Optimal Monetary Policy with Heterogeneous Firms

Beatriz González, Galo Nuño, Dominik Thaler†
Banco de España

March 2021

Abstract

We analyze optimal monetary policy in a New Keynesian model with heterogeneous firms. Firms differ in their productivity and net worth, and face collateral constraints which cause capital misallocation. TFP is endogenous in this economy and it depends on the productivity distribution of firms. We introduce a new algorithm to compute optimal policies in continuous-time heterogeneous-agent models. Our results show that a central bank without pre-commitments engineers an unexpected monetary expansion to increase the profits of most productive firms, allowing them to relax their financial constraints. This reduces capital misallocation and increases TFP. Contrary to the case with complete markets, 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.


*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 thank Klaus Adam, Ben Moll, Ricardo Reis and Gianluca Violante for excellent comments and suggestions, as well as participants at several seminars and conferences. All remaining errors are ours.

†Corresponding e-mails: beatrizgonzalez@bde.es (Beatriz González), galo.nuno@bde.es (Galo Nuño) and dominik.thaler@bde.es (Dominik Thaler).
1 Introduction

Interest rates have remained at historically low levels in most advanced economies since the financial crisis of 2008, as central banks attempted to raise inflation to its target. This has raised concerns among some policymakers and analysts about the effects of such an expansionary monetary policy. In their view, this policy might have detrimental effects on the allocative efficiency of the economy. These arguments were at the core of the recent sentence by the German Constitutional Court which states that, by lowering interest rates, the ECB "allows economically unviable companies to stay on the market" (German Federal Constitutional Court (2020)). If this is the case, there is a non-trivial trade-off between the traditional central bank objective of inflation stabilization and long-run aggregate productivity. But what does this trade-off imply for the optimal design of monetary policy? Should the central bank renounce to stabilize the economy after a negative shock on the grounds that it would decrease allocative efficiency? Do concerns about fueling misallocation justify a more hawkish monetary policy stance? These are the important questions that we try to answer in this paper.

To this end, we propose a continuous-time New Keynesian model with heterogeneous firms. The model features a continuum of firms with idiosyncratic time-varying productivity levels which are subject to borrowing constraints. In this environment, firms endogenously decide whether to enter or exit the market: if a firm’s productivity is above a certain threshold, the firm hires workers and invests in capital, otherwise it leaves the market. Only a fraction of firms at the top of the productivity distribution will thus operate, introducing dispersion in the marginal product of capital (MPK), i.e. capital misallocation.

Firm heterogeneity and incomplete markets change the behavior of the economy relative to its complete-markets representative-agent New-Keynesian (RANK) counterpart in a meaningful way. The borrowing constraints implies that only self-financing can undo financial frictions, such that the distribution of net worth across firms matters for allocative efficiency. While in the RANK economy aggregate TFP is exogenous, in our economy TFP evolves endogenously as a result of the heterogeneous responses of firms. Due to nominal rigidities, the central bank can influence the productivity distribution of firms operating by changing nominal rates, and hence

\footnote{Kaplan et al. (2018) refer to HANK and RANK to distinguish between heterogeneous and representative household models. Here we use the same names to differentiate between heterogeneous and representative firms.}
the level of misallocation in the economy. We name this effect as the 'misallocation channel' of monetary policy.

Finding the optimal monetary policy in models with heterogeneous firms is a challenge, as the productivity-capital distribution is an infinite-dimensional object. To overcome this problem, we propose a novel methodology to compute optimal policies in models featuring non-trivial heterogeneity. The core idea is to convert the original continuous-time, continuous-space problem into a discrete-time, discrete-space problem using a finite difference method similar to the one introduced in Achdou et al. (2017). Then we compute the first-order conditions of the Ramsey planner on the modified, finite-dimensional, problem using standard software packages for symbolic differentiation. Finally, we solve the resulting system of nonlinear equilibrium conditions in the sequence space using a Newton solver. We provide a proposition that shows how this methodology can be applied to a general class of heterogeneous-agent models. Our algorithm is easy to code using Dynare, for instance.

We analyze first time-0 optimal policy and uncover a new source of time inconsistency. Though zero inflation is optimal in the long run, the central bank engineers a temporary monetary expansion. This policy surprise increases the profits of highly-productive firms, allowing them to accumulate more capital and thus to grow. This, in turn, reduces capital misallocation and increases TFP in the medium term. Firm heterogeneity thus represents a new source of time inconsistency that is absent in the complete-market representative-firm New Keynesian model. In order to understand this result, we decompose the effects of monetary policy on misallocation into direct effects, i.e., those operating through interest rates, and indirect or general equilibrium effects, i.e., those affecting prices and wages. We show how a reduction in real rates increases misallocation by allowing less productive firms to enter the market, but this effect is more than compensated by general equilibrium effects.

We turn next to 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 TFP shock. In this case, the divine coincidence that characterizes RANK models (Gali (2008)) still holds approximately, and the central bank pursues a zero inflation policy. This is in stark contrast with the case of a cost-push shock. The prescription in the RANK model is that the central bank should lean against the wind – by tightening the monetary policy stance but tolerating some inflation in order to minimize the reduction in output. In the
case of firm heterogeneity, the central bank should instead lean with the wind. It loosens monetary policy despite the rise in inflation, as the increase in demand boost firms' profits and increases TFP, amplifying the expansionary demand effect on output. The misallocation channel of monetary policy thus makes optimal policy more dovish.

**Related literature.** This paper contributes to several strands of the literature. First, three recent papers have focused on the role of firm heterogeneity in monetary policy transmission. Ottonello and Winberry (2018) analyze the effect of monetary policy on firm investment in a model with endogenous default. They find that expansionary monetary policy causes an increase in investment both because it affects the cost of capital and because it relaxes the borrowing constraint for riskier firms. Jeenas (2020) analyzes the role of firms' balance sheet liquidity in the transmission of monetary policy to investment. study the conditions under which 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 capital misallocation through endogenous entry/exit, a channel absent in these papers. Second, and more importantly, our paper is normative, not positive. We analyze optimal monetary policy in a model with non-trivial firm heterogeneity. 2

Second, we add to the literature analyzing optimal monetary policies in models with heterogeneous agents. First, Nuño and Thomas (2016) employ calculus of variations to analyze optimal monetary policy in a model with heterogeneous agents, incomplete markets and Fisherian redistribution through long-term nominal debt. Bhandari et al. (2017) analyze optimal monetary and fiscal policies in a HANK model using perturbation techniques. Bilbiie and Ragot (2017), Acharya et al. (2019) and Le Grand et al. (2020) analyze optimal monetary policy in HANK models in which the wealth distribution is finite dimensional, and hence tractable using standard techniques. Beyond the fact that we focus on heterogeneous firms and not households, the key difference with these papers is that we introduce a new method that is both easy to code and can deal with relatively complex models with heterogeneous agents, including exogenous borrowing limits or other nonlinear features that cannot be tackled with perturbation techniques.

---

2Other strands of the literature have analyzed the links between monetary policy and firm heterogeneity through heterogeneity in markups ( e.g. Meier et al. (2020), Bilbiie et al. (2014), Andrés et al. (2021) or Baqaee et al. (2021)) or in firm-level productivity trends (e.g. Adam and Weber (2019)).
Finally, our model is related to the extensive literature on capital misallocation, and the different channels that may affect this, such as Hsieh and Klenow (2009) and 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-firm model to study how the nature of the idiosyncratic shocks impacts the speed of transitions. We build on this paper and enrich his model by introducing a New Keynesian monetary setting, since our focus is to understand how monetary policy affects aggregates through its impact on heterogeneous firms. Focusing on the impact of monetary policy 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. Here, instead, we focus on a closed-economy general equilibrium setting where real rates depends on the endogenous reaction of the central bank.

2 Model

We propose a New Keynesian closed economy model with heterogeneous good producers building on Moll (2014). Time is continuous and there is no aggregate uncertainty. The economy is populated by five types of agents: households, the central bank, input good firms, retail and final goods producers. The representative household works and saves. Input goods firms combine capital and labor to produce the input good. They are heterogeneous in their net worth and face idiosyncratic uncertainty about their productivity. The input good is then differentiated by imperfectly competitive retail goods producers, whose output is aggregated by the final goods producer. The latter two firms are standard in New Keynesian models. Retailers face sticky prices à la Rotemberg (1982).

2.1 Heterogeneous input good firms

There is a continuum of input good producers – firms for short. They are heterogeneous in their productivity $z_t$ and in their net worth, which they hold in units of capital $a_t$. Productivity is subject to idiosyncratic shocks to productivity $z_t$. Each

---

3For 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.
firm owns a technology that produces input goods with a constant returns to scale production function in capital $k_t$ and labor $l_t$,

$$y_t = f_t(z_t, k_t, l_t) = (\Gamma z_t k_t)^{\alpha}(l_t)^{1-\alpha},$$  \hspace{1cm} (1)

with $\alpha \in (0, 1)$. $\Gamma$ is aggregate productivity. 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 saving in the economy, both households and other firms. Firms receive a risk-free return of $R_t/q_t - \delta$ on their net worth $q_t a_t$. The law of motion of net worth in units of capital, $a_t$, is

$$\dot{a}_t = \frac{1}{q_t} \left[ \max\{m_t f_t(z_t, k_t, l_t) - w_t l_t - R_t k_t, 0\} + \frac{(R_t/q_t - \delta)}{\text{Return on net worth}} q_t a_t - \frac{d_t}{\text{Dividends}} \right],$$  \hspace{1cm} (2)

where $q_t$ is the real price of capital, $m_t = p_t^y / P_t$ 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. Note that profits and the return on net worth can be used either to distribute as dividends $d_t$, or for savings $\dot{a}_t$. We assume that the firms are owned by the household, hence she receives the distributed dividends. Importantly, firms can operate with more capital $k_t$ than they own $a_t$. That is they can borrow additional capital $b_t = k_t - a_t$. However, firms face a collateral constraint, such that the capital used in production cannot exceed $\gamma > 1$ of their net worth,

$$q_t k_t \leq \gamma q_t a_t.$$  \hspace{1cm} (3)

Productivity follows a diffusion process, which is the continuous time analog of a Markov process,

$$dz_t = \mu(z_t)dt + \sigma(z_t)dW_t,$$  \hspace{1cm} (4)

where $\mu(z)$ is the drift and $\sigma(z)$ the diffusion of the process. Firms close down according to an exogenous Poisson process with arrival rate $\eta$. Once they close down, they pay back all their assets, $q_t a_t$, to the households, and they are replaced by a new firm. Firms maximize their lifetime value (5), which is given by
subject to the budget constraint (2), the collateral constraint (3), and the process followed by productivity (4). Future profits are discounted at the real rate, which, as we will see later, coincides with the household’s stochastic discount factor.

We can separate the firms’ problem into two: a profit maximization problem, and a savings-dividend distribution problem. First, firms maximize profits given their productivity and net worth, subject to the borrowing constraint:

\[
\Phi_t(z_t, a_t) = \max \left\{ \max_{k_t, l_t} \left\{ \Gamma z_t k_t(z_t, a_t) - w_t l_t - R_t k_t \right\}, 0 \right\},
\]

s.t. \( q_t k_t \leq \gamma q_t a_t \). Since the production function follows constant returns to scale, factor demands and profits are linear in net worth, and there exists a productivity cut-off \( z^* \) such that the firm operates whenever \( z_t \geq z^*_t \) and remains inactive otherwise:

\[
k_t(z_t, a_t) = \begin{cases} 
\gamma a_t, & \text{if } z_t \geq z^*_t, \\
0, & \text{if } z_t < z^*_t,
\end{cases}
\]

Inactive firms still can rent the capital they own to other firms. The first order condition gives an expression for the demand of labor,

\[
l_t(z_t, a_t) = \left( \frac{(1 - \alpha) m_t}{w_t} \right)^{1/\alpha} \Gamma z_t k_t(z_t, a_t). \]

Firm’s profits are then given by

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

and the productivity cut-off is given by

\[
\Gamma z_t^* \varphi_t = R_t.
\]
Next, they decide how to split their profits between dividends $d_t$ and savings $s_t^a$. The law of motion of a firm’s net worth (in units of capital) (2) can be rewritten as

$$\dot{a}_t \equiv s_t^a(z_t, a_t, d_t) = \frac{1}{q_t} \left[ \Phi_t(z_t, a_t) + (R_t - \delta q_t) a_t - d_t \right] = \frac{1}{q_t} \left[ (\gamma \max \{ \Gamma z_t \varphi_t - R_t, 0 \} + R_t - \delta q_t) a_t - d_t \right]. \quad (11)$$

The solution to this problem is shown in Appendix A.1. There we show how firms never distribute dividends while they are active, and only once they liquidate they distribute as dividends to the household all their own capital. The intuition is simple: one unit of capital inside the firms receives at least a return of $(R_t - \delta q_t)$, while outside the firm, the return of this unit of capital is exactly $(R_t - \delta q_t)$. Since the recovery after liquidation is all the net worth of the firm, $q_t a_t$, it is always worthwhile to keep the funds inside the firm. The household collects all these dividends after the liquidation of the firm, and to keep things simple, we assume she uses a fraction $\psi$ of these dividends to finance the net worth of the new entrants, so net dividends are $(1 - \psi)$ of the net worth of exiting firms.

2.2 Final good producers

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

$$Y_t = \left( \int_0^1 \frac{\varepsilon+1}{\varepsilon} y_{j,t}^{j} dj \right)^{\frac{\varepsilon}{\varepsilon+1}} \quad (12)$$

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

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

2.3 Retailers

We assume that the 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 \( \tau \) 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 retailer \( T_t \). The adjustment costs are quadratic in the rate of price change \( \left( \frac{\dot{p}_{j,t}}{p_{j,t}} \right) \) and expressed as a fraction of output, \( Y_t \)

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

where \( \theta > 0 \). Suppressing notational dependence on \( j \), each retailers chooses \( \{p_t\} \) to maximize the expected profit stream, discounted at the real rate

\[
\int_0^{\infty} e^{-\int_0^t r_s ds} \left\{ \Pi_t (p_t) - \Theta_t \left( \frac{\dot{p}_t}{p_t} \right) \right\} dt
\]

where

\[
\Pi_t (p_t) = \left( \frac{p_t}{\tilde{m}_t} \right) \left( 1 - \frac{m_t}{\tilde{m}_t} \right) Y_t - T_t,
\]

are per-period profits gross of price adjustment costs. The symmetric solution to the pricing problem is the yields the New Keynesian Phillips curve (see Appendix A.2), which is given by

\[
\left( r_t - \frac{\dot{Y}_t}{Y_t} \right) \pi_t = \frac{\varepsilon}{\theta} (\tilde{m}_t - m^*) + \dot{\pi}_t, \quad m^* = \frac{\varepsilon - 1}{\varepsilon}.
\]

where \( \pi_t \) denotes the inflation rate \( \pi_t = \dot{P}_t/P_t \). The total profit of retailers, net of the lump sum tax, which is transferred to the households lump sum, is

\[
\Pi_t = (1 - m_t) Y_t - \frac{\theta}{2} \pi_t^2 Y_t.
\]

### 2.4 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 he takes as given. His cost function is given by \( (\iota_t + \Phi (\iota_t)) K_t \) where \( \iota_t \) is the investment rate

\[\text{footnote}{4}\] This fiscal scheme is introduced to eliminate the distortions caused by imperfect competition in steady state, as common in the optimal policy literature.

\[\text{footnote}{5}\] We assume that input good producers discount profits at the risk-free rate \( r_t \). Our rationale for this choice is that firms are owned by households, and given the lack of aggregate risk, households’ stochastic discount factor coincides with the risk free rate, as discussed below.
and $\Phi(t_t)$ is a capital adjustment cost function. He maximizes the expected profit stream, discounted at the real rate. Profits are paid to the household lump sum.

$$W_t = \max_{i_t, K_t} \mathbb{E}_0 \int_0^\infty e^{-\int_0^t r_s ds} (q_t i_t - i_t - \Phi(t_t)) K_t dt. \quad (16)$$

s.t. $\dot{K}_t = (t_t - \delta) K_t. \quad (17)$

The optimality conditions imply (see Appendix A.3)

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

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

$$\Phi(t_t) = \frac{k}{2} (t_t - \delta)^2. \quad (18)$$

### 2.5 Households

There is a representative household that supplies labor $L_t$ and saves in capital $D_t$ or in nominal instantaneous bonds whose real value is denoted by $B_t^N$. The household maximizes

$$W_t = \max_{C_t, L_t, B_t^N, D_t} \mathbb{E}_0 \int_0^\infty e^{-\rho h t} u(C_t, L_t) dt. \quad (19)$$

s.t. $\dot{D}_t q_t + \dot{B}_t^N + C_t = (R_t - \delta q_t) D_t + (i_t - \pi_t) B_t^N + w_t L_t + T_t, \quad (20)$

and $T_t$ are the profits received by the household, which is the sum of the profits from retail goods producers, the profits of the capital producer and the dividends received from input-good firms.

