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: November 2021

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. Although a reduction in real rates increases misallocation in partial equilibrium, general-equilibrium effects overturn this result around: a monetary expansion increases the investment of high-productivity firms relatively more than that of low-productivity 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. We show how a central bank without pre-commitments engineers an unexpected monetary expansion to increase TFP in the medium run. In the event of a cost-push shock, the central bank *leans with the wind* to increase demand and reduce misallocation.

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, Veronica Guerrieri, Masashige Hamano, Ben Moll, Ricardo Reis, Gianluca Violante 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. 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. Each entrepreneur decides whether to operate her firm or not: if it is profitable to operate the firm, the entrepreneur hires workers and rents capital in order to produce. Otherwise, she does not operate, lending her funds to other entrepreneurs. We assume that entrepreneurs have access to a constant returns to scale technology: since returns to capital are constant, only those entrepreneurs above a certain productivity threshold operate. We embed this heterogeneous firm sector into an otherwise standard continuous-time New Keynesian model. This economy allows for an aggregate representation akin to that in the complete-market New Keynesian model. The key difference is that, while in the complete-market economy aggregate Total Factor Productivity (TFP) is exogenous, in our economy TFP evolves endogenously as a result of the heterogeneous investment and production decisions of firms, which drive the extent to which capital is misallocated. TFP thus depends on the distribution of capital across firms or, equivalently, net worth across those entrepreneurs with productivities above the threshold.

First, we study the positive properties of this economy. By changing interest rates, the central bank influences the capital allocation through both changes in the productivity threshold and in the net-worth distribution. We call these two channels the *productivity-threshold channel* and the *net-worth distribution channel* of monetary pol-

icy. We show how, in partial equilibrium, a reduction in real rates reduces the productivity threshold, thus crowding in low-productivity firms into production and worsening capital misallocation: when real interest rates decrease, low-productivity projects can turn into profitable firms due to their lower costs of capital. This is the "direct effect" of monetary policy on aggregate productivity.[1] In general equilibrium, however, this result does not necessarily hold, as changes in other prices have different effects on TFP through both the productivity-threshold and net-worth distribution channels.

We analyze the overall effect of monetary policy on aggregate productivity by calibrating the model to the U.S. and solving it numerically. We find that an expansionary monetary policy shock produces an *increase* in TFP, as general equilibrium effects dominate the direct effect. This increase in TFP is brought about through both of the two aforementioned channels: after a monetary policy surprise high-productivity firms increase their investment relatively more than low-productivity ones, crowding out the latter. This increases the market share of high-productivity firms and reduces capital misallocation.

We present empirical evidence supporting this mechanism. We show how high-productivity firms invest more relative to low-productivity ones in response to an expansionary monetary policy shock. 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, proxied by the marginal revenue product of capital (MRPK), using an empirical specification that follows Ottonello and Winberry (2020) closely. We find that having one standard deviation higher MRPK implies an additional 29pp increase in the investment rate in response to a 1pp cut in interest rates.

We turn next to the normative prescriptions of the model. We analyze the Ramsey problem of a benevolent central bank. In the steady-state of the Ramsey plan inflation is zero, 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. Whereas the optimal policy in the case with complete markets is time consistent, financial frictions introduce a source of time inconsistency: the central bank engineers a temporary monetary expansion in the short run to increase medium-term TFP through the thresh-

[1]The separation between direct and indirect effects of monetary policy is discussed in Kaplan et al. (2018).

old and net-worth distribution channels. 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), 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.

Finally, we analyze the optimal response of monetary policy to shocks. We study optimal monetary policy from a 'timeless perspective' (Woodford, 2003) in which the central bank has to honor its pre-commitments when the economy is hit by a shock. We consider first a cost-push shock. The prescription in the complete-markets model is that the central bank should lean *against* the wind (Gali, 2008) – by tightening the monetary policy stance but tolerating some inflation to minimize the reduction in output. In the case of financial frictions, the central bank should instead lean *with* the wind. It loosens monetary policy despite the rise in inflation, as the increase in demand boosts high-productivity firms' investment and thus increases TFP, amplifying the expansionary demand effect on output. We analyze then a temporary demand shock. In this case, the optimal response is strict price stability (zero inflation), the same policy as 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.[2]

[2]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,

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.[3] Focusing on the impact of lower interest rates in a small open economy, Reis (2013) and Gopinath et al. (2017) analyze 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.

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) 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. Finding the optimal monetary policy in a model with heterogeneous firms is a challenge, as the net-worth distribution is an infinite-dimensional object. To overcome this problem, 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 using Dynare and can be easily generalized to other problems.

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).

[3] 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.

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 rms

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.[4]

Entrepreneurs are heterogeneous in two dimensions: their net worth a_t and in their idiosyncratic productivity z_t.[5] Each entrepreneur owns a technology which uses capital k_t and labor l_t to produce input good y_t:

$$y_t = f_t(z_t; k_t; l_t) = (z_t k_t)^{\quad} (l_t)^{1} \quad : \tag{1}$$

The labor share $\quad 2\ (0; 1)$ is the same across entrepreneurs. Idiosyncratic productivity

[4] 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.

[5] 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.

z_t follows a diffusion process,

$$dz_t = \quad (z_t)dt + \quad (z_t)dW_t; \tag{2}$$

where (z) is the drift and (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 real price q_t. Capital depreciates at rate . An entrepreneur's flow budget constraint can be expressed as follows

$$a_{t+1} = \frac{1}{q_t}\Big[\underbrace{m_t f_t(z_t;k_t;l_t) \quad w_t l_t \quad R_t k_t}_{\text{Firm's profits}} + \underbrace{(R_t + q_t \quad)}_{\text{Return on net worth}} q_{t-1}a_t \quad \underbrace{d_t}_{\text{Dividends}}\Big]: \tag{3}$$

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

$$q_t k_t \qquad q_t a_t: \tag{4}$$

Entrepreneurs retire and return to the household according to an exogenous Poisson process with arrival rate . 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

$$V_0(z;a) = \max_{k_t;l_t;d_t} E_0 \int_0^1 e^{\quad t} \, {}_{0;t} \Big[\underbrace{d_t}_{\text{Dividends}} + \underbrace{q_t a_t}_{\text{Terminal value}}\Big] dt; \tag{5}$$

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 $_{0;t}$. Below we show that $_{0;t} = e^{\int_0^t r_s ds}$; 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,

$$\max_{k_t;l_t} f m_t f_t(z_t; k_t; l_t) \quad w_t l_t \quad R_t k_t g; \tag{6}$$

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 entrepreneurs remain inactive. Firm's demand for capital and labor is :

$$k_t(z_t; a_t) = \begin{cases} a_t; & \text{if } z_t \quad z_t; \\ 0; & \text{if } z_t < z_t; \end{cases} \tag{7}$$

$$l_t(z_t; a_t) = \frac{(1 \quad)m_t}{w_t}^{1=} z_t k_t(z_t; a_t): \tag{8}$$

Firm's profits are then given by

$$_t(z_t; a_t) = \max f z_t' _t \quad R_t; 0g \; a_t; \quad \text{where} \quad ' _t = \frac{(1 \quad)}{w_t}^{(1 \quad)=} m_t^{\frac{1}{}}; \tag{9}$$

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

$$z_t ' _t = R_t: \tag{10}$$

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

$$\begin{aligned} \dot{a}_t &= \frac{1}{q_t}[\ _t(z_t; a_t) + (R_t \quad q_t)a_t \quad d_t] \\ &= \frac{1}{q_t}[(\ \max\{z_t' \ _t \quad R_t; 0\} + R_t \quad q_t)a_t \quad d_t]: \end{aligned} \tag{11}$$

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* $(R_t \quad q_t)$; while for the household the return of this unit of capital is *exactly* $(R_t \quad q_t)$. 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 $\quad 2 (0; 1)$ of these dividends to finance the net worth of the new entrepreneurs, so net dividends are $(1 \quad)$ 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

$$W_t = \max_{C_t; L_t; B_t^N; D_t} \int_0^1 e^{\ \int_t^h t} u(C_t; L_t)dt: \tag{12}$$

$$\begin{aligned} \text{s:t:} \quad D_t q_t &= (R_t \quad q_t) D_t \quad S_t^N \quad C_t + w_t L_t + T_t; \\ \dot{B}_t^N &= (i_t \quad _t) B_t^N + S_t^N; \end{aligned} \tag{13}$$

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 ($[\ _t q_t \quad _t \quad (\ _t)] K_t$), the profits from retail goods producers ($\ _t$ from equation 21) and net dividends received from entrepreneurs ($(1 \quad) q_t A_t$).

We assume separable utility of CRRA form, i.e., $u(C_t; L_t) = \frac{C_t^1}{1} \quad \frac{L_t^{1+\#}}{1+\#}$. Solving

this problem (see Appendix (A.4) for details), we obtain the Euler equation,

$$\frac{C_t}{C_t} = \frac{r_t \quad {}^{h}_{t}}{}; \tag{14}$$

the labor supply condition

$$w_t = \frac{L^{\#}_t}{C_t}; \tag{15}$$

and the Fisher equation

$$r_t = i_t \quad {}_t; \tag{16}$$

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

$$r_t \quad \frac{R_t \quad q_t + q_t}{q_t}; \tag{17}$$

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 ${}_{0;t}$ can be defined as

$${}_{0;t} \quad e^{\int_0^t {}^{h}_{t} ds} \frac{u^0_c(C_t)}{u^0_c(C_0)} = e^{\int_0^t r_s ds}.$$

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 $j \ 2 \ [0;1]$,

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

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

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

2.4 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 T_t.[6] The adjustment costs are quadratic in the rate of price change $(p_{j,t}/p_{j,t-1})$ and expressed as a fraction of output (Y_t),

$$\Phi_t\left(\frac{p_{j,t}}{p_{j,t-1}}\right) = \frac{\phi}{2}\left(\frac{p_{j,t}}{p_{j,t-1}}\right)^2 Y_t;$$

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

$$\int_0^1 Q_{0,t}\left[\Psi_t(p_t) - \Phi_t\left(\frac{p_t}{p_{t-1}}\right)\right] dt; \tag{19}$$

where

$$\Psi_t(p_t) = \left(\frac{p_t}{P_t}\right)^{1-\varepsilon}\left(\tilde{m}_t - \left(\frac{p_t}{P_t}\right)^{\varepsilon}\right) Y_t - T_t$$

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

$$r_t - \left(\frac{Y_{t+1}}{Y_t}\right)^{\gamma} \pi_t = \frac{\varepsilon}{\phi}(\tilde{m}_t - m) + \lambda_t; \qquad m = \frac{\varepsilon - 1}{\varepsilon}. \tag{20}$$

where π_t denotes the inflation rate $\pi_t = P_t/P_{t-1}$. We exploit the fact that, given the lack of aggregate risk, the household's stochastic discount factor can be expressed as $Q_{0,t} = e^{-\int_0^t r_s ds}$ (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

$$T_t = (1 - \tilde{m}_t) Y_t - \frac{\phi}{2}\pi_t^2 Y_t. \tag{21}$$

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

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 $(\ _t + \ (\ _t))\ K_t$ where $\ _t$ is the investment rate and $\ (\ _t)$ 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.

$$W_t = \max_{\ _t;K_t} \int_0^1 \ _{0;t}\,(q_t\ _t \quad _t \quad (\ _t))\ K_t dt: \tag{22}$$

$$\text{s.t.} \quad \dot{K}_t = (\ _t \quad)\,K_t: \tag{23}$$

The optimality conditions imply (see Appendix A.3)

$$r_t = (\ _t \quad) + \frac{q_t \quad ^{00}(\ _t)\ _t}{q_t \quad 1 \quad ^{0}(\ _t)} \quad \frac{q_t\ _t \quad _t \quad (\ _t)}{q_t \quad 1 \quad ^{0}(\ _t)}.$$

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 $\ < \ 1$ 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

$$\frac{@g_t(z;a)}{@t} = \underbrace{-\frac{@}{@a}[g_t(z;a)s_t(z)a]}_{\text{Retained earnings}} \underbrace{-\frac{@}{@z}[g_t(z;a) \quad (z)] + \frac{1}{2}\frac{@^2}{@z^2}[g_t(z;a) \quad ^2(z)]}_{\text{Productivity changing randomly}}$$

$$\underbrace{g_t(z;a)}_{\text{Entrepreneurs retiring}} \quad + \underbrace{-g_t(z;\bar{a}))}_{\text{Entrepreneurs entering}} \quad ; \tag{24}$$

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

$$s_t(z) \quad \frac{1}{q_t}\Big(\underbrace{\max\{z_t' \quad R_t; 0\}}_{\text{Profit rate from operating the firm}} \quad + R_t \quad q_t\Big); \tag{25}$$

and $1= g_t(z; a=)$ is the density of new entrepreneurs entering.

Using this two-dimensional distribution we can define the one-dimensional distribution of *net-worth shares* as $! _t(z) \quad \frac{1}{A_t} \int_0^1 ag_t(z; a)da$. 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)

$$\frac{@!_t(z)}{@t} = \left[s_t(z) \quad \frac{A_t}{A_t} \quad (1 \quad) \quad ! _t(z) \right] \quad \frac{@}{@z}(z)! _t(z) + \frac{1}{2}\frac{@^2}{@z^2} \quad ^2(z)! _t(z): \tag{26}$$

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 $b_t = k_t \quad a_t$, 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 (z) be the cumulative distribution of net-worth shares, i.e. $_t(z) = \int_0^z ! _t(x)\, dx$: By combining equations (27), (28), aggregating capital used by firms (7), and solving for A_t; we obtain

$$A_t = \frac{D_t}{(1 \quad _t(z_t)) \quad 1}: \tag{29}$$

Labor market clearing implies

$$L_t = \int_0^1 l_t(z;a)dG_t(z;a): \tag{30}$$

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

$$Y_t = Z_t K_t \, L_t^{1} \quad ; \tag{31}$$

where aggregate TFP Z_t is an endogenous variable given by

$$Z_t = \quad E_{! \, t(\,)}[z \, j \, z > z_t] \quad = \quad \frac{\int_{z_t}^{1} x! \, _t(x) \, dx}{1 \qquad _t(z_t)}^{!} \quad : \tag{32}$$

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. 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. The degree to which capital is misallocated is endogenous and implies that aggregate TFP Z_t evolves over time and, importantly, depends on monetary policy. This mechanism linking aggregate TFP and monetary policy differs from the one in the endogenous growth literature, such as Benigno and Fornaro (2018) or Moran and Queralto (2018).

Factor prices are

$$w_t = (1 \qquad) m_t Z_t K_t \, L_t \quad ; \tag{33}$$

$$R_t = m \, _t Z_t K_t \quad ^{1} L_t^{1} \quad \frac{z_t}{E_{! \, t(\,)}[z \, j \, z > z_t]}: \tag{34}$$

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

$$\frac{A_t}{A_t} = \frac{1}{q_t} \quad (1 \qquad _t(z_t)) \quad m \, _t Z_t K_t \quad ^{1} L_t^{1} \qquad R_t \quad + R_t \quad q_t \quad q_t(1 \qquad) \;) \; : \tag{35}$$

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

$$di = \quad i_t \quad ^{h} + \quad (\quad _t \quad) + \quad dt; \tag{36}$$

where is the inflation target, is the sensitivity to inflation deviations and 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 Misallocation and monetary policy

3.1 The misallocation channels of monetary policy

As discussed above, aggregate TFP is endogenous and depends on the allocation of net-worth across entrepreneurs. What is the link between monetary policy and TFP in this economy? Equation (32) above shows how aggregate TFP Z_t only depends on the mass of the net-worth distribution $!_t(\,)$ above the productivity threshold z_t (the shaded area in Figure 1),

$$Z_t = \quad E_{!_t(\,)}[z \mid z > z_t] \quad : \tag{37}$$

Entrepreneurs to the left of z remain inactive and rent out their net worth to active entrepreneurs to the right of the cut-off (those in the shaded area). Equation (37) allows us to identify the two 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.

Productivity-threshold channel . The productivity-threshold channel captures the fact that, by changing factor prices, monetary policy affects the productivity threshold below which entrepreneurs remain inactive. Combining equation (10) with the definitions (9) and (17), we can express this productivity threshold as

$$z_t = \quad \frac{(q_t r_t + \quad q_t \quad q_t)}{\frac{(1 \quad)}{w_t} \quad (1 \quad)= \quad m_t^{\frac{1}{}}}:$$

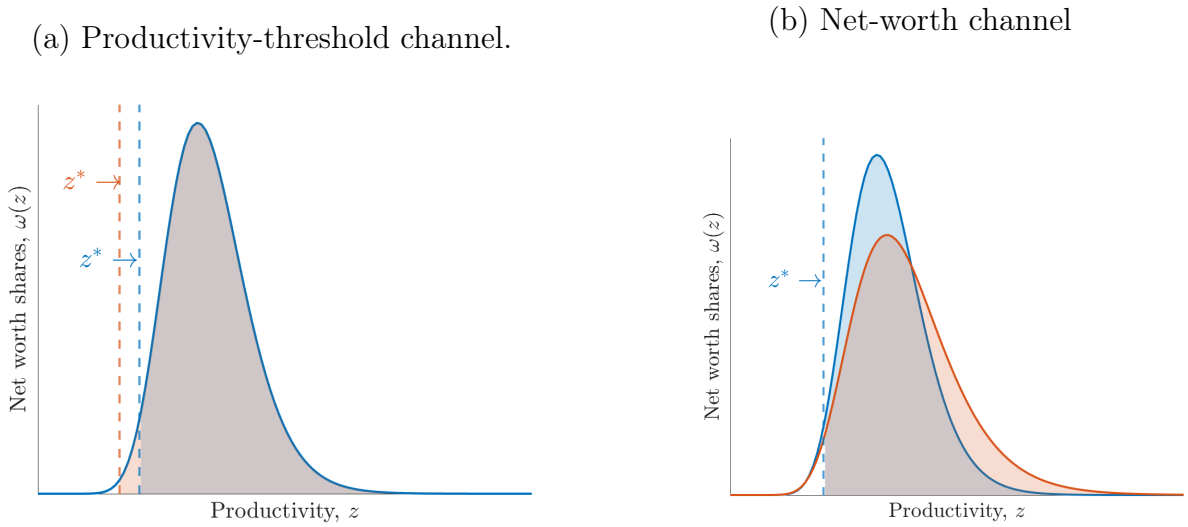

Figure 1: Net-worth distribution and productivity-threshold channels.

