Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy

Eduardo Dávila* Andreas Schaab†

First version: March, 2022

This version: October, 2023

Abstract

This paper characterizes optimal monetary policy in a canonical heterogeneous-agent New Keynesian (HANK) model with wage rigidity. Under discretion, a utilitarian planner faces the incentive to redistribute towards indebted, high marginal utility households, which is a new source of inflationary bias. With commitment, i) zero inflation is the optimal long-run policy, ii) time-consistent policy requires both inflation and distributional penalties, and iii) the planner trades off aggregate stabilization against distributional considerations, so Divine Coincidence fails. We compute optimal stabilization policy in response to productivity, demand, and cost-push shocks using sequence-space methods, which we extend to Ramsey problems and welfare analysis.

JEL codes: E52, E61

Keywords: optimal monetary policy, heterogeneous-agent New Keynesian model, policy under discretion, timeless Ramsey approach, timeless penalties, inflation target, sequence-space Jacobians, sequence-space Hessians

*Yale University and NBER. Email: eduardo.davila@yale.edu. Website: https://eduardodavila.com.

†Toulouse School of Economics. Email: andreas.schaab@tse-fr.eu. Website: https://andreasschaab.com.

We are grateful to Mark Aguiar, Manuel Amador, Marios Angeletos, Adrien Auclert, Adrien Bilal, Larry Christiano, Marty Eichenbaum, Emmanuel Farhi, Jordi Galí, Mark Gertler, Simon Gilchrist, Mike Golosov, Greg Kaplan, Patrick Kehoe, Alisdair McKay, Benjamin Moll, Dmitry Muhkin, Emi Nakamura, Galo Nuño, Matt Rognlie, Tom Sargent, Stephanie Schmitt-Grohé, Jón Steinsson, Ludwig Straub, Pedro Teles, Dominik Thaler, and Christian Wolf for helpful conversations, as well as seminar participants at Columbia, Northwestern, the College de France Conference in Honor of Emmanuel Farhi, NBER SI Monetary Economics, Bank of Finland, NYU, Bank of Portugal, Oslo, LSE, 2023 AEA New Orleans Meetings, and ECB. An earlier version of this paper circulated as "Optimal Monetary Policy with Heterogeneous Agents: A Timeless Ramsey Approach".

1 Introduction

There is large heterogeneity in households' exposure to business cycle fluctuations. At the same time, there is now a growing consensus that monetary policy has distributional consequences—a view supported by mounting empirical evidence (Doepke and Schneider, 2006; Coibion et al., 2017; Ampudia et al., 2018) and the burgeoning heterogeneous-agent New Keynesian (HANK) literature (McKay et al., 2016; Kaplan et al., 2018; Auclert, 2019; Auclert et al., 2023). Household heterogeneity may therefore be an important determinant of the welfare impact of monetary policy and should inform the study of optimal policy design. However, accounting for rich heterogeneity and incomplete markets in dynamic optimal policy problems remains challenging.

In this paper, we characterize optimal monetary policy in a canonical one-asset HANK economy with wage rigidity, which represents a minimal departure from the representative-agent New Keynesian (RANK) model. Our goal is to systematically revisit the canonical New Keynesian consensus on optimal monetary policy (Clarida et al., 1999; Woodford, 2003; Galí, 2015). To do so, we structure our analysis of optimal monetary policy to parallel that of Clarida et al. (1999), starting with policy under discretion in Section 3 and studying optimal policy under commitment in Section 4. Concluding with a quantitative analysis in Section 5, we compute optimal monetary policy both non-linearly and using sequence-space perturbation methods (Boppart et al., 2018; Auclert et al., 2021), which we extend to Ramsey problems and welfare analysis.

Optimal monetary policy under discretion. Under discretion, a utilitarian planner in a HANK economy has an incentive to raise output above natural output and overheat the economy, even in the absence of markup distortions. This occurs because the planner values redistribution toward indebted, high marginal utility households via lower interest rates. At the optimum, the planner trades off this novel redistribution motive against aggregate stabilization. However, when agents anticipate the planner's incentives to lower interest rates, inflationary bias in the sense of Barro and Gordon (1983) emerges in equilibrium. Quantitatively, the redistribution motive dominates the standard markup distortion as a source of inflationary bias. The gains from commitment are consequently larger in heterogeneous-agent economies.

Optimal monetary policy with commitment. Motivated by the results under discretion, we study optimal policy under commitment in three steps. Each step isolates an important dimension of optimal monetary policy design: long-run policy, time consistency, and stabilization policy. We study the implications of household heterogeneity for optimal monetary policy along each of these dimensions.

In the first step, after introducing the standard Ramsey problem and characterizing the associated Ramsey plan, we study optimal long-run policy. We show that the optimal stationary equilibrium under commitment features zero inflation, eliminating the inflationary bias of policy under discretion in the long run. This result is due to the fact that inflation and the nominal interest rate affect households' financial income symmetrically, which can be seen as a relevant benchmark. Therefore, since the long-run real interest rate is invariant to policy, but inflation is costly while adjusting the nominal rate is not, the planner finds it optimal to keep inflation at zero in the stationary Ramsey plan.

In the second step, we show that while the standard Ramsey problem eliminates inflationary bias in the long run, it still suffers from inflationary bias in the short run. This is a due to a "time-0 problem" (Kydland and Prescott, 1980) associated with two dimensions of time inconsistency. The first source of time inconsistency is the forward-looking Phillips curve, through which inflation expectations enter the Ramsey problem. This time consistency problem has been widely studied in RANK economies by the literature following Barro and Gordon (1983). In the presence of household heterogeneity, a new second time consistency problem emerges because forward-looking individual Bellman equations appear as constraints in the Ramsey problem.

While the standard Ramsey planner chooses policy with commitment from time 0 onwards, time inconsistency still manifests at time 0. In order to find a "timeless" planning solution, we extend the approach of Marcet and Marimon (2019) to our setting (i.e., continuous-time heterogeneous-agent economies) by introducing timeless penalties for each forward-looking implementability condition. We then define a timeless Ramsey problem, which augments the standard Ramsey problem with such timeless penalties, and prove that it no longer suffers from a time-0 problem: the planner has no incentive to deviate from the stationary Ramsey plan in the absence of shocks. Hence, the timeless Ramsey problem eliminates inflationary bias in both the short and the long run.

We analytically characterize the two timeless penalties required by the timeless Ramsey problem: an inflation penalty and a distributional penalty. We first show that the inflation penalty, which is already present in RANK economies, depends on novel distributional considerations in HANK. When households are heterogeneous, changes in aggregate economic activity have distributional consequences. The standard Ramsey planner's incentive to generate inflation out of steady state at time 0, which the inflation penalty is designed to counteract, is consequently also governed by distributional considerations. Second, we show that the new distributional penalty penalizes the welfare gains of indebted, high marginal utility households. While it may seem counterintuitive that a utilitarian planner penalizes high marginal utility households, this is precisely to counteract the planner's time inconsistent incentive to redistribute towards such households, which becomes a source of inflationary bias under discretion. Finally, we show that the distributional penalty solves a novel promise-keeping Kolmogorov forward equation.

Concluding our discussion of time consistency, we explore whether a central bank that sets policy under discretion can still implement the optimal commitment solution under an appropriate institutional arrangement or with the appropriate penalties or targets. We first show that the timeless Ramsey plan can be implemented under discretion if the planner's objective is augmented to incorporate the appropriate time-varying inflation and distributional penalties. Moreover, a strict zero-inflation target implements the timeless Ramsey plan in the absence of shocks, while a modified flexible inflation target around zero inflation can implement optimal stabilization policy under commitment.

In the third and final step, we study optimal stabilization policy under the timeless Ramsey problem, which allows us to separate the pure stabilization motive from the time-0 problem. We characterize an analytical targeting rule for optimal stabilization policy in response to demand, supply, and cost-push shocks, and use it to illustrate the departures from optimal policy in RANK, which it nests. In a RANK economy, no tradeoff emerges between inflation and output in the absence of cost-push shocks; the planner finds it optimal to simultaneously close both the inflation and output gaps. In HANK economies, on the other hand, this Divine Coincidence result generically fails even in the absence of cost-push shocks. The planner now accounts for the distributional impact of policies and perceives a tradeoff between aggregate stabilization and distributional considerations. That is, even in the absence of cost-push shocks and with the appropriate employment subsidy, the planner finds it optimal not to simultaneously close the inflation and output gaps in response to shocks.

Quantitative analysis in sequence space. This paper extends the sequence-space approach (Boppart et al., 2018; Auclert et al., 2021) to Ramsey problems and welfare analysis. We develop a general sequence-space representation of (timeless) Ramsey plans that builds on Auclert et al. (2021)'s sequence-space representation of competitive equilibrium in heterogeneous-agent economies. Under this representation, a Ramsey plan is a system of equations that take as inputs the time paths of aggregate allocations and prices, aggregate multipliers, policies, and shocks.

While our approach allows us to characterize and compute timeless Ramsey plans non-linearly, an important contribution of this paper is to bring sequence-space perturbation methods to bear on optimal policy questions in heterogeneous-agent economies. We extend the fake-news algorithm of Auclert et al. (2021) to compute optimal policy and show that our timeless Ramsey approach is critical for the validity of sequence-space perturbations.

We show how to leverage both the primal and dual forms of the timeless Ramsey problem.[1] In the primal representation, we compute an extended set of sequence-space Jacobians and solve for the time paths of the multipliers that comprise a Ramsey plan. In the dual representation, we

[1] Throughout the paper, we say that a planning problem is in primal form when allocations or prices are explicit control variables for a planner, perhaps in addition to policy instruments. We say that a planning problem is in dual form when the only explicit control variables are the policy instruments. This terminology is consistent with standard use in related environments, e.g., Chari and Kehoe (1999) and Ljungqvist and Sargent (2018).

avoid having to compute the time paths of multipliers. However, approximating optimal policy in the dual is no longer possible in terms of sequence-space Jacobians and instead requires a second-order analysis. To that end, we introduce sequence-space Hessians as the natural, second-order generalization of sequence-space Jacobians.

Leveraging these methodological results, we conclude with a quantitative analysis of optimal monetary stabilization policy. Our approach allows us to compute transition dynamics under optimal policy—under discretion and with commitment—both non-linearly and to first order. We contrast optimal policy dynamics in HANK and RANK in response to demand shocks (Section 5.2), and productivity and cost-push shocks (Appendix F).

Related literature. Our paper contributes to multiple strands of the literature on optimal monetary policy, in particular recent work on optimal policy in HANK economies. Our continuous-time approach is most closely related to the work of Nuño and Thomas (2020), on which we build.[2,3] Nuño and Thomas (2020) study optimal monetary policy under commitment in a small open economy, in which short-term real interest rates and output are unaffected by policy.[4] Our paper studies optimal monetary policy in a closed economy that features the classic output-inflation tradeoff, which is central to the New Keynesian literature. Farhi and Werning (2016) study optimal monetary and macroprudential policies in a general heterogeneous-agent environment, highlighting the importance of labor wedges for optimal policies in the presence of nominal rigidities. Bhandari et al. (2021) introduce a small-noise expansion method to compute optimal monetary and fiscal policy in a HANK model with aggregate risk. Acharya et al. (2020) study optimal monetary policy in closed form in a HANK economy with constant absolute risk aversion (CARA) preferences and normally distributed shocks. Le Grand et al. (2021) study optimal monetary and fiscal policy keeping heterogeneity finite-dimensional by truncating idiosyncratic histories. González et al. (2021) characterize optimal monetary policy with heterogeneous firms. McKay and Wolf (2022) study optimal monetary policy with heterogeneous households in linear-quadratic environments.[5] Smirnov (2022) computes optimal monetary policy using a variational approach.

Our contribution relative to this body of work is fivefold. First, we provide the first analysis

[2] Nuño and Moll (2018) solve constrained-efficiency problems treating the cross-sectional distribution as a control.

[3] As emphasized by Werning (2011), continuous time Ramsey problems in New Keynesian economies are particularly tractable.

[4] Formally, the open economy setup in Nuño and Thomas (2020) immediately implies that both the Lagrange multipliers of the households' HJB equation and their optimality condition—which correspond to $\phi_t(a,z)$ and $\chi_t(a,z)$ in our paper, see equation (29)—are zero by construction. Hence, in their model, the planner would make the same savings decisions as households. Characterizing and computing these multipliers is a novel contribution of our paper.

[5] Several other papers study optimal monetary policy in environments with heterogeneity, typically relying on a second-order approximation to aggregate welfare. In particular, in two-agent New Keynesian environments, Bilbiie (2008, 2018) study optimal monetary policy without and with idiosyncratic risk, respectively; Cúrdia and Woodford (2016) study optimal monetary policy in a model with credit frictions; and Benigno et al. (2020) study optimal monetary policy at the zero lower bound.

of optimal policy under discretion in HANK economies, which shows that a utilitarian planner trades off aggregate stabilization against a novel redistribution motive. This redistribution motive is a new source of time inconsistency and exacerbates inflationary bias. Second, we jointly characterize three important dimensions of optimal monetary policy design: long-run policy, time consistency, and stabilization policy. Third, we introduce and analytically characterize the timeless penalties that resolve the time-0 problem of the standard Ramsey problem. A planner under discretion can implement the timeless Ramsey policy when confronted with the appropriate penalties. Relative to RANK, time consistent policy requires a novel distributional penalty. Fourth, the analytical targeting rules we derive for optimal monetary stabilization policy allow us to contrast policy prescriptions in HANK and RANK, which they nest as a special case. Finally, we extend the sequence-space approach to Ramsey problems, which allows us to compute optimal policy efficiently and fast.