We assume separable utility of CRRA form, i.e., $u(C_t, L_t) = \frac{C_t^{1-\xi}}{1-\xi} - \frac{Y^{1+\theta}}{1+\theta}$. Solving this problem (see Appendix A.4 for detail), we obtain the Euler equation, the labor supply condition and the Fisher equation, respectively:
\[
\frac{\dot{C}_t}{C_t} = \frac{r_t - \rho_t^h}{\zeta},
\]
\(21\)

\[
w_t = \frac{\Upsilon L_t^\theta}{C_t^{-\sigma}},
\]
\(22\)

\[
r_t = i_t - \pi_t,
\]
\(23\)

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

\[
r_t \equiv \frac{R_t - \delta q_t + \dot{q}_t}{q_t}.
\]
\(24\)

2.6 Distribution

We assume that for each dying firm, a new firm is born with the same productivity level. This new firm receives a startup capital stock from the household in a lump sum fashion. As previously explained, we assume that the initial net worth of each new firm is equal to a fraction \(\psi < 1\) of the net worth of the firm it replaces. The joint distribution of net worth and productivity is then given by the Kolmogorov Forward equation

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

(25)

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

Note that we can express the distribution also in terms of net-worth shares defined as \(\omega_t(z) \equiv \frac{1}{A_t} \int_0^\infty a g_t(z, a) da\). Given this definition and the structure of the problem, wealth shares are non-negative, bounded, continuous, once differentiable everywhere and they integrate up to 1. The law of motion of net-worth shares is given by (see in Appendix A.5)
\[ \frac{\partial \omega_t(z)}{\partial t} = \left[ s_t(z) - \frac{\dot{A}_t}{A_t} (1 - \psi) \eta \right] \omega_t(z) - \frac{\partial}{\partial z} \mu(z) \omega_t(z) + \frac{1}{2} \frac{\partial^2}{\partial z^2} \sigma^2(z) \omega_t(z). \] 

(26)

2.7 Market Clearing and Aggregation

Borrowing of an input good firm is the extra capital used for production on top of their net worth, \( b_t = k_t - a_t \), where \( b_t > 0 \) if the firm is borrowing and \( b_t < 0 \) if it is saving. Adding up, we get

\[ \int k_t(z,a) dG_t(z,a) = \int b_t(z,a) dG_t(z,a) + \int a_t dG_t(z,a), \] 

(27)

Asset market clearing requires that net borrowing of firms, \( B_t \), equals net savings of the household, \( D_t \),

\[ B_t = D_t. \] 

(28)

Let \( \Omega(z) \) be the cumulative distribution of net-worth shares, i.e. \( \Omega_t(z) = \int_0^z \omega_t(x) \, dx \).

By combining equations (27), (28), aggregating capital used by firms, and solving for \( A_t \), we obtain

\[ A_t = \frac{D_t}{\gamma (1 - \Omega(z^*_t)) - 1}. \] 

(29)

Labor market clearing implies

\[ L_t = \int_0^\infty l_t(z,a) dG_t(z,a). \] 

(30)

Output is given by

\[ Y_t = \tilde{Z}_t K_t^\alpha L_t^{1-\alpha}, \] 

(31)

where the TFP term \( \tilde{Z}_t \) is

\[ \tilde{Z}_t = (\Gamma E [z \mid z > z^*_t])^\alpha. \] 

(32)

Factor prices are
\[ w_t = (1 - \alpha) m_t \tilde{Z}_t K_t^{\alpha} L_t^{-\alpha}, \]  
(33) 

\[ R_t = \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1 - \alpha} \frac{z_t^*}{\mathbb{E}[z | z > z_t^*]}. \]  
(34)

We could equivalently write equation (34) in terms of real rate of return \( r_t \),

\[ r_t = \frac{1}{q_t} \left( \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1 - \alpha} \frac{z_t^*}{\mathbb{E}[z | z > z_t^*]} \right) - \delta + \hat{q}. \]  
(35)

Finally, the law of motion of aggregate wealth of firms is given by

\[ \frac{\dot{A}_t}{A_t} = \frac{1}{q_t} \left[ \gamma (1 - \Omega(z_t^*)) \left( \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1 - \alpha} - R_t \right) + R_t - \delta q_t - q_t (1 - \psi) \eta \right]. \]  
(36)

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

### 2.8 Central Bank

The central bank controls nominal interest rates \( i_t \) on nominal bonds held by households. The central bank solves the following Ramsey problem

\[ \max_{\{g(a,z), s(z), w,r,\psi,q,K,A,L,C,D,Z,X,z^*,\pi,m,i,Y,T\}} \mathbb{E}_0 \int_0^\infty e^{-\rho t} u(C_t, L_t) dt \]

subject to the private equilibrium conditions, described in detail in Appendix A.7. Notice that \( g_t(a,z) \) and \( s_t(z) \) not only depend on time, but also on the idiosyncratic state variables. This poses some difficulties when solving optimal monetary policy. In the next section we deal with them in a general environment.

### 2.9 The link between misallocation and real rates

The financial friction in the firm sector implies that capital is not optimally allocated. 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 \( \tilde{Z}_t \) fluc-
tuates over time and, importantly, depends on monetary policy. As a first step to understand how monetary policy interacts with firm heterogeneity, it is useful to understand how aggregate TFP depends on the real interest rate, that is, how monetary policy directly affects misallocation.

As we show next, holding everything else constant, a decrease in real interest rates increases misallocation. This result is intuitive: lower costs of capital makes production cheaper and crowds in firms that would otherwise find it unprofitable to operate. This result was first illustrated numerically in by Gopinath et al. (2017). In our simpler framework, we can prove it analytically: a fall in real interest rates decreases the productivity cut-off $z^*$, which, in turn, induces a decline in aggregate TFP.

To see this, we plug the definition of $z^*$ into the definition of TFP (32), and take the partial derivative of TFP with respect to $r_t$, holding the other prices constant ($\varphi_t = \varphi, q_t = q, \dot{q} = 0$):

$$\frac{\partial \bar{Z}_t}{\partial r_t} = \frac{\partial z^*_t}{\partial r_t} \frac{\partial \bar{Z}_t}{\partial z^*_t} = \alpha \left( \Gamma \mathbb{E}[z | z > z^*_t] \right)^{\alpha-1} \frac{q \omega(z^*_t)}{\varphi (1 - \Omega(z^*_t))} (\mathbb{E}[z | z > z^*_t] - z^*_t) \geq 0. \quad (37)$$

The derivative of TFP with respect to the interest rate is always non-negative, and it is strictly positive as long as the distribution $\omega(z)$ is non-zero for $z > z^*_t$. This means that, ceteris paribus, if interest rates decrease so does TFP. Note that the term $(\mathbb{E}[z | z > z^*_t] - z^*_t)$ is a measure of the dispersion of productivity of active firms: the larger the difference between the average productivity of active firms and the cut-off productivity, the larger the impact of a change in interest rates is.

However, in general equilibrium, the response of TFP depends not only on the direct effect of monetary policy on the real rate $r_t$, but also the indirect effects on the price of capital $q_t$, wages $w_t$ and relative prices $m_t$—which change the value of $\varphi_t$—and the evolution of the net worth distribution. We discuss how these general equilibrium forces shape the central bank’s optimal policy in section 5.

---

6This mechanism linking aggregate TFP and monetary policy differs from the one in Benigno and Fornaro (2018) or Moran and Queralto (2018), who focus instead on endogenous R&D.
3 Computing optimal policies in heterogeneous agents models

Solving optimal policy in models with heterogeneous agents poses a certain challenge, since the state in such 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 the 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.

We start reviewing the three existing approaches to analyze optimal monetary policy in models with non-trivial heterogeneity. Le Grand et al. (2020) employ the finite-memory algorithm originally 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. Bhandari et al. (2017), instead, make the continuous cross-sectional distribution finite dimensional by assuming that there are \( N \) agents instead of a continuum. They then derive standard first-order conditions (FOCs) for the planner. In order to cope with the large dimensionality of their problem, they employ a perturbation technique. This precludes the use of their algorithm to problems with kinks, such as the one presented here, or with exogenous borrowing limits, as in the standard Aiyagari-Bewley-Hugget framework. 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.\(^7\) They then solve the problem using the upwind finite-difference method of Achdou et al. (2017). The problem with this method is that it requires solving by hand the first-order conditions, which can be demanding in medium-scale models such as the one presented in this paper.

\(^7\)The Lagrange multipliers associated to the HJB are states that evolve according to a law of motion similar to the KF equation, whereas those associated with the KF equation are controls that evolve according to a HJB-like equation, which can be interpreted as the value function that the planner assigns to each agent.
The algorithm proposed in this paper is distinct from the previous ones. If any, it can be seen as the mirror image of Nuño and Thomas (2016). Instead of first computing by hand the planner’s FOCs and then discretizing them using finite differences, we propose to first discretize the private equilibrium conditions using finite differences and then to find the planner’s FOCs by symbolic differentiation. This avoids the cumbersome mathematical derivations and allows us to solve the dynamic problem nonlinearly in a few seconds using Dynare.

(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 $\mu_t$ at period $t$ are given by

$$
\mu_t = \Pi_t^{-1} \mu_{t-1} \text{ where } \Pi_t \equiv (I - \Delta t A_t^T),
$$

(38)

where $\Pi_t$ is the Markov matrix of the discretized states, $\Delta t$ is the time step between periods and $A_t$ is a matrix whose elements depend nonlinearly and in closed form on the idiosyncratic and aggregate variables at period $t$.\(^8\) Similarly, the HJB equation is approximated as

$$
v_t = u_{t+1} \Delta t + \beta \Pi_{t+1} v_{t+1},
$$

(39)

which is the Bellman equation of the discretized problem, with a discount $\beta \equiv (1 - \rho \Delta t)$.\(^9\) Together with some additional static equations, such as market clearing conditions or budget constraints, and some aggregate dynamic equations, including the Euler equations of representative agents (if any) and 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 (38) and Bellman equation (39) present two advantages with respect 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

\(^8\)Technically, this matrix results from the discretization of the infinitesimal generator of the idiosyncratic states. In the example of Section 2, $\mu_t = \omega_t$ and $A_t = B_t$.

\(^9\)This comes from the discretized HJB equation $\rho v_{t+1} = u_{t+1} + A_{t+1} v_{t+1} + (v_{t+1} - v_t) / \Delta t$, assuming that $A_{t+1} / \beta \simeq A_t$.  

15
step at a time” making the matrix $\Pi_t$ sparse. 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 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 agents 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 the Lagrange multipliers feature no jump for $t > 0$, the solution of the "discretize-optimize" and the "optimize-discretize" algorithms converge to each other as the time step $\Delta t$ goes towards 0. The maximum error between the two solutions at each moment $t$ is $\| \varrho (\lambda_t - \lambda_{t-\Delta t}) \|_{\infty}$, where $\lambda_t$ is the vector of Lagrange multipliers associated to dynamic equations and $\varrho$ the discount factor of the planner.

Proof: See appendix C.

The proposition guarantees that both strategies coincide when $\Delta t$ 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 heterogeneous 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 general case, as well as to showcase its generality in dealing with different problems, in Appendix C we compute the optimal monetary policy in the HANK model of Nuño and Thomas.
(2016) using our method in Dynare. We compare our results with those using their "optimize-discretize" algorithm and conclude that both approaches essentially coincide when the model is solved at monthly frequency $\Delta t = 1/12$.

(iii) Newton algorithm to solve the optimal policy problem non-linearly in the sequence space Finally, we use our discretized planner’s optimality conditions to compute the optimal responses to "MIT shocks" non-linearly using a Newton algorithm. That is, we compute transitional dynamics, but unlike it is common in the literature, we do not iterate over guesses for aggregate sequences that are used together with time iterations. Instead 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 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.\footnote{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.} 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.\footnote{To find the steady state, we provide Dynare with the steady state of the private equilibrium conditions as a function of the policy instrument.} 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 a temporary change in parameters (a "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 different shocks.

Finally, it is important to highlight that our solution approach is different to 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 equation 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 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.

4 Calibration

We solve the model using the method described above under a illustrative calibration. Table 1 summarizes our calibration. We work at quarterly frequency (time period $\Delta t = 1/4$). The rate of time preference of the household $\rho_h$ is 0.025, which targets an average real rate of return of 2.5%. The capital depreciation rate $\delta$ is set at 0.065, equal to the aggregate depreciation rate in NIPA. The fraction of assets of exiting firms reinvested in new firms ($\psi$) is 0.1, so that average size of entrants is 10% that of incumbents, in line with US data. Firms’ death rate ($\eta$) is 0.12, which together with $\psi$, implies an average real return on equity of 0.11, similar to that of the S&P500 from 2009 to 2019. The borrowing constraint parameter $\gamma$ is 1.43, implying that entrepreneurs can borrow up to 43% of their net worth, which targets the level of aggregate corporate debt as percentage of their net worth of the US from 2009 to 2019. The capital share $\alpha$ is set at a standard value of 0.3. We assume log-utility in consumption ($\eta = 1$) and the inverse Frisch elasticity $\vartheta$ to be 1. We set the constant multiplying the disutility of labor $\Upsilon$ such that aggregate labor supplied in SS is equal to 1. Capital adjustment costs, $\phi^k$, are set to 10, such that the response of investment to output a monetary policy shock at its peak 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 $\epsilon$ is set to 10, so that the steady state mark-up is $1/(1-\epsilon) = 0.11$. The Rotemberg
cost parameter $\theta$ is set to 100, so that the slope of the Phillips curve is $\epsilon/\theta = 0.1$ as in Kaplan et al. (2018)

### Table 1: Calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Source/target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho^{hh}$ Rate of time preference of HH</td>
<td>0.025</td>
<td>Av. 10Y bond return of 2.5% (FRED)</td>
</tr>
<tr>
<td>$\delta$ Capital depreciation rate</td>
<td>0.065</td>
<td>Aggregate depreciation rate (NIPA)</td>
</tr>
<tr>
<td>$\psi$ Fraction firms’ assets at entry</td>
<td>0.1</td>
<td>Av. size at entry 10% (OECD)</td>
</tr>
<tr>
<td>$\eta$ Firms’ death rate</td>
<td>0.12</td>
<td>Av. real return on equity 11% (S&amp;P500)</td>
</tr>
<tr>
<td>$\gamma$ Borrowing constraint parameter</td>
<td>1.43</td>
<td>Corporate debt to net worth of 43% (FRED)</td>
</tr>
<tr>
<td>$\alpha$ Capital share in production function</td>
<td>0.3</td>
<td>Standard</td>
</tr>
<tr>
<td>$\zeta$ Relative risk aversion parameter HH</td>
<td>1</td>
<td>Log utility in consumption</td>
</tr>
<tr>
<td>$\vartheta$ Inverse Frisch Elasticity</td>
<td>1</td>
<td>Kaplan et al. (2018)</td>
</tr>
<tr>
<td>$\Upsilon$ Constant in disutility of labor</td>
<td>0.71</td>
<td>Normalization</td>
</tr>
<tr>
<td>$\phi^k$ Capital adjustment costs</td>
<td>10</td>
<td>VAR evidence</td>
</tr>
<tr>
<td>$\epsilon$ Elasticity of substitution retail goods</td>
<td>10</td>
<td>Mark-up of 0.11</td>
</tr>
<tr>
<td>$\theta$ Price adjustment costs</td>
<td>100</td>
<td>Slope of PC of 0.1</td>
</tr>
<tr>
<td>$\Gamma$ SS Aggregate Productivity</td>
<td>1</td>
<td>Normalization</td>
</tr>
<tr>
<td>$\varsigma_z$ Mean reverting parameter</td>
<td>0.8</td>
<td>Quarterly persistence Gilchrist et al. (2014)</td>
</tr>
<tr>
<td>$\sigma_z$ Volatility of the shock</td>
<td>0.15</td>
<td>Quarterly volatility Gilchrist et al. (2014)</td>
</tr>
</tbody>
</table>

Aggregate productivity term $\Gamma$ is normalized to 1 in SS. We assume that individual productivity $z$ follows an Orstein-Uhlenbeck process in logs

$$d \log(z) = -\varsigma_z \log(z) dt + \sigma_z dW_t. \quad (40)$$

By Itô’s lemma, this implies that $z$ in levels follows the diffusion process

$$dz = \mu(z) dt + \sigma(z)dW_t, \quad (41)$$

where $\mu(z) = z \left( -\varsigma_z \log z + \frac{\sigma_z^2}{2} \right)$ and $\sigma(z) = \sigma_z z$. We calibrate the productivity process using the estimates from Gilchrist et al. (2014), who find quarterly persistence of 0.8 and volatility of 0.15 (0.3 annualized).