Notes: The gure shows the net-worth share distribution ! (z) and the productivity-threshold z (blue). The light blue area is the initial mass of active rms. 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 active rms after the change is depicted by the shaded orange area in both panels.

This equation reflects how the threshold is *directly* affected by monetary policy through changes in the real rate r_t, and *indirectly* through changes in the other prices. Holding everything else constant, a decrease in real interest rates decreases the cut-off, $\frac{@z_t}{@r_t} = \frac{q_t}{'_t} > 0$. If we plug the definition of z (equation 10) into the definition of TFP (equation 32), and take the partial derivative of TFP with respect to r_t, holding the other prices constant ($'_t = '$, $q_t = q$, $\underline{q} = 0$), we obtain

$$\frac{@Z_t}{@r_t} = \underbrace{\frac{! (z_t)}{Z_t^{\frac{1}{}} (1 \quad _t(z_t))}}_{\frac{@z_t}{@r_t}} \underbrace{E_{!_t(\cdot)}[z \mid z > z_t] \quad z_t}_{\frac{@z_t}{@r_t}} \frac{q}{'} > 0: \quad (38)$$

The derivative of TFP with respect to the interest rate is always non-negative, and it is strictly positive as long as the distribution ! (z) is non-zero for $z > z_t$. This means that, *ceteris paribus*, if interest rates decrease so does TFP. [7] The intuition is the following. A lower cost of capital makes production cheaper, but since active entrepreneurs are

[7] Note that the term $E_{!_t(\cdot)}[z \mid z > z_t] \quad z_t$ is a measure of the dispersion of productivity of active firms: the larger the difference between the average productivity of active firms and the productivity threshold, the larger the impact of a change in interest rates is.

constrained by the borrowing limit, this reduction can only stimulate investment by those entrepreneurs that would otherwise find it unprofitable to operate. The increase in the share of low-productivity entrepreneurs consequently reduces aggregate TFP. Panel (a) in Figure 1 illustrates how a reduction in the threshold increases the share of active firms by crowding in low-productivity entrepreneurs.

In general equilibrium, the response of the threshold z depends not only on the direct effect of monetary policy on the real rate r_t, but also on the *indirect effects* coming from changes in the other factor prices, namely the wage w_t, the price of capital q_t, and that of the input good m_t. These effects can amplify or mitigate the direct effect of the real rate. The impact of the different factor prices on the threshold is heterogeneous. For instance, an increase in the wage increases the threshold, $\frac{@z}{@w} = \frac{(1 \quad)}{} \frac{z_t}{w_t} > 0$; as it increases production costs, whereas an increase in the input price decreases the threshold, $\frac{@z}{@m} = \quad \frac{1}{} \frac{z_t}{m_t} < 0$: Taking into account direct and indirect effects, the threshold could shift to the right or to the left, therefore increasing or decreasing endogenous TFP *ceteris paribus*.

Net-worth distribution channel. So far we have 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. To see this, notice how aggregate TFP (eq. 32) depends on the conditional mean of the distribution of *active* firms, $E_{! \ t( )}[z \ j \ z > z_t]$. Changes in this distribution, such as shifts or changes in skewness or kurtosis, can change this conditional mean. Panel (b) of Figure 1 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. This implies that monetary policy only affects the net-worth distribution through general equilibrium effects. The intuition for this result is that all active 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

$$\tilde{\tau}_t(z) \qquad \frac{\pi_t}{q_t k_t} = \max\left\{ \frac{1}{q_t} \left(\frac{(1-\alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}} (z - z_t); 0 \right\}; \tag{39}$$

where we have employed the definition of profits π_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 π_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 $\tilde{\tau}_t(z)$ as the investment rate of a firm with productivity z in excess of the investment rate of the marginal firm with productivity z .[8] 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.

The impact of prices on the slope of the excess investment rate is heterogeneous. For instance, an increase in the price of the input good increases profits and hence the excess investment rate of active entrepreneurs

$$\frac{\partial \tilde{\tau}_t(z)}{\partial m_t} = \frac{(1-\alpha)}{w_t} \left(\frac{(1-\alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}-1} (z - z_t) > 0, \text{ for } z > z_t :$$

The increase in the excess investment rate is particularly strong for the most productive firms, $\frac{\partial^2 \tilde{\tau}_t(z)}{\partial m_t \partial z} > 0$, who hence grow relatively more than the least productive firms. An increase in wages or the price of capital, however, decreases the excess investment rate, $\frac{\partial \tilde{\tau}_t(z)}{\partial w_t}, \frac{\partial \tilde{\tau}_t(z)}{\partial q_t} < 0$, as it increases firms' costs. The reduction of the investment rate affects high-productivity firms the most $\frac{\partial^2 \tilde{\tau}_t(z)}{\partial w_t \partial z}, \frac{\partial^2 \tilde{\tau}_t(z)}{\partial q_t \partial z} < 0$.

Given how the different factor prices affect TFP through both channels in opposite directions, we cannot derive the overall effect of a monetary policy shock on TFP analytically. For this reason, in the next sections we turn to the numerical solution of the model.

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

3.2 Numerical solution and calibration

Numerical algorithm. We solve the model numerically using the 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 $!(z)$, 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 as in Boppart et al. (2018) or Auclert et al. (2019).[9] 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.

Calibration. Table 1 summarizes our calibration. The rate of time preference of the household ^{h} is 0.025, which targets an average real rate of return of 2.5 percent. The capital depreciation rate is set at 0.065, equal to the aggregate depreciation rate in NIPA. The fraction of assets of exiting entrepreneurs reinvested ( ) 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 ( ) is 0.12 which, together with , implies an average real return on equity of 11 percent, the return of the S&P500 from 2009 to 2019. The borrowing constraint parameter 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 is set at a standard value of 0.3. We assume log-utility in consumption $(= 1)$ and the inverse Frisch elasticity # is also set to 1, standard values in the literature. We normalize the

[9] 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.

Table 1: Calibration

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

constant multiplying the disutility of labor such that aggregate labor supply in steady state is equal to one.

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

$$(\ _t) \ = \ \frac{k}{2}(\ _t \quad)^2 : \tag{40}$$

Capital adjustment costs, ^{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 is set to 10, so that the steady state mark-up is $1=(1 \quad) = 0:11$. The Rotemberg cost parameter is set to 100, so that the slope of the Phillips curve is $= \ = 0.1$ as in Kaplan et al. (2018)

The Taylor rule parameters take the following values: $= 0; \quad = 1:25$ and $= 0:8$. These values are explained when dealing with the optimal policy in Section 4.2 below.

We assume that individual productivity z follows an Ornstein-Uhlenbeck process in

logs[10]

$$d\log(z) = \quad \& \log(z)dt + \quad {}_z dW_t: \tag{42}$$

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).

3.3 Response to shocks

To understand how the different channels identified above interact, we next analyze the transmission of shocks quantitatively.

Monetary policy shock . We analyze the response of the economy to a temporary expansionary monetary policy shock where the nominal interest rate decreases 20 basis points, starting from the steady state. See the blue solid lines in Figure 2.[11] 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. Furthermore, aggregate TFP *increases* (panel g). This is a consequence of both the productivity-threshold channel, as the threshold z moves up (panel h), and the net-worth distribution channel, as the slope of the excess investment rate increases (panel i).

Next, we dig into the two channels that drive the aggregate response of TFP just described. The blue solid line of panel (a) in Figure 3 shows the initial excess investment rate before the shock arrival $\tilde{}_0(z)$, which coincides with the one in steady state.[12] For productivity values below the cut-off z_0, this excess rate is zero, since entrepreneurs with such low productivity prefer to remain inactive and not invest. From z onwards, entrepreneurs operate firms, and their investment rate increases linearly in productivity, as shown in equation (39) above. The green solid line shows the excess investment rate one year after the shock arrival, $\tilde{}_1(z)$. The cut-off after one year z_1 shifts rightwards, so TFP increases through the productivity-threshold channel.

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

$$dz = \quad (z)dt + \quad (z)dW_t; \tag{41}$$

where $(z) = z \quad \& \log z + \frac{2}{2}$ and $(z) = \quad {}_z z$.

[11] We simulate the model at quarterly frequency (time period $t = 1=4$).

[12] For exposition purposes, Figure 3 plots the responses to a larger monetary policy shock (7 pp decrease in nominal interest rates). The shifts of the curves scale linearly for a smaller shock.

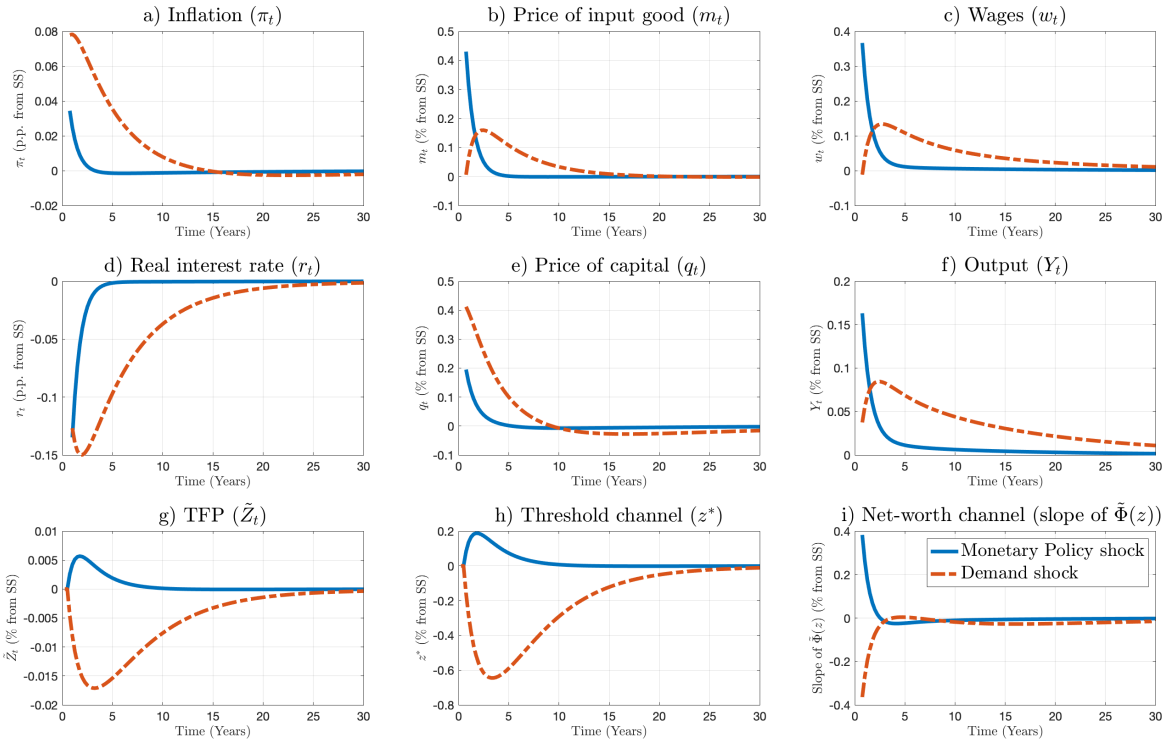

Figure 2: Impulse responses to a monetary and a demand shock.

Notes: The gure 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 20 basis points. The orange dotted line is the response of the baseline economy to a 10% decrease of the discount factor of the household, $^{hh}_{t}$, that reverts to its steady state value following an autoregressive process with yearly persistence of 0.8. In both cases the central bank follows the Taylor rule.

(a) Excess investment rate before the shock $\tilde{}_0(z)$ and after one year $\tilde{}_1(z)$.

(b) Deviations of net-worth shares after one year

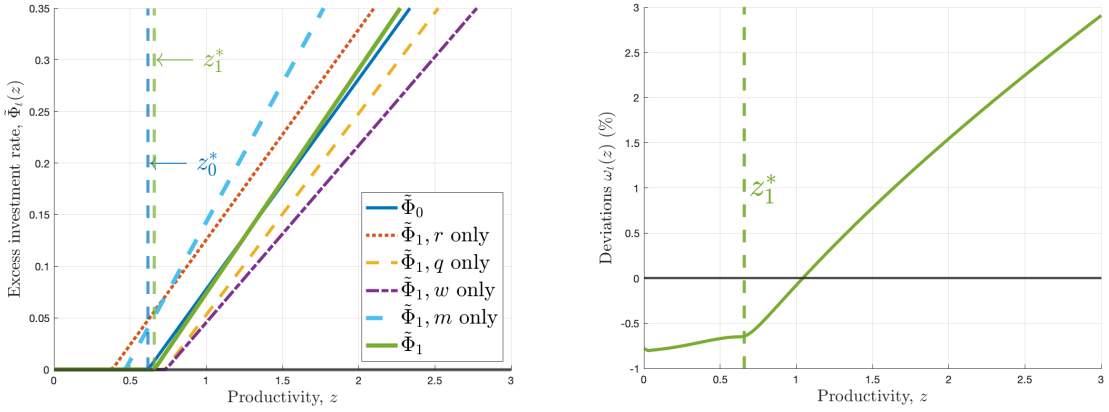

Figure 3: Channels driving the firms' response one year after the shock hits.

Notes: Panel (a): Idiosyncratic productivity is shown on the x-axis, the excess investment rate ţ z) on the y-axis. The solid blue line is the excess investment rate function before the shock arrival at time 0, and the solid green line is the same function in year 1. The rest of the lines show the excess investment rate function when only one price is changed at a time to its year 1 value, keeping the rest of the prices constant to the pre-shock, steady state value. Panel (b): deviations from steady state of the net-worth shares for each idiosyncratic productivity level z one year after the shock, i.e. $\frac{!_{t=1}(z) \quad !_0(z)}{!_0(z)}$.

Moreover, the slope of the excess investment rate $\tilde{}_1(z)$ increases one year after the shock arrival. This implies that investment increases *relatively more* the more productive a firm is. High-productivity firms thus invest more and operate at a larger scale, which improves the allocation of resources through the net-worth distribution channel. Panel (b) of Figure 3 displays the percent deviations of net-worth shares, $[!_1(z) \quad !_0(z)] =!_0(z)$, after one year. It illustrates how the distribution tilts to the right. The share of firms with productivities slightly above one increases, whereas the share of firms below that threshold declines. As a result, production now concentrates more on high-productivity firms.

Panel (a) of Figure 3 also shows a decomposition of the partial-equilibrium impact of each of the prices on the excess investment rate one year after the shock $\tilde{}_1(z)$. The red dotted line illustrates the *direct effect* of monetary policy, that is, the response to a decrease in the real rate. As discussed above in Section (3), in partial equilibrium a decrease in the real interest rate r_t increases misallocation through the productivity threshold channel, but has no effect through the net-worth distribution channel. That is, it affects the cut-off but not the slope. We thus observe a parallel shift of the excess

investment rate function to the left. The (partial equilibrium) effect is thus an increase in misallocation or, equivalently, a reduction in TFP, as shown in equation (38). This coincides with the results of Gopinath et al. (2017), who find, in a partial equilibrium setting, that a reduction in real rates increases misallocation.

In general equilibrium, the overall response of TFP also depends on the *indirect effects* coming from changes in the other factor prices already described above. The increase in the price of capital q_t counteracts the direct effect on the cut-off, which reduces misallocation through the productivity-threshold channel; but it also reduces the slope of the excess investment rate, which increases misallocation through the net-worth distribution channel (yellow dashed line in 3). Similarly, the increase in wages w_t both shifts the cut-off to the right and decreases its slope (purple dashed-dotted line). The increase in the price of the input good m_t has the exact opposite effect: It shifts the cut-off to the left and increases the slope (light blue dashed line).

As discussed above, which of these channels prevails is a quantitative question. For our particular calibration, the result (green solid line) is a counterclockwise tilt, implying a rise in the threshold z_t and an increase in the slope of the excess investment function, which shifts the net-worth distribution to the right. Expansionary monetary policy thus increases TFP. It does so through both the threshold and the net-worth distribution channel. This quantitative result is robust to alternative realistic calibrations of the model. The bottom line is that, by expanding demand through a more accommodative monetary policy stance, an expansionary monetary policy increases the share of production carried out by high-productivity firms, reducing misallocation and increasing TFP. We test this theoretical prediction in the data after analyzing the effects of a demand shock.

Demand shock . Can we then conclude that a reduction in real rates reduces misallocation? Not necessarily. This can be seen in Figure 2, where we also display the impulse responses to a demand shock that temporarily reduces real rates (orange dashed line). In particular, we consider a temporary decrease in households' discount factor ^{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 h), reflecting an increase in misallocation. Since a decrease in $^{h}_{t}$ implies that the household becomes more patient, consumption decreases and savings (in capital) 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 imply

that TFP now falls as misallocation increases through both the productivity-threshold and the net-worth distribution channel. This result is in line with the findings of Asriyan et al. (2021), who show how a decline in the real rate may produce an increase in capital misallocation through the general equilibrium effect of the price of capital in a model that shares our modeling of firm heterogeneity and financial frictions, but embeds those elements in a small open economy with fixed capital that abstracts from labor and nominal rigidities.

Summing up, the overall effect of a reduction of the real rate depends on the source of its movement.

3.4 Empirical evidence: the e ect of monetary policy shocks at rm level

In the previous subsection we showed that, according to our model, an expansionary monetary policy shock allows the most productive firms to increase their investment relatively more than low-productivity ones, which reduces misallocation and increases TFP. This result is, however, of quantitative nature: the overall effect of a monetary expansion on firm investment depends on the responses of different equilibrium prices, whose individual effects can be positive or negative. We now test empirically whether, in response to a monetary expansion, high-productivity firms indeed increase investment relatively more compared to low-productivity ones. This would provide evidence supporting the *net-worth channel* identified in the model.

To address this question 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. [13]

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

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.

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

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 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 productive firms' investment is more responsive, we estimate the following equation:

$$\log k_{j;t} = \alpha_j + \alpha_{s;t} + \beta (MRPK_{j;t-1} - E_j[MRPK_j])\varepsilon_t^{MP} + \gamma' Z_{j;t-1} + u_{j;t}. \tag{43}$$

The dependent variable $\log k_{j;t}$ 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 ε. We demean $MRPK_{j;t-1}$ by the firm average across time $E_j[MRPK_j]$ 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 ($_{s;t}$) 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: Heterogeneous responses of investment to monetary policy in MRPK

	(1)	(2)
$^{MP\ 1}_{t}$ 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 coe cient 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 nancial 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 $"^{MP}_{t}$ so that a positive shock corresponds to a decrease in interest rates. We have standardize $(MRPK_{jt\ \ 1}\ \ E_j\ [MRPK\ ])$ over the entire sample.

