, making it difficult for them to increase their share of wealth. This implies that the effect of a monetary shock on output and investment is hump-shaped in steady-state wealth inequality (equivalently, in the persistence of private returns): low at the extremes of little inequality and high inequality, and peaked in the middle. It also implies that an increase in wealth inequality, as has been observed in the US over the past forty years (e.g. Piketty and Saez, 2014), can render monetary policy *less* effective.

To more rigorously study these mechanisms, I develop an analytically tractable New Keynesian model with heterogeneous entrepreneurs who face collateral constraints and issue nominal debt in order to finance their operations. In my model, entrepreneurs accumulate wealth in order to consume and invest in their business, and employ hand-to-mouth workers in the operation of their firms. I show that in this framework, an unexpected monetary easing (modeled as a decrease in the nominal interest rate) redistributes wealth from low-productivity to high-productivity entrepreneurs. I demonstrate a tight link between the steady-state wealth distribution among entrepreneurs and the impact of a shock: wealth is redistributed towards more productive entrepreneurs following a shock if the underlying conditions are such that more productive entrepreneurs are wealthier to begin with. If wealth is concentrated among high-productivity entrepreneurs when the shock hits, then the effect of the shock will be to increase this concentration. My model admits an aggregate production function, with aggregate productivity being determined by the distribution of wealth among entrepreneurs. As a result of this dependence, an accommodative monetary shock increases aggregate productivity to the extent that it reallocates wealth towards higher-productivity entrepreneurs.

1.1 Related Literature

My paper bridges several strands in the literature. The first concerns the role of entrepreneurial and firm-level heterogeneity, in concert with financial market frictions, in determining aggregate activity and the response of the economy to policies. In these models, production is carried out by two or more firms who differ in their productivity.[1] The key result in these models is that in a world with heterogeneous establishments, financial market imperfections lead to output loss through *misallocation*: capital market frictions such as leverage constraints prevent the most productive firms from scaling up to their socially-efficient size, and thus some output is produced by less-productive firms. This literature finds that the impact of misallocation is empirically meaningful: in a calibrated model of firm heterogeneity, Restuccia and Rogerson (2008) find that policies which distort the efficient allocation of capital, shifting resources from productive firms to unproductive, can lead to losses in aggregate TFP of 30-50%. Conversely, Baqaee and Farhi (2020) estimate that half of aggregate TFP growth from 1997-2015 was due to reallocation of capital towards more productive firms. On the policy side, Boar and Midrigan (2019) achieve substantial welfare gains through a tax reform which redistributes capital towards larger firms. Crucially, their framework finds that such a policy is optimal even when households have unequal exposure to the profits from these firms: general equilibrium effects, such as an increase in the wage rate, outweigh the regressive nature of their policy. More closely related to my work is Baqaee et al. (2021), who augment a standard representative-agent New Keynesian model (as in, e.g., Galí, 2015) with firms of heterogeneous productivity. The mechanism at work here is the pass-through of marginal costs to prices: in response to a demand shock that raises marginal costs, high-productivity firms raise their prices by less than do low-productivity firms, leading to a shift in production towards efficient producers and a concomitant increase in overall TFP. My work complements theirs by demonstrating that, in addition to the reallocation of *labor*, monetary policy engenders a reallocation of

[1] See Hopenhayn (2014), for instance, for a survey.

capital between heterogeneous entrepreneurs. As a result, the wealth distribution is also a key determinant of the effect of these changes in interest rates.

The second strain of literature to which I contribute is the growing study of monetary policy in economies with household heterogeneity. The early papers in this literature focused on the role of *labor* income risk in altering the effects of a monetary shock relative to a representative-agent framework. McKay et al. (2016) and Kaplan et al. (2018) employed a model with heterogeneous agents in the style of Aiyagari (1994) to address inconsistencies between representative agent New Keynesian models and data, providing theoretical arguments that distributional concerns, precautionary savings motives, and household borrowing constraints are all relevant to a complete understanding the operation of monetary policy. These early "HANK" papers, and others, focused on the presence of borrowing-constrained agents, who have higher marginal propensities to consume out of current income shocks than do their unconstrained counterparts. This emphasis on the role of *poor* households has made tremendous strides in examining the role of heterogeneity in MPCs in the transmission of policy, but has left relatively under-explored the role of *wealthy* households in transmitting policy. My contribution is to fill in this gap, and to characterize the role played by rich entrepreneurs in translating a policy shock to aggregate output. I focus on heterogeneity in marginal propensities to *invest*, rather than to *consume*, in transmitting shocks. My paper contributes as well to a growing literature which aims to develop analytically tractable models combining household heterogeneity and monetary policy, in order to gain a better intuitive understanding of the mechanisms at play and the distinction between these models with heterogeneity and their representative-agent counterparts—for instance, Bilbiie (2021, 2020); Bilbiie et al. (2021); Acharya et al. (2020). As in the early HANK literature, these papers primarily focus on heterogeneity in labor income, and specifically, on the degree to which the income risk of various groups co-moves with the business cycle. I develop a complementary framework, showing in an analytically tractable framework the importance of entrepreneurial return heterogeneity in determining the effects of Central Bank policy.

Of course, my paper is not the first to study monetary policy with entrepreneurs. The celebrated "Financial Accelerator" literature—as in, for example, Carlstrom and Fuerst (1997), Bernanke et al. (1999), and Carlstrom and Fuerst (2001)—uses entrepreneurs as a mechanism to link aggregate activity to financial market fluctuations. The papers in this literature argue that shocks to financial markets affect the asset values of entrepreneurs, the activities of whose firms are linked to their personal wealth. In this way, adverse shocks which affect asset values are amplified, as these shocks reduce entrepreneurial investment, and thus output. Indeed, Kiyotaki (1998) mentions but does not study a simple version of the mechanism which lies at the heart of my paper: with entrepreneurs who are *ex-ante* heterogeneous (the bulk of other papers in this literature assumed only ex-post, IID heterogeneity), a monetary shock can potentially redistribute assets between entrepreneurs. I show, in a model that retains much of the analytical tractability of these earlier papers, that this is indeed the case. As a result, my work extends and refines the results of this literature: where others have found that aggregate entrepreneurial wealth is an important determinant of monetary transmission, I go further and show that the *distribution* of wealth among entrepreneurs matters as well.

Finally, a recent group of studies has focused on the role of monetary policy in reallocating resources across heterogeneous firms or investors. Ottonello and Winberry (2020) and Jeenas (2023) both study monetary policy with heterogeneous firms, finding that heterogeneity in firm balance sheets affects monetary transmission. In the former, accommodative shocks shift investment towards firms with lower default risk, and the latter that these shocks reallocate towards firms with more liquid balance sheets. I complement their analyses by showing that firm productivity is also a meaningful dimension along which monetary policy acts to reallocate assets. Kekre and Lenel (2022) show that with heterogeneity in risk tolerance, a decrease in interest rates decreases risk premia by shifting wealth to agents with higher willingness to invest in risky assets. Similarly, Melcangi and Sterk (2024) argue that monetary expansions increase wealth inequality by benefitting the small subset of the population active in the stock market, and conversely, have a larger effect when wealth is more concentrated in the hands of stockholders, as has increasingly become the case in the US data. Both of the latter two models would predict that, as wealth inequality increases, the efficacy of monetary policy should always increase. However, the data do not support this idea: over the past four decades, as wealth inequality has increased, what evidence there is of a change in the strength of monetary policy has shown that its potency has *decreased* (see e.g. Boivin et al. (2010); Boivin and Giannoni (2002)). While there are many ways of accounting for this potential change, my model provides a rationalization: beyond a certain point, increases in inequality dampen the redistributive channel of policy by making it harder for wealthy entrepreneurs to earn returns in excess of the aggregate return on capital. This channel is lost when the risky asset in question evolves exogeneously, rather than as a function of allocations.

The two papers most similar to mine are Matusche and Wacks (2021) and González et al. (2024). Matusche and Wacks (2021) construct a computational model of heterogeneous entrepreneurs who face diminishing returns, and show that an accommodative shock shifts wealth towards wealthier entrepreneurs, thereby increasing aggregate investment and amplifying the response of the economy to the monetary shift. They also demonstrate by means of a numerical example that shifting wealth towards entrepreneurs in the steady state leads to larger effects of monetary policy. González et al. (2024) also demonstrate that an easing in monetary policy shifts resources towards firms with higher productivity, and derive a prescription for optimal monetary policy in this economy. They also provide empirical evidence using a representative sample of the universe of Spanish firms—both public and private—that heterogeneity in marginal returns to capital better explains the response of investment to monetary shocks, as opposed to other balance sheet, revenue, or productivity measures. Relative to these papers, the contribution of mine is to go further in deriving closed-form, analytic results on the response of investment, productivity, and the wealth distribution to a monetary shock. My analysis offers several advantages, relative to a computational approach. It provides clear, direct insights into the relationships between aggregate variables and policy shocks, allowing for an intuitive understanding of the way in which monetary policy transmits to economic activity. Furthermore, my analytical approach facilitates comparative statics analysis, making it easier to study how this transmission depends on the process of entrepreneurial productivity, and the wealth distribution to which it gives rise. I also emphasize the role of *nominal* leverage in reallocating assets towards entrepreneurs, an important channel which this literature has thus far largely overlooked. Auclert (2019) and Doepke and Schneider (2006) show that empirically, one of the primary means by which monetary policy affects household choices is by generating inflation, which redistributes from nominal creditors to nominal debtors. This force is also present, and prominent, in my model: to the extent that more productive entrepreneurs are wealthier than those who are less productive, these wealthy entrepreneurs will issue more nominal debt to increase their scale, and thus will benefit to a larger degree from unexpected inflation.

My paper proceeds as follows. In section 2 I construct my model, and describe optimal behavior of all of the agents therein. I also study selected properties of the steady state in my model. I will argue that the wealth distribution in the steady state, and the assumptions underlying it, are crucial in shaping the transmission of monetary policy. As such, it is important to build an understanding of how the steady state varies across parameters. Section (3) contains my main results: responses of aggregate variables to an unanticipated change in monetary policy. In addition, Section (3) studies how these responses change in the wealth distribution at the time of the shock. Section (4) concludes.

2 Model

I consider a model with two types of agents: workers and entrepreneurs. It is not possible for a worker to become an entrepreneur, or vice versa. All workers are identical. Entrepreneurs are indexed by $i \in [0,1]$. If entrepreneur i chooses to be active in period t, she hires workers on the spot labor market. The private firm i produces output according to

$$y_{it} = \max_{n_{it}^d} \left(z_{it} k_{it}\right)^{\alpha} \left(n_{it}^d\right)^{1-\alpha} \tag{1}$$

Here, k_{it} is the capital stock of household i, and n_{it}^d is the quantity of labor hired by firm i on the spot market. Entrepreneurs produce a homogeneous good y, which may be used for either consumption or capital investment. The entrepreneurial talent or productivity of household i in period t is given by z_{it}.

2.1 Entrepreneurs' Problem and Collateral Constraints

I follow papers such as Buera et al. (2011) and Moll (2014) in assuming that entrepreneurs have the ability to save in one of two assets: capital used to run their firm, and risk-free nominal bonds.[2]Entrepreneurs maximize their lifetime expected utility

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U^e\left(c_{it}\right) \tag{2}$$

subject to their nominal budget constraint,

$$P_t \left\{c_{it} + q_{it}\right\} = P_{tx} y_{it} - W_t l_{it} - \left(1 + i_t\right) D_{it} + D_{it+1} \tag{3}$$

where P_t is the aggregate price level, W_t the *nominal* wage (taken as given by the firm), and i_t the nominal interest rate between $t-1$ and t, set by the monetary authority at time $t-1$. D_{it+1} is the quantity of nominal bonds *issued* by firm i at time t; so $D_{it} < 0$ indicates that the household is a net lender, or purchaser of bonds. These agents split their total income between purchases of consumption goods c_{it} and investment (capital) goods q_{it}, which I assume are identical and hence share nominal price P_t. Entrepreneur i's capital stock evolves according to

$$k_{it+1} = (1-\delta) k_{it} + q_{it} \tag{4}$$

where q_{it} is the quantity of investment goods purchased. Following Buera and Moll (2015), I assume that households are subject to a collateral constraint of the form

$$D_{it+1} \leq \theta P_t k_{it+1}, \qquad \theta \in [0,1] \tag{5}$$

This collateral constraint implies that only a proportion θ of the nominal value of the next-period capital stock may be externally financed. As I will demonstrate, this framework is isomorphic to one in which entrepreneurs save only in risk-free bonds, and borrow their entire period-t capital stock from an intermediary. I also follow Buera and Moll (2015) in assuming that next-period productivity z_{it+1} is revealed to household at the end

[2]As will become clear later, these bonds are risk-free in the sense that their *nominal* rate of return between periods t and $t+1$ is predetermined in period t. However, their *real* return is subject to risk in the event of unanticipated inflation between these two periods.

of period t, before they issue bonds D_{it+1}.