We relate our results to the vast literature on monetary policy in RANK models and provide analytical insights into the departures of optimal policy from the RANK benchmark (Clarida et al., 1999; Woodford, 2003; Galí, 2015). At an abstract level, our approach is closest to Khan et al. (2003), who initially characterize standard and augmented (timeless) Ramsey plans and then use perturbation methods to characterize stabilization policy. Schmitt-Grohé and Uribe (2010) and Woodford (2010) systematically study and review optimal long-run policy and optimal stabilization policy in RANK economies. Our goal is to systematically revisit the canonical New Keynesian consensus on optimal monetary policy in the presence of household heterogeneity.

We formalize Woodford (1999)'s timeless perspective in our heterogeneous-agent setting by introducing a timeless penalty that resolves the standard Ramsey planner's time-0 problem (Kydland and Prescott, 1980).[6] Our characterization of the timeless penalty builds on the recursive multiplier approach of Marcet and Marimon (2019), which we extend to continuous-time heterogeneous-agent economies. Relative to RANK, we show that a time consistent implementation of monetary policy requires a novel distributional penalty and an inflation penalty augmented by distributional considerations. One contribution of our paper is to show that the distributional penalty solves a promise-keeping Kolmogorov forward equation.

Finally, we extend the sequence-space apparatus to Ramsey problems and welfare analysis, contributing to recent work on computational methods in heterogeneous-agent environments (Boppart et al., 2018; Auclert et al., 2021). We develop a sequence-space representation of timeless Ramsey plans, which we can compute non-linearly and using sequence-space perturbation methods. In particular, we extend the fake-news algorithm of Auclert et al. (2021) to compute Ramsey problems in both primal and dual forms. We also introduce and define sequence-space Hessians as the natural, second-order generalization of sequence-space Jacobians.

[6] See Woodford (2003, 2010) and Benigno and Woodford (2012) for expositions of the timeless perspective.

2 Model

Our baseline model is a one-asset heterogeneous-agent New Keynesian (HANK) model with wage rigidity (Auclert et al., 2023). It represents a minimal departure from a representative-agent New Keynesian (RANK) model (Clarida et al., 1999; Woodford, 2003; Galí, 2015).

Time is continuous and indexed by $t \in [0, \infty)$. There is no aggregate uncertainty and we focus on one-time, unanticipated shocks. Following much of the New Keynesian literature, we allow for demand, productivity, and cost-push shocks.

2.1 Households

The economy is populated by a unit mass of households whose lifetime utility is

$$V_0(\cdot) = \max \mathbb{E}_0 \int_0^\infty e^{-\int_0^t \rho_s ds} U_t(c_t, n_t) dt, \tag{1}$$

where U_t denotes the instantaneous utility flow from consumption c_t and labor n_t. Households discount at a potentially time-varying rate ρ_t, which represents a source of demand shocks. They can trade a single bond a_t and face a budget constraint

$$\dot{a}_t = r_t a_t + z_t w_t n_t + T_t(z_t) - c_t, \tag{2}$$

where r_t is the real interest rate and w_t the real wage rate. Beside financial and labor income, households may receive a lump-sum transfer $T_t(z_t)$ from the government, which will be zero in equilibrium, as described below. Finally, households face the borrowing constraint $a_t \geq \underline{a}$.

While there is no aggregate uncertainty, households face idiosyncratic earnings risk, captured by the exogenous Markov process z_t. Since we abstract from permanent heterogeneity, we can index individual households by their idiosyncratic state variables (a, z). We denote the mass of households in state (a, z) by $g_t(a, z)$, which we also refer to as the cross-sectional distribution.

2.2 Labor Market

As is standard in the New Keynesian sticky-wage literatures without heterogeneity (Erceg et al., 2000; Schmitt-Grohé and Uribe, 2005) and with heterogeneity (Auclert et al., 2023), labor unions determine work hours.[7] While Appendix A.1 details the union problem, we only summarize its relevant implications to study optimal monetary policy here. Labor is rationed, so all households supply the same hours,

$$n_t = N_t, \tag{3}$$

[7] It is possible to rederive our results in a model with price rigidity. The assumptions of sticky wages and symmetric labor rationing make our model more tractable. Moreover, firm profits are zero in equilibrium instead of counter-cyclical.

where N_t is aggregate labor. Nominal wages are sticky, and unions pay a quadratic Rotemberg (1982) adjustment cost to change wages. We assume this cost is passed to households as a utility cost, so that instantaneous flow utility in equation (1) takes the form

$$U_t(c_t, n_t) = u(c_t) - v(n_t) - \frac{\delta}{2}(\pi_t^w)^2, \tag{4}$$

where π_t^w denotes wage inflation and the parameter $\delta \geq 0$ modulates the degree of wage rigidity.

Unions choose wages to maximize stakeholder value—the private lifetime values of households. We show in Appendix A.1 that the union problem gives rise to the non-linear New Keynesian wage Phillips curve

$$\dot{\pi}_t^w = \rho_t \pi_t^w + \underbrace{\frac{\epsilon_t}{\delta}}_{\text{NKPC slope}} \iint \Big(\underbrace{\frac{\epsilon_t - 1}{\epsilon_t} \overbrace{(1+\tau^L)}^{\text{Employment Subsidy}}}_{\text{Desired Markup}} zu'(c_t(a,z)) - \frac{v'(n_t)}{A_t} \Big) w_t n_t g_t(a,z)\, da\, dz, \tag{5}$$

where ϵ_t, the elasticity of substitution across unions, is potentially time-varying and a source of cost-push shocks. As is standard in the New Keynesian literature, we allow for a time-invariant employment subsidy τ^L to potentially offset unions' market power. This Phillips curve illustrates that labor wedges, which will play a key role in our welfare analysis, are key determinants of the dynamics of inflation. Formally, we define *individual inflation-relevant labor wedges* as $\tau_t(a,z) = (\frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L)zu'(c_t(a,z)) - \frac{v'(n_t)}{A_t})w_t n_t$, and refer to $zu'(c_t(a,z)) - \frac{v'(n_t)}{A_t}$ as *individual* (welfare-relevant) *labor wedges*.[8] This definition allows us to rewrite the Phillips curve as

$$\dot{\pi}_t^w = \rho_t \pi_t^w + \frac{\epsilon_t}{\delta} \iint \tau_t(a,z) g_t(a,z)\, da\, dz, \tag{6}$$

which highlights that unions target an aggregate inflation-relevant labor wedge of zero.[9]

[8] Note that the inflation-relevant labor wedges are proportional to labor wedges when $\frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L) = 1$. In the limit as wages become flexible, $\delta \to 0$, there are no cost-push shocks, $\epsilon_t = \epsilon$, and we allow for the appropriate employment subsidy, the aggregate inflation- and welfare-relevant labor wedges coincide and are both zero.

[9] Equation (6) implies that an increase in the aggregate inflation-relevant labor wedge (on the RHS) leads to an increase in the rate of change of inflation (on the LHS). Since the Phillips curve is a forward-looking equation with a terminal condition, an increase in the rate of change of inflation requires a fall in the actual level of inflation. This is consistent with the interpretation of negative aggregate labor wedges indicating a recession, which generates deflationary pressure. The aggregate inflation-relevant labor wedge can rise either if individual inflation-relevant labor wedges $\tau_t(a,z)$ increase or mass $g_t(a,z)$ shifts to states with high $\tau_t(a,z)$.

2.3 Final Good Producer

A representative firm produces the final consumption good using labor,

$$Y_t = A_t N_t, \tag{7}$$

where total factor productivity (TFP) A_t is potentially time-varying and a source of exogenous productivity shocks. Under perfect competition and flexible prices, profits from production are zero and the marginal cost of labor is equal to its marginal product, with

$$w_t = A_t, \tag{8}$$

so the real wage w_t is equal to the marginal rate of transformation (MRT) A_t.

2.4 Government

The role of fiscal policy is deliberately minimal. There is no government spending and no debt, with bonds in zero net supply. The fiscal authority pays an employment subsidy $\tau^L z_t w_t n_t$ to households with labor productivity z_t. To balance the budget, it raises a lump-sum tax also proportional to labor productivity as well as aggregate labor income. Households therefore receive a net fiscal rebate of $T_t(z_t) = \tau^L z_t w_t n_t - \tau^L z_t w_t N_t = 0$. Our focus is instead on the monetary authority, which optimally sets the path of nominal interest rates $\{i_t\}_{t\geq 0}$. This is the only policy instrument of the planner. A Fisher relation holds, with

$$r_t = i_t - \pi_t, \tag{9}$$

where π_t is consumer price inflation. Finally, we can relate price inflation to wage inflation by differentiating equation (8), which yields

$$\pi_t = \pi_t^w - \frac{\dot{A}_t}{A_t}. \tag{10}$$

2.5 Equilibrium and Implementability

Definition 1. (**Competitive Equilibrium**) *Given an initial distribution over household bond holdings and idiosyncratic labor productivities, $g_0(a,z)$, and given predetermined sequences of monetary policy $\{i_t\}$ and shocks $\{A_t, \rho_t, \epsilon_t\}$, an equilibrium is defined as paths for prices $\{\pi_t^w, \pi_t, w_t, r_t\}$, aggregates $\{Y_t, N_t, C_t, B_t\}$, individual allocation rules $\{c_t(a,z)\}$, and the distribution $\{g_t(a,z)\}$ such that households optimize, unions* *optimize, labor is rationed, firms optimize, and markets for goods and bonds clear, that is,*

$$Y_t = C_t = \iint c_t(a,z)\, g_t(a,z)\, da\, dz \tag{11}$$

$$0 = B_t = \iint a\, g_t(a,z)\, da\, dz. \tag{12}$$

The Ramsey problems we study in Sections 3 and 4 take a primal approach: the planner chooses among those competitive equilibria that are implementable by policy. We formally derive these implementability conditions in Lemma 12 in Appendix A and show there that they comprise five equations—three at the individual level and two at the aggregate level.

At the individual level, the planner must respect individuals' consumption-savings decisions. These are encoded in the household's standard first-order condition for consumption, which equates marginal utility of consumption with the private marginal value of wealth,

$$u'(c_t(a,z)) = \partial_a V_t(a,z). \tag{13}$$

Because equation (13) features the private lifetime value $V_t(a,z)$, the planner must also respect its evolution over time. A standard Bellman equation—or Hamilton-Jacobi-Bellman (HJB) equation in continuous time—relates current lifetime value $V_t(a,z)$ to flow utility and continuation value,

$$\rho_t V_t(a,z) = \underbrace{u(c_t(a,z)) - v(N_t) - \frac{\delta}{2}(\pi_t^w)^2}_{\text{Flow Utility}} + \underbrace{\partial_t V_t(a,z) + \mathcal{A}_t V_t(a,z)}_{\text{Continuation Value}}. \tag{14}$$

The continuation value from state (a,z) at time t is $\mathbb{E}_t\left[\frac{dV_t(a,z)}{dt}\right] = \partial_t V_t(a,z) + \mathcal{A}_t V_t(a,z)$, where $\mathcal{A}_t$ denotes the infinitesimal generator of the process (a_t, z_t), formally defined in equation (60) in Appendix A.4.[10] Finally, the planner internalizes that a change in policy affects the evolution of the cross-sectional household distribution, characterized in continuous time by the Kolmogorov forward (KF) equation

$$\partial_t g_t(a,z) = \mathcal{A}_t^* g_t(a,z), \tag{15}$$

where $\mathcal{A}_t^*$ denotes the adjoint of $\mathcal{A}_t$.[11] The KF equation tracks the movement of households across individual states (a,z) over time. The relationship between the generator $\mathcal{A}_t$ and its adjoint $\mathcal{A}_t^*$ connects the Bellman equation (14) and the KF equation (15): Under a law of large numbers, a household's rational expectations over future transitions across states must be consistent with the actual evolution of the cross-sectional distribution.

[10] Using households' Bellman equations and consumption-savings optimality conditions as implementability conditions instead of consumption Euler equations is critical to derive analytical insights.

[11] The adjoint of an operator can be seen as a generalization of the transpose of a matrix.

At the aggregate level, the implementability conditions comprise the aggregate resource constraint that combines the goods market clearing condition (11) with the production function (7),

$$\iint c_t(a,z)\, g_t(a,z)\, da\, dz = A_t N_t, \tag{16}$$

as well as the New Keynesian wage Phillips curve (6).

Comparison benchmarks. Our HANK model nests two useful benchmarks. First, we relate our results to the RANK limit of our model (Appendix E). Second, we define the flexible-wage benchmark as the limit of our economy as $\delta \to 0$, analogous to the flexible-price limit in the canonical New Keynesian analysis (Appendix A.3). We refer to natural output as the output that obtains in the flexible-wage limit, denoted $\tilde{Y}_t$.

2.6 Sources of Suboptimality

To conclude the description of the model, we discuss its four sources of suboptimality.[12]

1. First, monopolistic competition drives a wedge between the real wage, which is equal to the marginal rate of transformation (MRT), A_t, and households' average marginal rate of substitution (MRS) between consumption and labor. The appropriate employment subsidy may offset this wedge in steady state.

2. Second, nominal wage rigidity implies that the economy's average MRS can converge only gradually to the MRT in response to shocks. Moreover, wage adjustment costs represent a direct deadweight loss (utility cost to households).

3. Third, our model also features labor rationing. In the absence of aggregate shocks and with the appropriate employment subsidy, an appropriate notion of average MRS is equal to the MRT in our economy. However, individual MRS are not equalized across households because all households are required to work the same hours.

4. Finally, and most importantly, there are incomplete markets for risk: Noncontingent bonds are the only financial asset in this economy and households face a borrowing constraint, jointly restricting their ability to self-insure against idiosyncratic earnings risk. Both forms of incompleteness imply that households' marginal rates of substitution are not equalized across periods and states.

[12] This subsection is meant to parallel Section 4 of Khan et al. (2003) and Chapter 4.2 of Galí (2015), which discuss sources of suboptimality in RANK economies.