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, ; is easily interpretable. First, we standardize $(MRPK_{jt\ \ 1}\ \ E_j\ [MRPK\ ])$ 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 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 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 this 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 finding 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, which is the key mechanism behind the net-worth distribution channel. 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 normative question of 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:

$$\max_{\{!_t(z),s_t(z),w_t,r_t,q_t,\ _t,K_t,A_t,L_t,C_t,D_t,Z_t,\ _t,z_t,\ _t,\ _t,m_t,\dot{m}_t,i_t,Y_t,T_t\}_{t=0}^{1}} E_0 \int_0^1 e^{\ ^ht} u(C_t, L_t)dt \quad (44)$$

subject to the all the equilibrium conditions derived above and listed in Appendix A.7 and the initial conditions. The equilibrium conditions include, among others, the law of motion of the net-worth distribution (equation 26), as the central bank internalizes the impact of her decisions on it.

Numerical approach . Notice that $!_t(z)$ 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, which in this case may take the form of distribution and (social) value functions. They then solve the problem using the upwind finite-difference method of Achdou et al. (2017). The problem with this approach is that it requires deriving the FOCs by hand, which can be demanding in medium-scale models such as the one presented in this paper.

The algorithm proposed here can be seen as the mirror image of Nuño and Thomas (2016). Instead of first computing *by hand* the planner's FOCs in continuous variables and then discretizing them using finite differences, we propose to first discretize the central bank's objective and constraints (the private equilibrium conditions) using finite differences, and then to find the planner's FOCs *by symbolic differentiation*. The first step (discretization using finite differences) was already described in Section 3.2 and Appendix C. The second step (symbolic differentiation to obtain the FOCs and nonlinear solution of the dynamic system) can conveniently be executed using Dynare.[14] Relative to Nuño and Thomas (2016), this avoids the cumbersome mathematical derivations and allows us to solve for nonlinear transitional dynamics under the Ramsey policy in a few seconds, leading to the same results, as we show in the appendix. Details are provided in Appendix D.

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

4.2 Optimal Ramsey policy

We compare the dynamic Ramsey plan in our baseline heterogeneous-firm economy to that in the complete-market economy. The complete-market economy is the standard representative agent New Keynesian model 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 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). We stress the fact that the central bank's only instrument is the nominal interest rate. For simplicity, we calibrate the tax/subsidy such that it undoes the New Keynesian mark-up distortion in the steady state of both economies.

Steady state . We compute first the steady state under the optimal policy, that is, we solve the Ramsey problem and compute its steady state. It is well known that the New Keynesian economy with complete markets features zero inflation in steady state under the optimal policy. The baseline economy also features zero inflation 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.

Aggregate dynamics . 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. We assume that the central bank faces no pre-commitments, 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).

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

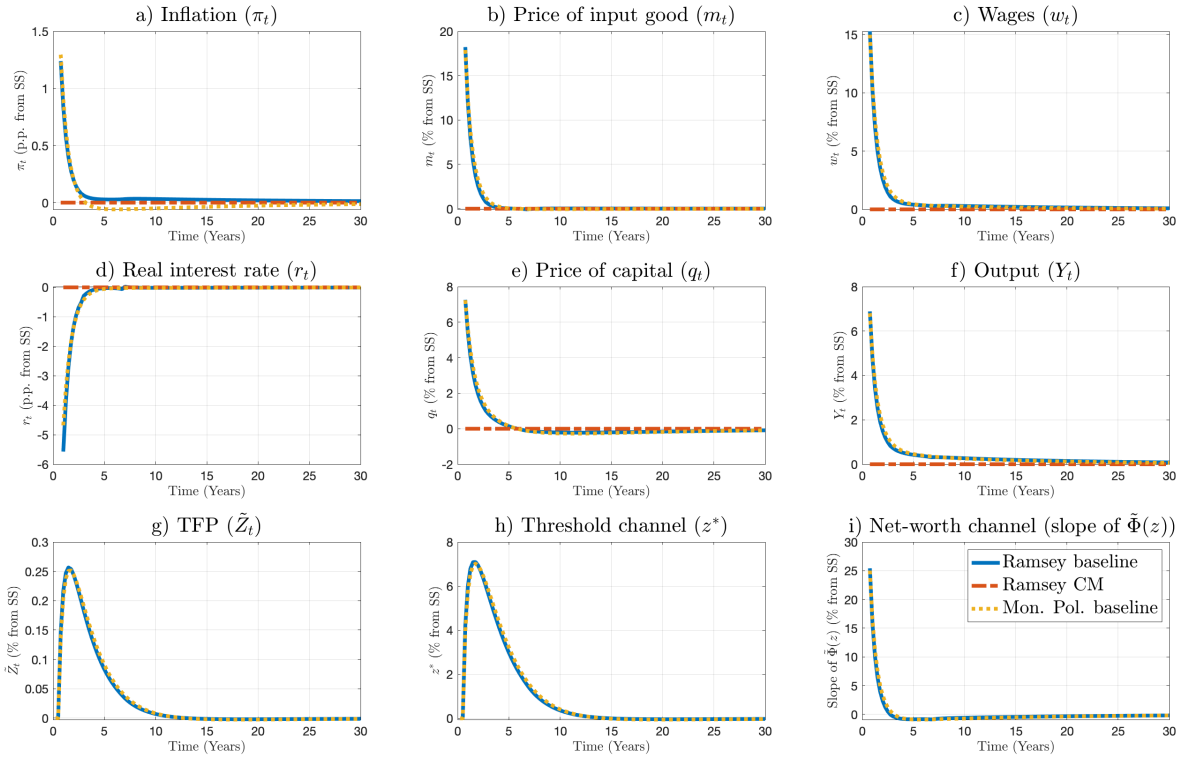

Figure 4: Time 0 optimal monetary policy.

Notes: The gure 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 $di = \quad i_t \quad {}^{h}_{t} + \quad (\quad _t \quad) + \quad dt$, with $\quad = 0:8$, and $\quad = 1:25$.

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 d). The dotted yellow line displays a monetary policy rule with the same calibration as in Section 3.3. 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 3.3: the change in factor prices increases TFP (panel g) though the productivity-threshold and net-worth channels (panels h and i). 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

Cost-push shock . Next, we analyze the optimal policy response when an unexpected mark-up shock hits the economy that was previously in its zero-inflation Ramsey 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.[15] This is a concept that only makes sense in the presence of aggregate risk. As discussed in Section 3.2, building on the argument by Boppart et al. (2018) one can however reinterpret the timeless response to MIT shocks as a first order approximation

[15]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.

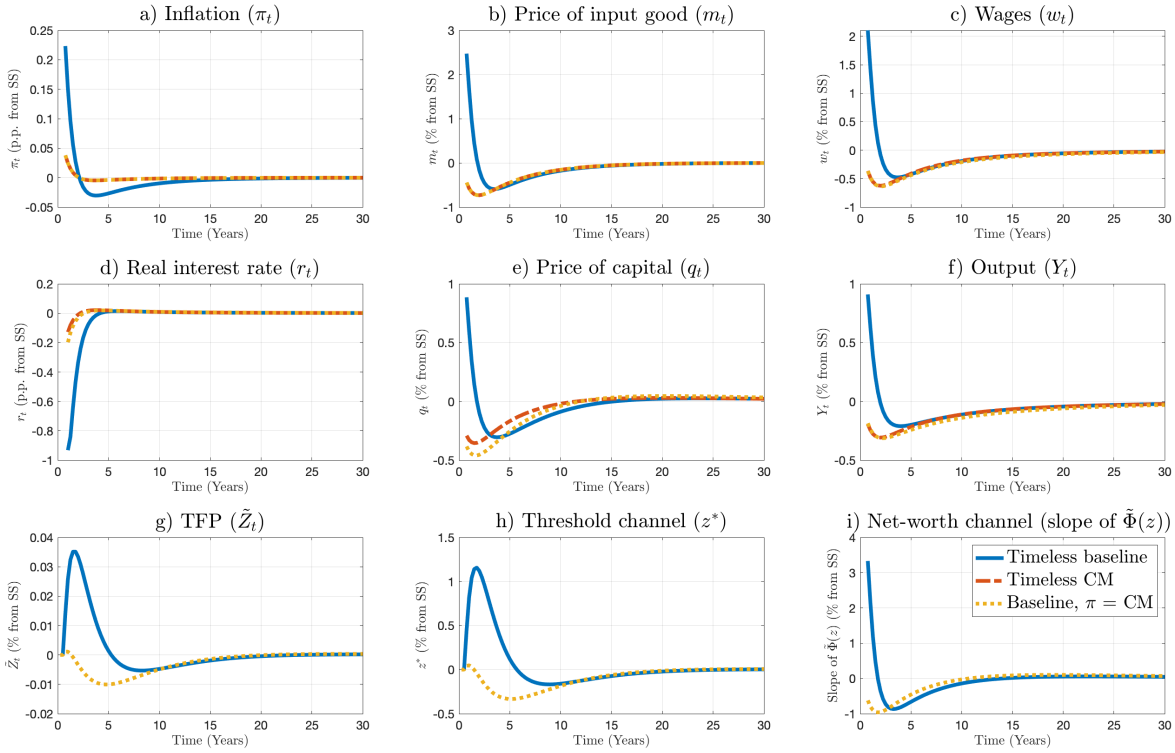

Figure 5: Optimal monetary policy response to a cost-push shock.

Notes: The gure shows the optimal response from a timeless perspective (in deviations from steady state) to a 10% decrease in the elasticity of substitution 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 in ation is that of the complete markets economy.

to the response in a model with aggregate uncertainty under the ex-ante optimal time-invariant state-contingent policy rule.

Figure 5 shows the optimal timeless response of the central bank to a cost-push shock caused by a sudden unexpected temporary decrease in the elasticity of substitution ( ) of 10% that is mean-reverting with yearly persistence of 0.8. This shock increases retailers' markup, pushing down the price of the input goods sold by heterogeneous firms. Each panel shows the response of different equilibrium variables. The dashed red line in Figure 5 shows the optimal response in the complete-market economy to this cost-push shock. This shock induces inflationary pressures due to the increase in markups (panel a). This induces a trade off between inflation and output gap stabilization. The central bank reacts optimally by driving output below its efficient level to dampen inflation (panels a and f). This is the well-known policy of *leaning against the wind* (Gali, 2008).

The optimal response of the monetary authority is, however, very different in the

baseline economy. In addition to the short-run trade-off between inflation and output gap, the central bank also influences misallocation and TFP in the medium run. This motivates the central bank to adopt a *leaning with the wind* policy. Instead of containing inflation at the cost of a fall in output, the central bank allows inflation to rise well above the optimal level in the case with complete markets (solid blue line panel a). Thus, output increases (panel f). By increasing inflation, the central bank generates a demand expansion, increasing input good prices, wages, and real rental rates (panels b and c, rental rates not shown). The net effect is to increase both the productivity-threshold (panel h) and the slope of excess investment rate (panel i). This raises TFP (panel g), aggregate capital and aggregate output (panel f).

To isolate the effects due to the different optimal policy prescriptions from those driven by the market incompleteness, we also consider a scenario where the central bank in the baseline economy targets the path for inflation of the model with complete markets (dotted yellow line). The aggregate dynamics of the economy in this case differ significantly from those in the baseline under the optimal policy and are very similar to the case with complete markets. Instead of an expansion, aggregate output falls below its steady state value (panel f), and so do capital and TFP. The difference between the case with and without complete markets is thus almost exclusively driven by the different optimal policy prescriptions.

These results illustrate that financial frictions and firm heterogeneity alter the policy prescriptions regarding the optimal response to cost push shocks. In this case, the central bank should reduce rates aggressively to expand demand, as this redistributes resources towards entrepreneurs, thus increasing TFP in the medium run.

Demand shock . Finally, we analyze the optimal response to a demand shock of the type considered in Section 3.3, i.e. a temporary fall in the household's discount factor of 20%. Figure 6 shows that the optimal response in the baseline economy (blue solid line) mimics that under complete markets (orange dashed line). It is characterized by inflation stability at its steady state value of zero (panel a). To this end, the central bank lets the real interest rate decline (panel d), while the price of capital increases significantly due to the higher demand for capital (savings) of households (panel e). This generates an output boom, which is lower than in the complete markets economy (panel f). In the baseline model with incomplete markets, the optimal response produces a fall in aggregate TFP (panel g), as it was in the case with a Taylor rule analyzed in Figure 3.3, which explains the more muted response of output compared to the case

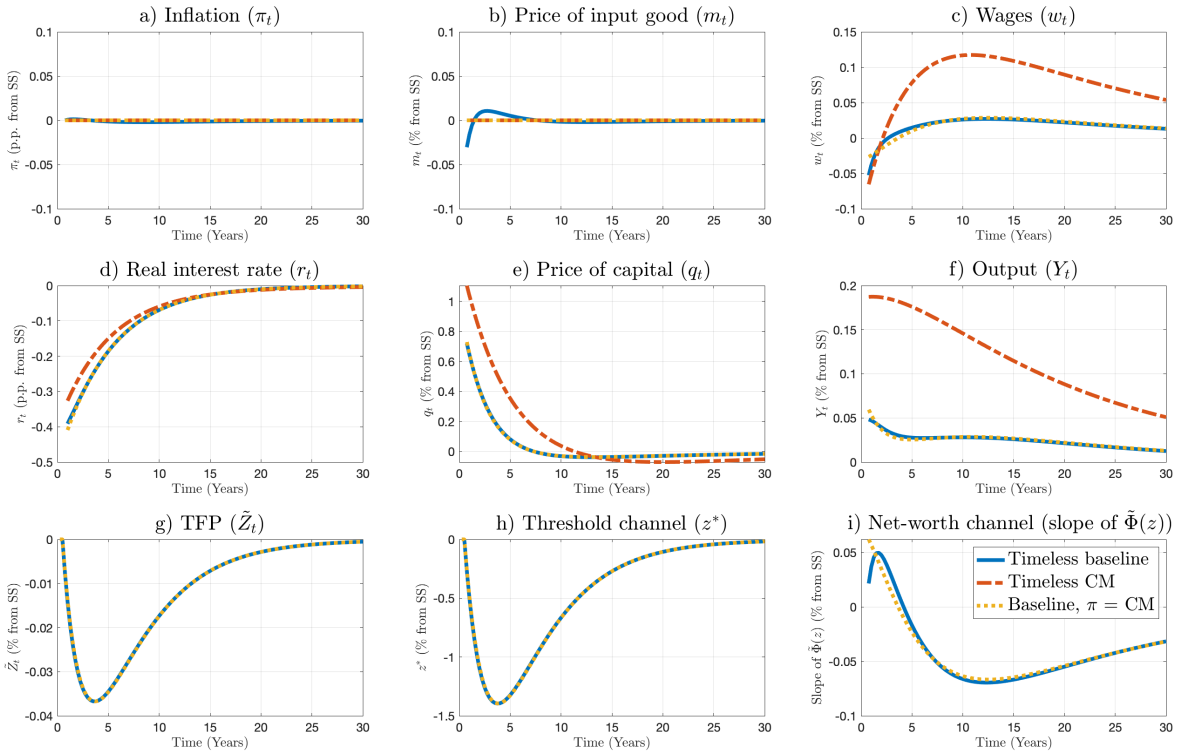

Figure 6: Optimal monetary policy response to a demand shock.

Notes: The gure 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 ^{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 in ation is that of the complete markets economy.

with complete markets.

These results point at the importance of understanding the underlying shock to which monetary policy is responding to, since the prescriptions of optimal monetary policy in the presence of misallocation might significantly differ – or not – from those of the complete markets economy.

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 profitable, and (ii) the net-worth distribution of firms. Although in partial equilibrium an expansionary monetary policy

shock increases misallocation by reducing the productivity threshold and crowding in low-productivity firms, this result can be overturned when also considering general equilibrium effects. We find that this is the case in a calibrated numerical exercise: changes in other factor prices favor high-productivity firms, allowing them 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 cost-push shock, the optimal prescription is to lean with the wind, tolerating more inflation in exchange for a boom in demand that raises TFP further down the road. This result contrasts with the case of a temporary demand shock, in which price stability is the optimal policy, just as under complete markets.