Each entrepreneurial household maximizes its nominal capital income:

$$P_t y_{it} = \max_{n_{it}^d} \; P_{tx} \left(z_{it} k_{it} \right)^{\alpha} \left(n_{it}^d \right)^{1-\alpha} - W_t l_{it} \tag{6}$$

The following lemma is a well-known result in problems such as (6):

Lemma 1. *Entrepreneurial labor demand is linear in capital:*

$$n_{it}^d = \left(\frac{1-\alpha}{W_t / P_t} \frac{P_{tx}}{P_t} \right)^{\frac{1}{\alpha}} z_{it} k_{it} \tag{7}$$

Lemma 1 is the result of the fact that the problem in (6) is static: entrepreneurs hire labor on the spot market to maximize their profit given their state (z_{it}, k_{it}) and the wage and price W_t and P_{tx}, which they take as given. Defining

$$\omega_t \equiv \alpha \left(\frac{1-\alpha}{w_t} \right)^{\frac{1-\alpha}{\alpha}} p_{tx}^{\frac{1}{\alpha}} \tag{8}$$

where w_t is the real wage and p_{tx} the real entrepreneurs' price, the budget constraint can be written as

$$P_t c_{it} + P_t k_{it+1} = P_t \left[\omega_t z_{it} + (1-\delta) \right] k_{it} - (1 + i_t) D_{it} + D_{it+1} \tag{9}$$

It is useful to write the entrepreneurs' budget constraint in real terms. To do so, define real bond issuance as

$$d_{it+1} \equiv \frac{D_{it+1}}{P_t} \tag{10}$$

With this definition, the budget constraint for entrepreneurial household i in real terms is

$$c_{it} + k_{it+1} = \left[\omega_t z_{it} + (1-\delta) \right] k_{it} - (1 + r_t) d_{it} + d_{it+1} \tag{11}$$

Here, r_t is the time-t ex-post real interest rate, defined by the Fisher equation:

$$1 + r_t = (1 + i_{t-1}) \frac{P_{t-1}}{P_t} = \frac{1 + i_t}{1 + \pi_t} \tag{12}$$

Note that this interest rate depends on the realized inflation rate π_t. Define the real net worth as

$$a_{it} = k_{it} - d_{it} \tag{13}$$

With this definition, I can write the borrowing constraint in real terms:

$$d_{t+1} \leq \theta k_{t+1} \tag{14}$$

or

$$k_{t+1} \leq \lambda a_{t+1}, \qquad \lambda \equiv \frac{1}{1-\theta} \tag{15}$$

Then, I have the following lemma, similar to Moll (2014) and Buera and Moll (2015):

Lemma 2. *Entrepreneurs' capital and bond choices are at corners:*

$$k_{t+1} = \begin{cases} \lambda a_{t+1} & z_{t+1} \geq \underline{z}_{t+1} \\ 0 & z_{t+1} < \underline{z}_{t+1} \end{cases} \tag{16}$$

$$d_{t+1} = \begin{cases} (\lambda - 1)\, a_{t+1} & z_{t+1} \geq \underline{z}_{t+1} \\ -a_{t+1} & z_{t+1} < \underline{z}_{t+1} \end{cases} \tag{17}$$

where the cutoff $\underline{z}$ is such that

$$\omega_{t+1}\underline{z}_{t+1} = r_{t+1} + \delta \tag{18}$$

Lemma 2 shows that entrepreneurs are essentially divided into two groups: those above the productivity threshold in (2), who are active, and those below it, who are inactive. Active entrepreneurs, who earn excess returns on their investment above the risk-free rate, borrow up to their limit and are thus bound by the collateral constraint. Inactive entrepreneurs save at the risk-free rate, lending to active entrepreneurs. Due to the linearity of the production technology, the cutoff productivity $\underline{z}_t$ is independent of wealth; instead, $\underline{z}_t$ is a linear function of the risk-free rate r_t and ω_t, which can be thought of as the private return per effective unit of capital zk.

2.2 Nominal Rigidities

To introduce nominal rigidities while maintaining tractability, I follow Bernanke et al. (1999) in assuming a three-tiered production structure. Entrepreneurs produce a homogenous good x_t, which is then sold to retailers. Retailers, a continuum of whom are indexed by $j \in [0, 1]$, in turn costlessly differentiate these goods. Retailers sell their output y_{tj} to a final good producer, who aggregates them using a CES technology:

$$Y_t = \left[\int_0^1 y_{tj}^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{19}$$

This assumption on the structure of production allows me to introduce price stickiness in a way that preserves the tractability of the entrepreneurs' problem. It is analytically convenient to assume that entrepreneurs are price takers; otherwise, their investment and savings choices would be intermingled with a forward-looking pricing problem, which would complicate my model without providing any obvious upside.

Optimal behavior by the final good aggregator in (19) implies that the demand for variety j is

$$y_{t,j} = \left(\frac{P_{t,j}}{P_t} \right)^{-\varepsilon} Y_t \tag{20}$$

where

$$P_t = \left(\int P_{t,j}^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}} \tag{21}$$

is the overall price level. Retailer j produces output y_{tj} according to

$$y_{tj} = x_{tj} \tag{22}$$

therefore, their marginal cost is $m_t = p_{tx}$. In addition, retailers incur Rotemberg (1982)-style quadratic adjustment costs to change their price:

$$\Theta\left(P_{tj}, P_{tj}^{R}\right) = \frac{\theta}{2}\left(\frac{P_{tj}}{P_{tj}^{R}} - 1\right)^{2} Y_t \tag{23}$$

P_{tj}^{R} is the time-t reference price for retailer j. Typically, $P_{tj}^{R} = P_{t-1,j}$; that is, the firm incurs an adjustment cost when it wants to update its price relative to its own lagged price. As shown in lemma 3 below, I consider this case, as well as a "static" case chosen for additional gains in tractability.

Defining inflation as

$$\pi_t \equiv \frac{P_t}{P_{t-1}} - 1 \tag{24}$$

the following lemma describes the behavior of inflation over time:

Lemma 3. *Inflation evolves according to the New Keynesian Phillips Curve, which arises under optimal behavior by retailers:*

$$\pi_t = \frac{\varepsilon}{\theta}\left(p_{tx} - m^*\right) + \beta_f \mathbb{E}_t \pi_{t+1} \tag{25}$$

Here, $m^* = \varepsilon / \left(\varepsilon - 1\right)$ *is the inverse of the optimal markup in the absence of price rigidities, and* β_f *is the rate at which retailers discount future profits. In the two cases that I consider,*

$$\beta_f = \begin{cases} \beta & P_{tj}^{R} = P_{t-1,j} \\ 0 & P_{tj}^{R} = P_{t-1} \end{cases} \tag{26}$$

The intuition in Lemma 3 is standard. Iterating forward on equation (25) gives

$$\pi_t = \frac{\varepsilon}{\theta}\sum_{s=0}^{\infty} \beta_f^s \left[p_{t+s,x} - m^*\right] \tag{27}$$

In the presence of price stickiness, retailers raise their prices when they believe that future marginal costs will exceed their long-run optimal level—equivalently, retailer raise current prices when they believe that, in the future, markups will fall *below* their long-run optimal level. Additionally, the retailers' discount rate depends on the reference price on which their adjustment cost is based. Under the typical assumption that a retailer's adjustment costs are a function of the deviation between its own current and lagged prices ($P_{tj}^{R} = P_{t-1,j}$), retailers share a discount factor with the households by whom they are owned. If, on the other hand, the reference price for the firm is the lagged aggregate price, then the New Keynesian Phillips Curve is static, and current inflation depends only on the current deviation of marginal cost from its optimal long-run level. As noted in Bilbiie (2021), this assumption is empirically unrealistic; it implies that firms do not consider future profits in setting their current price. Nevertheless, it allows for a contemporaneous tradeoff between inflation and real output, and as such can offer a convenient alternative to the forward-looking assumption.

2.3 Equilibrium

There are two actors needed to close the model: workers, and a monetary authority. For expositional purposes, I assume for the time being that workers—who supply their labor to entrepreneurs at real wage w_t—cannot borrow or save, and are thus constrained to be hand-to-mouth. Workers are, however, free to adjust their labor supply in response to movements in the real wage. Worker households are all identical, and have preferences as in Greenwood et al. (1988):

$$U^w\left(C_t^w, N_t^w\right) = \frac{1}{1-\gamma}\left(C_t^w - \frac{\left(N_t^w\right)^{1+\varphi}}{1+\varphi}\right)^{1-\gamma} \tag{28}$$

In addition, worker households own the retailers, and receive the profits of these firms as real dividends T_t. Workers' budget constraint in real terms is simply $C_t = w_t N_t + T_t$. The labor supplied by the households is given by

$$N_t^w = w_t^{1/\varphi} \tag{29}$$

I follow the literature in assuming that the monetary authority sets the nominal interest rate i_t according to a Taylor rule:

$$i_{t+1} = \overline{r} + \phi_\pi \pi_t + \nu_t \tag{30}$$

Recall that i_t dictates the nominal cost that an entrepreneur pays for outside financing. In order to ensure notational consistency, I date all interest rates according to when they are earned, rather than when they are set. As such, the nominal rate i_{t+1} in (30) is set at time t, and dictates the interest rate on nominal debt issued in period t and maturing in period $t+1$. The term ν_t is an exogenous, stochastic innovation; I will use this shock to measure the impact of monetary policy in my model.

Definition 4. An equilibrium is a sequence of prices $\{P_t, P_{tx}, W_t\}$, aggregates $\{C_t^e, C_t^w, N_t, Y_t, K_t, Z_t\}$, interest rates $\{i_t, r_t\}$, a path for inflation π_t, a sequence of aggregate shocks ν_t, and a sequence of distributions $\{g_t(a,z)\}$ over the idiosyncratic states for entrepreneurs such that:

1. Entrepreneurs, workers, retailers, and the final good producer all maximize their respective objectives,
2. The monetary authority sets the nominal interest rate in accordance with the Taylor rule in (30), given an exogenous sequence for the shock ν_t,
3. Prices clear markets:

$$K_t = \int_0^{\overline{z}} \int_0^\infty a g_t(a,z)\, da dz \tag{31}$$

$$N_t^w = N_t^d \tag{32}$$

$$C_t^e + C_t^w + K_{t+1} + \Theta(\pi_t) = Y_t + (1-\delta) K_t \tag{33}$$

I will study the properties of the equilibrium defined above using *wealth shares*:

$$s_t(z) \equiv \frac{1}{K_t}\int_0^\infty a g_t(a,z)\, da \tag{34}$$

As in Moll (2014), among others, the wealth share $s_t(z)$ denotes the share of aggregate wealth held by agents of type z. There are a number of reasons why these objects are a convenient tool for studying the behavior of the model. First, the shares $s_t(z)$ can be thought of as a density: they are nonnegative for all z, and integrate to one: $\int_0^{\overline{z}} s_t(z)\, dz = 1$ for all t. As such, I can define the analogous cumulative share:

$$S_t(z) \equiv \int_0^z s_t(\hat{z})\, d\hat{z} \tag{35}$$

Second, note that because returns are linear in wealth, individual wealth follows a random growth process.[3] As a result, the joint distribution $g_t(a,z)$ does *not* admit a stationary measure: the log of individual wealth a_t follows a random walk, and thus the cross-sectional variance of a_t grows without bound in t. However, it can be demonstrated that the *wealth shares* $s_t(z)$ do admit a stationary measure. This result is convenient: it allows me to study the long-run properties of my model without needing to augment the model with an assumption to deliver a stationary measure over wealth, such as random death and annuity markets (as in Gouin-Bonenfant and Toda, 2019) or a hard borrowing limit.

