Firm Heterogeneity, Capital Misallocation and Optimal Monetary Policy*

Beatriz González — BdE
Galo Nuño — BdE
Dominik Thaler — ECB
Silvia Albrizio — IMF

First version: October 2020. This version: May 2022

Abstract

We analyze monetary policy in a New Keynesian model with heterogeneous firms and financial frictions. Firms differ in their productivity and net worth and face collateral constraints that cause capital misallocation. TFP endogenously depends on the time-varying distribution of firms. A monetary expansion increases the investment of constrained firms with a high marginal revenue product of capital (MRPK) relatively more than that of low-MRPK ones, crowding out the latter and increasing TFP. We provide empirical evidence based on Spanish granular data supporting this mechanism. This has important implications for optimal monetary policy design. First, a central bank without pre-commitments engineers an unexpected monetary expansion to increase TFP in the medium run. Second, the divine coincidence holds after a demand shock. Third, if nominal rates are constrained by the zero lower bound, the optimal policy prescribes that rates should remain low for much longer than under complete markets.

Keywords: Monetary policy, firm heterogeneity, financial frictions, misallocation.
JEL classification: E12, E22, E43, E52, L11.

---

*Beatriz González (Banco de España): beatrizgonzalez@bde.es. Galo Nuño (Banco de España): galo.nuno@bde.es. Dominik Thaler (European Central Bank): dominik.thaler@eui.eu. Silvia Albrizio (International Monetary Fund): salbrizio@imf.org. Two version of this paper were previously circulated as “Monetary Policy and Capital Misallocation” and “Optimal Monetary Policy with Heterogeneous Firms”. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the Bank of Spain or the Eurosystem. We are specially grateful to Anmol Bhandari, Saki Bigio and Jesús Fernández-Villaverde for their suggestions. We thank Klaus Adam, Paco Buera, Jim Costain, Luis Franjo, Veronica Guerrieri, Masashige Hamano, Ben Moll, Morten Ravn, Ricardo Reis, Gianluca Violante, Chris Wolf, and Francesco Zanetti for excellent comments, as well as participants at several seminars and conferences. The opinions and analysis are the responsibility of the authors, and therefore, do not necessarily coincide with those of the International Monetary Fund, the ECB, Banco de España or the Eurosystem. All errors are our own.

1 Introduction

Firms' investment decisions are one of the key transmission channels of monetary policy. In the presence of firm heterogeneity and financial frictions, the distribution of capital across firms matters for aggregate productivity, as the literature on capital misallocation documents. This opens the door to the possibility of monetary policy affecting productivity through its impact on the endogenous investment decisions of firms, which raises important questions for policymakers. First, what are the channels through which monetary policy affects capital misallocation and endogenous TFP? Second, how do these channels modify the optimal conduct of monetary policy? To answer these questions, we introduce a framework that combines the workhorse model of monetary policy – the New Keynesian model – with a tractable model of firm heterogeneity in which capital misallocation arises from financial frictions.

We consider an economy populated by a continuum of firms owned by entrepreneurs, who have access to a constant returns to scale technology. Entrepreneurs are heterogeneous in their net worth and receive idiosyncratic productivity shocks. They face financial frictions, as they can only borrow subject to a collateral constraint, which restricts their borrowing to a multiple of their net worth. Entrepreneurs above a certain idiosyncratic productivity threshold are constrained: they rent as much capital as possible since their marginal revenue product of capital (MRPK) is higher than their cost of capital. Entrepreneurs below the threshold are unconstrained: their optimal size is zero and they decide to lend their net worth to other entrepreneurs. This can be seen as the limit case of an economy with decreasing returns to scale at the individual level, in which unconstrained firms are optimally very small and the bulk of production is carried out by constrained firms. We embed this heterogeneous-firm block into an otherwise standard continuous-time New Keynesian model. This economy allows for an aggregate representation akin to that in the standard New Keynesian model with capital, except that in this case aggregate Total Factor Productivity (TFP) is endogenous and depends on the distribution of capital across firms.

Our model nests the complete markets New Keynesian model as a particular case. If the borrowing limit disappears, the most productive firm carries out all the production. In this case TFP is exogenous and the distribution of net worth across firms does not affect macroeconomic variables.[1]

[1] To be precise, the mean of the distribution, which pins down aggregate capital does, but given a

Monetary policy affects aggregate TFP in our incomplete-market baseline model by modifying both the distribution of net-worth across entrepreneurs and the productivity threshold above which entrepreneurs are constrained. We jointly refer to these two mechanism as the "misallocation channel of monetary policy". We calibrate the model to the US and compute the response to a monetary policy shock when the central bank follows a Taylor rule. The response of the baseline economy features the same demand driven expansion as the complete markets economy. However, in the baseline economy, the shock additionally causes an increase in TFP. This response is in line with empirical evidence (Jordà et al. (2020), Moran and Queralto, 2018; Meier et al., 2020; or Baqaee et al., 2021), which has explained the link between monetary policy and TFP through different mechanisms such as endogenous growth through R&D or markup heterogeneity. Here, instead, this is the consequence of changes in the capital distribution across firms. An expansionary monetary policy shock leads high-productivity firms to increase their investment relatively more than low-productivity ones, crowding out the latter. This increases the market share of high-productivity firms, thus reducing capital misallocation.

We present empirical evidence supporting this mechanism. We use micro panel data of the quasi-universe of Spanish firms during the period 2000-2016, and construct the monetary policy shocks using the high-frequency event-study approach of Jarociński and Karadi (2020). We then estimate to what extent the firms' investment response to monetary policy shocks depends on firms' productivity, using an empirical specification that follows Ottonello and Winberry (2020) closely. We find that high-productivity firms invest more relative to low-productivity ones in response to an expansionary monetary policy shock. [2]

We turn next to the normative prescriptions of the model. We analyze the Ramsey problem of a benevolent central bank. This is a nontrivial endeavor, as the net-worth distribution is an infinite-dimensional state in the central bank's problem. We introduce a new algorithm that leverages the computational advantages of continuous time. The idea is to discretize the continuous-time, continuous-state problem into a discrete-time, discrete-space one as in Achdou et al., 2017, and then to use symbolic differentiation to

level of aggregate capital, the shape of the distribution is irrelevant at the macro level.

[2]These findings are in line with those of Caglio et al. (2021) for the US, who find that an expansionary monetary policy shock increases highly levered SMEs' demand for credit and their borrowing capacity.

obtain the first-order conditions of the central bank. This produces a high-dimensional nonlinear dynamic system, which is efficiently solved using a Newton algorithm in the sequence space.

The steady-state of the Ramsey plan features zero inflation, as in the complete-markets case. We study Ramsey optimal policy in the absence of shocks when the initial state coincides with the zero-inflation steady state and the central bank has no pre-commitments. Whereas the optimal policy in the case with complete markets is time consistent, financial frictions introduce a new source of time inconsistency as the central bank engineers a temporary monetary expansion in the short run while committing to price stability in the long-run. This strategy allows the central bank to exploit the misallocation channel, as the monetary expansion gives rise to a temporary, endogenous increase in TFP. The desire of the central bank to redistribute resources towards productive entrepreneurs in order to promote firm growth is reminiscent of the case with optimal fiscal policy analyzed by Itskhoki and Moll (2019), who find how optimal fiscal policy initially redistributes from households towards entrepreneurs to promote capital accumulation. In our case, and given the lack of time-varying fiscal instruments, it is the central bank that engineers this redistribution through an expansion in aggregate demand.

We analyze next the optimal response of monetary policy to shocks from a 'timeless perspective' (Woodford, 2003), in which the central bank adopts a behavior to which it would have wished to commit itself to at a date far in the past. We consider a temporary demand shock that reduces households' intertemporal discount rate. In this case, the optimal response is price stability (zero inflation). This is the same policy as with complete markets. The "divine coincidence" (Gali, 2008) thus extends to our model. The implementation of this policy, however, differs in both cases. Under incomplete markets, the demand shock leads to an endogenous fall in TFP through the misallocation channel, in line with the findings of Gopinath et al. (2017) or Asriyan et al. (2021). The fall in TFP, in turn, amplifies the reduction of the natural rate brought about by the demand shock itself, such that the natural rate drops more than in the case with complete markets. As real rates mimic natural rates, the result is that real rates decline more, and more persistently, than in the standard New Keynesian model.

Finally, we analyze timeless optimal monetary policy when nominal rates are constrained by the zero lower bound (ZLB). We consider again a demand shock due to a decline in households' discount rates. The optimal policy, originally proposed by Eggertsson and Woodford (2004), is "low for longer": nominal rates should remain at the ZLB longer that what would be prescribed if the ZLB were not present. In the case with incomplete markets, the larger and more persistent decline in natural rates due to the endogenous fall in TFP leads to what we call a "low for even longer" optimal policy, as nominal rates should remain at the ZLB for more than twice the time they do under complete markets.

Related literature. This paper contributes to several strands of the literature. First, we contribute to the emerging literature on the role of financial frictions and firm heterogeneity in monetary policy transmission. Ottonello and Winberry (2020) analyze the effect of monetary policy on firm investment in a model with endogenous default. They find that low-risk firms are more responsive to monetary shocks because they face a flatter marginal cost curve for financing investment. Jeenas (2020) analyzes the role of firms' balance sheet liquidity in the transmission of monetary policy to investment. Koby and Wolf (2020) study the conditions under which the lumpiness of firm-level investment matters for aggregate investment dynamics and, as an application, analyze monetary policy transmission with heterogeneous firms. We contribute to this nascent literature on two fronts. First, we focus on the link between monetary policy and capital misallocation. Second, we analyze optimal monetary policy in a model with non-trivial firm heterogeneity.[3]

Second, our model is related to the extensive literature on capital misallocation, and the different channels that may affect it, such as Hsieh and Klenow (2009) or Midrigan and Xu (2014) – see Restuccia and Rogerson (2017) for a review on this literature. Our paper builds on Moll (2014), who introduces a heterogeneous producer model to study how the nature of the idiosyncratic shocks impacts the speed of transitions. We enrich this model by introducing aggregate capital adjustment costs and a New Keynesian monetary block since our focus is to understand how monetary policy affects aggregates through its impact on heterogeneous firms.[4] Focusing on the impact of lower interest rates in a small open economy, Reis (2013) and Gopinath et al. (2017) analyze

[3]Other strands of the literature have analyzed the links between monetary policy and firm heterogeneity through heterogeneity in markups and entry-exit (e.g. Meier et al., 2020, Bilbiie et al., 2014, Zanetti and Hamano (2020), Andrés et al., 2021, Nakov and Webber, 2021 or Baqaee et al., 2021), in cyclicality (David and Zeke, 2021) or in firm-level productivity trends (Adam and Weber, 2019).

[4]Buera and Nicolini (2020) employ a discrete-time version of Moll (2014) with cash-in-advance constraints to analyze the impact of different monetary and fiscal policies after a credit crunch.

how an exogenous increase in the availability of cheap foreign funds or an exogenous decrease in real interest rates may increase capital misallocation among firms facing financial frictions. Asriyan et al. (2021) extend these results to a general equilibrium environment. Acharya et al. (2021) analyze the links between zombie lending and monetary policy. Here, instead, we focus on the interactions between monetary policy and capital misallocation in a nominal economy with price rigidities.

Third, to the best of our knowledge this is the first paper to analyze optimal policy at the ZLB with heterogeneous firms. We show that the "low for longer" prescription highlighted by the literature under complete markets (Eggertsson and Woodford, 2004; Adam and Billi, 2006 or Nakov et al., 2008) is amplified due to the interaction between the misallocation channel and the optimal path of interest rates.

Finally, we add to the literature analyzing optimal monetary policies in models with heterogeneous agents. Nuño and Thomas (2016), Bilbiie and Ragot (2020), Bhandari et al. (2021), Acharya et al. (2019), Bigio and Sannikov (2021), Dávila and Schaab, 2022 and Le Grand et al. (2020) analyze optimal monetary policy in models with heterogeneous households using different techniques. Here, instead, our focus is on heterogeneous firms. We propose a novel methodology to compute optimal policies nonlinearly in models featuring non-trivial heterogeneity, including exogenous borrowing limits or other nonlinear features. Our algorithm is simple to code and can be easily generalized to other problems. It can be implemented using several available software packages. In our case, we employ Dynare.

2 Model

We propose a New Keynesian closed economy model with financial frictions and heterogeneous firms based on Moll (2014). Time is continuous and there is no aggregate uncertainty. Later we discuss how we introduce aggregate shocks. The economy is populated by five types of agents: households, the central bank, input-good firms, retail, and final goods producers. The representative household is composed of two types of members: workers and entrepreneurs. Workers rent their labor whereas entrepreneurs operate the input good firms, which combine capital and labor to produce the input good. Entrepreneurs are heterogeneous in their net worth and productivity. The input good is differentiated by imperfectly competitive retail goods producers facing sticky prices, whose output is aggregated by the final goods producer. The latter two firms are standard in New Keynesian models.

2.1 Heterogeneous input good firms

There is a continuum of entrepreneurs. Each entrepreneur owns some net worth, which they hold in units of capital. They can use this capital for production in their own input-good producing firm – firm for short – or rent it out to other entrepreneurs. Similar to Gertler and Karadi (2011), we assume that entrepreneurs are members of the representative household, to whom they may transfer dividends.[5]

Entrepreneurs are heterogeneous in two dimensions: their net worth a_t and in their idiosyncratic productivity z_t.[6] Each entrepreneur owns a technology which uses capital k_t and labor l_t to produce input good y_t:

$$y_t = f_t(z_t, k_t, l_t) = (z_t k_t)^{\alpha}(l_t)^{1-\alpha}. \tag{1}$$

The labor share $\alpha \in (0, 1)$ is the same across entrepreneurs. Idiosyncratic productivity z_t follows a diffusion process,

$$dz_t = \mu(z_t)dt + \sigma(z_t)dW_t, \tag{2}$$

where $\mu(z)$ is the drift and $\sigma(z)$ the diffusion of the process.

Entrepreneurs can use their technology to produce or not. If they do, we say they run a firm and call them active. If they do not, they lend their net worth to firms owned by other entrepreneurs. Firms hire workers at the real wage w_t and rent capital at the real rental rate of capital R_t. Capital is rented from the agents which save, i.e. both households and inactive entrepreneurs. Firms sell the input good at the real price $m_t = p_t^y/P_t$, which is the inverse of the gross markup associated to retail products over input goods, being p_t^y the nominal price of the input good and P_t the price of the final good, i.e. the numeraire. Entrepreneurs use the return on their activities to distribute (non-negative) dividends d_t to the household and to invest in additional capital at the

[5]This assumption is the only relevant difference between the real side of our model and the model of Moll (2014). We consider it to avoid having to deal with redistributive issues between households and entrepreneurs when analyzing optimal monetary policy. Both models produce linear dividend policies, so they can be seen as equivalent from a positive perspective.

[6]For notational simplicity, we use x_t instead of $x(t)$ for the variables depending on time. Furthermore, we suppress the input goods firm's index.

real price q_t. Capital depreciates at rate δ. An entrepreneur's flow budget constraint can be expressed as follows

$$\dot{a}_t = \frac{1}{q_t}\left[\underbrace{m_t f_t(z_t, k_t, l_t) - w_t l_t - R_t k_t}_{\text{Firm's profits}} + \underbrace{(R_t/q_t - \delta)}_{\text{Return on net worth}} q_t a_t - \underbrace{d_t}_{\text{Dividends}}\right]. \tag{3}$$

Note that we have rearranged the budget constraint to yield the law of motion of net worth *in units of capital.*

Entrepreneurs can borrow additional capital $b_t = k_t - a_t$ to use in production. However, they face a collateral constraint, such that the value of capital used in production cannot exceed $\gamma > 1$ of their net worth,

$$q_t k_t \leq \gamma q_t a_t. \tag{4}$$

Entrepreneurs retire and return to the household according to an exogenous Poisson process with arrival rate η. Upon retirement they pay all their net worth, valued $q_t a_t$, to the household, and they are replaced by a new entrepreneur with the same productivity level. Entrepreneurs maximize the discounted flow of dividends, which is given by

$$V_0(z, a) = \max_{k_t, l_t, d_t} \mathbb{E}_0 \int_0^{\infty} e^{-\eta t} \Lambda_{0,t} \left(\underbrace{d_t}_{\text{Dividends}} + \eta \underbrace{q_t a_t}_{\text{Terminal value}} \right) dt, \tag{5}$$

subject to the budget constraint (3), the collateral constraint (4), and the process followed by productivity (2). Future profits are discounted by the household's stochastic discount factor $\Lambda_{0,t}$. Below we show that $\Lambda_{0,t} = e^{-\int_0^t r_s ds}$, where r_t is the real interest rate.

We can split the entrepreneurs' problem into two parts: a static profit maximization problem and a dynamic dividend-distribution problem. First, entrepreneurs maximize firm profits given their productivity and net worth,

$$\max_{k_t, l_t} \left\{ m_t f_t(z_t, k_t, l_t) - w_t l_t - R_t k_t \right\}, \tag{6}$$

subject to the collateral constraint (4). Since the production function has constant returns to scale, entrepreneurs find it optimal to operate a firm at the maximum scale defined by the collateral constraint whenever their idiosyncratic productivity is high enough. Else they remain inactive, because they cannot run a profitable firm given their low productivity. Factor demands and profits of operating firms are thus linear in net worth, and there exists a productivity cut-off z_t^* below which the optimal size of the firm is $k^*(z) = 0$, so this entrepreneur is unconstrained. From now on, we refer to active entrepreneurs as 'constrained' and to inactive as 'unconstrained'.

Firm's demand for capital and labor is :

$$k_t(z_t, a_t) = \begin{cases} \gamma a_t, & \text{if } z_t \geq z_t^*, \\ 0, & \text{if } z_t < z_t^*, \end{cases} \tag{7}$$

$$l_t(z_t, a_t) = \left(\frac{(1-\alpha) m_t}{w_t} \right)^{1/\alpha} z_t k_t(z_t, a_t). \tag{8}$$

Firm's profits are then given by

$$\Phi_t(z_t, a_t) = \max \{ z_t \varphi_t - R_t, 0 \} \gamma a_t, \quad \text{where} \quad \varphi_t = \alpha \left(\frac{(1-\alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}, \tag{9}$$

and the productivity cut-off, above which firms are profitable, is given by

$$z_t^* \varphi_t = R_t. \tag{10}$$

Second, entrepreneurs decide the dividends d_t that they pay to the household. Using (9), the law of motion of an entrepreneur's net worth (in units of capital) (3) can be rewritten as

$$\begin{aligned} \dot{a}_t &= \frac{1}{q_t} \left[\Phi_t(z_t, a_t) + (R_t - \delta q_t) a_t - d_t \right] \\ &= \frac{1}{q_t} \left[(\gamma \max \{ z_t \varphi_t - R_t, 0 \} + R_t - \delta q_t) a_t - d_t \right]. \end{aligned} \tag{11}$$

The solution to this problem is derived in Appendix A.1. There we show how entrepreneurs never distribute dividends until retirement, $d_t = 0$, when they bring all their net worth home to the household. The intuition is the following. The return on one unit of capital in the hands of the entrepreneur is *at least* $(R_t - \delta q_t)$, while for the household the return of this unit of capital is *exactly* $(R_t - \delta q_t)$. It is thus always worthwhile for entrepreneurs to keep their funds. The household collects all these funds as dividends once the entrepreneur retires. To keep things simple, we assume the representative household uses a fraction $\psi \in (0,1)$ of these dividends to finance the net worth of the new entrepreneurs, so net dividends are $(1-\psi)$ of the net worth of retiring entrepreneurs.

2.2 Households

There is a representative household, composed of workers and entrepreneurs, that saves in capital D_t or in nominal instantaneous bonds whose real value is denoted by B_t^N. Nominal bonds B_t^N are in zero net supply. Workers supply labor L_t. The household maximizes

$$W_t = \max_{C_t, L_t, B_t^N, D_t} \int_0^\infty e^{-\rho_t^h t} u(C_t, L_t) dt. \tag{12}$$

$$\begin{aligned} s.t. \quad \dot{D}_t q_t &= (R_t - \delta q_t) D_t - S_t^N - C_t + w_t L_t + T_t, \\ \dot{B_t^N} &= (i_t - \pi_t) B_t^N + S_t^N, \end{aligned} \tag{13}$$

where S_t^N is the investment into nominal bonds and T_t are the profits received by the household, which is the sum of the profits of the capital producer ($[\iota_t q_t - \iota_t - \Xi(\iota_t)] K_t$), the profits from retail goods producers (Π_t from equation 21) and net dividends received from entrepreneurs ($(1-\psi)\eta q_t A_t$).

We assume separable utility of CRRA form, i.e., $u(C_t, L_t) = \frac{C_t^{1-\zeta}}{1-\zeta} - \Upsilon \frac{L_t^{1+\vartheta}}{1+\vartheta}$. Solving this problem (see Appendix (A.4) for details), we obtain the Euler equation,

$$\frac{\dot{C}_t}{C_t} = \frac{r_t - \rho_t^h}{\zeta}, \tag{14}$$

the labor supply condition

$$w_t = \frac{\Upsilon L_t^{\vartheta}}{C_t^{-\zeta}}, \tag{15}$$

and the Fisher equation

$$r_t = i_t - \pi_t, \tag{16}$$

where, for convenience, we have made use of the following definition of the real rate of interest

$$r_t \equiv \frac{R_t - \delta q_t + \dot{q}_t}{q_t}, \tag{17}$$

which equals the real return on capital adjusted by capital gains and depreciation.

Integrating the Euler equation (14), we can verify that the stochastic discount factor $\Lambda_{0,t}$ can be defined as

$$\Lambda_{0,t} \equiv e^{-\int_0^t \rho_t^h ds} \frac{u_c'(C_t)}{u_c'(C_0)} = e^{-\int_0^t r_s ds}.$$

2.3 Final good producers

As usual in New Keynesian models, a competitive representative final goods producer aggregates a continuum of output produced by retailer $j \in [0,1]$,

$$Y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{18}$$

where $\varepsilon > 0$ is the elasticity of substitution across goods. Cost minimization implies

$$y_{j,t}(p_{j,t}) = \left(\frac{p_{j,t}}{P_t} \right)^{-\varepsilon} Y_t, \text{ where } P_t = \left(\int_0^1 p_{j,t}^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}}.$$

2.4 Retailers

Following Ottonello and Winberry (2020) and Jeenas (2020), we differentiate between heterogeneous input-good firms and retailers. We assume that monopolistic competition occurs at the retail level. Retailers purchase input goods from the input-good firms, differentiate them and sell them to final good producers. Each retailer j chooses the sales price $p_{j,t}$ to maximize profits subject to price adjustment costs as in Rotemberg (1982), taking as given the demand curve $y_{j,t}(p_{j,t})$ and the price of input goods, p_t^y. We assume the government pays a proportional constant subsidy τ on input good, so that the net real price for the retailer is $\tilde{m}_t = m_t(1-\tau)$. This subsidy is financed by a lump-sum tax on the retailers Ψ_t.[7] The adjustment costs are quadratic in the rate of price change $(\dot{p}_{j,t}/p_{j,t})$ and expressed as a fraction of output (Y_t),

$$\Theta_t \left(\frac{\dot{p}_{j,t}}{p_{j,t}} \right) = \frac{\theta}{2} \left(\frac{\dot{p}_{j,t}}{p_{j,t}} \right)^2 Y_t,$$

where $\theta > 0$. Suppressing notational dependence on j, each retailers chooses $\{p_t\}_{t\geq 0}$ to maximize the expected profit stream, discounted at the stochastic discount factor of

[7] This fiscal scheme is introduced to eliminate the distortions caused by imperfect competition in steady state, as common in the optimal policy literature.

the household,

$$\int_0^\infty \Lambda_{0,t}\left[\Pi_t\left(p_t\right)-\Theta_t\left(\frac{\dot{p}_t}{p_t}\right)\right]dt, \tag{19}$$

where

$$\Pi_t\left(p_t\right)=\left(\frac{p_t}{P_t}-\tilde{m}_t\right)\left(\frac{p_t}{P_t}\right)^{-\varepsilon}Y_t-\Psi_t$$

are per-period profits gross of price adjustment costs.

The symmetric solution to the pricing problem yields the New Keynesian Phillips curve (see Appendix A.2), which is given by

$$\left(r_t-\frac{\dot{Y}_t}{Y_t}\right)\pi_t=\frac{\varepsilon}{\theta}\left(\tilde{m}_t-m^*\right)+\dot{\pi}_t, \qquad m^*=\frac{\varepsilon-1}{\varepsilon}. \tag{20}$$

where π_t denotes the inflation rate $\pi_t=\dot{P}_t/P_t$. We exploit the fact that, given the lack of aggregate risk, the household's stochastic discount factor can be expressed as $\Lambda_{0,t}=e^{-\int_0^t r_s ds}$ (see derivation in Section 2.2). The total profit of retailers, net of the lump-sum tax, which is transferred to the households lump sum, is

$$\Pi_t=\left(1-m_t\right)Y_t-\frac{\theta}{2}\pi_t^2Y_t. \tag{21}$$

2.5 Capital producers

A representative capital producer owned by the representative household produces capital and sells it to the household and the firms at price q_t, which she takes as given. Her cost function is given by $\left(\iota_t+\Xi\left(\iota_t\right)\right)K_t$ where ι_t is the investment rate and $\Xi\left(\iota_t\right)$ is a capital adjustment cost function. She maximizes the expected profit stream, discounted at the stochastic discount factor of the household. Profits are paid to the household.

$$W_t=\max_{\iota_t,K_t}\int_0^\infty \Lambda_{0,t}\left(q_t\iota_t-\iota_t-\Xi\left(\iota_t\right)\right)K_t dt. \tag{22}$$

$$\text{s.t.} \quad \dot{K}_t \;=\; \left(\iota_t-\delta\right)K_t. \tag{23}$$

The optimality conditions imply (see Appendix A.3)

$$r_t = (\iota_t - \delta) + \frac{\dot{q}_t - \Xi''(\iota_t)\, i_t}{q_t - 1 - \Xi'(\iota_t)} - \frac{q_t \iota_t - \iota_t - \Xi(\iota_t)}{q_t - 1 - \Xi'(\iota_t)}.$$

2.6 Distribution

As previously explained, we assume that, for each entrepreneur returning to the household, a new entrepreneur arrives operating the same technology, that is, with the same productivity level. This new entrepreneur receives a startup capital stock from the household in a lump-sum fashion, equal to a fraction $\psi < 1$ of the net worth of the entrepreneur she replaces. Let $G_t(z,a)$ be the joint distribution of net worth and productivity. The evolution of its density $g_t(z,a)$ is given by the Kolmogorov Forward equation

$$\frac{\partial g_t(z,a)}{\partial t} = \underbrace{-\frac{\partial}{\partial a}[g_t(z,a)s_t(z)a]}_{\text{Retained earnings}} \underbrace{-\frac{\partial}{\partial z}[g_t(z,a)\mu(z)] + \frac{1}{2}\frac{\partial^2}{\partial z^2}[g_t(z,a)\sigma^2(z)]}_{\text{Productivity changing randomly}} \underbrace{-\eta g_t(z,a)}_{\text{Entrepreneurs retiring}} \underbrace{+\frac{\eta}{\psi} g_t(z, \frac{a}{\psi}))}_{\text{Entrepreneurs entering}}, \quad (24)$$

where $s_t(z)$ is the entrepreneurs' investment rate (11)

$$s_t(z) \equiv \frac{1}{q_t}(\underbrace{\gamma \max\{z_t\varphi_t - R_t, 0\}}_{\text{Profit rate from operating the firm}} + R_t - \delta q_t), \quad (25)$$

and $1/\psi g_t(z, a/\psi)$ is the density of new entrepreneurs entering.

Using this two-dimensional distribution we can define the one-dimensional distribution of *net-worth shares* as $\omega_t(z) \equiv \frac{1}{A_t}\int_0^\infty a g_t(z,a)da$. This distribution measures the share of net worth held by entrepreneurs with productivity z. It contains all the relevant information in a more compact form, which is why we shall work with it. Given this definition and the structure of the problem, net-worth shares are non-negative, continuous, once differentiable everywhere and they integrate up to 1. The law of motion of net worth shares is given by (see in Appendix A.5)

$$\frac{\partial \omega_t(z)}{\partial t} = \left[s_t(z) - \frac{\dot{A}_t}{A_t} - (1-\psi)\eta \right] \omega_t(z) - \frac{\partial}{\partial z}\mu(z)\omega_t(z) + \frac{1}{2}\frac{\partial^2}{\partial z^2}\sigma^2(z)\omega_t(z). \tag{26}$$

2.7 Market Clearing and Aggregation

Market clearing. Define aggregate capital used in production as $K_t = \int k_t(z,a)dG_t(z,a)$, aggregate firm net worth as $A_t = \int a_t dG_t(z,a)$, and aggregate net debt as $B_t = \int b_t(z,a)dG_t(z,a)$. Since the capital borrowed by an individual entrepreneur equals that used in production minus its net worth $b_t = k_t - a_t$, we have that

$$K_t = A_t + B_t, \tag{27}$$

Asset market clearing requires that net borrowing of entrepreneurs B_t equals net savings of the household D_t,

$$B_t = D_t. \tag{28}$$

Let $\Omega(z)$ be the cumulative distribution of net-worth shares, i.e. $\Omega_t(z) = \int_0^z \omega_t(x)\,dx$. By combining equations (27), (28), aggregating capital used by firms (7), and solving for A_t, we obtain

$$A_t = \frac{D_t}{\gamma(1-\Omega_t(z_t^*)) - 1}. \tag{29}$$

Labor market clearing implies

$$L_t = \int_0^\infty l_t(z,a)dG_t(z,a). \tag{30}$$

Aggregation. Aggregating up, one can express output as a function of aggregate factors and aggregate TFP

$$Y_t = Z_t K_t^\alpha {L_t}^{1-\alpha}, \tag{31}$$

where aggregate TFP Z_t is an endogenous variable given by

$$Z_t = \left(\mathbb{E}_{\omega_t(\cdot)}\left[z \mid z > z_t^*\right]\right)^\alpha = \left(\frac{\int_{z_t^*}^\infty x\omega_t(x)\,dx}{1-\Omega_t(z_t^*)}\right)^\alpha. \tag{32}$$

This highlights that, in terms of output, the model is isomorphic to a standard representative-agent New Keynesian model with capital and TFP Z_t. TFP is endogenous and evolves over time and, as we will discuss below, depends on monetary policy.