The paper also makes what we deem as a methodological contribution. It introduces a new algorithm to compute optimal policies in heterogeneous-agent models. The algorithm leverages on 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 representative agent 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. 3.2, 4.1, C.1, 18, D.1

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

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. 3.2, D.1

Albrizio, S., B. González, and D. Khametshin (2021). A tale of two margins: Monetary policy and misallocation. Technical report. 3.4

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*. 2

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, 3.3

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. 3.2, 9, 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. 16, 22

Baqaee, D., E. Farhi, and K. Sangani (2021). The supply-side effects of monetary policy. Technical report, National Bureau of Economic Research. 2

Benigno, G. and L. Fornaro (2018). Stagnation traps. *The Review of Economic Studies 85*(3), 1425–1470. 2.7

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. 2

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. 3.2, 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. 3

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. 13

Christiano, L. J., M. S. Eichenbaum, and M. Trabandt (2016). Unemployment and business cycles. *Econometrica 84*(4), 1523–1569. 3.2

David, J. M. and D. Zeke (2021). Risk-taking, capital allocation and optimal monetary policy. Technical report. 2

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

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. 3.2, 3.2

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, 3.3

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, 3.4, B.2

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

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. 1, 3.2, 3.2, 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. 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, 3, 2, 4

Moran, P. and A. Queralto (2018). Innovation, productivity, and monetary policy. *Journal of Monetary Economics 93*, 24–41. 2.7

Nakov, A. and H. Webber (2021). Micro heterogeneity, misallocation and monetary policy. Technical report. 2

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. 3.2, 3.2

Ottonello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica 88*(6), 2473–2502. 1, 3.4, 3.4, 3.4, 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. 3.2, 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. 3.2, 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. 2

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

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

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

$$\psi_t(z) - 1 = \psi_d \text{ and } \min\{\psi_d; d_t\} = 0,$$

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

$$\begin{aligned}&(r_t + \lambda)\psi_t(z)q_t = \\ &q_t + (\lambda - \max\{z_t - \psi_t, R_t; 0\} + R_t - q_t)\psi_t(z) + \mu(z)q_t\frac{\partial \psi_t}{\partial z} + \frac{\sigma^2(z)}{2}q_t\frac{\partial^2 \psi_t}{\partial z^2} + \frac{\partial(q_t\psi_t)}{\partial t}.\end{aligned} \tag{45}$$

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

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

$$s_t^a = \frac{1}{q_t}[(\lambda - \max\{z_t - \psi_t, R_t; 0\} + R_t - q_t] - \frac{R_t - q_t}{q_t}$$

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

$$E_0 a_t = E_0 a_0 e^{\int_0^t s_u^a du} > a_0 e^{\int_0^t \frac{R_s - q_s}{q_s} ds}. \tag{46}$$

The value function is then

$$
\begin{aligned}
{}_{t_0}(z)\, q_{t_0} a_{t_0} &= V_{t_0}(z; a_{t_0}) \\
&= E_{t_0} \int_0^1 e^{\int_{t_0}^t (r_s + \quad) ds} (d_t + \quad q_t a_t)\, dt \\
& \qquad\qquad \int_0^1 e^{\int_{t_0}^t \left(\frac{R_s \quad q_s}{q_s} + \frac{\dot{q}_s}{q_s} \right) ds} \\
E_{t_0} \int_0^1 e^{\int_{t_0}^t (r_s + \quad) ds} q_t a_t dt &= E_{t_0} \int_0^1 e^{\qquad\qquad\qquad} q_t a_t dt \\
&= E_{t_0} \int_0^1 e^{\int_{t_0}^t \frac{R_s \quad q_s}{q_s} + \quad ds \ \log \frac{q_t}{q_{t_0}}} q_t a_t dt = E_{t_0} \int_0^1 e^{\int_{t_0}^t \frac{R_s \quad q_s}{q_s} + \quad ds} q_{t_0} a_t dt \\
&> E_{t_0} \int_0^1 e^{\int_{t_0}^t \frac{R_s \quad q_s}{q_s} + \quad ds} q_{t_0} a_{t_0} e^{\int_0^t \frac{R_s \quad q_s}{q_s} ds} dt = \int_0^1 e^{\quad t} q_{t_0} a_{t_0} dt = q_{t_0} a_{t_0};
\end{aligned}
$$

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 ${}_{t_0}(z) > 1$ 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

$$
r_t V_t^r(p) = \max \quad \frac{p \quad P_t^y (1 \quad)}{P_t} \qquad \frac{p}{P_t} \quad " \quad Y_t \quad \frac{}{2} \quad {}^2 Y_t + \quad p \frac{@\dot{V}^r}{@p} + \frac{@\dot{V}^r}{@t};
$$

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

$$
\begin{aligned}
Y_t &= p \frac{@\dot{V}^r}{@p}; \\
(r \quad) \frac{@\dot{V}^r}{@p} &= \frac{p}{P_t} \quad " \quad \frac{Y_t}{P_t} \quad " \quad \frac{p \quad P_t^y (1 \quad)}{P_t} \qquad \frac{p}{P_t} \quad " \quad {}^1 \frac{Y_t}{P_t} + \quad p \frac{@^2 V^r}{@p^2} + \frac{@^2 V^r}{@t @p};
\end{aligned}
$$

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

$$\frac{\partial V^r}{\partial p} = -\frac{Y_t}{p}; \tag{47}$$
$$(r -) \frac{\partial V^r}{\partial p} = -\frac{Y_t}{p} - \frac{p - P_t^y(1 -)}{p} \frac{Y_t}{p} + p \frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p}.$$

Deriving (47) with respect to time gives

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

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

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

Finally, rearranging we obtain the New Keynesian Phillips curve

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

A.3 Capital producers' problem

The problem of the capital producer is

$$W_t = \max_{t; K_t} E_0 \int_0^1 e^{-\int_0^t r_s ds} (q_t - t - (t)) K_t dt. \tag{48}$$

$$\dot{K}_t = (t -) K_t; \tag{49}$$

We construct the Hamiltonian

$$H = (q_t - t - (t)) K_t + t (t -) K_t$$

with first-order conditions

$$(q_t - 1 - {}^0(t)) + t = 0 \tag{50}$$

$$(q_t \quad _t \quad _t \quad (\ _t)) + \quad _t(\ _t \quad) = r_t \quad _t \quad -_t \tag{51}$$

Taking the time derivative of equation (50)

$$-_t = \quad (q_t \quad ^{00}(\ _t)\ _t)$$

which, combined with (51), yields

$$(q_t \quad _t \quad _t \quad (\ _t)) \quad (q_t \quad 1 \quad ^0(\ _t))(\ _t \quad r_t) = (q_t \quad ^{00}(\ _t)\ _t)$$

Rearranging we get

$$r_t = (\ _t \quad) + \frac{q_t \quad ^{00}(\ _t)\ _t}{q_t \quad 1 \quad ^0(\ _t)} \quad \frac{q_t \quad _t \quad _t \quad (\ _t)}{q_t \quad 1 \quad ^0(\ _t)}:$$

A.4 Household's problem

We can rewrite the household's problem as

$$W_t = \max_{C_t;L_t;D_t;B_t^N;S_t^N} E_0 \int_0^1 e^{\ _t t} \left(\frac{C_t^1}{1} \quad \frac{L_t^{1+\#}}{1+\#} \right)^{!} dt: \tag{52}$$

$$\text{s:t:} \quad D_t = (R_t \quad q_t) D_t + w_t L_t \quad C_t \quad S_t^N + \quad _t = q_t; \tag{53}$$

$$B_t^N = S_t^N + (i_t \quad _t) B_t^N; \tag{54}$$

where S_t^N is the investment into nominal bonds.

The Hamiltonian is

$$H = \frac{C_t^1}{1} \quad \frac{L_t^{1+\#}}{1+\#}^{!}$$

$$+ \%_t \quad (R_t \quad q_t) D_t + w_t L_t \quad C_t \quad S_t^N + (q_t \quad _t \quad _t \quad (\ _t)) K_t + \quad _t = q_t + \quad _t S_t^N + (i_t \quad _t) B_t^N$$

The first order conditions are

$$C_t \quad \%_t = q_t = 0 \tag{55}$$

$$L_t^{\#} + \%_p w_t = q \quad (56)$$

$$\%_p = q + \quad _t = 0 \quad (57)$$

$$\%_p = \quad _t^h \%_p \quad \%_p (R_t \quad q_t) = q \quad (58)$$

$$\_t = \quad _t^h \quad _t \quad \quad _t [(i_t \quad \quad _t)] \quad (59)$$

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

$$w_t = \frac{L_t^{\#}}{C_t};$$

(55) can be rewritten as

$$\%_p = C_t^{-} \quad q_t$$

Now take derivative with respect to time

$$\%_p = \quad C_t^{-} \quad ^{-1} C_t q_t + C_t^{-} \quad q_t$$

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

$$\frac{C_t}{C_t} = \frac{\frac{R_t \quad q_t + q_t}{q_t} \quad _t^h}{}$$

(57) can be rewritten as

$$_t = \%_p = q$$

Now take derivative with respect to time

$$\_t = \frac{\%_p q_t \quad \%_p q}{q_t^2}$$

Use these two expressions and the definition of $\%_p$ in (59) to get the second Euler equation

$$\frac{C_t}{C_t} = (i_t \quad _t) \quad _t^h$$

Combining the two Euler equations, we get the Fisher equation

$$\frac{R_t \quad q_t + g_t}{q_t} = (i_t \quad _t)$$

Finally using the definition of $r_t \quad \frac{R_t \quad q_t + g_t}{q_t}$ 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

$$\frac{@g(z;a)}{@t} = \quad \frac{@}{@a}[g_t(z;a)s_t(z)a] \quad \frac{@}{@z}[g_t(z;a) \quad (z)] + \frac{1}{2}\frac{@^2}{@z^2}[g_t(z;a) \quad ^2(z)] \quad g_t(z;a) + = g_{\ t}(z;a=); \tag{60}$$

where $1= g_t(z;a=)$ is the distribution of entry firms.

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

$$\frac{@!_t(z)}{@t} = \quad \frac{A_t}{A_t^2} \int_0^1 ag_t(z;a)da + \frac{1}{A_t} \int_0^1 a\frac{@g(z;a)}{@t}da: \tag{61}$$

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

$$\begin{aligned}\frac{@!_t(z)}{@t} = \quad & \frac{A_t}{A_t^2} \int_0^1 ag_t(z;a)da + \frac{1}{A_t} \int_0^1 a \quad \frac{@}{@a}[g_t(z;a)s_t(z)a] \quad da \\ & \frac{@}{@z}(z)\frac{1}{A_t} \int_0^1 ag_t(z;a)da + \frac{1}{2}\frac{@^2}{@z^2} \ ^2(z)\frac{1}{A_t} \int_0^1 ag_t(z;a)da \\ & \frac{1}{A_t} \int_0^1 ag_t(z;a)da + \frac{1}{A_t} \int_0^1 a= g_{\ t}(z;a=)da:\end{aligned}$$

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,

$$\frac{@!_t(z)}{@t} = \left[s_t(z) \quad \frac{A_t}{A_t} \quad (1 \quad) \right] ! _t(z) \quad \frac{@}{@z}(z)! _t(z) + \frac{1}{2}\frac{@^2}{@z^2} \ ^2(z)! _t(z): \tag{62}$$

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

$$0 = (s(z) \quad (1 \quad) \ ) ! (z) \quad \frac{@}{@z}(z)! (z) + \frac{1}{2}\frac{@^2}{@z^2} \ ^2(z)! (z): \tag{63}$$

Remember that $s_t^a(z_t; a_t; c_t) = \frac{1}{a_t}[\ _t(z_t; a_t) + (R_t \quad q_t)a_t]$, since entrepreneurs distribute zero dividends while active.

A.6 Market clearing and aggregation

Define the cumulative function of net-worth shares as

$$_t(z) = \int_0^z ! _t(z)dz: \tag{64}$$

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

$$K_t = \int k_t(z; a)dG_t(z; a) = \int_{z_t}^1 a \frac{1}{A_t} g_t(z; a)dadzA_t = (1 \quad (z_t))A_t: \tag{65}$$

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

$$A_t = \frac{D_t}{(1 \quad (z_t)) \quad 1}; \tag{66}$$

Labor market clearing implies

$$L_t = \int_0^1 l_t(z; a)dG_t(z; a): \tag{67}$$

Define the following auxiliary variable,

$$X_t \int_{z_t}^1 z! _t(z)dz = E[z \mid z > z_t](1 \quad (z_t)): \tag{68}$$

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

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

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

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

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

$$Z_t = (\varphi X_t)^{\alpha}: \tag{71}$$

Aggregate profits of retailers are given by

$$\Pi_t^{Agg} = \int \max\{z_t \varphi_t - R_t; 0\} a_t dG_t(z;a) = [\varphi_t X_t - R_t(1 - \Gamma(z_t))] A_t: \tag{72}$$

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

$$Z_t = \frac{X_t}{(1 - \Gamma(z_t))^{\alpha}} = (E[z \mid z > z_t])^{\alpha}: \tag{74}$$

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

$$\begin{aligned} A_{t+1} &= \int a dG(z;a;t) \\ &= \int_0^1 \frac{1}{q_t}\left(\max\{z_t \varphi_t - R_t; 0\} + R_t - q_t - q_t(1-\delta)\right) a_t dG(z;a); \end{aligned}$$

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

$$\frac{A_{t+1}}{A_t} = \frac{1}{q_t}\left(\varphi_t X_t - R_t(1 - \Gamma(z_t)) + R_t - q_t - q_t(1-\delta)\right): \tag{75}$$

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

$$\frac{A_{t+1}}{A_t} = \frac{1}{q_t}\left(m_t Z_t A_t^{\ 1} L_t^{1} \quad R_t \ (1 \quad (z_t)) + R_t \quad q_t \quad q_t(1 \quad)\right): \tag{76}$$

Finally, we can obtain factor prices

$$w_t = (1 \quad) m_t Z_t A_t \ L_t \tag{77}$$

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

where wages come from substituting the definition of $'_t$ 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 :

$$r_t = \frac{1}{q_t} \quad m_t Z_t A_t^{\ 1} L_t^{1} \ \frac{z_t}{X_t} \quad + \frac{q_{t-1}}{q_t} \tag{79}$$

We can easily get these equations in terms of capital instead of net worth by simply using equation (65), i.e. $A_t = \frac{K_t}{(1 \quad (z_t))}$, and using that $E[z \mid z > z_t] = \frac{X_t}{(1 \quad (z_t))} = \frac{\int_{z_t}^{1} z \, _t(z)dz}{(1 \quad (z_t))}$ (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 $\{ ! (z); s(z); w; r; q; '; K; A; L; C; D; Z; \quad E[z \mid z > z_t]; \quad ; z \ ; \ ; \ ; m; \ m; i; Y; T \}$. Remember that $(z) = z \quad \& \log z + \frac{^2}{2}$ and $(z) = \ _z z$, 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.

$$\frac{@!_t(z)}{@t} = s_t(z) \quad (1 \quad) \quad \frac{\dot{A}_t}{A_t} \quad !_t(z) \quad \frac{@}{@z}[\ (z)!_t(z)] + \frac{1}{2}\frac{@^2}{@z^2} \quad ^2(z)!_t(z)$$

$$s_t(z) = \frac{1}{q_t}(\ \max f z_t' \ _t \quad R_t; 0g + R_t \quad q_t)$$

$$_t(z \ ) = \int_0^z !_t(z)\,dz$$

$$' \ _t = \quad \frac{(1 \quad)}{w_t}^{(1 \quad)=} \quad m_t^{\frac{1}{}}$$

$$\dot{m}_t = m_t(1 \quad)$$

$$w_t = (1 \quad)m_t\tilde{Z}_t K_t \ L_t$$

$$R_t = \ m \ _t\tilde{Z}_t K_t \ ^1 L_t^{1} \quad \frac{z_t}{E[z \mid z > z_t]}$$

$$\frac{\dot{A}_t}{A_t} = \frac{1}{q_t} \quad (1 \quad (\ z_t)) \quad m \ _t Z_t K_t \ ^1 L_t^{1} \quad R_t \ + R_t \quad q_t \quad q_t(1 \quad) \ )$$

$$K_t = A_t + D_t$$

$$\dot{K}_t = (\ _t \quad)K_t$$

$$A_t = \frac{D_t}{(1 \quad (\ z_t)) \quad 1}$$

$$\tilde{Z}_t = (\ E[z \mid z > z_t])$$

$$E[z \mid z > z_t] = \frac{\int_{z_t}^{1} z!_t(z)dz}{(1 \quad (\ z_t))}$$

$$\frac{\dot{C}_t}{C_t} = \frac{r_t \quad \ ^h_t}{}$$

$$w_t = \frac{L \ ^{\#}_t}{C_t}$$

$$\dot{D}_t = [(R_t \quad q_t) D_t + w_t L_t \quad C_t + T_t]=q_t$$

$$r_t = i_t \quad _t$$

$$r_t = \frac{R_t \quad q_t + \dot{q}_t}{q_t}$$

$$(q_t \quad 1 \quad ^0(\ _t))(r_t \quad (\ _t \quad)) = \dot{q}_t \quad ^{00}(\ _t) \perp \quad (q_t \ _t \quad _t \quad (\ _t))$$

$$r_t \quad \frac{\dot{Y}_t}{Y_t} \quad _t = \ -(\dot{m}_t \quad m \ ) + \_t; \quad m \ = \frac{"}{"} \quad \frac{1}{}$$

$$Y_t = Z_t K_t \ L_t^{1}$$

$$T_t = (1 \quad m_t) Y_t \quad \frac{}{2} \ ^2_t Y_t + (1 \quad) A_t + \ _t q_t \quad _t \quad \frac{k}{2}(\ _t \quad)^2 \ K_t$$

$$di = \quad i_t \quad \ ^h_t + \quad (\ _t \quad) + \quad dt:$$

A.8 The net-worth channel of monetary policy

TFP is given by equation (32)

$$Z_t = \left(\int_{z_t}^{1} z\, !_t(z)\, dz \right)^{!} \Big/ \int_{z_t}^{1} !_t(z)\, dz \quad :$$

As the distribution $!_t(z)$ 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

$$\frac{1}{Z_t}\frac{dZ_t}{dt}_{z} = \frac{d \log Z_t}{dt}_{z} = \left[\frac{d}{dt} \log \int_{z_t}^{1} z\, !_t(z)\, dz \right]^{!} - \left[\frac{d}{dt} \log \int_{z_t}^{1} !_t(z)\, dz \right]^{!\#}_{z}$$
$$= \frac{\int_{z_t}^{1} z\, !\__t(z)\, dz}{\int_{z_t}^{1} z\, !_t(z)\, dz} - \frac{\int_{z_t}^{1} !\__t(z)\, dz}{\int_{z_t}^{1} !_t(z)\, dz}.$$

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

$$\frac{\partial}{\partial x_t}\frac{d \log Z_t}{dt}_{z} = \frac{\int_{z_t}^{1} z \frac{\partial !\__t(z)}{\partial x_t} dz}{\int_{z_t}^{1} z\, !_t(z)\, dz} - \frac{\int_{z_t}^{1} \frac{\partial !\__t(z)}{\partial x_t} dz}{\int_{z_t}^{1} !_t(z)\, dz},$$

where

$$\frac{\partial !\__t(z)}{\partial x_t}_{z} = \frac{\partial}{\partial x_t}\left(\tilde{}_t(z) + \tilde{}_t \right)^{(1 \quad)=} !_t(z);$$
$$\tilde{}_t(z) = \max\left\{ \frac{(1 \quad)}{q_t} \left(\frac{}{w_t} \right)^{(1 \quad)=} m_t^{\frac{1}{}}(z \quad z);\, 0 \right\};$$
$$\tilde{}_t = \frac{R_t \quad q_t}{q_t} \frac{A_t}{} (1 \quad) = \frac{(1 \quad {}_t(z))\ m\ {}_t Z_t K_t^{\ 1} L_t^{1} \quad R_t}{q_t}.$$

Then we have:

$$\frac{\partial}{\partial x_t}\frac{d \log Z_t}{dt}_{z} = \frac{\int_{z_t}^{1} z \frac{\partial !(z)}{\partial x_t} !_t(z)\, dz}{\int_{z_t}^{1} z\, !_t(z)\, dz} - \frac{\int_{z_t}^{1} \frac{\partial !(z)}{\partial x_t} !_t(z)\, dz}{\int_{z_t}^{1} !_t(z)\, dz}^{!}$$
$$+ \underbrace{\frac{\partial \tilde{}_t}{\partial x_t}}_{|} \underbrace{\left[\frac{\int_{z_t}^{1} z\, !_t(z)\, dz}{\int_{z_t}^{1} z\, !_t(z)\, dz} - \frac{\int_{z_t}^{1} !_t(z)\, dz}{\int_{z_t}^{1} !_t(z)\, dz} \right]}_{\{z\ 0} \quad :$$