### 5 Quantitative Results

We now analyze optimal policy under commitment, i.e. we solve the central bank’s Ramsey problem. We compare the optimal responses in our heterogeneous-firm
problem (HANK) to the responses in the representative-agent version model (RANK). For simplicity we introduce a subsidy on production factors that undoes the New Keynesian mark-up distortion in steady state in both economies. The RANK economy is the standard New Keynesian model with capital. It is a special case of the HANK economy where the borrowing constraint is set to infinite, so that the firm 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 HANK economy, in which the net worth distribution across firms matters due to financial frictions and determines the endogenous component of TFP (see Appendix A.8 for more details regarding the RANK versus HANK model). We stress the fact that the central bank only instrument is the nominal interest rate. The way monetary policy affects real allocations is through its impact on prices in the New Keynesian Phillips curve.

We start by analyzing how the central bank would behave if it is allowed to re-optimize without pre-commitments, starting at the steady state of the Ramsey solution. This is the "time 0 optimal policy" (Woodford (2003)). Next, we analyze the optimal policy response when an unexpected TFP shock or mark-up shock hits the economy that was previously in its steady state. In this case we adopt a "timeless perspective". The timeless perspective means that central bank cannot exploit the initial state of the economy, but rather stick 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.

Before we get to the dynamics, we analyze the steady state of the Ramsey problem. It is well known that the RANK economy features zero inflation in steady state, since the zero inflation steady state is first best. We find numerically that, for our calibration and several robustness checks, the HANK economy also features zero inflation in the steady state of the Ramsey problem. Our reading of this results is that, though the long-run Phillips curve allows the monetary authority to modify the relative factor prices and thus the degree of capital misallocation in steady state, it is not optimal to do so given the price adjustment costs and the markup in the retail sector.
5.1 Time-0 optimal policy

Firm heterogeneity causes a time inconsistency problem that does not exists in the standard RANK. Figure 1 shows this time inconsistency problem: starting at the steady state of the Ramsey problem, if the central bank is allowed to re-optimize, it will take advantage of the lack of pre-commitments and would engineer a monetary expansion. It does so by reducing the nominal rate (not shown) which leads to a reduction in the real rate (red dotted line, Panel 5) and an increase in inflation and output (Panel 1 and 9) through the standard New Keynesian channels. In the RANK, absent capital misallocation, the steady state is first best, so the central bank does nothing. But why exactly is it optimal to engineer an expansion in HANK?

The surprise expansion is socially optimal because it shifts factor prices in such a way that the allocation of resources improves, which leads to a temporary increase in aggregate TFP (Panel 7). This improvement is the result of two forces: First, the threshold $z^*$ moves up, making less productive firms abstain from producing. Second, firms’ profits increase for the most productive firms, such that they can accumulate more net worth, partially undoing financial frictions. Both effects lead to an increase in endogenous TFP, which amplifies the boom generated by the initial monetary policy expansion.

Next we dig in the heterogeneity of the responses that drive the aggregate responses just explained. We define the firm’s excess return on capital $\tilde{\Phi}_t(z)$ as

$$\tilde{\Phi}_t(z) \equiv \frac{\Phi_t}{k_t} = \max \left\{ \Gamma z_t \alpha \left( \frac{(1 - \alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{1/\alpha} - \frac{\dot{q}_t}{R_t}, 0 \right\},$$

where the latter equality comes from equation (9). We speak of the 'excess' return here since it is the return that a firm makes in excess of the remuneration of its own net worth ($R_t/q_t - \delta$) $q_t a_t$. The blue solid line of Panel 1 in Figure 2 shows the excess return on capital $\tilde{\Phi}_{SS}(z)$ in steady state. For low values of productivity, this value is 0, since these firms prefer not to operate. From $z^*$ onwards, firms operate, and their profits increase linearly in productivity. The green solid line shows the excess return on capital looks like one year after the implementation of the time 0 optimal policy. The threshold for operating has moved to the right, and the slope of the return function $\tilde{\Phi}_t(z)$ has increased: this implies that profits increase relatively more
Figure 1: Time 0 optimal monetary policy.

Notes: The figure shows the deviations from steady state of the economy when the planner is allowed to reoptimize with no precommitments in response to no shock. RANK is solid blue line, HANK is orange dotted line.
for more productive firms. Hence, more productive firms can undo financial frictions faster, operating at a larger scale; while at the same time less productive firms find it desirable to stop operating and rent their net worth to the active firms. This, in turn, decreases the net worth share held by less productive firms, and increases that of the most productive firms, hence improving the allocation of resources. Panel 2 of Figure 2 displays the deviations of net worth shares, \[
\left[ \omega_1(z) - \omega_{SS}(z) \right] / \omega_{SS}(z),
\]
after 1 year. Firms with productivities slightly above 1 see their shares increase, whereas those below that threshold experience a decline. As a result, production now concentrates more on high productivity firms.

Panel 1 of Figure 2 also shows a decomposition of the impact of each of the prices on \( \Phi(z) \) one year after the shock. The decrease in the real rate \( r_t \) (orange dotted line) shifts the return function parallel to the left. The reduction in the real rate makes capital cheaper ceteris paribus, which increases profits and crowds in less productive firms as we already showed analytically in section 2.9. The changes in the price of capital \( q_t \) has the exact opposite effect (yellow dashed line). The increase in wages \( w_t \) (purple dashed-dotted line) both shifts the kink of the return function to the right and decreases its slope, \( \frac{\partial \delta}{\partial w} < 0 \). This reflects the reduction in returns as wages increase. However, the increase in relative prices \( m_t \) shifts \( z^* \) to the left, and increases the slope of the return function significantly, \( \frac{\partial \delta}{\partial m} > 0 \). As relative prices increase, firms’ returns go up, specially those of high-productivity firms. Which of these channels prevail is a quantitative question. In our particular case, the result (green solid line) is a shift to the right, implying a rise in \( z_t^* \) and an increase in the slope of the profit function. As commented above, the higher \( z_t^* \) implies that less productive firms (those with productivities \( z_{ss}^* < z < z_{t=1}^* \)) stop operating, while the increase in the slope means that more productive firms see their profits rise.
Panel 1 - Excess return on capital \( \Phi_1(z) \) after 1 year

Panel 2 - Deviations of net-worth shares after 1 year

Figure 2: Heterogeneity 1 years after the shock hits.

Notes: Panel 1: Idiosyncratic productivity is shown on the X-axis, the excess return on capital \( \Phi(z) \) on the Y-axis. The solid blue line is the return function in the SS, and solid green line is the same function in year 1. The rest of the lines show the excess return function when only one price is changed at a timeto its year 1 value, keeping the rest of the prices constant to the SS value. Panel 2: deviations from steady state of the net-worth shares for each idiosyncratic productivity level \( z \) 1 year after the shock, i.e. \( \omega_{t+1}(z) - \omega_{SS}(z) \).

The bottom line is that, by expanding demand through a more accommodative monetary policy stance, the central bank increases the share of production carried out by high-productivity firms, which in turn raises TFP.

5.2 Timeless optimal policy response to a TFP shock

We turn next to the optimal response to shocks. Figure 3 shows the optimal timeless response of the central bank when there is an sudden unexpected decrease in exogenous TFP (\( \Gamma \)) of 10% that is mean-reverting, with yearly persistence of 0.8. Each panel shows the response of different equilibrium variables, and each line in a different economy: the solid blue line is the optimal response in the representative-agent model (RANK), the dotted orange one is the optimal response of our heterogeneous-firm model (HANK), and the dashed yellow line is the response of the HANK economy if we assume that the central bank implements the same (suboptimal) path of inflation as in RANK, i.e. zero inflation at all times.
Figure 3: Optimal monetary policy response to a TFP shock.

Notes: The figure shows the optimal response from a timeless perspective (in deviations from steady state) to a 10% decrease in exogenous TFP $\Gamma$ that is mean reverting with a yearly persistence of 0.8. RANK is solid blue line, HANK is orange dotted line, the response of the HANK model when fed exogenously the path of $\pi$ obtained in the optimal policy of RANK is the yellow dashed-dotted line.

As the solid blue lines show, the *divine coincidence* holds in the RANK economy: there is no trade-off between the stabilization of inflation and the stabilization of the output gap (Gali (2008)). The responses of prices and quantities are simply the reaction to the exogenous fall in TFP, that reverts to its long run average.

In the HANK economy (dashed orange line), the divine coincidence *nearly* holds: inflation only deviates from its steady state value of 0 slightly at the beginning, which translates in a slight decrease in the relative price of the input good (or the inverse markup). Note that due to the decrease in the exogenous component of TFP, the threshold $z^*$ moves to the right (see equation 10), and hence the endogenous TFP component increases (Panel 7 of Figure 3). The exogenous fall in aggregate TFP decreases proportionally more the profits of the most productive firms (see the yellow dashed-dotted line of Panel 1, Figure 4). The endogenous change in prices partly undo this negative impact of TFP on profits (orange dotted line of

---

12Output gap is defined as deviations of output from the efficient output level.
Panel 1 - Excess return on capital by \( z \) after 5 years

Panel 2 - Deviations of net-worth shares

Notes: Panel 1: Idiosyncratic productivity is shown on the X-axis, the excess return on capital \( \Phi(z) \) on the Y-axis. The solid blue line is the return function in the SS, and solid green line is the same function in year 5. The orange dotted line shows the return if only endogenous prices change as they do in year 5, but keeping the exogenous TFP component constant at the SS value. The yellow dashed line shows profits if only the exogenous productivity component changes as it does in year 5, but keeping all the prices constant at their SS values. Panel 2: deviations from steady state of the net-worth shares of each idiosyncratic productivity level \( z \) 5 years after the shock, i.e. \( \omega_t(z) - \omega_{SS}(z) \).

Panel 1, Figure 4). In net, the shock and the changes in prices increase slightly the productivity threshold for being active \( z^* \), but decreases the slope of the profit function, hence more productive firms are relatively worst after the shock. This translates into a lower net-worth share of the most productive firms, but a larger net-worth share of the least productive firms (Panel 2, Figure 4). If we plug the optimal path of inflation in the RANK model (which is 0 in this case) into the HANK model, there are nearly no differences with the HANK model responding optimality, due to this divine coincidence nearly holding in this case.

5.3 Timeless optimal policy response to a cost-push shock

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 (\( \epsilon \)) of 10% that is mean-reverting with yearly persistence of 0.8. This shock increases retailers’ markup, reducing the price of the goods sold by the heterogeneous firms.
The blue solid lines if Figure 5 show the optimal response in the RANK model to this cost-push shock: since a positive mark-up shock has inflationary pressures, the central bank reacts by tightening monetary policy, thus driving output below its efficient level with the objective of dampening inflation, the well-known *leaning against the wind* Gali (2008).

However, the optimal response of the monetary authority is very different in the HANK model, since it follows a *leaning with the wind* policy, that is, instead of containing inflation at the cost of a fall in output, it allows inflation to rise well above the optimal RANK inflation such that output actually increases (see orange dotted lines in Figure 5). The intuition and mechanisms behind this are very close to those explained in Section 5.1. By increasing inflation, the central bank is able to generate a demand expansion, increasing relative prices, but also wages and real rental rates. This increases the excess returns to capital particularly for the most productive firms, allowing them to undo financial frictions faster, which increases endogenous TFP, aggregate capital and aggregate output.

If instead the central bank were to target the path for inflation that is optimal in RANK (yellow dashed-dotted line), the behaviour of the economy would be significantly different: instead of an expansion, aggregate output would fall below its steady state value, and so would profits, capital and endogenous TFP.

Introducing heterogeneous firms thus affects significantly the timeless response to markup shocks, but not to TFP shocks. The intuition is simple. In the RANK model the planner pursues two objectives: to minimize the costs of inflation and to minimize the deviation of output from the efficient level. In the case of TFP shocks the planner faces no trade-off, as zero inflation minimizes both costs. In the face of markup shocks there is however a trade-off, and optimal policy carefully balances the two opposing objectives. Adding firm heterogeneity adds a third objective: to reduce capital misallocation. This third objective is too weak quantitatively to significantly influence optimal policy when both of the traditional objectives are aligned, as it is the case of the TFP shock. But it is strong enough to shift the fragile balance of the two objectives in the case of the cost push shock. And it does so in the direction of less inflation stabilization.
6 Conclusions

This paper analyzes optimal monetary policy in a model with heterogeneous firms, financial frictions and nominal rigidities. The model features a link between monetary policy and capital misallocation.

We identify a new source of time-inconsistency in monetary policy: Though zero inflation is optimal in the long run, a benevolent central bank without pre-commitments engineers a temporary surprise expansion. It does so because the surprise expansion modifies equilibrium prices in such a way that capital misallocation is reduced. This is despite the fact that a drop in the real rate has an unambiguously negative direct effect on capital misallocation by crowding in less productive firms. However, the associated changes in other equilibrium prices favor highly-productive firms, allowing them to increase their share of total output. Overall the latter indirect effects dominate and the capital allocation becomes more efficient.
We also illustrate how the central bank’s optimal response to shocks can be similar, or radically different, depending on the nature of the shock, to the case with complete markets. When an exogenous TFP shock hits, the divine coincidence still holds approximately. However, when faced with a cost-push shock, the optimal prescription in this case is to lean with the wind, tolerating more inflation in exchange for a boom in demand that will raise TFP further down the road.

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 shaping optimal monetary policy, 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.

The paper also makes what we deem as an important methodological contribution. It introduces a new algorithm to compute optimal policies in heterogeneous-agent models using Dynare. The algorithm leverages on the numerical advantages of continuous time and will allow researchers to solve optimal policy in heterogeneous agents models with or without aggregate shocks in an efficient and simple way. It is our hope that this will spur a new wave of research on the normative implications of heterogeneous-agent models in the years to come.
References


Appendix

A Further details on the model

A.1 Firm’s intertemporal problem

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

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

We guess and verify a value function of the form

\[ V_t(z,a) = \kappa_t(z) q_t a. \]

The first order condition is

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

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

\[ (r_t + \eta) \kappa_t(z) q_t = \eta q_t + (\gamma \max \{ \Gamma z_t \varphi_t - R_t, 0 \} + R_t - \delta q_t) \kappa_t(z) + \mu(z) q_t \kappa_t \frac{\partial \kappa_t}{\partial z} + \frac{\sigma(z)^2}{2} q_t \frac{\partial^2 \kappa_t}{\partial z^2} + \frac{\partial (q_t \kappa_t)}{\partial t}. \]

(42)

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

**Proof.** The drift of the firms capital holdings is

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

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

\[ E_0 a_t = E_0 a_0 e^{\int_0^t s_u^0 du} > a_0 e^{\int_0^t \frac{R_s - \delta q_s}{q_s} ds}. \]

(43)
The value function is then
\[
\kappa_{t_0}(z)q_{t_0}a_{t_0} = V_{t_0}(z, a_{t_0}) = \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^{t} r_s \, ds} \left( d_t + \eta q_t a_t \right) \frac{R_s - \delta q_t + q_s}{q_s} ds \\
\geq \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^{t} r_s \, ds} \eta q_t a_t dt = \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^{t} \left( \frac{R_s - \delta q_t + q_s}{q_s} + \eta \right) ds} \eta q_t a_t dt \\
= \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^{t} \left( \frac{R_s - \delta q_t + q_s}{q_s} + \eta \right) ds} \eta q_t a_t dt \eta q_{t_0} a_{t_0} dt = \mathbb{E}_{t_0} \int_0^\infty e^{-\int_{t_0}^{t} \left( \frac{R_s - \delta q_t + q_s}{q_s} + \eta \right) ds} \eta q_{t_0} a_{t_0} dt,
\]
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 24 and in the last line the inequality (43). Hence \(\kappa_{t_0}(z) > 1\) for any \(t_0\).