2.3.1 Aggregation

Using the definition of wealth shares in (34), aggregate quantities are easily derived:

Proposition 5. *Aggregate quantities satisfy*

$$Y_t = (Z_t K_t)^{\alpha} N_t^{1-\alpha} \tag{36}$$

$$K_{t+1} = \beta \left\{ \alpha p_{tx} Y_t + (1-\delta) K_t \right\} \tag{37}$$

Aggregate productivity is a function of the wealth distribution $s_t(z)$:

$$\begin{aligned} Z_t &= \frac{\int_{\underline{z}_t}^{\overline{z}} z s_t(z)\, dz}{\int_{\underline{z}_t}^{\overline{z}} s_t(z)\, dz} \\ &= \mathbb{E}_{s_t}\left[z | z > \underline{z}_t \right] \end{aligned} \tag{38}$$

Given a path for wealth shares $s_t(z)$, the cutoff productivity $\underline{z}_t$ is pinned down by capital market clearing:

$$1 = \lambda \left(1 - S_t(\underline{z}_t)\right) \tag{39}$$

Factor prices are

$$w_t = (1-\alpha) p_{tx} \left(\frac{Z_t K_t}{N_t} \right)^{\alpha} \tag{40}$$

$$\mathbb{E}_{t-1} r_t = \alpha p_{tx} Z_t^{\alpha} \left(\frac{N_t}{K_t} \right)^{1-\alpha} \frac{\underline{z}_t}{Z_t} - \delta \tag{41}$$

Returns are given by

$$\omega_t = \alpha p_{tx} \left(\frac{N_t}{Z_t K_t} \right)^{1-\alpha} \tag{42}$$

$$R_{tK} = 1 - \delta + \alpha p_{tx} Z_t^{\alpha} \left(\frac{N_t}{K_t} \right)^{1-\alpha} \tag{43}$$

The return to an entrepreneur of type z is given by

$$R_t(z) = 1 + i_{t-1} - \pi_t + \lambda \begin{cases} 0 & z < \underline{z}_t \\ \omega_t z - (i_{t-1} - \pi_t + \delta) & z > \underline{z}_t \end{cases} \tag{44}$$

[3] See Gabaix (2009) for a study of random growth processes in economics, and Benhabib et al. (2015) for an example of how this process gives rise to wealth distributions in models that share the "fat-tailed" (Pareto) nature of their empirical counterparts.

Proposition 5 has largely the same interpretation as its counterparts in Moll (2014) and Buera and Moll (2015); I refer the reader there for excellent discussions. Equations (36) and (37) show that this economy behaves as one with a representative firm, with the key difference being that aggregate TFP is an endogenous result of the wealth distribution among entrepreneurs, as in (38). Per equation (40), the real wage is given by the real value of the aggregate marginal product of labor hired by entrepreneurs. The same is generally *not* true for the ex-ante expected risk-free real rate $\mathbb{E}r_t$: in equation (41), this object is equal to the aggregate marginal return to capital, weighted by $\underline{z}_t/Z_t$.[4] In this economy, capital market frictions prevent the return from investment to be equated with outside savings. As a corollary, the two are equated in the case that $\underline{z}_t = Z_t$, which is the case when capital markets are frictionless, $\lambda = \infty$. Note as well that both factor prices, as well as the returns R_{tK} and ω_t, move with the price paid to entrepreneurs by retailers for their goods.

In equation (43), the aggregate return to capital is derived from

$$R_{tK} = \int_0^{\overline{z}} R_t\left(z\right) s_t\left(z\right) dz \tag{45}$$

$$= \mathbb{E}_{s_t}\left[R_t\left(z\right)\right] \tag{46}$$

Thus, the aggregate return on capital is the average return across all entrepreneurs, weighted by their respective wealth shares. Finally, entrepreneurs' returns, per equation (44), exhibit a few key properties that will later drive my results. First, entrepreneurs with $z > \underline{z}_t$ are able to earn excess returns above the ex-ante risk-free rate $i_{t-1} - \pi_t$, due to their ability to make leveraged investments into their inside firm. In partial equilibrium, equation (44) previews the differential impact of a change in real rates on entrepreneurs of different productivities. Returns can also be written as

$$R_t\left(z\right) = \begin{cases} 1 + i_{t-1} - \pi_t & z \leq \underline{z}_t \\ 1 - \left(\lambda - 1\right)\left(i_{t-1} - \pi_t\right) + \lambda\left(\omega_t z - \delta\right) & z > \underline{z}_t \end{cases} \tag{47}$$

From (47) it is immediately obvious that a fall in the real rate $i_{t-1} - \pi_t$ lowers the returns of savers, who *earn* this rate, and raises that of active entrepreneurs, who *pay* this return to borrow capital. Additionally, the expression of returns in (47) makes clear the role of inflation in driving redistribution in this model: an unexpected increase in π_t reduces nominal obligations, redistributing from low types (lenders) to high types (borrowers). Crucially, π_t is realized one period *after* $\underline{z}_t$ has been determined: entrepreneurs operate their firms at time t with capital stock chosen at time $t-1$, and thus agents cannot switch from being inactive to active following unexpected inflation.

2.4 Persistence in Returns

To study analytically the effects of a monetary shock in my model economy, I make the following assumption on individual entrepreneurs' productivities:

Assumption 1. *Individual entrepreneurial productivities are distributed according to some differentiable, time-invariant function $F(z)$. With probability p, an entrepreneur will maintain his productivity from one*

[4] The expectations operator indicates that the ex-post real rate is subject to inflation risk. In the absence of nominal rigidities, Equation (41) would always hold. In order to study the redistributive effects of unanticipated inflation, I leave open the possibility that the ex-post real rates may differ from their ex-ante expectations. In the event that inflation is not equal to its ex-ante expectation, this equation will hold for the expected risk-free rate, upon which time-t contracts are based, but *not* for the ex-post rate r_{t+1}.

period to the next, $z_{t+1} = z_t$. With probability $1-p$, meanwhile, he draws his next-period productivity at random from the time-invariant distribution given by $F(z)$.

Assumption 1 allows for gains in tractability while maintaining rich heterogeneity in the model. With this assumption, the autocorrelation of z_t and z_{t+1} is parsimoniously given by

$$\rho\left(z_t, z_{t+1}\right) = p \tag{48}$$

This persistence is also incorporated in an appealing way: conditional on $z_{t+1} \neq z_t$, the distribution of z_{t+1} is independent of z_t. Thus, I maintain many of the desirable properties of an IID process while still allowing for a positive autocorrelation in returns. As pointed out by Moll (2014), persistence in returns is the empirically relevant case (see, e.g., DeBacker et al., 2023), and the case that leads to a correlation between entrepreneurs' wealth and their productivity. With autocorrelated productivities, more productive entrepreneurs will accumulate more wealth over time, using their own wealth as a complement to outside credit. I also assume that the distribution $F(z)$ has support on $\mathcal{Z} = [0, \overline{z}]$ with $\overline{z} < \infty$.

Under assumption 1, the behavior of the wealth shares $s_t(z)$ defined in (34) can be characterized in a clean and intuitive fashion:

Proposition 6. *The wealth share of type z, $s_t(z)$, evolves according to*

$$s_{t+1}(z) = p\frac{R_t(z)}{R_{tK}} s_t(z) + (1-p) f(z) \tag{49}$$

where R_{tK} is the aggregate (wealth-share weighted average) return to capital, as defined in Lemma 5.

Proposition (6) has an intuitive interpretation. There are two sources of change in the wealth share $s_{t+1}(z)$: entrepreneurs who retain their type ($z_{t+1} = z_t$), and entrepreneurs who transition to type z at $t+1$ from some other $z_t \neq z_{t+1}$. For each source of change, the sign of its contribution (whether it increases or decreases $s_{t+1}(z)$ relative to $s_t(z)$) depends on the returns of the agents in question, relative to the aggregate return on capital. For agents who retain their type: if the time-t return $R_t(z)$ is greater than the aggregate return to capital, then the wealth of agents of type z grows faster than the overall capital stock, and their share increases. Agents transitioning to type z from some other z' (the $(1-p) f(z)$) term in (49) on average earn, by definition, the aggregate return R_{tK}, hence the coefficient of 1 on this term.

2.5 Steady State

Here I define the steady-state concept that I will use for my analysis. When studying the effect of monetary policy, I begin the economy in its long-run, zero-inflation steady state. Lemma 7 outlines the behavior of aggregate quantities and prices in this steady state, where quantities without a time subscript denote their steady-state values:

Lemma 7. *Quantities and prices in the zero-inflation steady state satisfy*

$$Y = (ZK)^{\alpha} N^{1-\alpha} \tag{50}$$

$$R_K = 1 - \delta + \alpha p_x Z^{\alpha} (N/K)^{1-\alpha} = \frac{1}{\beta} \tag{51}$$

The steady-state wealth shares are given by

$$s(z) = \frac{1-p}{1-\beta p R(z)} f(z) \tag{52}$$

where $R(z)$ denotes the return on private capital to type z in the steady state, given by:

$$R(z) = 1 + \alpha p_x Z^{\alpha} \left(\frac{N}{K}\right)^{1-\alpha} \frac{z_{\lambda}(z)}{Z} - \delta \tag{53}$$

where

$$z_{\lambda}(z) \equiv \begin{cases} z + (\lambda - 1)(z - \underline{z}) & z > \underline{z} \\ \underline{z} & z \leq \underline{z} \end{cases} \tag{54}$$

and the cutoff is implicitly defined again by $1 = \lambda(1 - S(\underline{z}))$. Aggregate productivity is

$$Z = \lambda \int_{\underline{z}}^{\overline{z}} z s(z)\, dz \tag{55}$$

The steady-state real interest rate is given by

$$r = \alpha p_x Z^{\alpha} \left(\frac{N}{K}\right)^{1-\alpha} \frac{\overline{z}}{Z} - \delta \tag{56}$$

The price for entrepreneurial goods is equal to its optimal level in the absence of nominal rigidities:

$$p_x = m^* = \frac{\varepsilon - 1}{\varepsilon} \tag{57}$$

As will become clear, the nature of the steady state at the time of a change in policy—in particular, the steady state shares $s(z)$—is a key driver of the aggregate response to policy. Equation (51) shows that stationarity of the capital stock implies that the steady state return to capital is the inverse of entrepreneurs' discount rate. From the Phillips curve (Equation 25), in the long run prices and quantities adjust such that retailers' markups reach their optimal level $\mathcal{M}^* = \varepsilon/(\varepsilon - 1)$, and thus the price they pay entrepreneurs for inputs will equal the inverse of the optimal markup, as in Equation (57). Steady-state entrepreneurial returns and wealth shares, laid out in equations (52)-(54), have forms that are simple and intuitive thanks to the linearity of entrepreneurs' technology and the stochastic process for z in Assumption 1. Equations (53) and (54) show that in the absence of unexpected movements in the price level (and thus the real interest rate), active entrepreneurs earn an excess return above the risk-free rate. The size of this excess return is given by $z_{\lambda}(z)$, which denotes the *effective* productivity of an entrepreneur of type z, taking into account the fact that he borrows the full amount available to him, given his net worth.