The first two sources of inefficiency exactly mirror those in the standard New Keynesian model. Labor rationing and incomplete markets, on the other hand, are unique to the heterogeneous-agent environment. While labor rationing is not critical for the insights of this paper, and could be eliminated at the cost of complicating the labor market block, the presence of incomplete markets is central to our analysis. Due to these two inefficiencies, the flexible-wage allocation is no longer first-best in a HANK economy even with the appropriate employment subsidy.

3 Optimal Monetary Policy under Discretion

We structure our analysis of optimal monetary policy to parallel that of Clarida et al. (1999), starting with policy under discretion in Section 3 and studying policy with commitment in Section 4. Throughout the paper, we adopt an equal-weighted utilitarian welfare criterion. While this assumption is not innocuous, it is a natural starting point—see Dávila and Schaab (2022) for a systematic study of welfare criteria in general dynamic stochastic environments.[13]

Ramsey problem with finite commitment horizon. Under discretion, a planner has control over policy in the present and takes future policy—under the control of a future planner—as well as agents' expectations as given. In discrete time, it is straightforward to associate the present with period t and the future with periods $t+1$ and onwards (Clarida et al., 1999). To stay as close as possible to this notion of policy under discretion in continuous time, we introduce a *Ramsey problem with finite commitment horizon*.

Formally, we consider a planner who exercises control (and has commitment) over policy over some finite time horizon—the analog of the time interval $[t, t+1)$ in discrete time. At the transition time, which occurs at the transition rate ψ, the present planner is replaced by another who sets policy going forward until she herself is again replaced. Planners do not honor promises made by previous planners. We denote the times at which planners transition by $\{\tau_n\}_{n=0}^{\infty}$, with $\tau_0 = 0$ the starting time of the first planner.[14]

Definition 2. (**Ramsey Problem with Finite Commitment Horizon**) *A Ramsey planner with finite commitment horizon $[0, \tau_1)$ chooses allocations, prices, and policy*

$$\boldsymbol{X} = \{c_t(a,z), V_t(a,z), g_t(a,z), N_t, \pi_t^w, i_t\}_{t=0}^{\tau_1}$$

[13] In ongoing work, we study optimal monetary policy and central bank mandates under alternative welfare criteria (Dávila and Schaab, 2023).

[14] Formally, our infinitesimal discretion approach merges insights from Marcet and Marimon (2019) with the continuous-time results of Harris and Laibson (2013). Schaumburg and Tambalotti (2007) study a similar planning problem with finite commitment horizon in a RANK model in discrete time. See Appendix C for details.

as well as multipliers

$$\boldsymbol{M} = \{\phi_t(a,z), \chi_t(a,z), \lambda_t(a,z), \mu_t, \theta_t\}_{t=0}^{\tau_1}$$

to maximize social welfare subject to implementability conditions (6, 13 – 16), taking as given the initial cross-sectional distribution $g_0(a,z)$ as well as future policy. That is,

$$\mathcal{W}_0(g_0) = \min_{\boldsymbol{M}} \max_{\boldsymbol{X}} \mathbb{E}_0 \left[L(0,\tau_1,g_0) + e^{-\int_0^{\tau_1} \rho_s ds} \mathcal{W}_{\tau_1}(g_{\tau_1}) \right] \tag{17}$$

where the expectation $\mathbb{E}_0$ is over the transition time τ_1, $\mathbb{E}_0\left[e^{-\int_0^{\tau_1}\rho_s ds}\mathcal{W}_{\tau_1}(g_{\tau_1})\right]$ denotes the expected discounted continuation value, and $L(t_1,t_2,g_{t_1})$ is the planner's Lagrangian over the horizon $[t_1,t_2)$, given an initial cross-sectional distribution $g_{t_1}(a,z)$:

$$\begin{aligned} L(t_1,t_2,g_{t_1}) = \int_{t_1}^{t_2} e^{-\int_{t_1}^{t} \rho_s ds} \Bigg\{ \iint \Bigg\{ & U_t(a,z)\, g_t(a,z) \\ & + \phi_t(a,z)\Big[-\rho_t V_t(a,z) + U_t(a,z) + \partial_t V_t(a,z) + \mathcal{A}_t V_t(a,z) \Big] \\ & + \chi_t(a,z)\Big[u'(c_t(a,z)) - \partial_a V_t(a,z) \Big] \\ & + \lambda_t(a,z)\Big[-\partial_t g_t(a,z) + \mathcal{A}_t^* g_t(a,z) \Big] \Bigg\}\, da\, dz \\ & + \mu_t \Big[\iint (c_t(a,z) - A_t z N_t) g_t(a,z)\, da\, dz \Big] \\ & + \theta_t \Big[-\partial_t \pi_t^w + \rho_t \pi_t^w + \frac{\epsilon_t}{\delta} \iint \tau_t(a,z) g_t(a,z)\, da\, dz \Big] \Bigg\} dt, \end{aligned} \tag{18}$$

where $U_t(a,z) = u(c_t(a,z)) - v(N_t) - \frac{\delta}{2}(\pi_t^w)^2$. The operators $\mathcal{A}_t$ and $\mathcal{A}_t^*$ are defined in Appendix A.4.

In the remainder of the paper, we focus on two limits of this finite-horizon Ramsey problem: First, as $\psi \to 0$, planners never transition. In fact, the first planner stays in power forever. The resulting Ramsey problem is thus simply the standard Ramsey problem with an infinite commitment horizon as we show in Section 4.1. Second, as $\psi \to \infty$, planners transition increasingly frequently and their commitment horizon becomes vanishingly small. This is the limit we associate with *policy under discretion* in continuous time.[15]

[15] Formally, for a given ψ, we study the Markov perfect equilibrium of the game played by a sequence of Ramsey planners with finite commitment horizon. It comprises i) paths for prices, π_t^w, aggregates, N_t, individual consumption allocations and value functions, $c_t(a,z)$ and $V_t(a,z)$, as well as cross-sectional distributions, $g_t(a,z)$, that satisfy the competitive equilibrium conditions (6, 13 – 16) given paths for policy, i_t, and shocks, $(A_t, \rho_t, \epsilon_t)$, as well as an initial distribution $g_0(a,z)$; ii) a path of interest rate policy i_t; and iii) a sequence of multiplier functions, $\phi_t(a,z)$, $\chi_t(a,z)$,

The implementability conditions encoded in the Lagrangian (18) include two forward-looking constraints: the individual Bellman equations and the Phillips curve, respectively associated with the multipliers $\phi_t(a,z)$ and θ_t. In Clarida et al. (1999), lack of commitment implies the planner takes as given next-period inflation expectations. In our continuous-time formulation, the planner similarly takes as given expectations about inflation and value assignments beyond the current commitment horizon. Formally, this is encoded in the terminal conditions π_{τ_1} and $V_{\tau_1}(a,z)$ that the planner faces, and which are themselves part of the solution of Markov perfect equilibrium as in discrete time.[16]

Optimal monetary policy under discretion: optimality conditions and interpretation. We now summarize the necessary first-order conditions that characterize optimal monetary policy under discretion, i.e., in the limit of the finite-horizon Ramsey problem (17) as $\psi \to \infty$.

Proposition 1. (**Policy under Discretion: Optimality Conditions**) *The necessary first-order conditions that characterize optimal monetary policy under discretion are given by*

$$\rho_t \lambda_t(a,z) = U_t(a,z) + \mu_t(c_t(a,z) - A_t z N_t) + \partial_t \lambda_t(a,z) + \mathcal{A}_t \lambda_t(a,z) \tag{19}$$

$$0 = u'(c_t(a,z)) - \partial_a \lambda_t(a,z) + \mu_t - \tilde{\chi}_t(a,z) \tag{20}$$

$$0 = \iint z \partial_a \lambda_t(a,z) g_t(a,z)\, da\, dz + \underline{z}\, \zeta_t^{HTM} g_t(\underline{a},\underline{z})\, da\, dz - \mu_t - \frac{v'(N_t)}{A_t} \tag{21}$$

$$0 = \iint a \partial_a \lambda_t(a,z) g_t(a,z)\, da\, dz + \underline{a}\, \zeta_t^{HTM} g_t(\underline{a},\underline{z})\, da\, dz \tag{22}$$

where

$$\zeta_t^{HTM} \equiv u'(c_t(\underline{a},\underline{z})) - \partial_a \lambda_t(\underline{a},\underline{z}) + \mu_t \quad \text{and} \quad \tilde{\chi}_t(a,z) \equiv -u''(c_t(a,z)) \frac{\chi_t(a,z)}{g_t(a,z)}.$$

In the limit as $\psi \to \infty$, the paths of the multipliers on forward-looking implementability conditions converge to $\theta_t \to 0$ and $\phi_t(a,z) \to 0$ for all t and (a,z).

Equations (19) through (22) correspond to the first-order conditions for the cross-sectional distribution $g_t(a,z)$, individual consumption $c_t(a,z)$, aggregate activity N_t, and the nominal interest rate i_t,

$\lambda_t(a,z)$, μ_t, and θ_t that solve (17). Policy under discretion corresponds to the limit of this equilibrium as $\psi \to \infty$ and planners have vanishingly small commitment horizons.

[16] In this problem, there are two direct linkages between the present finite-horizon Ramsey problem and future policy. The first is encoded in the continuation value $\mathcal{W}_{\tau_1}(g_{\tau_1})$: the present Ramsey planner internalizes that choosing policy today affects the evolution of the cross-sectional distribution and, therefore, the initial condition $g_{\tau_1}(a,z)$ of the future planner at the time of transition. Second, taking future policy as given implies that the present planner faces terminal conditions for each forward-looking constraint. Concretely, the planner takes as given inflation π_{τ_1} and values $V_{\tau_1}(a,z)$ at the time of transition. This is analogous to the setup in Clarida et al. (1999), where the present planner takes as given inflation expectations.

respectively. Judiciously combining these conditions allows us to characterize the properties of optimal monetary policy under discretion.

To that end, we start by providing an economic interpretation of the three non-zero multipliers $\chi_t(a,z)$, $\lambda(a,z)$, and μ_t. First, the multiplier $\chi_t(a,z)$ corresponds to the social shadow value of relaxing households' consumption-savings decisions. When $\chi_t(a,z) > (<)\ 0$, the planner perceives that households in state (a,z) save (consume) too much relative to an environment in which the planner could perfectly manage consumption-savings decisions. The multiplier $\chi_t(a,z)$ acts as the shadow penalty that ensures the planner respects private consumption-savings decisions. Second, the multiplier $\lambda_t(a,z)$ corresponds to the social shadow value of increasing the mass of households in state (a,z). As we show below, this multiplier represents the social lifetime value of a household in state (a,z). Third, the multiplier μ_t corresponds to the social shadow value of increasing aggregate excess demand. When $\mu_t > (<)\ 0$, the planner perceives that increasing (reducing) aggregate demand or reducing (increasing) aggregate supply is socially beneficial.[17]

After introducing the non-zero multipliers, we interpret the four optimality conditions (19) through (22). First, equation (19) implies that $\lambda_t(a,z)$ defines the social lifetime value of a household.[18] The difference between private lifetime value (14) and social lifetime value under discretion (19) is given by the term $\mu_t(c_t(a,z) - A_t z N_t)$, which captures the contribution of a household in state (a,z) to aggregate excess demand. Intuitively, households for whom $c_t(a,z) > A_t z N_t$ put positive pressure on aggregate excess demand since their contribution to aggregate demand, $c_t(a,z)$, is higher than their contribution to aggregate supply, $w_t z N_t$, which is socially desirable (undesirable) when $\mu_t > (<)\ 0$. Equation (19) allows us to characterize the social marginal value of wealth—a key input for the remaining optimality conditions—as

$$\partial_a \lambda_t(a,z) = \partial_a V_t(a,z) + \mathcal{M}_t(a,z)\mu_t, \tag{23}$$

where $\mathcal{M}_t(a,z)$ denotes an operator, introduced in Appendix A.4, that acts on the path of multipliers μ_t. The difference between the private and the social marginal value of wealth, $\mathcal{M}_t(a,z)\mu_t$, can be interpreted as the present discounted value of the contribution of future consumption to aggregate excess demand induced by an increase in the household's wealth at time t.[19]

Second, equation (20) has the interpretation of a social consumption-savings optimality con-

[17] The interpretation of ξ_t^{HTM} is the social marginal value of giving a dollar of (unearned) income to every hand-to-mouth household, whose mass is $g_t(\underline{a},\underline{z})$. For our proofs, we assume that $z_t \in \{\underline{z}, \overline{z}\}$ follows a two-state Markov chain, so only households at the borrowing constraint $\underline{a}$ and with the low earnings realization $\underline{z}$ are hand-to-mouth (Achdou et al., 2022). For the perturbations that feature ξ_t^{HTM}, we are holding fixed consumption for all households when, e.g., perturbing i_t, except for the hand-to-mouth households: The planner must respect the borrowing constraint and cannot freely choose the consumption of households at the borrowing constraint. We therefore consider perturbations where, only for the hand-to-mouth household, a change in income leads to a change in consumption.

[18] The multiplier $\lambda_t(a,z)$ takes the form of an HJB equation and can alternatively be written as $\rho\lambda_t = U_t + \mu_t(c_t - w_t z N_t) + \mathbb{E}_t\left[\frac{d\lambda_t}{dt}\right]$, suppressing the dependence on (a,z).

[19] $\mathcal{M}_t(a,z)$ is positive and bounded between 0 and 1 under mild regularity conditions. See Appendix A.4.

dition. Like households, the planner trades off the direct benefit of increasing consumption, $u'(c_t(a,z))$, against the marginal value of having higher future assets from savings, which the planner values at $\partial_a \lambda_t(a,z)$. Moreover, increasing consumption increases aggregate excess demand, which is socially desirable (undesirable) when $\mu_t > (<) \ 0$. Since the planner must also respect private consumption-savings decisions, $\tilde{\chi}_t(a,z) > (<) \ 0$ acts as an additional social shadow cost (benefit) of consumption that ensures that the individual consumption-savings optimality condition is satisfied.