Note that TFP Z_t serves as a measure of misallocation. The financial frictions faced by entrepreneurs imply that capital is not optimally allocated. The entrepreneur operating the most productive firm does not have enough net worth to operate the whole capital stock, hence less productive firms operate as well, which is suboptimal and which reduces TFPZ_t. Thus the more misallocated capital is, the lower is TFP Z_t.

Factor prices are

$$w_t = (1-\alpha) m_t Z_t K_t^{\alpha} {L_t}^{-\alpha}, \tag{33}$$

$$R_t = \alpha m_t Z_t K_t^{\alpha-1} {L_t}^{1-\alpha} \frac{z_t^*}{\mathbb{E}_{\omega_t(\cdot)}\left[z \mid z > z_t^*\right]}. \tag{34}$$

Finally, the law of motion of the aggregate net-worth of entrepreneurs is given by

$$\frac{\dot{A}_t}{A_t} = \frac{1}{q_t}\left[\gamma(1-\Omega_t(z_t^*))\left(\alpha m_t Z_t K_t^{\alpha-1} {L_t}^{1-\alpha} - R_t\right) + R_t - \delta q_t - q_t(1-\psi)\eta)\right]. \tag{35}$$

Appendix A.6 derives step by step these aggregate formulae.

2.8 Central Bank

The central bank controls nominal interest rates i_t on nominal bonds held by households. For the positive analysis in Section 3 we assume that the central bank sets the nominal rate according to a Taylor rule of the form form

$$di = -\upsilon\left(i_t - \left(\rho^h + \phi\left(\pi_t - \bar{\pi}\right) + \bar{\pi}\right)\right)dt, \tag{36}$$

where $\bar{\pi}$ is the inflation target, ϕ is the sensitivity to inflation deviations and υ is a parameter related to the persistence. For the normative analysis in Section 4 we assume that the central bank implements the Ramsey-optimal policy.

3 Monetary policy transmission

3.1 Numerical solution and calibration

Numerical algorithm. We solve the model numerically using the a new method, described in Appendix C.3. It combines a discretization of the model using an upwind finite-difference method similar to the one in Achdou et al. (2017) with a Newton algorithm that computes non-linear transitional dynamics. This can be easily implemented using Dynare's perfect foresight solver. Notice that the variables of the model include the distribution $\omega(z)$, which is an infinite-dimensional object. The finite-difference discretization turns this continuous variable into a finite dimensional vector.

It is important to highlight that our solution approach is different from the one in Winberry (2018) or Ahn et al. (2018). These papers analyze heterogeneous-agent models with aggregate shocks building on the seminal contribution by Reiter (2009). To this end, they linearize the model around the deterministic steady state. Winberry (2018) illustrates how this can be also implemented using Dynare and Ahn et al. (2018) extend the methodology to continuous-time problems. Here, instead, we compute the nonlinear transitional dynamics in the the sequence space, as Boppart et al. (2018) or Auclert et al. (2019). Boppart et al. (2018) show how the perfect-foresight transitional dynamics to a (small) MIT shock, such as the ones we compute here, coincide with the impulse responses obtained by a first-order perturbation approach in the model with aggregate uncertainty. We solve the model using a Newton solver. An important technical difference with Auclert et al. (2019) is that we do not guess a path of prices and iterate over time to find the path of all other variables, but we update all variables in a single step.

Calibration. Table 1 summarizes our calibration. The rate of time preference of the household ρ^h is 0.025, which targets an average real rate of return of 2.5 percent. The capital depreciation rate δ is set at 0.065, equal to the aggregate depreciation rate in NIPA. The fraction of assets of exiting entrepreneurs reinvested (ψ) is 0.1, so that the average size of entrants is 10 percent of that of incumbents, in line with US data (OECD, 2001). Entrepreneurs' exit rate (η) is 0.12 which, together with ψ, implies an average real return on equity of 11 percent, the return of the S&P500 from 2009 to 2019. The borrowing constraint parameter γ is 1.43, implying that entrepreneurs can borrow up to 43% of their net worth, which targets the level of aggregate US corporate debt as a percentage of net worth from 2009 to 2019. The capital share α is set at a standard value of 0.3. We assume log-utility in consumption ($\zeta = 1$) and the inverse Frisch elasticity ϑ is also set to 1, standard values in the literature. We normalize the constant multiplying the disutility of labor Υ such that aggregate labor supply in steady state is equal to one.

We assume adjustment costs are quadratic, i.e.,

$$\Xi\left(\iota_t\right) = \frac{\phi^k}{2}\left(\iota_t - \delta\right)^2. \tag{37}$$

Capital adjustment costs, ϕ^k, are set to 10, such that the peak response of investment to output after a monetary policy shock is around 2, in line with the VAR evidence of Christiano et al. (2016).

Regarding the New Keynesian block, the elasticity of substitution of retailer goods ϵ is set to 10, so that the steady state mark-up is $1/(\epsilon - 1) = 0.11$. The Rotemberg cost parameter θ is set to 100, so that the slope of the Phillips curve is $\epsilon/\theta = 0.1$ as in Kaplan et al. (2018)

The Taylor rule parameters take the following values: $\bar{\pi} = 0$, $\phi = 1.25$ and $\upsilon = 0.8$. These values are explained when dealing with the optimal policy in Section 4.2 below.

We assume that individual productivity z in logs follows an Ornstein-Uhlenbeck process[8]

$$d\log(z) = -\varsigma_z \log(z)dt + \sigma_z dW_t. \tag{39}$$

We calibrate the productivity process using the estimates from Gilchrist et al. (2014), who find a quarterly persistence of 0.8 and a volatility of 0.15 (0.3 annualized).

3.2 Misallocation and monetary policy

In this section we show the impulse response functions to a monetary policy shock and a demand shock when the central bank follows a Taylor rule. We compare our baseline heterogeneous-firm economy to that in the complete-market economy. The complete-market economy is the standard representative agent New Keynesian model

[8]By Ito's lemma, this implies that z in levels follows the diffusion process

$$dz = \mu(z)dt + \sigma(z)dW_t, \tag{38}$$

where $\mu(z) = z\left(-\varsigma_z \log z + \frac{\sigma^2}{2}\right)$ and $\sigma(z) = \sigma_z z$.

Table 1: **Calibration**

	Parameter	**Value**	**Source/target**
ρ^h	Household's discount factor	0.025	Av. 10Y bond return of 2.5% (FRED)
δ	Capital depreciation rate	0.065	Aggregate depreciation rate (NIPA)
ψ	Fraction firms' assets at entry	0.1	Av. size at entry 10% (OECD, 2001)
η	Firms' death rate	0.12	Av. real return on equity 11% (S&P500)
γ	Borrowing constraint parameter	1.43	Corporate debt to net worth of 43% (FRED)
α	Capital share in production function	0.3	Standard
ζ	Relative risk aversion parameter HH	1	Log utility in consumption
ϑ	Inverse Frisch Elasticity	1	Kaplan et al. (2018)
Υ	Constant in disutility of labor	0.71	Normalization $L = 1$
ϕ^k	Capital adjustment costs	10	VAR evidence
ϵ	Elasticity of substitution retail goods	10	Mark-up of 11%
θ	Price adjustment costs	100	Slope of PC of 0.1
$\bar{\pi}$	Inflation target	0	-
ϕ	Slope Taylor rule	1.25	-
υ	Persistence Taylor rule	0.8	-
	SS Aggregate Productivity	1	Normalization
ς_z	Mean reverting parameter	0.8	Persistence Gilchrist et al. (2014)
σ_z	Volatility of the shock	0.30	Volatility Gilchrist et al. (2014)

with capital. It can be seen as a special case of the baseline economy where the collateral constraint is set to infinite, so that the productivity/net-worth distribution becomes irrelevant and only the most productive firm operates. In this case, capital allocation is efficient (no misallocation) and TFP is exogenous. This contrasts with the baseline economy, in which the distribution of capital across firms matters due to financial frictions and determines the endogenous component of TFP. See Appendix A.9 for more details regarding the baseline versus complete-market model.

Impulse response to a monetary policy shock. We start our analysis of the model with the response to an expansionary monetary policy shock. The blue solid lines in Figure 1 show the response to a temporary 20 basis points reduction in the nominal rate (not shown). The shock produces a temporary fall in the nominal rate (not shown) which leads to a reduction in the real rate (panel d) and an increase in inflation and output (panels a and f) through the standard New Keynesian channels.

Note that, compared to the complete markets economy, the responses of inflation, price of input goods or wages (panels a-c) is very similar. Nonetheless, in the heterogeneous-firms economy these movements in prices allow high-productivity entrepreneurs to increase their investment relatively more, so that endogenous TFP *increases* (panel g). This is in line with the empirical evidence based on time series analysis, such as the one provided by Jordà et al. (2020), Moran and Queralto (2018), Meier et al. (2020) or Baqaee et al. (2021), among others. These papers explain the dynamics of TFP as the result of endogenous growth, hysteresis effects, or heterogeneity in markups. Here, instead, we focus on a complementary mechanism, namely capital misallocation. This increase in TFP, although moderate, allows the increase in output to be slightly larger in the baseline economy (panel f) than in the complete-markets case. The difference between both economies is larger in the medium term (2-4 years after impact), due to the persistence in TFP dynamics.

The misallocation channel of monetary policy. Why does TFP change in response to a monetary policy shock? As discussed above, aggregate TFP depends on the allocation of net-worth across entrepreneurs. In particular, by equation (32) which we reproduce here, TFP is given by the average productivity of constrained entrepreneurs, weighted by net worth shares:

$$Z_t = \left(\mathbb{E}_{\omega_t(\cdot)}\left[z \mid z > z_t^*\right]\right)^{\alpha}. \tag{40}$$

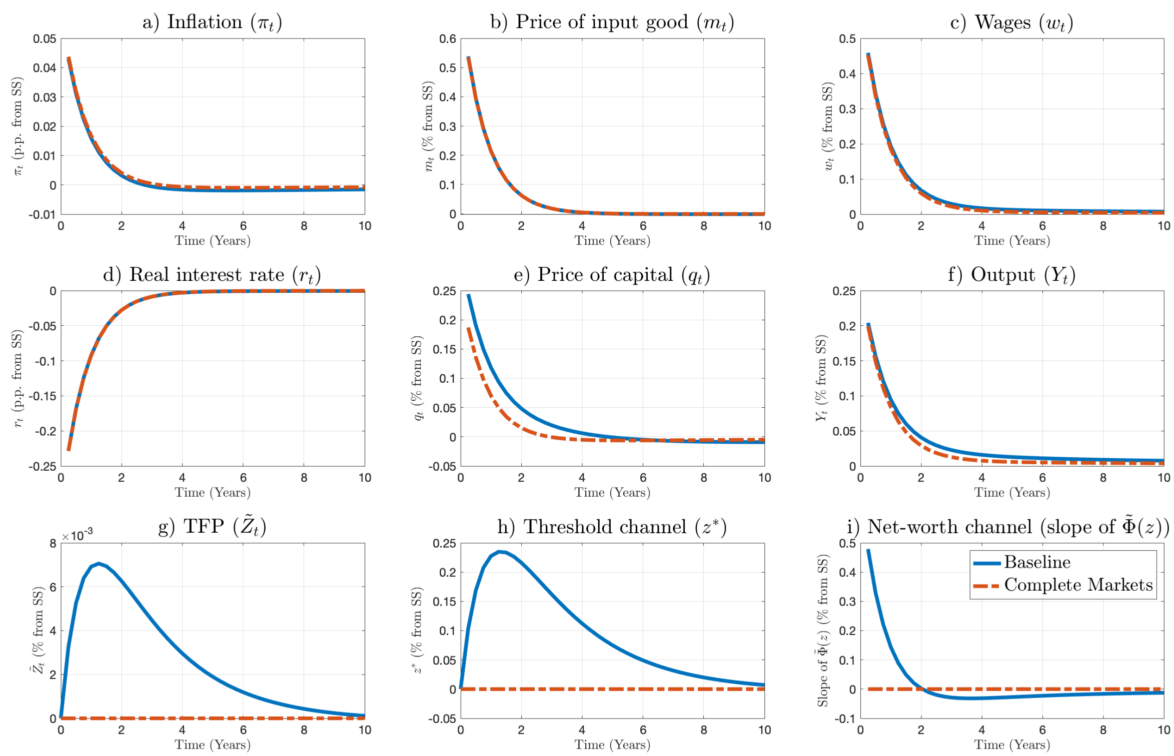

Figure 1: Impulse response to a monetary shock.

Notes: The figure shows the deviations from steady state of the economy. The solid blue line is the response of the baseline economy to a monetary policy shock of 25 basis points. The orange dotted line is the response in the complete markets economy to the same 25 basis point decrease in nominal rates. In both cases the central bank follows the Taylor rule.

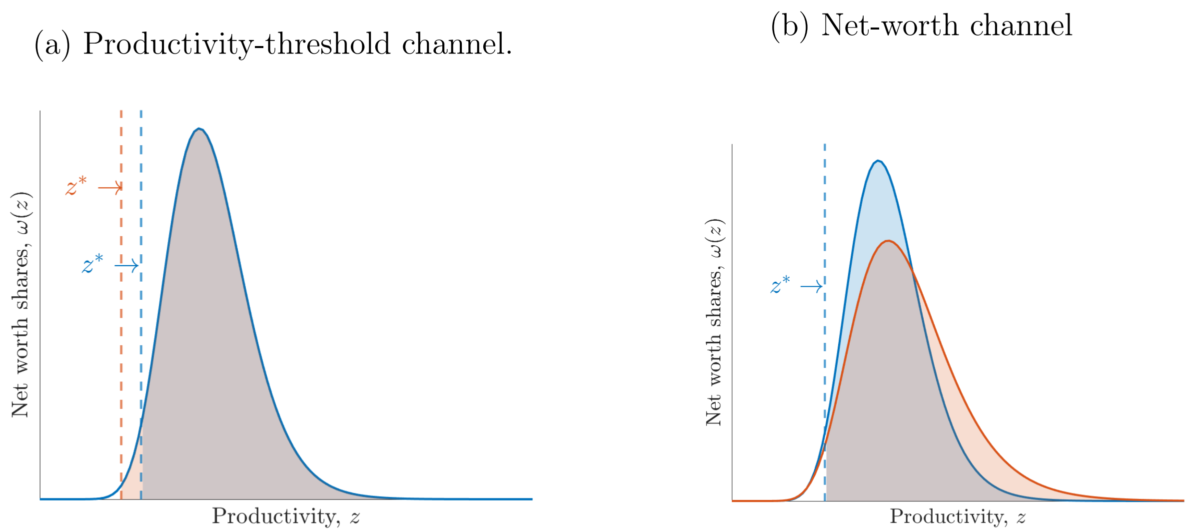

Figure 2: Net-worth distribution and productivity-threshold channels.
Notes: The figure shows the net-worth share distribution $\omega(z)$ and the productivity-threshold z^* (blue). The light blue area is the initial mass of constrained firms. Panel (a) shows the impact of a change in the threshold channel (orange dashed line), which shifts the threshold to the left. Panel (b) shows the impact of a change in the net-worth distribution, i.e. the net-worth channel. The new mass of constrained firms after the change is depicted by the shaded orange area in both panels.

TFP Z_t thus depends on the mass of the net-worth distribution $\omega_t(\cdot)$ above the productivity threshold z_t^* (the shaded area in Figure 2). Entrepreneurs to the left of z^* stay at their optimal size $k(z)^* = 0$, and rent out their net worth to constrained entrepreneurs to the right of the cut-off (those in the shaded area). Equation (40) allows us to identify the two sub-channels through which monetary policy affects aggregate TFP: (i) the *productivity-threshold channel*, related to changes in the threshold; and (ii) the *net-worth distribution channel*, related to changes in the net-worth distribution. We jointly refer to them as the "misallocation channel of monetary policy"

Productivity threshold channel. The productivity-threshold channel captures the fact that, by changing factor prices, monetary policy affects the productivity threshold above which entrepreneurs become constrained. Panel (a) in Figure 2 illustrates how a reduction in the threshold increases the share of constrained firms by crowding in low-productivity entrepreneurs. Analytically, take the partial derivative of TFP (equation 32) with respect to z_t^*, we obtain

$$\frac{\partial Z_t}{\partial z_t^*} = \frac{\alpha\omega\left(z_t^*\right)}{Z_t^{\frac{1-\alpha}{\alpha}}\left(1-\Omega_t\left(z_t^*\right)\right)}\left(\mathbb{E}_{\omega_t(\cdot)}\left[z \mid z > z_t^*\right] - z_t^*\right) > 0. \quad (41)$$

The derivative of TFP with respect to the threshold is always non-negative, and it is strictly positive as long as the distribution $\omega(z)$ is non-zero for $z > z_t^*$. This means that, *ceteris paribus*, if the threshold decreases so does TFP. This is the result of a larger share of capital in the hands on low-productivity entrepreneurs. Note that the term $\left(\mathbb{E}_{\omega_t(\cdot)}\left[z \mid z > z_t^*\right] - z_t^*\right)$ is a measure of the dispersion of productivity of constrained firms: the larger the difference between the average productivity of constrained firms and the productivity threshold, the larger the impact of a change in the threshold is.

Combining equation (10) with the definitions (9) and (17), we can express the productivity threshold as

$$z_t^* = \frac{(q_t r_t + \delta q_t - \dot{q}_t)}{\alpha \left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}}.$$

This equation reflects how the threshold is affected by monetary policy through changes in factor prices. Panel (h) in Figure 1 shows how, in response to a monetary policy shock, the threshold increases. This contributes positively to the the increase in TFP.

The interpretation of the productivity threshold channel is the following. As discussed by Moll (2014), the assumption of constant return to scale in firms' production function (1) can be seen as the limiting case of decreasing returns to scale (DRS), $y_t = \left[(z_t k_t)^\alpha (l_t)^{1-\alpha}\right]^\nu$, $\nu < 1$, when $\nu \to 1$. In the case with DRS, there is a threshold $z^*(a)$ which depends on net-worth, such that for those firms with $z \leq z^*(a)$, the firm is unconstrained and it produces at its optimal level $(k^*(z))$, whereas for those with $z > z^*(a)$, the firm is constrained. When $\nu \to 1$, the optimal size of the firm, and hence its production, are very small, $k^*(z), y^*(z) \to 0$. Hence, both in the DRS and in the limiting case of CRS, the productivity threshold channel modifies the share of constrained versus unconstrained firms in the economy. This mechanism is different from the extensive margin mechanism. Hence, it is not meant to capture firm entry and exit, which in our model is exogenously given by the retiring probability η.

Net-worth distribution channel. So far, we have implicitly kept the net worth distribution constant. But by changing firms' profits and investment, monetary policy also affects the dynamics of the net-worth distribution, and hence of aggregate TFP through equation (40). Changes in this distribution, such as shifts or changes in skewness or kurtosis, can change this conditional mean. Panel (b) of Figure 2 illustrates the effect of a rightward shift and tilt in the distribution.

On impact, the only operating channel is the productivity-threshold one, as the net-worth distribution is predetermined. The net-worth distribution channel thus only affects TFP as time goes by.

Appendix A.8 derives two results concerning this channel. First, we show how, conditional on a constant cut-off z^*, only changes in the wage w_t, price of capital q_t, and input-good price m_t affect TFP dynamics. The intuition for this result is that all constrained firms benefit the same from lower capital costs. Second, we prove how the sign of the impact on TFP growth of a change in each of these prices depends exclusively on the effect of the price on the firm's *excess investment rate*, defined as

$$\tilde{\Phi}_t(z) \equiv \frac{\gamma \Phi_t}{q_t k_t} = \max \left\{ \frac{\gamma \alpha}{q_t} \left(\frac{(1-\alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}} \left(z - z_t^* \right), 0 \right\}, \tag{42}$$

where we have employed the definition of profits Φ_t (equation 9), as well the definitions of the rental rate R_t and the threshold z_t^* (equations 17 and 10, respectively). Notice that Φ_t/k_t is the return that a firm makes over the cost of capital R_t. Since entrepreneurs do not distribute dividends until they retire, these returns are reinvested in firms' capital. Hence we can understand $\tilde{\Phi}_t(z)$ as the investment rate of a firm with productivity z in excess of the investment rate of the marginal firm with productivity z^*.[9] The excess investment rate captures the heterogeneity in investment across productivity levels. Its shape informs us about how the net worth distribution evolves over time and hence, *ceteris paribus*, how TFP does. The steeper it is, the more do high-productivity firms outgrow low-productivity ones, and the faster increases TFP.

Panel (i) in Figure 1 shows that the slope of the excess investment rate increases after a monetary expansion. High-productivity firms thus invest relatively more than low-productivity ones.

Impulse response to a demand shock. Can we then conclude that a reduction in real rates reduces misallocation? Not necessarily. This can be seen in Figure 3, where we display the impulse responses to a demand shock that temporarily reduces real rates. In particular, we consider a temporary decrease in households' discount factor ρ^h calibrated to match the decline in real rates on impact of the expansionary monetary policy shock. In contrast to the case of a monetary policy shock, TFP now decreases (panel g), reflecting an increase in misallocation. Since a decrease in ρ_t^h implies that the household becomes more patient, consumption decreases and savings (in capital)

[9]Note that the investment rate of the marginal firm with productivity z^* is equal to $\frac{Ra}{k} = R/\gamma$.

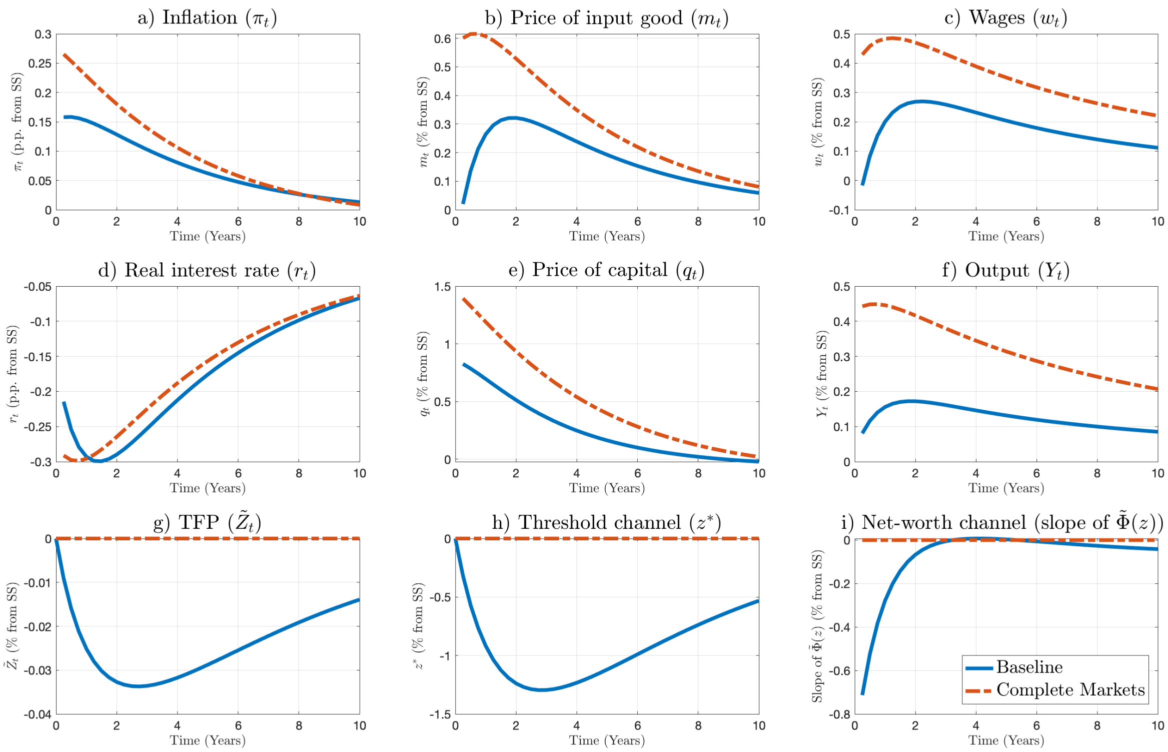

Figure 3: Impulse response to a a demand shock.

Notes: The figure shows the deviations from steady state of the economy. The solid blue line is the response of the baseline economy to a 20% decrease of the discount factor of the household, ρ_t^{hh}, that reverts to its steady state value following an autoregressive process with yearly persistence of 0.8. The dotted orange line is the response of the complete market economy to the same shock. In both cases the central bank follows the Taylor rule.

increase. This triggers price adjustments of the same sign as the monetary policy shock (panels b-e). However, the difference in magnitudes of these price movements implies that TFP now falls as misallocation increases through both the productivity-threshold and the net-worth distribution channels (panels h-i). This result is in line with the findings of Gopinath et al. (2017) or Asriyan et al. (2021), who show how a decline in the real rate may produce an increase in capital misallocation.

3.3 Empirical evidence: the effect of monetary policy shocks at firm level

As discussed above, the positive impact of monetary policy shocks on TFP can be rationalized by different explanations, such as endogenous growth or markup heterogeneity. In our case, it is due to the fact that a larger share of capital is in the hands of high-productivity firms. This is both the result of a higher productivity threshold and more investment by high-productivity firms. While the first channel has a less obvious mapping to the data, the second one can be empirically tested.

To this end, we combine Spanish firm-level panel data with a time series measure of exogenous monetary policy shocks. We use yearly balance-sheet and cash-flow data from the quasi-universe of Spanish firms from 2000 to 2016 from the *Central de Balances Integrada* (see Appendix B.1 for further details on the data). The main advantage of this dataset is that it covers the quasi-universe of Spanish firms, including not only large firms with access to stock and bond markets, but also medium and small firms more reliant on bank credit and internal financing. This contrasts with most papers in this literature, which use data from publicly traded firms (e.g. Compustat). These are generally large firms with access to the equity market, which can potentially behave very differently from the rest of firms in the economy. [10]

Our key variable of interest is firm level productivity, which we proxy by the marginal revenue product of capital ($MRPK_{j,t-1}$). In our context, this measure has two advantages compared to other empirical productivity measures. First, it is a measure directly linked to capital productivity, and hence to investment in capital. Furthermore, in the model the MRPK is proportional to firm productivity z.

$$MRPK_t = \frac{\partial m_t f_t(z,k,l^*)}{\partial k} = \left[\left(\frac{1-\alpha}{w_t} \right)^{\frac{1-\alpha}{\alpha}} m_t^{\frac{1}{\alpha}} \right] z \propto z.$$

Second, its computation from the data is straightforward and it does not rely on estimation. The monetary policy shock ε_t^{MP} is taken from Jarociński and Karadi (2020). They use high-frequency data and sign restrictions in a SVAR to identify monetary policy shocks in the Euro area at the monthly frequency. The key idea behind their identification strategy is that movements of interest rates and stock markets within a narrow window around monetary policy announcements can help disentangle monetary policy shocks from information surprises. While an unexpected policy tightening raises interest rates and reduces stock prices, a positive central bank information shock (i.e. unexpected positive assessment of the economic outlook) raises both. We need to aggregate their shocks to yearly frequency, as in our data. To do so, we follow a

[10] Caglio et al. (2021), for instance, show how monetary policy transmission and risk taking differ across SMEs and large listed firms.

methodology that resembles the one employed by Ottonello and Winberry (2020) to aggregate to quarterly frequency. Appendix B.2 provides more details on the construction of the monetary policy shock.

In order to test whether high-productivity firms' investment is more responsive to monetary policy shocks, we estimate the following equation:

$$\Delta log\ k_{j,t} = \alpha_j + \alpha_{s,t} + \beta \left(MRPK_{j,t-1} - \mathbb{E}_j \left[MRPK_j \right] \right) \varepsilon_t^{MP} + \Lambda' Z_{j,t-1} + u_{j,t}. \tag{43}$$

The dependent variable $\Delta log\ k_{j,t}$ is the log increase in the capital stock of firm j from $t-1$ to t. The key parameter of interest in equation (43) is the coefficient β multiplying the interaction term between productivity and the monetary policy shock . We demean $MRPK_{j,t-1}$ by the firm average across time $\mathbb{E}_j \left[MRPK_j \right]$ to ensure that the results are not driven by permanent heterogeneity in responsiveness across firms. We lag $MRPK_{jt-1}$ to address reverse causality concerns. A positive value of β indicates that high-productivity firms' investment responds more to a monetary expansion. We also include firm fixed effects (α_j) to capture permanent differences in investment patterns, sector-year fixed effects ($\alpha_{s,t}$) to control for aggregate shocks at the sector level, and a vector of lagged controls $Z_{j,t-1}$ that includes the demeaned MRPK measure, total assets, sales growth, leverage, net financial assets as a share of total assets, and the interaction of demeaned MRPK with GDP growth.