In models such as mine, aggregate productivity is determined by the allocation of wealth among entrepreneurs, as described by the shares $s(z)$. Here, misallocation in the long run is determined by two parameters: the efficiency of capital markets λ, and the persistence of entrepreneurial productivities p. Proposition (8) describes how persistence in returns z determines the properties of the steady state that will later influence the response of my economy to monetary policy:

Proposition 8. *Steady-state productivity Z is increasing in the autocorrelation p. In particular, given $\lambda > 1$*

and finite, $Z(p)$ is an increasing function of p, with

$$Z(0) = \mathbb{E}_f [z | z \geq \underline{z}] \tag{58}$$

$$Z(1) = \overline{z} \tag{59}$$

Additionally, the capital stock K, risk-free rate R, and cutoff $\underline{z}$ are all increasing in p. An increase in p also increases the concentration of wealth in the hands of entrepreneurs who earn effective return $z_\lambda(z)$ sufficiently greater than average productivity Z to overcome general equilibrium effects:

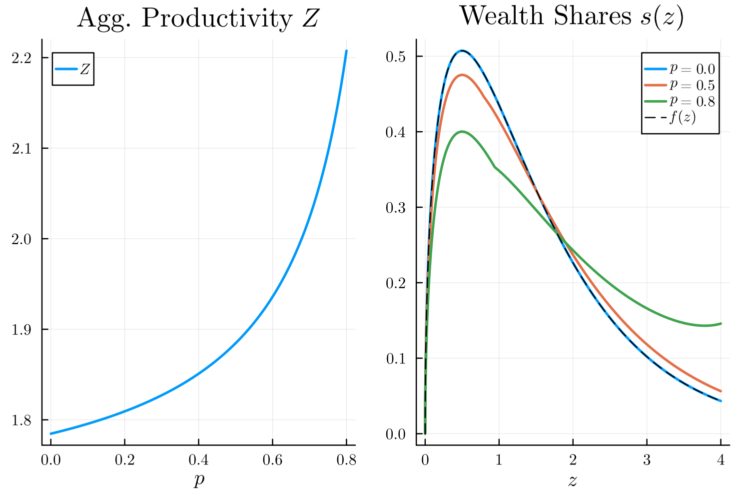
$$z_\lambda(z) - Z > p(1-p)\left\{(\lambda - 1)\frac{d\underline{z}}{dp} + \frac{z_\lambda(z)}{Z}\frac{dZ}{dp}\right\} \implies \frac{\partial}{\partial p} s(z) > 0 \tag{60}$$

Intuitively, an increase in p reduces the degree of misallocation in the steady state by allowing productive entrepreneurs to *self-finance*, accumulating wealth over time in order to scale up their projects. The more persistent is productivity, the greater the concentration of wealth in the hands of high-productivity entrepreneurs, and thus the greater is steady-state TFP Z. At the upper limit of $p = 1$, $z_{t+1} = z_t$ for all t, and TFP will be equal to its first-best level: the marginal productivity of the entrepreneurs of the highest type, $\overline{z}$. In this case, it is clear from Equations (51) and (56) that the risk-free rate equates with the aggregate return to capital, as in an economy with perfect financial markets. In this sense, full persistence in returns allows wealth accumulation by high-productivity firm owners *completely* negates the effect of financial market frictions: in the long run, these entrepreneurs hold all of the capital in the economy, and production operates with first-best TFP. At the other end, if $p = 0$, productivities z_t are IID across entrepreneurs and time, and Equation (52) implies that $s(z) = f(z)$. In this case, then, TFP Z is simply the expectation of z, conditional on $z > \underline{z}$, taken with respect to the stationary distribution $f(z)$. In between these extremes, Z takes on intermediate values: with imperfect but nonzero autocorrelation in types, wealth accumulation counteracts financial frictions, allowing productivity to increase.

The final piece of Proposition 8 concerns which entrepreneurs benefit—as measured by an increase in their wealth share—from an increase in persistence. It is not sufficient to simply say that an increase in persistence unequivocally benefits active entrepreneurs, for a few reasons. First, not all active entrepreneurs accumulate assets; it can be shown that the savings rate $\beta R(z) < 1$ even for some $z > \underline{z}$. For these entrepreneurs, an increase in persistence will tend to decrease their wealth shares, as they de-accumulate assets for longer on average. Second, as p increases, the cutoff $\underline{z}$ and TFP Z increase as well. The effect of the change in $\underline{z}$ is to push some entrepreneurs below the cutoff, reducing their wealth shares in the steady state. Additionally, the increase in overall Z reduces the *excess* returns that active entrepreneurs earn, which tends to drag on their wealth shares as well. These latter two general equilibrium forces, both of which exert downward pressure on wealth shares, are captured in the final two terms of Equation (60). Entrepreneurs whose effective returns are sufficiently above Z so as to overcome these forces, will see their long-run shares increase following a rise in p. In Section 3, I demonstrate the importance of entrepreneurial persistence, and the wealth distribution that it engenders, in determining the effect of a change in monetary policy.

An increase in the collateral constraint, λ, will also decrease misallocation, but in a different way from p. Whereas an increase in persistence undoes financial frictions via wealth accumulation, an increase λ reduces misallocation by increasing the scope for *within-period* wealth transfers via the capital market. At the limit of $\lambda = \infty$, entrepreneurs are unconstrained in their ability to take on leverage, and can borrow as much as the market will bear. In this case, the real interest rate is again bid up to the marginal product of the

Figure 1: Steady State Comparison: p



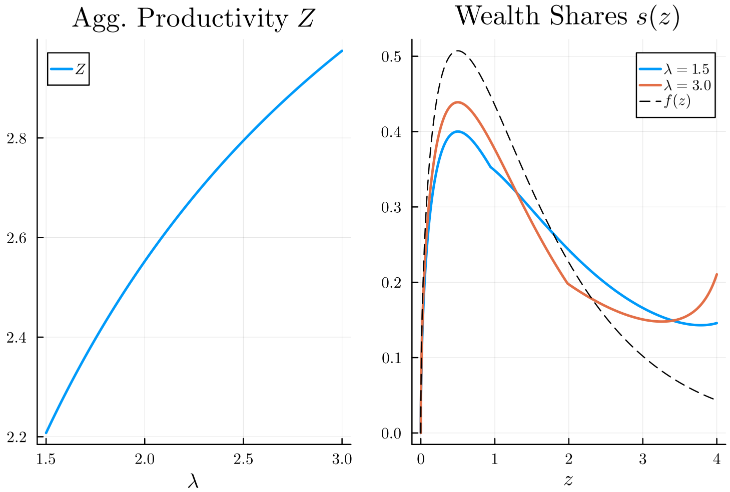
highest-type entrepreneurs, $\underline{z}$, and the economy once again attains its first-best allocation. However, despite the fact that an increase in both λ and p increase steady-state productivity by reducing misallocation, from the perspective of monetary policy, the two are *not* isomorphic. As will become clear in Section 3, whether productivity is increased as a result of λ or p implies different responses to a policy change: it matters whether high productivity entrepreneurs accumulate capital over time, or borrow it on spot markets.

To further these points, Figures 1 and 2 show comparative statics in the steady state for a given parameterization of my model.[5] Figure 1 demonstrates the results of Proposition 8: steady state TFP Z is an increasing function of the persistence of entrepreneurial productivities, p. This increase in persistence also increases the concentration of wealth, shifting the shares $s_t(z)$ to the right.

Figure 2, meanwhile, shows how productivity Z and the wealth shares $s(z)$ are affected by an increase in the quality of financial markets, modeled as an increase in λ. Here again, Z is increasing in λ. The wealth distributions underlying these gains in productivity, however, are markedly different. An increase in λ benefits low and high-productivity agents, shifting wealth away from the middle. Intuitively, a loosening of credit constraints allows for high-type entrepreneurs to borrow more capital from inactive entrepreneurs ($z < \underline{z}$). This has the effect of increasing the risk-free real rate: recall that in the absence of capital market frictions ($\lambda = \infty$), the return to outside savings is equated with the marginal product of the highest types $\overline{z}$, who undertake all of the investment. The increase in the risk-free rate benefits types who do not invest, who now earn a higher return on their savings. The increase in λ also benefits high-z types, who still earn excess returns, but are now able to take on greater leverage to increase their capital income. Looser financial frictions shift wealth away from those with z in the middle of the support, near the cutoff. These types earn small excess returns, and to them, the benefit of looser capital markets is undone by the concomitant increase in the cost of external financing.

[5] I assume $\beta = 0.96$, and that $F(z)$ is distributed according to a truncated Gamma $(1.5, 1)$ distribution on $[0, 4]$. Households' Frisch elasticity $\phi = 2$, and the elasticity of substitution $\varepsilon = 10$, implying that $p_x = 0.9$. Under the baseline, $p = \rho(z_t, z_{t+1}) = 0.5$, and $\lambda = 1.5$.

Figure 2: Steady State Comparison: λ



3 Effect of a Monetary Shock

Here, I consider the effect of an unanticipated change in the stance of monetary policy. I assume that, prior to the shock, the economy is in its long-run, zero-inflation steady state as outlined in Section (2.5). Then, at time $t = 0$, there is an unanticipated innovation to the Taylor rule:

$$i_{t+1} = \overline{r} + \phi_\pi \pi_t + \nu_t \tag{61}$$

The shock $\nu_0 < 0$ decays at rate ρ_ν:

$$\nu_t = \rho_\nu \nu_{t-1} = \rho_\nu^t \nu_0 \tag{62}$$

Although agents do not anticipate the initial shock ν_0, they understand that it will decay according to the process above, and thus for $t > 0$ we return to a *perfect foresight* equilibrium, where there are no further aggregate shocks and agents perfectly anticipate the evolution of all aggregate variables.

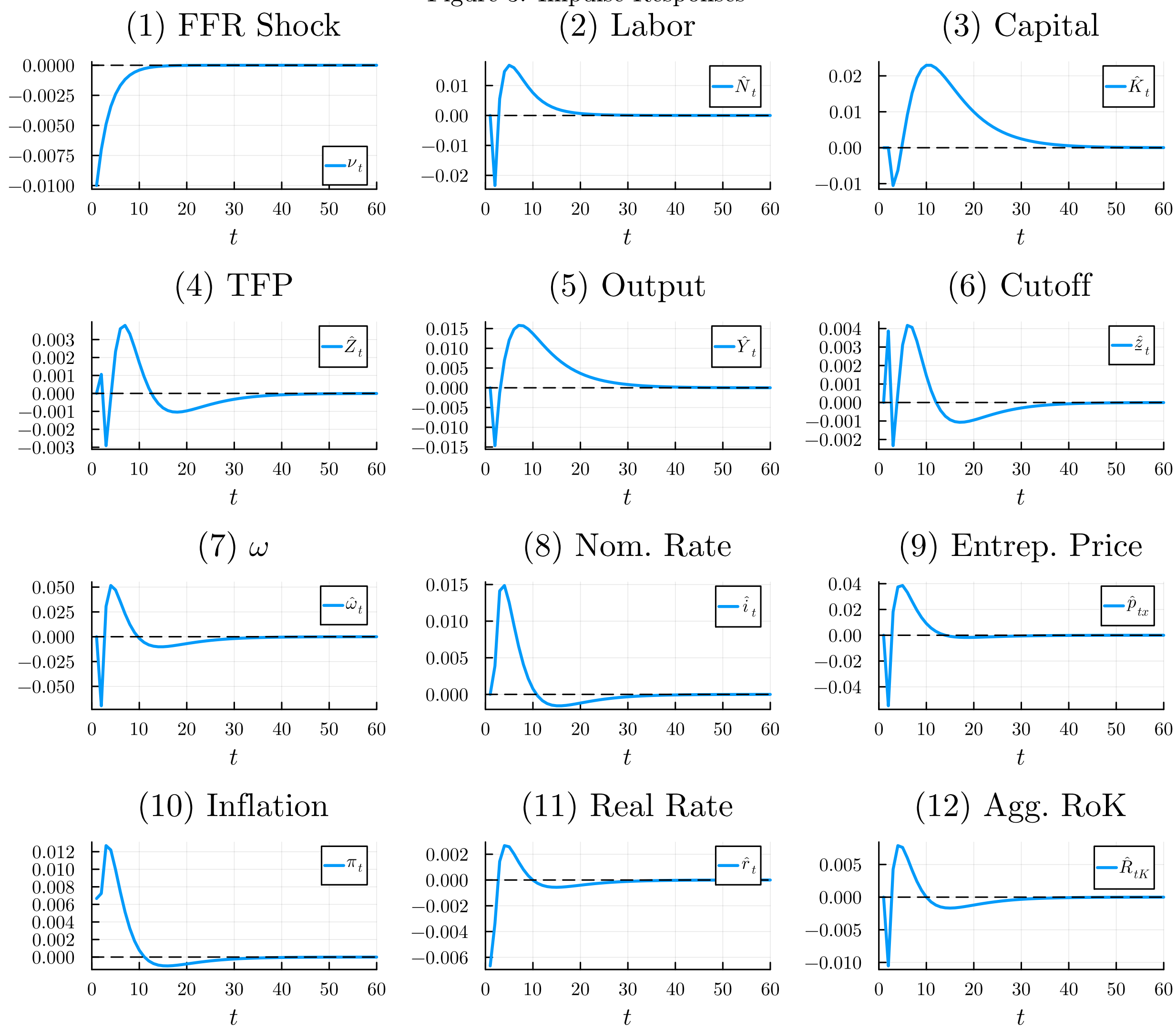
Figure 3 shows the impact of the shock ν_0 on aggregate quantities, prices, and returns in my model. By way of notation, for a variable W_t, I denote by $\hat{W}_t \equiv \ln W_t - \ln W$ its log-deviation from steady state. At present, the impulse responses in my economy cannot be entirely solved for in closed-form. However, the simplicity of my model renders solving for the full, nonlinear solution remarkably easy: Given a path for the nominal interest rate i_t, the model is recursive in $(K_t, s_t(z))$, and the shares $s_t(z)$ can be easily calculated and updated. As such, solving for transition paths in my model boils down to a solution of T equations in T unknowns, for some large T after which I assume that the model has returned to steady state.