Third, equation (21) has the interpretation of an aggregate activity condition, and represents the planner's valuation of an increase in hours worked by all households, which has three components.[20] First, household wealth increases in proportion to the effective wage $zw_t = zA_t$. The planner values this effect using the social marginal value of wealth for unconstrained households, $\partial_a \lambda_t(a,z)$, and the social marginal value of consumption for constrained households, $\partial_a \lambda_t(\underline{a},\underline{z}) + \xi_t^{HTM}(\underline{a},\underline{z})$.[21] Second, aggregate supply increases by $\iint zA_t g_t(a,z)\,da\,dz = A_t$, which the planner values at the shadow value of aggregate excess demand, μ_t. Third, the planner accounts for households' direct disutility from working more, $\iint v'(N_t) g_t(a,z)\,da\,dz = v'(N_t)$. When choosing optimal aggregate economic activity, the planner trades off these three forces.

Finally, equation (22) represents the planner's valuation of an increase in the nominal interest rate. In this environment, an increase in the interest rate redistributes dollars across households in proportion to their bond holdings a. The planner values such redistribution in dollars according to $\partial_a \lambda_t(a,z)$ for unconstrained households and $\partial_a \lambda_t(\underline{a},\underline{z}) + \xi_t^{HTM}$ for constrained households. This term captures the distributive pecuniary effect of a change in interest rates, which is central to the determination of optimal monetary policy, as we show next.[22]

Targeting rule and inflationary bias. Combining the optimality conditions just described allows us to characterize a *targeting rule* for optimal monetary policy under discretion in Proposition 2. This targeting rule illustrates the forces that optimal monetary policy under discretion balances in our HANK model and facilitates the comparison of our results to the canonical analysis of monetary policy in RANK, which we show to be nested as a special case by our targeting rule. We

[20] Since the planner must respect how the union allocates labor, the planner can only consider perturbations that change hours worked for all households symmetrically.

[21] The social value of increasing the consumption of a household at the borrowing constraint corresponds to

$$\partial_a \lambda_t(\underline{a},\underline{z}) + \xi_t^{\text{HTM}} = u'(c_t(\underline{a},\underline{z})) + \mu_t. \tag{24}$$

Intuitively, the planner internalizes that a marginal change in the wealth of a household at the borrowing constraint leads to a one-for-one change in consumption, whose social value is given by the sum of the direct utility benefit, $u'(c_t(\underline{a},\underline{z}))$, and the impact on aggregate excess demand, μ_t.

[22] We use the terminology distributive pecuniary effects as in Dávila and Korinek (2018). That paper shows that distributive pecuniary effects are characterized by i) changes in net asset positions, here a, and ii) differences in valuation, here $\partial_a \lambda_t(a,z)$. As shown in that paper, if the planner valued a dollar across households identically, market clearing would imply that distributive pecuniary effects are zero, so $\iint a \partial_a \lambda_t(a,z) g_t(a,z)\,da\,dz + \underline{a}\,\xi_t^{\text{HTM}}\, g_t(\underline{a},\underline{z})\,da\,dz = 0$.

present a constructive proof of Proposition 2 in Appendix C.2.

Proposition 2. (**Targeting Rule under Discretion**) *Optimal monetary policy under discretion is characterized by the non-linear targeting rule*

$$\underbrace{\iint \left(z u'(c_t(a,z)) - \frac{v'(N_t)}{A_t} \right) g_t(a,z)\, da\, dz}_{\text{Aggregate Labor Wedge}} = \Omega_t^{\mathrm{D}} \underbrace{\iint a u'(c_t(a,z)) g_t(a,z)\, da\, dz}_{\text{Distributive Pecuniary Effect}}, \tag{25}$$

where Ω_t^{D}*, given by*

$$\Omega_t^{\mathrm{D}} = \frac{\iint z(1-\mathcal{M}_t(a,z)) g_t(a,z)\, da\, dz - (1-\mathcal{M}_t(\underline{a},\underline{z}))\, \underline{z}\, g_t(\underline{a},\underline{z})}{\iint a(1-\mathcal{M}_t(a,z)) g_t(a,z)\, da\, dz - (1-\mathcal{M}_t(\underline{a},\underline{z}))\, \underline{a}\, g_t(\underline{a},\underline{z})},$$

is positive under mild regularity conditions.

This non-linear targeting rule shows that, under discretion, the utilitarian planner trades off aggregate stabilization against a novel redistribution motive. The LHS of equation (25) is the aggregate labor wedge and represents the aggregate stabilization motive of the planner, while the RHS is the distributive pecuniary effect of interest rate changes. Crucially, the marginal utility of consumption falls with household wealth, so that

$$\iint a u'(c_t(a,z)) g_t(a,z)\, da\, dz = \mathbb{C}ov_{g_t(a,z)}\left(a, u'(c_t(a,z))\right) < 0,$$

where $\mathbb{C}ov_{g_t(a,z)}(a, u'(c_t(a,z)))$ is the cross-sectional covariance between wealth and marginal utility.[23]

Therefore, optimal monetary policy under discretion targets a negative aggregate labor wedge, which is associated with an overheated economy. To illustrate, consider a level of interest rates at which the aggregate labor wedge is zero and policy attains aggregate stabilization. The negative RHS of (25) implies that, relative to the policy stance under consideration, the planner finds it valuable to lower the real interest rate in order to redistribute towards indebted, high marginal utility households. To lower the real rate, the planner lowers the nominal policy rate, which results in a negative aggregate labor wedge, i.e., an overheated economy. Proposition 2 thus offers a novel perspective on optimal monetary policy under discretion in a heterogeneous-agent environment.[24]

[23] While the positive HANK literature has concluded that marginal propensities to consume (MPC) are central for monetary policy transmission, equation (22) highlights that marginal utilities are instead the key direct determinant of the targeting rules for optimal monetary policy. This is the case under discretion, but also with commitment—see Section 4.

[24] In the RANK limit of our economy, the planner's motive to redistribute via distributive pecuniary interest rate effects vanishes. Formally, the RHS of equation (25) goes to 0 in that limit and, as a result, optimal monetary policy under discretion focuses solely on aggregate stabilization, targeting an aggregate labor wedge of 0.

When assuming isoelastic (CRRA) preferences, with $u(c) = \frac{1}{1-\gamma}c^{1-\gamma}$ and $v(n) = \frac{1}{1+\eta}n^{1+\eta}$, the targeting rule under discretion takes the form

$$Y_t = \tilde{Y}_t \times \underbrace{\left(\frac{\epsilon_t}{\epsilon_t - 1} \frac{1}{1+\tau^L} \right)^{\frac{1}{\gamma+\eta}}}_{\text{Desired Markup: } \geq 1} \times \underbrace{\left(1 - \Omega_t^{\mathrm{D}} \frac{\iint a u'(c_t(a,z)) g_t(a,z)\, da\, dz}{\iint z u'(c_t(a,z)) g_t(a,z)\, da\, dz} \right)^{\frac{1}{\gamma+\eta}}}_{\text{Desired Redistribution: } > 1} \tag{26}$$

where $\tilde{Y}_t$ denotes natural output as defined in equation (59). While this output gap targeting rule is more closely connected to the canonical results on optimal monetary policy in RANK, equations (25) and (26) have the same content.

Equation (26) shows that, under discretion, monetary policy targets output to be equal to natural output, i.e., to close the output gap, up to two wedges. The first wedge is the familiar one deriving from monopolistic competition and unions' desired markups, due to which employment may be inefficiently low. Whenever the employment subsidy τ^L is not sufficiently large, this wedge is positive, motivating the planner to raise output above potential to raise employment.

In HANK, a second redistribution wedge emerges, since marginal utility of consumption falls with wealth, i.e., $\mathbb{C}ov_{g_t(a,z)}(a, u'(c_t(a,z))) < 0$. This wedge is therefore strictly positive, encouraging the utilitarian planner under discretion to overheat the economy even further.

An important conclusion of the canonical monetary policy analysis in RANK is that there are no gains from commitment when the planner sets the correct steady state employment subsidy and there are no cost-push shocks. Indeed, in the RANK limit of our economy, $\Omega_t^{\mathrm{D}} \to 0$ and the redistribution wedge vanishes. And when $\frac{\epsilon_t}{\epsilon_t - 1}\frac{1}{1+\tau^L} = 1$, equation (26) collapses to $Y_t = \tilde{Y}_t$: In that case, in RANK, monetary policy under discretion closes the output gap, which also closes the inflation gap and Divine Coincidence obtains even without commitment. In HANK, this is no longer the case as a planner under discretion always has an incentive to overheat the economy due to distributional considerations.

In equilibrium, agents anticipate the planner's incentive to raise output above natural output, rendering the planner's attempt to stimulate the economy futile. Instead, inflation ensues. Proposition 3 shows that a Markov perfect stationary equilibrium features inflationary bias, now exacerbated by the novel redistribution motive.

Proposition 3. (**Inflationary Bias**) *The Markov perfect stationary equilibrium with optimal monetary policy under discretion features inflationary bias, given by*

$$\pi_{ss}^w = \frac{\epsilon}{\delta} A_{ss} N_{ss} \Bigg[\underbrace{\left(1 - \frac{\epsilon - 1}{\epsilon}(1+\tau^L) \right) \Lambda_{ss}}_{\textit{Markup: } \geq 0} - \Omega_{ss}^D \underbrace{\mathbb{C}ov_{g_{ss}(a,z)}\Big(a, u'(c_{ss}(a,z))\Big)}_{\textit{Redistribution: } < 0} \Bigg]. \tag{27}$$

Consistent with our discussion of targeting rules under discretion, inflationary bias emerges from two sources: inefficiently low employment due to markups, as in the RANK limit, and redistribution. Quantitatively, the contribution of the novel redistribution motive is over 4 times larger than that of markups in our calibration exercise—see Figure 1 in Section 4.3.

Our stylized model features a single distributive pecuniary effect associated with adjusting interest rates. In richer models, optimal policy would account for all pecuniary effects and inflationary bias would be determined by the covariance between marginal utility and the aggregate of those effects. While our approach and the logic behind our results apply more generally, the exact quantitative conclusions—including whether policy under discretion features an inflationary or deflationary bias—may not.

Practical implications of optimal monetary policy under discretion. Summing up, our analysis of optimal monetary policy under discretion yields three main takeaways. First, a utilitarian planner has an incentive to run an overheated economy. This occurs because the planner values redistribution toward indebted, high marginal utility households via lower interest rates. The planner trades off this novel redistribution motive against aggregate stabilization. Second, the economy features inflationary bias in the sense of Barro and Gordon (1983). The standard motive to stimulate the economy due to markups is exacerbated by a novel desire to redistribute towards indebted households. When agents anticipate these incentives in equilibrium, both result in elevated inflation. Quantitatively, the redistribution motive is the dominant source of inflationary bias. Third, the markup-correcting employment subsidy, $\frac{\epsilon-1}{\epsilon}(1+\tau^L)=1$, that eliminates inflationary bias in RANK is no longer sufficient to address inflationary bias in HANK.[25]

4 Optimal Monetary with Policy with Commitment

We have shown in Section 3 that the desire to redistribute exacerbates inflationary bias when a utilitarian planner sets optimal policy under discretion. In this section, we characterize optimal monetary policy under commitment. We proceed in three steps.

First, in Section 4.1, we introduce the standard Ramsey problem and characterize the associated Ramsey plan and stationary Ramsey plan. In particular, we show in Section 4.2 that the optimal stationary equilibrium under commitment features zero inflation, eliminating the inflationary bias of policy under discretion in the long run.

Second, in Section 4.3, we show that while the full-commitment standard Ramsey problem eliminates inflationary bias in the long run, it still suffers from inflationary bias in the short run. This is due to a "time-0 problem" (Kydland and Prescott, 1980) associated with two dimensions

[25] In principle, it is possible to set a sufficiently large employment subsidy τ_L so that inflationary bias is zero.

of time inconsistency. In order to find a "timeless" planning solution, we extend the approach of Marcet and Marimon (2019) to our setting (i.e., continuous-time heterogeneous-agent economies) by introducing timeless penalties for each forward-looking implementability condition. We then define a timeless Ramsey problem, which augments the standard Ramsey problem with the timeless penalties, and prove that it no longer suffers from a time-0 problem—there is no incentive to deviate from the stationary Ramsey plan in the absence of shocks. Hence, the timeless Ramsey problem resolves inflationary bias in both the short run and the long run. In Sections 4.4 and 4.5, we study the determinants of the timeless penalty and contrast implementations of optimal policy based on penalties and targets.

Finally, in Section 4.6, we characterize optimal stabilization policy under the timeless Ramsey problem, which allows us to separate the pure stabilization motive from the time-0 problem.

Each of these three steps isolates one important dimension of optimal monetary policy design. First, characterizing the stationary Ramsey plan allows us to solve for optimal long-run policy, with which the planner addresses distortions in a stationary equilibrium. Second, characterizing the timeless Ramsey plan and the timeless penalties that support it allows us to isolate the planner's incentives to deviate from the stationary Ramsey plan in the short run due to the time-0 problem. Finally, by characterizing optimal stabilization policy with the appropriate timeless penalties, we isolate the planner's pure stabilization motive, no longer confounded by long-run distortions and time inconsistency considerations.

4.1 Standard Ramsey Problem and Ramsey Plan

The standard Ramsey problem corresponds to the limit of problem (17) as $\psi \to 0$, i.e., as the commitment horizon becomes infinite. We state the full problem in Appendix B.1 for convenience.