We follow Ottonello and Winberry (2020) in both preparing and cleaning the data set (see Appendix B.1 for details) and in designing the estimation equation (43), where we just switch the variable of interest. In doing so we aim to maximize transparency and comparability with previous studies.

Table 2 shows the main results of the estimation. We perform the same normalization as in Ottonello and Winberry (2020), so that the coefficient of interest, β, is easily interpretable. First, we standardize $\left(MRPK_{jt-1} - \mathbb{E}_j \left[MRPK \right] \right)$ over the entire sample, which implies that the units are standard deviations in our sample. Second, we normalize the shock, so that the interpretation of β can be read as the response to an expansionary monetary policy shock of 100bps (or in other words, a decrease of 1pp in the EONIA rate). Results show that firms with high productivity, proxied by high MRPK, respond more to expansionary monetary policy shocks. Our baseline specification, column (2), shows that a surprise reduction of 1pp in real interest rates (expansionary monetary policy shock) implies a further 29pp increase in the investment

Table 2: **Heterogeneous responses of investment to monetary policy in MRPK**

	(1)	(2)
ϵ_t^{MP1} x $MRPK_{t-1}$	0.141**	0.293***
	(0.06)	(0.07)
Observations	5,567,706	4,169,950
R^2	0.267	0.285
MRPK control	YES	YES
Controls	NO	YES
Time-sector FE	YES	YES
Time-sector clustering	YES	YES

Notes: The table shows the coefficient β that results of estimating equation (43). Column (1) only includes the standardized demeaned MRPK as control, while column (2) introduces the all the controls $Z_{j,t-1}$ (standardized demeaned MRPK, total assets, leverage, sales growth, and net financial assets as a share of total assets; and the interaction of demeaned MRPK with GDP growth). Standard errors are clustered at the sector-year level. We have normalized the sign of the monetary shock ε_t^{MP} so that a positive shock corresponds to a decrease in interest rates. We have standardize $(MRPK_{jt-1} - \mathbb{E}_j [MRPK])$ over the entire sample.

rate of a firm that is one standard deviation more productive than the average in our sample (in terms of MRPK). When we do not include firm controls (column 1), this effect is still positive and significant, although of lower magnitude. Appendix B.3 shows that the result is robust to several alternative specifications. It is worth noticing that this heterogeneous response is not driven by changes in the composition of firms in the data, since keeping a balanced sample of firms we find even larger results (see Appendix B.3).

Summing up, the empirical evidence supports the model prediction that the impact of monetary policy on investment is increasing in the productivity of the firm. Albrizio et al. (2021) show that, after an expansionary monetary policy shock, aggregate measures of misallocation decrease. This provides further evidence pointing at a decrease in misallocation as the net effect of the different general equilibrium forces after an expansionary monetary policy shock. This result is key to understand how firm heterogeneity shapes optimal monetary policy in the next section.

4 Optimal monetary policy

4.1 Central bank objective and numerical approach

Ramsey problem. Having analyzed the interactions between monetary policy, firm heterogeneity and financial frictions, we turn next to the main focus of this paper, namely how these interactions affect optimal monetary policy. We assume that the central bank sets its policy instrument – the nominal interest rate i_t – such as to maximize household utility under full commitment. That is, the central bank solves the following Ramsey problem:

$$\max_{\{\omega_t(z),s_t(z),w_t,r_t,q_t,\varphi_t,K_t,A_t,L_t,C_t,D_t,Z_t,\Omega_t,z_t^*,\iota_t,\pi_t,m_t,\tilde{m}_t,i_t,Y_t,T_t\}_{t\geq 0}} \mathbb{E}_0 \int_0^\infty e^{-\rho^h t} u(C_t, L_t) dt \quad (44)$$

subject to the all the equilibrium conditions derived above and listed in Appendix A.7 and the initial conditions $\{\omega_0(z), K_0, Z_0\}$. The equilibrium conditions include, among others, the law of motion of the net-worth distribution $\omega_t(z)$ (equation 26), as the central bank internalizes the impact of her decisions on it. We stress the fact that the central bank's only instrument is the nominal interest rate. For simplicity, we calibrate the labor tax/subsidy τ such that it undoes the New Keynesian mark-up distortion in the steady state.

Numerical approach. Notice that $\omega_t(z)$ and $s_t(z)$ not only depend on time, but also on individual productivity. This poses a challenge when solving optimal monetary policy, as we need to compute the first order conditions (FOCs) with respect to infinite-dimensional objects. There are a number of proposals in the literature to deal with this problem. Bhandari et al. (2021) make the continuous cross-sectional distribution finite-dimensional by assuming that there are N agents instead of a continuum. They then derive standard FOCs for the planner. In order to cope with the large dimensionality of their problem, they employ a perturbation technique. Le Grand et al. (2020) employ the finite-memory algorithm proposed by Ragot (2019). It requires changing the original problem such that, after K periods, the state of each agent is reset. This way the cross-sectional distribution becomes finite-dimensional. Nuño and Thomas (2016) deal with the full infinite-dimensional problem in continuous time. This implies that the continuous Kolmogorov forward (KF) and the Hamilton-Jacobi-Bellman (HJB) equations form part of the constraints faced by the central bank. They derive the planner's FOCs using calculus of variations, thus expanding the original problem to also include the Lagrange multipliers. They then solve the problem using the upwind finite-difference method of Achdou et al. (2017).

Here we proposed a new algorithm, detailed in in Appendix D. We first discretize the central bank's objective and constraints (the private equilibrium conditions) using finite differences, then find the planner's FOCs *by symbolic differentiation* and finally solve them non-linearly in the sequence space. The first step (discretization using finite differences) was already described in Section 3.1 and Appendix C. The idea is to transform the original continuous-time, continuous-state planner's problem into a discrete-time, discrete-space one. This problem is high-dimensional, as infinite-dimensional objects become large vectors. The second and third steps can conveniently be implemented using several available software packages. In our case, we employ Dynare.[11] This algorithm can be employed to compute optimal policies in a large class of heterogeneous agent models and stands out for being extremely easy to implement.

4.2 Optimal Ramsey policy

Steady state. We first consider the steady state of the Ramsey problem. It is well known that the standard New Keynesian economy with an efficient steady state features zero inflation in steady state under the optimal policy. Due to capital misallocation, our baseline economy does not feature steady state efficiency. Yet, inflation is zero in the steady state of the Ramsey problem. This result mirrors a similar result from the textbook New Keynesian model with a distorted steady state (Woodford, 2003; Gali, 2008). Though the long-run Phillips curve allows monetary policy to affect misallocation in the long run through positive trend inflation, the benefits of this policy are compensated for by the cost of the anticipation of this policy.

Time-0 optimal policy. We turn next to the deterministic dynamics under the Ramsey optimal plan. We solve for the Ramsey plan when the initial state of the economy coincides with the steady state under the optimal policy, i.e., that with zero inflation. As explained in the previous section, the Ramsey planner faces no precommitments, which is equivalent to assume that the Lagrange multipliers associated to the forward-looking variables are initially zero. This is commonly referred to as the "time-0 optimal policy" (Woodford, 2003).

[11] Dynare includes the command `ramsey_policy` that automatically performs these steps.

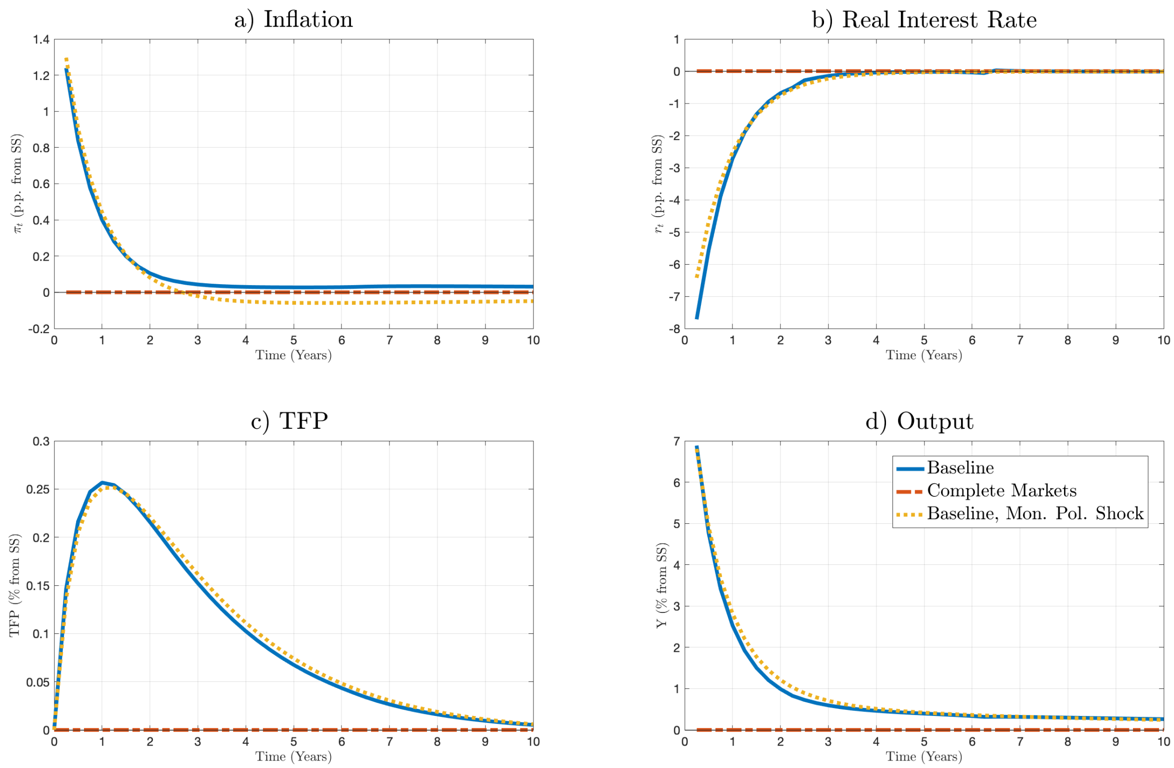

Figure 4: Time 0 optimal monetary policy.

Notes: The figure shows the deviations from steady state of the economy when the planner is allowed to re-optimize with no pre-commitments in response to no shock. The baseline economy is the solid blue line, and the complete markets economy (CM) the dashed orange line. The dotted yellow line is the response of the baseline model in general equilibrium to a monetary policy shock of 700 basis points, where the central bank follows the Taylor rule $di = -\upsilon \left(i_t - \left(\rho_t^h + \phi\left(\pi_t - \bar{\pi}\right) + \bar{\pi}\right)\right) dt$, with $\upsilon = 0.8$, and $\phi = 1.25$.

The Ramsey plan in the model with complete markets is time-consistent. Hence, inflation and the rest of variables remain constant at their steady state values. This is displayed by the dashed red lines in Figure 4. Market incompleteness, however, introduces a new motive for time inconsistency, urging the central bank to temporally deviate from the zero-inflation policy. The solid blue lines in Figure 4 show how the central bank engineers a surprise monetary expansion, by reducing real rates (panel b). The dotted yellow line displays a monetary policy rule with the same calibration as in Section 3.2. The calibration of the policy rule was chosen to replicate the dynamics of the optimal monetary policy. The resulting dynamics are almost identical to those caused by an expansionary monetary policy shock, which were described in detail in Section 3.2: the change in factor prices increases TFP (panel c). The central bank is thus willing to tolerate a temporary increase in inflation to achieve a persistent rise in TFP, brought about by a more efficient allocation of capital.

The desire of the central bank to redistribute resources towards entrepreneurs in order to promote firm growth is reminiscent of the case with optimal fiscal policy analyzed by Itskhoki and Moll (2019). They find that optimal fiscal policies in economies starting at below steady-state net-worth levels initially redistributes from households towards entrepreneurs in order to speed up net worth accumulation. In our case, and given the lack of fiscal instruments, it is the central bank who engineers this redistribution through an expansion in aggregate demand.

4.3 Timeless optimal policy response

Next, we analyze the optimal policy response when an unexpected shock hits the economy that was previously in its zero-inflation steady state. In this case, we adopt a "timeless perspective" (Woodford, 2003, Gali, 2008). The optimal timeless Ramsey policy implies that the central bank sticks to its pre-commitments, implementing the policy that it would have chosen to implement if it had been optimizing from a time period far in the past. The Lagrange multipliers associated to forward-looking equations in this case are initially set to their steady state values.[12] This is a concept that

[12]Dynare allows to compute optimal policies from a timeless perspective. First, the `ramsey_model` command computes the FOCs for the Ramsey problem by symbolic differentiation. Second, the `steady` command computes the steady state of the Ramsey problem. Finally, the `perfect_foresight_solver` command uses the Newton method to solve simultaneously all the non-linear equations for every period, using sparse matrices.

only makes sense in the presence of aggregate risk. As discussed in Section 3.1, building on the argument by Boppart et al. (2018) one can reinterpret the timeless response to MIT shocks as a first order approximation to the response in a model with aggregate uncertainty under the ex-ante optimal time-invariant state-contingent policy rule.

Demand shock. We analyze the optimal response to a demand shock, modeled as a temporary fall in the household's discount factor of 20%. Figure 5 shows that the optimal response in the baseline economy (blue solid line) mimics that under complete markets (orange dashed line).

The response both with complete and incomplete markets is characterized by what Gali (2008) described as "divine coincidence": the optimal response by the central bank is to stabilize inflation at its steady state value of zero (panel a), which also keeps the output gap at its optimal steady state value of zero (panel d).[13] This, however, requires that the central bank lowers the real and nominal rates more aggressively in the baseline with incomplete markets (panel b). The reason is that the original demand shock leads to a negative 'supply shock' through its impact on aggregate TFP (panel c), which depresses output and natural rates relative to the complete markets case. As the central bank has to replicate the path of natural rates (not shown), it is forced to reduce real rates more persistently than with complete markets.

In contrast to these results, if the central bank follows a suboptimal Taylor rule, as in Figure (3) above, the central bank reduces rates by less, and hence inflation increases in comparison to the optimal policy.

Zero lower bound. The fact that the central bank responds more persistently to the demand shock under the optimal policy has important implications when the zero lower bound constrains its room for maneuver. Figure 6 displays the optimal response from a timeless perspective to a large negative demand shock that forces nominal rates to hit the zero lower bound (ZLB). The optimal response under complete markets was originally analyzed by Eggertsson et al. (2003). It prescribes a "low for longer" strategy, namely that nominal rates (dashed orange line in panel a) should remain at the ZLB for a longer period that would be prescribed in the absence of the ZLB (dotted yellow line). This delay in the lift-off date of nominal rates is of one quarter, approximately, and nominal rates abandon the ZLB in the three quarters after the shock arrival.

This contrasts with the case with the baseline economy with incomplete markets.

[13] The output gaps is defined as the difference between observed output and the counterfactual output in an economy without nominal rigidities.

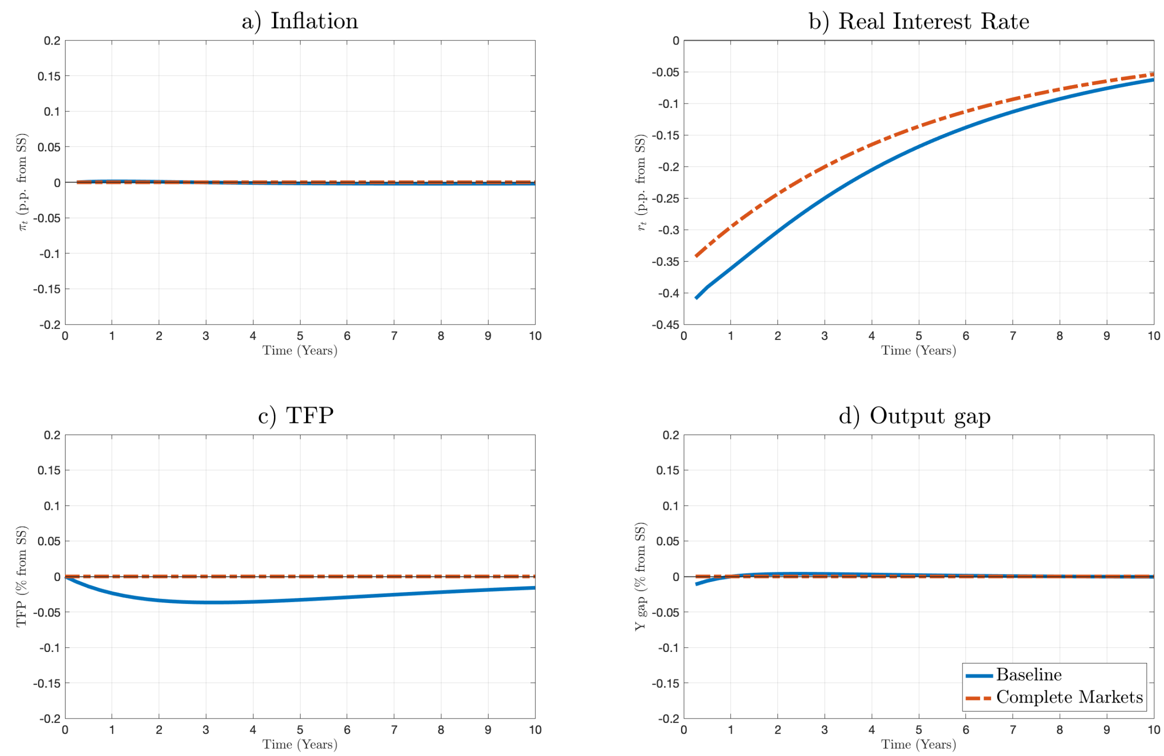

Figure 5: Optimal monetary policy response to a demand shock.

Notes: The figure shows the optimal response from a timeless perspective (in deviations from steady state) to a 20% decrease in the rate of time preference of the household ρ^h that is mean reverting with a yearly persistence of 0.8. The baseline economy is the solid blue line, and the complete markets economy (CM) the dashed orange line. The dotted yellow line is the response of the baseline model when the path of inflation is that of the complete markets economy.

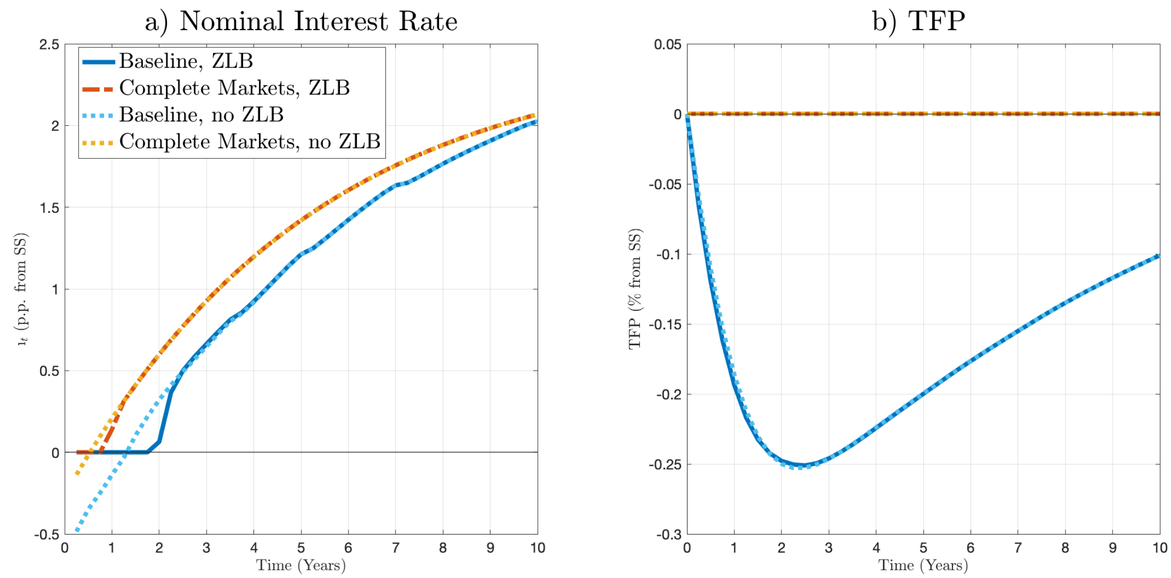

Figure 6: Optimal monetary policy response to a demand shock with the zero lower bound.

Notes: The figure shows the optimal response from a timeless perspective (in deviations from steady state) to a 4pp decrease in the rate of time preference of the household ρ^h that is mean reverting with a yearly persistence of 0.8. The baseline economy with the zero lower bound is the solid blue line, and the complete markets economy with he zero lower bound is the dashed orange line. The dotted light blue line is the optimal response in the baseline economy without zero lower bound, and the yellow dotted line is the optimal response in the complete markets economy without the zero lower bound..

This case is also characterized by a low for longer optimal policy (solid blue line in panel a), but the lift-off date is now delayed by more than twice later than with complete markets, happening at the end of the second year after the shock. We call this a "low for *even* longer" policy. There are two reasons for this delay. First, as discussed above, natural rates fall more persistently in the case with incomplete markets, and so do nominal rates under the optimal policy without the ZLB (dotted light blue line). This explains a delay of a year and a quarter. Second, the delay in the lift-off with respect to the counterfactual no-ZLB case is also amplified in this case, in order to compensate for the larger decline in real rates that could not be attained due to the presence of the ZLB.

5 Conclusions

This paper analyzes monetary policy in a model with heterogeneous firms, financial frictions, and nominal rigidities. The model features a link between monetary policy and capital misallocation. Monetary policy affects aggregate misallocation by changing (i) the productivity threshold above which firms are credit-constrained, and (ii) the net-worth distribution of firms. We find that expansionary monetary policy reduces capital misallocation by allowing high-productivity firms to increase investment and grow faster. Using granular information about Spanish firms, we provide empirical evidence that this mechanism is indeed present in the data: high-productivity firms are more responsive to monetary policy shocks.

We analyze optimal monetary policy for a benevolent central bank. We show how a central bank without pre-commitments engineers an unexpected monetary expansion to increase TFP in the medium run. We also illustrate how, when faced with a temporary demand shock, price stability is the optimal policy, just as under complete markets. If the ZLB constraints the path of nominal rates, then the optimal policy is a "low for even longer", delaying the lift-off in nominal rates much longer than under complete markets.

The paper also makes a methodological contribution. It introduces a new algorithm to compute optimal policies in heterogeneous-agent models. The algorithm leverages the numerical advantages of continuous time and will allow researchers to solve optimal policy in heterogeneous-agent models with or without aggregate shocks in an efficient and simple way using Dynare.

Finally, the model presented in this paper abstracts from several relevant mechanisms driving firm dynamics, such as endogenous default, size-varying capital constraints, or decreasing returns to scale, among many others. This helps us to provide a clear understanding of the different forces linking monetary policy with capital misallocation, as well as highlighting the similarities and differences with the standard New Keynesian model. A natural extension would be to add more of these features to study their impact on the optimal conduct of monetary policy.

References

Acharya, V. V., T. Eisert, C. Eufinger, and C. Hirsch (2019). Whatever it takes: The real effects of unconventional monetary policy. *The Review of Financial Studies 32*(9), 3366–3411. 1

Acharya, V. V., S. Lenzu, and O. Wang (2021). Zombie lending and policy traps. *Available at SSRN 3936064*. 1

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017). Income and wealth distribution in macroeconomics: A continuous-time approach. Technical report, National Bureau of Economic Research. 1, 3.1, 4.1, C.1, 16, D.1

Adam, K. and R. M. Billi (2006). Optimal monetary policy under commitment with a zero bound on nominal interest rates. *Journal of Money, credit and Banking*, 1877–1905. 1

Adam, K. and H. Weber (2019). Optimal trend inflation. *American Economic Review 109*(2), 702–37. 3

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018). When inequality matters for macro and macro matters for inequality. *NBER macroeconomics annual 32*(1), 1–75. 3.1, D.1

Albrizio, S., B. González, and D. Khametshin (2021). A tale of two margins: Monetary policy and misallocation. Technical report. 3.3

Almunia, M., D. Lopez Rodriguez, and E. Moral-Benito (2018). Evaluating the macro-representativeness of a firm-level database: an application for the spanish economy. *Banco de España Ocassional Paper* (1802). B.1

Andrés, J., O. Arce, and P. Burriel (2021). Market polarization and the phillips curve. *Documentos de trabajo/Banco de España, 2106*. 3

Armenter, R. and V. Hnatkovska (2017). Taxes and capital structure: Understanding firms' savings. *Journal of Monetary Economics 87*, 13–33. B.1

Asriyan, V., L. Laeven, A. Martin, A. Van der Ghote, and V. Vanasco (2021). Falling interest rates and credit misallocation: Lessons from general equilibrium. Technical report. 1, 3.2

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2019). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. Technical report, National Bureau of Economic Research. 3.1, C, D.1

Auclert, A., M. Rognlie, and L. Straub (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. Technical report, National Bureau of Economic Research. 14, 20

Baqaee, D., E. Farhi, and K. Sangani (2021). The supply-side effects of monetary policy. Technical report, National Bureau of Economic Research. 1, 3, 3.2

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021). Inequality, business cycles, and monetary-fiscal policy. *Econometrica, forthcoming*. 1, 4.1

Bigio, S. and Y. Sannikov (2021). A model of credit, money, interest, and prices. Technical report, National Bureau of Economic Research. 1

Bilbiie, F. O., I. Fujiwara, and F. Ghironi (2014). Optimal monetary policy with endogenous entry and product variety. *Journal of Monetary Economics 64*, 1–20. 3

Bilbiie, F. O. and X. Ragot (2020). Optimal monetary policy and liquidity with heterogeneous households. *Review of Economic Dynamics*. 1

Boppart, T., P. Krusell, and K. Mitman (2018). Exploiting mit shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control 89*, 68–92. 3.1, 4.3, D.1

Buera, F. J. and J. P. Nicolini (2020). Liquidity traps and monetary policy: Managing a credit crunch. *American Economic Journal: Macroeconomics 12*(3), 110–38. 4

Caglio, C. R., R. M. Darst, and S. Kalemli-Özcan (2021). Risk-taking and monetary policy transmission: Evidence from loans to smes and large firms. Technical report, National Bureau of Economic Research. 2, 10

Christiano, L. J., M. S. Eichenbaum, and M. Trabandt (2016). Unemployment and business cycles. *Econometrica 84*(4), 1523–1569. 3.1

David, J. M. and D. Zeke (2021). Risk-taking, capital allocation and optimal monetary policy. Technical report. 3

Dávila, E. and A. Schaab (2022). Optimal monetary policy with heterogeneous agents: A timeless ramsey approach. *Working Paper*. 1

Eberly, J., S. Rebelo, and N. Vincent (2012). What explains the lagged-investment effect? *Journal of Monetary Economics 59*(4), 370–380. B.3

Eggertsson, G. B. et al. (2003). Zero bound on interest rates and optimal monetary policy. *Brookings papers on economic activity 2003*(1), 139–233. 4.3

Eggertsson, G. B. and M. Woodford (2004). Policy options in a liquidity trap. *American Economic Review 94*(2), 76–79. 1

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425. B.1

Gali, J. (2008). Monetary policy, inflation, and the business cycle: An introduction to the new keynesian framework. *Princeton University Press*. 1, 4.2, 4.3

Gertler, M. and P. Karadi (2011). A model of unconventional monetary policy. *Journal of monetary Economics 58*(1), 17–34. 2.1

Gilchrist, S., J. W. Sim, and E. Zakrajšek (2014). Uncertainty, financial frictions, and investment dynamics. Technical report, National Bureau of Economic Research. 3.1, 3.1

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital allocation and productivity in south europe. *The Quarterly Journal of Economics 132*(4), 1915–1967. 1, 3.2

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics 124*(4), 1403–1448. 1

Itskhoki, O. and B. Moll (2019). Optimal development policies with financial frictions. *Econometrica 87*(1), 139–173. 1, 4.2

Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises: the role of information shocks. *American Economic Journal: Macroeconomics 12*(2), 1–43. 1, 3.3, B.2

Jeenas, P. (2020). Firm balance sheet liquidity, monetary policy shocks, and investment dynamics. Technical report, Universidad Pompeu Fabra. 1, 2.4

Jordà, Ò., S. R. Singh, and A. M. Taylor (2020). The long-run effects of monetary policy. Technical report, National Bureau of Economic Research. 1, 3.2

Juillard, M., D. Laxton, P. McAdam, and H. Pioro (1998). An algorithm competition: First-order iterations versus newton-based techniques. *Journal of Economic Dynamics and Control 22*(8-9), 1291–1318. C, D.1

Kaplan, G., B. Moll, and G. L. Violante (2018). Monetary policy according to hank. *American Economic Review 108*(3), 697–743. 3.1, 3.1, A.2

Koby, Y. and C. Wolf (2020). Aggregation in heterogeneous-firm models: Theory and measurement. Technical report. 1

Le Grand, F., A. Martin-Baillon, and X. Ragot (2020). What is monetary policy about? optimal monetary policy with heterogeneous agents. Technical report, Paris School of Economics. 1, 4.1

Meier, M., T. Reinelt, et al. (2020). Monetary policy, markup dispersion, and aggregate tfp. Technical report, University of Bonn and University of Mannheim, Germany. 1, 3, 3.2