The initial effect of the policy shock (panel 1) is to engender a temporary *decrease* in activity, along with a rise in inflation. The rise in inflation is driven by future inflation expectations. To see this, recall that solving the Phillips curve forward, as in Section (2.2), gives

$$\pi_t = \frac{\varepsilon}{\theta} \sum_{s=0}^{\infty} \beta^s \left\{ p_{tx} - \frac{\varepsilon - 1}{\varepsilon} \right\} \tag{63}$$

Thus, inflation is driven by expected future deviations of the price of entrepreneurial goods p_{tx}—the intermediate good for the retailers—from its long-run value. As is evident from panel 9 of Figure 3, following the

Figure 3: Impulse Responses



initial dip this price remains elevated for quite some time. Anticipating this path of marginal costs, retailers preemptively raise prices beginning at the time of the shock.

The initial dip in output, productivity, and returns to capital can be most clearly explained by considering an extreme limiting case of my economy: fixed prices, and IID shocks. If prices are fixed ($\theta \to \infty$ in the Phillips Curve), then $\pi_t = 0$ always, and the monetary authority has full control over the real interest rate, $r_t = i_t$. Meanwhile, if the z_t shocks are IID across time ($p = 0$), then $s_t(z) = f(z)$ for all t, as discussed in Section 2.5. An immediate corollary of the independence of wealth and productivity is that the cutoff $\underline{z}_t$ is *fixed* in time: substituting $S_t(z) = F(z)$ into capital market clearing (Equation (39) in Section 2.3.1) gives

$$F(\underline{z}) = \frac{\lambda - 1}{\lambda} \tag{64}$$

which is fixed across time and invariant to shocks, provided the leverage constraint λ remains unchanged. In this world, the path of aggregates can be easily solved:

Lemma 9. *If prices are fixed ($\theta = \infty$) and returns are serially independent ($p = 0$), returns and the evolution of the capital stock are given by*

$$\omega_t = \frac{i_t + \delta}{\underline{z}} \tag{65}$$

$$R_{tK} = 1 - \delta + \omega_t Z \tag{66}$$

$$\begin{aligned} K_{t+1} &= \beta \left\{ \alpha p_{tx} Y_t + (1-\delta) K_t \right\} \\ &= \beta K_t \left\{ \frac{Z}{\underline{z}} (i_t + \delta) + (1 - \delta) \right\} \end{aligned} \tag{67}$$

In the economy of Lemma (9), the monetary authority effectively has full leverage over both the real interest rate, and the aggregate return to capital. So, following the shock $\nu_0 < 0$, ω_t falls, and thus investment, output, and labor demand must fall. The same forces are at work with partially rigid prices ($\theta < \infty$) and autocorrelated z_t shocks. Rearranging the definition for $\underline{z}_t$ gives

$$\omega_t = \frac{i_t - \pi_t}{\underline{z}_t} \tag{68}$$

From the perspective of time t, $\underline{z}_t$ and i_t are state variables. As such, an decrease in the real interest rate $i_t - \pi_t$—driven by a combination of the initial shock $\nu_0 < 0$ and the ensuing inflation—must be accompanied by a decrease in the return to effective capital, ω_t.

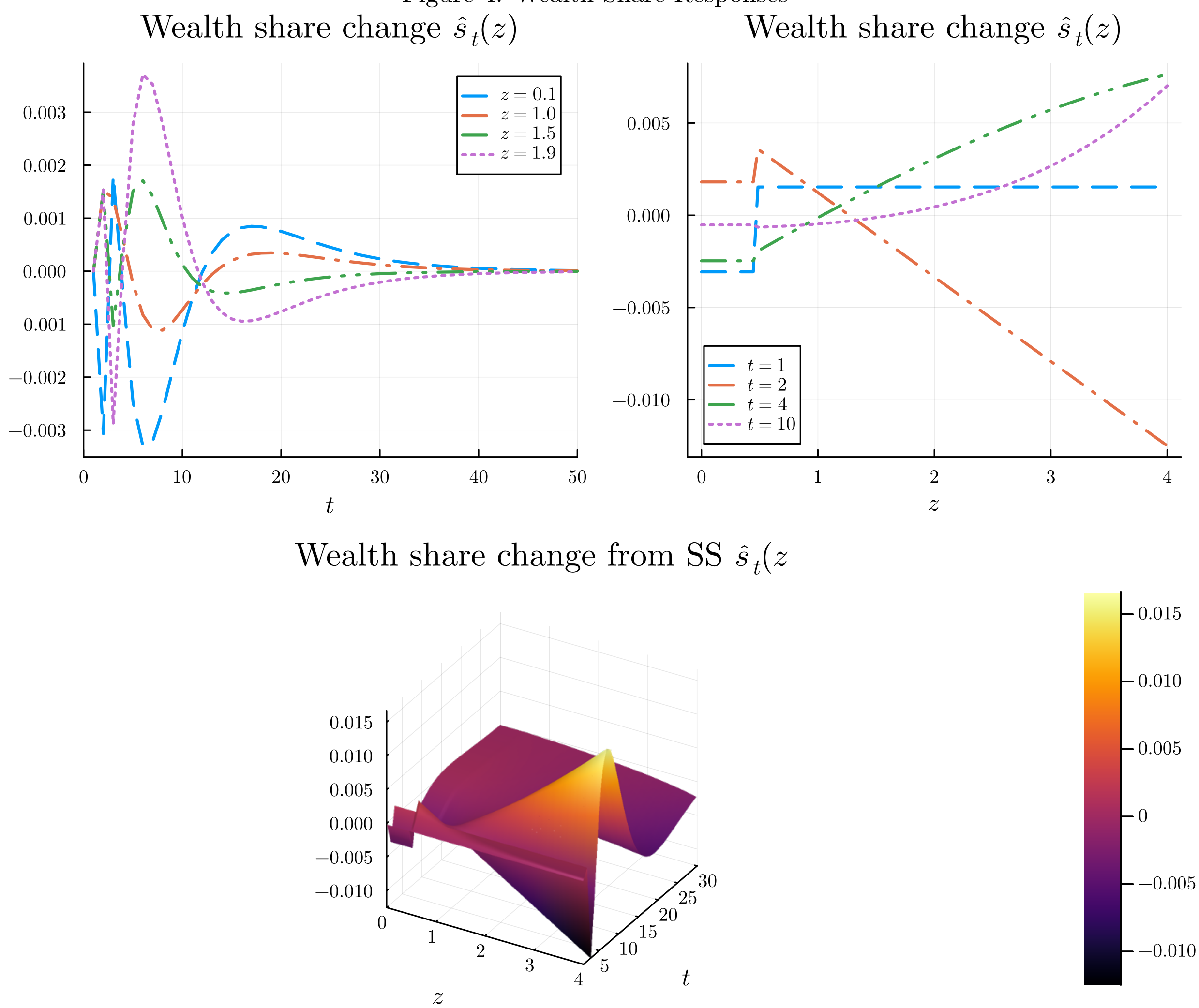
Turning to the behavior of the cutoff $\underline{z}_t$ and TFP Z_t, the model produces oscillatory responses to the monetary shock. These can be explained, however, by changes in wealth accumulation. To begin, observe that $\underline{z}_1$ increases, relative to $\underline{z}_0$ (which is equal to the steady-state $\underline{z}$ by definition). TFP in the period following the shock Z_1 increases as well. This is the consequence of the *Fisher channel*. Recall that I can write returns at the time of the shock as

$$R_0(z) = \begin{cases} 1 + i_0 - \pi_0 & z \leq \underline{z}_0 \\ 1 - (\lambda - 1)(i_0 - \pi_0) + \lambda(\omega_0 z - \delta) & z > \underline{z}_0 \end{cases} \tag{69}$$

As we see in Figure 3, ω_0 is equal to steady-state ω. Inflation, however, responds contemporaneously to the shock, increasing in anticipation of future accelerated activity, as discussed above. This inflation has the usual effect, first suggested by Fisher (1933): the increase in the price level devalues nominal debt, redistributing from lenders to borrowers. Here, because lenders are less productive entrepreneurs, and borrowers more productive, the Fisher channel has the effect of increasing concentration by shifting wealth to higher-z types. This shift comes with an increase in the cutoff $\underline{z}_t$, as well; capital market clearing implies that the mass of wealth above the cutoff, $1 - S_t(\underline{z}_t)$, must remain constant. As such, redistribution to agents with higher productivities must be accompanied by an increase in the cutoff in order to clear the capital markets. In the next period, however, the pattern reverses itself: as the real interest rate falls, the cutoff $\underline{z}_t$ falls with it. By the same market clearing logic in reverse, the fall in $\underline{z}_t$ is associated with downward redistribution of wealth, and thus a fall in TFP.

Following fall in real interest rates as a result of the monetary shock, and the accompanying fall in the cutoff, the stage is set for expansionary effects. To further study the way in which changes in the wealth distribution drive the effects in Figure 3, Figure (4) shows the evolution of the wealth shares $s_t(z)$ following the shock. The upper right and left panels of Figure (4) show cross-sections of the surface plot in the bottom panel, in t and z respectively. The top right panel shows the reversal clearly: following the redistributive inflation at $t = 0$, we see the downward redistribution mentioned above, brought about as $\underline{z}_t$ falls. As time goes on, however, this trend reverses itself: the low risk-free rate redistributes wealth from low productivity entrepreneurs, who now earn lower returns on their bond portfolios, to high productivity, who face reduced financing costs. This redistributive channel has the effect, of course, of raising aggregate productivity, further amplifying the effect on output. It is here that we see the importance of the wealth distribution in determining the overall effect of the change in policy: not only does investment increase (an effect which obtains even with identical entrepreneurs, or IID shocks), productivity also changes as the composition of investment is altered, with the additional investment being carried out by entrepreneurs who are more productive than average. The following section expands on this point.