Definition 3. (Standard Primal Ramsey Problem / Ramsey Plan)

a) *The standard primal Ramsey problem solves*

$$\min_{\{\phi_t(a,z),\chi_t(a,z),\lambda_t(a,z),\mu_t,\theta_t\}} \max_{\{c_t(a,z),V_t(a,z),g_t(a,z),N_t,\pi_t^w,i_t\}} L^{\text{SP}}(g_0) \tag{28}$$

where $L^{\text{SP}}(g_0)$ denotes the standard primal Lagrangian, given an initial distribution of bond holdings and idiosyncratic labor productivity $g_0(a,z)$:

$$L^{\text{SP}}(g_0) = \lim_{T\to\infty} L(0,T,g_0). \tag{29}$$

b) *A Ramsey plan corresponds to the solution of this problem and comprises i) paths for prices, π_t^w, aggregates, N_t, individual consumption allocations and value functions, $c_t(a,z)$ and $V_t(a,z)$, and* *the cross-sectional distribution, $g_t(a,z)$, that satisfy the implementability conditions given paths for interest rates, i_t, and shocks, $(A_t, \rho_t, \epsilon_t)$, as well as an initial distribution, $g_0(a,z)$; ii) a path of interest rate policy i_t; and iii) paths for the multiplier functions, $\phi_t(a,z)$, $\chi_t(a,z)$, $\lambda_t(a,z)$, μ_t, and θ_t that solve (28).*

Proposition 4 summarizes the optimality conditions that characterize the standard Ramsey plan. Our derivation relies on a variational approach, formally developed in Appendix B.[26]

Proposition 4. (**Standard Primal Ramsey Problem: Optimality Conditions**) *The optimality conditions for the standard primal Ramsey problem are given by*

$$\partial_t \phi_t(a,z) = -\mathcal{A}_t^* \phi_t(a,z) + \partial_a \chi_t(a,z) \tag{30}$$

$$\rho_t \lambda_t(a,z) = U_t(a,z) + \mu_t(c_t(a,z) - A_t z N_t) + \theta_t \frac{\epsilon_t}{\delta} \tau_t(a,z) + \partial_t \lambda_t(a,z) + \mathcal{A}_t \lambda_t(a,z) \tag{31}$$

$$0 = u'(c_t(a,z)) - \partial_a \lambda_t(a,z) + \mu_t + \theta_t \frac{\epsilon_t}{\delta} \frac{d\tau_t(a,z)}{dc_t(a,z)} - \tilde{\chi}_t(a,z) \tag{32}$$

$$\begin{aligned} 0 = & \iint z \partial_a \lambda_t(a,z) g_t(a,z)\, da\, dz + \underline{z} \xi_t^{HTM} g_t(\underline{a}, \underline{z})\, da\, dz - \mu_t - \frac{v'(N_t)}{A_t} \\ & + \iint \phi_t(a,z) \left(z u'(c_t(a,z)) - \frac{v'(N_t)}{A_t} \right) da\, dz + \theta_t \frac{\epsilon_t}{\delta} \iint \frac{1}{A_t} \frac{d\tau_t(a,z)}{dN_t} g_t(a,z)\, da\, dz \end{aligned} \tag{33}$$

$$\dot{\theta}_t = \delta \pi_t^w \left(1 + \iint \phi_t(a,z)\, da\, dz \right) \tag{34}$$

$$0 = \iint \left(a \partial_a \lambda_t(a,z) g_t(a,z) + a \phi_t(a,z) \partial_a V_t(a,z) \right) da\, dz + \underline{a} \xi_t^{HTM} g_t(\underline{a}, \underline{z}) da\, dz \tag{35}$$

where

$$\xi_t^{HTM} = u'(c_t(\underline{a}, \underline{z})) - \partial_a \lambda_t(\underline{a}, \underline{z}) + \mu_t + \theta_t \frac{\epsilon_t}{\delta} \frac{d\tau_t(\underline{a}, \underline{z})}{dc_t(\underline{a}, \underline{z})} \quad \text{and} \quad \tilde{\chi}_t(a,z) = -u''(c_t(a,z)) \frac{\chi_t(a,z)}{g_t(a,z)}$$

as well as a set of initial conditions for the multipliers on forward-looking implementability conditions

$$0 = \theta_0 \tag{36}$$

$$0 = \phi_0(a,z). \tag{37}$$

The optimality conditions (30) – (32) hold everywhere in the interior of the idiosyncratic state space. For a formal treatment of boundary conditions, see Appendices B.2 through B.4

[26] In Appendix B.1, we first present a heuristic derivation of Proposition 4 in continuous time for the interior of the idiosyncratic state space. A formal treatment of boundary conditions follows in Appendices B.2 through B.4.

Equations (30) through (34) respectively correspond to the optimality conditions for i) the value function, ii) the cross-sectional distribution, iii) consumption, iv) aggregate labor, and v) wage inflation. Equation (35) corresponds to the optimality condition for the nominal interest rate.

The optimality conditions for the standard Ramsey problem (Proposition 4) can be seen as an augmented version of the optimality conditions for policy under discretion (Proposition 1). In particular, the multipliers $\chi_t(a,z)$, $\lambda_t(a,z)$, and μ_t, as well as ζ_t^{HTM}, have the same interpretation as in the discretion case—see pages 13 and 14. With commitment, the planner can counteract inflationary bias in the long run by making promises. These promises are encoded in the multipliers on the two forward-looking implementability conditions: θ_t for the Phillips curve and $\phi_t(a,z)$ for households' Bellman equations. Under discretion, these multipliers vanish. In fact, equations (31), (32), (33), and (35) correspond exactly to the optimality conditions (19) – (22) for policy under discretion when $\theta_t = 0$ and $\phi_t(a,z) = 0$.

The multiplier associated with the Phillips curve, θ_t, has the interpretation of a penalty (reward) associated with increasing inflation when $\theta_t > (<)\ 0$. Hence, we refer to θ_t as an *inflation penalty*.[27] Any perturbation that increases (decreases) the aggregate inflation-relevant labor wedge leads to deflationary (inflationary) pressure through the Phillips curve (see footnote 9). Changes in this labor wedge at date t are thus valued both directly and indirectly due to their effect on past inflation. The inflation penalty θ_t encodes the cumulative valuation of these resulting changes in past inflation. It thus appears in all perturbations that affect the aggregate inflation-relevant labor wedge: (i) The equation for social lifetime value (31) considers an increase in the mass of households in state (a,z). If the individual inflation-relevant labor wedge is positive (negative) for these states, $\tau_t(a,z) > (<)\ 0$, this perturbation puts negative (positive) pressure on inflation. (ii) In the social consumption-savings optimality condition (32), increasing consumption generates inflationary pressure since it reduces individual inflation-relevant labor wedges, $\frac{d\tau_t(a,z)}{dc_t(a,z)} < 0$. Finally (iii) in the aggregate activity condition (33), increasing hours worked also generates inflationary pressure by reducing individual inflation-relevant labor wedges, $\frac{d\tau_t(a,z)}{dN_t} < 0$.[28]

The multipliers associated with households' Bellman equation, $\phi_t(a,z)$, have the interpretation of a penalty (reward) associated with an increase in lifetime utility when $\phi_t(a,z) < (>)\ 0$. Hence, we

[27] As the multiplier on a forward-looking equation, θ_t encodes the impact on the Lagrangian (welfare) at time 0 from a change in inflation at time t. This is analogous to multipliers on backward-looking equations; for example, the multiplier on the capital accumulation equation in the neoclassical growth model encodes the present discounted value of a change in the capital stock. A change in inflation at time t affects inflation at all prior dates $s \in [0,t)$ through the forward-looking Phillips curve. These changes in inflation appear in the time-0 Lagrangian and are valued by the planner. If the planner were to reoptimize at time t, she would disregard these effects on past inflation. The inflation penalty θ_t encodes the associated welfare impact so that, if the planner reoptimizes at time t when confronted with θ_t, she will behave consistently with time-0 optimization. In summary, θ_t captures the cumulative, backward-looking impact on time-0 welfare resulting from a change in inflation at time t, appropriately discounted to period t.

[28] The inflation penalty θ_t also appears in the definition of ζ_t^{HTM}, the social valuation of increasing wealth for households in state $(\underline{a},\underline{z})$. As in the social consumption-savings condition, shifting wealth towards and thus increasing the consumption of hand-to-mouth households leads to inflationary pressure by lowering the aggregate inflation-relevant labor wedge.

refer to $\phi_t(a,z)$ as *distributional penalties*.[29] The distributional penalties appear in all perturbations that affect date-t lifetime values: (i) In the aggregate activity condition (33), increasing hours worked leads to a change in all households' flow utility that is captured by the individual labor wedges, $zu'(c_t(a,z)) - \frac{v'(N_t)}{A_t}$. (ii) And in the optimality condition for interest rates (35), the distributive pecuniary effect of a rate increase changes a household's utility by $a\partial_a V_t(a,z)$.

Relative to the optimality conditions for policy under discretion, Proposition 4 also features two new equations, (30) and (34). These define the laws of motion for $\phi_t(a,z)$ and θ_t. Equation (30) is central to this paper. It takes the form of a Kolmogorov forward equation augmented to account for births and deaths. This equation shows that the evolution of the distribution of distributional penalties $\phi_t(a,z)$ must be consistent with the evolution of households across idiosyncratic states, via $\mathcal{A}^*$. It also accounts for births of penalties, captured by the term $\partial_a \chi_t(a,z)$, as we explain in Section 4.4.

The optimality condition for inflation (34), which defines the law of motion for θ_t, simplifies to

$$\dot{\theta}_t = \delta \pi_t^w \tag{38}$$

whenever the distributional penalties add up to zero, $\iint \phi_t(a,z)\, da\, dz = 0$. We have proven that this is the case at a stationary Ramsey plan (defined below) and strongly conjecture that this condition holds at all times, which we have verified numerically.

4.2 Optimal Long-Run Policy

We start unpacking the implications of Proposition 4 by characterizing the optimal long-run policy under commitment. To that end, we define a stationary Ramsey plan, towards which a Ramsey plan may converge when all shocks converge as $t \to \infty$.

Definition 4. (**Stationary Ramsey Plan**) *A stationary Ramsey plan, with* $(A_t, \rho_t, \epsilon_t) = (A_{ss}, \rho_{ss}, \epsilon_{ss})$ *constant, is given by (i) an inflation rate,* π_{ss}^w*, aggregate hours,* N_{ss}*, stationary individual consumption allocations and value functions,* $c_{ss}(a,z)$ *and* $V_{ss}(a,z)$*, and a stationary cross-sectional distribution,* $g_{ss}(a,z)$*; (ii) a stationary Ramsey policy,* i_{ss}*; and (iii) a set of stationary multipliers,* $\phi_{ss}(a,z)$*,* $\lambda_{ss}(a,z)$*,* $\chi_{ss}(a,z)$*,* μ_{ss}*, and* θ_{ss}*, such that the optimality conditions and the implementability conditions for a Ramsey plan are satisfied.*

What are the implications of household heterogeneity for optimal long-run inflation? When policy

[29] Like θ_t, $\phi_t(a,z)$ is a multiplier on a forward-looking implementability condition. Any perturbation that shifts lifetime values $V_t(a,z)$ at date t affects past lifetime values $V_s(a,z)$, for $s \in [0,t)$, through the Bellman equation and, therefore, past consumption decisions $c_s(a,z)$ through the individual consumption-savings condition. The multiplier $\phi_t(a,z)$ encodes the cumulative, backward-looking social valuation of these indirect effects on past consumption that result from a change in value $V_t(a,z)$ at date t.

is set with discretion, the planner's redistribution motive substantially exacerbates inflationary bias. Proposition 5 shows that the stationary Ramsey plan features zero inflation even in the presence of household heterogeneity. Optimal policy under commitment therefore addresses the inflationary bias associated with discretion in the long run.[30]

Proposition 5. (**Optimal Long-Run Policy**) *With commitment, optimal long-run price inflation in both HANK and RANK is zero. That is,* $\pi_{\text{ss}} = \pi_{\text{ss}}^{\text{RA}} = 0$.

Our HANK model features long-run neutrality of monetary policy: In any competitive stationary equilibrium, the real interest rate and the allocation are pinned down by real forces. The only choice that the planner has is the split between nominal interest rate and nominal price inflation for a given real interest rate. Crucially, inflation and the nominal interest rate symmetrically affect households' financial income, which itself is proportional to the real interest rate $r_{\text{ss}} = i_{\text{ss}} - \pi_{\text{ss}}$. Therefore, since maintaining non-zero inflation is costly due to nominal rigidities while adjusting the nominal rate is not, the planner finds it optimal to exclusively use the nominal interest rate in the stationary Ramsey plan while promising to keep inflation at zero.[31] Formally, any stationary Ramsey plan must feature $\dot{\theta}_{\text{ss}} = 0$ since θ_{ss} is constant. Equation (38) then directly implies that optimal long-run inflation is zero, $\pi_{\text{ss}} = \pi_{\text{ss}}^{w} = 0$.

Importantly, the planner can only maintain zero inflation in the long run under commitment. As we discuss in Section 3, the planner always faces a time inconsistent incentive to overheat the economy, both to address the markup distortion and to redistribute toward indebted, high marginal utility households. Under discretion, this incentive is self-defeating as it simply results in inflationary bias. With commitment, the planner promises to keep inflation at zero in the long run in the absence of shocks.

4.3 Time Inconsistency, Timeless Penalty, and the Timeless Ramsey Problem

A planning problem is time inconsistent if the optimality conditions pinning down policy, allocation, and prices at some time t depend on the time at which the optimization takes place. The implementability conditions that constrain the Ramsey problem (28) include two sets of forward-looking conditions: individual Bellman equations and the New Keynesian Phillips curve. Each of these conditions is a source of time inconsistency.[32] While the standard Ramsey planner chooses

[30] Lemma 20 in the Appendix explicitly describes the conditions that characterize a stationary Ramsey plan.

[31] The fact that the optimal long-run policy features zero inflation should be understood as a benchmark. When inflation and the nominal interest rate have a differential impact across households, we may expect an optimal long-run policy that features non-zero inflation. Other frictions, such as those studied in Chari and Kehoe (1999), Khan et al. (2003), and Schmitt-Grohé and Uribe (2010), could also imply a non-zero optimal long-run inflation.