This expression shows how only the excess investment rate $\tilde{\tau}(z)$ 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 $\frac{\partial \tilde{\tau}_t(z)}{\partial r_t}_z = 0$; and $\frac{\partial \tilde{\tau}_t}{\partial R_t}_z = (1 - \tau_t(z))$, thus

$$\frac{\partial}{\partial r_t}\frac{d\log Z_t}{dt}_z = 0:$$

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 $\frac{\partial \tilde{\tau}_t(z)}{\partial w_t}_z = \frac{(1-\alpha)}{q_t w_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}(z - z) < 0$; and $\frac{\partial \tilde{\tau}_t}{\partial w_t}_z = 0$, thus

$$\frac{\partial}{\partial w_t}\frac{d\log Z_t}{dt}_z = \frac{(1-\alpha)}{q_t w_t}\left(\frac{(1-\alpha)}{w_t}\right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}}\left(\frac{\int_{z_t}^1 (z-z)\, z\omega_t(z)}{\int_{z_t}^1 z\omega_t(z)\,dz} - \left(\frac{\int_{z_t}^1 (z-z)\,\omega_t(z)}{\int_{z_t}^1 \omega_t(z)\,dz}\right)^{\alpha}\right):$$

To uncover the sign, we analyze the term

$$\frac{\int_z^1 (z-z)\, z\omega_t(z)}{\int_{z_t}^1 z\omega_t(z)\,dz} - \left(\frac{\int_z^1 (z-z)\,\omega_t(z)}{\int_{z_t}^1 \omega_t(z)\,dz}\right)^{\alpha} = \frac{\int_z^1 z^2\omega_t(z)}{\int_{z_t}^1 z\omega_t(z)\,dz} - \left(\frac{\int_z^1 z\omega_t(z)}{\int_{z_t}^1 \omega_t(z)\,dz}\right)^{\alpha}:$$

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

$$l(z) = \frac{\vdash_t(z)}{\omega_t(z)} = z\frac{\int_z^1 z\omega_t(z)\,dz}{\int_z^1 \omega_t(z)\,dz}$$

is non decreasing. This implies that function $\vdash_t(z)$ dominates $\omega_t(z)$ first-order stochastically. Hence

$$\frac{\int_z^1 z\omega_t(z)}{\int_{z_t}^1 \omega_t(z)\,dz} = E_{\omega_t(z)}[z] = \int_z^1 z\omega_t(z)zdz < \int_{z_t}^1 z\vdash_t(z)\,dz = E_{\vdash_t(z)}[z] = \frac{\int_z^1 z^2\omega_t(z)}{\int_{z_t}^1 z\omega_t(z)\,dz}:$$

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

$$\frac{\partial \tilde{\tau}_t(z)}{\partial w_t}_z < 0) \frac{\partial}{\partial w_t}\frac{d\log Z_t}{dt}_z < 0:$$

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 $\tilde{}_t(z)$. Take, for instance, input prices. We have $\frac{@\tilde{}_t(z)}{@m_t}_z = \frac{}{q_t}\frac{(1 \quad)}{w_t}^{(1 \quad)=} m_t^{\frac{1}{}}(z \quad z) > 0$; and $\frac{@\tilde{}_t}{@m_t}_z = \frac{(1 \quad {}_t(z)) Z_t K_t^{\ 1} L_t^{1}}{q_t}$, thus

$$\frac{@}{@m_t}\frac{d\log Z_t}{dt}_z = \frac{}{q_t}\frac{(1 \quad)}{w_t}^{(1 \quad)=} m_t^{\frac{1}{}} \frac{\int_z^1 (z \quad z) z\,!_t(z)}{\int_{z_t}^1 z\,!_t(z)\,dz} \left(\frac{\int_z^1 (z \quad z)\,!_t(z)}{\int_{z_t}^1 !_t(z)\,dz}\right)^{!} > 0:$$

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 $Z_t = (z_t^{max})$ .In contrast, in the baseline model with incomplete markets, entrepreneurs' firms can only use capital up to a multiple of their net worth , i.e. $a_t \quad k_t$. 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 active firms, $Z = (E[z \mid z > z_t])$. 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 $\{w; r; q; '; K; L; C; D; Z; \ ; \ ; m; \quad m; i; Y; T\}$:

$$\varphi_t = \left(\frac{(1-\theta)}{w_t}\right)^{(1-\theta)/\theta} m_t^{\frac{1}{\theta}}$$

$$\tilde{m}_t = m_t(1-\theta)$$

$$w_t = (1-\theta)m_t Z_t K_t^{\theta} L_t^{-\theta}$$

$$R_t = \theta m_t Z_t K_t^{\theta-1} L_t^{1-\theta}$$

$$K_t = D_t$$

$$K_{t+1} = (1-\delta_t)K_t$$

$$Z_t = \Gamma(\delta_t)$$

$$\frac{C_{t+1}}{C_t} = \frac{r_t - \rho_t^h}{\gamma}$$

$$w_t = \frac{L_t^{\#}}{C_t}$$

$$D_{t+1} = [(R_t - q_t)D_t + w_t L_t - C_t + T_t]/q_t$$

$$r_t = i_t - \pi_t$$

$$r_t = \frac{R_t - q_t + q_{t+1}}{q_t}$$

$$(q_t - 1 - \Gamma'(\delta_t))(r_t - (\delta_t - \delta)) = q_t \Gamma''(\delta_t)\dot{\delta}_t - (q_{t+1} - \pi_t - (\delta_t - \delta))$$

$$r_t - \frac{\dot{Y}_t}{Y_t} - \pi_t = -\varepsilon(\tilde{m}_t - m) + \varepsilon_t; \quad m = \frac{\varepsilon - 1}{\varepsilon}$$

$$Y_t = Z_t K_t^{\theta} L_t^{1-\theta}$$

$$T_t = (1-\theta)m_t Y_t - \frac{\varepsilon}{2}\theta_t^2 Y_t + \theta_t q_t \pi_t - \frac{k}{2}(\delta_t - \delta)^2 K_t$$

$$di = (i_t - \rho_t^h + \pi(\delta_t - \delta) + \varepsilon)dt.$$

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 3.4. 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 $!_a(m)$, and increasing weights in the last quarter of the previous year $!_b(m)$, i.e.

$$\varepsilon_t^{MP} = \sum_{m \in t} !_a(m)\varepsilon_m^{MP} + \sum_{m \in q4_{t-1}} !_b(m)\varepsilon_m^{MP} :$$

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.

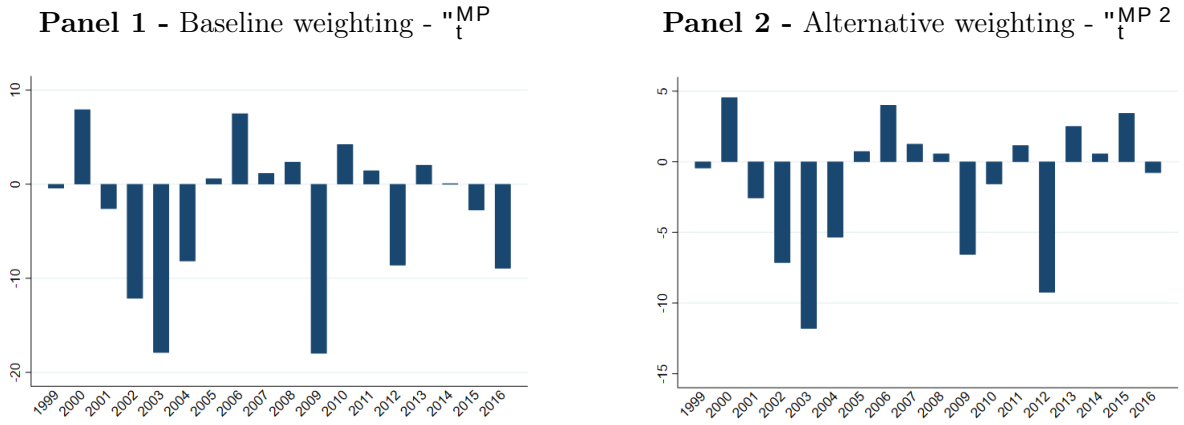

Figure 7: Monetary policy shocks at annual frequency.

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.

$$\log k_{j;t} = {}_{j} + {}_{s;t} + (MRPK_{j;t\ 1} - E_j[MRPK_j])"_t^{MP} + {}^{0}Z_{j;t\ 1} + u_{j;t}:$$

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, $"_t^{MP\ 2}$. Results are still significant and of slightly larger

magnitude. Finally, Columns (7) and (8) show the results using the baseline monetary policy shock $"^{MP}$, 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)
$^{MP}_{t}$ 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)						
$^{MP\ 2}_{t}$ x MRPK $_{t\ 1}$					0.166	0.345		
					(0.10)	(0.10)		
$^{MP}_{t}$ 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 speci cations of the estimation in the main text (Section 3.4). 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 nancial 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.[16] 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 di erence 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 $! _t (z)$ on a finite grid $z \ 2 \ f \ z_1; :::; z_J g$, $t \ 2 \ f \ t_1; :::; t_N g$ with steps z and time steps t. We use the notation $! _j^n := ! _{n \ t}(z_j); \ j \ = 1; :::; J; \ n = 0; ::; N$: The KF equation is then approximated as

$$\frac{! _j^n \quad ! _j^{n \ 1}}{t} = s_n(z_j) \quad \frac{A_n}{A_n} \quad (1 \qquad) \quad ! \ _n(z_j) \quad \frac{! _j^n \ (z_j) \quad ! _j^{n \ 1} \ (z_j \ 1)}{z} + \frac{! _{j+1}^n e^2(z_{j+1}) + ! _j^{n \ 1} e^2(z_j \ 1) \quad 2! _j^n e^2(z_j)}{2(\ z)^2};$$

[16] 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

$$\frac{\lambda_j^n - \lambda_j^{n-1}}{t} = \left[s_n(z_j) \frac{A_n}{A_n} (1 - \quad) - \frac{(z_j)}{z} - \frac{e^2(z_j)}{(z)^2} \right] \lambda_n(z_j)$$
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{z_j^n}$$
$$+ \underbrace{\frac{(z_{j-1})}{z} + \frac{e^2(z_{j-1})}{2(z)^2}}_{\theta_{j-1}} \lambda_{j-1}^n + \underbrace{\frac{e^2(z_{j+1})}{2(z)^2}}_{z_{j+1}^n} \lambda_{j+1}^n :$$

The boundary conditions are the ones associated with a reflected process z at the boundaries:[17]

$$\frac{\lambda_1^n - \lambda_1^{n-1}}{t} = (\quad_1^n + \quad_1^n)\lambda_n(z_1) + \quad_2^n \lambda_{j+1}^n ;$$
$$\frac{\lambda_J^n - \lambda_J^{n-1}}{t} = (\quad_J^n + \theta_J)\lambda_n(z_J) + \theta_{J-1}\lambda_j^n{}_1 :$$

If we define matrix

$$B^n = \begin{bmatrix} \quad_1^n + \quad_1^n & \quad_2^n & 0 & 0 & \cdots & 0 & 0 & 0 \\ \theta_1 & \quad_2^n & \quad_3^n & 0 & \cdots & 0 & 0 & 0 \\ 0 & \theta_2 & \quad_3^n & \quad_4^n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \theta_{J-2} & \quad_{J-1}^n & \quad_J^n \\ 0 & 0 & 0 & 0 & \cdots & 0 & \theta_{J-1} & \quad_J^n + \theta_J \end{bmatrix};$$

then we can express the KF equation as

$$\frac{\lambda^n - \lambda^{n-1}}{t} = B^{n-1}\lambda^n;$$

or

$$\lambda^n = \left[I \quad tB^{n-1} \right]^{-1} \lambda^{n-1}; \tag{80}$$

where $\lambda^n = \begin{bmatrix} \lambda_1^n & \lambda_2^n & \cdots & \lambda_{J-1}^n & \lambda_J^n \end{bmatrix}^T$; and I is the identity matrix of dimension J:

[17] It is easy to check that this formulation preserves the fact that matrix 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.[18] 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 $!$ grid – no matter whether we add them where most of the mass of $!(z)$ is located or in the range in which z_t moves.

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

$$\frac{!_j^n - !_j^{n-1}}{t} = s_n(z) - (1 -) \frac{A_n}{A_n} ! \, _t(z_j) \frac{(z_{j+1})! \, _t(z_{j+1}) - (z_{j-1})! \, _t(z_{j-1})}{z_{j-1;j+1}}$$
$$+ \frac{1}{2} \frac{z_{j-1;j}^{\,2}(z_{j+1})! \, _t(z_{j+1}) + z_{j;j+1}^{\,2}(z_{j-1})! \, _t(z_{j-1}) - z_{j-1;j+1}^{\,2}(z_j)! \, _t(z_j)}{\frac{1}{2}(z_{j-1;j+1}) z_{j;j+1} z_{j-1;j}}$$

[18] 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

$$\frac{!_j^n - !_j^{n-1}}{t} = \Big[s_n(z) \quad (1 \quad) \quad \frac{A_n}{A_n} \Big] \{z_j^n\} \; !_t(z) + \frac{^2(z_j)\,!_t(z_j)}{z_{j;j+1} \quad z_{j-1;j}} \; !_n(z_j)$$

$$+ \Big[\frac{(z_j \quad 1)!_t(z_j \quad 1)}{z_{j-1;j+1}} + \frac{^2(z_{j+1})!_t(z_{j+1})}{(z_{j-1;j+1}) \quad z_{j;j+1}} \Big] \{z_{j-1}^n\} \; !_{j-1}^n$$

$$+ \Big[\frac{(z_{j+1})!_t(z_{j+1})}{z_{j-1;j+1}} + \frac{^2(z_{j+1})!_t(z_{j+1})}{(z_{j-1;j+1}) \quad z_{j;j+1}} \Big] \{z_{j+1}^n\} \; !_{j+1}^n :$$

The law of motion of $!$ can equivalently be written in matrix form

$$\frac{!^n - !^{n-1}}{t} = B^{n-1} !^n$$

where

$$B^n = \begin{bmatrix} \frac{n}{1} + \frac{n}{1} & \frac{n}{2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{n}{1} & \frac{n}{2} + \frac{n}{3} & \frac{n}{3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{n}{2} & \frac{n}{3} + \frac{n}{4} & \frac{n}{4} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{n}{J-2} & \frac{n}{J-1} & \frac{n}{J} \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{n}{J-1} & \frac{n}{J} + \frac{n}{J} \end{bmatrix};$$

Abstracting for brevity from the term $s_n(z) \quad (1 \quad) \quad \frac{A_n}{A_n}$, which is independent of the grid, and spelling out B^n we have

$$\frac{!^n - !^{n-1}}{t} = \begin{bmatrix} \frac{(z_1)}{z_{1;2}} & \frac{(z_1)}{z_{1;2}\, z_{1;2}=2} + \frac{(z_1)}{z_{1;2} \quad z_{1;2}} & \frac{(z_2)}{z_{1;2}} + \frac{(z_2)}{z_{1;2} \quad z_{1;2}} & 0 \\ \frac{(z_1)}{z_{1;3}} + \frac{(z_1)}{z_{1;3} \quad z_{1;2}} & & \frac{(z_2)}{z_{1;2} \quad z_{2;3}} & \frac{(z_3)}{z_{1;3}} + \frac{(z_3)}{z_{1;3} \quad z_{2;3}} \\ 0 & & \frac{(z_2)}{z_{2;4}} + \frac{(z_2)}{z_{2;4} \quad z_{2;3}} & \frac{(z_3)}{z_{2;3} \quad z_{3;4}} \\ 0 & & 0 & \frac{(z_3)}{z_{3;5}} + \frac{(z_3)}{z_{3;4} \quad z_{3;5}} \\ \vdots & & \vdots & \vdots \quad \ddots \end{bmatrix} !^n :$$

We can rewrite this as follows

$$\frac{!^{n} - !^{n-1}}{t} = \begin{bmatrix} \frac{(z_1)}{z_{1;2}} & \frac{(z_2)}{z_{1;2}\,(z_{1;2} - z_{1;2})} & \mathbf{0} \\ \frac{(z_1)}{z_{1;3}} + \frac{(z_1)}{z_{1;3}\,(z_{1;3} - z_{1;2})} & \frac{(z_2)}{z_{1;3}\,(z_{1;2} + z_{2;3})} + \frac{z_{2;3}\,(z_2)}{z_{2;3}(z_{1;3} - z_{1;2})\,(z_{1;2} + z_{2;3})} & \frac{(z_3)}{z_{1;3}} + \frac{z_{3;4}\,(z_3)}{z_{3;4}(z_{1;3} - z_{2;3})\,(z_{2;3} + z_{3;4})} \\ \mathbf{0} & \frac{(z_2)}{z_{2;4}} + \frac{z_{1;2}\,(z_2)}{z_{1;2}(z_{2;4} - z_{2;3})} & \frac{(z_3)}{z_{2;4}\,(z_{2;3} - z_{3;4})} + \frac{z_{2;3}\,(z_3)}{z_{2;3}(z_{3;4} - z_{3;5})} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \end{bmatrix} !^{n}.$$

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

$$\sum_j !_j^n z_j = \sum_j !_j^{n-1} z_j$$

where $\sum_j !_j^n z_j$ is a trapezoid approximation of the integral $\int !^n(z)dz$.