Figure 4: Wealth Share Responses



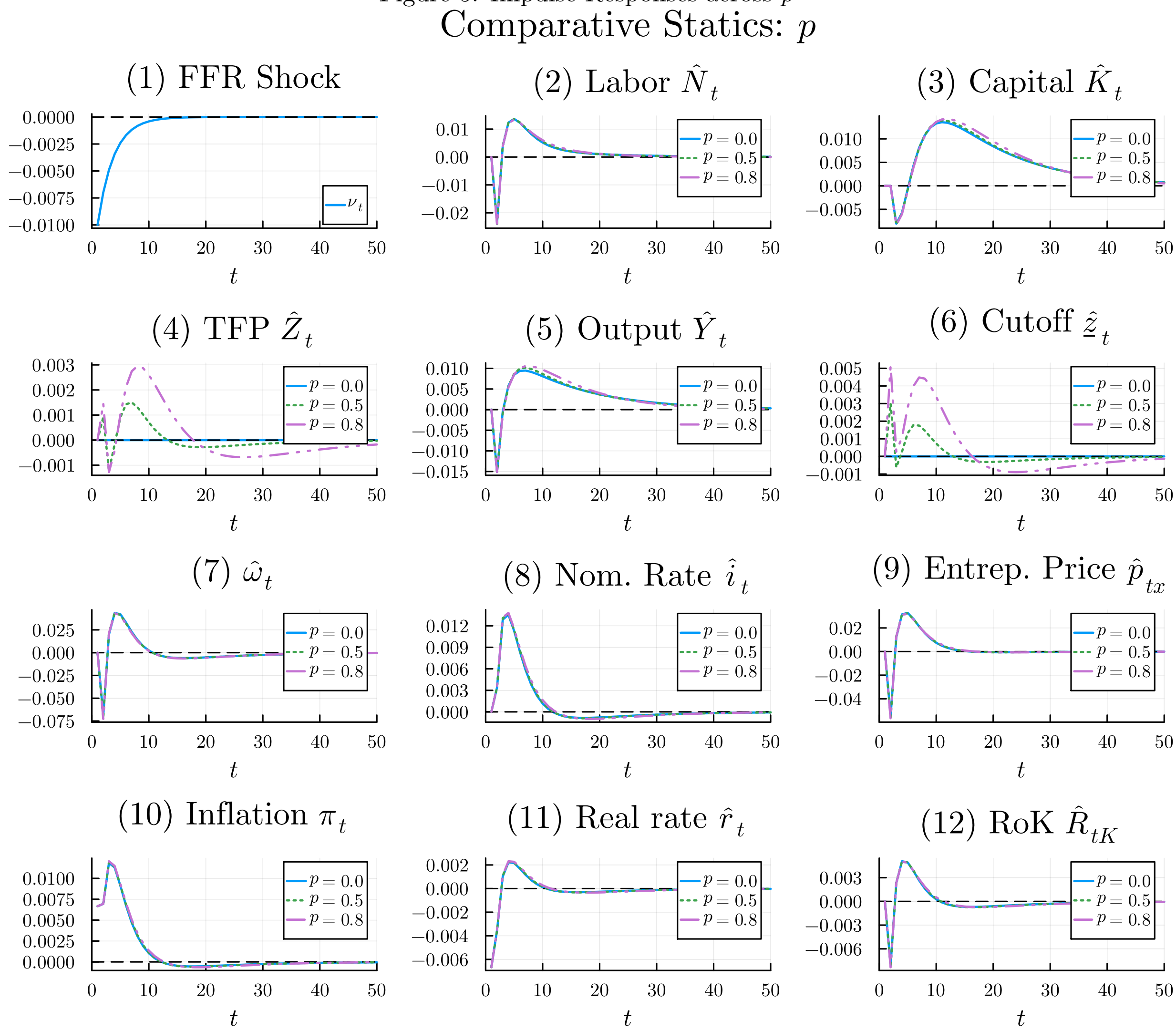
3.1 Persistence and Wealth Inequality

As outlined in Section 2.5, an increase in the persistence p of entrepreneurs' idiosyncratic productivities shifts steady-state wealth shares to the right, redistributing wealth from low to high-productivity agents and increasing productivity and investment in the process. Here, I argue that an increase in persistence also leads to a larger aggregate response to the monetary policy shock, both in terms of peak level and the time after the shock for which activity remains elevated. Figure 5 shows the responses of my model economy to the same shock as in Section 3, across three different values for the autocorrelation parameter p. I begin with the blue lines, for which shocks to z_t are IID across time ($p = 0$). In this economy, wealth and productivity are uncorrelated, and the wealth shares $s_t(z)$, cutoff $\underline{z}_t$, and TFP Z_t are all fixed in time. Here we see that the expansionary effects of monetary policy are still present: following the initial dip, there is an increase in investment and output through the usual channel: active entrepreneurs face reduced financing costs, and thus are able to scale up to a greater degree than before the shock. With $p > 0$, however, the effect on output is *amplified*. When productivities are persistent, wealth and productivity will be correlated in the steady state, as laid out in Section 2.5. Importantly, away from the steady state persistence in z_t also implies that the *distribution* of wealth changes in response to a shock (see Figure 4). As we can see in panels 4 and 6 of Figure 5, both the cutoff $\underline{z}_t$ and concentration of wealth (as measured by Z_t) are now able to change. As a result of the redistributive effects of low real rates that I describe above, aggregate productivity Z_t increases following the shock, which results in a larger peak response of output in panel 5 than with IID shocks. Furthermore, the greater is the persistence of entrepreneurs' productivities, the greater is the amplification: as p increases, productive agents accumulate more wealth out of the windfalls that they enjoy from reduced borrowing costs. As a result of this channel of wealth accumulation and redistribution, not only are the effects of the accommodative monetary shock larger, they are also *longer lasting*. This result is related to Moll (2014): with more persistence in types, the transition to the steady state takes longer, as wealth accumulation and de-accumulation take time. In brief: the more persistent are productivity shocks, the more concentrated wealth will be in the steady state, and the larger and more long-lived will be the effects of a monetary policy shock.

3.2 The Role of Financial Market Frictions

In Section 2.5, I stressed that although a relaxing of financial frictions (decrease in λ) generates an aggregate response similar to that following from an increase in persistence, the pattern of wealth distributions is markedly different. Here, I demonstrate how this difference in the steady-state wealth distribution manifests in a different path of aggregates in response to the shift in policy. Figure 6 plots the impulse responses in two economies: one with the baseline $\lambda = 1.5$, and one with looser financing constraints, $\lambda = 4$. Generally, it is *not* the case that looser financing frictions lead to greater amplification. This result follows from Figure 2 and the surrounding discussion, and is rooted in the different responses of the wealth distribution. In panel 5, we see that the response of output Y_t to the shock peaks at a higher level in the presence of *tighter* financial constraints, but is longer lived in the world with better financial markets. The reason for this can be seen in the response of the real interest rate r_t. Recall from the discussion in Section 2.5 that less restrictive financial markets benefit low-type entrepreneurs as well as high-type, as an increase in λ increases the real interest rate, bringing it closer to the marginal returns of the high types to whom capital is lent. This same effect can be seen in panel 11 of Figure 2. With looser financial frictions, the real interest rate peaks at a higher level following the shocks, for the same reason as in the steady state. The higher path for

Figure 5: Impulse Responses across p

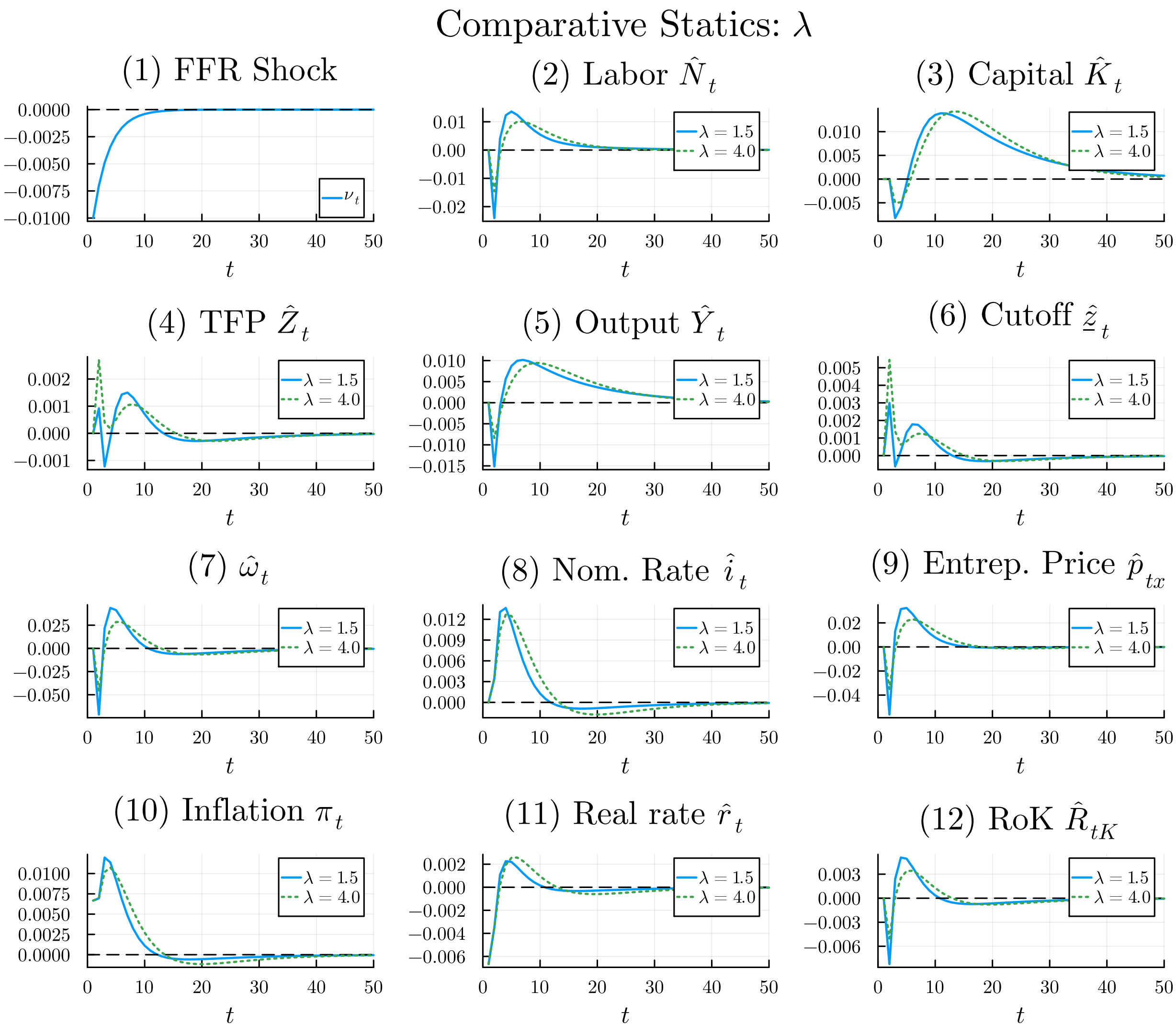


real rates mutes the redistributive effects discussed earlier, resulting in lower peaks for TFP and output in panels 4 and 5 respectively. The impulse responses in Figures 5 and 6 stress my main finding. The effect of monetary policy on aggregates *cannot* be predicted from aggregates alone; across these figures, we see that two economies which produce similar aggregates in the steady state produce different responses to a change in interest rate policy. Instead, the distribution of wealth—and the blend of entrepreneurial shocks and financial market conditions giving rise to this distribution—is crucial to determining both the size of the stimulative effect of looser policy, and how long the effects may be expected to last.

3.3 Further Studying the Responses: The Linearized Economy

An advantage of my formulation is that it allows for many of the effects discussed thus far to be studied analytically. As is often the case in models of monetary policy, it is easier to undertake this study in a linearized version of the model, and so it is here that I now turn my attention. I linearize my model around the zero-inflation steady state described in section 2.5. As a reminder, for a variable W_t, W (without a

Figure 6: Impulse Responses across λ



time subscript) denotes its steady state level, and $\hat{W}_t$ its time-t deviation in logs from that steady-state level. The parsimonious nature of my model makes linearization straightforward, with one small exception. One state variable in my model is the wealth shares $\{s_t(z)\}_{z\in[0,\overline{z}]}$, which is of course a function rather than a single value. In order to study a linearized version of my economy, I need to take a stand on what it means to construct a linear approximation of this state. For this purpose, I use the notion of linearization in the function space using Fréchet derivatives—see, for instance, Childers (2018) and Bhandari et al. (2021). As it turns out, the linearized economy that this approach produces is quite intuitive, and the additional computation is negligible.

I begin with the log-linearized law of motion for the wealth share $s_t(z)$, given by

$$\hat{s}_{t+1}(z) = p\beta R(z)\left\{\hat{R}_t(z) - \hat{R}_{tK} + \hat{s}_t(z)\right\} \tag{70}$$

Equation (70) simply reiterates an earlier intuition: given a path for aggregate real returns, entrepreneurs of productivity z increase their wealth share to the extent that they earn real returns $\hat{R}_t(z)$ in excess of the aggregate return to capital $\hat{R}_{tK}$, which recall is the wealth-weighted average return among all active entrepreneurs. This effect is magnified the greater is the persistence p in returns, and the greater is the entrepreneurs' marginal propensity to save out of current wealth, equal in the steady state to $\beta R(z)$.

Next, linearizing the expression for TFP Z_t in Equation (38) and combining with capital market clearing (39) gives

$$\hat{Z}_t = \frac{\lambda}{Z}\left\{-\underline{z}^2 s(\underline{z})\,\hat{\underline{z}}_t + \int_{\underline{z}}^{\overline{z}} zs(z)\,\hat{s}_t(z)\,dz\right\} \tag{71}$$

Equation (71) demonstrates that to first order, the evolution of total productivity can be decomposed into the sum of two movements: changes in the threshold $\hat{\underline{z}}_t$ and changes in wealth shares $\hat{s}_t(z)$ among entrepreneurs who are active in the steady state. Note that the equation on $\hat{\underline{z}}_t$ is negative; lowering the threshold reduces productivity. This reflects both partial and general equilibrium effects. In partial equilibrium, lowering the threshold for active entrepreneurship necessarily reduces total productivity, which is simply the weighted average of idiosyncratic productivities above the threshold. Additionally, as discussed in Section (3), capital market clearing requires that there always remain a constant share of wealth above the cutoff. As such, a fall in the cutoff must come with a downward shift in wealth shares of the productive entrepreneurs above the cutoff, further reducing aggregate productivity.