A.2 New Keynesian Philips curve

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

\[
r(t)V^r(t, p) = \max_{\pi} \left( \frac{p - P^y(t)(1 - \tau)}{P(t)} \right)^{-\varepsilon} \left( \frac{p}{P(t)} \right)^{-\theta} Y(t) - \frac{\theta}{2} \pi^2 Y(t) + \pi p \frac{\partial V^r}{\partial p} + \frac{\partial V^r}{\partial t},
\]

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

\[
\theta \pi Y(t) = p \frac{\partial V^r}{\partial p},
\]

\[
(r - \pi) \frac{\partial V^r}{\partial p} = \left( \frac{p}{P(t)} \right)^{-\varepsilon} Y(t) - \varepsilon \left( \frac{p - P^y(t)(1 - \tau)}{P(t)} \right) \left( \frac{p}{P(t)} \right)^{-\varepsilon - 1} Y(t) + \pi p \frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p}.
\]

In a symmetric equilibrium we will have \(p = P\), and hence

\[
\frac{\partial V^r}{\partial p} = \frac{\theta \pi Y(t)}{p},
\]

\[
(r - \pi) \frac{\partial V^r}{\partial p} = \frac{Y(t)}{p} - \varepsilon \left( \frac{p - P^y(t)(1 - \tau)}{p} \right) \frac{Y(t)}{p} + \pi p \frac{\partial^2 V^r}{\partial p^2} + \frac{\partial^2 V^r}{\partial t \partial p}.
\]

35
Deriving (44) with respect to time gives

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

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

\[
\left( r - \frac{\dot{Y}}{Y} \right) \pi = \frac{1}{\theta} \left( 1 - \varepsilon \left( 1 - \frac{P_y(t)}{p} (1 - \tau) \right) \right) + \dot{\pi}.
\]

Finally, rearranging we obtain the New Keynesian Phillips curve

\[
\left( r - \frac{\dot{Y}}{Y} \right) \pi = \frac{\varepsilon}{\theta} \left( \frac{1 - \varepsilon}{\varepsilon} + \ddot{m}(t) \right) + \dot{\pi}.
\]

### A.3 Capital producers

The problem of the capital producer is

\[
W_t = \max_{i_t, K_t} \mathbb{E}_0 \int_0^\infty e^{-\int_0^t r_s ds} (q_t i_t - \Phi (i_t)) K_t dt.
\]

\[\dot{K}_t = (i_t - \delta) K_t,\]  \hspace{1cm} (45)

We can construct the Hamiltonian

\[
H = (q_t i_t - \Phi (i_t)) K_t + \lambda_t (i_t - \delta) K_t
\]

with first-order conditions

\[
(q_t - 1 - \Phi' (i_t)) + \lambda_t = 0 \hspace{1cm} (47)
\]

\[
(q_t i_t - \Phi (i_t)) + \lambda_t (i_t - \delta) = r_t \lambda_t - \dot{\lambda}_t \hspace{1cm} (48)
\]

Taking the time derivative of equation (47)

\[
\dot{\lambda}_t = - (q_t - \Phi'' (i_t) i_t)
\]
which, combined with (48), yields

$$ (q_{t+1} - q_t - \Phi'(t)) - (q_t - 1 - \Phi'(t)) (q_t - \delta - r_t) = (\dot{q}_t - \Phi''(t) i_t) $$

Rearranging we get

$$ r_t = (\delta) + \frac{\dot{q}_t - \Phi''(t) i_t}{q_t - 1 - \Phi'(t)} - \frac{q_{t+1} - q_t - \Phi(t)}{q_t - 1 - \Phi'(t)}. $$

### A.4 Household’s problem

We can rewrite the household’s problem as

$$ W_t = \max_{C_t, L_t, L_t, S_t^N, B_t^N} \mathbb{E}_0 \int_0^\infty e^{-\rho t} \left( \frac{C_t^{1-\zeta}}{1-\zeta} - \frac{\rho L_t^{1+\phi}}{1+\phi} \right) dt. $$

(49)

s.t.  

$$ \dot{D}_t = \left[ (R_t - \delta q_t) D_t + w_t L_t - C_t - S_t^N + \Pi_t \right] / q_t, $$

(50)

$$ \dot{B}_t^N = S_t^N + (i_t - \pi_t) B_t^N, $$

(51)

The Hamiltonian is

$$ H = \left( \frac{C_t^{1-\zeta}}{1-\zeta} - \frac{\rho L_t^{1+\phi}}{1+\phi} \right) + \rho_t \left[ \left( (R_t - \delta q_t) D_t + w_t L_t - C_t - S_t^N + (q_{t+1} - q_t) - \Phi'(t) \right) K_t + \Pi_t \right] / q_t \right] + \eta_t \left[ S_t^N - \right. $$

The first order conditions are

$$ C_t^{1-\zeta} - \rho_t / q_t = 0 $$

(52)

$$ - \gamma L_t^{\phi} + \rho_t w_t / q_t = 0 $$

(53)

$$ - \rho_t / q_t + \eta_t = 0 $$

(54)

$$ \dot{\rho}_t = \rho_t' \rho_t - \rho_t (R_t - \delta q_t) / q_t $$

(55)
\[ \dot{\eta}_t = \rho^h_t \eta_t - \eta_t [(i_t - \pi_t)] \quad (56) \]

(52) and (53) combine to the optimality condition for labor

\[ w_t = \frac{L^\theta_t}{C_t^{-\eta_t}} \]

(52) can be rewritten as

\[ \varrho_t = C_t^{\eta_t} q_t \]

Now take derivative wrt. time

\[ \dot{\varrho}_t = -\eta C_t^{\eta_t - 1} \dot{C}_t q_t + C_t^{\eta_t} \dot{q}_t \]

and plug this into (55) and rearrange to get the first Euler equation

\[ \frac{\dot{C}_t}{C_t} = \frac{R_t - \delta q_t + \dot{q}_t}{q_t} - \rho^h_t \]

(54) can be rewritten as

\[ \eta_t = \varrho_t / q_t \]

Now take derivative wrt. time

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

Use these two expressions and the definition of \( \dot{\varrho}_t \) in (56) and rearrange to get the second Euler equation

\[ \frac{\dot{C}_t}{C_t} = \frac{(i_t - \pi_t) - \rho^h_t}{\eta_t} \]

Combining the 2 Euler equations we get the Fisher equation

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

Finally using the definition of \( r_t \equiv \frac{R_t - \delta q_t + \dot{q}_t}{q_t} \) we can rewrite the first Euler equation and the Fisher equation as in the main text.
A.5 Distribution

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

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

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

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

$$\frac{\partial \omega_t(z)}{\partial t} = -\frac{\dot{A}_t}{A_t^2} \int_0^\infty ag_t(z,a) da + \frac{1}{A_t} \int_0^\infty a \frac{\partial g_t(z,a)}{\partial t} da.$$  (58)

Next, we plug in the derivative of $g_t(z,a)$ wrt time from equation (57) into equation (58),

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

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

$$\frac{\partial \omega_t(z)}{\partial t} = \left[ s_t(z) - \frac{\dot{A}_t}{A_t} - (1 - \psi)\eta \right] \omega_t(z) - \frac{\partial}{\partial z}[\mu(z)\omega_t(z)] + \frac{1}{2} \frac{\partial^2}{\partial z^2}[\sigma^2(z)\omega_t(z)].$$  (59)

The stationary distribution is therefore given by the following second order partial differential equation,
\begin{align*}
0 &= (s(z) - (1 - \psi)\eta) \omega(z) - \frac{\partial}{\partial z} \mu(z)\omega(z) + \frac{1}{2} \frac{\partial^2}{\partial z^2} \sigma^2(z)\omega(z). \tag{60}
\end{align*}

Remember that \( s_t(z_t, a_t, c_t) = \frac{1}{q_t} [\Phi_t(z_t, a_t) + (R_t - \delta q_t) a_t] \), since firms distribute 0 dividends.

### A.6 Market clearing and aggregation

Define the cumulative function of net-worth shares as

\begin{equation}
\Omega_t(z) = \int_0^z \omega_t(z) dz. \tag{61}
\end{equation}

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

\begin{equation}
K_t = \int k_t(z, a) dG_t(z, a) = \int_{z_t^*}^{\infty} \int \gamma a \frac{1}{A_t} g_t(z, a) d\alpha dA_t = \gamma (1 - \Omega(z_t^*)) A_t. \tag{62}
\end{equation}

By combining equations (27), (28) and (62), and solving for \( A_t \), we obtain

\begin{equation}
A_t = \frac{D_t}{(1 - \Omega(z_t^*)) - 1}. \tag{63}
\end{equation}

Labor market clearing implies

\begin{equation}
L_t = \int_{0}^{\infty} l_t(z, a) dG_t(z, a). \tag{64}
\end{equation}

Define the following auxiliary variable,

\begin{equation}
X_t \equiv \int_{z_t^*}^{\infty} z \omega_t(z) dz = \mathbb{E} [z \mid z > z_t^*] (1 - \Omega(z_t^*)). \tag{65}
\end{equation}

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

\begin{equation}
L_t = \int_{0}^{\infty} \left( \frac{\varphi_t}{\alpha m_t} \right)^{\frac{1}{1-\alpha}} \Gamma z_t \gamma a_t dG_t(z, a) = \left( \frac{\varphi_t}{\alpha m_t} \right)^{\frac{1}{1-\alpha}} \Gamma A_t X_t. \tag{66}
\end{equation}

Plugging in (8) into production function (1), and using again the definition of shares, we obtain...
\[ Y_t = \int \frac{\Gamma z_t \varphi_t}{\alpha m_t} \gamma a \, dG_t(z,a) = \Gamma \frac{\varphi_t}{\alpha m_t} X_t \gamma A_t = Z_t A_t^\alpha L_t^{1-\alpha}, \quad (67) \]

where in the last equality we have used equation (66), and we have defined

\[ Z_t = (\Gamma \gamma X_t)^\alpha. \quad (68) \]

Aggregate profits of retailers are given by

\[ \Phi_t^{Agg} = \int \gamma \max \{\Gamma z_t \varphi_t - R_t, 0\} a_t dG_t(z,a) = [\Gamma \varphi_t X_t - R_t \left(1 - \Omega(z^*)\right)] \gamma A_t. \quad (69) \]

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

\[ Y_t = \tilde{Z}_t K_t^{\alpha} L_t^{1-\alpha}, \quad (70) \]

where the TFP term \( \tilde{Z}_t \) is defined as

\[ \tilde{Z}_t = \left( \frac{\Gamma X_t}{(1 - \Omega(z^*))} \right)^\alpha = (\Gamma \mathbb{E}[z \mid z > z^*_t])^\alpha. \quad (71) \]

Aggregating the budget constraint of all input good firms, using the linearity of savings policy (11) and using (63), we obtain

\[ \hat{A}_t = \int \hat{a} dG(z,a,t) - \eta \int (1 - \psi)a_t dG(z,a,t) = \int_0^\infty \frac{1}{q_t} \left( \gamma \max \{\Gamma z_t \varphi_t - R_t, 0\} + R_t - \delta q_t - q_t (1 - \psi) \eta \right) a_t dG(z,a), \]

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

\[ \frac{\hat{A}_t}{A_t} = \frac{1}{q_t} \left( \gamma \varphi_t \Gamma X_t - R_t \gamma (1 - \Omega(z^*_t)) + R_t - \delta q_t - q_t (1 - \psi) \eta \right). \quad (72) \]

Using the definition of \( X_t \), and substituting \( \varphi_t \) using equation (66), we can simplify equation (72) as

\[ \frac{\hat{A}_t}{A_t} = \frac{1}{q_t} \left( \alpha m_t Z_t A_t^{\alpha-1} L_t^{1-\alpha} - R_t \gamma (1 - \Omega(z^*_t)) + R_t - \delta q_t - q_t (1 - \psi) \eta \right). \quad (73) \]
Finally, we can obtain factor prices

\[ w_t = (1 - \alpha) m_t Z_t A_t^{\alpha} L_t^{-\alpha} \quad (74) \]

\[ R_t = \alpha m_t Z_t A_t^{\alpha - 1} L_t^{1 - \alpha} \frac{z_t^*}{\gamma X_t} \quad (75) \]

where wages come from substituting the definition of \( \varphi_t \) into equation (66); and interest rates come from plugging in the wage expression (74) into the cut-off rule (10) and using equation (63). We could equivalently write equation (75) in terms of real rate of return \( r_t \):

\[ r_t = \frac{1}{q_t} \left( \alpha m_t Z_t A_t^{\alpha - 1} L_t^{1 - \alpha} \frac{z_t^*}{\gamma X_t} \right) - \delta + \frac{\dot{q}}{q_t} \quad (76) \]

We can easily get these equations in terms of capital instead of net worth by simply using equation (62), i.e. \( A_t = \frac{K_t}{\gamma (1 - \Omega(z_t^*))} \), and using that \( \mathbb{E} [ z | z > z_t^* ] = \frac{X_t}{(1 - \Omega(z_t^*))} = \frac{\int_{z_t^*}^{\infty} z \omega_t(z) dz}{\gamma (1 - \Omega(z_t^*))} \) (see equation (68) and (71)).
The competitive equilibrium economy is described by the following 22 equations:

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

\[ s_t(z) = \frac{1}{q_t} \left( \gamma \max \{ \Gamma z_t \phi_t - R_t, 0 \} \right) + R_t - \delta q_t \]

\[ \Omega_t(z) = \int_0^z \omega_t(x) \, dx \]

\[ \varphi_t = \alpha \left( \frac{1 - \alpha}{w_t} \right)^{\frac{1}{\alpha}} m_t^\alpha \]

\[ \tilde{m}_t = m_t(1 - \tau) \]

\[ w_t = (1 - \alpha)m_t \tilde{Z}_t K_t^\alpha L_t^{-\alpha} \]

\[ R_t = \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1 - \alpha} \left( \frac{z_t^*}{E[z \mid z > z_t^*]} \right) \]

\[ \frac{\dot{A}_t}{A_t} = \frac{1}{q_t} \left[ \gamma (1 - \Omega(z_t^*)) \left( \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1 - \alpha} - R_t \right) + R_t - \delta q_t - q_t(1 - \psi)\eta \right] \]

\[ K_t = A_t + D_t \]

\[ \dot{K}_t = (\iota_t - \delta)K_t \]

\[ A_t = \frac{\gamma(1 - \Omega(z_t^*)) - 1}{\gamma(1 - \Omega(z_t^*))} \]

\[ \tilde{Z}_t = \left( \Gamma E[z \mid z > z_t^*] \right)^\alpha \]

\[ E[z \mid z > z_t^*] = \frac{\int_{z_t^*}^{\infty} z \omega_t(z) \, dz}{(1 - \Omega(z_t^*))} \]

\[ \dot{C}_t = \frac{r_t - \rho_t^h}{\eta} \]

\[ w_t = \frac{\Gamma L_t^\alpha}{C_t^{-\eta}} \]

\[ \dot{D}_t = \left[ (R_t - \delta q_t) D_t + w_t L_t - C_t + \Pi_t \right] / q_t \]

\[ r_t = \iota_t - \pi_t \]

\[ r_t = \frac{R_t - \delta q_t + \dot{q}_t}{q_t} \]

\[ (q_t - 1 - \Phi'(\iota_t)) (r_t - (\iota_t - \delta)) = \dot{q}_t - \Phi''(\iota_t) \iota_t - (q_t \iota_t - \iota_t - \Phi(\iota_t)) \]

\[ \left( r_t - \frac{\dot{Y}_t}{Y_t} \right) \pi_t = \frac{\varepsilon}{\theta} (\bar{m}_t - m^*) + \bar{\pi}_t, \quad m^* = \frac{\varepsilon - 1}{\varepsilon} \]

\[ Y_t = \tilde{Z}_t K_t^\alpha L_t^{1 - \alpha} \]

\[ \Pi_t = (1 - m_t) Y_t - \frac{\theta}{2} \sigma_t^2 Y_t \]

\[ \dot{d} = -\nu \left( \dot{e} - \left( \rho_t^h + \phi (\pi_t - \bar{\pi}) + \bar{\pi} \right) \right) \, dt \]
for the 22 variables \( \{\omega, s, w, r, q, \varphi, K, A, L, C, D, \tilde{Z}, \Omega, z^*, \iota, \pi, m, \tilde{m}, i, Y, \Pi\} \).