[32] The conditions under which forward-looking implementability conditions lead to time inconsistency in planning problems are well understood since Kydland and Prescott (1977).

policy with commitment from time 0 onwards, time inconsistency still manifests at time 0. This is often referred to as the "time-0 problem" (Kydland and Prescott, 1980).

Formally, the time-0 problem materializes as follows: The optimality conditions for the standard Ramsey problem require the initial conditions $\theta_0 = 0$ and $\phi_0(a,z) = 0$ for all (a,z) because initial inflation π_0^w and lifetime value $V_0(a,z)$ respectively are free. But any stationary Ramsey plan will generically feature $\theta_{ss} \neq 0$ and $\phi_{\text{ss}}(a,z) \neq 0$. This follows directly from the stationary version of equations (33) and (30). Intuitively, the standard Ramsey planner benefits from making promises for inflation and lifetime values in the long run; these are encoded in θ_{ss} and $\phi_{ss}(a,z)$. But at time 0, there are no such past promises. Hence, even if we initialize the economy at the allocation that obtains at the stationary Ramsey plan, i.e., $g_0(a,z) = g_{\text{ss}}(a,z)$, the planner will not set policy to $i_0 = i_{\text{ss}}$ in the absence of shocks, which would keep the economy at the stationary Ramsey plan. This would violate the initial conditions of the standard Ramsey problem, $\theta_0 = 0$ and $\phi_0(a,z) = 0$, and precisely formalizes the time-0 problem in our setting.

We illustrate the time-0 problem in Figure 1, which plots optimal inflation in the absence of shocks under different planning problems.[33] The solid and dashed red lines illustrate the inflationary bias associated with policy under discretion. The solid green line plots inflation under the standard Ramsey plan in the absence of shocks. Inflation converges to zero, thus addressing inflationary bias in the long run (Proposition 5). However, the time-0 problem implies that the standard Ramsey plan still features inflationary bias in the short run. As the world starts at time 0, no past promises constrain the planner, and so she finds it optimal to boost inefficiently low employment due to markups and redistribute towards indebted households, generating inflation in the short run.

Motivated by these observations, we now present a *timeless Ramsey approach* to resolve the time-0 problem. The associated timeless Ramsey plan—the solid blue line in Figure 1—resolves inflationary bias in both the short run and the long run. To that end, we introduce a particular time-0 penalty, the *timeless penalty*. Intuitively, confronting the Ramsey planner with the timeless penalty makes her internalize the promises she herself would like to make for the future. In Proposition 6, we show that the timeless penalty formalizes Woodford (1999)'s timeless perspective in our setting, so that the planner at time 0 behaves as if she had chosen policy with commitment infinitely long ago.

Definition 4. (**Time-0 Penalty**) *We define the time-0 penalty as*

$$\mathcal{T}(\phi,\theta) = \underbrace{\iint \phi(a,z) V_0(a,z)\, da\, dz}_{\text{Distributional Penalty}} \quad - \quad \underbrace{\theta \pi_0^w}_{\text{Inflation Penalty}} \tag{39}$$

[33] Figure 1 extends Figure 7.1 in Woodford (2003) and Figure 2 in Woodford (2010) to an environment with household heterogeneity.

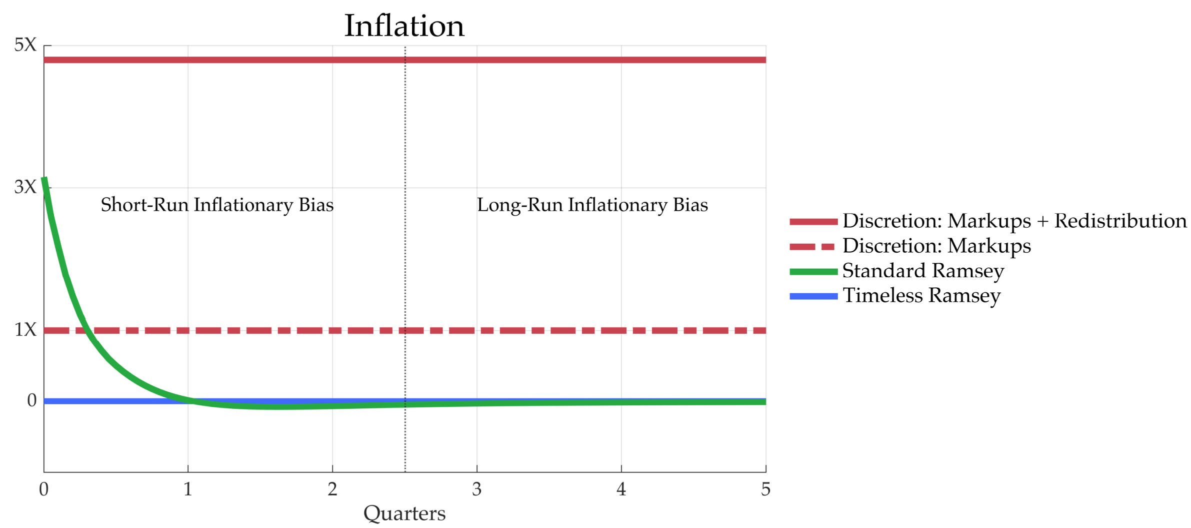


Figure 1. Inflationary Bias in the Short Run and the Long Run

Note. Figure 1 plots optimal inflation in the absence of shocks under different planning problems. We normalize to 1 the standard inflationary bias associated with monopolistic competition (red, dashed). The redistribution motive a utilitarian planner faces under discretion exacerbates long-run inflationary bias by a factor of 4 (red solid). Under the standard Ramsey problem (green), there is no inflationary bias in the long run. Due to the time-0 problem, however, there is still short-run inflationary bias. Only the timeless Ramsey problem (blue) resolves the time-0 problem and addresses inflationary bias in both the short run and the long run.

where we refer to $\phi(a,z)$ as a (per unit) distributional penalty and θ as a (per unit) inflation penalty.[34]

Building on Marcet and Marimon (2019), we introduce a penalty at time 0 for each forward-looking implementability condition that the Ramsey planner faces. We then define the *augmented Ramsey problem* in primal form as a modification of the standard Ramsey problem that confronts the planner with a time-0 penalty. Finally, we define the *timeless Ramsey problem* as the augmented Ramsey problem in which the time-0 penalty is chosen to resolve the time-0 problem. That is, a timeless Ramsey planner has no incentive to deviate from the stationary Ramsey plan at time 0 in the absence shocks.[35]

Definition 5.

(a) (**Augmented Ramsey Problem**) *The augmented Ramsey problem in primal form solves*

$$\min_{\{\phi_t(a,z),\, \chi_t(a,z), \lambda_t(a,z), \mu_t, \theta_t\}} \max_{\{c_t(a,z), V_t(a,z), g_t(a,z), N_t, \pi_t^w, i_t\}} L^{\mathrm{AP}}(g_0, \phi, \theta), \tag{40}$$

[34] To streamline the exposition, we use the term penalty to refer to i) $\mathcal{T}(\phi,\theta)$; ii) its components $\iint \phi(a,z)V_0(a,z)\,da\,dz$ and $\theta\pi_0^w$; and iii) the values of $\phi(a,z)$ and θ. The first three terms are total penalties, while the last two correspond to penalties per unit of lifetime utility and inflation, respectively.

[35] In Appendix B.6, we also characterize the dual form of the augmented and timeless Ramsey problems.

where $L^{\text{AP}}(g_0, \phi, \theta)$ denotes the augmented primal Lagrangian, given an initial distribution g_0 as well as initial penalties ϕ and θ. The augmented primal Lagrangian is defined as

$$L^{\text{AP}}(g_0, \phi, \theta) = L^{\text{SP}}(g_0) + \mathcal{T}(\phi, \theta) \tag{41}$$

where $L^{\text{SP}}(g_0)$ is the standard primal Lagrangian (29) and $\mathcal{T}(\phi, \theta)$ a time-0 penalty (39).

(b) (**Timeless Ramsey Problem**) *The timeless Ramsey problem in primal form is an augmented Ramsey problem in which the time-0 penalty takes the form $\mathcal{T}(\phi, \theta) = \mathcal{T}(\phi_{ss}, \theta_{ss})$. We refer to $\mathcal{T}(\phi_{ss}, \theta_{ss})$ as the timeless penalty, and we define the timeless primal Lagrangian as*

$$L^{\text{TP}} = L^{\text{AP}}(g_{ss}, \phi_{ss}, \theta_{ss}). \tag{42}$$

The Lagrangian of the augmented Ramsey problem $L^{\text{AP}}(g_0, \phi, \theta)$ is defined for arbitrary initial penalties ϕ and θ. For example, the augmented Ramsey problem nests the standard one when we set $\phi(a, z) = 0$ and $\theta = 0$, implying $L^{\text{AP}}(g_0, 0, 0) = L^{\text{SP}}(g_0)$. It will become clear in the following that only the timeless Ramsey problem, in which $\phi(a, z) = \phi_{\text{ss}}(a, z)$ and $\theta = \theta_{\text{ss}}$, resolves the time-0 problem. In that case, the timeless penalty encodes precisely the promises that the Ramsey planner would like to make in the long run, i.e., in the stationary Ramsey plan. Intuitively, the timeless penalty introduces an artificial cost that, on the margin, exactly offsets the marginal benefit of time-inconsistent deviations from the stationary Ramsey plan at time 0. Our approach shows that it is possible to transform the standard Ramsey problem into a timeless problem by simply augmenting it with the timeless penalty $\mathcal{T}(\phi_{ss}, \theta_{ss})$.[36]

Formally, a time-0 penalty enforces a new set of initial conditions on the two multipliers associated with forward-looking implementability conditions. In continuous time, the choice of initial lifetime values $V_0(a, z)$ and inflation π_0^w is free under the standard Ramsey problem, which gives rise to the initial conditions (36) and (37) of the standard Ramsey plan. Indeed, $\phi_0(a, z) = 0$ and $\theta_0 = 0$ is precisely an expression of the fact that the standard Ramsey planner is not bound by past promises on lifetime values and inflation at time 0, even though she would like to bind her future self by making such promises. By augmenting the Ramsey problem with the time-0 penalty

[36] Formally, a planner solving problem (42) sets policy at time 0 as if she had set policy with commitment infinitely far in the past. The timeless Ramsey plan associated with (42) corresponds exactly to optimal policy from a timeless perspective (Woodford, 1999). A timeless policy, as defined by Woodford (2010), represents a policy that

> "*even if not what the policy authority would choose if optimizing afresh at a given date t, [...] it should have been willing to commit itself to follow from that date t onward if the choice had been made at some indeterminate point in the past, when its choice would have internalized the consequences of the policy for expectations prior to date t.*"

$\mathcal{T}(\phi, \theta)$, we technically enforce new initial conditions,

$$\phi_0(a,z) = \phi(a,z) \tag{43}$$

$$\theta_0 = \theta, \tag{44}$$

which in turn constrain the planner's choice of initial lifetime values and inflation. The optimality conditions of the augmented Ramsey problem comprise exactly the same equations as in Proposition 4, i.e., equations (30) through (35), except that the initial conditions for the multipliers are now given by $\phi_0(a,z) = \phi(a,z)$ and $\theta_0 = \theta$. In economic terms, it is as if the penalties $\phi(a,z)$ and θ enforce artificial past promises. And when we initialize the time-0 penalty at $\phi = \phi_{\text{ss}}$ and $\theta = \theta_{\text{ss}}$, it is as if the timeless Ramsey planner is confronted with the same promises at time 0 that she herself would like to make in the long run.

Having introduced the timeless Ramsey problem, we now prove that it resolves the time-0 problem in Proposition 6, which is the main result of this subsection. Anticipating the sequence-space representation of our model (Section 5), we can interpret all endogenous variables as functions of the time paths of (i) policy, which we denote by $\boldsymbol{i} = \{i_t\}$, and (ii) exogenous shocks, which we denote by $\boldsymbol{Z} = \{A_t, \rho_t, \epsilon_t\}$, as well as (iii) initial conditions for the distribution $g_0(a,z)$ and penalties $(\phi(a,z), \theta)$. Given a sequence of shocks, an initial distribution, and initial promises, the planner then chooses among those competitive equilibria that are implementable by policy. In particular, we can evaluate the objective L^{TP} at any feasible policy path $\boldsymbol{i}$. A policy is then locally optimal when the derivative of L^{TP} with respect to any feasible perturbation of the policy path $d\boldsymbol{i}$ is 0.

Proposition 6. (**Timeless Ramsey Problem Resolves Time-0 Problem**) *Optimal policy under the timeless Lagrangian is time-consistent at the stationary Ramsey plan under the timeless penalty $\mathcal{T}(\phi_{ss}, \theta_{ss})$. That is,*

$$\frac{d}{d\boldsymbol{i}} L^{\text{TP}}\Big(g_{\text{ss}}, \phi_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}\Big) = 0. \tag{45}$$

Equation (45) says that, when we initialize the economy at the stationary Ramsey plan, i.e., $g_0(a,z) = g_{\text{ss}}(a,z)$, and set the time-0 penalty using the stationary multipliers, i.e., $\phi(a,z) = \phi_{\text{ss}}(a,z)$ and $\theta = \theta_{\text{ss}}$, then in the absence of shocks, i.e., $\boldsymbol{Z} = \boldsymbol{Z}_{\text{ss}}$, the stationary Ramsey policy is optimal, i.e., $\frac{d}{d\boldsymbol{i}} L^{\text{TP}}(\cdot) = 0$ when we set $\boldsymbol{i} = \boldsymbol{i}_{\text{ss}}$. Proposition 6 proves that a timeless Ramsey planner has no incentive to deviate from the stationary Ramsey plan in the absence of shocks. The timeless Ramsey problem resolves the time-0 problem and addresses inflationary bias in both the short run and the long run. The solid blue line in Figure 1 shows that the timeless Ramsey planner sets inflation to 0 in the absence of shocks.