C.2 Finite difference approximation of the integrals

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

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

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

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

and

$$\int_{z_1}^{z} f(z)\,dz = \begin{bmatrix} \frac{1}{2} & 1 & 1 & \cdots & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} f(z_1) \\ f(z_2) \\ \vdots \\ f(z_j) \end{bmatrix} + f(z_{j-1}) + \frac{1}{2}[f(z_j) - f(z_{j-1})] \frac{(z - z_{j-1})}{z} (z - z_{j-1})$$

$$\text{where } j = \arg\min_j \{ j \mid z_j > z \}$$

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)

$$r^{ss} = {}^{h}: \tag{81}$$

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

$$^{ss} = : \tag{82}$$

This means that, from equation (17) and the functional form we assumed for the capital adjustment costs (40),

$$(q_t \quad 1 \quad {}^{0}(\ {}_t))(r_t \quad (\ {}_t \quad)) = \underline{g}_t \quad {}^{00}(\ {}_t)\ \underline{\ }_t \quad (q_t\ {}_t \quad {}_t \quad (\ {}_t)) \tag{83}$$

$$q^{ss} \quad 1 \quad {}^{k}(\ {}^{ss} \quad) \quad {}^{hh} \quad (\ {}^{ss} \quad) = 0 \quad {}^{k} \quad 0 \quad q^{ss}\ {}^{ss} \quad {}^{ss} \quad {}^{k}(\ {}^{ss} \quad)$$

$$^{hh}(q^{ss} \quad 1) = \quad (1 \quad q^{ss})$$

.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

$$^{ss} = i^{ss} \quad ^{h}: \tag{85}$$

In SS, $_{t} = 0$ and $Y_t = 0$. Hence, from equation (20) we obtain

$$m^{ss} = \quad m \; + \; ^{h} \; ^{ss}_{} \; : \tag{86}$$

Using equation (34) and (81),

$$^{h} = \frac{1}{q^{ss}} \quad m \; _{t} Z_t A_t \quad ^{1} L^{1} \quad \frac{z_t}{X \; _{t}} \tag{87}$$

From equation (35) and (81),

$$\frac{A_t}{A_t} = 0 = \frac{1}{q_t}(m \; _{t} Z_t A_t \quad ^{1} L_t^{1} \qquad R_t(1 \quad (\; z_t)) + R_t \quad q_t \quad q_t(1 \qquad)): \tag{88}$$

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

$$^{h} + \quad = \quad (\; ^{h} + \quad)(\quad (1 \quad (\; z_t)) \quad 1) + (1 \qquad) \quad + \quad \frac{z}{X}: \tag{89}$$

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

- Get $r^{ss} = \quad ^{h}$, $^{ss} = \quad$ and $i^{ss} = \quad ^{h} + \quad ^{ss}$ and $R^{ss} = q^{ss}(\; ^{h} + \quad)$ and $m^{ss} = m \; + \; ^{h} \; ^{ss}_{-}$.
- 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 $!^{0}$

  1. Use a non-linear equation solver on equation (89) to obtain z from equation (89).
  2. Obtain $Z_n = (\; _{n} X_n) \; :$

3. Find A from equation (33),

$$A^n = \left[\frac{q^{ss}\ h + q^{ss}}{m\ _n Z_n L_m{}^1\ \ \frac{Z_t}{X\ _t}} \right]^{\frac{1}{1}} :$$

4. Find the stocks $K_n = (1 \quad {}^n(z\ ))A^n$, $D_n = K_n \quad A_n$.

5. Compute $w_n = (1 \quad)m^{ss}Z_n A_n\ L_n \quad ;' _n = \frac{(1 \quad)}{w_n}^{(1 \quad)=} m^{ss\frac{1}{}}$.

6. Get aggregate output $Y = Z_n A_n L_n{}^1$, transfers $T_n = (1 \quad m^{ss})\ Y_n$ $\frac{}{2}(\ {}^{ss})^2 Y_n + (1 \quad)\ A_t$, and consumption $C_n = w_n L_m + r^{ss} D_n + T_n$.

7. Update $S_j^n = \frac{1}{q^{ss}}(\ \max f z' _n \quad R_n; 0g + R_n \quad q^{ss})$ and employ it to construct matrix $B^{n\ \ 1}$.

8. Update $!^{\ n+1}$ using equation $\frac{!^{\ n+1} \quad !^{\ n}}{t} = B^n !^{\ n+1}$.

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

ˆ Set $\quad = w_{L=1} C_{L=1}$ 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 $_t$ at period t are given by

$$_t = \quad I \qquad tA_t^T \quad ^1 \quad _{t\ 1}; \tag{90}$$

where t is the time step between periods and A_t is a matrix whose entries depend *nonlinearly* and in *closed form* on the idiosyncratic and aggregate variables in period t.[19] Similarly, the HJB equation is approximated as[20]

$$v_{t+1} = u_{t+1} + A_{t+1}v_{t+1}; \quad (v_{t+1} \quad v_t) = \ t: \tag{91}$$

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 $_t$ sparse.[21] 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

[19] Technically, this matrix results from the discretization of the infinitesimal generator of the idiosyncratic states. In the example of Section 2, $_t = \ ! \ _t$ and $A_t = B_t$.

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

[21] The introduction of Poisson shocks would not change the sparsity of matrix $_t$.

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* t *goes towards 0.*

Proof : See Appendix D.2.

The proposition guarantees that both strategies coincide when t 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 4 $t = 1=12$: 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 al-

gorithm. 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.[22] 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.[23] 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

[22] 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.

[23] 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

$$\max_{Z_t;u_t(x);\ _t(x);v_t(x)} \int_0^1 \exp(\ \%t) f_0(Z_t)dt \tag{92}$$

s:t: $8t$

$$X_t = f_1(Z_t) \tag{93}$$

$$U_t = f_2(Z_t) \tag{94}$$

$$0 = f_3(Z_t) \tag{95}$$

$$U_t = \int f_4(x;u_t(x);Z_t)\ _t(x)\ dx \tag{96}$$

$$v_t(x) = \underline{v}_t(x) + f_5(x;u_t(x);Z_t) + \sum_{i=1}^{I} b_i(x;u_t(x);Z_t)\frac{@v_t(x)}{@x_i} + \sum_{i=1}^{I}\sum_{k=1}^{I} \frac{(x)\ (x)^>}{2}_{i;k} \frac{@^2v_t(x)}{@x_i@x_k};\ 8x \tag{97}$$

$$0 = \frac{@f_5}{@u_{j;t}} + \sum_{i=1}^{I} \frac{@b_i}{@u_{j;t}}\frac{@v_t(x)}{@x_i};\quad j = 1;:::;J;8x: \tag{98}$$

$$\dot{}_t(x) = -\sum_{i=1}^{I} \frac{@}{@x_i}[b_i(x;u_t(x);Z_t)\ _t(x)] + \frac{1}{2}\sum_{i=1}^{I}\sum_{k=1}^{I} \frac{@^2}{@x_i@x_k}\left[(x)\ (x)^>_{i;k}\ _t(x)\right];\ 8x \tag{99}$$

$$X_0 = X_0 \tag{100}$$

$$_0(x) = \ _0(x) \tag{101}$$

$$\lim_{t!1} U = U_1 \tag{102}$$

$$\lim_{t!1} v(x) = v(x)_1 \tag{103}$$

where we have adopted the following notation:

ˆ 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
- (x) distribution of agents across states

$v(x)$ value function as function of individual state

X aggregate state vector (other than)

$\hat{U}$ aggregate control vector of purely contemporaneous variables

U aggregate control vector of intertemporal variables

Υ control vector of aggregator variables

$Z_t = \{\Upsilon_t, U_t, U_t, X_t\}$ vector of all aggregate variables

$\hat{}$ 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.[24] 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.

[24] Notice that the planner's discount factor, %;can be different to that of individual agents, .

2.a The Lagrangian The Lagrangian for this problem is given by:[25]

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

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

[25] 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:

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

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

$$
\begin{aligned}
& \int_0^1 e^{-\rho t}\lambda_{1,t}\dot{X}_t\, dt + \int_0^1 e^{-\rho t}\lambda_{2,t}\dot{U}_t\, dt + \int_0^1 e^{-\rho t}\lambda_{5,t}(x)\dot{v}_t(x)\, dt\, dx \\
& - \int_0^1 e^{-\rho t}\lambda_{7,t}(x)\dot{\mu}_t(x)\, dt\, dx
\end{aligned}
$$

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

$$
\begin{aligned}
\int_0^1 e^{-\rho t} a_t \dot{b}_t\, dt &= e^{-\rho t} a_t b_t \Big|_0^1 - \int_0^1 e^{-\rho t}(\dot{a}_{1,t} - \rho a_{1,t}) b_t\, dt \\
&= \lim_{t\to\infty} e^{-\rho t} a_t b_t - a_0 b_0 - \int_0^1 e^{-\rho t}(\dot{a}_t - \rho a_t) b_t\, dt
\end{aligned}
$$

to get

$$\lim_{t\to\infty} e^{-\rho t}\,\mu_{1,t}X_t \;-\; \mu_{1,0}X_0 \;-\; \int_0^\infty e^{-\rho t}(\dot{\mu}_{1,t} - \rho\,\mu_{1,t})X_t\,dt + \lim_{t\to\infty} e^{-\rho t}\,\mu_{2,t}U_t \;-\; \mu_{2,0}U_0$$
$$- \int_0^\infty e^{-\rho t}(\dot{\mu}_{2,t} - \rho\,\mu_{2,t})U_t\,dt\,dx$$
$$+ \int \lim_{t\to\infty} e^{-\rho t}\,\mu_{5,t}(x)v_t(x) \;-\; \mu_{5,0}(x)v_0(x)\;dx \;-\; \int_0^\infty e^{-\rho t}(\dot{\mu}_{5,t}(x) - \rho\,\mu_{5,t}(x))v_t(x)\,dt\,dx$$
$$\int \lim_{t\to\infty} e^{-\rho t}\,\mu_{7,t}(x)\;\gamma_t(x) \;-\; \mu_{7,0}(x)\;\gamma_0(x)\,dx + \int_0^\infty e^{-\rho t}(\dot{\mu}_{7,t}(x) - \rho\,\mu_{7,t}(x))\;\gamma_t(x)\,dt\,dx$$

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

$$+ \lim_{t\to\infty} e^{-\rho t}\,\mu_{1,t}X_t \;-\; \mu_{2,0}U_0 \;-\; \int_0^\infty e^{-\rho t}(\dot{\mu}_{1,t} - \rho\,\mu_{1,t})X_t\,dt \;-\; \int_0^\infty e^{-\rho t}(\dot{\mu}_{2,t} - \rho\,\mu_{2,t})U_t\,dt$$
$$\mu_{5,0}(x)v_0(x)dx + \int_0^\infty e^{-\rho t}(\dot{\mu}_{5,t}(x) - \rho\,\mu_{5,t}(x))v_t(x)\,dt\,dx$$
$$\int \lim_{t\to\infty} e^{-\rho t}\,\mu_{7,t}(x)\;\gamma_t(x)dx + \int_0^\infty e^{-\rho t}(\dot{\mu}_{7,t}(x) - \rho\,\mu_{7,t}(x))\;\gamma_t(x)\,dt\,dx$$

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

$$+ \int \Big(\mu_{5,t}(x)v_t(x) + f_5(x;u_t(x);Z_t) \sum_{i=1}^{I} \frac{\partial b_i(x;u_t(x);Z_t)\;\mu_{5,t}(x)}{\partial x_i} v_t(x) \Big)\,dx$$
$$+ \int \Big(+ \frac{1}{2}\sum_{i=1}^{I} \frac{\partial^2}{\partial x_i^2}\;\sigma_i^2(x)\;\mu_{5,t}(x)\;v_t(x) \Big)\,dx$$
$$+ \int \Big[\sum_{j=1}^{J} \Big(\mu_{6,j,t}(x)\frac{\partial f_{5,t}}{\partial \mu_{j,t}} \sum_{i=1}^{I} \frac{\partial \left(\mu_{6,j,t}(x)\frac{\partial b_i}{\partial \mu_{j,t}} \right)}{\partial x_i} v_t(x) \Big) \Big]\,dx$$
$$+ \int \Big(\sum_{i=1}^{I} \frac{\partial\,\mu_{7,t}(x)}{\partial x_i}\,[b_i(x;u_t(x);Z_t)\;\gamma_t(x)] + \sum_{i=1}^{I} \frac{\partial^2\,\mu_{7,t}(x)}{\partial x_i^2}\,\frac{\sigma_i^2(x)}{2}\;\gamma_t(x) \Big)\,dx\,dt$$

Putting this all together the Lagrangian has become:

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

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:

$$U_t: \quad 0 = \quad (-_{2;t} \quad \%_{2;t}) \tag{104}$$

$$+ \frac{@f_{0;t}}{@U_t} \quad {}_{1;t}\frac{@f_{1;t}}{@U_t} \quad {}_{2;t}\frac{@f_{2;t}}{@U_t} \quad {}_{3;t}\frac{@f_{3;t}}{@U_t} \quad {}_{4;t} \quad \frac{@f_{4;t}}{@U_t} \quad {}_t(x) \; dx \tag{105}$$

$$+ \quad {}_{5;t}(x) \quad \frac{@f_{5;t}}{@U_t} + \sum_{i=1}^{I} \frac{@h_{i;t}}{@U_t}\frac{@v_t(x)}{@x_i} \quad dx \tag{106}$$

$$+ \sum_{j=1}^{J} \quad {}_{6;j;t}(x) \quad \frac{@^2 f_{5;t}}{@\mu_{j;t}\,@U_t} + \sum_{i=1}^{I} \frac{@h_{i;t}}{@\mu_{j;t}\,@U_t}\frac{@v_t(x)}{@x_i} \quad dx \tag{107}$$

$$+ \quad {}_{7;t}(x) \quad \sum_{i=1}^{I} \frac{@}{@x_i}\frac{@h_{i;t}}{@U_t} \quad {}_t(x) \quad dx; \tag{108}$$

$$8t > 0; \tag{109}$$

$$0 = \quad {}_{2;0}: \tag{110}$$

$$X_t: \quad 0 = \quad (-_{1;t} \quad \%_{1;t})$$

$$+ \frac{@f_{0;t}}{@X_t} \quad {}_{1;t}\frac{@f_{1;t}}{@X_t} \quad {}_{2;t}\frac{@f_{2;t}}{@X_t} \quad {}_{3;t}\frac{@f_{3;t}}{@X_t} \quad {}_{4;t} \quad \frac{@f_{4;t}}{@X_t} \quad {}_t(x) \; dx$$

$$+ \quad {}_{5;t}(x) \quad \frac{@f_{5;t}}{@X_t} + \sum_{i=1}^{I} \frac{@h_{i;t}}{@X_t}\frac{@v_t(x)}{@x_i} \quad dx$$

$$+ \sum_{j=1}^{J} \quad {}_{6;j;t}(x) \quad \frac{@^2 f_{5;t}}{@\mu_{j;t}\,@X_t} + \sum_{i=1}^{I} \frac{@h_{i;t}}{@\mu_{j;t}\,@X_t}\frac{@v_t(x)}{@x_i} \quad dx$$

$$+ \quad {}_{7;t}(x) \quad \sum_{i=1}^{I} \frac{@}{@x_i}\frac{@h_{i;t}}{@X_t} \quad {}_t(x) \quad dx;$$

$$8t \quad 0;$$

$$0 = \lim_{t!1} e^{\%t} \quad {}_{1;t}(x):$$

$$
\begin{aligned}
\hat{U}_t : \quad 0 = \; & 0 \\
& + \frac{\partial f_{0,t}}{\partial \hat{U}_t} + \lambda_{1,t} \frac{\partial f_{1,t}}{\partial \hat{U}_t} + \lambda_{2,t} \frac{\partial f_{2,t}}{\partial \hat{U}_t} + \lambda_{3,t} \frac{\partial f_{3,t}}{\partial \hat{U}_t} + \lambda_{4,t} \frac{\partial f_{4,t}}{\partial \hat{U}_t} \psi_t(x) \, dx \\
& + \int_{\mathcal{X}} \lambda_{5,t}(x) \left(\frac{\partial f_{5,t}}{\partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial h_{i,t}}{\partial \hat{U}_t} \frac{\partial \psi_t(x)}{\partial x_i} \right) dx \\
& + \sum_{j=1}^{J} \int_{\mathcal{X}} \lambda_{6,j,t}(x) \left(\frac{\partial^2 f_{5,t}}{\partial \mu_{j,t} \partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial h_{i,t}}{\partial \mu_{j,t} \partial \hat{U}_t} \frac{\partial \psi_t(x)}{\partial x_i} \right) dx \\
& + \int_{\mathcal{X}} \lambda_{7,t}(x) \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left(\frac{\partial h_{i,t}}{\partial \hat{U}_t} \psi_t(x) \right) dx; \\
& \forall t \geq 0.
\end{aligned}
$$

$$
\begin{aligned}
\tilde{U}_t : \quad 0 = \; & \lambda_{4,t} \\
& + \frac{\partial f_{0,t}}{\partial \tilde{U}_t} + \lambda_{1,t} \frac{\partial f_{1,t}}{\partial \tilde{U}_t} + \lambda_{2,t} \frac{\partial f_{2,t}}{\partial \tilde{U}_t} + \lambda_{3,t} \frac{\partial f_{3,t}}{\partial \tilde{U}_t} + \lambda_{4,t} \frac{\partial f_{4,t}}{\partial \tilde{U}_t} \psi_t(x) \, dx \\
& + \int_{\mathcal{X}} \lambda_{5,t}(x) \left(\frac{\partial f_{5,t}}{\partial \tilde{U}_t} + \sum_{i=1}^{I} \frac{\partial h_{i,t}}{\partial \tilde{U}_t} \frac{\partial \psi_t(x)}{\partial x_i} \right) dx \\
& + \sum_{j=1}^{J} \int_{\mathcal{X}} \lambda_{6,j,t}(x) \left(\frac{\partial^2 f_{5,t}}{\partial \mu_{j,t} \partial \tilde{U}_t} + \sum_{i=1}^{I} \frac{\partial h_{i,t}}{\partial \mu_{j,t} \partial \tilde{U}_t} \frac{\partial \psi_t(x)}{\partial x_i} \right) dx \\
& + \int_{\mathcal{X}} \lambda_{7,t}(x) \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left(\frac{\partial h_{i,t}}{\partial \tilde{U}_t} \psi_t(x) \right) dx; \\
& \forall t \geq 0.
\end{aligned}
$$