Midrigan, V. and D. Y. Xu (2014, February). Finance and misallocation: Evidence from plant-level data. *American Economic Review 104*(2), 422–58. 1

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review 104*(10), 3186–3221. 1, 4, 2, 5, 3.2

Moran, P. and A. Queralto (2018). Innovation, productivity, and monetary policy. *Journal of Monetary Economics 93*, 24–41. 1, 3.2

Nakov, A. et al. (2008). Optimal and simple monetary policy rules with zero floor on the nominal interest rate. *International Journal of Central Banking 4*(2), 73–127. 1

Nakov, A. and H. Webber (2021). Micro heterogeneity, misallocation and monetary policy. Technical report. 3

Nuño, G. and C. Thomas (2016). Optimal monetary policy with heterogeneous agents. *Documentos de trabajo/Banco de España, 1624*. 1, 4.1, D.1, D.2, D.3, 8

OECD (2001). Productivity and firm dynamics: evidence from microdata. Technical Report 69(1), OECD. 3.1, 3.1

Ottonello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica 88*(6), 2473–2502. 1, 2.4, 3.3, 3.3, 3.3, B.1, B.2, B.3

Ragot, X. (2019). Managing inequality over the business cycles: Optimal policies with heterogeneous agents and aggregate shocks. 2019 Meeting Papers 1090, Society for Economic Dynamics. 4.1

Reis, R. (2013). The portuguese slump and crash and the euro crisis. *Brookings Papers on Economic Activity*, 143–193. 1

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control 33*(3), 649–665. 3.1, D.1

Restuccia, D. and R. Rogerson (2017). The causes and costs of misallocation. *Journal of Economic Perspectives 31*(3), 151–74. 1

Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy 90*(6), 1187–1211. 2.4

Trimborn, T., K.-J. Koch, and T. Steger (2008). Multidimensional transitional dynamics: A simple numerical procedure. *Macroeconomic Dynamics 12*(3), 301â319. C, D.1

Winberry, T. (2018). A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics 9*(3), 1123–1151. 3.1, D.1, D.1

Woodford, M. (2003). Interest and prices: Foundations of a theory of monetary policy. *Princeton University Press*. 1, 4.2, 4.3

Zanetti, F. and M. Hamano (2020). Monetary policy, firm heterogeneity, and product variety. Technical report, University of Oxford, Department of Economics. 3

Online appendix

A Further details on the model

A.1 Entrepreneur's intertemporal problem

The Hamilton-Jacobi-Bellman (HJB) equation of the entrepreneur is given by

$$r_t V_t(z,a) = \max_{d_t \geq 0} d_t + s_t^a(z,a,d)\frac{\partial V}{\partial a} + \mu(z)\frac{\partial V}{\partial z} + \frac{\sigma^2(z)}{2}\frac{\partial^2 V}{\partial z^2} + \eta\left(q_t a_t - V_t(z,a)\right) + \frac{\partial V}{\partial t}.$$

We guess and verify a value function of the form $V_t(z,a) = \kappa_t(z)\, q_t a$. The first order condition is

$$\kappa_t(z) - 1 = \lambda_d \text{ and } \min\{\lambda_d, d_t\} = 0,$$

where $\lambda_d = 0$ if $\kappa_t(z) = 1$. If $\kappa_t(z) > 1\ \forall z, t$, then $d_t = 0$ and the firm does not pay dividends until it closes down. If this is the case, then the value of $\kappa_t(z)$ can be obtained from

$$\begin{aligned}&(r_t + \eta)\,\kappa_t(z)\, q_t = \\ &\eta q_t + (\gamma \max\{z_t\varphi_t - R_t, 0\} + R_t - \delta q_t)\kappa_t(z) + \mu(z) q_t \frac{\partial \kappa_t}{\partial z} + \frac{\sigma^2(z)}{2} q_t \frac{\partial^2 \kappa_t}{\partial z^2} + \frac{\partial\left(q_t \kappa_t\right)}{\partial t}.\end{aligned} \tag{45}$$

Lemma. $\kappa_t(z) > 1\ \forall z, t$

Proof. The drift of the entrepreneur's capital holdings is

$$s_t^a = \frac{1}{q_t}\left[(\gamma \max\{z_t\varphi_t - R_t, 0\} + R_t - \delta q_t\right] \geq \frac{R_t - \delta q_t}{q_t}$$

which is expected to hold with strict inequality eventually if $\exists\ \mathbb{P}(z_t \geq z_t^*) > 0$ (which is satisfied in equilibrium since z is unbounded), and hence

$$\mathbb{E}_0 a_t = \mathbb{E}_0 a_0 e^{\int_0^t s_u^a du} > a_0 e^{\int_0^t \frac{R_s - \delta q_s}{q_s} ds}. \tag{46}$$

The value function is then

$$
\begin{aligned}
\kappa_{t_0}(z)\, q_{t_0} a_{t_0} &= V_{t_0}(z, a_{t_0}) \\
&= \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t (r_s+\eta) ds} \left(d_t + \eta q_t a_t\right) dt \\
&\geq \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t (r_s+\eta) ds} \eta q_t a_t dt = \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t \left(\overbrace{\frac{R_s - \delta q_s + \dot{q}_s}{q_s}}^{r_s} + \eta\right) ds} \eta q_t a_t dt \\
&= \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t \left(\frac{R_s - \delta q_s}{q_s} + \eta\right) ds - \log \frac{q_t}{q_{t_0}}} \eta q_t a_t dt = \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t \left(\frac{R_s - \delta q_s}{q_s} + \eta\right) ds} \eta q_{t_0} a_t dt \\
&> \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^t \left(\frac{R_s - \delta q_s}{q_s} + \eta\right) ds} \eta q_{t_0} a_{t_0} e^{\int_0^t \frac{R_s - \delta q_s}{q_s} ds} dt = \int_0^\infty e^{-\eta t} \eta q_{t_0} a_{t_0} dt = q_{t_0} a_{t_0},
\end{aligned}
$$

where in the first equality we have employed the linear expression of the value function, in the second equation (5), in the third the fact that dividends are non-negative, in the fourth the definition of the real rate 17 and in the last line the inequality (46). Hence $\kappa_{t_0}(z) > 1$ for any t_0.

A.2 New Keynesian Philips curve

The proof is similar to that of Lemma 1 in Kaplan et al. (2018). The Hamilton-Jacobi-Bellman (HJB) equation of the retailer's problem is

$$
r_t V_t^r(p) = \max_\pi \left(\frac{p - P_t^y(1-\tau)}{P_t}\right) \left(\frac{p}{P_t}\right)^{-\varepsilon} Y_t - \frac{\theta}{2}\pi^2 Y_t + \pi p \frac{\partial V^r}{\partial p} + \frac{\partial V^r}{\partial t},
$$

where where $V_t^r(p)$ is the real value of a retailer with price p. The first order and envelope conditions for the retailer are

$$
\begin{aligned}
\theta \pi Y_t &= p \frac{\partial V^r}{\partial p}, \\
(r - \pi) \frac{\partial V^r}{\partial p} &= \left(\frac{p}{P_t}\right)^{-\varepsilon} \frac{Y_t}{P_t} - \varepsilon \left(\frac{p - P_t^y(1-\tau)}{P_t}\right) \left(\frac{p}{P_t}\right)^{-\varepsilon-1} \frac{Y_t}{P_t} + \pi p \frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p}.
\end{aligned}
$$

In a symmetric equilibrium we will have $p = P$, and hence

$$\begin{aligned}
\frac{\partial V^r}{\partial p} &= \frac{\theta\pi Y_t}{p}, \\
(r-\pi)\frac{\partial V^r}{\partial p} &= \frac{Y_t}{p} - \varepsilon\left(\frac{p - P_t^y(1-\tau)}{p}\right)\frac{Y_t}{p} + \pi p\frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p}.
\end{aligned} \tag{47}$$

Deriving (47) with respect to time gives

$$\pi p\frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p} = \frac{\theta\pi\dot{Y}}{p} + \frac{\theta\dot{\pi}Y}{p} - \frac{\theta\pi^2 Y}{p},$$

and substituting into the envelope condition and dividing by $\frac{\theta Y}{p}$ we obtain

$$\left(r - \frac{\dot{Y}}{Y}\right)\pi = \frac{1}{\theta}\left(1 - \varepsilon\left(1 - \frac{P_t^y(1-\tau)}{p}\right)\right) + \dot{\pi}.$$

Finally, rearranging we obtain the New Keynesian Phillips curve

$$\left(r - \frac{\dot{Y}}{Y}\right)\pi = \frac{\varepsilon}{\theta}\left(\frac{1-\varepsilon}{\varepsilon} + \tilde{m}\right) + \dot{\pi}.$$

A.3 Capital producers' problem

The problem of the capital producer is

$$W_t = \max_{\iota_t, K_t} \mathbb{E}_0 \int_0^\infty e^{-\int_0^t r_s ds}\left(q_t\iota_t - \iota_t - \Xi(\iota_t)\right)K_t dt. \tag{48}$$

$$\dot{K}_t = (\iota_t - \delta)K_t, \tag{49}$$

We construct the Hamiltonian

$$H = \left(q_t\iota_t - \iota_t - \Xi(\iota_t)\right)K_t + \lambda_t(\iota_t - \delta)K_t$$

with first-order conditions

$$\left(q_t - 1 - \Xi'(\iota_t)\right) + \lambda_t = 0 \tag{50}$$

$$\left(q_t\iota_t - \iota_t - \Xi\left(\iota_t\right)\right) + \lambda_t\left(\iota_t - \delta\right) = r_t\lambda_t - \dot{\lambda}_t \tag{51}$$

Taking the time derivative of equation (50)

$$\dot{\lambda}_t = -\left(\dot{q}_t - \Xi''\left(\iota_t\right) i_t\right)$$

which, combined with (51), yields

$$\left(q_t\iota_t - \iota_t - \Xi\left(\iota_t\right)\right) - \left(q_t - 1 - \Xi'\left(\iota_t\right)\right)\left(\iota_t - \delta - r_t\right) = \left(\dot{q}_t - \Xi''\left(\iota_t\right) i_t\right)$$

Rearranging we get

$$r_t = \left(\iota_t - \delta\right) + \frac{\dot{q}_t - \Xi''\left(\iota_t\right) i_t}{q_t - 1 - \Xi'\left(\iota_t\right)} - \frac{q_t\iota_t - \iota_t - \Xi\left(\iota_t\right)}{q_t - 1 - \Xi'\left(\iota_t\right)}.$$

A.4 Household's problem

We can rewrite the household's problem as

$$W_t = \max_{C_t, L_t, D_t, B_t^N, S_t^N} \mathbb{E}_0 \int_0^\infty e^{-\rho_t^h t}\left(\frac{C_t^{1-\zeta}}{1-\zeta} - \Upsilon\frac{L_t^{1+\vartheta}}{1+\vartheta}\right) dt. \tag{52}$$

$$\begin{aligned} s.t. \quad \dot{D}_t &= \left[\left(R_t - \delta q_t\right) D_t + w_t L_t - C_t - S_t^N + \Pi_t\right] / q_t, && (53)\\ \dot{B_t^N} &= S_t^N + \left(i_t - \pi_t\right) B_t^N, && (54) \end{aligned}$$

where S_t^N is the investment into nominal bonds.

The Hamiltonian is

$$H = \left(\frac{C_t^{1-\zeta}}{1-\zeta} - \Upsilon\frac{L_t^{1+\vartheta}}{1+\vartheta}\right)$$
$$+\varrho_t\left[\left(\left(R_t - \delta q_t\right) D_t + w_t L_t - C_t - S_t^N + \left(q_t\iota_t - \iota_t - \Phi\left(\iota_t\right)\right) K_t + \Pi_t\right) / q_t\right] + \eta_t\left[S_t^N + \left(i_t - \pi_t\right) B_t^N\right]$$

The first order conditions are

$$C_t^{-\zeta} - \varrho_t / q_t = 0 \tag{55}$$

$$-\Upsilon L_t^{\vartheta}+\varrho_t w_t/q_t=0 \tag{56}$$

$$-\varrho_t/q_t+\eta_t=0 \tag{57}$$

$$\dot{\varrho}_t=\rho_t^h\varrho_t-\varrho_t\left(R_t-\delta q_t\right)/q_t \tag{58}$$

$$\dot{\eta}_t=\rho_t^h\eta_t-\eta_t\left[(i_t-\pi_t)\right] \tag{59}$$

(55) and (56) combine to the optimality condition for labor

$$w_t=\frac{L_t^{\vartheta}}{C_t^{-\eta}},$$

(55) can be rewritten as

$$\varrho_t=C_t^{-\eta}q_t$$

Now take derivative with respect to time

$$\dot{\varrho}_t=-\eta C_t^{-\eta-1}\,\dot{C}_t q_t+C_t^{-\eta}\,\dot{q}_t$$

and plug this into (58) and rearrange to get the first Euler equation

$$\frac{\dot{C}_t}{C_t}=\frac{\frac{R_t-\delta q_t+\dot{q}_t}{q_t}-\rho_t^h}{\eta}$$

(57) can be rewritten as

$$\eta_t=\varrho_t/q_t$$

Now take derivative with respect to time

$$\dot{\eta}_t=\frac{\dot{\varrho}_t q_t-\varrho_t\dot{q}}{q_t^2}$$

Use these two expressions and the definition of $\dot{\varrho}_t$ in (59) to get the second Euler equation

$$\frac{\dot{C}_t}{C_t}=\frac{(i_t-\pi_t)-\rho_t^h}{\eta}$$

Combining the two Euler equations, we get the Fisher equation

$$\frac{R_t - \delta q_t + \dot{q}_t}{q_t} = (i_t - \pi_t)$$

Finally using the definition of $r_t \equiv \frac{R_t - \delta q_t + \dot{q}_t}{q_t}$ we can rewrite the first Euler equation and the Fisher equation as in the main text.

A.5 Distribution

The joint distribution of net worth and productivity is given by the Kolmogorov Forward equation

$$\frac{\partial g_t(z,a)}{\partial t} = -\frac{\partial}{\partial a}[g_t(z,a)s_t(z)a] - \frac{\partial}{\partial z}[g_t(z,a)\mu(z)] + \frac{1}{2}\frac{\partial^2}{\partial z^2}[g_t(z,a)\sigma^2(z)] - \eta g_t(z,a) + \eta/\psi g_t(z,a/\psi), \tag{60}$$

where $1/\psi g_t(z, a/\psi)$ is the distribution of entry firms.

To characterize the law of motion of net-worth shares, defined as $\omega_t(z) = \frac{1}{A_t}\int_0^\infty a g_t(z,a)da$, first we take the derivative of $\omega_t(z)$ wrt time

$$\frac{\partial \omega_t(z)}{\partial t} = -\frac{\dot{A}_t}{A_t^2}\int_0^\infty a g_t(z,a)da + \frac{1}{A_t}\int_0^\infty a\frac{\partial g_t(z,a)}{\partial t}da. \tag{61}$$

Next, we plug in the derivative of $g_t(z,a)$ wrt time from equation(60) into equation (61),

$$\begin{aligned}\frac{\partial \omega_t(z)}{\partial t} = &-\frac{\dot{A}_t}{A_t^2}\int_0^\infty a g_t(z,a)da + \frac{1}{A_t}\int_0^\infty a\left(-\frac{\partial}{\partial a}[g_t(z,a)s_t(z)a]\right)da \\ &-\frac{\partial}{\partial z}\mu(z)\frac{1}{A_t}\int_0^\infty a g_t(z,a)da + \frac{1}{2}\frac{\partial^2}{\partial z^2}\sigma^2(z)\frac{1}{A_t}\int_0^\infty a g_t(z,a)da \\ &-\frac{1}{A_t}\int_0^\infty \eta a g_t(z,a)da + \frac{1}{A_t}\int_0^\infty \eta a/\psi g_t(z,a/\psi)da.\end{aligned}$$

Using integration by parts and the definition of net worth shares, we obtain the second order partial differential equation that characterizes the law of motion of net-worth shares,

$$\frac{\partial \omega_t(z)}{\partial t} = \left[s_t(z) - \frac{\dot{A}_t}{A_t} - (1-\psi)\eta \right] \omega_t(z) - \frac{\partial}{\partial z}\mu(z)\omega_t(z) + \frac{1}{2}\frac{\partial^2}{\partial z^2}\sigma^2(z)\omega_t(z). \tag{62}$$

The stationary distribution is therefore given by the following second order partial differential equation,

$$0 = (s(z) - (1-\psi)\eta)\,\omega(z) - \frac{\partial}{\partial z}\mu(z)\omega(z) + \frac{1}{2}\frac{\partial^2}{\partial z^2}\sigma^2(z)\omega(z). \tag{63}$$

Remember that $s_t^a(z_t, a_t, c_t) = \frac{1}{q_t}\left[\Phi_t(z_t, a_t) + (R_t - \delta q_t)a_t\right]$, since entrepreneurs distribute zero dividends while constrained.

A.6 Market clearing and aggregation

Define the cumulative function of net-worth shares as

$$\Omega_t(z) = \int_0^z \omega_t(z)dz. \tag{64}$$

Using the optimal choice for k_t from equation (7), we obtain

$$K_t = \int k_t(z,a)dG_t(z,a) = \int_{z_t^*}^{\infty} \int \gamma a \frac{1}{A_t} g_t(z,a)dadzA_t = \gamma(1 - \Omega(z_t^*))A_t. \tag{65}$$

By combining equations (27), (28) and (65), and solving for A_t,we obtain

$$A_t = \frac{D_t}{(1 - \Omega(z_t^*)) - 1}, \tag{66}$$

Labor market clearing implies

$$L_t = \int_0^{\infty} l_t(z,a)dG_t(z,a). \tag{67}$$

Define the following auxiliary variable,

$$X_t \equiv \int_{z_t^*}^{\infty} z\omega_t(z)dz = \mathbb{E}\left[z \mid z > z_t^*\right](1 - \Omega(z_t^*)). \tag{68}$$

Using labor demand from (8) , X_t and using the definition of φ_t, we obtain

$$L_t = \int_0^\infty \left(\frac{\varphi_t}{\alpha m_t}\right)^{\frac{1}{1-\alpha}} z_t \gamma a_t dG_t(z,a) = \left(\frac{\varphi_t}{\alpha m_t}\right)^{\frac{1}{1-\alpha}} \gamma A_t X_t. \tag{69}$$

Plugging in (8) into production function (1), and using again the definition of shares, we obtain

$$Y_t = \int \underbrace{\frac{z_t \varphi_t}{\alpha m_t}\gamma a}_{y_t(z,a)} dG_t(z,a) = \frac{\varphi_t}{\alpha m_t} X_t \gamma A_t = Z_t A_t^\alpha {L_t}^{1-\alpha}, \tag{70}$$

where in the last equality we have used equation (69), and we have defined

$$Z_t = (\gamma X_t)^\alpha . \tag{71}$$

Aggregate profits of retailers are given by

$$\Phi_t^{Agg} = \int \gamma \max\{z_t\varphi_t - R_t, 0\} a_t dG_t(z,a) = [\varphi_t X_t - R_t(1-\Omega(z^*))]\gamma A_t. \tag{72}$$

We can also write the aggregate production in terms of physical capital,

$$Y_t = Z_t K_t^\alpha {L_t}^{1-\alpha}, \tag{73}$$

where the TFP term Z_t is defined as

$$Z_t = \left(\frac{X_t}{(1-\Omega(z_t^*))}\right)^\alpha = \left(\mathbb{E}\left[z \mid z > z_t^*\right]\right)^\alpha . \tag{74}$$

Aggregating the budget constraint of all input good firms, using the linearity of savings policy (11) and using (66), we obtain

$$\begin{aligned}\dot{A}_t &= \int \dot{a} dG(z,a,t) - \eta \int (1-\psi) a_t dG(z,a,t) \ = \\ &= \int_0^\infty \frac{1}{q_t}(\gamma \max\{z_t\varphi_t - R_t, 0\} + R_t - \delta q_t - q_t(1-\psi)\eta) a_t dG(z,a),\end{aligned}$$

Dividing by A_t both sides of this equation, using the definition of net worth shares and the fact that these integrate up to one, we obtain

$$\frac{\dot{A}_t}{A_t} = \frac{1}{q_t}(\gamma\varphi_t X_t - R_t\gamma(1-\Omega(z_t^*)) + R_t - \delta q_t - q_t(1-\psi)\eta). \tag{75}$$

Using the definition of X_t, and substituting φ_t using equation (69), we can simplify equation (75) as

$$\frac{\dot{A}_t}{A_t} = \frac{1}{q_t}(\alpha m_t Z_t {A_t}^{\alpha-1} {L_t}^{1-\alpha} - R_t\gamma(1-\Omega(z_t^*)) + R_t - \delta q_t - q_t(1-\psi)\eta). \tag{76}$$

Finally, we can obtain factor prices

$$w_t = (1-\alpha) m_t Z_t {A_t}^{\alpha} {L_t}^{-\alpha} \tag{77}$$

$$R_t = \alpha m_t Z_t A_t^{\alpha-1} {L_t}^{1-\alpha} \frac{z_t^*}{\gamma X_t} \tag{78}$$

where wages come from substituting the definition of φ_t into equation (69); and interest rates come from plugging in the wage expression (77) into the cut-off rule (10) and using equation (66). We could equivalently write equation (78) in terms of real rate of return r_t :

$$r_t = \frac{1}{q_t}\left(\alpha m_t Z_t A_t^{\alpha-1} {L_t}^{1-\alpha} \frac{z_t^*}{\gamma X_t}\right) - \delta + \frac{\dot{q}}{q_t} \tag{79}$$

We can easily get these equations in terms of capital instead of net worth by simply using equation (65), i.e. $A_t = \frac{K_t}{\gamma(1-\Omega(z_t^*))}$, and using that $\mathbb{E}[z \mid z > z_t^*] = \frac{X_t}{(1-\Omega(z_t^*))} = \frac{\int_{z_t^*}^{\infty} z\omega_t(z)dz}{(1-\Omega(z_t^*))}$ (see equation (71) and (74)).

A.7 Full set of equations

The competitive equilibrium economy is described by the following 22 equations, for the 22 variables $\{\omega(z), s(z), w, r, q, \varphi, K, A, L, C, D, Z, \mathbb{E}[z \mid z > z_t^*], \Omega, z^*, \iota, \pi, m, \tilde{m}, i, Y, T\}$. Remember that $\mu(z) = z\left(-\varsigma_z \log z + \frac{\sigma^2}{2}\right)$ and $\sigma(z) = \sigma_z z$, and that government bonds are in zero net supply ($B_t^N = 0$, hence $X_t = 0$). Except from the last equation (Taylor rule), the other 21 equations are the constraints of the Ramsey problem described in Section 2.8.

$$
\begin{aligned}
\frac{\partial \omega_t\left(z\right)}{\partial t} &= \left(s_t(z) - (1-\psi)\eta - \frac{\dot{A}_t}{A_t}\right)\omega_t\left(z\right) - \frac{\partial}{\partial z}\left[\mu(z)\omega_t\left(z\right)\right] + \frac{1}{2}\frac{\partial^2}{\partial z^2}\left[\sigma^2(z)\omega_t\left(z\right)\right] \\
s_t(z) &= \frac{1}{q_t}\left(\gamma \max\left\{z_t\varphi_t - R_t, 0\right\} + R_t - \delta q_t\right) \\
\Omega_t(z^*) &= \int_0^{z^*} \omega_t\left(z\right)dz \\
\varphi_t &= \alpha\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}} \\
\tilde{m}_t &= m_t(1-\tau) \\
w_t &= (1-\alpha)m_t\tilde{Z}_t {K_t}^{\alpha}{L_t}^{-\alpha} \\
R_t &= \alpha m_t\tilde{Z}_t K_t^{\alpha-1}{L_t}^{1-\alpha}\frac{z_t^*}{\mathbb{E}\left[z \mid z > z_t^*\right]} \\
\frac{\dot{A}_t}{A_t} &= \frac{1}{q_t}\left[\gamma(1-\Omega(z_t^*))\left(\alpha m_t Z_t K_t^{\alpha-1}{L_t}^{1-\alpha} - R_t\right) + R_t - \delta q_t - q_t(1-\psi)\eta)\right] \\
K_t &= A_t + D_t \\
\dot{K}_t &= (\iota_t - \delta)K_t \\
A_t &= \frac{D_t}{\gamma(1-\Omega(z_t^*)) - 1} \\
\tilde{Z}_t &= \left(\mathbb{E}\left[z \mid z > z_t^*\right]\right)^{\alpha} \\
\mathbb{E}\left[z \mid z > z_t^*\right] &= \frac{\int_{z_t^*}^{\infty} z\omega_t(z)dz}{(1-\Omega(z_t^*))} \\
\frac{\dot{C}_t}{C_t} &= \frac{r_t - \rho_t^h}{\eta} \\
w_t &= \frac{\Upsilon L_t^{\vartheta}}{C_t^{-\eta}} \\
\dot{D}_t &= \left[\left(R_t - \delta q_t\right)D_t + w_t L_t - C_t + T_t\right]/q_t \\
r_t &= i_t - \pi_t \\
r_t &= \frac{R_t - \delta q_t + \dot{q}_t}{q_t} \\
\left(q_t - 1 - \Phi'\left(\iota_t\right)\right)\left(r_t - \left(\iota_t - \delta\right)\right) &= \dot{q}_t - \Phi''\left(\iota_t\right)i_t - \left(q_t\iota_t - \iota_t - \Phi\left(\iota_t\right)\right) \\
\left(r_t - \frac{\dot{Y}_t}{Y_t}\right)\pi_t &= \frac{\varepsilon}{\theta}\left(\tilde{m}_t - m^*\right) + \dot{\pi}_t, \qquad m^* = \frac{\varepsilon - 1}{\varepsilon} \\
Y_t &= Z_t K_t^{\alpha}{L_t}^{1-\alpha} \\
T_t &= \left(1 - m_t\right)Y_t - \frac{\theta}{2}\pi_t^2 Y_t + (1-\psi)\eta A_t + \left[\iota_t q_t - \iota_t - \frac{\phi^k}{2}(\iota_t - \delta)^2\right]K_t \\
di &= -\upsilon\left(i_t - \left(\rho_t^h + \phi\left(\pi_t - \bar{\pi}\right) + \bar{\pi}\right)\right)dt.
\end{aligned}
$$

A.8 The net-worth channel of monetary policy

TFP is given by equation (32)

$$Z_t = \left(\frac{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz}{\int_{z_t^*}^{\infty} \omega_t(z)\,dz} \right)^{\alpha}.$$

As the distribution $\omega_t(z)$ is predetermined at time t, the net-worth channel does not operate on impact. It may affect, however, TFP dynamics. We compute the growth rate of TFP keeping z^* constant as

$$\begin{aligned}
\left.\frac{1}{Z_t}\frac{dZ_t}{dt}\right|_{z^*} &= \left.\frac{d\log Z_t}{dt}\right|_{z^*} = \alpha\left.\left[\frac{d}{dt}\left(\log\int_{z_t^*}^{\infty} z\omega_t(z)\,dz\right) - \frac{d}{dt}\left(\log\int_{z_t^*}^{\infty}\omega_t(z)\,dz\right)\right]\right|_{z^*} \\
&= \frac{\int_{z^*}^{\infty} z\dot{\omega}_t(z)\,dz}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty} \dot{\omega}_t(z)\,dz}{\int_{z_t^*}^{\infty} \omega_t(z)\,dz}.
\end{aligned}$$

The derivative of with respect to a price $x_t = \{r_t, w_t, m_t, q_t\}$ is

$$\frac{\partial}{\partial x_t}\left.\frac{d\log Z_t}{dt}\right|_{z^*} = \frac{\int_{z^*}^{\infty} z\frac{\partial \dot{\omega}_t(z)}{\partial x_t}dz}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty} \frac{\partial \dot{\omega}_t(z)}{\partial x_t}dz}{\int_{z_t^*}^{\infty} \omega_t(z)\,dz},$$

where

$$\begin{aligned}
\left.\frac{\partial \dot{\omega}_t(z)}{\partial x_t}\right|_{z^*} &= \left.\frac{\partial}{\partial x_t}\left(\tilde{\Phi}_t(z) + \tilde{\Xi}_t\right)\right|_{z^*}\omega(z), \\
\tilde{\Phi}_t(z) &= \max\left\{\frac{\gamma\alpha}{q_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}(z - z^*), 0\right\}, \\
\tilde{\Xi}_t &= \frac{R_t - \delta q_t}{q_t} - \frac{\dot{A}_t}{A_t} - (1-\psi)\eta = -\frac{\gamma(1-\Omega_t(z^*))\left(\alpha m_t Z_t K_t^{\alpha-1} L_t^{1-\alpha} - R_t\right)}{q_t}.
\end{aligned}$$

Then we have:

$$\begin{aligned}
\frac{\partial}{\partial x_t}\left.\frac{d\log Z_t}{dt}\right|_{z^*} &= \frac{\int_{z^*}^{\infty} z\frac{\partial \tilde{\Phi}(z)}{\partial x_t}\omega_t(z)\,dz}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty} \frac{\partial \tilde{\Phi}(z)}{\partial x_t}\omega_t(z)\,dz}{\int_{z_t^*}^{\infty} \omega_t(z)\,dz} \\
&+ \frac{\partial \tilde{\Xi}_t}{\partial x_t}\underbrace{\left(\frac{\int_{z^*}^{\infty} z\omega_t(z)\,dz}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty} \omega_t(z)\,dz}{\int_{z_t^*}^{\infty} \omega_t(z)\,dz}\right)}_{0}.
\end{aligned}$$

This expression shows how only the excess investment rate $\tilde{\Phi}(z)$ matters to understand the impact of changes in prices on the growth rate of TFP.