Remember that \( \mu(z) = z \left(-\varsigma z \log z + \frac{a_z^2}{2}\right) \) and \( \sigma(z) = \sigma_z z \), and that government bonds are in zero net supply \( (B^N_t \equiv 0, \text{ hence } X_t = 0) \). Except from the last equation (Taylor rule), these 20 equations are the constraints of the Ramsey problem described in section 2.8.

### A.8 RANK vs HANK

In this appendix we want to highlight what the differences between the heterogeneous-agent New Keynesian model (HANK ) presented in this paper and the standard representative agent New Keynesian model with capital (RANK). The first difference is that in RANK there are no financial frictions, i.e. the representative firm can rent the desired amount of capital to produce, while it is the household who owns all the capital, i.e. \( D_t = K_t \). In contrast, in the HANK model, firms can only use capital that exceeds in a factor of \( \gamma \) the net worth they own, i.e. \( \gamma a_t \leq k_t \).

This feature makes that the heterogeneous units of production need to accumulate net worth in the form of capital inside the firm to alleviate these financial frictions. Hence, in the HANK model, the aggregate capital is held by the heterogeneous producing units \( (A_t) \) and the final household \( (D_t) \), i.e. \( K_t = D_t + A_t \). Secondly, in RANK there is no producer heterogeneity in productivity, hence the representative producer produces with exogenous productivity, \( TFP = \Gamma^0_t \). However, in HANK each producing unit has an idiosyncratic productivity, which evolves stochastically. These firm will only produce if their productivity is over a certain threshold \( z^* \), and hence aggregate productivity depends on the expected productivity of active firms, \( TFP = (\Gamma E [z \mid z > z^*_t])^0 \). The rest of the agents (retailers, final good producers, capital producers) are identical in both economies.

The competitive equilibrium of the RANK model with capital consists of the following equations 16 equations:
\[ \varphi_t = \alpha \left( \frac{(1 - \alpha)}{w_t} \right)^{(1-\alpha)/\alpha} m_t^{\frac{1}{\alpha}} \]
\[ \tilde{m}_t = m_t(1 - \tau) \]
\[ w_t = (1 - \alpha) m_t \tilde{Z}_t K_t^\alpha L_t^{-\alpha} \]
\[ R_t = \alpha m_t \tilde{Z}_t K_t^{\alpha - 1} L_t^{1-\alpha} \]
\[ K_t = D_t \]
\[ \dot{K}_t = (\iota_t - \delta) K_t \]
\[ \tilde{Z}_t = (\Gamma_t)^\alpha \]
\[ \dot{C}_t = \frac{r_t - \rho_t^h}{\eta} \]
\[ w_t = \frac{\gamma L_t^0}{C_t^{-\eta}} \]
\[ \dot{D}_t = \frac{[(R_t - \delta q_t) D_t + w_t L_t - C_t + \Pi_t]}{q_t} \]
\[ r_t = \iota_t - \pi_t \]
\[ r_t = \frac{R_t - \delta q_t + \dot{\pi}_t}{q_t} \]
\[ (q_t - 1 - \Phi' (\iota_t)) (r_t - (\iota_t - \delta)) = \dot{q}_t - \Phi'' (\iota_t) \iota_t - (q_t \iota_t - \iota_t - \Phi (\iota_t)) \]
\[ \left( r_t - \frac{Y_t}{\bar{Y}_t} \right) \pi_t = \frac{\varepsilon}{\theta} (\tilde{m}_t - m^*) + \dot{\pi}_t, \quad m^* = \frac{\varepsilon - 1}{\varepsilon} \]
\[ Y_t = \tilde{Z}_t K_t^\alpha L_t^{1-\alpha} \]
\[ \Pi_t = (1 - m_t) Y_t - \frac{\theta}{2} \pi_t^2 Y_t \]
\[ di = -v \left( \iota_t - \left( \rho_t^h + \phi (\pi_t - \bar{\pi}) + \bar{\pi} \right) \right) dt \]
for the 16 variables \( \{w, r, q, \varphi, K, L, C, D, \tilde{Z}, \iota, \pi, m, \tilde{m}, i, Y, \Pi\} \).
B Numerical Appendix

B.1 Finite difference approximation of the Kolmogorov Forward equation

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

\[
\frac{\omega^n_j - \omega^{n-1}_j}{\Delta t} = \left( s_n(z_j) - \frac{\dot{A}_n}{A_n} - (1 - \psi)\eta \right) \omega_n(z_j) - \frac{\omega^n_j \mu(z_j) - \omega^{n-1}_j \mu(z_{j-1})}{\Delta z} + \frac{\omega^n_{j+1} \tilde{\sigma}^2(z_{j+1}) + \omega^n_{j-1} \tilde{\sigma}^2(z_{j-1}) - 2 \omega^n_j \tilde{\sigma}^2(z_j)}{2 (\Delta z)^2},
\]

which, grouping, results in

\[
\frac{\omega^n_j - \omega^{n-1}_j}{\Delta t} = \left[ s_n(z_j) - \frac{\dot{A}_n}{A_n} - (1 - \psi)\eta \right] \omega_n(z_j) - \frac{\mu(z_j)}{\Delta z} - \frac{\tilde{\sigma}^2(z_j)}{(\Delta z)^2} \omega_n(z_j) - \frac{\mu(z_{j-1}) + \tilde{\sigma}^2(z_{j-1})}{\Delta z} \omega_{j-1}^n + \frac{\tilde{\sigma}^2(z_{j+1})}{2 (\Delta z)^2} \omega_{j+1}^n + \frac{\tilde{\sigma}^2(z_{j+1})}{2 (\Delta z)^2} \omega_{j+1}^n.
\]

The boundary conditions are the ones associated with a reflected process \( z \) at the boundaries:\textsuperscript{13}

\[
\frac{\omega^n_1 - \omega^{n-1}_1}{\Delta t} = (\beta^n_1 + \chi^n_1) \omega_n(z_1) + \chi^n_2 \omega^n_{j+1},
\]

\[
\frac{\omega^n_J - \omega^{n-1}_J}{\Delta t} = (\beta^n_J + \gamma^n_J) \omega_n(z_J) + \gamma^n_{j-1} \omega^n_{j-1}.
\]

\textsuperscript{13} 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.
If we define matrix
\[
B^n = \begin{bmatrix} 
\beta_1^n + \chi_1^n & \chi_2^n & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\rho_1^n & \beta_2^n & \chi_3^n & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \rho_2^n & \beta_3^n & \chi_4^n & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\rho^n & \beta^n & \chi^n & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\end{bmatrix},
\]
then we can express the KF equation as
\[
\frac{\omega^n - \omega^{n-1}}{\Delta t} = B^{n-1} \omega^n,
\]
or
\[
\omega^n = (I - \Delta t B^{n-1})^{-1} \omega^{n-1},
\]
where \( \omega^n = \begin{bmatrix} \omega_1^n & \omega_2^n & \cdots & \omega_{J-1}^n & \omega_J^n \end{bmatrix}^T \), and \( I \) is the identity matrix of dimension \( J \).

B.2 Algorithm to solve for the SS

We know that in SS consumption does not grow, hence from (21)
\[
r^{ss} = \rho^h.
\]
(78)

We also know that in SS, the investment rate is equal to the depreciation,
\[
i^{ss} = \delta.
\]
(79)

This means that, from equation (35) and the functional form we assumed for the capital adjustment costs (18),
\[
(q_t - 1 - \Phi'(t_t)) (r_t - (\iota_t - \delta)) = \Phi''(t_t) \dot{i}_t - (\mu t_t - \iota_t - \Phi(t_t))
\]
(80)

\[
(q^{ss} - 1 - \phi^k(i^{ss} - \delta)) \left( \rho^{hh} - (i^{ss} - \delta) \right) = 0 - \phi^k \ast 0 - \left( q^{ss} i^{ss} - i^{ss} - \phi^k(i^{ss} - \delta) \right).
\]

48
\[
\rho^{hh}(q^{ss} - 1) = \delta(1 - q^{ss})
\]

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

\[
q^{ss} = 1. \tag{81}
\]

We have that in steady state, all variables are constant, including \( i^{ss} \). For the nominal interest rate to be constant, we need that

\[
\pi^{SS} = \bar{\pi}, \tag{82}
\]

and the nominal interest rate is then given by

\[
i^{ss} = \rho^h + \bar{\pi}. \tag{83}
\]

In SS, \( \dot{\pi}_t = 0 \) and \( \dot{Y}_t = 0 \). Hence, from equation (14) we obtain

\[
m^{ss} = \left( m^* + \rho^h \frac{\theta}{\bar{\pi}} \right). \tag{84}
\]

Using equation (34) and (78),

\[
\rho^h = \frac{1}{q^{ss}} \left( \alpha m_t Z_t A_t^{\alpha - 1} L_t^{1-\alpha} \frac{z^*}{\gamma X_t} \right) - \delta \tag{85}
\]

From equation (36) and (78),

\[
\frac{\dot{A}_t}{A_t} = 0 = \frac{1}{q_t} (\alpha m_t Z_t A_t^{\alpha - 1} L_t^{1-\alpha} - R_t \Gamma(1 - \Omega(z^*_t)) + R_t - \delta q_t - q_t (1 - \psi) \eta). \tag{86}
\]

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

\[
\rho^h + \delta = \left[ (\rho^h + \delta) \left( \frac{1 - \Omega(z^*_t)}{\gamma X^*} \right) - 1 \right] + (1 - \psi) \eta + \delta \left[ \frac{z^*}{\gamma X^*} \right]. \tag{87}
\]

In the algorithm, we will use this equation to solve for \( z^* \). We will compute the RHS (which is an increasing function) for the values of the grid, and find where
they intersect the LHS (a scalar) by interpolation, in order to obtain \( z^* \).

**The Algorithm.**

Assume that now \( n \) denotes iteration round of the inner loop, and \( m \) denotes the iteration round of the outer loop.

- Get \( r^{ss} = \rho^h \), \( \pi^{ss} = \bar{\pi} \) and \( i^{ss} = \rho^h + \pi^{ss} \) and \( m^{ss} = m^* + \rho^h \bar{\pi} \).

- \( A - \) For an initial guess for labor supply \( L_0 \),
  1. Compute the vector \( \Omega^n_j = \sum_{i=1}^J \omega^n_i \Delta z \).
  2. Compute a vector of \( X^n_{j} \) for each value in the grid of \( z \), such that \( X^n_{j} = \sum_{j=j}^{\hat{j}} \omega^n_j \Delta z \).
  3. Compute the RHS of equation (87), and find \( z^* \) by interpolation.
  4. Interpolate \( X^*_{n} \) and \( \Omega^n_{n}(z^*) \) too, and obtain \( Z^n = (\gamma \Gamma_n X^*_{n})^\alpha \).
  5. Find \( A \) from equation (33),
     \[
     A^n = \left[ q^{ss} \rho^h + \delta q^{ss} \right]^{\frac{1}{1-\alpha}}.
     \]
  6. Find the stocks \( K^n = \gamma(1 - \Omega^n(z^*))A^n \), \( D^n = K^n - A^n \).
  7. Compute \( \pi^{SS} = \bar{\pi}, i^{ss} = \rho^h + \bar{\pi} \) and \( m^{ss} = (m^* + \rho^h \bar{\pi}) \).
  8. Compute \( R^n = q^{ss}(\rho^h + \delta) \).
  9. Compute \( w^n = (1 - \alpha)m^{ss} Z^n A^n \alpha L^n - \alpha \varphi^n = \alpha \left( \frac{(1-\alpha)}{w^n} \right)^{(1-\alpha)/\alpha} m^{ss} \frac{1}{\bar{\pi}} \).
  10. Get aggregate output \( Y = Z_n A^n \alpha L^{1-\alpha}, \) transfers \( T_n = (1 - m^{ss}) Y_n - \frac{\theta}{2} \pi^{ss} Y_n \), and consumption \( C^n = w_n L_m + r^{ss} D_n + T_n \).
  11. Update \( s^n_j = \frac{1}{q^n}(\gamma \max \{ \Gamma z \varphi_n - R_n, 0 \} + R_n - \delta q^{ss}) \) and employ it to construct matrix \( B^{n-1} \).
  12. Update \( \omega^{n+1} \) using equation \( B^{n-1} \omega^n = 0 \).
  13. If net worth do not coincide with the guess, update them using a relaxation scheme, set \( n + 1 \) and return to point 1

- Once the net-worth shares have converged for a given \( L_m \), compute labor supply from the HH FOC 22

\[
L^{new}_m = (w_m C^{m-1})^{\frac{1}{\bar{\pi}}}
\]
Minimize the loss function \( fcn_{\text{loss}} = L_m - L_{m}^{\text{new}} \).

We set the parameter \( \Upsilon \) such that in steady state \( L = 1 \), i.e. \( \Upsilon = \left( w_{L=1} C_{L=1}^{-\eta} \right) \).

\[ \text{C Proof of proposition 3} \]

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

1. **The generic problem**  The planner’s problem in an economy with heterogeneity among one agent type (e.g. households or firms) can be written as
\[
\begin{align*}
\max_{Z_t, u_t(x), \mu_t(x), v_t(x)} & \quad \int_0^\infty \exp(-\rho t) f_0(Z_t) dt \\
\text{s.t. } & \forall t \\
\dot{X}_t & = f_1(Z_t) \\
\dot{U}_t & = f_2(Z_t) \\
0 & = f_3(Z_t) \\
\ddot{U}_t & = \int f_4(x, u_t(x), Z_t) \mu_t(x) \, dx \\
\rho v_t(x) & = \ddot{v}_t(x) + f_5(x, u_t(x), Z_t) \\
& \quad + \sum_{i=1}^{I} b_i(x, u_t(x), Z_t) \frac{\partial v_t(x)}{\partial x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{1}{2} \frac{\partial^2 v_t(x)}{\partial x_i \partial x_k}, \forall x \\
0 & = \frac{\partial f_5}{\partial u_{j,t}} + \sum_{i=1}^{I} \frac{\partial b_i}{\partial u_{j,t}} \frac{\partial v_t(x)}{\partial x_i}, \quad j = 1, \ldots, J, \forall x. \\
\dot{\mu}_t(x) & = -\sum_{i=1}^{I} \frac{\partial}{\partial x_i} [b_i(x, u_t(x), Z_t) \mu_t(x)] \\
& \quad + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\partial^2}{\partial x_i \partial x_k} \left[ \left( \sigma(x) \sigma(x)^\top \right)_{i,k} \mu_t(x) \right], \forall x \\
X_0 & = \bar{X}_0 \\
\mu_0(x) & = \bar{\mu}_0(x) \\
\lim_{t \to \infty} U & = \bar{U}_\infty \\
\lim_{t \to \infty} v(x) & = \bar{v}(x) \infty
\end{align*}
\]

where we have adopted the following notation:

- Variables (capitals are reserved for aggregate variables):
  - \( x \) individual state vector with \( I \) elements
  - \( u \) individual control vector with \( J \) elements
  - \( v \) individual value function vector with 1 element
  - \( u(x) \) control vector as function of individual state
  - \( \mu(x) \) distribution of agents across states
  - \( v(x) \) value function as function of individual state
- $X$ aggregate state vector (other than $\mu$)
- $\tilde{U}$ aggregate control vector of purely contemporaneous variables
- $U$ aggregate control vector of intertemporal variables
- $\bar{U}$ control vector of aggregator variables
- $Z_t = \{\tilde{U}_t, U_t, \bar{U}_t, X_t\}$ vector of all aggregate variables

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

Line (88) is the planner’s objective function. Equations (89)-(91) 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 (92) links aggregate and individual variables, such as the definition of aggregate TFP in our model. Equations (93) and (94) 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 individually optimal choice. The Kolmogorov Forward equation (25) determines the evolution of the distribution of agents. Finally (96)-(99) 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.