Purpose of the timeless Ramsey approach. It should be evident that, from a time-0 perspective, the standard Ramsey plan attains higher welfare than the timeless Ramsey plan. The timeless Ramsey problem may thus be viewed as an inferior guide for policy design. However, there are at least three reasons why the timeless Ramsey approach is valuable. First, Woodford (1999)'s concerns about the time-0 problem remain valid: From a policymaker's perspective, access to new information and advances in modeling often necessitate a reevaluation of the framework used for policy design. If optimal policy is then recomputed each time under the standard Ramsey problem, it repeatedly suffers from the time-0 problem, which Woodford (1999) argues is an impractical guide for policy design. Second, the timeless Ramsey approach allows us to isolate the planner's pure stabilization motive in response to business cycle shocks and separate it from the time inconsistent incentive to deviate from the stationary Ramsey plan at time 0. We develop this argument in Sections 4.6 and 5. Third, we show in Section 5 that perturbation methods only yield valid approximations of optimal stabilization policy under the timeless Ramsey problem.

4.4 Properties of the Timeless Inflation and Distributional Penalties

We introduced the timeless penalty in Section 4.3 and showed that it resolves the time-0 problem, disincentivizing the planner from generating inflation in the short run. In this subsection, we explore this timeless penalty analytically. We establish two main results. First, we show that the inflation penalty, θ_{ss}, which is already present in RANK economies, depends on novel distributional considerations in HANK. Second, we show that the new distributional penalty that we introduce in this paper penalizes the welfare gains of indebted, high marginal utility households. The distributional penalty solves a novel promise-keeping Kolmogorov forward equation.

Timeless inflation penalty. Proposition 7 introduces an analytical characterization of the inflation penalty that resolves the time-0 problem in HANK. This expression combines the optimality conditions for consumption and hours worked with the Phillips curve.

Proposition 7. (**Timeless Inflation Penalty**) *The timeless penalty on inflation in both RANK and HANK economies satisfies*

$$\theta_{ss} = \frac{\Omega^1_{ss} + \Omega^2_{ss}}{-\frac{\epsilon}{\delta} \iint z \left(\frac{d\tau_{ss}(a,z)}{dc_{ss}(a,z)} + \frac{1}{A} \frac{d\tau_{ss}(a,z)}{dN_{ss}} \right) g_{ss}(a,z)\, da\, dz} \tag{46}$$

where Ω^1_{ss} *and* Ω^1_{ss} *are given by*

$$\Omega^1_{ss} = \overbrace{\left(1 - \frac{\epsilon - 1}{\epsilon}(1 + \tau^L)\right) \iint zu'(c_{ss}(a,z))g_{ss}(a,z)\, da\, dz}^{=0 \text{ with appropriate employment subsidy}}$$

$$\Omega^2_{ss} = \underbrace{\iint \left(zu'(c_{ss}(a,z)) - \frac{v'(n_{ss})}{A}\right)\phi_{ss}(a,z)\, da\, dz - \iint z\tilde{\chi}_{ss}(a,z)g_{ss}(a,z)\, da\, dz}_{=0 \text{ in RANK}}$$

Ω^1_{ss} *is* 0 *under the appropriate employment subsidy and* Ω^2_{ss} *is* 0 *in the RANK limit. Since* Ω^2_{ss} *is typically non-zero, distributional considerations shape the inflation penalty in HANK economies.*

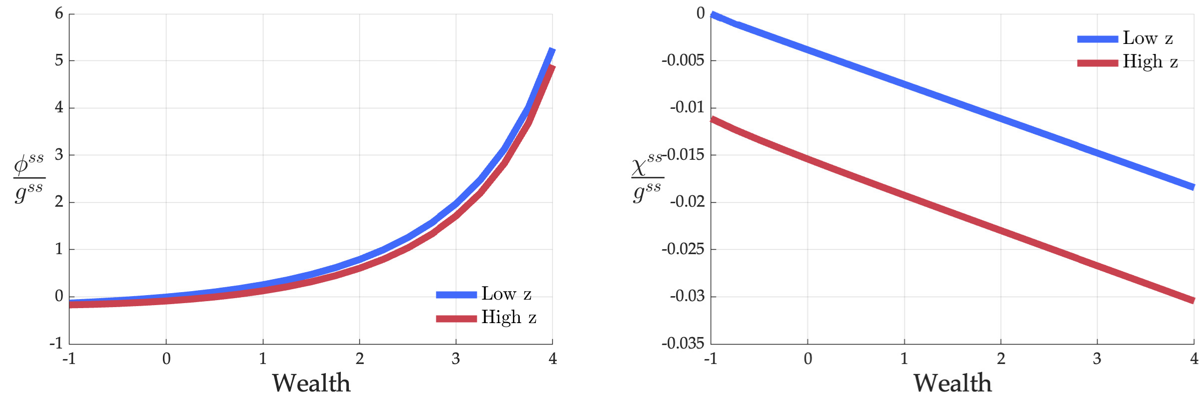
As we show in the Appendix, the denominator of equation (46) is always positive, so the sign of the inflation penalty depends on the signs of Ω^1_{ss} and Ω^2_{ss}. In the RANK limit, distributional considerations disappear and Ω^2_{ss} vanishes. In this case, the inflation penalty inherits the sign of $(1 - \frac{\epsilon-1}{\epsilon}(1 + \tau^L))$. In RANK, time inconsistency only emerges when employment is inefficiently low due to markups in a distorted steady state. In fact, with the appropriate employment subsidy, $\frac{\epsilon-1}{\epsilon}(1 + \tau^L) = 1$, equation (46) implies that no inflation penalty is required because no time consistency problem emerges in that case.

These same forces that shape the inflation penalty in RANK also appear in our HANK economy. However, the inflation penalty in HANK is also shaped by distributional considerations. Even with the correct employment subsidy to address the markup distortion in steady state, the time consistency problem on inflation does not disappear. Intuitively, θ_{ss} impacts the planner's desire to perturb aggregate economic activity by penalizing inflation at time 0. When households are heterogeneous, changes in aggregate economic activity have distributional consequences. In particular, the first term in Ω^2_{ss} captures the differential impact of an increase in hours worked on households' flow utility, while the second term accounts for the differential impact on households' consumption-savings decisions.

In other words, the two sources of time inconsistency—markups and redistribution—meaningfully interact now. A corollary of this result is that the choice of an appropriate inflation penalty takes on a distributional dimension whenever the planner has a utilitarian objective.

Timeless distributional penalty. In a RANK economy, the nominal interest rate is sufficient to fully correct the representative household's consumption-savings decision. Formally, we show that the Ramsey planner in RANK sets $\phi^{\text{RA}}_t = 0$ at all times—see equation (84) in the Appendix. This implies that no time consistency problem separately emerges from the Bellman equation.

With heterogeneous households, the policy instrument still corrects households' decisions, but



(a) Distributional penalties

(b) Consumption-savings multiplier

Figure 2. Timeless Distributional Penalty

Note. The left panel of Figure 2 shows the steady state values of the timeless distributional penalties, $\phi_{\text{ss}}(a, z)$, normalized by the mass of households, $g_{\text{ss}}(a, z)$. The right panel of Figure 2 shows the steady state values of the multiplier on households' consumption-savings optimality condition, $\chi_{\text{ss}}(a, z)$, also normalized by the mass of households, $g_{\text{ss}}(a, z)$. See Section 5 for the calibration.

only on average, implying $\iint \phi_t(a, z)\, da\, dz = 0$. Unlike in RANK economies, the nominal interest rate is no longer sufficient to correct the entire cross section of consumption-savings decisions. The Ramsey planner consequently finds that households privately consume too much or too little. Under commitment, the planner then finds it valuable to make promises about the future in order to influence consumption allocations today. These promises—encoded in the time-varying multiplier $\phi_t(a, z)$—open the door to time inconsistency. Our next result characterizes the timeless distributional penalty $\phi_{\text{ss}}(a, z)$ that addresses this time consistency problem.[37]

Proposition 8. (**Timeless Distributional Penalty**) *The timeless distributional penalty $\phi_{ss}(a, z)$ solves the promise-keeping Kolmogorov forward equation*

$$0 = -\mathcal{A}^*_{ss}\phi_{ss}(a, z) + \partial_a \chi_{ss}(a, z), \tag{47}$$

where $\mathcal{A}^*_{ss}$ is the Kolmogorov forward operator associated with the stationary cross-sectional distribution.

Proposition 8 shows that the multipliers on households' Bellman equations are themselves characterized by Kolmogorov forward equations. We refer to equation (30) and its stationary counterpart (47) as *promise-keeping Kolmogorov forward* equations because they characterizes the evolution of

[37] Chari et al. (2020) study optimal capital taxation with heterogeneous agents. They address the time-0 problem by augmented the Ramsey problem with a "wealth constraint" that prevents taxing away free rents. Our penalty performs a similar function but its form derives from the recursive multiplier approach of Marcet and Marimon (2019).

the planner's promises on individual lifetime values. These promises, which also have the interpretation of penalties, as explained above, are encoded in the multiplier $\phi_{\text{ss}}(a,z)$. The evolution of distributional penalties in the cross section must be consistent with the law of motion of households, which is why the Kolmogorov forward operator $\mathcal{A}^*_{\text{ss}}$ appears in equation (47). Intuitively, penalties are associated with households in a given state, and so if a household transitions from one individual state to another, the penalty moves with her.[38] If individuals' optimal consumption-savings decisions change as they transition between different wealth states, then $\partial_a \chi_t(a,z)$ can be interpreted as capturing "births" and "deaths" of relative promises in the cross section. The reason why $\partial_a \chi_t(a,z)$ enters the promise-keeping Kolmogorov equations is because changes in lifetime utility for a given state (a,z) impact the consumption of households at a and slightly above (or below), per the households' consumption-savings optimality condition. Solving for $\phi_t(a,z)$ and $\chi_t(a,z)$ jointly and characterizing how they are linked via this promise-keeping Kolmogorov forward equation is one of the contributions of this paper.[39]

Figure 2 illustrates the distributional penalty $\phi_{\text{ss}}(a,z)$. Panel (a) shows that $\phi_{\text{ss}}(a,z) < 0$ for indebted households, which is consistent with an interpretation of $\phi_{\text{ss}}(a,z)$ as a penalty on redistribution. Intuitively, in order to counteract the time-inconsistent incentive to redistribute towards high marginal utility households, the planner with commitment sets distributional penalties that penalize welfare assessments that benefit such households. While it may seem counterintuitive that a utilitarian planner penalizes high marginal utility, this is the way in which a planner with commitment fights the inflationary bias present under discretion. Panel (b) of Figure 2 displays the stationary consumption-savings multiplier, $\chi_{\text{ss}}(a,z)$, normalized by the mass of households, $g_{\text{ss}}(a,z)$. In this particular calibration, $\chi_{\text{ss}}(a,z) < 0$ for all households, which implies that the planner, if unconstrained, would like households to consume less. This is consistent with the fact that $\mu_{ss} < 0$ at the stationary Ramsey plan of our calibrated model.

4.5 Penalties vs. Targets

Sections 4.3 and 4.4 focus on resolving the time-0 problem that emerges in the standard Ramsey problem. To that end, we introduce the timeless penalty and study the timeless Ramsey problem (42). Implementing the resulting Ramsey plan still relies on an infinite sequence of promises, however, which may be unrealistic in practice. We now explore whether a central bank that sets policy under discretion can still implement the optimal commitment solution under an appropriate institutional arrangement or with the appropriate penalties or targets (Clarida et al., 1999).

[38] Note that the operator $\mathcal{A}^*_t$ is mass-preserving, i.e., $\iint \mathcal{A}^*_t \phi_t(a,z)\, da\, dz = 0$, which allows us to interpret $\phi_t(a,z)$ as a distribution (of penalties).

[39] We conjecture that promise-keeping Kolmogorov forward equations will appear in other models in which a continuum of Bellman equations act as constraints. We also conjecture that if households had additional margins of adjustment besides consumption-savings, new birth and death terms would augment the Kolmogorov equation.

Time-varying penalties. First, consider again problem (17), where a Ramsey planner sets policy with commitment over a finite horizon. In Section 3, we identify policy under discretion with the limit of this problem as the commitment horizon becomes vanishingly small, i.e., as $\psi \to \infty$. Leveraging the observation that a timeless penalty can resolve the time-0 problem of a standard Ramsey planner, we now confront each finite-horizon Ramsey planner with a penalty at the time of transition, just like we confronted the standard Ramsey planner with a timeless penalty at time 0. Intuitively, this sequence of timeless penalties ensures that each successive finite-horizon planner behaves as if she had committed to policy in the infinite past. The resulting sequence problem is given by

$$\tilde{\mathcal{W}}_0(g_0, \phi_0, \theta_0) = \min_{\boldsymbol{M}} \max_{\boldsymbol{X}} \mathbb{E}_0 \left[\underbrace{L(0, \tau_1, g_0) + \mathcal{T}_0(\phi_0, \theta_0)}_{\text{Timeless Lagrangian: } L^{\text{TP}}} + e^{-\int_0^{\tau_1} \rho_s ds} \tilde{\mathcal{W}}_{\tau_1}(g_{\tau_1}, \phi_{\tau_1}, \theta_{\tau_1}) \right], \tag{48}$$

where $\boldsymbol{M}$, $\boldsymbol{X}$, and $L(0, T, g_0)$ are defined as in (17). Crucially, the evolution of ϕ_t and θ_t is given by equations (30) and (34), and we initialize the timeless penalties at $\phi_0(a,z) = \phi_{\text{ss}}(a,z)$ and $\theta_0 = \theta_{\text{ss}}$. By modifying the flow payoff in equation (48), we confront each planner with the appropriate timeless penalty at the time of transition. And in the limit as $\psi \to \infty$, where planners transition instantaneously, the penalties are also "active" in every instant. In the discrete-time analysis of, e.g., Galí (2015), we would say that the Markov planner faces these penalties in every period.