In the case of the real rate (direct effect), $x_t = r_t$, we have $\left.\frac{\partial \tilde{\Phi}_t(z)}{\partial r_t}\right|_{z^*} = 0$, and $\left.\frac{\partial \tilde{\Xi}_t}{\partial R_t}\right|_{z^*} = \gamma(1-\Omega_t(z^*))$, thus

$$\left.\frac{\partial}{\partial r_t}\frac{d\log Z_t}{dt}\right|_{z^*} = 0.$$

This implies that changes in the real rate do not affect the growth rate of TFP.

In the case of the wage, $x_t = w_t$, we have $\left.\frac{\partial \tilde{\Phi}_t(z)}{\partial w_t}\right|_{z^*} = -\frac{\gamma(1-\alpha)}{q_t w_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}(z-z^*) <$ 0, and $\left.\frac{\partial \tilde{\Xi}_t}{\partial w_t}\right|_{z^*} = 0$, thus

$$\left.\frac{\partial}{\partial w_t}\frac{d\log Z_t}{dt}\right|_{z^*} = -\frac{\gamma(1-\alpha)}{q_t w_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}\left(\frac{\int_{z^*}^{\infty}(z-z^*)\,z\omega_t(z)}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty}(z-z^*)\,\omega_t(z)}{\int_{z_t^*}^{\infty}\omega_t(z)\,dz}\right).$$

To uncover the sign, we analyze the term

$$\frac{\int_{z^*}^{\infty}(z-z^*)\,z\omega_t(z)}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty}(z-z^*)\,\omega_t(z)}{\int_{z_t^*}^{\infty}\omega_t(z)\,dz} = \frac{\int_{z^*}^{\infty} z^2\omega_t(z)}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty} z\omega_t(z)}{\int_{z_t^*}^{\infty}\omega_t(z)\,dz}.$$

We define $\bar{\omega}_t(z) \equiv \frac{\omega_t(z)}{\int_{z^*}^{\infty}\omega_t(z)dz}\mathbb{I}_{z>z^*}$ and $\tilde{\omega}_t(z) \equiv \frac{z\omega_t(z)}{\int_{z^*}^{\infty}\omega_t(z)zdz}\mathbb{I}_{z>z^*}$. These are continuous probability density functions over the domain $[z^*, \infty)$, as they are non-negative and sum up to 1. They satisfy the monotone likelihood ratio condition as

$$I(z) = \frac{\tilde{\omega}_t(z)}{\bar{\omega}_t(z)} = z\frac{\int_{z^*}^{\infty} z\omega_t(z)\,dz}{\int_{z^*}^{\infty}\omega_t(z)\,dz}$$

is non decreasing. This implies that function $\tilde{\omega}_t(z)$ dominates $\bar{\omega}_t(z)$ first-order stochastically. Hence

$$\frac{\int_{z^*}^{\infty} z\omega_t(z)}{\int_{z_t^*}^{\infty}\omega_t(z)\,dz} = \mathbb{E}_{\bar{\omega}_t(z)}[z] = \int_{z^*}^{\infty} z\bar{\omega}_t(z)zdz \quad < \quad \int_{z_t^*}^{\infty} z\tilde{\omega}_t(z)\,dz = \mathbb{E}_{\tilde{\omega}_t(z)}[z] = \frac{\int_{z^*}^{\infty} z^2\omega_t(z)}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz}.$$

The sign of a change in wages on TFP growth thus coincides with that of the excess investment rate:

$$\left.\frac{\partial \tilde{\Phi}_t(z)}{\partial w_t}\right|_{z^*} < 0 \Rightarrow \left.\frac{\partial}{\partial w_t}\frac{d\log Z_t}{dt}\right|_{z^*} < 0.$$

It is trivial to check that the same happens in the case of other prices, m_t, q_t, that is, that the sign of their impact on TFP growth is captured by the slope of the excess investment rate $\tilde{\Phi}_t(z)$. Take, for instance, input prices. We have $\left.\frac{\partial \tilde{\Phi}_t(z)}{\partial m_t}\right|_{z^*} = \frac{\gamma}{q_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1-\alpha}{\alpha}}(z-z^*) > 0$, and $\left.\frac{\partial \tilde{\Xi}_t}{\partial m_t}\right|_{z^*} = -\frac{\gamma(1-\Omega_t(z^*))\alpha Z_t K_t^{\alpha-1} L_t{}^{1-\alpha}}{q_t}$, thus

$$\left.\frac{\partial}{\partial m_t}\frac{d\log Z_t}{dt}\right|_{z^*} = \frac{\gamma}{q_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1-\alpha}{\alpha}}\left(\frac{\int_{z^*}^{\infty}(z-z^*)\, z\omega_t(z)}{\int_{z_t^*}^{\infty} z\omega_t(z)\,dz} - \frac{\int_{z^*}^{\infty}(z-z^*)\,\omega_t(z)}{\int_{z_t^*}^{\infty}\omega_t(z)\,dz}\right) > 0.$$

A.9 Baseline vs complete markets

In this appendix we want to highlight the differences between the model presented in this paper and the standard representative agent New Keynesian model with capital (complete markets). Note first that the baseline economy collapses to the standard complete market economy if the collateral constraint is made infinitely slack (assuming that the support of entrepreneurs productivity distribution is bounded above). In that case entrepreneurial net worth becomes irrelevant and only the entrepreneur with the highest level of productivity z_t produces, since she can frictionlessly rent all the capital in the economy. Her productivity determines aggregate productivity $Z_t = (z_t^{max})^{\alpha}$.In contrast, in the baseline model with incomplete markets, entrepreneurs' firms can only use capital up to a multiple γ of their net worth , i.e. $\gamma a_t \leq k_t$. Thus entrepreneurs need to accumulate net worth (in units of capital) to alleviate these financial frictions. Hence, in the baseline model, the distribution of aggregate capital across entrepreneurs and the representative household matters and aggregate productivity depends on the expected productivity of constrained firms, $Z = (\mathbb{E}[z \mid z > z_t^*])^{\alpha}$. The rest of the agents (retailers, final good producers, capital producers) are identical in both economies.

Below we report the equilibrium conditions in the complete markets economy. Comparing them with those of the baseline economy reveals that they are identical up to the fact that in the baseline $\tilde{Z}_t$ is endogenous (and determined by a bunch of extra equations) and up to a term in the condition equating the rental rate of capital R_t with the marginal return on capital.

The competitive equilibrium of the complete market model with capital consists of the following equations 16 equations, for the 16 variables $\{w, r, q, \varphi, K, L, C, D, Z, \iota, \pi, m, \tilde{m}, i, Y, T\}$:

$$
\begin{aligned}
\varphi_t &= \alpha \left(\frac{(1-\alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}} \\
\tilde{m}_t &= m_t(1-\tau) \\
w_t &= (1-\alpha) m_t \tilde{Z}_t {K_t}^{\alpha} {L_t}^{-\alpha} \\
R_t &= \alpha m_t \tilde{Z}_t K_t^{\alpha-1} {L_t}^{1-\alpha} \\
K_t &= D_t \\
\dot{K}_t &= (\iota_t - \delta) K_t \\
\tilde{Z}_t &= (_t)^{\alpha} \\
\frac{\dot{C}_t}{C_t} &= \frac{r_t - \rho_t^h}{\eta} \\
w_t &= \frac{\Upsilon L_t^{\vartheta}}{C_t^{-\eta}} \\
\dot{D}_t &= \left[\left(R_t - \delta q_t \right) D_t + w_t L_t - C_t + T_t \right] / q_t \\
r_t &= i_t - \pi_t \\
r_t &= \frac{R_t - \delta q_t + \dot{q}_t}{q_t} \\
\left(q_t - 1 - \Phi'\left(\iota_t\right) \right) \left(r_t - \left(\iota_t - \delta \right) \right) &= \dot{q}_t - \Phi''\left(\iota_t\right) i_t - \left(q_t \iota_t - \iota_t - \Phi\left(\iota_t\right) \right) \\
\left(r_t - \frac{\dot{Y}_t}{Y_t} \right) \pi_t &= \frac{\varepsilon}{\theta} \left(\tilde{m}_t - m^* \right) + \dot{\pi}_t, \qquad m^* = \frac{\varepsilon - 1}{\varepsilon} \\
Y_t &= Z_t K_t^{\alpha} {L_t}^{1-\alpha} \\
T_t &= \left(1 - m_t\right) Y_t - \frac{\theta}{2} \pi_t^2 Y_t + \left[\iota_t q_t - \iota_t - \frac{\phi^k}{2} (\iota_t - \delta)^2 \right] K_t \\
di &= -\upsilon \left(i_t - \left(\rho_t^h + \phi \left(\pi_t - \bar{\pi} \right) + \bar{\pi} \right) \right) dt.
\end{aligned}
$$

B Empirical Appendix

B.1 Firm level data

The empirical exercise relies on annual firm balance-sheet data from the *Central de Balances Integrada* database (Integrated Central Balance Sheet Data Office Survey). Being a detailed administrative dataset, the main advantage is that it covers the quasi-universe of Spanish firms (see Almunia et al., 2018 for further details on the representativeness of this dataset). Our dependent variable, the investment rate, is defined as the log difference of firm's tangible capital between periods t and $t-1$. Firm's marginal revenue product of capital (MRPK) is proxied by the log of the ratio of value added over tangible capital. Leverage is computed as total debt (short-term plus long-term debt) divided by total assets. Net financial assets are constructed as the log difference between financial assets and financial liabilities, where financial assets include short-term financial investment, trade receivables, inventories and cash holdings; and financial liabilities include short-term debt, trade payables and long-term debt. We proxy for size using log total assets. Real sales growth is defined as the log difference of sales in two consecutive years. Variables are deflated using industry price level to preserve the firms' level price changes and consider a revenue-based measure of MRPK (Foster et al., 2008). We use the value-added price deflator for value added and sales, and the investment price deflator for capital and total assets. Descriptive statistics are reported in Table 3.

Data is cleaned following closely Ottonello and Winberry (2020). In particular, (i) observations with negative capital or value added are dropped; (ii) the investment rate and MRPK are winsorized at 0.5%; (ii) we use net financial assets over as a share of total assets to control for firms' savings, following Armenter and Hnatkovska, 2017, instead of net current assets (as Ottonello and Winberry (2020) do), and we drop values in absolute terms greater than 10; and (iii) negative values of leverage are dropped, as well as values higher than 10. While Ottonello and Winberry (2020) drop firms for which the time spell is shorter than 10 years, we prefer to consider the full sample of firms without imposing an arbitrary threshold, and we show that our results are robust considering a balanced sample where we keep only firms that are present in our dataset for the whole time period considered.

Table 3: **Descriptive statistics**

	mean	sd	min	max
ε_t^{MP}	-2.90	7.77	-17.99	7.94
ε_t^{MP} x $MRPK_{j,t-1}$	-0.00	0.08	-1.60	1.82
$MRPK_t$	-0.00	1.00	-10.09	10.25
g_t^{GDP} x $MRPK_{j,t}$	0.22	3.07	-40.36	46.81
$MRPK_{j,t}$ (not demeaned)	0.56	2.09	-5.47	6.22
Sales growth$_{j,t}$	0.00	1.00	-17.84	13.56
Total assets$_{j,t}$	0.00	1.00	-5.57	7.07
Leverage$_{j,t}$	-0.00	1.00	-0.57	25.95
Observations	9485676			

Notes: The table shows the mean (column 1), standard deviation (column 2), minimum and maximum value (column 3 and 4 respectively) of the main variables used in the analysis. ε_t^{MP} is the annualized monetary policy shock, renormalized so that a positive value is an expansionary shock.MRPK stands for the demeaned measure of MRPK explained in Section 3.3. MRPK, sales growth, total assets and leverage are standardized, as in Ottonello and Winberry (2020). MRPK (not demeaned) is the raw variable of MRPK. g_t^{GDP} stands for GDP growth.

B.2 Monetary policy shocks

We construct our yearly monetary policy shocks aggregating the monthly monetary policy shocks of Jarociński and Karadi (2020). Since firms have less time to react to shocks happening at the end of the year, ignoring this issue would lead to biased estimates. Therefore, similar to Ottonello and Winberry (2020), but on a month-year level instead of month-quarter, we apply a weighting scheme that aggregates the shocks happening in the fourth quarter of the previous year with increasing linear weight, and uses linear and decreasing weights in the current year. Namely, we add them using decreasing weights within the year $\omega_a(m)$, and increasing weights in the last quarter of the previous year $\omega_b(m)$, i.e.

$$\varepsilon_t^{MP} = \sum_{m \in t} \omega_a(m)\varepsilon_m^{MP} + \sum_{m \in q4_{t-1}} \omega_b(m)\varepsilon_m^{MP}.$$

This is equivalent to say that a shock in January of period t has more weight than a shock in December of the same year, exactly because firms take time to adjust their investment plans. Panel 1 of Figure 7 shows the time series of the shock built in this way. As a robustness check, as well as in order to reduce concerns about potential autocorrelation in the residuals, we use an alternative weighting scheme that aggregates the shocks in the same year with a simple linear decreasing weight, without considering previous year's shocks. Panel 2 of Figure 7 shows the time series of the shock built with this alternative weighting.

Panel 1 - Baseline weighting - ε_t^{MP} **Panel 2 -** Alternative weighting - ε_t^{MP2}

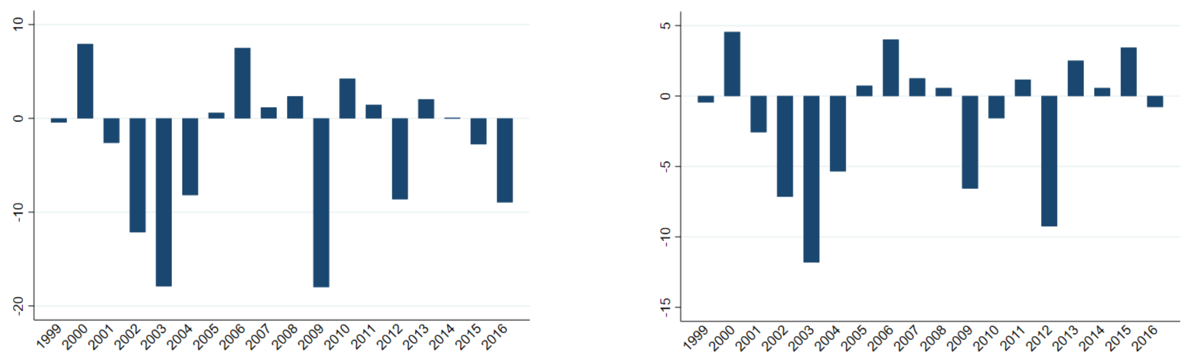

Figure 7: Monetary policy shocks at annual frequency.

Notes: **Panel 1** shows the monetary policy shocks at an annual frequency, applying a weighting scheme at aggregation that includes the shock in the fourth quarter of the previous period with an increasing linear weight and uses linear and decreasing weights in the current year. **Panel 2** shows the monetary policy shocks at an annual frequency, applying an alternative weighting, that is, a weighting scheme at aggregation with linear and decreasing weights in the current year only.

B.3 Robustness

In this section we check the robustness of our empirical results. We perform variations of the main empirical specification explained in the main text, equation (43), which we repeat here for the sake of completeness.

$$\Delta log\ k_{j,t} = \alpha_j + \alpha_{s,t} + \beta \left(MRPK_{j,t-1} - \mathbb{E}_j \left[MRPK_j \right] \right) \varepsilon_t^{MP} +' Z_{j,t-1} + u_{j,t}.$$

Following Ottonello and Winberry (2020) and Eberly et al. (2012), we control for the lagged of the dependent variable, i.e. firms' lagged investment rate, since it has been shown that it is a good predictor of a firm's current investment. Columns (1) and (2) in Table 4 show that results are robust to adding this variable, even stronger in magnitude, and R^2 does not change significantly. Columns (3) and (4) in Table 4 show the results considering the balanced panel, i.e. keeping only firms that we observe during the entire time sample period, in order to focus on pure incumbents. This does not only confirm the baseline results, but it shows that the effect can be even larger for incumbent firms. Columns (5) and (6) in Table 4 use the monetary policy shocks constructed using the alternative weighting scheme, ε_t^{MP2}. Results are still significant and of slightly larger magnitude. Finally, Columns (7) and (8) show the results using the baseline monetary policy shock ε^{MP}, but interacting this shock with the lagged MRPK in levels, instead of the demeaned standardized measure. The coefficients are still positive and significant. Summing up, all these exercises point at the robustness of the empirical support of the main mechanism of the model, that is, a higher heterogeneous response of investment for high MRPK firms to a monetary policy shock.

Table 4: **Robustness**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ϵ_t^{MP} x $MRPK_{t-1}$	0.238***	0.299***	0.177**	0.432***				
	(0.06)	(0.07)	(0.07)	(0.09)				
Inv_{t-1}	-0.0310***	-0.0259***						
	(0.00)	(0.00)						
ϵ_t^{MP2} x $MRPK_{t-1}$					0.166*	0.345***		
					(0.10)	(0.10)		
ϵ_t^{MP} x $MRPK_{t-1}$ (not demeaned)							0.0906**	0.243***
							(0.04)	(0.04)
Observations	4,162,114	4,094,537	283,835	263,397	5,567,706	4,169,950	5,567,706	4,169,950
R^2	0.279	0.283	0.153	0.162	0.267	0.285	0.267	0.286
MRPK control	YES	YES	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
Time-sector FE	YES	YES	YES	YES	YES	YES	YES	YES
Time-sector clustering	YES	YES	YES	YES	YES	YES	YES	YES
Panel	FULL	FULL	BALANCED	BALANCED	FULL	FULL	FULL	FULL

Notes: Results of estimating equation (43), departing from some of the specifications of the estimation in the main text (Section 3.3). Columns (1) and (2) include as control the lag of the investment rate ($log(k_{t-1}) - log(k_{t-2})$). Columns (3) and (4) restrict the sample to a balanced panel. Columns (5) and (6) consider the alternative yearly aggregation of the monetary policy shocks. Columns (7) and (8) use MRPK in levels, $MRPK$ (not demeaned), instead of the demeaned standardized value. Columns (1) , (3) and (5) use only lagged MRPK as controls, while columns (2), (4) and (6) include all the controls, lagged: MRPK, total assets, sales growth, leverage and net financial assets as a share of total assets; and the interaction of MRPK with GDP growth. Columns (1),(2), (5) and (6) use the demeaned standardized measure of MRPK explained in the main text, while columns (3)-(4) do not demean MRPK.

C Numerical Appendix

We discretize the model using a finite difference approach and compute non-linearly the responses to temporary change in parameters (an "MIT shock") using a Newton algorithm. Instead of time iterations over guesses for aggregate sequences, as is common in the literature, we use a global relaxation algorithm. This approach has been made popular in discrete-time models by Juillard et al. (1998) thanks to Dynare, but it is somewhat less common in continuous-time models (e.g. Trimborn et al., 2008). This approach helps to overcome the curse of dimensionality since in the sequence space the complexity of the problem grows only linearly in the number of aggregate variables, whereas the complexity of the state-space solution grows exponentially in the number of state variables. Recently Auclert et al. (2019) have exploited a particularly efficient variant of this approach in the context of heterogeneous-agent models.[14] We build on these contributions when we compute the optimal transition path. Again we make use of Dynare. We use its nonlinear Newton solver to compute both the steady state of the Ramsey problem and the optimal transition path under perfect foresight. To find the steady state, we provide Dynare with the steady state of the private equilibrium conditions as a function of the policy instrument.

C.1 Finite difference approximation of the Kolmogorov Forward equation

The KF equation is solved by a finite difference scheme following Achdou et al. (2017). It approximates the density $\omega_t(z)$ on a finite grid $z \in \{z_1, ..., z_J\}$, $t \in \{t_1, ..., t_N\}$ with steps Δz and time steps Δt. We use the notation $\omega_j^n := \omega_{n\Delta t}(z_j)$, $j = 1, ..., J$, $n = 0, .., N$. The KF equation is then approximated as

$$\frac{\omega_j^n - \omega_j^{n-1}}{\Delta t} = \left(s_n(z_j) - \frac{\dot{A}_n}{A_n} - (1-\psi)\eta\right)\omega_n(z_j) - \frac{\omega_j^n \mu(z_j) - \omega_{j-1}^n \mu(z_{j-1})}{\Delta z} + \frac{\omega_{j+1}^n \widetilde{\sigma}^2(z_{j+1}) + \omega_{j-1}^n \widetilde{\sigma}^2(z_{j-1}) - 2\omega_j^n \widetilde{\sigma}^2(z_j)}{2(\Delta z)^2},$$

[14]Compared to Auclert et al. (2020), who break the solution procedure into two steps, first solving for the idiosyncratic variables given the aggregate variables, we solve for the path of all aggregate and idiosyncratic variables at once. Note that, besides the nonlinear perfect foresight method we refer to here (see their Section 6), they also propose a linear method.

which, grouping, results in

$$\frac{\omega_j^n - \omega_j^{n-1}}{\Delta t} = \underbrace{\left[\left(s_n(z_j) - \frac{\dot{A}_n}{A_n} - (1-\psi)\eta\right) - \frac{\mu(z_j)}{\Delta z} - \frac{\widetilde{\sigma}^2(z_j)}{(\Delta z)^2}\right]}_{\beta_j^n}\omega_n(z_j)$$

$$+ \underbrace{\left[\frac{\mu(z_{j-1})}{\Delta z} + \frac{\widetilde{\sigma}^2(z_{j-1})}{2\,(\Delta z)^2}\right]}_{\varrho_{j-1}^n}\omega_{j-1}^n + \underbrace{\left[\frac{\widetilde{\sigma}^2(z_{j+1})}{2\,(\Delta z)^2}\right]}_{\chi_{j+1}^n}\omega_{j+1}^n.$$

The boundary conditions are the ones associated with a reflected process z at the boundaries:[15]

$$\frac{\omega_1^n - \omega_1^{n-1}}{\Delta t} = (\beta_1^n + \chi_1^n)\,\omega_n(z_1) + \chi_2^n\omega_{j+1}^n,$$
$$\frac{\omega_J^n - \omega_J^{n-1}}{\Delta t} = (\beta_J^n + \varrho_J^n)\,\omega_n(z_J) + \varrho_{J-1}^n\omega_{j-1}^n.$$

If we define matrix

$$\mathbf{B}^n = \begin{bmatrix} \beta_1^n + \chi_1^n & \chi_2^n & 0 & 0 & \cdots & 0 & 0 & 0 \\ \varrho_1^n & \beta_2^n & \chi_3^n & 0 & \cdots & 0 & 0 & 0 \\ 0 & \varrho_2^n & \beta_3^n & \chi_4^n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \varrho_{J-2}^n & \beta_{J-1}^n & \chi_J^n \\ 0 & 0 & 0 & 0 & \cdots & 0 & \varrho_{J-1}^n & \beta_J^n + \varrho_J^n \end{bmatrix},$$

then we can express the KF equation as

$$\frac{\boldsymbol{\omega}^n - \boldsymbol{\omega}^{n-1}}{\Delta t} = \mathbf{B}^{n-1}\boldsymbol{\omega}^n,$$

or

$$\boldsymbol{\omega}^n = \left(\mathbf{I} - \Delta t\mathbf{B}^{n-1}\right)^{-1}\boldsymbol{\omega}^{n-1}, \tag{80}$$

where $\boldsymbol{\omega}^n = \begin{bmatrix} \omega_1^n & \omega_2^n & \ldots & \omega_{J-1}^n & \omega_J^n \end{bmatrix}^T$, and $\mathbf{I}$ is the identity matrix of dimension J.

[15]It is easy to check that this formulation preserves the fact that matrix $\mathbf{B}^n$ below is the transpose of the matrix associated with the infinitesimal generator of the process.

Extension to non-homogeneous grids Our model has been solved using a homogeneous grid and all the results presented in the paper have been computed using homogeneous grids. However, in some robustness tests that we have performed to assess the accuracy of the method, we have used non-homogeneous grid for the state $z$to economize on grid points. . We could not find a universally applicable way to implement non-homogeneous grids in the economics literature, so we propose the following discretization scheme.[16] We have used this scheme to verify that our numerical results are accurate in the sense that they do not change if we add additional grid points to the ω grid – no matter whether we add them where most of the mass of $\omega(z)$ is located or in the range in which z_t^* moves.

Be $z = \begin{bmatrix} z_1, & z_2, & ... & z_{J-1} & z_J \end{bmatrix}$ the grid. Define $\Delta z_{a,b} = z_b - z_a$ and let $\Delta z = \frac{1}{2}\begin{bmatrix} \Delta z_{1,2}, & \Delta z_{1,3}, & \Delta z_{2,4}, & ..., & \Delta z_{J-2,J} & \Delta z_{J-1,J} \end{bmatrix}$. We approximate the KFE (26) using central difference for both the first derivative and the second derivative.

$$\begin{aligned} \frac{\omega_j^n - \omega_j^{n-1}}{\Delta t} \quad &= \quad \left(s_n(z) - (1-\psi)\eta - \frac{\dot{A}_n}{A_n} \right) \omega_t\left(z_j\right) - \left[\frac{\mu(z_{j+1})\omega_t\left(z_{j+1}\right) - \mu(z_{j-1})\omega_t\left(z_{j-1}\right)}{\Delta z_{j-1,j+1}} \right] \\ &+ \frac{1}{2} \frac{\Delta z_{j-1,j}\sigma^2(z_{j+1})\omega_t\left(z_{j+1}\right) + \Delta z_{j,j+1}\sigma^2(z_{j-1})\omega_t\left(z_{j-1}\right) - \Delta z_{j-1,j+1}\sigma^2(z_j)\omega_t\left(z_j\right)}{\frac{1}{2}\left(\Delta z_{j-1,j+1}\right)\Delta z_{j,j+1}\Delta z_{j-1,j}} \end{aligned}$$

[16] Our approach builds on the one in the appendix to Achdou et al., 2017. It differs from theirs in two ways. First, it can be derived as a finite difference scheme to the KFE. Their approach delivers a finite difference approximation for the HJB, but not for the KFE, and hence it requires the grid to be constructed such that the step size to both sides of any grid point converge to one another. Furthermore, our approach is not an upwind scheme and has only been tested in the current model, which features no endogenous drift.

which, grouping, results in

$$
\begin{aligned}
\frac{\omega_j^n - \omega_j^{n-1}}{\Delta t} &= \underbrace{\left[\left(s_n(z) - (1-\psi)\eta - \frac{\dot{A}_n}{A_n}\right)\omega_t\left(z\right) + \frac{\sigma^2(z_j)\omega_t\left(z_j\right)}{\Delta z_{j,j+1}\Delta z_{j-1,j}}\right]}_{\beta_j^n}\omega_n(z_j) \\
&\quad + \underbrace{\left[\frac{\mu(z_{j-1})\omega_t\left(z_{j-1}\right)}{\Delta z_{j-1,j+1}} + \frac{\sigma^2(z_{j+1})\omega_t\left(z_{j+1}\right)}{\left(\Delta z_{j-1,j+1}\right)\Delta z_{j,j+1}}\right]}_{\varrho_{j-1}^n}\omega_{j-1}^n \\
&\quad + \underbrace{\left[-\frac{\mu(z_{j+1})\omega_t\left(z_{j+1}\right)}{\Delta z_{j-1,j+1}} + \frac{\sigma^2(z_{j+1})\omega_t\left(z_{j+1}\right)}{\left(\Delta z_{j-1,j+1}\right)\Delta z_{j,j+1}}\right]}_{\chi_{j+1}^n}\omega_{j+1}^n.
\end{aligned}
$$

The law of motion of ω can equivalently be written in matrix form

$$
\frac{\boldsymbol{\omega}^n - \boldsymbol{\omega}^{n-1}}{\Delta t} = \mathbf{B}^{n-1}\boldsymbol{\omega}^n
$$

where

$$
\mathbf{B}^n = \begin{bmatrix}
\beta_1^n + \chi_1^n & \chi_2^n & 0 & 0 & \cdots & 0 & 0 & 0 \\
\varrho_1^n & \beta_2^n & \chi_3^n & 0 & \cdots & 0 & 0 & 0 \\
0 & \varrho_2^n & \beta_3^n & \chi_4^n & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & \varrho_{J-2}^n & \beta_{J-1}^n & \chi_J^n \\
0 & 0 & 0 & 0 & \cdots & 0 & \varrho_{J-1}^n & \beta_J^n + \varrho_J^n
\end{bmatrix},
$$

Abstracting for brevity from the term $\left(s_n(z) - (1-\psi)\eta - \frac{\dot{A}_n}{A_n}\right)$, which is independent of the grid, and spelling out $\mathbf{B}^n$ we have

$$
\frac{\boldsymbol{\omega}^n - \boldsymbol{\omega}^{n-1}}{\Delta t} = \begin{bmatrix}
-\frac{\mu(z_1)}{\Delta z_{1,2}} - \frac{\sigma(z_1)}{\Delta z_{1,2}\Delta z_{1,2}/2} + \frac{\sigma(z_1)}{\Delta z_{1,2}\Delta z_{1,2}} & -\frac{\mu(z_2)}{\Delta z_{1,2}} + \frac{\sigma(z_2)}{\Delta z_{1,2}\Delta z_{1,2}} & 0 & \cdots \\
\frac{\mu(z_1)}{\Delta z_{1,3}} + \frac{\sigma(z_1)}{\Delta z_{1,3}\Delta z_{1,2}} & -\frac{\sigma(z_2)}{\Delta z_{1,2}\Delta z_{2,3}} & -\frac{\mu(z_3)}{\Delta z_{1,3}} + \frac{\sigma(z_3)}{\Delta z_{1,3}\Delta z_{2,3}} & \cdots \\
0 & \frac{\mu(z_2)}{\Delta z_{2,4}} + \frac{\sigma(z_2)}{\Delta z_{2,4}\Delta z_{2,3}} & -\frac{\sigma(z_3)}{\Delta z_{2,3}\Delta z_{3,4}} & \cdots \\
0 & 0 & \frac{\mu(z_3)}{\Delta z_{3,5}} + \frac{\sigma(z_3)}{\Delta z_{3,4}\Delta z_{3,5}} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}\boldsymbol{\omega}^n.
$$

We can rewrite this as follows

$$\frac{\boldsymbol{\omega}^n - \boldsymbol{\omega}^{n-1}}{\Delta t} = \begin{bmatrix} -\frac{\mu(z_1)}{\Delta z_{1,2}} - \frac{\sigma(z_2)}{\boldsymbol{\Delta z_{1,2}}\Delta z_{1,2}} & -\frac{\mu(z_2)}{\Delta z_{1,2}} + \frac{\Delta z_{2,3}\sigma(z_2)}{\Delta z_{2,3}(\boldsymbol{\Delta z_{1,2}}\Delta z_{1,2})} & 0 & \cdots \\ \frac{\mu(z_1)}{\Delta z_{1,3}} + \frac{\sigma(z_1)}{\boldsymbol{\Delta z_{1,3}}\Delta z_{1,2}} & -\frac{(\Delta z_{1,2}+\Delta z_{2,3})\sigma(z_2)}{\boldsymbol{\Delta z_{1,3}}(\Delta z_{1,2}\Delta z_{2,3})} & -\frac{\mu(z_3)}{\Delta z_{1,3}} + \frac{\Delta z_{3,4}\sigma(z_3)}{\Delta z_{3,4}(\boldsymbol{\Delta z_{1,3}}\Delta z_{2,3})} & \cdots \\ 0 & \frac{\mu(z_2)}{\Delta z_{2,4}} + \frac{\Delta z_{1,2}\sigma(z_2)}{\Delta z_{1,2}(\boldsymbol{\Delta z_{2,4}}\Delta z_{2,3})} & -\frac{(\Delta z_{2,3}+\Delta z_{3,4})\sigma(z_3)}{\boldsymbol{\Delta z_{2,4}}(\Delta z_{2,3}\Delta z_{3,4})} & \cdots \\ 0 & 0 & \frac{\mu(z_3)}{\Delta z_{3,5}} + \frac{\Delta z_{2,3}\sigma(z_3)}{\Delta z_{2,3}(\Delta z_{3,4}\boldsymbol{\Delta z_{3,5}})} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \boldsymbol{\omega}^n.$$

Note that the bold terms in line i are equal to $1/\Delta z_i$. Thus the columns of $\mathbf{B}^n \Delta z$ sum up to 1 and the operation is mass preserving, in the sense that the above relationship guarantees that

$$\sum \omega_j^n \Delta z_j = \sum \omega_j^{n-1} \Delta z_j$$

where $\sum \omega_j^n \Delta z_j$ is a trapezoid approximation of the integral $\int \omega^n(z) dz$.