\footnote{Notice that the planner’s discount factor, $\varrho$, can be different to that of individual agents, $\rho$.}
2.a The Lagrangian  The Lagrangian for this problem is given by:

\[ L = \int_0^\infty \left\{ e^{-\sigma t} \right\} f_0(Z_t) + \lambda_{1,t} \left( X_t - f_1(Z_t) \right) + \lambda_{2,t} \left( \dot{U}_t - f_2(Z_t) \right) + \lambda_{3,t} \left( f_3(Z_t) \right) + \lambda_{4,t} \left( \dot{U}_t - \int f_4(x, u_t(x), Z_t) \mu_t(x) \, dx \right) \]

\[ + \int \left[ \lambda_{5,t}(x) \left( -\rho v_t(x) + \dot{v}_t(x) + f_5(x, u_t(x), Z_t) + \sum_{i=1}^I b_i(x, u_t(x), Z_t) \frac{\partial v_t(x)}{\partial x_i} + \sum_{i=1}^I \sum_{k=1}^I \frac{(\sigma(x)\sigma(x)^\top)_{i,k}}{2} \right) \right] dx \]

\[ + \int \left[ \lambda_{6,t}(x) \left( \frac{\partial f_5}{\partial u_{j,t}} + \sum_{i=1}^I \frac{\partial b_i}{\partial u_{j,t}} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \]

\[ + \int \left[ \lambda_{7,t}(x) \left( -\dot{\mu}_t(x) + \left( -\sum_{i=1}^I \frac{\partial }{\partial x_i} [b_i(x, u_t(x), Z_t) \mu_t(x)] \right) + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^I \frac{\partial^2}{\partial x_i \partial x_k} \left[ (\sigma(x)\sigma(x)^\top)_{i,k} \right] \mu_t(x) \right] \]

where \( \lambda_1 \) to \( \lambda_7 \) denote the multipliers on the respective constraints. For convenience
we write the time derivatives in a separate line at the end. The Lagrangian becomes:

\[
\mathcal{L} = \int_0^\infty \left\{ e^{-\rho t} \left[ f_0(Z_t) + \lambda_{1,t} (-f_1(Z_t)) + \lambda_{2,t} (-f_2(Z_t)) + \lambda_{3,t} (-f_3(Z_t)) + \lambda_{4,t} \left( U_t - \int f_4(x,u_t(x),Z_t) \mu_t(x) \, dx \right) \right] \right\} dt
\]

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

\[
\int_0^\infty \left\{ e^{-\rho t} \left[ \lambda_{5,t}(x) \left( -\rho v_t(x) + f_5(x,u_t(x),Z_t) + \sum_{i=1}^I b_i(x,u_t(x),Z_t) \frac{\partial v_t(x)}{\partial x_i} \right) \right] \right\} \, dx
\]

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

\[
\int_0^\infty e^{-\rho t} \left[ a_t b_t \right] dt = \left[ e^{-\rho t} a_t b_t \right]_0^\infty - \int_0^\infty \left[ e^{-\rho t} (\dot{a}_t - \rho a_t) b_t \right] dt
\]

\[
= \lim_{t \to \infty} e^{-\rho t} a_t b_t - a_0 b_0 - \int_0^\infty \left[ e^{-\rho t} (\dot{a}_t - \rho a_t) b_t \right] dt
\]

55
Putting this all together the Lagrangian has become:

\[
\lim_{t \to \infty} e^{-\theta t} \lambda_{1,t} X_t - \lambda_{1,0} X_0 - \int_0^\infty e^{-\theta t} (\dot{\lambda}_{1,t} - \varrho \lambda_{1,t}) \, dt + \lim_{t \to \infty} e^{-\theta t} \lambda_{2,t} U_t - \lambda_{2,0} U_0 - \int_0^\infty e^{-\theta t} (\dot{\lambda}_{2,t} - \varrho \lambda_{2,t}) \, dt \\
+ \int \left( \lim_{t \to \infty} e^{-\theta t} \lambda_{5,t}(x) v_t(x) - \lambda_{5,0}(x) v_0(x) \right) \, dx - \int \int_0^\infty e^{-\theta t} (\dot{\lambda}_{5,t}(x) - \varrho \lambda_{5,t}(x)) v_t(x) \, dt \, dx \\
- \int \lim_{t \to \infty} e^{-\theta t} \lambda_{7,t}(x) \mu_t(x) - \lambda_{7,0}(x) \mu_0(x) \, dx + \int \int_0^\infty e^{-\theta t} (\dot{\lambda}_{7,t}(x) - \varrho \lambda_{7,t}(x)) \mu_t(x) \, dt \, dx
\]

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

\[
+ \lim_{t \to \infty} e^{-\theta t} \lambda_{1,t} X_t - \lambda_{2,0} U_0 - \int_0^\infty e^{-\theta t} (\dot{\lambda}_{1,t} - \varrho \lambda_{1,t}) \, dt - \int_0^\infty e^{-\theta t} (\dot{\lambda}_{2,t} - \varrho \lambda_{2,t}) \, dt \\
- \int \lambda_{5,0}(x) v_0(x) \, dx + \int \int_0^\infty e^{-\theta t} (\dot{\lambda}_{5,t}(x) - \varrho \lambda_{5,t}(x)) v_t(x) \, dt \, dx \\
- \int \lim_{t \to \infty} e^{-\theta t} \lambda_{7,t}(x) \mu_t(x) \, dx + \int \int_0^\infty e^{-\theta t} (\dot{\lambda}_{7,t}(x) - \varrho \lambda_{7,t}(x)) \mu_t(x) \, dt \, dx
\]

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

\[
+ \int \left[ \left( -\rho \lambda_{5,t}(x) v_t(x) + f_5(x, u_t(x), Z_t) - \sum_{i=1}^I \frac{\partial b_i(x, u_t(x), Z_t)}{\partial x_i} \lambda_{5,t}(x) v_t(x) \right) + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^J \frac{\partial^2 (\sigma(x)\sigma(x)^\top)}{\partial x_i \partial x_k} \right] \, dx \\
+ \sum_{j=1}^J \int \left[ \lambda_{6,j,t}(x) \frac{\partial f_{5,t}}{\partial u_j} - \sum_{i=1}^I \frac{\partial}{\partial x_i} \left[ \lambda_{6,j,t}(x) \frac{\partial b_i}{\partial u_j} \right] v_t(x) \right] \, dx \\
+ \int \left[ \left( \sum_{i=1}^I \frac{\partial \lambda_{7,t}(x)}{\partial x_i} [b_i(x, u_t(x), Z_t) \mu_t(x)] + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^J \frac{\partial^2 \lambda_{7,t}(x)}{\partial x_i \partial x_k} \left[ (\sigma(x)\sigma(x)^\top)_{i,k} \mu_t(x) \right] \right) \right] \, dx \right) \right) \right) \right)
\]

Putting this all together the Lagrangian has become:
\[
\mathcal{L} = \int_0^\infty \left\{ e^{-\varrho t} \left[ f_0(Z_t) \right] + \lambda_{1,t} (-f_1(Z_t)) + \lambda_{2,t} (-f_2(Z_t)) + \lambda_{3,t} (-f_3(Z_t)) + \lambda_{4,t} \left( \dot{U}_t - \int f_4(x, u_t(x), Z_t) \mu_t(x) \, dx \right) \right. \\
+ \int \left[ \left( -\rho \lambda_{5,t}(x)v_t(x) + \lambda_{5,t}(x)f_5(x, u_t(x), Z_t) - \sum_{i=1}^{I} \partial \left[ b_i (x, u_t(x), Z_t) \lambda_{5,t}(x) \right] v_t(x) + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \partial^2 \left[ \left( \sigma(x)\sigma(x)\top \right)_{i,k} \mu_t(x) \right] \right) \right] dx \\
+ \int \left[ \left( \sum_{i=1}^{I} \frac{\partial \lambda_{6,j,t}(x)}{\partial x_i} \left[ b_i (x, u_t(x), Z_t) \mu_t(x) \right] + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\partial^2 \lambda_{6,j,t}(x)}{\partial x_i \partial x_k} \left[ \left( \sigma(x)\sigma(x)\top \right)_{i,k} \mu_t(x) \right] \right) \right] dx \\
+ \lim_{t \to \infty} e^{-\varrho t} \lambda_{7,t}(x)X_t - \lambda_{2,0} U_0 - \int_0^\infty e^{-\varrho t} (\dot{\lambda}_{1,t} - \varrho \lambda_{1,t}) X_t dt - \int_0^\infty e^{-\varrho t} (\dot{\lambda}_{2,t} - \varrho \lambda_{2,t}) U_t dt \\
+ \int -\lambda_{5,0}(x)v_0(x)dx + \int \int e^{-\varrho t} (\dot{\lambda}_{5,t}(x) - \varrho \lambda_{5,t}(x)) v_t(x)dt dx \\
- \int \lim_{t \to \infty} e^{-\varrho t} \lambda_{7,t}(x) \mu_t(x) dx + \int \int e^{-\varrho t} (\dot{\lambda}_{7,t}(x) - \varrho \lambda_{7,t}(x)) \mu_t(x) dt dx.
\]

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:
\( U_t : 0 = -(\dot{\lambda}_{2,t} - \varrho \lambda_{2,t}) \)
\[ + \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial U_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t} \int \frac{\partial f_{4,t}}{\partial U_t} \mu_t(x) \, dx \tag{101} \]
\[ + \int \left[ \lambda_{5,t}(x) \left( \frac{\partial f_{5,t}}{\partial U_t} + \sum_{i=1}^{l} \frac{\partial b_{i,t}}{\partial U_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \tag{102} \]
\[ + \sum_{j=1}^{J} \int \left[ \lambda_{6,j,t}(x) \left( \frac{\partial^2 f_{5,t}}{\partial u_{j,t} \partial U_t} + \sum_{i=1}^{l} \frac{\partial b_{i,t}}{\partial U_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \tag{103} \]
\[ + \int \left[ \lambda_{7,t}(x) \left( - \sum_{i=1}^{l} \frac{\partial}{\partial x_i} \left[ \frac{\partial b_{i,t}}{\partial U_t} \mu_t(x) \right] \right) \right] dx, \tag{104} \]
\( \forall t > 0, \tag{105} \)
\( 0 = \lambda_{2,0}. \tag{106} \)

\( X_t : 0 = -(\dot{\lambda}_{1,t} - \varrho \lambda_{1,t}) \)
\[ + \frac{\partial f_{0,t}}{\partial X_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial X_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial X_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial X_t} - \lambda_{4,t} \int \frac{\partial f_{4,t}}{\partial X_t} \mu_t(x) \, dx \]
\[ + \int \left[ \lambda_{5,t}(x) \left( \frac{\partial f_{5,t}}{\partial X_t} + \sum_{i=1}^{l} \frac{\partial b_{i,t}}{\partial X_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \]
\[ + \sum_{j=1}^{J} \int \left[ \lambda_{6,j,t}(x) \left( \frac{\partial^2 f_{5,t}}{\partial u_{j,t} \partial X_t} + \sum_{i=1}^{l} \frac{\partial b_{i,t}}{\partial X_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] dx \]
\[ + \int \left[ \lambda_{7,t}(x) \left( - \sum_{i=1}^{l} \frac{\partial}{\partial x_i} \left[ \frac{\partial b_{i,t}}{\partial X_t} \mu_t(x) \right] \right) \right] dx, \tag{107} \]
\( \forall t \geq 0, \tag{108} \)
\( 0 = \lim_{t \to \infty} e^{-\varrho t} \lambda_{1,t}(x). \)
\( \hat{U}_t : 0 = 0 \)
\[\begin{eqnarray*}
&+& \frac{\partial f_{0,t}}{\partial \hat{U}_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial \hat{U}_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial \hat{U}_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial \hat{U}_t} - \lambda_{4,t} \int \frac{\partial f_{4,t}}{\partial \hat{U}_t} \mu_t(x) \, dx \\
&+& \int \left[ \lambda_{5,t}(x) \left( \frac{\partial f_{5,t}}{\partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial \hat{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] \, dx \\
&+& \sum_{j=1}^{J} \int \left[ \lambda_{6,j,t}(x) \left( \frac{\partial^2 f_{5,t}}{\partial u_{j,t} \partial \hat{U}_t} + \sum_{i=1}^{I} \frac{\partial b_{i,t}}{\partial u_{j,t} \partial \hat{U}_t} \frac{\partial v_t(x)}{\partial x_i} \right) \right] \, dx \\
&+& \int \left[ \lambda_{7,t}(x) \left( - \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left[ \frac{\partial b_{i,t}}{\partial \hat{U}_t} \mu_t(x) \right] \right) \right] \, dx,
\end{eqnarray*}\]
\( \forall t \geq 0. \)

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

Value function, distribution and policy functions
\[ v_t(x) : 0 = \left( -\lambda_{5,t}(x) \rho - \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left[ \lambda_{5,t}(x) b_i(x, u_t(x), Z_t) \right] \right) + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\partial^2}{\partial x_i \partial x_k} \left[ \left( \sigma(x) \sigma(x)^T \right)_{i,k} \lambda_{5,t}(x) \right] \]

\[ - \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{\partial}{\partial x_i} \left( \lambda_{6,j,t}(x) \frac{\partial b_i(x, u_t(x), Z_t)}{\partial u_{j,t}} \right) \]

\[ - (\lambda_{5,t}(x) - \rho \lambda_{5,t}(x)), \quad \forall t > 0, \]

\[ 0 = \lambda_{5,0}(x). \]

\[
\mu_t(x) : 0 = -\lambda_{4,t} f_4(x, u_t(x), Z_t) \\
+ \lambda_{7,t}(x) \left( \sum_{i=1}^{I} \frac{\partial \lambda_{7,t}(x)}{\partial x_i} b_i(x, u_t(x), Z_t) \right) + \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\partial^2 \lambda_{7,t}(x)}{\partial x_i \partial x_k} \left( \sigma(x) \sigma(x)^T \right)_{i,k} \\
+ (\lambda_{7,t}(x) - \rho \lambda_{7,t}(x)), \quad \forall t \geq 0, \]

\[ 0 = \lim_{t \to \infty} e^{-\rho t} \lambda_{7,t}(x). \]

\[ u_{i,t}(x) : 0 = -\lambda_{4,t} \frac{\partial f_4}{\partial u_{i,t}} \mu_t(x) \]

\[ + \left[ \lambda_{5,t}(x) \left( \frac{\partial f_5}{\partial u_{i,t}} + \sum_{i=1}^{I} \frac{\partial b_i}{\partial u_{i,t}} \frac{\partial v_t(x)}{\partial x_i} \right) \right] \]

\[ + \sum_{j=1}^{J} \lambda_{6,k,t}(x) \left( \frac{\partial^2 f_5}{\partial u_{i,t} \partial u_{j,t}} + \sum_{i=1}^{I} \frac{\partial^2 b_i}{\partial u_{i,t} \partial u_{j,t}} \frac{\partial v_t(x)}{\partial x_i} \right) \]

\[ - \left( \sum_{i=1}^{I} \frac{\partial \lambda_{7,t}(x)}{\partial x_i} \frac{\partial b_{i,t}}{\partial u_{i,t}} \mu_t(x) \right). \]
2.c Discretized optimality conditions

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

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

\[
\nabla_i [v^n_t] \equiv \left[ \frac{v^{n_i+1}_t - v^n_t}{\Delta x_i} + \frac{v^n_t - v^{n_i-1}_t}{\Delta x_i} \right],
\]

\[
\hat{\nabla}_i [\mu^n_t] \equiv \left[ \frac{\|b^n_{i,t} > 0\mu^{n_i+1}_t\| - \|b^n_{i,t} < 0\mu^n_t + \|b^n_{i,t} < 0\mu^{n_i-1}_t\|}{\Delta x_i} + \frac{\|b^n_{i,t} > 0\mu^{n_i-1}_t\| - \|b^n_{i,t} > 0\mu^{n_i-1}_t\|}{\Delta x_i} \right],
\]

for any discretized functions \(v^n_t, \mu^n_t\). We simplify the notation for sums \(\sum_n \equiv \sum_{n_1 \in \{1, ..., N_1\}, ..., n_I \in \{1, ..., N_I\}}\). We maintain the subscript \(t\) even if it refers now to discrete time with a step \(\Delta t\), that is, \(X_{t+1}\) is the shortcut for \(X_{t+\Delta t}\).