It is also through the lens of Equation (71) that we can see analytically the way in which the wealth distribution shapes the size and duration of the response to policy. In the second term, we see that the effect of a change in wealth shares of active entrepreneurs is increasing in the steady-state wealth share-weighted productivity of these entrepreneurs, $zs(z)$. As such, a policy shift that redistributes to higher-z entrepreneurs will be amplified to the extent that the steady-state shares $s(z)$ place more weight on these types to begin with. Put simply: a policy that redistributes wealth to higher types will be more potent if these types are wealthier to begin with. We can see this effect in Figure (5): higher values of persistence lead to a steady-state wealth distribution with more weight $s(z)$ on high-z types, those who earn large excess returns. This idiosyncratic productivity process, and the wealth distribution it creates, engender a larger, longer-lasting response.

Turning to output, the linearized production function has the usual form:

$$\hat{Y}_t = \alpha\left(\hat{Z}_t + \hat{K}_t\right) + (1-\alpha)\,\hat{N}_t \tag{72}$$

This expression for output makes clear the amplification in Figure (5), and the surrounding discussion. If shocks are IID, and wealth and productivity are uncorrelated in the steady state, then $\hat{\underline{z}}_t = \hat{s}_t(z) = 0$ for all t, and Equation (71) immediately implies $\hat{Z}_t = 0$ as well. In this case, Equation (72) becomes $\hat{Y}_t = \alpha\hat{K}_t + (1-\alpha)\,\hat{N}_t$, the standard form with fixed productivity. In this world, there is still amplification through investment: $\hat{K}_t$ will increase, as in e.g. Bernanke et al. (1999), which creates an elevated and long-lived response of output to the policy shift. In the case that $p > 0$, however, the results of Section (3) show that the pattern of redistribution $\hat{s}_t(z)$ is such that $\hat{Z}_t > 0$ for a period following the shock. This creates further amplification of the shock, through an increase in productivity, which also occurs in the data (Christiano et al. 2005, Baqaee et al. 2021). Equations (70) and (71) then imply that this additional amplification also leads to longer-lasting effects: as the shock fades, entrepreneurs slowly spend down their wealth to return to steady state, leaving both investment and productivity elevated for some time after the shock has faded.

The effect of the steady-state wealth distribution on the economy's response to monetary policy can also be seen through the lens of inflation. The linearized Phillips curve is

$$\pi_t = \kappa_p \hat{p}_{tx} + \beta \pi_{t+1} \tag{73}$$

where $\kappa_p > 0$ determines the slope of the Phillips curve as a function of the elasticity of substitution and adjustment costs. Clearing in the markets for labor and entrepreneurial goods together imply that

$$\hat{p}_{tx} = \frac{\alpha + \eta}{1 - \alpha} \hat{Y}_t - \frac{\alpha (\eta + 1)}{1 - \alpha} \left(\hat{Z}_t + \hat{K}_t \right) \tag{74}$$

Combining the two gives

$$\pi_t = \kappa_p \left\{ \frac{\alpha + \eta}{1 - \alpha} \hat{Y}_t - \frac{\alpha (\eta + 1)}{1 - \alpha} \left(\hat{Z}_t + \hat{K}_t \right) \right\} + \beta_f \pi_{t+1} \tag{75}$$

Above and beyond the impact of additional investment, the change in productivity $\hat{Z}_t$—itself a function of reallocation, per Equation (71)—has the effect of *lowering* the Phillips curve, implying less inflation for a given pattern of economic activity. This anti-inflationary effect can be seen in Figure (5): as p increases, peak output following the shock increases but peak inflation remains constant. This is due to the disinflationary effect of redistribution, visible in Equation (75): although more output ($\hat{Y}_t$) in principle pushes inflation up by the standard Phillips Curve logic, this effect is offset by the larger response of $\hat{Z}_t$. The size of this disinflationary force is driven by the response of $\hat{Z}_t$, and so it inherits from Equation (71) its dependence on the initial distribution. The more concentrated is the initial distribution, the larger the gains in productivity from redistribution, and the larger the *downward* shift in the Phillips curve.

4 Conclusion

I argue here two key points concerning the effect of monetary policy on economies with unequal wealth distributions generated by entrepreneurs who earn persistently different returns on their businesses. First, in this framework, redistribution of wealth among entrepreneurs is a key component of the transmission of monetary policy: in particular, a reduction in interest rates ultimately redistributes from low-productivity entrepreneurs to those with higher productivity. Second, the size and duration of the economy's response to monetary policy is determined by the wealth distribution, and the process for entrepreneurs' productivity that generates it. The more persistent are entrepreneurs' idiosyncratic shocks, the more concentrated the wealth distribution will be prior to a shock, and the larger and longer-lasting the response of aggregates to this shock will be.

This paper reconciles two findings in the data: that expansionary monetary policy increases productivity, and that economies are more responsive to monetary policy when wealth inequality is greater. It also has important implications for the *optimal* conduct of monetary policy (see González et al. 2024 for further discussion in a similar framework). My model can also make some headway in explaining empirical evidence that suggests the effects of monetary shocks have changed over time (e.g. Canova and Gambetti, 2009; Boivin et al., 2010), as in the US the concentration of wealth in the hands of the most successful entrepreneurs has increased. The most important implication, in my opinion, is this: my paper contributes to a growing notion in the literature on monetary policy that measuring and predicting responses to changes in interest rates cannot be done by observing aggregates alone, and that distributions play an equally important role. Where many early papers in this strain emphasize the importance of heterogeneities in marginal propensities to *consume*, I stress that marginal propensities to *invest*, and their correlation with wealth, are of equal importance.

References

Sushant Acharya, Edouard Challe, and Keshav Dogra. Optimal monetary policy according to hank. 2020.

S Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684, 1994.

John Asker, Joan Farre-Mensa, and Alexander Ljungqvist. Corporate investment and stock market listing: A puzzle? *The Review of Financial Studies*, 28(2):342–390, 2015.

Adrien Auclert. Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–67, 2019.

Adrien Auclert, Matthew Rognlie, and Ludwig Straub. Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. Technical report, National Bureau of Economic Research, 2020.

David Baqaee, Emmanuel Farhi, and Kunal Sangani. The supply-side effects of monetary policy. Technical report, National Bureau of Economic Research, 2021.

David Rezza Baqaee and Emmanuel Farhi. Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163, 2020.

Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The wealth distribution in bewley economies with capital income risk. *Journal of Economic Theory*, 159:489–515, 2015.

Ben S Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393, 1999.

Anmol Bhandari, David Evans, Mikhail Golosov, and Thomas J Sargent. Inequality, business cycles, and monetary-fiscal policy. *Econometrica*, 89(6):2559–2599, 2021.

Florin Bilbiie. Monetary Policy and Heterogeneity. 2021.

Florin O. Bilbiie. The New Keynesian cross. *Journal of Monetary Economics*, 114:90–108, October 2020. ISSN 03043932. doi: 10.1016/j.jmoneco.2019.03.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S0304393219300492`.

Florin O Bilbiie, Diego R Känzig, and Paolo Surico. Capital and Income Inequality: an Aggregate-Demand Complementarity. 2021.

Corina Boar and Virgiliu Midrigan. Markups and inequality. Technical report, National Bureau of Economic Research, 2019.

Jean Boivin and Marc P Giannoni. Assessing changes in the monetary transmission mechanism: A var approach. *Economic Policy Review*, 8(1), 2002.

Jean Boivin, Michael T Kiley, and Frederic S Mishkin. How has the monetary transmission mechanism evolved over time? In *Handbook of monetary economics*, volume 3, pages 369–422. Elsevier, 2010.

Francisco J Buera and Benjamin Moll. Aggregate implications of a credit crunch: The importance of heterogeneity. *American Economic Journal: Macroeconomics*, 7(3):1–42, 2015.

Francisco J Buera, Joseph P Kaboski, and Yongseok Shin. Finance and development: A tale of two sectors. *American economic review*, 101(5):1964–2002, 2011.

Marco Cagetti and Mariacristina De Nardi. Entrepreneurship, frictions, and wealth. *Journal of political Economy*, 114(5):835–870, 2006.

Fabio Canova and Luca Gambetti. Structural changes in the us economy: Is there a role for monetary policy? *Journal of Economic dynamics and control*, 33(2):477–490, 2009.

Charles T Carlstrom and Timothy S Fuerst. Agency costs, net worth, and business fluctuations: A computable general equilibrium analysis. *The American Economic Review*, pages 893–910, 1997.

Charles T Carlstrom and Timothy S Fuerst. Monetary policy in a world without perfect capital markets. Technical report, 2001.

David Childers. Solution of rational expectations models with function valued states. *Manuscript, Carnegie Mellon University*, 2018.

Lawrence J Christiano, Martin Eichenbaum, and Charles L Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45, 2005.

Andrea Colciago, Anna Samarina, and Jakob de Haan. Central bank policies and income and wealth inequality: A survey. *Journal of Economic Surveys*, 33(4):1199–1231, 2019.

Mariacristina De Nardi, Phil Doctor, and Spencer D Krane. Evidence on entrepreneurs in the united states: Data from the 1989-2004 survey of consumer finances. *Economic Perspectives*, 31(4), 2007.

Jason DeBacker, Vasia Panousi, and Shanthi Ramnath. A risky venture: Income dynamics among pass-through business owners. *American Economic Journal: Macroeconomics*, 15(1):444–474, 2023.

Matthias Doepke and Martin Schneider. Inflation and the redistribution of nominal wealth. *Journal of Political Economy*, 114(6):1069–1097, 2006.

Ethan Feilich. Monetary policy and the dynamics of wealth inequality. 2021.

Irving Fisher. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society*, pages 337–357, 1933.

Xavier Gabaix. Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294, 2009.

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

Beatriz González, Galo Nuño, Dominik Thaler, and Silvia Albrizio. Firm heterogeneity, capital misallocation and optimal monetary policy. 2024.

Emilien Gouin-Bonenfant and Alexis Akira Toda. Pareto extrapolation: An analytical framework for studying tail inequality. *Available at SSRN 3260899*, 2019.

Jeremy Greenwood, Zvi Hercowitz, and Gregory W Huffman. Investment, capacity utilization, and the real business cycle. *The American Economic Review*, pages 402–417, 1988.

Hugo A Hopenhayn. Firms, misallocation, and aggregate productivity: A review. *Annu. Rev. Econ.*, 6(1): 735–770, 2014.

Priit Jeenas. *Firm balance sheet liquidity, monetary policy shocks, and investment dynamics*. Universitat Pompeu Fabra, Department of Economics and Business, 2023.

Greg Kaplan, Benjamin Moll, and Giovanni L Violante. Monetary policy according to hank. *American Economic Review*, 108(3):697–743, 2018.

Rohan Kekre and Moritz Lenel. Monetary policy, redistribution, and risk premia. *Econometrica*, 90(5): 2249–2282, 2022.

Nobuhiro Kiyotaki. Credit and business cycles. *The Japanese Economic Review*, 49(1):18–35, 1998.

Alexander Matusche and Johannes Wacks. Monetary policy and wealth inequality: The role of entrepreneurs. 2021.

Alexander Matusche and Johannes Wacks. Does wealth inequality affect the transmission of monetary policy? *Journal of Macroeconomics*, 75:103474, 2023.

Alisdair McKay, Emi Nakamura, and Jón Steinsson. The power of forward guidance revisited. *American Economic Review*, 106(10):3133–3158, 2016.

Aaron Medlin. Federal reserve monetary policy and wealth inequality: An instrumental-variable local projections approach. *Available at SSRN 4650805*, 2023.

Davide Melcangi and Vincent Sterk. Stock market participation, inequality, and monetary policy. *Review of Economic Studies*, page rdae068, 2024.

Benjamin Moll. Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221, 2014.

Pablo Ottonello and Thomas Winberry. Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502, 2020.

Thomas Piketty and Emmanuel Saez. Inequality in the long run. *Science*, 344(6186):838–843, 2014.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720, 2008.