C.2 Finite difference approximation of the integrals

To approximate the integrals in $\int_0^z \omega_t(z)\, dz$ and $\int_{z_t^*}^{\infty} z\omega_t(z) dz$ we use the trapezoid rule. I.e. if $f(z)$ is either $\omega_t(z)$ or $z\omega_t(z)$ and $z_j \leq \bar{z} \leq z_{j+1}$ then the integral from the closest lower gridpoint is given by

$$\int_{z_j}^{\bar{z}} f(z)\, dz = \left[f(z_j) + \frac{1}{2} \left[f(z_{j+1}) - f(z_j) \right] \frac{(\bar{z} - z_j)}{\Delta z} \right] (\bar{z} - z_j)$$

We use this formula to construct the integrals over a larger range piecewise. For example:

$$\int_{z_1}^{z_N} f(z)\, dz = \begin{bmatrix} \frac{1}{2} & 1 & 1 & \cdots & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} f(z_1) \\ f(z_2) \\ \vdots \\ f(z_N) \end{bmatrix}$$

and

$$\int_{z_1}^{z^*} f(z)\,dz = \left[\begin{array}{cccccc} \frac{1}{2} & 1 & 1 & \cdots & 1 & \frac{1}{2} \end{array}\right] \left[\begin{array}{c} f(z_1) \\ f(z_2) \\ \vdots \\ f(z_{j^*}) \end{array}\right]$$
$$+ \left[f(z_{j^*-1}) + \frac{1}{2}\left[f(z_{j^*}) - f(z_{j^*-1})\right]\frac{(z^* - z_{j^*-1})}{\Delta z}\right](z^* - z_{j^*-1})$$
$$\text{where } j^* = \arg\min_j \{j \leq J | z_{j^*} > z^*\}$$

C.3 Algorithm to solve for the SS

Here we present how to solve for the SS of the private equilibrium, that is for the SS when the central bank sets a certain level of the nominal interest rate in SS i^{ss}.

We know that in SS consumption does not grow, hence from (14)

$$r^{ss} = \rho^h. \tag{81}$$

We also know that in SS, the investment rate is equal to the depreciation,

$$\iota^{ss} = \delta. \tag{82}$$

This means that, from equation (17) and the functional form we assumed for the capital adjustment costs (37),

$$(q_t - 1 - \Phi'(\iota_t))(r_t - (\iota_t - \delta)) = \dot{q}_t - \Phi''(\iota_t)\,\dot{\iota}_t - (q_t \iota_t - \iota_t - \Phi(\iota_t)) \tag{83}$$

$$\left(q^{ss} - 1 - \phi^k(\iota^{ss} - \delta)\right)\left(\rho^{hh} - (\iota^{ss} - \delta)\right) = 0 - \phi^k * 0 - \left(q^{ss}\iota^{ss} - \iota^{ss} - \phi^k(\iota^{ss} - \delta)\right)$$

$$\rho^{hh}(q^{ss} - 1) = \delta(1 - q^{ss})$$

.From here we can solve for the steady state value of q^{ss}, which is given by

$$q^{ss} = 1. \tag{84}$$

Furthermore, combining (81) with the fisher equation and the fact that the planner sets a certain nominal rate i^{ss} we get that

$$\pi^{ss} = i^{ss} - \rho^h. \tag{85}$$

In SS, $\dot{\pi}_t = 0$ and $\dot{Y}_t = 0$. Hence, from equation (20) we obtain

$$m^{ss} = \left(m^* + \rho^h \pi^{ss} \frac{\theta}{\varepsilon} \right). \tag{86}$$

Using equation (34) and (81),

$$\rho^h = \frac{1}{q^{ss}} \left(\alpha m_t Z_t A_t^{\alpha-1} L^{1-\alpha} \frac{z_t^*}{\gamma X_t} \right) - \delta \tag{87}$$

From equation (35) and (81),

$$\frac{\dot{A}_t}{A_t} = 0 = \frac{1}{q_t} (\alpha m_t Z_t {A_t}^{\alpha-1} {L_t}^{1-\alpha} - R_t(1 - \Omega(z_t^*)) + R_t - \delta q_t - q_t(1-\psi)\eta). \tag{88}$$

Plugging the latter equation into the former, using $q^{SS} = 1$ and using the definition of r_t we obtain:

$$\rho^h + \delta = \left[(\rho^h + \delta) \left(\gamma(1 - \Omega(z_t^*)) - 1 \right) + (1-\psi)\eta + \delta \right] \frac{z^*}{\gamma X^*}. \tag{89}$$

In the algorithm, we use a non-linear equation solver to obtain z^* from this equation. *The Algorithm.*

- Get $r^{ss} = \rho^h$, $\pi^{ss} = \bar{\pi}$ and $i^{ss} = \rho^h + \pi^{ss}$ and $R^{ss} = q^{ss}(\rho^h + \delta)$ and $m^{ss} = m^* + \rho^h \pi^{ss} \frac{\theta}{\epsilon}$.
- Given that our calibration target for $L^{ss} = 1$, we "guess" $L^{ss} = 1$
- Let n now denote the iteration counter. Make an initial guess for the net worth distribution $\boldsymbol{\omega}^0$

  1. Use a non-linear equation solver on equation (89) to obtain z^* from equation (89).
  2. Obtain $Z_n = (\gamma_n X_n^*)^\alpha$.

3. Find A from equation (33),

$$A^n = \left[\frac{q^{ss}\rho^h + \delta q^{ss}}{\alpha m_n Z_n {L_m}^{1-\alpha} \frac{z_t^*}{\gamma X_t}}\right]^{\frac{1}{\alpha-1}}.$$

4. Find the stocks $K_n = \gamma(1-\Omega^n(z^*))A^n$, $D_n = K_n - A_n$.

5. Compute $w_n = (1-\alpha)m^{ss} Z_n {A_n}^{\alpha} {L_n}^{-\alpha}$,$\varphi_n = \alpha\left(\frac{(1-\alpha)}{w_n}\right)^{(1-\alpha)/\alpha} {m^{ss}}^{\frac{1}{\alpha}}$.

6. Get aggregate output $Y = Z_n A_n^{\alpha} {L_n}^{1-\alpha}$, transfers $T_n = (1-m^{ss})\,Y_n - \frac{\theta}{2}\left(\pi^{ss}\right)^2 Y_n + (1-\psi)\eta A_t$, and consumption $C_n = w_n L_m + r^{ss} D_n + T_n$.

7. Update $\hat{s}_j^n = \frac{1}{q^{ss}}(\gamma \max\{z\varphi_n - R_n, 0\} + R_n - \delta q^{ss})$ and employ it to construct matrix $\mathbf{B}^{n-1}$.

8. Update $\boldsymbol{\omega}^{n+1}$ using equation $\frac{\boldsymbol{\omega}^{n+1}-\boldsymbol{\omega}^n}{\Delta t} = \mathbf{B}^n \boldsymbol{\omega}^{n+1}$.

9. If the net worth distribution do not coincide with the guess, set $n = n+1$ and return to point 1

- Set $\Upsilon = \left(w_{L=1} C_{L=1}^{-\eta}\right)$ to ensure our "guess" for L^{ss} is correct.

D Computing optimal policies in heterogeneous-agent models

D.1 General algorithm

Solving for the optimal policy in models with heterogeneous agents poses a certain challenge since the state in such a model contains a distribution, which is an infinite-dimensional object. In this section, we explain how such models can be solved in a relatively straightforward manner. Our approach relies on three main conceptual ingredients: (i) finite difference approximation of continuous time and continuous idiosyncratic states, (ii) symbolic derivation of the planner's first-order conditions, and (iii) use of a Newton algorithm to solve the optimal policy problem non-linearly in the sequence space. Here we present a general overview which goes beyond the particular model presented in the paper.

(i) Finite difference approximation A continuous-time, continuous-space heterogeneous-agent model discretized using an upwind finite-difference method becomes a discrete-time, discrete-space model. In this discretized model the dynamics of the (now finite-dimensional) distribution $\boldsymbol{\mu}_t$ at period t are given by

$$\boldsymbol{\mu}_t = \left(\mathbf{I} - \Delta t \boldsymbol{A}_t^T\right)^{-1} \boldsymbol{\mu}_{t-1}, \tag{90}$$

where Δt is the time step between periods and $\boldsymbol{A}_{\mathbf{t}}$ is a matrix whose entries depend *nonlinearly* and in *closed form* on the idiosyncratic and aggregate variables in period t.[17] Similarly, the HJB equation is approximated as[18]

$$\rho \mathbf{v}_{t+1} = \boldsymbol{u}_{t+1} + \mathbf{A}_{t+1}\mathbf{v}_{t+1,} - \left(\mathbf{v}_{t+1} - \mathbf{v}_t\right)/\Delta t. \tag{91}$$

Together with additional static equations, such as market clearing conditions or budget constraints, and aggregate dynamic equations, including the Euler equations of representative agents (if any) and the dynamics of aggregate states, they define the discretized model.

Though we have ended up with a discrete-time approximation, casting the original model in continuous time is central to our method. The discretized dynamics of the distribution (90) and Bellman equation (91) present two advantages compared to their counterparts in the discrete-time continuous-state formulation typically employed in the literature. First, the analytical tractability of the original continuous-time model implies that the agents' optimal choices in the discretized version are always "on the grid", avoiding the need for interpolation, and are "one step at a time" making the matrix $\mathit{\Pi}_t$ sparse.[19] Second, the private agent's FOCs hold with equality even at the exogenous boundaries (see Achdou et al. (2017) for a detailed discussion of these advantages).

(ii) Symbolic derivation of planner's FOCs Once we have a finite-dimensional discrete-time discrete-space model, we can derive the planner's FOCs by *symbolic differentiation* using standard software packages. For convenience, we rely on Dynare's

[17] Technically, this matrix results from the discretization of the *infinitesimal generator* of the idiosyncratic states. In the example of Section 2, $\boldsymbol{\mu}_t = \boldsymbol{\omega}_{\mathbf{t}}$ and $\boldsymbol{A}_t = \mathbf{B}_{\mathbf{t}}$.

[18] In the model presented in this paper the HJB can be solved analytically and hence there is no need to solve it computationally.

[19] The introduction of Poisson shocks would not change the sparsity of matrix $\mathit{\Pi}_t$.

toolbox for Ramsey optimal policy to do this task for us. To this end, we simply provide the discretized version of our model's private equilibrium conditions to Dynare (the discretized counterpart to the equations in Appendix A.7), making use of loops for the heterogeneous-agent block, as in Winberry (2018). We furthermore provide the discretized objective function, and Dynare then takes symbolic derivatives to construct the set of optimality conditions of the planner for us.

A natural question at this stage is under which conditions the optimal policies of the discrete-time, discrete-space problem coincide with those of the original problem. The following proposition shows that, if the time interval is small enough (the standard condition when approximating continuous-time models), then the two solutions coincide.

Proposition 1 : *Provided that all the Lagrange multipliers associated to the equilibrium conditions are continuous for $t > 0$, the solution of the "discretize-optimize" and the "optimize-discretize" algorithms converge to each other as the time step Δt goes towards 0.*

Proof: See Appendix D.2.

The proposition guarantees that both strategies coincide when Δt goes towards zero and provides an error bound that depends on the value of the maximum change in the Lagrange multipliers. This proposition is quite general, as most continuous-time, perfect-foresight, general equilibrium models do not feature discontinuities for $t > 0$.

The model presented in Section 2 is arguably simpler than the general heterogenous-agent model covered by Proposition 1, as it features an analytic solution for the HJB equation. To get an idea of the performance of our method in a case in which the HJB is also a constraint in the planner's problem, as well as to showcase its generality in dealing with different problems, we compute the optimal monetary policy in the HANK model of Nuño and Thomas (2016) using our method in Dynare (see Appendix D.3). We compare our results with those using their "optimize-discretize" algorithm at monthly frequency $\triangle t = 1/12$. We conclude that both approaches essentially coincide.

(iii) Newton algorithm to solve the optimal policy problem non-linearly in the sequence space Finally, we use the discretized optimality conditions of the planner to compute non-linearly the *optimal responses* a temporary change in parameters (an "MIT shock") using a Newton algorithm. Instead of time iterations over guesses for aggregate sequences, as is common in the literature, we use a global relaxation algorithm. This approach has been made popular in discrete-time models by Juillard et al. (1998) thanks to Dynare, but it is somewhat less common in continuous-time models (e.g. Trimborn et al., 2008). This approach helps to overcome the curse of dimensionality since in the sequence space the complexity of the problem grows only linearly in the number of aggregate variables, whereas the complexity of the state-space solution grows exponentially in the number of state variables. Recently Auclert et al. (2019) have exploited a particularly efficient variant of this approach in the context of heterogeneous-agent models.[20] We build on these contributions when we compute the optimal transition path. Again we make use of Dynare. We use its nonlinear Newton solver to compute both the steady state of the Ramsey problem and the optimal transition path under perfect foresight.[21] Our hope is that the convenience of using Dynare will make optimal policy problems in heterogeneous-agent models easily accessible to a large audience of researchers.

The solution to the perfect foresight problem can be easily adapted to the case with aggregate shocks. As Boppart et al. (2018) show, the perfect-foresight transitional dynamics to an "MIT shock" coincides with the solution of the model with aggregate uncertainty using a first-order perturbation approach. We follow this approach to analyze the optimal response to a cost-push shock below.

Finally, it is important to highlight that our solution approach is different from the one in Winberry (2018) or Ahn et al. (2018). These papers expand the seminal contribution by Reiter (2009), based on a two-stage algorithm that (i) first finds the nonlinear solution of the steady state of the model and (ii) then applies perturbation techniques to produce a linear system of equations describing the dynamics around the steady state. Winberry (2018) illustrates how this can be also implemented using Dynare and Ahn et al. (2018) extend the methodology to continuous-time problems. However, these methods were not created to deal with the problem of finding the optimal policies, the focus of our algorithm, as the first stage requires the computation of the steady state, which in our case is the steady state of the problem under optimal policies. Our algorithm finds the steady state of the planner's problem, including the

[20] Compared to Auclert et al. (2020), who break the solution procedure into two steps, first solving for the idiosyncratic variables given the aggregate variables, we solve for the path of all aggregate and idiosyncratic variables at once. Note that, besides the nonlinear perfect foresight method we refer to here (see their Section 6), they also propose a linear method.

[21] To find the steady state, we provide Dynare with the steady state of the private equilibrium conditions as a function of the policy instrument.

Lagrange multipliers. Naturally, this steady does not need to coincide with the steady state that can be found by looking for the value of the planner's policy that maximizes steady-state welfare.

D.2 Proof of proposition D.1

Proof: The proof has the following structure. First, we set up a generic planner's problem in a continuous-time heterogeneous-agent economy without aggregate uncertainty. Second, we derive the continuous time optimality conditions of the planner's problem and discretize them. Third, we discretize the planners problem and the derive the optimality conditions. Fourth, we compare the two sets of discretized optimality conditions.

1. The generic problem The planner's problem in an economy with heterogeneity among one agent type (e.g. households or firms) can be written as

$$
\begin{aligned}
\max_{Z_t,u_t(x),\mu_t(x),v_t(x)} \quad & \int_0^\infty \exp(-\varrho t) f_0(Z_t) dt && (92)\\
s.t.\ \forall t & \\
\dot{X}_t &= f_1(Z_t) && (93)\\
\dot{U}_t &= f_2(Z_t) && (94)\\
0 &= f_3(Z_t) && (95)\\
\tilde{U}_t &= \int f_4\left(x, u_t(x), Z_t\right) \mu_t\left(x\right)\ dx && (96)\\
\rho v_t(x) &= \dot{v}_t(x) + f_5(x, u_t(x), Z_t) && (97)\\
& + \sum_{i=1}^{I} b_i\left(x, u_t(x), Z_t\right) \frac{\partial v_t(x)}{\partial x_i} + \sum_{i=1}^{I}\sum_{k=1}^{I} \frac{\left(\sigma(x)\sigma(x)^{\top}\right)_{i,k}}{2} \frac{\partial^2 v_t(x)}{\partial x_i \partial x_k},\ \forall x \\
0 &= \frac{\partial f_5}{\partial u_{j,t}} + \sum_{i=1}^{I} \frac{\partial b_i}{\partial u_{j,t}} \frac{\partial v_t(x)}{\partial x_i}, \quad j = 1, ..., J, \forall x. && (98)\\
\dot{\mu}_t\left(x\right) &= -\sum_{i=1}^{I} \frac{\partial}{\partial x_i}\left[b_i\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\right] && (99)\\
& + \frac{1}{2}\sum_{i=1}^{I}\sum_{k=1}^{I} \frac{\partial^2}{\partial x_i \partial x_k}\left[\left(\sigma(x)\sigma(x)^{\top}\right)_{i,k}\mu_t\left(x\right)\right],\ \forall x \\
X_0 &= \bar{X}_0 && (100)\\
\mu_0\left(x\right) &= \bar{\mu}_0\left(x\right) && (101)\\
lim_{t\to\infty} U &= \bar{U}_\infty && (102)\\
lim_{t\to\infty} v(x) &= \bar{v}(x)_\infty && (103)
\end{aligned}
$$

where we have adopted the following notation:

- Variables (capitals are reserved for aggregate variables):
  - x individual state vector with I elements
  - u individual control vector with J elements
  - v individual value function vector with 1 element
  - $u(x)$ control vector as function of individual state
  - $\mu(x)$ distribution of agents across states

- $v(x)$ value function as function of individual state
- X aggregate state vector (other than μ)
- $\hat{U}$ aggregate control vector of purely contemporaneous variables
- U aggregate control vector of intertemporal variables
- $\tilde{U}$ control vector of aggregator variables
- $Z_t = \left\{\tilde{U}_t, U_t, \bar{U}_t, X_t\right\}$ vector of all aggregate variables

- Functions
  - b function that determines the drift of x
  - f_0 welfare function
  - f_1, f_2, f_3 aggregate equilibrium conditions
  - f_4 aggregator function
  - f_5 individual utility function

Line (92) is the planner's objective function.[22] Equations (93)-(95) are the aggregate equilibrium conditions for aggregate states, jump variables and contemporaneous variables. In our model, examples for each of these three types of equations are the law of motion of aggregate capital, the household's Euler equation and the household's labor supply condition, respectively. Equation (96) links aggregate and individual variables, such as the definition of aggregate TFP in our model. Equations (97) and (98) are the individual agent's value function and first order conditions, which must hold across the whole individual state vector x. In our model we do not have these two types of equations since we can analytically solve the individual optimal choice. The Kolmogorov Forward equation (24) determines the evolution of the distribution of agents. Finally (100)-(103) are the initial and terminal conditions for the aggregate and individual state and dynamic control variables. In our model these are the initial capital stock and firm distribution and the terminal conditions for variables such as consumption.

2. Optimize, then discretize First we consider the approach introduced in Nuño and Thomas (2016), namely to compute the first order conditions using calculus of variations and then to discretize the problem using an upwind finite difference scheme.

[22]Notice that the planner's discount factor, ϱ, can be different to that of individual agents, ρ.

2.a The Lagrangian The Lagrangian for this problem is given by:[23]

$$
\begin{aligned}
\mathcal{L} &= \int_0^\infty \left\{ e^{-\varrho t} f_0(Z_t) \right.\\
&+ \lambda_{1,t}\left(\dot{X}_t - f_1(Z_t)\right)\\
&+ \lambda_{2,t}\left(\dot{U}_t - f_2(Z_t)\right)\\
&+ \lambda_{3,t}\left(f_3(Z_t)\right)\\
&+ \lambda_{4,t}\left(\tilde{U}_t - \int f_4\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\ dx\right)\\
&+ \int \left[\lambda_{5,t}(x)\left(-\rho v_t(x) + \dot{v}_t(x) + f_5(x, u_t(x), Z_t) + \sum_{i=1}^{I} b_i\left(x, u_t(x), Z_t\right)\frac{\partial v_t(x)}{\partial x_i} + \sum_{i=1}^{I}\frac{\sigma_i^2(x)}{2}\frac{\partial^2 v_t(x)}{\partial^2 x_i}\right)\right] dx\\
&+ \sum_{j=1}^{J}\int\left[\lambda_{6,j,t}(x)\left(\frac{\partial f_5}{\partial u_{j,t}} + \sum_{i=1}^{I}\frac{\partial b_i}{\partial u_{j,t}}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx\\
&+ \left.\int\left[\lambda_{7,t}(x)\left(-\dot{\mu}_t\left(x\right) + \left(-\sum_{i=1}^{I}\frac{\partial}{\partial x_i}\left[b_i\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\right] + \frac{1}{2}\sum_{i=1}^{I}\frac{\partial^2}{\partial^2 x_i}\left[\sigma_i^2(x)\mu_t\left(x\right)\right]\right)\right)\right] dx\right\} dt
\end{aligned}
$$

where λ_1 to λ_7 denote the multipliers on the respective constraints. For convenience,

[23] For simplicity, we assume that the Wiener processes driving the dynamics of the state x are independent, though the proof can be trivially extended to that case, at the cost of a more cumbersome notation.

we write the time derivatives in a separate line at the end. The Lagrangian becomes:

$$
\begin{aligned}
\mathcal{L} \;=\; & \int_0^\infty \left\{ e^{-\varrho t}\, f_0(Z_t) \right. \\
+ \; & \lambda_{1,t}\left(-f_1(Z_t)\right) \\
+ \; & \lambda_{2,t}\left(-f_2(Z_t)\right) \\
+ \; & \lambda_{3,t}\left(-f_3(Z_t)\right) \\
+ \; & \lambda_{4,t}\left(\tilde{U}_t - \int f_4\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\; dx\right) \\
+ \; & \int \left[\lambda_{5,t}(x)\left(-\rho v_t(x) + f_5(x, u_t(x), Z_t) + \sum_{i=1}^{I} b_i\left(x, u_t(x), Z_t\right)\frac{\partial v_t(x)}{\partial x_i} + \sum_{i=1}^{I}\frac{\sigma_i^2(x)}{2}\frac{\partial^2 v_t(x)}{\partial^2 x_i}\right)\right] dx \\
+ \; & \sum_{j=1}^{J}\int\left[\lambda_{6,j,t}(x)\left(\frac{\partial f_{5,t}}{\partial u_{j,t}} + \sum_{i=1}^{I}\frac{\partial b_i}{\partial u_{j,t}}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx \\
+ \; & \left. \int\left[\lambda_{7,t}(x)\left(-\sum_{i=1}^{I}\frac{\partial}{\partial x_i}\left[b_i\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\right] + \frac{1}{2}\sum_{i=1}^{I}\frac{\partial^2}{\partial^2 x_i}\left[\sigma_i^2(x)\mu_t\left(x\right)\right]\right)\right] dx \right\} dt \\
+ \; & \int_0^\infty \left\{ e^{-\varrho t}\,\lambda_{1,t}\dot{X}_t + \lambda_{2,t}\dot{U}_t + \int\left[\lambda_{5,t}\dot{v}_t(x)\right] dx - \int\left[\lambda_{7,t}\dot{\mu}_t\left(x\right)\right] dx \right\} dt.
\end{aligned}
$$

We have ignored the terminal and initial conditions but we will account for them later on. Now we manipulate the Lagrangian using integration by parts in order to bring it into a more convenient form. We start with the last line. Switching the order of integration, the last line becomes

$$
\begin{aligned}
\int_0^\infty e^{-\varrho t}\,\lambda_{1,t}\dot{X}_t\; dt + \int_0^\infty e^{-\varrho t}\,\lambda_{2,t}\dot{U}_t\; dt + \int\int_0^\infty \left[e^{-\varrho t}\lambda_{5,t}(x)\dot{v}_t(x)\right] dt\; dx & \\
- \int\int_0^\infty \left[e^{-\varrho t}\lambda_{7,t}(x)\dot{\mu}_t\left(x\right)\right] dt\; dx &
\end{aligned}
$$

Now we integrate this expression by parts with respect to time t, using

$$
\begin{aligned}
\int_0^\infty e^{-\varrho t}\, a_t\dot{b}_t\; dt &= \left[e^{-\varrho t}a_t b_t\right]_0^\infty - \int_0^\infty e^{-\varrho t}(\dot{a}_{1,t} - \varrho a_{1,t})b_t\; dt \\
&= \lim_{t\to\infty} e^{-\varrho t}a_t b_t - a_0 b_0 - \int_0^\infty e^{-\varrho t}(\dot{a}_t - \varrho a_t)b_t\; dt
\end{aligned}
$$

to get

$$
\begin{aligned}
& \lim_{t\to\infty} e^{-\varrho t}\lambda_{1,t}X_t - \lambda_{1,0}X_0 - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{1,t} - \varrho\lambda_{1,t})X_t dt + \lim_{t\to\infty} e^{-\varrho t}\lambda_{2,t}U_t - \lambda_{2,0}U_0 \\
& - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{2,t} - \varrho\lambda_{2,t})U_t dtx \\
+ & \int \left(\lim_{t\to\infty} e^{-\varrho t}\lambda_{5,t}(x)v_t(x) - \lambda_{5,0}(x)v_0(x)\right) dx - \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{5,t}(x) - \varrho\lambda_{5,t}(x))v_t(x)dtdx \\
- & \int \lim_{t\to\infty} e^{-\varrho t}\lambda_{7,t}(x)\mu_t(x) - \lambda_{7,0}(x)\mu_0(x)dx + \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{7,t}(x) - \varrho\lambda_{7,t}(x))\mu_t\,(x)\,dtdx
\end{aligned}
$$