We start with the optimality condition for \(U_t\)
\( U_t : 0 = - \left( \frac{\lambda_{2,t} - \lambda_{2,t-1}}{\Delta t} - \varrho \lambda_{2,t} \right) \) \tag{107}

+ \frac{\partial f_0}{\partial U_t} - \lambda_{1,t} \frac{\partial f_1}{\partial U_t} - \lambda_{2,t} \frac{\partial f_2}{\partial U_t} - \lambda_{3,t} \frac{\partial f_3}{\partial U_t} - \lambda_{4,t} \sum_{n=1}^{N} \frac{\partial f_n}{\partial U_t} \mu_t^n \) \tag{108}

+ \sum_n \left[ \lambda_{5,t} \left( \frac{\partial f_5^n}{\partial U_t} \right) + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial U_t} \nabla_i \left[ v_t^n \right] \right] \tag{109}

∀ t \geq 0.

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

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

+ \frac{\partial f_0}{\partial X_t} - \lambda_{1,t} \frac{\partial f_1}{\partial X_t} - \lambda_{2,t} \frac{\partial f_2}{\partial X_t} - \lambda_{3,t} \frac{\partial f_3}{\partial X_t} - \lambda_{4,t} \sum_{n=1}^{N} \frac{\partial f_n}{\partial X_t} \mu_t^n \)

+ \sum_n \left[ \lambda_{5,t} \left( \frac{\partial f_5^n}{\partial X_t} \right) + \sum_{i=1}^{I} \frac{\partial b_i^n}{\partial X_t} \nabla_i \left[ v_t^n \right] \right] \tag{109}

∀ t > 0.
\[ \dot{U}_t : 0 = 0 + \frac{\partial f_0}{\partial U_t} - \lambda_{1,t} \frac{\partial f_1}{\partial U_t} - \lambda_{2,t} \frac{\partial f_2}{\partial U_t} - \lambda_{3,t} \frac{\partial f_3}{\partial U_t} - \lambda_{4,t} \sum_n \frac{\partial f^n}{\partial U_t} \nu^n_t \]

\[ + \sum_n \left[ \lambda_{5,n}^n \left( \frac{\partial f^n}{\partial U_t} + \sum_{i=1}^I \frac{\partial b^n_i}{\partial U_t} \nabla_i [v^n_t] \right) \right] \]

\[ + \sum_{j=1}^J \sum_n \left[ \lambda_{6,j,n}^n \left( \frac{\partial^2 f^n}{\partial u_j \partial U_t} + \sum_{i=1}^I \frac{\partial b^n_i}{\partial u_j \partial U_t} \nabla_i [v^n_t] \right) \right] \]

\[ + \sum_n \left[ -\lambda_{7,n}^n \sum_{i=1}^I \nabla_i \left[ \frac{\partial b^n_{i,t}}{\partial U_t} \nu^n_t \right] \right] \]

\[ \forall t \geq 0. \]

\[ \ddot{U}_t : 0 = \lambda_{4,t} \]

\[ + \frac{\partial f_0}{\partial U_t} - \lambda_{1,t} \frac{\partial f_1}{\partial U_t} - \lambda_{2,t} \frac{\partial f_2}{\partial U_t} - \lambda_{3,t} \frac{\partial f_3}{\partial U_t} - \lambda_{4,t} \sum_{n=1}^N \frac{\partial f^n}{\partial U_t} \nu^n_t \]

\[ + \sum_n \left[ \lambda_{5,n}^n \left( \frac{\partial f^n}{\partial U_t} + \sum_{i=1}^I \frac{\partial b^n_i}{\partial U_t} \nabla_i [v^n_t] \right) \right] \]

\[ + \sum_{j=1}^J \sum_n \left[ \lambda_{6,j,n}^n \left( \frac{\partial^2 f^n}{\partial u_j \partial U_t} + \sum_{i=1}^I \frac{\partial b^n_i}{\partial u_j \partial U_t} \nabla_i [v^n_t] \right) \right] \]

\[ + \sum_n \left[ -\lambda_{7,n}^n \sum_{i=1}^I \nabla_i \left[ \frac{\partial b^n_{i,t}}{\partial U_t} \nu^n_t \right] \right] \]

\[ \forall t \geq 0. \]

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

\[
v_t(x) : 0 = -\lambda_{5,t}^n \rho - \sum_{i=1}^I \hat{\nabla}_i \left[ \lambda_{5,t}^n b_i^n \right] + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^I \nabla_i \left[ \sigma_{i,k}^n \lambda_{5,t}^n \right] - \sum_{j=1}^J \sum_{i=1}^I \left( \hat{\nabla}_i \left[ \lambda_{6,i,t}^n \frac{\partial b_i^n}{\partial u_{j,t}^n} \right] \right) - \frac{\lambda_{5,t}^n - \lambda_{5,t}^{n-1}}{\Delta t} - \rho \lambda_{5,t}^n,
\]

where

\[
\Delta_{i,k}^2 \left[ \sigma_{i,k}^n \lambda_{5,t}^n \right] \equiv \left[ \frac{(\sigma_{i,k}^{n+1} \lambda_{5,t}^{n+1}) + (\sigma_{i,k}^{n-1} \lambda_{5,t}^{n-1}) - 2(\sigma_{i,k}^n \lambda_{5,t}^n)}{\Delta x_i \Delta x_k} \right].
\]

\[
\mu_t(x) : 0 = -\lambda_{4,t}^n f_{4,t}^n + \lambda_7(x) \left( \sum_{i=1}^I b_i(x, u_t(x), Z_t) \nabla_i \left[ \lambda_{7,t}^n \right] + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^I (\sigma_{i,k}^2)^n \Delta_{i,k}^2 \left[ \lambda_{7,t}^n \right] \right) + \frac{\lambda_{7,t}^n - \lambda_{7,t}^{n-1}}{\Delta t} - \rho \lambda_{7,t}^n.
\]

\[
u_{t,t}(x) : 0 = -\lambda_{4,t}^n \frac{\partial f_{4,t}^n}{\partial u_{t,t}^n} \mu_t^n + \sum_{j=1}^J \lambda_{6,j,t}^n \left( \frac{\partial^2 f_{5,t}^n}{\partial u_{t,t}^n \partial u_{j,t}^n} + \sum_{i=1}^I \frac{\partial^2 b_{i,t}^n}{\partial u_{t,t}^n \partial u_{j,t}^n} \nabla_i [v_t^n] \right) - \sum_{i=1}^I \nabla_i \left[ \lambda_{7,t}^n \right] \frac{\partial b_{i,t}^n}{\partial u_{t,t}^n} \mu_t^n.
\]

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 (timestep \( \Delta t \)) and the idiosynchratic state \( (N \text{ grid} \)
points, grid step $\Delta x_i$). We define the discount factor $\beta \equiv (1 + \rho \Delta t)^{-1}$.

\[
\max_{z_t, u^n_t, µ^n_t, v^n_t} \sum_t \beta^t f_0(Z_t)
\]

s.t. \forall t

\[
\frac{X_{t+1} - X_t}{\Delta t} = f_1(Z_t)
\tag{113}
\]

\[
\frac{U_{t+1} - U_t}{\Delta t} = f_2(Z_t)
\tag{114}
\]

\[
0 = f_3(Z_t)
\tag{115}
\]

\[
\dot{U}_t = \sum_{n=1}^N f_4(x^n, u^n_t, Z_t) µ^n_t
\tag{116}
\]

\[
\rho v^n_t = \frac{v^n_{t+1} - v^n_t}{\Delta t} + f_5(x^n, u^n_t, Z_t) + \sum_{i=1}^I b_i(x^n, u^n_t, Z_t) \nabla_i [v^n_t]
\tag{117}
\]

\[
+ \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^I (\sigma_{i,k}^2)^n \Delta_{i,k}^2 [v^n_t], \ \forall n
\]

\[
0 = \frac{\partial f_5^n}{\partial u^n_{j,t}} + \sum_{i=1}^I \frac{\partial b^n_i}{\partial u^n_{j,t}} \nabla_i [v^n_t], \ \forall j, n.
\tag{118}
\]

\[
\frac{µ^n_{t+1} - µ^n_t}{\Delta t} = - \sum_{i=1}^I \hat{\Delta} [b^n_{i,t}µ^n_t]
\tag{119}
\]

\[
+ \sum_{i=1}^I \sum_{k=1}^I \Delta_{i,k}^2 [\tilde{σ}_{i,k} µ^n_t]
\tag{120}
\]

\[
X_0 = \bar{X}_0
\tag{121}
\]

\[
µ_0^n = \bar{µ}_0^n
\tag{122}
\]

3.b The Lagrangian  The Lagrangian is
\[ L = \sum_t \beta^t f_0(Z_t) \]
\[ + \sum_t \beta^t \lambda_{1,t} \left\{ \frac{X_{t+1} - X_t}{\Delta t} - f_1(Z_t) \right\} \]
\[ + \sum_t \beta^t \lambda_{2,t} \left\{ \frac{U_{t+1} - U_t}{\Delta t} - f_2(Z_t) \right\} \]
\[ + \sum_t \beta^t \lambda_{3,t} \{-f_3(Z_t)\} \]
\[ + \sum_t \beta^t \lambda_{4,t} \left\{ \bar{U}_t - \sum_n f_4(x^n, u^n_t, Z_t) \mu^n_t \right\} \]
\[ + \sum_t \lambda_{5,t} \left\{ \frac{\partial f_{5,t}}{\partial u_{n,j,t}} + \sum_{i=1}^I \frac{\partial b_{n,i,t}}{\partial u_{n,j,t}} \nabla_i [v^n_t] \right\} \]
\[ + \sum_t \lambda_{6,t} \left\{ \frac{\partial^2 f_{5,t}}{\partial u_{n,j,t}^2} + \sum_{i=1}^I \frac{\partial^2 b_{n,i,t}}{\partial u_{n,j,t}^2} \nabla_i [v^n_t] \right\} \]
\[ 3.c The optimality conditions \quad The FOCs are \]
\[ \frac{\partial L}{\partial U_t} : 0 = \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} + \lambda_{2,t} \left\{-\frac{1}{\Delta t} \frac{\partial f_{2,t}}{\partial U_t} \right\} + \beta^{-1} \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t} \sum_n \frac{\partial f_{4,n,t}}{\partial U_t} \]
\[ + \sum_n \lambda_{5,n,t} \left\{ \frac{\partial f_{5,t}}{\partial U_t} + \sum_{i=1}^I \frac{\partial b_{n,i,t}}{\partial U_t} \nabla_i [v^n_t] \right\} \]
\[ + \sum_n \lambda_{6,n,t} \left\{ \frac{\partial^2 f_{5,t}}{\partial U_t^2} + \sum_{i=1}^I \frac{\partial^2 b_{n,i,t}}{\partial U_t^2} \nabla_i [v^n_t] \right\} \]
\[ + \sum_n \left\{ \lambda_{7,n,t} - \lambda_{7,n,t}^{-1} \right\} \left[ \|b_{n,t}^{n_i} < 0 \frac{\partial b_{n,t}^{n_i}}{\partial U_t} \Delta x_i \right] + \sum_n \left( \lambda_{7,n+1,t} - \lambda_{7,t} \right) \left[ \|b_{n,t}^{n_i} > 0 \frac{\partial b_{n,t}^{n_i}}{\partial U_t} \Delta x_i \right] \]
\[ \forall t \geq 0 \]
\[
\frac{\partial L}{\partial X_t} : 0 = \frac{\partial f_{0,t}}{\partial X_t} - \lambda_{1,t} \left( \frac{1}{\Delta t} + \frac{\partial f_{1,t}}{\partial X_t} \right) + \beta^{-1} \lambda_{1,t-1} \frac{1}{\Delta t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial X_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial X_t} - \lambda_{4,t} \sum_n \frac{\partial f_{4,n,t}}{\partial X_t} \mu_t^n
\]
\[+ \sum_n \lambda_{5,t} \left( \frac{\partial f_{5,n,t}}{\partial X_t} + \frac{\partial f_{5,t}}{\partial X_t} \nabla_i [v^n_i] \right) \]
\[+ \sum_n \sum_j \lambda_{6,j,t} \left( \frac{\partial f_{6,n,t}}{\partial u^n_j,t} + \frac{\partial f_{6,t}}{\partial X_t} \nabla_i [v^n_i] \right) \]
\[+ \sum_n \left\{ \sum_{i=1}^I \left( \lambda_{7,t} - \lambda_{7,t}^{n_i-1} \right) \left[ \mathbb{I}_{b_{i,t}<0} \frac{\partial b_{i,t}}{\partial X_t} \Lambda x_i \right] + \sum_{i=1}^I \left( \lambda_{7,t}^{n_i+1} - \lambda_{7,t} \right) \left[ \mathbb{I}_{b_{i,t}>0} \frac{\partial b_{i,t}}{\partial X_t} \Lambda x_i \right] \right\} \forall t > 0 \]

\[
\frac{\partial L}{\partial U_t} : 0 = \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial U_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t} \sum_n \frac{\partial f_{4,n,t}}{\partial U_t} \mu_t^n
\]
\[+ \sum_n \lambda_{5,t} \left( + \frac{\partial f_{5,n,t}}{\partial U_t} \right) \]
\[+ \sum_n \sum_j \lambda_{6,j,t} \left( \frac{\partial f_{6,n,t}}{\partial u^n_j,t} \frac{\partial f_{6,t}}{\partial X_t} \nabla_i [v^n_i] \right) \]
\[+ \sum_n \left\{ \sum_{i=1}^I \left( \lambda_{7,t} - \lambda_{7,t}^{n_i-1} \right) \left[ \mathbb{I}_{b_{i,t}<0} \frac{\partial b_{i,t}}{\partial U_t} \Lambda x_i \right] + \sum_{i=1}^I \left( \lambda_{7,t}^{n_i+1} - \lambda_{7,t} \right) \left[ \mathbb{I}_{b_{i,t}>0} \frac{\partial b_{i,t}}{\partial U_t} \Lambda x_i \right] \right\} \forall t > 0 \]

\[
\frac{\partial L}{\partial U_t} : 0 = \frac{\partial f_{0,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial U_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \lambda_{4,t} \sum_n \frac{\partial f_{4,n,t}}{\partial U_t} \mu_t^n
\]
\[+ \sum_n \lambda_{5,t} \left( + \frac{\partial f_{5,n,t}}{\partial U_t} \right) \]
\[+ \sum_n \sum_j \lambda_{6,j,t} \left( \frac{\partial f_{6,n,t}}{\partial u^n_j,t} \frac{\partial f_{6,t}}{\partial U_t} \nabla_i [v^n_i] \right) \]
\[+ \sum_n \left\{ \sum_{i=1}^I \left( \lambda_{7,t} - \lambda_{7,t}^{n_i-1} \right) \left[ \mathbb{I}_{b_{i,t}<0} \frac{\partial b_{i,t}}{\partial U_t} \Lambda x_i \right] + \sum_{i=1}^I \left( \lambda_{7,t}^{n_i+1} - \lambda_{7,t} \right) \left[ \mathbb{I}_{b_{i,t}>0} \frac{\partial b_{i,t}}{\partial U_t} \Lambda x_i \right] \right\} \forall t > 0 \]
\[
\frac{\partial L}{\partial v^n_i} : 0 = \lambda^n_{5,t} \left\{ -\rho - \frac{1}{\Delta t} + \sum_{i=1}^{I} b^n_{i,t} \frac{1_{b^n_{i,t} < 0} - 1_{b^n_{i,t} > 0}}{\Delta x_i} - \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{2\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} \right\} \\
+ \lambda^n_{5,t-1} \beta^{-1} \frac{1}{\Delta t} \\
+ \sum_{i=1}^{I} \lambda^n_{5,t} b^n_{i,t} \frac{1_{b^n_{i,t} < 0}}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \lambda^n_{5,t} \frac{\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} - \sum_{i=1}^{I} \lambda^n_{5,t} b^n_{i,t} \frac{1_{b^n_{i,t} > 0}}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \lambda^n_{5,t} \frac{\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} \\
+ \sum_{j=1}^{I} \sum_{i=1}^{I} \left\{ \lambda^n_{6,j,t} \left( \frac{\partial b^n_{i,t}}{\partial u^n_{i,j,t}} 1_{b^n_{i,t} < 0} - 1_{b^n_{i,t} > 0} \right) \right\} + \lambda^n_{6,j,t} \left( \frac{\partial b^n_{i,t}}{\partial u^n_{i,j,t}} \frac{1_{b^n_{i,t} > 0}}{\Delta x_i} \right) - \lambda^n_{6,j,t} \left( \frac{\partial b^n_{i,t}}{\partial u^n_{i,j,t}} \frac{1_{b^n_{i,t} < 0}}{\Delta x_i} \right) - \lambda^n_{6,j,t} \left( \frac{\partial b^n_{i,t}}{\partial u^n_{i,j,t}} \frac{1_{b^n_{i,t} > 0}}{\Delta x_i} \right) \\
\forall t \geq 0
\]