Value function, distribution and policy functions

$$v_t(x): \quad 0 = \mu_{5;t}(x) - \sum_{i=1}^{I} \frac{\partial [\mu_{5;t}(x) b_i(x; u_t(x); Z_t)]}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{I} \frac{\partial^2}{\partial x_i^2} \sigma_i^2(x) \mu_{5;t}(x) - \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \mu_{6;j;t}(x) \frac{\partial b_i(x; u_t(x); Z_t)}{\partial \mu_{j;t}}$$
$$(\rho_{5;t}(x) + \psi_{5;t}(x));$$
$$\forall t > 0;$$

$$0 = \mu_{5;0}(x):$$

$$_t(x): \quad 0 = \mu_{4;t} f_4(x; u_t(x); Z_t) + \mu_{7;t}(x) \sum_{i=1}^{I} \frac{\partial \mu_{7;t}(x)}{\partial x_i} b_i(x; u_t(x); Z_t) + \sum_{i=1}^{I} \frac{\partial^2 \mu_{7;t}(x)}{\partial x_i^2} \frac{\sigma_i^2(x)}{2}$$
$$+ (\rho_{7t}(x) + \psi_{7;t}(x));$$
$$\forall t \geq 0;$$

$$0 = \lim_{t \to \infty} e^{-\psi t} \mu_{7;t}(x):$$

$$u_{l;t}(x): \quad 0 = \mu_{4;t} \frac{\partial f_4}{\partial u_{l;t}} \mu_t(x) + \mu_{5;t}(x) \left[\frac{\partial f_5}{\partial u_{l;t}} + \sum_{i=1}^{I} \frac{\partial b_i}{\partial u_{l;t}} \frac{\partial \ln \nu_t(x)}{\partial x_i} + \sum_{j=1}^{J} \mu_{6;k;t}(x) \left[\frac{\partial^2 f_5}{\partial u_{l;t} \partial u_{j;t}} + \sum_{i=1}^{I} \frac{\partial^2 b_i}{\partial u_{l;t} \partial u_{j;t}} \frac{\partial \ln \nu_t(x)}{\partial x_i} \right] + \sum_{i=1}^{I} \frac{\partial \mu_{7;t}(x)}{\partial x_i} \frac{\partial b_i}{\partial u_{l;t}} \mu_t(x) \right] :$$

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 $[N_1; :::; N_l]$ where $1; ::; l$ 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 x_i:We define $x^n \quad (x_{1;n_1}; :::; x_{i;n_i}; :::; x_{l;n_l})$, where $n_1 \; 2 \; f \, 1; N_1 g; :::; n_l \; 2 \; f \, 1; N_l g$. We are assuming that, due to state constraints and/or reflecting boundaries, the dynamics of idiosyncratic states are constrained to the compact set $[x_{1;1}; x_{1;N_1}]$ $[x_{2;1}; x_{2;N_2}]$ :::: $[x_{l;1}; x_{l;N_l}]$. We also define $x^{n_i+1} \quad (x_{1;n_1}; :::; x_{i;n_i+1}; :::; x_{l;n_l})$, $x^{n_i \; 1}$ $(x_{1;n_1}; :::; x_{i;n_i \; 1}; :::; x_{l;n_l})$ $f_t^n \quad f \; (x^n; u_t^n; Z_t)$, $f_t^{n_i \; 1} \quad f \; (x^{n_i \; 1}; u_t^n; Z_t)$ and $f_t^{n_i+1}$ $f \; (x^{n_i+1}; u_t^n; Z_t)$. I.e. the superscript n indicates a particular grid point and the superscript $n_i + 1$ and $n_i \quad 1$ 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 $r \;$ or $\hat{r}$:

$$r_i[v_t^n] \quad I_{b_{i;t}^n > 0} \frac{v_t^{n_i+1} \quad v_t^n}{x_i} + I_{b_{i;t}^n < 0} \frac{v_t^n \quad v_t^{n_i \; 1}}{x_i} \; ;$$

$$\hat{r}_i[\;{}_t^n] \quad \left[\frac{I_{b_{i;t}^{n_i+1} < 0} \; {}_t^{n_i+1} \quad I_{b_{i;t}^n < 0} \; {}_t^n}{x_i} + \frac{I_{b_{i;t}^n > 0} \; {}_t^n \quad I_{b_{i;t}^{n_i \; 1} > 0} \; {}_t^{n_i \; 1}}{x_i} \right] ;$$

for any discretized functions v_t^n; ${}_t^n$. We simplify the notation for sums $\sum_n \quad \sum_{n_1 2 f \, 1; :::; N_1 g; :::; n_l 2 f \, 1; :::; N_l g}$:

We maintain the subscript t even if it refers now to discrete time with a step t, that is, X_{t+1} is the shortcut for $X_{t+4\,t}$: The second-order derivative is approximated as

$$4_i[v_t^n] \quad \left[\frac{v_t^{n_i+1} \; + \; v_t^{n_i \; 1} \quad 2 \, (v_t^n)}{(\;x_i)^2} \right] :$$

We start with the optimality condition for U_t

$$U_t: \quad 0 = \frac{\beta_{2;t} - \beta_{2;t-1}}{t} - \pi_{2;t} \tag{111}$$

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

$$+ \sum_{n} \pi_{5;t}^n \left(\frac{\partial f_5^n}{\partial U_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial U_t} r_i[v_t^n] \right)$$

$$+ \sum_{j=1}^{J} \sum_{n} \mu_{6;j;t}^n \left(\frac{\partial f_5^n}{\partial \mu_j \partial U_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial \mu_j \partial U_t} r_i[v_t^n] \right)$$

$$+ \sum_{n} \sum_{i=1}^{I} \hat{r}_{i;t}^n \frac{\partial b_{i,t}^n}{\partial U_t} \tag{113}$$

$$\forall t \geq 0.$$

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

$$X_t: \quad 0 = \left(\frac{\beta_{1;t} - \beta_{1;t-1}}{t} - \pi_{1;t} \right)$$

$$+ \frac{\partial f_0}{\partial X_t} - \pi_{1;t}\frac{\partial f_1}{\partial X_t} - \beta_{2;t}\frac{\partial f_2}{\partial X_t} - \pi_{3;t}\frac{\partial f_3}{\partial X_t} - \sum_{n} \pi_{4;t}^n \frac{\partial f_4^n}{\partial X_t}$$

$$+ \sum_{n} \pi_{5;t}^n \left(\frac{\partial f_5^n}{\partial X_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial X_t} r_i[v_t^n] \right)$$

$$+ \sum_{j=1}^{J} \sum_{n} \mu_{6;j;t}^n \left(\frac{\partial f_5^n}{\partial \mu_j \partial X_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial \mu_j \partial X_t} r_i[v_t^n] \right)$$

$$+ \sum_{n} \sum_{i=1}^{I} \hat{r}_{i;t}^n \frac{\partial b_{i,t}^n}{\partial X_t}$$

$$\forall t > 0.$$

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

$$
\begin{aligned}
\breve{U}_t : \quad 0 = \; & \nabla_{4;t} \\
& + \frac{\partial f_0}{\partial \breve{U}_t} + \nabla_{1;t} \frac{\partial f_1}{\partial \breve{U}_t} + \nabla_{2;t} \frac{\partial f_2}{\partial \breve{U}_t} + \nabla_{3;t} \frac{\partial f_3}{\partial \breve{U}_t} + \nabla_{4;t} \sum_{n=1}^{N} \frac{\partial f_4^n}{\partial \breve{U}_t} \\
& + \sum_{n} \nabla_{5;t}^n \left(\frac{\partial f_5^n}{\partial \breve{U}_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial \breve{U}_t} r_i[v_t^n] \right) \\
& + \sum_{j=1}^{J} \sum_{n} \nabla_{6;j;t}^n \left(\frac{\partial^2 f_5^n}{\partial \mu_j \partial \breve{U}_t} + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial \mu_j \partial \breve{U}_t} r_i[v_t^n] \right) \\
& + \sum_{n} \nabla_{7;t}^n \sum_{i=1}^{I} \hat{r}_i \frac{\partial b_{i,t}^n}{\partial \breve{U}_t} \\
& \forall t \geq 0.
\end{aligned}
$$

The discretized optimality condition with respect to the value function $v_t(x)$, the

distribution $_t(x)$ and the individual jump variable $u_{j;t}(x)$ are.

$$v_t(x): \quad 0 = \quad \overset{n}{}_{5;t} \quad \overset{X^{I}}{} \quad \hat{r}_i \quad \overset{n}{}_{5;t} b^{n}_{i;t} \tag{114}$$
$$+ \frac{1}{2} \overset{X^{I}}{}_{i=1} \overset{X^{I}}{}_{k=1} r_i \quad \overset{n}{}_{i;k} \quad \overset{n}{}_{5;t}$$
$$\overset{X^{J}}{}_{j=1} \overset{X^{I}}{}_{i=1} \hat{r}_i \quad \overset{n}{}_{6;j;t} \frac{@ b^{n}_{i;t}}{@ \mu_{j;t}}$$
$$\left(\frac{\overset{n}{}_{5;t} \quad \overset{n}{}_{5;t} \quad 1}{t} \quad \% \overset{n}{}_{5;t}\right):$$

$$_t(x): \quad 0 = \quad \overset{n}{}_{4;t} f^{n}_{4;t} \tag{115}$$
$$+ \quad \overset{n}{}_{7;t}(x) \overset{X^{I}}{}_{i=1} b_i(x; u_t(x); Z_t) r_i \quad \overset{n}{}_{7;t} + \frac{1}{2} \overset{X^{I}}{}_{i=1} \quad \overset{2}{}_i \quad \overset{n}{} 4^{2}_i \quad \overset{n}{}_{7;t} \quad !$$
$$+ \frac{\overset{n}{}_{7;t} \quad \overset{n}{}_{7;t} \quad 1}{t} \quad \% \overset{n}{}_{7;t}$$

$$u_{l;t}(x): \quad 0 = \quad \overset{n}{}_{4;t} \frac{@ f_4}{@ \mu_{l;t}} \quad \overset{n}{}_t \tag{116}$$
$$+ \overset{X^{J}}{}_{j=1} \quad \overset{n}{}_{6;k;t} \quad \frac{@^{2} f^{n}_{5;t}}{@ \mu_{l;t} @ \mu_{j;t}} + \overset{X^{I}}{}_{i=1} \quad \frac{@^{2} b^{n}_{i;t}}{@ \mu_{l;t} @ \mu_{j;t}} r_i [v^{n}_t] \quad !$$
$$\overset{X^{I}}{}_{i=1} r_i \quad \overset{n}{}_{7;t} \quad \frac{@ b^{n}_{i;t}}{@ \mu_{l;t}} \quad \overset{n}{}_t$$

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 t) and the idiosyncratic state (N grid points, grid step x_i). We define the discount factor $(1 + \% \ t)^{-1}$.

$$\max_{Z_t;u_t^n;\ {}_t^n;v_t^n} \quad \sum_t {}^t f_0(Z_t)$$

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

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

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

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

$$U_t = \sum_{n=1}^{N} f_4(x^n; u_t^n; Z_t) \ {}_t^n \tag{120}$$

$$v_t^n = \frac{v_{t+1}^n - v_t^n}{t} + f_5(x^n; u_t^n; Z_t) + \sum_{i=1}^{I} b_i(x^n; u_t^n; Z_t) r_i[v_t^n] \tag{121}$$

$$+ \frac{1}{2} \sum_{i=1}^{I} {}_i^2 \ {}^n \ 4_i^2[v_t^n]; \ \forall n$$

$$0 = \frac{\partial f_{5;t}}{\partial u_{j;t}} + \sum_{i=1}^{I} \frac{\partial b_{i;t}}{\partial u_{j;t}} r_i[v_t^n]; \quad \forall j; n: \tag{122}$$

$$\frac{{}_{t+1}^n - {}_t^n}{t} = \sum_{i=1}^{I} \hat{r}_i \ b_{i;t}^n \ {}_t^n \tag{123}$$

$$+ \frac{1}{2} \sum_{i=1}^{I} 4_i \ {}_i^2 \ {}_t^n \tag{124}$$

$$X_0 = X_0 \tag{125}$$

$${}_0^n = {}_0^n \tag{126}$$

3.b The Lagrangian The Lagrangian is

$$
\begin{aligned}
L = & \sum_t {}^t f_0(Z_t) \\
& + \sum_t {}^t {}_{1;t} \frac{X_{t+1} - X_t}{t} - f_1(Z_t) \\
& + \sum_t {}^t {}_{2;t} \frac{U_{t+1} - U_t}{t} - f_2(Z_t) \\
& + \sum_t {}^t {}_{3;t} \{ - f_3(Z_t) \} \\
& + \sum_t {}^t {}_{4;t} \Big(U_t - \sum_n f_4(x^n; u_t^n; Z_t) \Big) \\
& + \sum_t \sum_n {}^t {}_{5;t} \Big(v_t^n + \frac{v_{t+1}^n - v_t^n}{t} + f_5(x^n; u_t^n; Z_t) + \sum_{i=1}^I b_i(x^n; u_t^n; Z_t) r_i[v_t^n] \\
& \qquad + \sum_{i=1}^I {}_i^2[v_t^n] \Big) \\
& + \sum_t \sum_n \sum_{j=1}^J {}^t {}_{6;j;t} \Big(\frac{\partial f_{5;t}^n}{\partial u_{j;t}} + \sum_{i=1}^I \frac{\partial b_{i;t}^n}{\partial u_{j;t}} r_i[v_t^n] \Big) \\
& + \sum_t \sum_n {}^t {}_{7;t}^n \Big(\frac{{}_{t+1}^n - {}_t^n}{t} - \sum_{i=1}^I \hat{r}_i b_{i;t}^n {}_t^n \\
& \qquad + \frac{1}{2} \sum_{i=1}^I {}_i [{}_i^2 {}_t^n] \Big)
\end{aligned}
$$

3.c The optimality conditions The FOCs are

$$
\begin{aligned}
\frac{\partial L}{\partial u_t} : 0 = & \frac{\partial f_{0;t}}{\partial u_t} - {}_{1;t} \frac{\partial f_{1;t}}{\partial u_t} + {}_{2;t} \frac{1}{t} \frac{\partial f_{2;t}}{\partial u_t} + {}^1 {}_{2;t} \, 1 \frac{1}{t} {}_{3;t} \frac{\partial f_{3;t}}{\partial u_t} - {}_{4;t} \sum_n \frac{\partial f_{4;t}^n}{\partial u_t} \\
& + \sum_n \Big({}_{5;t} + \frac{\partial f_{5;t}^n}{\partial u_t} + \sum_{i=1}^I \frac{\partial b_{i;t}^n}{\partial u_t} r_i[v_t^n] \Big) \\
& + \sum_n \sum_{j=1}^J {}^n {}_{6;j;t} \Big(\frac{\partial^2 f_{5;t}^n}{\partial u_{j;t} \partial u_t} + \sum_{i=1}^I \frac{\partial^2 b_{i;t}^n}{\partial u_{j;t} \partial u_t} r_i[v_t^n] \Big) \\
& + \sum_n \sum_{i=1}^I {}_{7;t}^n \, {}_{n_i \, 1} \, I_{b_{i;t}^n < 0} \frac{\partial b_{i;t}^n}{\partial u_t} \frac{{}_t^n}{x_i} + \sum_{i=1}^I {}_{7;t}^{n_i + 1} \, {}_{7;t}^n \, I_{b_{i;t}^n > 0} \frac{\partial b_{i;t}^n}{\partial u_t} \frac{{}_t^n}{x_i} \\
& 8t \quad 0
\end{aligned}
\tag{127}
$$

$$\frac{\partial L}{\partial x_t}: 0 = \frac{\partial f_{0;t}}{\partial x_t} \quad {}_{1;t}\left(\frac{1}{t} + \frac{\partial f_{1;t}}{\partial x_t}\right) + {}_{1}{}_{1;t}\left(1 - \frac{1}{t}\right) \quad {}_{2;t}\frac{\partial f_{2;t}}{\partial x_t} \quad {}_{3;t}\frac{\partial f_{3;t}}{\partial x_t} \quad {}_{4;t}\sum_{n} \frac{\partial f^{n}_{4;t}}{\partial x_t}\,{}^{n}_{t}$$
$$+ \sum_{n} {}^{n}_{5;t}\left(\frac{\partial f^{n}_{5;t}}{\partial x_t} + \sum_{i=1}^{I} \frac{\partial b^{n}_{i;t}}{\partial x_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n} {}^{n}_{6;j;t}\left(\frac{\partial^2 f^{n}_{5;t}}{\partial \mu_{j;t}\,\partial x_t} + \sum_{i=1}^{I} \frac{\partial^2 b^{n}_{i;t}}{\partial \mu_{j;t}\,\partial x_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n}\left(\sum_{i=1}^{I} {}^{n}_{7;t}\,{}^{n_i}_{7;t}\,{}_{1}\,\mathbb{I}_{b^{n}_{i;t}<0}\frac{\partial b^{n}_{i;t}}{\partial x_t}\frac{{}^{n}_{t}}{x_i} + \sum_{i=1}^{I} {}^{n_i+1}_{7;t}\,{}^{n}_{7;t}\,\mathbb{I}_{b^{n}_{i;t}>0}\frac{\partial b^{n}_{i;t}}{\partial x_t}\frac{{}^{n}_{t}}{x_i}\right)$$
$$8t > 0$$

$$\frac{\partial L}{\partial v_t}: 0 = \frac{\partial f_{0;t}}{\partial v_t} \quad {}_{1;t}\frac{\partial f_{1;t}}{\partial v_t} \quad {}_{2;t}\frac{\partial f_{2;t}}{\partial v_t} \quad {}_{3;t}\frac{\partial f_{3;t}}{\partial v_t} \quad {}_{4;t}\sum_{n} \frac{\partial f^{n}_{4;t}}{\partial v_t}\,{}^{n}_{t}$$
$$+ \sum_{n} {}^{n}_{5;t}\left(+\frac{\partial f^{n}_{5;t}}{\partial v_t} + \sum_{i=1}^{I} \frac{\partial b^{n}_{i;t}}{\partial v_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n} {}^{n}_{6;j;t}\left(\frac{\partial^2 f^{n}_{5;t}}{\partial \mu_{j;t}\,\partial v_t} + \sum_{i=1}^{I} \frac{\partial^2 b^{n}_{i;t}}{\partial \mu_{j;t}\,\partial v_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n}\left(\sum_{i=1}^{I} {}^{n}_{7;t}\,{}^{n_i}_{7;t}\,{}_{1}\,\mathbb{I}_{b^{n}_{i;t}<0}\frac{\partial b^{n}_{i;t}}{\partial v_t}\frac{{}^{n}_{t}}{x_i} + \sum_{i=1}^{I} {}^{n_i+1}_{7;t}\,{}^{n}_{7;t}\,\mathbb{I}_{b^{n}_{i;t}>0}\frac{\partial b^{n}_{i;t}}{\partial v_t}\frac{{}^{n}_{t}}{x_i}\right)$$
$$8t \quad 0$$