Julio J Rotemberg. Monopolistic price adjustment and aggregate output. *The Review of Economic Studies*, 49(4):517–531, 1982.

Ali Shourideh and Ariel Zetlin-Jones. External financing and the role of financial frictions over the business cycle: Measurement and theory. *Available at SSRN 2062357*, 2012.

5 Appendix

5.1 Proof of Proposition (5)

Proof. I begin with aggregate output. Given the optimal choice of labor in Lemma 1, the output of an entrepreneur with productivity $z$and capital k_t is

$$y_t(z,k_t) = \frac{\omega_t}{\alpha} z k_t$$

Integrating over entrepreneurs using the stationary distribution $g_t(a,z)$, and using the fact that $k_t(a,z) = \lambda a$ if $z > \underline{z}_t$ and 0 otherwise, gives

$$\begin{aligned} Y_t &= \frac{\omega_t}{\alpha}\lambda \int_{\underline{z}}^{\infty}\int_0^{\infty} a z g_t(a,z)\, da dz \\ &= \frac{\omega_t}{\alpha}\lambda K_t \int_{\underline{z}}^{\infty} z s_t(z)\, dz \\ &= \frac{\omega_t}{\alpha}\lambda K_t X_t \end{aligned}$$

where

$$X_t = \int_{\underline{z}}^{\infty} z s_t(z)\, dz$$

From the labor market, we have

$$\begin{aligned} N_t &= \left(\frac{\omega_t}{\alpha}\right)^{\frac{1}{1-\alpha}} \lambda \int_{\underline{z}}^{\infty}\int_0^{\infty} a z g_t(a,z)\, da dz \\ &= \left(\frac{\omega_t}{\alpha}\right)^{\frac{1}{1-\alpha}} \lambda K_t X_t \end{aligned}$$

which implies

$$\omega_t = \alpha \left(\frac{N_t}{\lambda K_t X_t}\right)^{1-\alpha}$$

so production is

$$\begin{aligned} Y_t &= \frac{\omega_t}{\alpha}\lambda K_t X_t \\ &= \left(\frac{N_t}{\lambda K_t X_t}\right)^{1-\alpha} \lambda K_t X_t \\ &= (\lambda X_t K_t)^{\alpha} N_t^{1-\alpha} \end{aligned}$$

In order to eliminate λ, note that capital market clearing requires

$$
\begin{aligned}
K_t &= \int_0^\infty \int_0^\infty k_t(a,z)\, g_t(a,z)\, dadz \\
&= \int_{\underline{z}}^\infty \int_0^\infty \lambda a g_t(a,z)\, dadz \\
&\downarrow \\
1 &= \lambda \int_{\underline{z}}^\infty s_t(z)\, dz
\end{aligned}
$$

and thus

$$
\lambda = \frac{1}{\int_{\underline{z}}^\infty s_t(z)\, dz}
$$

Replacing this into production gives

$$
Y_t = (Z_t K_t)^\alpha N_t^{1-\alpha}
$$

where

$$
\begin{aligned}
Z_t &= \lambda X_t \\
&= \frac{\int_{\underline{z}}^\infty z s_t(z)\, dz}{\int_{\underline{z}}^\infty s_t(z)\, dz} \\
&= \mathbb{E}_\omega \left[z | z > \underline{z} \right]
\end{aligned}
$$

Now, the law of motion for the aggregate capital stock is

$$
\begin{aligned}
K_{t+1} &= \int \int a_{t+1}(a,z)\, g_t(a,z)\, dadz \\
&= \int \int \beta R_t(z)\, a g_t(a,z)\, dadz \\
&= \beta K_t \int R_t(z)\, s_t(z)\, dz
\end{aligned}
$$

Recall that

$$
R(z_t) = 1 + r_t + \lambda \max\{\omega_t z_t - r_t - \delta, 0\}
$$

and so I can write this as

$$
\begin{aligned}
K_{t+1} &= \beta K_t \int \left[1 + r_t + \lambda \max\{\omega_t z_t - r_t - \delta, 0\}\right] s_t(z)\, dz \\
&= \beta K_t \left(1 + r_t + \lambda \int_{\underline{z}}^\infty (\omega_t z - r_t - \delta)\, s_t dz\right) \\
&= \beta K_t \left(1 + r_t + \lambda \int_{\underline{z}}^\infty z \omega_t s_t(z)\, dz - \lambda (r_t + \delta) \int_{\underline{z}}^\infty s_t dz\right)
\end{aligned}
$$

The two integrals are

$$\omega_t \lambda \int_{\underline{z}}^{\infty} z s_t (z)\, dz = \omega_t X_t \lambda$$
$$= \omega_t Z_t$$

and

$$\lambda \int_{\underline{z}}^{\infty} s_t dz = \lambda \left(1 - S\left(\underline{z}\right)\right)$$
$$= 1$$

and so the LoM is

$$K_{t+1} = \beta K_t \left(1 + \omega_t Z_t - \delta\right)$$

From labor market clearing,

$$\omega_t Z_t = \alpha p_{tx} Z_t^{\alpha} \left(\frac{N_t}{K_t}\right)^{1-\alpha}$$

and so the LoM becomes

$$\begin{aligned} K_{t+1} &= \alpha \beta p_{tx} \left(Z_t K_t\right)^{\alpha} N_t^{1-\alpha} + \beta \left(1 - \delta\right) K_t \\ &= \alpha \beta p_{tx} Y_t + \beta \left(1 - \delta\right) K_t \end{aligned}$$

The return on capital equals the average return across all entrepreneurs, calculated above:

$$\begin{aligned} R_{tK} &= \int_0^{\overline{z}} R_t (z)\, s_t (z) \\ &= 1 - \delta + \omega_t Z_t \\ &= 1 - \delta + \alpha p_{tx} Z_t^{\alpha} \left(\frac{N_t}{K_t}\right)^{1-\alpha} \end{aligned}$$

Finally, the factor prices. The wage can be calculated from labor market clearing: recall that

$$N_t = \left(\frac{1-\alpha}{w_t} p_{tx}\right)^{\frac{1}{\alpha}} \lambda K_t X_t$$

Rearranging gives

$$w_t = \left(1 - \alpha\right) p_{tx} \left(\frac{Z_t K_t}{N_t}\right)^{\alpha}$$

as in the text. The net real interest rate comes from the definition of the cutoff $\underline{z}_t$:

$$r_t = \omega_t \underline{z}_t - \delta$$

Substituting the definition of ω_t from labor market clearing gives the form in Equation (41). Note that this only holds in expectation, as nominal debt contracts are negotiated at the end of period t, and the ex-post real return r_t depends on the realization of inflation. □

5.2 Proof of Proposition (6)

Proof. By definition, the law of motion for the *cumulative* wealth share $S_t(z)$ is

$$\begin{aligned} S_{t+1}(z) = & \frac{1}{K_{t+1}} p \int_0^z \int_0^\infty a'(a,\hat{z})\, g_t(a,\hat{z})\, da d\hat{z} + \\ & \frac{1}{K_{t+1}} (1-p) F(z) \int_0^\infty \int_0^\infty a'(a,\hat{z})\, g_t(a,\hat{z})\, da d\hat{z} \end{aligned}$$

Differentiating with respect to z gives

$$\begin{aligned} S'_{t+1}(z) = & \frac{1}{K_{t+1}} p \int_0^\infty a'(a,z)\, g_t(a,z)\, da + \\ & \frac{1}{K_{t+1}} (1-p) f(z) \int_0^\infty \int_0^\infty a'(a,\hat{z})\, g_t(a,\hat{z})\, da d\hat{z} \end{aligned}$$

With the policy functions:

$$\begin{aligned} s_{t+1}(z) = & \frac{1}{K_{t+1}} p \int_0^\infty \beta R(z)\, a g_t(a,z)\, da + \\ & \frac{1}{K_{t+1}} (1-p) f(z) \int_0^\infty \int_0^\infty \beta R(\hat{z})\, a g_t(a,\hat{z})\, da d\hat{z} \\ = & \frac{K_t}{K_{t+1}} p \beta R(z)\, s_t(z) + \\ & \frac{K_t}{K_{t+1}} (1-p) f(z) \int_0^\infty \beta R(\hat{z})\, s_t(\hat{z})\, d\hat{z} \end{aligned}$$

Leibniz:

$$\frac{d}{dz} \int_0^z \int_0^\infty \beta R(\hat{z})\, a g_t(a,\hat{z})\, da d\hat{z} = \int_0^\infty \beta R(z)\, a g_t(a,z)\, da$$

Furthermore,

$$\int R(z)\, s_t(z)\, dz = 1 - \delta + \alpha p_{tx} Z_t^\alpha \left(\frac{N_t}{K_t}\right)^{1-\alpha}$$

as derived above, i.e. the RoK. So the LoM is

$$s_{t+1}(z) = \frac{K_t}{K_{t+1}} \beta \left[pR(z)\, s_t(z) + (1-p) f(z) \left(1 - \delta + \alpha p_{tx} Z_t^\alpha \left(\frac{N_t}{K_t}\right)^{1-\alpha}\right)\right] \tag{76}$$

Recall that from the law of motion for capital,

$$\frac{K_{t+1}}{\beta K_t} = 1 - \delta + \alpha p_{tx} Z_t^\alpha \left(\frac{N_t}{K_t}\right)^{1-\alpha} = R_{tK}$$

Substituting this into (76) gives Equation (49) in the text. □

5.3 Proof of Proposition (8)

Proof. From the law of motion for wealth shares in Equation (49), if $p = 0$ then $s_t(z) = f(z)$ for all z, and then the (fixed) value for Z follows from the definition in (38). If $p = 1$, meanwhile, then the steady-state expression becomes

$$s(z) = \beta R(z)\, s_t(z)$$

There are two cases where this is possible: either $s(z) = 0$, or $R(z) = R_K = \beta^{-1}$. The wealth shares consistent with both this condition and capital market clearing are

$$s_t(z) = \begin{cases} 0 & z < \overline{z} \\ \infty & z = \overline{z} \end{cases}$$

In this case, $Z = \overline{z}$, and

$$\begin{aligned} \beta R(z) &= \beta \left\{1 + r + \lambda \left[\omega \overline{z} - r - \delta\right]\right\} \\ &= \beta \left\{1 + \alpha Z^{\alpha} \left(\frac{N}{K}\right)^{1-\alpha} - \delta \right\} \\ &= 1 \end{aligned}$$

as expected. Next, consider the steady-state wealth shares

$$s(z) = \frac{1-p}{1 - \beta p R(z)} f(z)$$

In the steady state, returns $R(z)$ are given by

$$\begin{aligned} R(z) &= 1 - \delta + \alpha p_x Z^{\alpha} \left(\frac{N}{K}\right)^{1-\alpha} \frac{z_{\lambda}(z)}{Z} \\ &= 1 - \delta + \left(\frac{1}{\beta} - (1-\delta)\right) \frac{z_{\lambda}(z)}{Z} \end{aligned}$$

where the second line follows from the definition of the long-run return to capital. As a corollary,

$$R_p(z) = \frac{dR(z)}{dp} = \left(\frac{1}{\beta} - (1-\delta)\right) \frac{-Z(\lambda-1)\frac{d\underline{z}}{dp} - z + (\lambda - 1)(z - \underline{z})\frac{dZ}{dp}}{Z^2} < 0$$

An increase in persistence decreases excess returns for entrepreneurs. Next,

$$\frac{ds(z)}{dp} = \frac{\beta p R(z) - 1 + (1-p)\left(\beta R(z) + \beta p R_p(z)\right)}{\left[1 - \beta p R(z)\right]^2} f(z)$$

The denominator is always positive. The numerator is positive if

$$\begin{gathered} \beta p R(z) - 1 + (1-p)\left(\beta R(z) + \beta p R_p(z)\right) > 0 \\ \Updownarrow \\ R(z) - \frac{1}{\beta} > -p(1-p) R_p(z) \end{gathered}$$

Substituting for $R(z)$ and the return to capital β^{-1}, along with $R_p(z)$, gives after some simplification

$$z_{\lambda}(z) - Z > p(1-p)\left\{(\lambda - 1)\frac{d\underline{z}}{dp} + \frac{z_{\lambda}(z)}{Z}\frac{dZ}{dp}\right\}$$

which is Equation (60) in the text. □