\[
\frac{\partial L}{\partial \mu^n_l} : 0 = -\lambda^n_{4,t} f^n_{l,t} \tag{125} \\
+ \lambda^n_{7,t} \left\{ \frac{1}{\Delta t} - \sum_{i=1}^{I} \left( \frac{1_{b^n_{i,t} > 0}}{\Delta x_i} \right) - \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{2\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} \right\} \\
+ \left\{ - \sum_{i=1}^{I} \lambda^n_{7,t} \frac{1_{b^n_{i,t} < 0}}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} \right\} \\
+ \left\{ - \sum_{i=1}^{I} \lambda^n_{7,t} \frac{-1_{b^n_{i,t} > 0}}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma^n_{i,k}}{2\Delta x_i \Delta x_k} \right\} \\
+ \beta^{-1} \lambda^n_{7,t-1} \left\{ - \frac{1}{\Delta t} \right\} \\
\forall t > 0
\]
\[
\frac{\partial L}{\partial u_{n,t}} : 0 = -\lambda_{4,t} \frac{\partial f_{4,t}^n}{\partial u_{n,t}^n} \mu_t^n \\
+ \beta^t \lambda_{5,t} \left\{ \frac{\partial f_{5,t}^n}{\partial u_{n,t}^n} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial u_{n,t}^n} \nabla_i [v_t^n] \right\} \\
+ \sum_j \lambda_{6,t} \left\{ \frac{\partial^2 f_{5,t}^n}{\partial u_{n,j,t}^n \partial u_{n,t}^n} + \sum_{i=1}^I \frac{\partial^2 b_i^n}{\partial u_{n,j,t}^n \partial u_{n,t}^n} \nabla_i [v_t^n] \right\} \\
+ \sum I \left( \lambda_{7,t}^n - \lambda_{7,t}^{n-1} \right) \left[ I_{v_t^n < 0} \frac{\partial b_i^n}{\partial u_{n,t}^n} \Delta x_i \right] + \sum I \left( \lambda_{7,t}^{n+1} - \lambda_{7,t}^n \right) \left[ I_{v_t^n > 0} \frac{\partial b_i^n}{\partial u_{n,t}^n} \Delta x_i \right]
\]
\forall t \geq 0

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 (107), which we reproduce here

\[
U_t : 0 = -\left( \frac{\lambda_{2,t} - \lambda_{2,t-1}}{\Delta} - \delta \lambda_{2,t} \right) \\
+ \frac{\partial f_0}{\partial U_t} - \lambda_{1,t} \frac{\partial f_1}{\partial U_t} - \lambda_{2,t} \frac{\partial f_2}{\partial U_t} - \lambda_{3,t} \frac{\partial f_3}{\partial U_t} - \lambda_{4,t} \sum_{n=1}^N \frac{\partial f_{4,t}^n}{\partial U_t} \mu_t^n \\
+ \sum_n \lambda_{5,t} \left\{ \frac{\partial f_{5,t}^n}{\partial U_t} + \sum_{i=1}^I \frac{\partial b_i^n}{\partial U_t} \nabla_i [v_t^n] \right\} \\
+ \sum_n \sum_j \lambda_{6,t} \left\{ \frac{\partial^2 f_{5,t}^n}{\partial u_{n,j,t}^n \partial U_t} + \sum_{i=1}^I \frac{\partial^2 b_i^n}{\partial u_{n,j,t}^n \partial U_t} \nabla_i [v_t^n] \right\} \\
+ \sum_n \left[ -\lambda_{7,t} \sum_{i=1}^I \nabla_i [\frac{\partial b_i^n}{\partial U_t} \mu_t^n] \right] \\
\forall t \geq 0
\]
The discretize-optimize condition (123), rearranges to

$$\frac{\partial L}{\partial U_t} : 0 = - \left( \frac{\lambda_{2,t} - \lambda_{2,t-1}}{\Delta t} - \frac{\beta - 1}{\Delta t} - \lambda_{2,t-1} \right)$$

$$+ \sum_{n=1}^{N} \lambda_{n,t} \frac{\partial f_{n,t}}{\partial U_t} - \lambda_{1,t} \frac{\partial f_{1,t}}{\partial U_t} - \lambda_{2,t} \frac{\partial f_{2,t}}{\partial U_t} - \lambda_{3,t} \frac{\partial f_{3,t}}{\partial U_t} - \sum_{n=1}^{N} \lambda_{4,t} \sum_{n=1}^{N} \frac{\partial f_{n,t} \mu_{n,t}^n}{\partial U_t}$$

$$+ \sum_{n=1}^{N} \sum_{j=1}^{J} \lambda_{n,j,t} \left( \frac{\partial^2 f_{n,j,t}}{\partial u_{j,t} \partial U_t} + \frac{\partial^2 b_{n,j,t}}{\partial u_{j,t} \partial U_t} \right.$$  

$$\left. \nabla_i[v^n_i] \right)$$

$$+ \sum_{n} \left\{ \sum_{i=1}^{I} \left( \lambda_{n,i,t} - \lambda_{n,i-1} \right) \left[ \frac{\partial b_{n,i,t}}{\partial U_t} \mu_{n,t}^n \right] + \right.$$  

$$\left. \sum_{i=1}^{I} \left( \lambda_{n,i+1,t} - \lambda_{n,i,t} \right) \left[ \frac{\partial b_{n,i+1,t}}{\partial U_t} \mu_{n,t}^n \right] \right\}$$

$$\forall t \geq 0$$

The second to fourth lines are evidently identical. The last lines also coincide once we take into account the definition of \( \tilde{\nabla}_i \left[ \frac{\partial b_{n,i,t}}{\partial U_t} \mu_{n,t}^n \right] \). The difference between these two equations hence is \( \| \varrho \left( \lambda_{2,t} - \lambda_{2,t-1} \right) \|. \) In the limit as \( \Delta t \to 0 \), and provided that \( \lambda_{2,t} \) features no jumps for \( t > 0 \), this difference converges to zero.

The same argument applies to the optimality conditions with respect to \( X_t \) with the difference now proportional to \( \| \varrho \left( \lambda_{1,t} - \lambda_{1,t-1} \right) \|. \) The optimality conditions with respect to \( \hat{U}_t \) and \( \tilde{U}_t \) are identical, that is, there is no difference.

Next consider the two discretized optimality conditions with respect to \( v^n_t \) (110) and (124). After some rearranging they are given by

\[ \begin{align*}
\text{Finally compare the first lines. Since } \beta &\equiv (1 + \varrho \Delta t)^{-1} \text{ we have that } \frac{\beta - 1}{\Delta t} = \varrho . \\
\text{The difference between these two equations hence is } &\| \varrho \left( \lambda_{2,t} - \lambda_{2,t-1} \right) \|. \\
\text{In the limit as } &\Delta t \to 0, \text{ and provided that } \lambda_{2,t} \text{ features no jumps for } t > 0, \text{this difference converges to zero}.
\end{align*} \]
Next, consider the two discretized optimality conditions wrt \( \mu \):

\[
v_t(x) : 0 = -\sum_{i=1}^{I} \left( \frac{\|b_{i,t}^n > 0 \lambda_{5,j,t}^n b_{i,t}^n - \|b_{i,t}^{n-1} > 0 \lambda_{5,j,t}^{n-1} b_{i,t}^{n-1}}{\Delta x_i} + \frac{\|b_{i,t}^{n+1} < 0 \lambda_{5,j,t}^{n+1} b_{i,t}^{n+1} - \|b_{i,t}^{n} < 0 \lambda_{5,j,t}^n b_{i,t}^n}{\Delta x_i} \right)
\]

\[
+ \frac{1}{2} \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma_{i,k}^{n+1} + \lambda_{5,t}^n + \sigma_{i,k}^{n-1} - 2 \sigma_{i,k}^n \lambda_{5,t}^n}{\Delta x_i \Delta x_k}
\]

\[
- \sum_{j=1}^{J} \sum_{i=1}^{I} \left( \frac{\|b_{i,t}^n > 0 \lambda_{6,j,t}^n \partial b_{i,t}^n \| - \|b_{i,t}^{n-1} > 0 \lambda_{6,j,t}^{n-1} \partial b_{i,t}^{n-1}}{\Delta x_i} + \frac{\|b_{i,t}^{n+1} < 0 \lambda_{6,j,t}^{n+1} \partial b_{i,t}^{n+1} - \|b_{i,t}^{n} < 0 \lambda_{6,j,t}^n \partial b_{i,t}^n}{\Delta x_i} \right)
\]

\[-\lambda_{5,t}^n \rho - \left( \frac{\lambda_{5,t}^n - \lambda_{5,t}^{n-1}}{\Delta t} - \rho \lambda_{5,t}^n \right)
\]

and

\[
\frac{\partial L}{\partial v_t^n} : 0 = \lambda_{5,t}^n \left\{ \sum_{i=1}^{I} \frac{\|b_{i,t}^n < 0 - \|b_{i,t}^n > 0}{\Delta x_i} - \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma_{i,k}^n}{\Delta x_i \Delta x_k} \right\}
\]

\[
+ \left\{ \sum_{i=1}^{I} \frac{\lambda_{5,t}^{n-1} b_{i,t}^{n-1} - \|b_{i,t}^{n-1} > 0}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma_{i,k}^n}{2 \Delta x_i \Delta x_k} \right\}
\]

\[
+ \left\{ - \sum_{i=1}^{I} \frac{\lambda_{5,t}^{n+1} b_{i,t}^{n+1} - \|b_{i,t}^{n+1} < 0}{\Delta x_i} + \sum_{i=1}^{I} \sum_{k=1}^{I} \frac{\sigma_{i,k}^n}{2 \Delta x_i \Delta x_k} \right\}
\]

\[
+ \sum_{j=1}^{J} \sum_{i=1}^{I} \left( \frac{\lambda_{6,j,t}^n \partial b_{i,t}^n \| - \|b_{i,t}^n > 0}{\Delta x_i} + \lambda_{6,j,t}^{n-1} \partial b_{i,t}^{n-1} - \lambda_{6,j,t}^{n+1} \partial b_{i,t}^{n+1}}{\Delta x_i} \right)
\]

\[-\rho \lambda_{5,t}^n - \left( \frac{\lambda_{5,t}^n - \lambda_{5,t}^{n-1}}{\Delta t} - \beta^{-1} - 1 \right) \lambda_{5,t}^{n-1} \right)
\]

(127)

Again these, two expressions are identical up to the last time index in the last line \( (\lambda_{5}^2) \), and thus the difference is \( \| \rho (\lambda_{5,t} - \lambda_{5,t-1}) \| \).

Next, consider the two discretized optimality conditions wrt \( \mu_t^n \) (111) and (125).
After some rearranging they are given by

$$\mu_t(x) : 0 = -\lambda_{4,t} f^n_{4,t}$$

(128)

$$+ \sum_{i=1}^I b^n_{i,t} \left[ \mathbb{I}_{b^n_{i,t} > 0} \frac{\lambda^n_{7,t} - \lambda^n_{7,t-1}}{\Delta x_i} + \mathbb{I}_{b^n_{i,t} < 0} \frac{\lambda^n_{7,t} - \lambda^n_{7,t-1}}{\Delta x_i} \right] + \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^I \sigma^n_{i,k} \frac{\lambda^n_{7,t} + 1 - 2\lambda^n_{7,t}}{2 \Delta x_i \Delta x_k}$$

$$\frac{\lambda^n_{7,t} - \lambda^n_{7,t-1}}{\Delta t} - \varrho \lambda^n_{7,t}$$

$$\frac{\partial L}{\partial \mu^n_t} : 0 = -\lambda_{4,t} f^n_{4,t}$$

(129)

$$+ \lambda_{7,t} \left\{ -\sum_{i=1}^I \left[ \mathbb{I}_{b^n_{i,t} > 0} - \mathbb{I}_{b^n_{i,t} < 0} \right] \frac{b^n_{i,t}}{\Delta x_i} - \sum_{i=1}^I \sum_{k=1}^I \frac{-2\sigma^n_{i,k}}{2 \Delta x_i \Delta x_k} \right\}$$

$$- \sum_{i=1}^I \left[ \lambda^n_{7,t} \frac{\mathbb{I}_{b^n_{i,t} < 0} b^n_{i,t}}{\Delta x_i} \right] + \sum_{i=1}^I \sum_{k=1}^I \lambda^n_{7,t} \frac{\sigma^n_{i,k}}{2 \Delta x_i \Delta x_k}$$

$$+ \frac{\lambda^n_{7,t} - \lambda^n_{7,t-1}}{\Delta t} - \frac{\beta^{-1}}{\Delta t} \lambda^n_{7,t-1},$$

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

Finally, consider the two discretized optimality conditions wrt. $u^n_{l,t}(x)$, (112) and (126). After some rearranging they are given by

$$u_{l,t}(x) : 0 = -\lambda_{4,t} \frac{\partial f^4_{4,t}}{\partial u^n_{l,t}}$$

(130)

$$+ \sum_{j=1}^J \lambda^n_{6,l,t} \left( \frac{\partial^2 f^4_{5,t}}{\partial u^n_{j,t} \partial u^n_{l,t}} + \sum_{i=1}^I \frac{\partial^2 b^n_{i,t}}{\partial u^n_{j,t} \partial u^n_{l,t}} \left[ \mathbb{I}_{b^n_{i,t} > 0} \frac{v^n_{i,t} + v^n_{i,t}}{\Delta x_i} + \mathbb{I}_{b^n_{i,t} < 0} \frac{v^n_{i,t} - v^n_{i,t}}{\Delta x_i} \right] \right)$$

$$- \sum_{i=1}^I \left[ \mathbb{I}_{b^n_{i,t} > 0} \lambda^n_{7,t} + \mathbb{I}_{b^n_{i,t} < 0} \lambda^n_{7,t} - \frac{\lambda^n_{7,t} - \lambda^n_{7,t-1}}{\Delta x_i} \right] \frac{\partial b^n_{i,t}}{\partial u^n_{l,t}} \mu^n_t$$

72
∂L
∂u_{l,t}^n : 0 = −λ_{4,t} \frac{∂f_{4,t}^n}{∂u_{l,t}^n} \mu_t^n + ∑_j \lambda_{6,t}^n \left( \frac{∂^2 f_{5,t}^n}{∂u_{j,t}^n} \frac{∂u_{j,t}^n}{∂u_{l,t}^n} \right) + ∑_{i=1}^I \frac{∂^2 b_{i,t}^n}{∂u_{j,t}^n} \frac{∂u_{j,t}^n}{∂u_{l,t}^n} \left[ I_{b_{i,t}^n < 0} \frac{v_{t+1}^n - v_t^n}{∆x_i} + I_{b_{i,t}^n > 0} \frac{v_t^n - v_{t-1}^n}{∆x_i} \right] \left[ I_{b_{i,t}^n < 0} \frac{1}{∆x_i} + I_{b_{i,t}^n > 0} \frac{1}{∆x_i} \right] \frac{∂b_{i,t}^n}{∂u_{l,t}^n} \mu_t^n ,

which are identical.

To summarize, whether one discretizes the optimality conditions of the planner and then discretizes them, or one discretizes the planner’s problem and 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 ϱλ_t. Provided that multipliers experience no jumps, the difference between the two approaches goes to 0 as ∆t → 0.

Note that this issue has nothing to do with heterogeneity.

D. Solving the Nuño and Thomas model using Dynare

Here we apply the “discretize-optimize” methodology outlined in Section 3 to the heterogeneous-agent model introduced in Nuño and Thomas (2016). This is a model à la Aiyagari-Bewley-Huggett and 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 6 displays the optimal policy (inflation) in this case, compared to the one obtained through the “optimize-discretize” methodology employed in Nuño and Thomas (2016).\(^{15}\) 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.

\(^{15}\)The calibration is the same as in Nuño and Thomas (2016), except for risk aversion, which we calibrate to 1.1.
Figure 6: Time-0 optimal monetary policy using the two approaches.

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