Now we use the initial and terminal conditions to drop some $\lim_{t\to\infty}$ and $t = 0$ terms,

$$
\begin{aligned}
+ & \lim_{t\to\infty} e^{-\varrho t}\lambda_{1,t}X_t - \lambda_{2,0}U_0 - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{1,t} - \varrho\lambda_{1,t})X_t dt - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{2,t} - \varrho\lambda_{2,t})U_t dt \\
- & \int \lambda_{5,0}(x)v_0(x)dx + \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{5,t}(x) - \varrho\lambda_{5,t}(x))v_t(x)dtdx \\
- & \int \lim_{t\to\infty} e^{-\varrho t}\lambda_{7,t}(x)\mu_t(x)dx + \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{7,t}(x) - \varrho\lambda_{7,t}(x))\mu_t\,(x)\,dtdx
\end{aligned}
$$

Next we integrate lines 6 to 8 by parts with respect to x. This yields:

$$
\begin{aligned}
+ & \int \Bigg\{ \left[\left(-\rho\lambda_{5,t}(x)v_t(x) + f_5(x, u_t(x), Z_t) - \sum_{i=1}^{I} \frac{\partial b_i\,(x, u_t(x), Z_t)\,\lambda_{5,t}(x)}{\partial x_i} v_t(x)\right)\right] dx \\
& + \int \left[\left(+\frac{1}{2}\sum_{i=1}^{I} \frac{\partial^2}{\partial^2 x_i}\left[\sigma_i^2(x)\lambda_{5,t}(x)\right] v_t(x)\right)\right] dx \\
+ & \sum_{j=1}^{J} \int \left[\lambda_{6,j,t}(x)\frac{\partial f_{5,t}}{\partial u_{j,t}} - \sum_{i=1}^{I} \frac{\partial \left[\lambda_{6,j,t}(x)\frac{\partial b_i}{\partial u_{j,t}}\right]}{\partial x_i} v_t(x)\right] dx \\
+ & \int \left[\left(\sum_{i=1}^{I} \frac{\partial \lambda_{7,t}(x)}{\partial x_i}\left[b_i\,(x, u_t(x), Z_t)\,\mu_t\,(x)\right] + \sum_{i=1}^{I} \frac{\partial^2 \lambda_{7,t}(x)}{\partial^2 x_i}\frac{\sigma_i^2(x)}{2}\mu_t\,(x)\right)\right] dx \Bigg\}\, dt
\end{aligned}
$$

Putting this all together the Lagrangian has become:

$$
\begin{aligned}
\mathcal{L} \;=\; & \int_0^\infty \left\{ e^{-\varrho t}\, f_0(Z_t) \right. \\
+ \; & \lambda_{1,t}\left(-f_1(Z_t)\right) \\
+ \; & \lambda_{2,t}\left(-f_2(Z_t)\right) \\
+ \; & \lambda_{3,t}\left(-f_3(Z_t)\right) \\
+ \; & \lambda_{4,t}\left(\tilde{U}_t - \int f_4\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\ dx\right) \\
+ \; & \int \left(-\rho\lambda_{5,t}(x)v_t(x) + \lambda_{5,t}(x)f_5(x, u_t(x), Z_t) - \sum_{i=1}^{I}\frac{\partial\left[b_i\left(x, u_t(x), Z_t\right)\lambda_{5,t}(x)\right]}{\partial x_i}v_t(x)\right)dx \\
+ \; & \int \left(\frac{1}{2}\sum_{i=1}^{I}\frac{\partial^2}{\partial^2 x_i}\left[\sigma_i^2(x)\lambda_{5,t}(x)\right]v_t(x)\right)dx \\
+ \; & \sum_{j=1}^{J}\int\left[\lambda_{6,j,t}(x)\frac{\partial f_{5,t}}{\partial u_{j,t}} - \sum_{i=1}^{I}\frac{\partial\left[\lambda_{6,j,t}(x)\frac{\partial b_i}{\partial u_{j,t}}\right]}{\partial x_i}v_t(x)\right]dx \\
+ \; & \int_0^\infty \left[\left(\sum_{i=1}^{I}\frac{\partial\lambda_{7,t}(x)}{\partial x_i}\left[b_i\left(x, u_t(x), Z_t\right)\mu_t\left(x\right)\right] + \sum_{i=1}^{I}\frac{\partial^2\lambda_{7,t}(x)}{\partial^2 x_i}\frac{\sigma_i^2(x)}{2}\mu_t\left(x\right)\right)\right]dx \left.\vphantom{\int}\right\} dt \\
+ \; & \lim_{t\to\infty} e^{-\varrho t}\lambda_{1,t}X_t - \lambda_{2,0}U_0 - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{1,t} - \varrho\lambda_{1,t})X_t dt - \int_0^\infty e^{-\varrho t}(\dot{\lambda}_{2,t} - \varrho\lambda_{2,t})U_t dt \\
+ \; & \int -\lambda_{5,0}(x)v_0(x)dx + \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{5,t}(x) - \varrho\lambda_{5,t}(x))v_t(x)dtdx \\
- \; & \int \lim_{t\to\infty} e^{-\varrho t}\lambda_{7,t}(x)\mu_t(x)dx + \int\int_0^\infty e^{-\varrho t}(\dot{\lambda}_{7,t}(x) - \varrho\lambda_{7,t}(x))\mu_t\left(x\right)dtdx.
\end{aligned}
$$

2.b Optimality conditions in the continuous state space We take the Gateaux derivatives in direction $h_t(x)$ for each endogenous variable x. These derivatives have to be equal to zero for any $h_t(x)$ in the optimum. This implies the following optimality conditions:

Aggregate variables:

$$
\begin{aligned}
U_t: \quad 0 = \quad & -(\dot{\lambda}_{2,t} - \varrho\lambda_{2,t}) & (104)\\
& + \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t}\frac{\partial f_{1,t}}{\partial U_t} - \lambda_{2,t}\frac{\partial f_{2,t}}{\partial U_t} - \lambda_{3,t}\frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t}\int \frac{\partial f_{4,t}}{\partial U_t}\mu_t(x)\ dx & (105)\\
& + \int \left[\lambda_{5,t}(x)\left(\frac{\partial f_{5,t}}{\partial U_t} + \sum_{i=1}^{I}\frac{\partial b_{i,t}}{\partial U_t}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx & (106)\\
& + \sum_{j=1}^{J}\int \left[\lambda_{6,j,t}(x)\left(\frac{\partial^2 f_{5,t}}{\partial u_{j,t}\partial U_t} + \sum_{i=1}^{I}\frac{\partial b_{i,t}}{\partial u_{j,t}\partial U_t}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx & (107)\\
& + \int \left[\lambda_{7,t}(x)\left(-\sum_{i=1}^{I}\frac{\partial}{\partial x_i}\left[\frac{\partial b_{i,t}}{\partial U_t}\mu_t(x)\right]\right)\right] dx, & (108)\\
& \forall t > 0, & (109)
\end{aligned}
$$

$$
0 = \quad \lambda_{2,0}. \qquad (110)
$$

$$
\begin{aligned}
X_t: \quad 0 = \quad & -(\dot{\lambda}_{1,t} - \varrho\lambda_{1,t})\\
& + \frac{\partial f_{0,t}}{\partial X_t} - \lambda_{1,t}\frac{\partial f_{1,t}}{\partial X_t} - \lambda_{2,t}\frac{\partial f_{2,t}}{\partial X_t} - \lambda_{3,t}\frac{\partial f_{3,t}}{\partial X_t} - \lambda_{4,t}\int \frac{\partial f_{4,t}}{\partial X_t}\mu_t(x)\ dx\\
& + \int \left[\lambda_{5,t}(x)\left(\frac{\partial f_{5,t}}{\partial X_t} + \sum_{i=1}^{I}\frac{\partial b_{i,t}}{\partial X_t}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx\\
& + \sum_{j=1}^{J}\int \left[\lambda_{6,j,t}(x)\left(\frac{\partial^2 f_{5,t}}{\partial u_{j,t}\partial X_t} + \sum_{i=1}^{I}\frac{\partial b_{i,t}}{\partial u_{j,t}\partial X_t}\frac{\partial v_t(x)}{\partial x_i}\right)\right] dx\\
& + \int \left[\lambda_{7,t}(x)\left(-\sum_{i=1}^{I}\frac{\partial}{\partial x_i}\left[\frac{\partial b_{i,t}}{\partial X_t}\mu_t(x)\right]\right)\right] dx,\\
& \forall t \geq 0,
\end{aligned}
$$

$$
0 = \quad \lim_{t\to\infty} e^{-\varrho t}\lambda_{1,t}(x).
$$

$$
\begin{aligned}
\hat{U}_t: \quad 0 = \quad & 0 \\
& + \frac{\partial f_{0,t}}{\partial \hat{U}_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial \hat{U}_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial \hat{U}_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial \hat{U}_t} - \lambda_{4,t} \int \frac{\partial f_{4,t}}{\partial \hat{U}_t} \mu_t \left(x \right) \; dx \\
& + \int \left[\lambda_{5,t}(x) \left(\frac{\partial f_{5,t}}{\partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial \hat{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \\
& + \sum_{j=1}^{J} \int \left[\lambda_{6,j,t}(x) \left(\frac{\partial^2 f_{5,t}}{\partial u_{j,t} \partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial u_{j,t} \partial \hat{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \\
& + \int \left[\lambda_{7,t}(x) \left(- \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left[\frac{\partial b_{i,t}}{\partial \hat{U}_t} \mu_t \left(x \right) \right] \right) \right] dx, \\
& \forall t \geq 0.
\end{aligned}
$$

$$
\begin{aligned}
\tilde{U}_t: \quad 0 = \quad & \lambda_{4,t} \\
& + \frac{\partial f_{0,t}}{\partial \tilde{U}_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial \tilde{U}_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial \tilde{U}_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial \tilde{U}_t} - \lambda_{4,t} \int \frac{\partial f_{4,t}}{\partial \tilde{U}_t} \mu_t \left(x \right) \; dx \\
& + \int \left[\lambda_{5,t}(x) \left(\frac{\partial f_{5,t}}{\partial \tilde{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial \tilde{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \\
& + \sum_{j=1}^{J} \int \left[\lambda_{6,j,t}(x) \left(\frac{\partial^2 f_{5,t}}{\partial u_{j,t} \partial \tilde{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial u_{j,t} \partial \tilde{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \\
& + \int \left[\lambda_{7,t}(x) \left(- \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left[\frac{\partial b_{i,t}}{\partial \tilde{U}_t} \mu_t \left(x \right) \right] \right) \right] dx, \\
& \forall t \geq 0.
\end{aligned}
$$

Value function, distribution and policy functions

$$
\begin{aligned}
v_t\left(x\right): \quad 0 = \quad & \left(-\lambda_{5,t}(x)\rho - \sum_{i=1}^{I}\frac{\partial\left[\lambda_{5,t}(x)b_i\left(x,u_t(x),Z_t\right)\right]}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{I}\frac{\partial^2}{\partial^2 x_i}\left[\sigma_i^2(x)\lambda_{5,t}(x)\right]\right) \\
& -\sum_{j=1}^{J}\sum_{i=1}^{I}\frac{\partial}{\partial x_i}\left(\lambda_{6,j,t}(x)\frac{\partial b_i\left(x,u_t(x),Z_t\right)}{\partial u_{j,t}}\right) \\
& -(\dot{\lambda}_{5,t}(x) - \varrho\lambda_{5,t}(x)), \\
& \forall t > 0,
\end{aligned}
$$

$$
0 = \quad \lambda_{5,0}(x).
$$

$$
\begin{aligned}
\mu_t\left(x\right): \quad 0 = \quad & -\lambda_{4,t}f_4\left(x,u_t(x),Z_t\right) \\
& +\lambda_{7,t}(x)\left(\sum_{i=1}^{I}\frac{\partial\lambda_{7,t}(x)}{\partial x_i}b_i\left(x,u_t(x),Z_t\right) + \sum_{i=1}^{I}\frac{\partial^2\lambda_{7,t}(x)}{\partial^2 x_i}\frac{\sigma_i^2(x)}{2}\right) \\
& +(\dot{\lambda}_{7t}(x) - \varrho\lambda_{7,t}(x)), \\
& \forall t \geq 0,
\end{aligned}
$$

$$
0 = \quad \lim_{t\to\infty} e^{-\varrho t}\lambda_{7,t}(x).
$$

$$
\begin{aligned}
u_{l,t}\left(x\right): \quad 0 = \quad & -\lambda_{4,t}\frac{\partial f_4}{\partial u_{l,t}}\mu_t\left(x\right) \\
& + \quad \overbrace{\left[\lambda_{5,t}(x)\left(\frac{\partial f_5}{\partial u_{l,t}} + \sum_{i=1}^{I}\frac{\partial b_i}{\partial u_{l,t}}\frac{\partial v_t(x)}{\partial x_i}\right)\right]}^{=0} \\
& + \quad \sum_{j=1}^{J}\lambda_{6,k,t}(x)\left(\frac{\partial^2 f_5}{\partial u_{l,t}\partial u_{j,t}} + \sum_{i=1}^{I}\frac{\partial^2 b_i}{\partial u_{l,t}\partial u_{j,t}}\frac{\partial v_t(x)}{\partial x_i}\right) \\
& - \quad \left(\sum_{i=1}^{I}\frac{\partial\lambda_{7,t}(x)}{\partial x_i}\frac{\partial b_{i,t}}{\partial u_{l,t}}\mu_t\left(x\right)\right).
\end{aligned}
$$

2.c Discretized optimality conditions Now we discretize these conditions with respect to time and idiosyncratic states.

The idiosyncratic state is discretized by a evenly-spaced grid of size $[N_1, ..., N_I]$ where $1, .., I$ are the dimensions of the state x. We assume that in each dimension there is no mass of agents outside the compact domain $[x_{i,1}, x_{i,N_i}]$. The state step size is Δx_i.We define $x^n \equiv (x_{1,n_1}, ..., x_{i,n_i}, ..., x_{I,n_I})$, where $n_1 \in \{1, N_1\}, ..., n_I \in \{1, N_I\}$. We are assuming that, due to state constraints and/or reflecting boundaries, the dynamics of idiosyncratic states are constrained to the compact set $[x_{1,1}, x_{1,N_1}] \times [x_{2,1}, x_{2,N_2}] \times .... \times [x_{I,1}, x_{I,N_I}]$. We also define $x^{n_i+1} \equiv (x_{1,n_1}, ..., x_{i,n_i+1}, ..., x_{I,n_I})$, $x^{n_i-1} \equiv (x_{1,n_1}, ..., x_{i,n_i-1}, ..., x_{I,n_I})$ $f_t^n \equiv f(x^n, u_t^n, Z_t)$, $f_t^{n_i-1} \equiv f(x^{n_i-1}, u_t^n, Z_t)$ and $f_t^{n_i+1} \equiv f(x^{n_i+1}, u_t^n, Z_t)$. I.e. the superscript n indicates a particular grid point and the superscript n_i+1 and n_i-1 indicate neighboring grid points along dimension i.

To discretize the problem we now replace (i) time derivatives of multipliers by backward derivatives, (ii) integrals by sums (iii) derivatives with respect to x by the upwind derivatives ∇ or $\hat{\nabla}$:

$$\begin{aligned}
\nabla_i [v_t^n] &\equiv \left[\mathbb{I}_{b_{i,t}^n>0} \frac{v_t^{n_i+1} - v_t^n}{\Delta x_i} + \mathbb{I}_{b_{i,t}^n<0} \frac{v_t^n - v_t^{n_i-1}}{\Delta x_i}\right], \\
\hat{\nabla}_i [\mu_t^n] &\equiv \left[\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}\mu_t^{n_i+1} - \mathbb{I}_{b_{i,t}^n<0}\mu_t^n}{\Delta x_i} + \frac{\mathbb{I}_{b_{i,t}^n>0}\mu_t^n - \mathbb{I}_{b_{i,t}^{n_i-1}>0}\mu_t^{n_i-1}}{\Delta x_i}\right],
\end{aligned}$$

for any discretized functions v_t^n, μ_t^n. We simplify the notation for sums $\sum_n \equiv \sum_{n1\in\{1,..,N_1\},..,n_I\in\{1,..,N_I\}}$.

We maintain the subscript t even if it refers now to discrete time with a step Δt, that is, $X_{t+1}$is the shortcut for $X_{t+\triangle t}$. The second-order derivative is approximated as

$$\triangle_i [v_t^n] \equiv \left[\frac{\left(v_t^{n_i+1}\right) + \left(v_t^{n_i-1}\right) - 2\left(v_t^n\right)}{(\Delta x_i)^2}\right].$$

We start with the optimality condition for U_t

$$U_t: \quad 0 = \quad -\left(\frac{\lambda_{2,t} - \lambda_{2,t-1}}{\Delta t} - \varrho\lambda_{2,t}\right) \tag{111}$$

$$+\frac{\partial f_0}{\partial U_t} - \lambda_{1,t}\frac{\partial f_1}{\partial U_t} - \lambda_{2,t}\frac{\partial f_2}{\partial U_t} - \lambda_{3,t}\frac{\partial f_3}{\partial U_t} - \lambda_{4,t}\sum_{n=1}^{N}\frac{\partial f_4^n}{\partial U_t}\mu_t^n \tag{112}$$

$$+\sum_n \left[\lambda_{5,t}^n \left(\frac{\partial f_5^n}{\partial U_t} + \sum_{i=1}^{I}\frac{\partial b_i^n}{\partial U_t}\nabla_i\left[v_t^n\right]\right)\right]$$

$$+\sum_{j=1}^{J}\sum_n \left[\lambda_{6,j,t}^n \left(\frac{\partial^2 f_5^n}{\partial u_j \partial U_t} + \sum_{i=1}^{I}\frac{\partial b_i^n}{\partial u_j \partial U_t}\nabla_i\left[v_t^n\right]\right)\right]$$

$$+\sum_n \left[-\lambda_{7,t}^n \sum_{i=1}^{I}\hat{\nabla}_i\left[\frac{\partial b_{i,t}^n}{\partial U_t}\mu_t^n\right]\right] \tag{113}$$

$$\forall t \geq 0.$$

The optimality conditions for the other aggregate variables look very much alike:

$$X_t: \quad 0 = \quad -\left(\frac{\lambda_{1,t} - \lambda_{1,t-1}}{\Delta} - \varrho\lambda_{1,t}\right)$$

$$+\frac{\partial f_0}{\partial X_t} - \lambda_{1,t}\frac{\partial f_1}{\partial X_t} - \lambda_{2,t}\frac{\partial f_2}{\partial X_t} - \lambda_{3,t}\frac{\partial f_3}{\partial X_t} - \lambda_{4,t}\sum_n\frac{\partial f_4^n}{\partial X_t}\mu_t^n$$

$$+\sum_n \left[\lambda_{5,t}^n \left(\frac{\partial f_5^n}{\partial X_t} + \sum_{i=1}^{I}\frac{\partial b_i^n}{\partial X_t}\nabla_i\left[v_t^n\right]\right)\right]$$

$$+\sum_{j=1}^{J}\sum_n \left[\lambda_{6,j,t}^n \left(\frac{\partial^2 f_5^n}{\partial u_j \partial X_t} + \sum_{i=1}^{I}\frac{\partial b_i^n}{\partial u_j \partial X_t}\nabla_i\left[v_t^n\right]\right)\right]$$

$$+\sum_n \left[-\lambda_{7,t}^n \sum_{i=1}^{I}\hat{\nabla}_i\left[\frac{\partial b_{i,t}^n}{\partial X_t}\mu_t^n\right]\right]$$

$$\forall t > 0.$$

$$
\begin{aligned}
\hat{U}_t: \quad 0 = \quad & 0 \\
& + \frac{\partial f_0}{\partial \hat{U}_t} - \lambda_{1,t} \frac{\partial f_1}{\partial \hat{U}_t} - \lambda_{2,t} \frac{\partial f_2}{\partial \hat{U}_t} - \lambda_{3,t} \frac{\partial f_3}{\partial \hat{U}_t} - \lambda_{4,t} \sum_n \frac{\partial f_4^n}{\partial \hat{U}_t} \mu_t^n \\
& + \sum_n \left[\lambda_{5,t}^n \left(\frac{\partial f_5^n}{\partial \hat{U}_t} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial \hat{U}_t} \nabla_i \left[v_t^n \right] \right) \right] \\
& + \sum_{j=1}^J \sum_n \left[\lambda_{6,j,t}^n \left(\frac{\partial^2 f_5^n}{\partial u_j \partial \hat{U}_t} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial u_j \partial \hat{U}_t} \nabla_i \left[v_t^n \right] \right) \right] \\
& + \sum_n \left[-\lambda_{7,t}^n \sum_{i=1}^I \hat{\nabla}_i \left[\frac{\partial b_{i,t}^n}{\partial \hat{U}_t} \mu_t^n \right] \right] \\
& \forall t \geq 0.
\end{aligned}
$$

$$
\begin{aligned}
\tilde{U}_t: \quad 0 = \quad & \lambda_{4,t} \\
& + \frac{\partial f_0}{\partial \tilde{U}_t} - \lambda_{1,t} \frac{\partial f_1}{\partial \tilde{U}_t} - \lambda_{2,t} \frac{\partial f_2}{\partial \tilde{U}_t} - \lambda_{3,t} \frac{\partial f_3}{\partial \tilde{U}_t} - \lambda_{4,t} \sum_{n=1}^N \frac{\partial f_4^n}{\partial \tilde{U}_t} \mu_t^n \\
& + \sum_n \left[\lambda_{5,t}^n \left(\frac{\partial f_5^n}{\partial \tilde{U}_t} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial \tilde{U}_t} \nabla_i \left[v_t^n \right] \right) \right] \\
& + \sum_{j=1}^J \sum_n \left[\lambda_{6,j,t}^n \left(\frac{\partial^2 f_5^n}{\partial u_j \partial \tilde{U}_t} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial u_j \partial \tilde{U}_t} \nabla_i \left[v_t^n \right] \right) \right] \\
& + \sum_n \left[-\lambda_{7,t}^n \sum_{i=1}^I \hat{\nabla}_i \left[\frac{\partial b_{i,t}^n}{\partial \tilde{U}_t} \mu_t^n \right] \right] \\
& \forall t \geq 0.
\end{aligned}
$$

The discretized optimality condition with respect to the value function $v_t(x)$, the distribution $\mu_t(x)$ and the individual jump variable $u_{j,t}(x)$ are.

$$\begin{aligned} v_t(x): \quad 0 = \quad & -\lambda_{5,t}^n \rho - \sum_{i=1}^{I} \hat{\nabla}_i \left[\lambda_{5,t}^n b_{i,t}^n\right] \\ & + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \nabla_i \left[\sigma_{i,k}^n \lambda_{5,t}^n\right] \\ & - \sum_{j=1}^{J} \sum_{i=1}^{I} \left(\hat{\nabla}_i \left[\lambda_{6,j,t}^n \frac{\partial b_{i,t}^n}{\partial u_{j,t}^n} \right] \right) \\ & - \left(\frac{\lambda_{5,t}^n - \lambda_{5,t-1}^n}{\Delta t} - \varrho \lambda_{5,t}^n \right). \end{aligned} \tag{114}$$

$$\begin{aligned} \mu_t(x): \quad 0 = \quad & -\lambda_{4,t} f_{4,t}^n \\ & + \lambda_{7,t}(x) \left(\sum_{i=1}^{I} b_i\left(x, u_t(x), Z_t\right) \nabla_i \left[\lambda_{7,t}^n\right] + \frac{1}{2} \sum_{i=1}^{I} \left(\sigma_i^2\right)^n \triangle_i^2 \left[\lambda_{7,t}^n\right] \right) \\ & + \frac{\lambda_{7,t}^n - \lambda_{7,t-1}^n}{\Delta t} - \varrho \lambda_{7,t}^n \end{aligned} \tag{115}$$

$$\begin{aligned} u_{l,t}(x): \quad 0 = \quad & -\lambda_{4,t} \frac{\partial f_4}{\partial u_{l,t}} \mu_t^n \\ & + \sum_{j=1}^{J} \lambda_{6,k,t}^n \left(\frac{\partial^2 f_{5,t}^n}{\partial u_{l,t}^n \partial u_{j,t}^n} + \sum_{i=1}^{I} \frac{\partial^2 b_{i,t}^n}{\partial u_{l,t}^n \partial u_{j,t}^n} \nabla_i \left[v_t^n\right] \right) \\ & - \sum_{i=1}^{I} \nabla_i \left[\lambda_{7,t}^n\right] \frac{\partial b_{i,t}^n}{\partial u_{l,t}} \mu_t^n \end{aligned} \tag{116}$$

3. Discretize, then optimize We follow here the reverse approach, discretizing first and optimizing next.3.a The discretized planner's problem

Now first discretize the optimization problem with respect to time (time step Δt) and the idiosyncratic state (N grid points, grid step Δx_i). We define the discount factor $\beta \equiv (1+\varrho\Delta t)^{-1}$.

$$\max_{Z_t, u_t^n, \mu_t^n, v_t^n} \quad \sum_t \beta^t f_0(Z_t)$$

$$s.t.\ \forall t$$

$$\frac{X_{t+1}-X_t}{\Delta t} = f_1(Z_t) \tag{117}$$

$$\frac{U_{t+1}-U_t}{\Delta t} = f_2(Z_t) \tag{118}$$

$$0 = f_3(Z_t) \tag{119}$$

$$\tilde{U}_t = \sum_{n=1}^{N} f_4\left(x^n, u_t^n, Z_t\right)\mu_t^n \tag{120}$$

$$\rho v_t^n = \frac{v_{t+1}^n - v_t^n}{\Delta t} + f_5(x^n, u_t^n, Z_t) + \sum_{i=1}^{I} b_i\left(x^n, u_t^n, Z_t\right)\nabla_i\left[v_t^n\right] \tag{121}$$

$$+\frac{1}{2}\sum_{i=1}^{I}\left(\sigma_i^2\right)^n \triangle_i^2\left[v_t^n\right],\ \forall n$$

$$0 = \frac{\partial f_{5,t}^n}{\partial u_{j,t}^n} + \sum_{i=1}^{I}\frac{\partial b_{i,t}^n}{\partial u_{j,t}^n}\nabla_i\left[v_t^n\right], \quad \forall j, n. \tag{122}$$

$$\frac{\mu_{t+1}^n - \mu_t^n}{\Delta t} = -\sum_{i=1}^{I}\hat{\nabla}_i\left[b_{i,t}^n \mu_t^n\right] \tag{123}$$

$$+\frac{1}{2}\sum_{i=1}^{I}\triangle_i\left[\sigma_i^2 \mu_t^n\right] \tag{124}$$

$$X_0 = \bar{X}_0 \tag{125}$$

$$\mu_0^n = \bar{\mu}_0^n \tag{126}$$

3.b The Lagrangian The Lagrangian is

$$
\begin{aligned}
L \;=\; & \sum_t \beta^t f_0(Z_t) \\
& + \sum_t \beta^t \lambda_{1,t} \left\{ \frac{X_{t+1} - X_t}{\Delta t} - f_1(Z_t) \right\} \\
& + \sum_t \beta^t \lambda_{2,t} \left\{ \frac{U_{t+1} - U_t}{\Delta t} - f_2(Z_t) \right\} \\
& + \sum_t \beta^t \lambda_{3,t} \left\{ -f_3(Z_t) \right\} \\
& + \sum_t \beta^t \lambda_{4,t} \left\{ \tilde{U}_t - \sum_n f_4\left(x^n, u_t^n, Z_t\right) \mu_t^n \right\} \\
& + \sum_t \sum_n \beta^t \lambda_{5,t}^n \left\{ \begin{array}{c} -\rho v_t^n + \frac{v_{t+1}^n - v_t^n}{\Delta t} + f_5(x^n, u_t^n, Z_t) + \sum_{i=1}^I b_i\left(x^n, u_t^n, Z_t\right) \nabla_i \left[v_t^n\right] \\ + \sum_{i=1}^I \triangle_i^2 \left[v_t^n\right] \end{array} \right\} \\
& + \sum_t \sum_n \sum_{j=1}^J \beta^t \lambda_{6,j,t}^n \left\{ \frac{\partial f_{5,t}^n}{\partial u_{j,t}^n} + \sum_{i=1}^I \frac{\partial b_{i,t}^n}{\partial u_{j,t}^n} \nabla_i \left[v_t^n\right] \right\} \\
& + \sum_t \sum_n \beta^t \lambda_{7,t}^n \left\{ \begin{array}{c} -\frac{\mu_{t+1}^n - \mu_t^n}{\Delta t} - \sum_{i=1}^I \hat{\nabla}_i \left[b_{i,t}^n \mu_t^n\right] \\ + \frac{1}{2} \sum_{i=1}^I \triangle_i \left[\sigma_i^2 \mu_t^n\right] \end{array} \right\}
\end{aligned}
$$