$$\frac{\partial L}{\partial \omega_t}: 0 = \frac{\partial f_{0;t}}{\partial \omega_t} \quad {}_{1;t}\frac{\partial f_{1;t}}{\partial \omega_t} \quad {}_{2;t}\frac{\partial f_{2;t}}{\partial \omega_t} \quad {}_{3;t}\frac{\partial f_{3;t}}{\partial \omega_t} \quad {}_{4;t}\sum_{n} \frac{\partial f^{n}_{4;t}}{\partial \omega_t}\,{}^{n}_{t}$$
$$+ \sum_{n} {}^{n}_{5;t}\left(+\frac{\partial f^{n}_{5;t}}{\partial \omega_t} + \sum_{i=1}^{I} \frac{\partial b^{n}_{i;t}}{\partial \omega_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n} {}^{n}_{6;j;t}\left(\frac{\partial^2 f^{n}_{5;t}}{\partial \mu_{j;t}\,\partial \omega_t} + \sum_{i=1}^{I} \frac{\partial^2 b^{n}_{i;t}}{\partial \mu_{j;t}\,\partial \omega_t} r_i[v^{n}_{t}]\right)$$
$$+ \sum_{n}\left(\sum_{i=1}^{I} {}^{n}_{7;t}\,{}^{n_i}_{7;t}\,{}_{1}\,\mathbb{I}_{b^{n}_{i;t}<0}\frac{\partial b^{n}_{i;t}}{\partial \omega_t}\frac{{}^{n}_{t}}{x_i} + \sum_{i=1}^{I} {}^{n_i+1}_{7;t}\,{}^{n}_{7;t}\,\mathbb{I}_{b^{n}_{i;t}>0}\frac{\partial b^{n}_{i;t}}{\partial \omega_t}\frac{{}^{n}_{t}}{x_i}\right)$$
$$8t \quad 0$$

$$\frac{@L}{@\nu_t}: 0 = {}^{n}_{5;t}\left(\frac{1}{t} + \sum_{i=1}^{I} b^{n}_{i;t}\frac{I_{b^{n}_{t}<0} \quad I_{b^{n}_{t}>0}}{x_i}\right) \quad \sum_{i=1}^{I}\frac{2(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}} \tag{128}$$

$$+ \ {}^{n}_{5;t\ \ 1} \quad {}^{1}\frac{1}{t}$$

$$+ \sum_{i=1}^{I} {}^{n_i\ \ 1}_{5;t} b^{n_i\ \ 1}_{i;t}\frac{I_{b^{n_i\ \ 1}_{i;t}>0}}{x_i} + \sum_{i=1}^{I} {}^{n_i\ \ 1}_{5;t}\frac{(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}}$$

$$\sum_{i=1}^{I} {}^{n_i+1}_{5;t} b^{n_i+1}_{i;t}\frac{I_{b^{n_i+1}_{i;t}<0}}{x_i} + \sum_{i=1}^{I} {}^{n_i+1}_{5;t}\frac{(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}}$$

$$+ \sum_{j=1}^{J}\sum_{i=1}^{I}\left({}^{n}_{6;j;t}\frac{@b^{n}_{i;t}}{@\mu^{n}_{j;t}}\frac{I_{b^{n}_{i;t}<0} \quad I_{b^{n}_{i;t}>0}}{x_i} + {}^{n_i\ \ 1}_{6;j;t}\frac{@b^{n_i\ \ 1}_{i;t}}{@\mu^{n_i\ \ 1}_{j;t}}\frac{I_{b^{n_i\ \ 1}_{i;t}>0}}{x_i}\right)$$

$$\sum_{j=1}^{J}\sum_{i=1}^{I}{}^{n_i+1}_{6;j;t}\left(\frac{@b^{n_i+1}_{i;t}}{@\mu^{n_i+1}_{j;t}}\frac{I_{b^{n_i+1}_{i;t}<0}}{x_i}\right) \tag{129}$$

$$8t \quad 0$$

$$\frac{@L}{@\ ^{n}_{t}}: 0 = {}^{4;t}f^{\ n}_{4;t} \tag{130}$$

$$+ \ {}^{n}_{7;t}\left(\frac{1}{t} \sum_{i=1}^{I} I_{b^{n}_{i;t}>0} \quad I_{b^{n}_{i;t}<0}\right)\frac{b^{n}_{i;t}}{x_i} \quad \sum_{i=1}^{I}\frac{2(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}}$$

$$+ \left(\sum_{i=1}^{I} {}^{n_i\ \ 1}_{7;t}\frac{I_{b^{n}_{i;t}<0}b^{n}_{i;t}}{x_i} + \sum_{i=1}^{I}\frac{(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}}\right)$$

$$+ \left(\sum_{i=1}^{I} {}^{n_i+1}_{7;t}\frac{I_{b^{n}_{i;t}>0}b^{n}_{i;t}}{x_i} + \sum_{i=1}^{I}\frac{(\ ^{2}_{i})^{n}}{2(\ x_i)^{2}}\right)$$

$$+ \ {}^{1\ \ n}_{7;t\ \ 1} \quad \frac{1}{t}$$

$$8t > 0$$

$$\frac{\partial L}{\partial \theta_{i,t}} : 0 = \quad (131)$$

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

$$U_t : \quad 0 =$$

The discretize-optimize condition (127), rearranges to

$$
\begin{aligned}
\frac{\partial L}{\partial \psi}: 0 \;=\; & \frac{\lambda_{2;t} - \lambda_{2;t-1}}{t} - \frac{1}{t}\lambda_{2;t-1} \\
& + \frac{\partial f_{0;t}}{\partial \psi} - \lambda_{1;t}\frac{\partial f_{1;t}}{\partial \psi} - \lambda_{2;t}\frac{\partial f_{2;t}}{\partial \psi} - \lambda_{3;t}\frac{\partial f_{3;t}}{\partial \psi} - \sum_{n=1}^{N} \lambda_{4;t}^{n}\frac{\partial f_{4;t}^{n}}{\partial \psi} \\
& + \sum_{n=1}^{N} \lambda_{5;t}^{n}\left(\frac{\partial f_{5;t}^{n}}{\partial \psi} + \sum_{i=1}^{I} \frac{\partial b_{i;t}^{n}}{\partial \psi} r_{i}[v_{t}^{n}]\right) \\
& + \sum_{n=1}^{N}\sum_{j=1}^{J} \lambda_{6;j;t}^{n}\left(\frac{\partial^{2} f_{5;t}^{n}}{\partial \psi_{j;t}\,\partial \psi} + \frac{\partial^{2} b_{i;t}^{n}}{\partial \psi_{j;t}\,\partial \psi} r_{i}[v_{t}^{n}]\right) \\
& + \sum_{n} \left(\lambda_{7;t}^{n} - \lambda_{7;t}^{n_i - 1}\right) I_{b_{i;t}^{n} < 0} \frac{\partial b_{i;t}^{n}}{\partial \psi}\frac{n}{t} + \sum_{i=1}^{I} \lambda_{7;t}^{n_i+1} I_{b_{i;t}^{n} > 0} \frac{\partial b_{i;t}^{n}}{\partial \psi}\frac{n}{t} \\
& \forall t \quad 0
\end{aligned}
$$

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

Finally compare the first lines. Since $(1 + \%\, t)^{-1}$ we have that $\frac{1}{t} - \frac{1}{t} = \%$. The difference between these two equations hence is $k\%(\lambda_{2;t} - \lambda_{2;t-1})k$. In the limit as $t \to 0$, 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 $k\%(\lambda_{1;t} - \lambda_{1;t-1})k$. The optimality conditions with respect to $\hat{u}_t$ and 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

$$v_t(x): \quad 0 = \sum_{i=1}^{I} \left(\frac{\mathbb{1}_{b^n_{i,t} > 0} \; {}^{n}_{5;j;t} b^n_{i,t} \quad \mathbb{1}_{b^{n_i - 1}_{i,t} > 0} \; {}^{n_i - 1}_{5;j;t} b^{n_i - 1}_{i,t}}{x_i} + \frac{\mathbb{1}_{b^{n_i+1}_{i,t} < 0} \; {}^{n_i+1}_{5;j;t} b^{n_i+1}_{i,t} \quad \mathbb{1}_{b^n_{i,t} < 0} \; {}^{n}_{5;j;t} b^n_{i,t}}{x_i} \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{I} \frac{\left({}^{2}_{i} \right)^{n_i+1} \; {}^{n_i+1}_{5;t} + \left({}^{2}_{i} \right)^{n_i - 1} \; {}^{n_i - 1}_{5;t} \quad 2 \left({}^{2}_{i} \right)^{n} \; {}^{n}_{5;t}}{\left(x_i \right)^2}$$

$$\sum_{j=1}^{J} \sum_{i=1}^{I} \left(\frac{\mathbb{1}_{b^n_{i,t} > 0} \; {}^{n}_{6;j;t} \frac{\partial b^n_{i,t}}{\partial u^n_{j,t}} \quad \mathbb{1}_{b^{n_i - 1}_{i,t} > 0} \; {}^{n_i - 1}_{6;j;t} \frac{\partial b^{n_i - 1}_{i,t}}{\partial u^{n_i - 1}_{j,t}}}{x_i} + \frac{\mathbb{1}_{b^{n_i+1}_{i,t} < 0} \; {}^{n_i+1}_{6;j;t} \frac{\partial b^{n_i+1}_{i,t}}{\partial u^{n_i+1}_{j,t}} \quad \mathbb{1}_{b^n_{i,t} < 0} \; {}^{n}_{6;j;t} \frac{\partial b^n_{i,t}}{\partial u^n_{j,t}}}{x_i} \right)$$

$$\; {}^{n}_{5;t} \left(\frac{{}^{n}_{5;t} \quad {}^{n}_{5;t-1}}{t} \quad \% \; {}^{n}_{5;t} \right)$$

and

$$\frac{\partial L}{\partial v_t}: 0 = \; {}^{n}_{5;t} \left(\sum_{i=1}^{I} b^n_{i,t} \frac{\mathbb{1}_{b^n_t < 0} \quad \mathbb{1}_{b^n_t > 0}}{x_i} \quad \sum_{i=1}^{I} \frac{\left({}^{2}_{i} \right)^{n}}{\left(x_i \right)^2} \right)$$

$$+ \left(\sum_{i=1}^{I} \; {}^{n_i - 1}_{5;t} b^{n_i - 1}_{i,t} \frac{\mathbb{1}_{b^{n_i - 1}_{i,t} > 0}}{x_i} + \sum_{i=1}^{I} \; {}^{n_i - 1}_{5;t} \frac{\left({}^{2}_{i} \right)^{n}}{2 \left(x_i \right)^2} \right)$$

$$+ \left(\sum_{i=1}^{I} \; {}^{n_i+1}_{5;t} b^{n_i+1}_{i,t} \frac{\mathbb{1}_{b^{n_i+1}_{i,t} < 0}}{x_i} + \sum_{i=1}^{I} \; {}^{n_i+1}_{5;t} \frac{\left({}^{2}_{i} \right)^{n}}{2 \left(x_i \right)^2} \right)$$

$$+ \sum_{j=1}^{J} \sum_{i=1}^{I} \; {}^{n}_{6;j;t} \frac{\partial b^n_{i,t}}{\partial u^n_{j,t}} \frac{\mathbb{1}_{b^n_t < 0} \quad \mathbb{1}_{b^n_t > 0}}{x_i} + {}^{n_i - 1}_{6;j;t} \frac{\partial b^{n_i - 1}_{i,t}}{\partial u^{n_i - 1}_{j,t}} \frac{\mathbb{1}_{b^{n_i - 1}_t > 0}}{x_i} \quad {}^{n_i+1}_{6;j;t} \frac{\partial b^{n_i+1}_{i,t}}{\partial u^{n_i+1}_{j,t}} \frac{\mathbb{1}_{b^{n_i+1}_t < 0}}{x_i} \Bigg)$$

$$\; {}^{n}_{5;t} \quad \frac{{}^{n}_{5;t} \quad {}^{n}_{5;t-1}}{t} \quad \frac{1 \quad 1}{t} \; {}^{n}_{5;t-1} \tag{132}$$

Again these, two expressions are identical up to the last time index in the last line (${}^{n}_{5}$), and thus the difference is $k\%(\; {}^{n}_{5;t} \quad {}^{n}_{5;t-1})k:$

Next, consider the two discretized optimality conditions with respect to ${}^{n}_{t}$ (115)

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

$$_t(x): \quad 0 = \quad _{4;t} f^n_{4;t} \; " \qquad (133)$$
$$+ \sum_{i=1}^{I} b^n_{i;t} \; I_{b^n_{i;t}>0} \frac{^{n_i+1}_{7;t} \quad ^n_{7;t}}{x_i} + I_{b^n_{i;t}<0} \frac{^n_{7;t} \quad ^{n_i}_{7;t} \; 1}{x_i} \; \# + \frac{1}{2} \sum_{i=1}^{I} \; ^2_i \; ^n \; \frac{^{n_i+1}_{7;t} + \; ^{n_i}_{7;t} \; 1 \quad 2 \; ^n_{7;t}}{2(\; x_i)^2}$$
$$+ \frac{^n_{7;t} \quad ^n_{7;t} \; 1}{t} \quad \% \; ^n_{7;t}$$

$$\frac{@L}{@ \; ^n_t} : 0 \; = \quad _{4;t} f^n_{4;t} \qquad (134)$$
$$(\; + \; ^n_{7;t} \sum_{i=1}^{I} I_{b^n_{i;t}>0} \quad I_{b^n_{i;t}<0} \frac{b^n_{i;t}}{x_i} \quad \sum_{i=1}^{I} \frac{2(\; ^2_i)^n}{2(\; x_i)^2} \;)$$
$$\sum_{i=1}^{I} \; ^{n_i}_{7;t} \; 1 \frac{I_{b^n_{i;t}<0} b^n_{i;t}}{x_i} + \sum_{i=1}^{I} \; ^{n_i}_{7;t} \; 1 \frac{(\; ^2_i)^n}{2(\; x_i)^2}$$
$$+ \sum_{i=1}^{I} \; ^{n_i+1}_{7;t} \frac{I_{b^n_{i;t}>0} b^n_{i;t}}{x_i} + \sum_{i=1}^{I} \; ^{n_i+1}_{7;t} \frac{(\; ^2_i)^n}{2(\; x_i)^2}$$
$$+ \frac{^n_{7;t} \quad ^n_{7;t} \; 1}{t} \quad \frac{^1 \quad 1}{t} \; ^n_{7;t} \; _1 ;$$

which again differ in $k\%(\; _{7;t} \quad _{7;t} \; _1)k:$

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

$$u_{l;t}(x): \quad 0 = \quad _{4;t} \frac{@f_4}{@ \; ^n_{l;t}} \; ^n_t \qquad (135)$$
$$+ \sum_{j=1}^{J} \; ^n_{6;l;t} \; " \; \frac{@^2 f^n_{5;t}}{@ \; ^n_{j;t} \; @ \; ^n_{l;t}} + \sum_{i=1}^{I} \frac{@^2 b^n_{i;t}}{@ \; ^n_{j;t} \; @ \; ^n_{l;t}} \; I_{b^n_{i;t}>0} \frac{v_t^{n_i+1} \quad v_t^n}{x_i} + I_{b^n_{i;t}<0} \frac{v_t^n \quad v_t^{n_i} \; 1}{x_i} \; ! \; \#$$
$$\sum_{i=1}^{I} I_{b^n_{i;t}>0} \frac{^{n_i+1}_{7;t} \quad ^n_{7;t}}{x_i} + I_{b^n_{i;t}<0} \frac{^n_{7;t} \quad ^{n_i}_{7;t} \; 1}{x_i} \; \frac{@ \; ^n_{i;t}}{@ \; ^n_{l;t}} \; ^n_t$$

$$\frac{\partial \mathcal{L}}{\partial \mu_{i,t}} : 0 = \quad {}_{4;t}\frac{\partial f^{n}_{4;t}}{\partial \mu_{i,t}}{}^{n}_{t} \Big(+ \sum_{n}{}_{6;t} \quad \frac{\partial^{2} f^{n}_{5;t}}{\partial \mu_{j,t}\,\partial \mu_{i,t}} + \sum_{i=1}^{I} \frac{\partial^{2} b^{n}_{i;t}}{\partial \mu_{j,t}\,\partial \mu_{i,t}} \; \mathbb{I}_{b^{n}_{i;t}>0} \frac{v_t^{n_i+1} - v_t^{n}}{x_i} + \mathbb{I}_{b^{n}_{i;t}<0} \frac{v_t^{n} - v_t^{n_i-1}}{x_i} \Big) + \sum_{i=1}^{I} {}^{n}_{7;t} \quad {}^{n_i-1}_{7;t} \; \mathbb{I}_{b^{n}_{i;t}<0} \frac{1}{x_i} + \sum_{i=1}^{I} {}^{n_i+1}_{7;t} \quad {}^{n}_{7;t} \; \mathbb{I}_{b^{n}_{i;t}>0} \frac{1}{x_i} \quad \frac{\partial b^{n}_{i;t}}{\partial \mu_{i,t}} \Big]^{n}_{t},$$

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 $\%_t$. Provided that multipliers experience no jumps, the difference between the two approaches goes to 0 as $t \to 0$. 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.

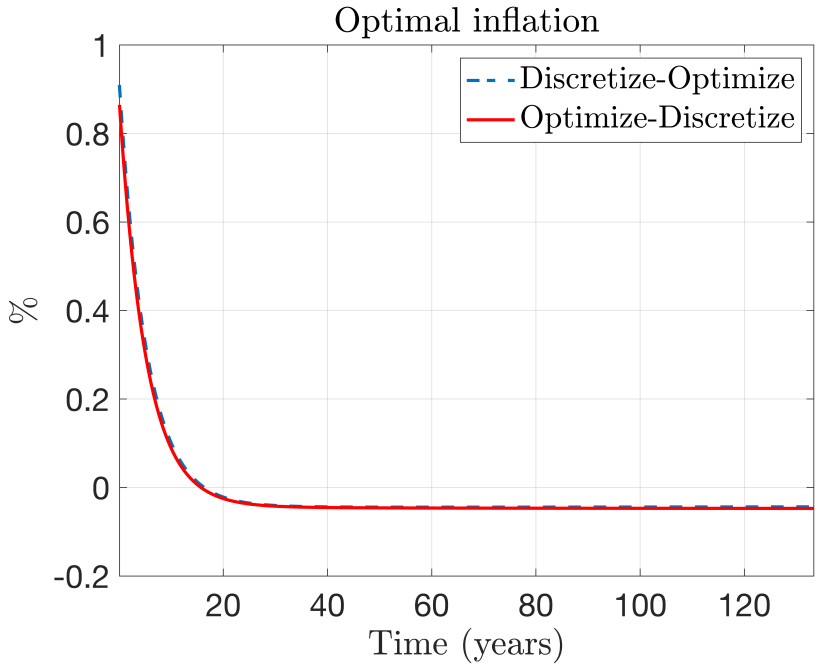

Figure 8: Time-0 optimal monetary policy using the two approaches.

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