3.c The optimality conditions The FOCs are

$$
\begin{aligned}
\frac{\partial L}{\partial U_t} : 0 \;=\; & \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} + \lambda_{2,t} \left\{ -\frac{1}{\Delta t} - \frac{\partial f_{2,t}}{\partial U_t} \right\} + \beta^{-1} \lambda_{2,t-1} \frac{1}{\Delta t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t} \sum_n \frac{\partial f_{4,t}^n}{\partial U_t} \mu_t^n \qquad (127) \\
& + \sum_n \lambda_{5,t}^n \left\{ + \frac{\partial f_{5,t}^n}{\partial U_t} + \sum_{i=1}^I \frac{\partial b_{i,t}^n}{\partial U_t} \nabla_i \left[v_t^n\right] \right\} \\
& + \sum_n \sum_{j=1}^J \lambda_{6,j,t}^n \left\{ \frac{\partial^2 f_{5,t}^n}{\partial u_{j,t}^n \partial U_t} + \sum_{i=1}^I \frac{\partial^2 b_{i,t}^n}{\partial u_{j,t}^n \partial U_t} \nabla_i \left[v_t^n\right] \right\} \\
& + \sum_n \left\{ \sum_{i=1}^I \left(\lambda_{7,t}^n - \lambda_{7,t}^{n_i - 1}\right) \left[\mathbb{I}_{b_{i,t}^n < 0} \frac{\partial b_{i,t}^n}{\partial U_t} \frac{\mu_t^n}{\Delta x_i} \right] + \sum_{i=1}^I \left(\lambda_{7,t}^{n_i + 1} - \lambda_{7,t}^n\right) \left[\mathbb{I}_{b_{i,t}^n > 0} \frac{\partial b_{i,t}^n}{\partial U_t} \frac{\mu_t^n}{\Delta x_i} \right] \right\} \\
& \forall t \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial X_t}:0 &= \frac{\partial f_{0,t}}{\partial X_t} - \lambda_{1,t}\left\{\frac{1}{\Delta t} + \frac{\partial f_{1,t}}{\partial X_t}\right\} + \beta^{-1}\lambda_{1,t-1}\frac{1}{\Delta t} - \lambda_{2,t}\frac{\partial f_{2,t}}{\partial X_t} - \lambda_{3,t}\frac{\partial f_{3,t}}{\partial X_t} - \lambda_{4,t}\sum_n \frac{\partial f^n_{4,t}}{\partial X_t}\mu^n_t \\
&+ \sum_n \lambda^n_{5,t}\left\{\frac{\partial f^n_{5,t}}{\partial X_t} + \sum_{i=1}^{I}\frac{\partial b^n_{i,t}}{\partial X_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\sum_j \lambda^n_{6,j,t}\left\{\frac{\partial^2 f^n_{5,t}}{\partial u^n_{j,t}\partial X_t} + \sum_{i=1}^{I}\frac{\partial^2 b^n_{i,t}}{\partial u^n_{j,t}\partial X_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\left\{\sum_{i=1}^{I}\left(\lambda^n_{7,t} - \lambda^{n_i-1}_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}<0}\frac{\partial b^n_{i,t}}{\partial X_t}\frac{\mu^n_t}{\Delta x_i}\right] + \sum_{i=1}^{I}\left(\lambda^{n_i+1}_{7,t} - \lambda^n_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}>0}\frac{\partial b^n_{i,t}}{\partial X_t}\frac{\mu^n_t}{\Delta x_i}\right]\right\} \\
&\forall t > 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \tilde{U}_t}:0 &= \frac{\partial f_{0,t}}{\partial \tilde{U}_t} - \lambda_{1,t}\frac{\partial f_{1,t}}{\partial \tilde{U}_t} - \lambda_{2,t}\frac{\partial f_{2,t}}{\partial \tilde{U}_t} - \lambda_{3,t}\frac{\partial f_{3,t}}{\partial \tilde{U}_t} - \lambda_{4,t}\sum_n \frac{\partial f^n_{4,t}}{\partial \tilde{U}_t}\mu^n_t \\
&+ \sum_n \lambda^n_{5,t}\left\{+\frac{\partial f^n_{5,t}}{\partial \tilde{U}_t} + \sum_{i=1}^{I}\frac{\partial b^n_{i,t}}{\partial \tilde{U}_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\sum_j \lambda^n_{6,j,t}\left\{\frac{\partial^2 f^n_{5,t}}{\partial u^n_{j,t}\partial \tilde{U}_t} + \sum_{i=1}^{I}\frac{\partial^2 b^n_{i,t}}{\partial u^n_{j,t}\partial \tilde{U}_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\left\{\sum_{i=1}^{I}\left(\lambda^n_{7,t} - \lambda^{n_i-1}_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}<0}\frac{\partial b^n_{i,t}}{\partial \tilde{U}_t}\frac{\mu^n_t}{\Delta x_i}\right] + \sum_{i=1}^{I}\left(\lambda^{n_i+1}_{7,t} - \lambda^n_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}>0}\frac{\partial b^n_{i,t}}{\partial \tilde{U}_t}\frac{\mu^n_t}{\Delta x_i}\right]\right\} \\
&\forall t \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \hat{U}_t}:0 &= \frac{\partial f_{0,t}}{\partial \hat{U}_t} - \lambda_{1,t}\frac{\partial f_{1,t}}{\partial \hat{U}_t} - \lambda_{2,t}\frac{\partial f_{2,t}}{\partial \hat{U}_t} - \lambda_{3,t}\frac{\partial f_{3,t}}{\partial \hat{U}_t} - \lambda_{4,t}\sum_n \frac{\partial f^n_{4,t}}{\partial \hat{U}_t}\mu^n_t \\
&+ \sum_n \lambda^n_{5,t}\left\{+\frac{\partial f^n_{5,t}}{\partial \hat{U}_t} + \sum_{i=1}^{I}\frac{\partial b^n_{i,t}}{\partial \hat{U}_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\sum_j \lambda^n_{6,j,t}\left\{\frac{\partial^2 f^n_{5,t}}{\partial u^n_{j,t}\partial \hat{U}_t} + \sum_{i=1}^{I}\frac{\partial^2 b^n_{i,t}}{\partial u^n_{j,t}\partial \hat{U}_t}\nabla_i\left[v^n_t\right]\right\} \\
&+ \sum_n\left\{\sum_{i=1}^{I}\left(\lambda^n_{7,t} - \lambda^{n_i-1}_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}<0}\frac{\partial b^n_{i,t}}{\partial \hat{U}_t}\frac{\mu^n_t}{\Delta x_i}\right] + \sum_{i=1}^{I}\left(\lambda^{n_i+1}_{7,t} - \lambda^n_{7,t}\right)\left[\mathbb{I}_{b^n_{i,t}>0}\frac{\partial b^n_{i,t}}{\partial \hat{U}_t}\frac{\mu^n_t}{\Delta x_i}\right]\right\} \\
&\forall t \geq 0
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial L}{\partial v_t^n}:0 \;\; = \;\; & \lambda_{5,t}^n\left\{-\rho-\frac{1}{\Delta t}+\sum_{i=1}^{I}b_{i,t}^n\frac{\mathbb{I}_{b_t^n<0}-\mathbb{I}_{b_t^n>0}}{\Delta x_i}-\sum_{i=1}^{I}\frac{2\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\} && (128)\\
& +\lambda_{5,t-1}^n\beta^{-1}\frac{1}{\Delta t}\\
& +\sum_{i=1}^{I}\lambda_{5,t}^{n_i-1}b_{i,t}^{n_i-1}\frac{\mathbb{I}_{b_{i,t}^{n_i-1}>0}}{\Delta x_i}+\sum_{i=1}^{I}\lambda_{5,t}^{n_i-1}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\\
& -\sum_{i=1}^{I}\lambda_{5,t}^{n_i+1}b_{i,t}^{n_i+1}\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}}{\Delta x_i}+\sum_{i=1}^{I}\lambda_{5,t}^{n_i+1}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\\
& +\sum_{j=1}^{J}\sum_{i=1}^{I}\left\{\lambda_{6,j,t}^n\left\{\frac{\partial b_{i,t}^n}{\partial u_{j,t}^n}\frac{\mathbb{I}_{b_{i,t}^n<0}-\mathbb{I}_{b_{i,t}^n>0}}{\Delta x_i}\right\}+\lambda_{6,j,t}^{n_i-1}\left\{\frac{\partial b_{i.t}^{n_i-1}}{\partial u_{j,t}^{n_i-1}}\frac{\mathbb{I}_{b_{i,t}^{n_i-1}>0}}{\Delta x_i}\right\}\right\}\\
& -\sum_{j=1}^{J}\sum_{i=1}^{I}\lambda_{6,j,t}^{n_i+1}\left\{\frac{\partial b_{i,t}^{n_i+1}}{\partial u_{j,t}^{n_i+1}}\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}}{\Delta x_i}\right\} && (129)\\
& \forall t\geq 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L}{\partial \mu_t^n}:0 \;\; = \;\; & -\lambda_{4,t}f_{4,t}^n && (130)\\
& +\lambda_{7,t}^n\left\{\frac{1}{\Delta t}-\sum_{i=1}^{I}\left[\left(\mathbb{I}_{b_{i,t}^n>0}-\mathbb{I}_{b_{i,t}^n<0}\right)\frac{b_{i,t}^n}{\Delta x_i}\right]-\sum_{i=1}^{I}\frac{-2\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\}\\
& +\left\{-\sum_{i=1}^{I}\lambda_{7,t}^{n_i-1}\left[\frac{\mathbb{I}_{b_{i,t}^n<0}b_{i,t}^n}{\Delta x_i}\right]+\sum_{i=1}^{I}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\}\\
& +\left\{-\sum_{i=1}^{I}\lambda_{7,t}^{n_i+1}\left[\frac{-\mathbb{I}_{b_{i,t}^n>0}b_{i,t}^n}{\Delta x_i}\right]+\sum_{i=1}^{I}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\}\\
& +\beta^{-1}\lambda_{7,t-1}^n\left\{-\frac{1}{\Delta t}\right\}\\
& \forall t>0
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial L}{\partial u_{l,t}^n}:0 \quad = \quad & -\lambda_{4,t}\frac{\partial f_{4,t}^n}{\partial u_{l,t}^n}\mu_t^n \qquad (131)\\
& +\beta^t\lambda_{5,t}^n\left\{\frac{\partial f_{5,t}^n}{\partial u_{l,t}^n}+\sum_{i=1}^{I}\frac{\partial b_{i,t}^n}{\partial u_{l,t}^n}\nabla_i\left[v_t^n\right]\right\}\\
& +\sum_j\lambda_{6,t}^n\left\{\frac{\partial^2 f_{5,t}^n}{\partial u_{j,t}^n\partial u_{l,t}^n}+\sum_{i=1}^{I}\frac{\partial^2 b_{i,t}^n}{\partial u_{j,t}^n\partial u_{l,t}^n}\nabla_i\left[v_t^n\right]\right\}\\
& +\sum_{i=1}^{I}\left(\lambda_{7,t}^n-\lambda_{7,t}^{n_i-1}\right)\left[\mathbb{I}_{b_{i,t}^n<0}\frac{\partial b_{i,t}^n}{\partial u_{l,t}^n}\frac{\mu_t^n}{\Delta x_i}\right]+\sum_{i=1}^{I}\left(\lambda_{7,t}^{n_i+1}-\lambda_{7,t}^n\right)\left[\mathbb{I}_{b_{i,t}^n>0}\frac{\partial b_{i,t}^n}{\partial u_{l,t}^n}\frac{\mu_t^n}{\Delta x_i}\right]\\
& \forall t\geq 0
\end{aligned}
$$

By the individual agents' optimality condition, line 2 of this expression is equal to 0.

4. Compare Finally, by comparing the respective discretized optimality conditions, we show that the two procedures yield the same equilibrium conditions in the limit. Consider first the condition for U_t. The optimize-discretize condition is given by (111), which we reproduce here

$$
\begin{aligned}
U_t:\quad 0= \quad & -\left(\frac{\lambda_{2,t}-\lambda_{2,t-1}}{\Delta}-\varrho\lambda_{2,t}\right)\\
& +\frac{\partial f_0}{\partial U_t}-\lambda_{1,t}\frac{\partial f_1}{\partial U_t}-\lambda_{2,t}\frac{\partial f_2}{\partial U_t}-\lambda_{3,t}\frac{\partial f_3}{\partial U_t}-\lambda_{4,t}\sum_{n=1}^{N}\frac{\partial f_4^n}{\partial U_t}\mu_t^n\\
& +\sum_n\lambda_{5,t}^n\left\{\frac{\partial f_5^n}{\partial U_t}+\sum_{i=1}^{I}\frac{\partial b_i^n}{\partial U_t}\nabla_i\left[v_t^n\right]\right\}\\
& +\sum_n\sum_{j=1}^{J}\lambda_{6,j,t}^n\left\{\frac{\partial^2 f_{5,t}^n}{\partial u_{j,t}^n\partial U_t}+\sum_{i=1}^{I}\frac{\partial^2 b_t^n}{\partial u_{j,t}^n\partial U_t}\nabla_i\left[v_t^n\right]\right\}\\
& +\sum_n\left[-\lambda_{7,t}^n\sum_{i=1}^{I}\hat{\nabla}_i\left[\frac{\partial b_{i,t}^n}{\partial U_t}\mu_t^n\right]\right]\\
& \forall t\geq 0
\end{aligned}
$$

The discretize-optimize condition (127), rearranges to

$$
\begin{aligned}
\frac{\partial L}{\partial U_t}: 0 \quad = \quad & -\left(\frac{\lambda_{2,t}-\lambda_{2,t-1}}{\Delta t}-\frac{\beta^{-1}-1}{\Delta t}\lambda_{2,t-1}\right) \\
& +\frac{\partial f_{0,t}}{\partial U_t}-\lambda_{1,t}\frac{\partial f_{1,t}}{\partial U_t}-\lambda_{2,t}\frac{\partial f_{2,t}}{\partial U_t}-\lambda_{3,t}\frac{\partial f_{3,t}}{\partial U_t}-\lambda_{4,t}\sum_{n=1}^{N}\frac{\partial f_{4,t}^n}{\partial U_t}\mu_t^n \\
& +\sum_{n=1}^{N}\lambda_{5,t}^n\left\{\frac{\partial f_{5,t}^n}{\partial U_t}+\sum_{i=1}^{I}\frac{\partial b_i^n}{\partial U_t}\nabla_i\left[v_t^n\right]\right\} \\
& +\sum_{n=1}^{N}\sum_{j=1}^{J}\lambda_{6,j,t}^n\left\{\frac{\partial^2 f_{5,t}^n}{\partial u_{j,t}^n\partial U_t}+\frac{\partial^2 b_t^n}{\partial u_{j,t}^n\partial U_t}\nabla_i\left[v_t^n\right]\right\} \\
& +\sum_{n}\left\{\sum_{i=1}^{I}\left(\lambda_{7,t}^n-\lambda_{7,t}^{n_i-1}\right)\left[\mathbb{I}_{b_{i,t}^n<0}\frac{\partial b_{i,t}^n}{\partial U_t}\frac{\mu_t^n}{\Delta x_i}\right]+\sum_{i=1}^{I}\left(\lambda_{7,t}^{n_i+1}-\lambda_{7,t}^n\right)\left[\mathbb{I}_{b_{i,t}^n>0}\frac{\partial b_{i,t}^n}{\partial U_t}\frac{\mu_t^n}{\Delta x_i}\right]\right\} \\
& \forall t\geq 0
\end{aligned}
$$

The second to fourth lines are evidently identical. The last lines also coincide once we take into account the definition of $\hat{\nabla}_i\left[\frac{\partial b_{i,t}^n}{\partial U_t}\mu_t^n\right] = \frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}\frac{\partial b_{i,t}^{n_i+1}}{\partial U_t}\mu_t^{n_i+1}-\mathbb{I}_{b_{i,t}^n<0}\frac{\partial b_{i,t}^n}{\partial U_t}\mu_t^n}{\Delta x_i} + \frac{\mathbb{I}_{b_{i,t}^n>0}\frac{\partial b_{i,t}^n}{\partial U_t}\mu_t^n-\mathbb{I}_{b_{i,t}^{n_i-1}>0}\frac{\partial b_{i,t}^{n_i-1}}{\partial U_t}\mu_t^{n_i-1}}{\Delta x_i}$.

Finally compare the first lines. Since $\beta\equiv(1+\varrho\Delta t)^{-1}$ we have that $\frac{\beta^{-1}-1}{\Delta t}=\varrho$. The difference between these two equations hence is $\|\varrho\left(\lambda_{2,t}-\lambda_{2,t-1}\right)\|$. In the limit as $\Delta t\to 0$, and provided that $\lambda_{2,t}$ features no jumps for $t>0$,this difference converges to zero.The same argument applies to the optimality conditions with respect to X_t with the difference now proportional to $\|\varrho\left(\lambda_{1,t}-\lambda_{1,t-1}\right)\|$. The optimality conditions with respect to $\hat{U}_t$ and $\tilde{U}_t$ are identical, that is, there is no difference.

Next consider the two discretized optimality conditions with respect to v_t^n (114) and (128). After some rearranging they are given by

$$
\begin{aligned}
v_t(x): \quad 0 = \quad & -\sum_{i=1}^{I}\left(\frac{\mathbb{I}_{b_{i,t}^{n}>0}\lambda_{5,j,t}^{n}b_{i,t}^{n}-\mathbb{I}_{b_{i,t}^{n_i-1}>0}\lambda_{5,j,t}^{n_i-1}b_{i,t}^{n_i-1}}{\Delta x_i}+\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}\lambda_{5,j,t}^{n_i+1}b_{i,t}^{n_i+1}-\mathbb{I}_{b_{i,t}^{n}<0}\lambda_{5,j,t}^{n}b_{i,t}^{n}}{\Delta x_i}\right) \\
& +\frac{1}{2}\sum_{i=1}^{I}\frac{\left(\sigma_i^2\right)^{n_i+1}\lambda_{5,t}^{n_i+1}+\left(\sigma_i^2\right)^{n_i-1}\lambda_{5,t}^{n_i-1}-2\left(\sigma_i^2\right)^{n}\lambda_{5,t}^{n}}{\left(\Delta x_i\right)^2} \\
& -\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\frac{\mathbb{I}_{b_{i,t}^{n}>0}\lambda_{6,j,t}^{n}\frac{\partial b_{i,t}^{n}}{\partial u_{j,t}^{n}}-\mathbb{I}_{b_{i,t}^{n_i-1}>0}\lambda_{6,j,t}^{n_i-1}\frac{\partial b_{i,t}^{n_i-1}}{\partial u_{j,t}^{n_i-1}}}{\Delta x_i}+\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}\lambda_{6,j,t}^{n_i+1}\frac{\partial b_{i,t}^{n_i+1}}{\partial u_{j,t}^{n_i+1}}-\mathbb{I}_{b_{i,t}^{n}<0}\lambda_{6,j,t}^{n}\frac{\partial b_{i,t}^{n}}{\partial u_{j,t}^{n}}}{\Delta x_i}\right) \\
& -\lambda_{5,t}^{n}\rho-\left(\frac{\lambda_{5,t}^{n}-\lambda_{5,t-1}^{n}}{\Delta t}-\varrho\lambda_{5,t}^{n}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial L}{\partial v_t^n}:0 \quad = \quad & \lambda_{5,t}^{n}\left\{\sum_{i=1}^{I}b_{i,t}^{n}\frac{\mathbb{I}_{b_t^n<0}-\mathbb{I}_{b_t^n>0}}{\Delta x_i}-\sum_{i=1}^{I}\frac{\left(\sigma_i^2\right)^n}{\left(\Delta x_i\right)^2}\right\} \\
& +\left\{\sum_{i=1}^{I}\lambda_{5,t}^{n_i-1}b_{i,t}^{n_i-1}\frac{\mathbb{I}_{b_{i,t}^{n_i-1}>0}}{\Delta x_i}+\sum_{i=1}^{I}\lambda_{5,t}^{n_i-1}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\} \\
& +\left\{-\sum_{i=1}^{I}\lambda_{5,t}^{n_i+1}b_{i,t}^{n_i+1}\frac{\mathbb{I}_{b_{i,t}^{n_i+1}<0}}{\Delta x_i}+\sum_{i=1}^{I}\lambda_{5,t}^{n_i+1}\frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2}\right\} \\
& +\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\lambda_{6,j,t}^{n}\frac{\partial b_{i,t}^{n}}{\partial u_{j,t}^{n}}\frac{\mathbb{I}_{b_t^n<0}-\mathbb{I}_{b_t^n>0}}{\Delta x_i}+\lambda_{6,j,t}^{n_i-1}\frac{\partial b_{i,t}^{n_i-1}}{\partial u_{j,t}^{n_i-1}}\frac{\mathbb{I}_{b_t^{n_i-1}>0}}{\Delta x_i}-\lambda_{6,j,t}^{n_i+1}\frac{\partial b_{i,t}^{n_i+1}}{\partial u_{j,t}^{n_i+1}}\frac{\mathbb{I}_{b_t^{n_i+1}<0}}{\Delta x_i}\right) \\
& -\rho\lambda_{5,t}^{n}-\left(\frac{\lambda_{5,t}^{n}-\lambda_{5,t-1}^{n}}{\Delta t}-\frac{\beta^{-1}-1}{\Delta t}\lambda_{5,t-1}^{n}\right) \qquad (132)
\end{aligned}
$$

Again these, two expressions are identical up to the last time index in the last line (λ_5^n), and thus the difference is $\|\varrho\left(\lambda_{5,t}-\lambda_{5,t-1}\right)\|$.

Next, consider the two discretized optimality conditions with respect to μ_t^n (115)

and (130). After some rearranging they are given by

$$\begin{aligned}\mu_t(x): \quad 0 = \quad & -\lambda_{4,t} f_{4,t}^n \\ & + \sum_{i=1}^{I} b_{i,t}^n \left[\mathbb{I}_{b_{i,t}^n > 0} \frac{\lambda_{7,t}^{n_i+1} - \lambda_{7,t}^n}{\Delta x_i} + \mathbb{I}_{b_{i,t}^n < 0} \frac{\lambda_{7,t}^n - \lambda_{7,t}^{n_i-1}}{\Delta x_i} \right] + \frac{1}{2} \sum_{i=1}^{I} \left(\sigma_i^2\right)^n \frac{\lambda_{7,t}^{n_i+1} + \lambda_{7,t}^{n_i-1} - 2\lambda_{7,t}^n}{2\left(\Delta x_i\right)^2} \\ & + \frac{\lambda_{7,t}^n - \lambda_{7,t-1}^n}{\Delta t} - \varrho \lambda_{7,t}^n \end{aligned} \tag{133}$$

$$\begin{aligned}\frac{\partial L}{\partial \mu_t^n} : 0 \quad = \quad & -\lambda_{4,t} f_{4,t}^n \\ & + \lambda_{7,t}^n \left\{ - \sum_{i=1}^{I} \left[\left(\mathbb{I}_{b_{i,t}^n > 0} - \mathbb{I}_{b_{i,t}^n < 0} \right) \frac{b_{i,t}^n}{\Delta x_i} \right] - \sum_{i=1}^{I} \frac{-2\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2} \right\} \\ & - \sum_{i=1}^{I} \left[\lambda_{7,t}^{n_i-1} \frac{\mathbb{I}_{b_{i,t}^n < 0} b_{i,t}^n}{\Delta x_i} \right] + \sum_{i=1}^{I} \lambda_{7,t}^{n_i-1} \frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2} \\ & + \sum_{i=1}^{I} \left[\lambda_{7,t}^{n_i+1} \frac{\mathbb{I}_{b_{i,t}^n > 0} b_{i,t}^n}{\Delta x_i} \right] + \sum_{i=1}^{I} \lambda_{7,t}^{n_i+1} \frac{\left(\sigma_i^2\right)^n}{2\left(\Delta x_i\right)^2} \\ & + \frac{\lambda_{7,t}^n - \lambda_{7,t-1}^n}{\Delta t} - \frac{\beta^{-1} - 1}{\Delta t} \lambda_{7,t-1}^n, \end{aligned} \tag{134}$$

which again differ in $\left\| \varrho \left(\lambda_{7,t} - \lambda_{7,t-1} \right) \right\|$.

Finally, consider the two discretized optimality conditions with respect to $u_{l,t}^n(x)$, (116) and (131). After some rearranging they are given by

$$\begin{aligned}u_{l,t}(x): \quad 0 = \quad & -\lambda_{4,t} \frac{\partial f_4}{\partial u_{l,t}} \mu_t^n \\ + \quad & \sum_{j=1}^{J} \lambda_{6,l,t}^n \left(\frac{\partial^2 f_{5,t}^n}{\partial u_{j,t}^n \partial u_{l,t}^n} + \sum_{i=1}^{I} \frac{\partial^2 b_{i,t}^n}{\partial u_{j,t}^n \partial u_{l,t}^n} \left[\mathbb{I}_{b_{i,t}^n > 0} \frac{v_t^{n_i+1} - v_t^n}{\Delta x_i} + \mathbb{I}_{b_{i,t}^n < 0} \frac{v_t^n - v_t^{n1}}{\Delta x_i} \right] \right) \\ - \quad & \sum_{i=1}^{I} \left[\mathbb{I}_{b_{i,t}^n > 0} \frac{\lambda_{7,t}^{n_i+1} - \lambda_{7,t}^n}{\Delta x_i} + \mathbb{I}_{b_{i,t}^n < 0} \frac{\lambda_{7,t}^n - \lambda_{7,t}^{n_i-1}}{\Delta x_i} \right] \frac{\partial b_{i,t}^n}{\partial u_{l,t}} \mu_t^n \end{aligned} \tag{135}$$

$$
\begin{aligned}
\frac{\partial L}{\partial u_{l,t}^{n}}:0 \;&=\; -\lambda_{4,t}\frac{\partial f_{4,t}^{n}}{\partial u_{l,t}^{n}}\mu_t^n \\
&+\sum_j \lambda_{6,t}^{n}\left\{\frac{\partial^2 f_{5,t}^{n}}{\partial u_{j,t}^{n}\partial u_{l,t}^{n}}+\sum_{i=1}^{I}\frac{\partial^2 b_{i,t}^{n}}{\partial u_{j,t}^{n}\partial u_{l,t}^{n}}\left[\mathbb{I}_{b_{i,t}^{n}>0}\frac{v_t^{n_i+1}-v_t^{n}}{\Delta x_i}+\mathbb{I}_{b_{i,t}^{n}<0}\frac{v_t^{n}-v_t^{n_i-1}}{\Delta x_i}\right]\right\} \\
&+\left[\sum_{i=1}^{I}\left(\lambda_{7,t}^{n}-\lambda_{7,t}^{n_i-1}\right)\left[\mathbb{I}_{b_{i,t}^{n}<0}\frac{1}{\Delta x_i}\right]+\sum_{i=1}^{I}\left(\lambda_{7,t}^{n_i+1}-\lambda_{7,t}^{n}\right)\left[\mathbb{I}_{b_{i,t}^{n}>0}\frac{1}{\Delta x_i}\right]\right]\frac{\partial b_{i,t}^{n}}{\partial u_{l,t}}\mu_t^n,
\end{aligned}
$$

which are identical.To summarize, whether one discretize the optimality conditions of the planner and then discretizes them, or one discretizes the planner's problem and then derives the optimality conditions, one arrives to a set of optimality conditions that coincide in everything but the timing of the multiplier in the term $\varrho\lambda_t$. Provided that multipliers experience no jumps, the difference between the two approaches goes to 0 as $\Delta t \to 0$. Note that this issue has nothing to do with heterogeneity.

D.3 Solving the Nuño and Thomas model using Dynare

Here we apply the "discretize-optimize" methodology outlined in Section D to the heterogeneous-agent model introduced in Nuño and Thomas (2016). This is a model *à la* Aiyagari-Bewley-Huggett with non-state-contingent long-term nominal debt contracts. Finding the optimal policy in this problem requires that the central bank takes into account not only the dynamics of the state distribution (given by the KF equation) but also the HJB equation. Figure 8 displays the time-0 optimal policy (inflation) in this case, compared to the one obtained through the "optimize-discretize" methodology employed in Nuño and Thomas (2016). Optimal inflation coincides in both cases, up to a numerical error that is reduced as we increase the number of grid points and we reduce the time step.

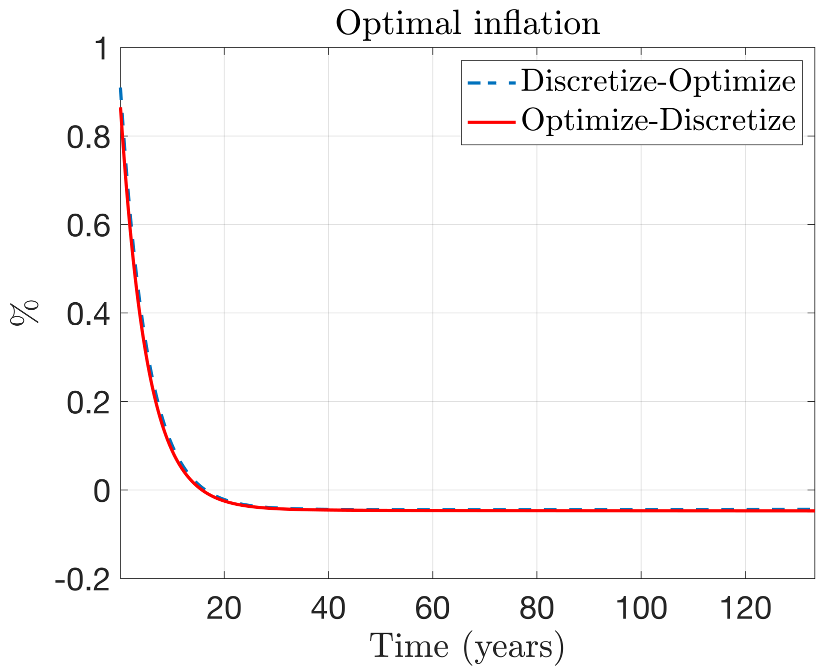

Figure 8: Time-0 optimal monetary policy using the two approaches.

Notes: The figure shows the optimal path of inflation in the Nuño and Thomas (2016) model using the "discretize-optimize" and "optimize-discretize" methods.