The timeless Ramsey plan can be implemented under discretion as long as the planner (central bank) faces the appropriate time-varying penalty $\mathcal{T}_t(\phi_t, \theta_t)$, which includes an inflation penalty $-\theta_t \pi_t^w$ and a distributional penalty $\iint \phi_t(a,z) V_t(a,z) \, da \, dz$. Intuitively, the difference between the commitment and the discretion solutions are the multipliers associated with the forward-looking constraints. By modifying the Markov planner's flow utility to account for these terms in the form of time-varying penalties, it is possible to implement the commitment solution under discretion (Svensson, 1997; Marcet and Marimon, 2019; Clayton and Schaab, 2022).

Inflation targeting. While confronting the discretionary planner (central bank) with the time-varying penalty $\mathcal{T}_t(\phi_t, \theta_t)$ can implement the timeless Ramsey solution, central bank design in practice is commonly based on targeting frameworks.

Proposition 5 highlights that a strict zero-inflation target implements the Ramsey plan in the absence of shocks.[40] In other words, household heterogeneity in our environment does not alter the longstanding view that an inflation target can successfully resolve inflationary bias in steady state. What is surprising is that an implementation of such an inflation target based on penalties requires

[40] Formally, a Markov planner will implement the optimal stationary Ramsey plan in the absence of shocks when confronted either with an infinite penalty for non-zero inflation or with an additional implementability condition (constraint) that requires $\pi_t = 0$.

two distinct penalties in our setting, whereas only the inflation penalty is required in RANK economies. Both penalties and targets can be used to implement a particular solution because of their duality relation: Constraints (targets) on optimal policy problems can be transformed into costs (penalties) in the objective function, and vice versa.[41] However, since the planner has a single aggregate instrument, once a path of aggregate variables (in this case inflation) is used as a target, the choice of instrument is automatically determined.

In the presence of shocks, our paper demonstrates that flexible inflation targeting is in principle still the appropriate framework for policy design. This target would be anchored around zero inflation in our setting. And in response to a shock, the target would prescribe the path of inflation that is optimal under the timeless Ramsey plan. Hence, an important takeaway of our analysis is that optimal policy in our HANK economy can also be implemented by a flexible inflation target around zero inflation, where the flexibility to stabilize business cycle shocks is now governed by distributional considerations as in our Ramsey problem.

4.6 Optimal Stabilization Policy

Characterizing optimal stabilization policy under the standard Ramsey plan would conflate the pure stabilization motive of policy with the time-0 problem, i.e., the planner's time inconsistent incentive to deviate from the stationary Ramsey plan even in the absence of shocks. A key motivation for setting up the timeless Ramsey problem is that it isolates the pure stabilization motive by resolving the time-0 problem. In this final subsection, we study optimal stabilization policy under commitment, we focus on the timeless Ramsey plan.

In Proposition 9 we characterize a non-linear, exact targeting rule for optimal monetary policy with commitment in response to demand, productivity, and cost-push shocks.[42] By considering special cases of this targeting rule it is possible to recover i) the discretion targeting rule introduced in Proposition 2, which allows us to highlight the role of inflation and distributional penalties counteracting the forces that drive discretionary policy, and ii) the well understood targeting rule in RANK, which allows us to identify the implications of household heterogeneity for optimal stabilization.

Proposition 9. (**Targeting Rule for Stabilization Policy under Commitment**) *Optimal monetary*

[41] A planner could conceivably target instead a particular path of lifetime utilities. This approach, which we do not explore in our paper, connects our results to the work on recursive contracting, as in Sannikov (2008) and Williams (2011), among many others.

[42] This targeting, as well as the targeting rule under discretion in Proposition 2, can be interpreted as a double perturbation in which the planner makes all households work an additional hour and consume the output generated, while also increasing interest rates to neutralize the intertemporal impact of the perturbation.

stabilization policy is summarized by the targeting rule

$$0 = \underbrace{\iint \left(zu'(c_t(a,z)) - \frac{v'(N_t)}{A_t} \right) g_t(a,z)\, da\, dz}_{\text{Aggregate Labor Wedge}} - \underbrace{\Omega_t^D \iint a u'(c_t(a,z)) g_t(a,z)\, da\, dz}_{\text{Redistribution Motive}} \tag{49}$$

$$+ \underbrace{\iint (z - \Omega_t^D a) u'(c_t(a,z)) \phi_t(a,z)\, da\, dz}_{\text{Distributional Penalty}} + \underbrace{\theta_t \frac{\epsilon_t}{\delta} \iint \left(\frac{z}{A_t} \frac{d\tau_t(a,z)}{dN_t} + (z - \Omega_t^D a) \mathcal{M}_t(a,z) \frac{d\tau_t(a,z)}{dc_t(a,z)} \right) g_t(a,z)\, da\, dz}_{\text{Inflation Penalty}}$$

where Ω_t^D is defined in Proposition 2.

The first and second terms in equation (49) respectively correspond to the aggregate labor wedge and the distributive pecuniary effect of interest rate changes. The third and fourth terms correspond to the distributional and inflation penalties. We leverage this equation to present four results.

First, note that when $\phi_t(a,z) = 0$ and $\theta_t = 0$, equation (49) collapses to the targeting rule under discretion introduced in Proposition 2. Intuitively, the new terms that shape the targeting rule under commitment act as penalties for the planner, counteracting the forces that drive discretionary policy.

Second, the targeting rule (49) also allows us to revisit optimal monetary stabilization policy in RANK, which it nests. In RANK, $\phi_t(a,z) = a = 0$, so the targeting rule simply trades off aggregate stabilization—encoded in the aggregate labor wedge—with an inflation penalty. Suppose we allow for the appropriate steady state employment subsidy, so that $\frac{\epsilon-1}{\epsilon}(1+\tau^L) = 1$. If we consider demand and TFP shocks, $\pi_t = \theta_t = 0$ are always feasible, so the targeting rule (49) implies that the aggregate labor wedge is zero: In response to demand and TFP shocks, the Ramsey planner in RANK closes both the inflation and output gaps at all times. This is the Divince Coincidence benchmark (Blanchard and Galí, 2007). In RANK, there is no tradeoff between inflation and output in the absence of cost-push shocks. In the case of cost-push shocks, Divine Coincidence breaks down, even in RANK. From the Phillips curve, it follows that at zero inflation, the only way in which the aggregate inflation-relevant labor wedge is zero, is when the aggregate labor wedge is non-zero. The planner consequently cannot close the inflation and output gaps at the same time.

Third, in a HANK economy, the targeting rule for aggregate stabilization policy is shaped by distributional considerations. Hence, even though it is feasible for the planner to close the inflation and output gaps at the same time in the absence of cost-push shocks, she finds it optimal not to do so. Divine Coincidence consequently fails even with the appropriate employment subsidy and in the absence of cost-push shocks. Formally, the aggregate labor wedge that makes equation (49) hold in response to a shock need not be zero. While pinpointing the source of departure from Divine Coincidence in general is difficult, we present a quantitative decomposition in Section 5.2.

We show in Appendix B.10 that the targeting rule (49) admits an alternative representation that augments the discretionary output gap targeting rule (26) with two penalty wedges. Using this decomposition, we demonstrate in Section 5.2 that departures from Divine Coincidence in response to demand shocks are due to changes in the redistribution wedge.

Finally, optimal stabilization policy under the timeless Ramsey plan always features inflation overshooting. This result applies to both RANK and HANK economies and follows from the fact that the inflation penalty has initial and terminal conditions $\theta_0 = \lim_{T\to\infty} \theta_T = \theta_{ss}$. Its evolution in response to a shock is characterized by $\dot{\theta}_t = \delta \pi_t^w$. Hence, integrating and using the boundary conditions, it must be the case that

$$\int_0^\infty \pi_t^w dt = 0 \tag{50}$$

in response to any shock. That is, if inflation is positive on impact in response to a shock, $\pi_0^w > 0$, it must turn negative at some point in the future, and vice versa if $\pi_0^w < 0$.

5 Quantitative Analysis in Sequence Space

In this section, we extend the sequence-space approach (Boppart et al., 2018; Auclert et al., 2021) to Ramsey problems and welfare analysis. This allows us to compute transition dynamics under optimal policy—under discretion and with commitment—both non-linearly and using perturbation methods. We extend the fake-news algorithm of (Auclert et al., 2021) to compute optimal policy and show that the timeless Ramsey approach of Section 4 is critical for the validity of sequence-space perturbations.[43]

5.1 Sequence-Space Methods for Optimal Policy in HANK

We work with an abstract sequence-space representation of our model. Competitive equilibrium can be summarized by an *equilibrium map* that take as inputs the time paths of aggregates,

$$\mathcal{H}(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z}) = 0, \tag{51}$$

where $\boldsymbol{i} = \{i_t\}_{t\geq 0}$ denotes the path of policy, $\boldsymbol{Z} = \{A_t, \rho_t, \epsilon_t\}_{t\geq 0}$ the path of exogenous shocks, and $\boldsymbol{X}$ the path of macroeconomic aggregates. Given an initial cross-sectional distribution $g_0(a,z)$, which is implicitly encoded in $\mathcal{H}(\cdot)$, the equilibrium map (51) characterizes macroeconomic aggregates in terms of policy $\boldsymbol{i}$ and shocks $\boldsymbol{Z}$, i.e., $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z})$. The sequence-space representation (51) is as in

[43] Our perturbation approach is closest to that of Khan et al. (2003) and Schmitt-Grohé and Uribe (2004a) who also first characterize the optimality conditions that define a Ramsey plan non-linearly and then approximate these. It is well understood that, at least in the standard model, alternative valid perturbation methods also include the linear-quadratic approach (Benigno and Woodford, 2012) and evaluating welfare under a higher-order approximation of the equilibrium conditions (Schmitt-Grohé and Uribe, 2004b, 2007).

Auclert et al. (2021), except that $\mathcal{H}(\cdot)$ here also takes the path of policy $\boldsymbol{i}$ as an input, which is set optimally by the planner. Optimal policy, in turn, is determined as part of a Ramsey plan, whose sequence-space representation we characterize next.

Proposition 10. (**Sequence-Space Representation of Ramsey Plans**) *Given an initial distribution $g_0(a,z)$, initial penalties $\phi(a,z)$ and θ, as well as a path for exogenous shocks $\boldsymbol{Z}$, a timeless Ramsey plan comprises aggregate allocations and prices $\boldsymbol{X}$, optimal policy $\boldsymbol{i}$, and multipliers $\boldsymbol{M}$. Its sequence-space representation is given by*

$$\mathcal{R}(\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{i}, \boldsymbol{Z}) = 0, \tag{52}$$

where we leave implicit the dependence of the Ramsey map $\mathcal{R}(\cdot)$ on $g_0(a,z)$, $\phi(a,z)$, and θ.

We prove the sequence-space representations of equilibrium (51) and Ramsey plans (52) in Appendices D.1 and D.2.

Our sequence-space representation of Ramsey plans is valid for any initial distribution $g_0(a,z)$ and initial penalties $\phi(a,z)$ and θ. Equation (52) therefore recovers the standard Ramsey plan of Proposition 4 when we set $\phi(a,z) = 0$ for all (a,z) and $\theta = 0$. Similarly, it follows from Proposition 6 that evaluating the Ramsey plan (52) at $(g_{\text{ss}}, \phi_{\text{ss}}, \theta_{\text{ss}})$ resolves the time-0 problem. In that case, we refer to it as a timeless Ramsey plan. In the following, we always initialize the penalties at $\phi(a,z) = \phi_{\text{ss}}(a,z)$ and $\theta = \theta_{\text{ss}}$, and focus on characterizing the response of optimal policy, $d\boldsymbol{i}$, to exogenous shocks, $d\boldsymbol{Z}$, under the timeless Ramsey plan.

The Ramsey plan representation (52) consists of two sets of equations. The first block is the system of equations (51), which characterizes aggregate allocations and prices $\boldsymbol{X}$ given policy $\boldsymbol{i}$ and shocks $\boldsymbol{Z}$. The second block comprises the first-order optimality conditions of the Ramsey problem that solve for aggregate multipliers $\boldsymbol{M}$ and policy $\boldsymbol{i}$. Crucially, the Ramsey equations that characterize optimal policy are coupled with those that describe the evolution of multipliers. Unlike the equilibrium map $\mathcal{H}(\cdot)$, which suffices to solve for transition dynamics given policy, the Ramsey map $\mathcal{R}(\cdot)$ takes as inputs the aggregate multipliers $\boldsymbol{M}$ and features the equations that characterize them.

In this sequence-space representation, we refer to a (timeless) Ramsey plan as the time paths of aggregates, $\boldsymbol{R} = (\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{i})$.[44] The system of equations (52) characterizes a Ramsey plan as a

[44] In Section 4.1, we defined a Ramsey plan as the time paths of both aggregates and individual objects—namely, individual allocations, the cross-sectional distribution, and individual multipliers. In the sequence-space representation of our economy, we can express these individual objects as functions of the time paths of aggregates, as we formally show in Appendices D.1 and D.2. We thus loosely refer to a Ramsey plan in sequence-space form as the time paths of aggregates $(\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{i})$ with the understanding that the remaining individual objects can be expressed and easily obtained as functions of these.

function of the exogenous shocks, i.e.,

$$\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{Z}),$$

implicitly taking as given an initial distribution $g_0(a,z)$ as well as initial penalties $\phi(a,z)$ and θ. The sequence-space representation of Ramsey plans in Proposition 10 is not unique. One minimal representation of our baseline economy, which we use in our numerical implementation, is $\boldsymbol{X} = \{\Lambda_t, N_t\}_{t\geq 0}$, $\boldsymbol{M} = \{\mu_t, \theta_t\}_{t\geq 0}$, and $\boldsymbol{i} = \{i_t\}_{t\geq 0}$, where Λ_t is the aggregate labor wedge. In that case, the Ramsey plan representation (52) becomes a system of five equations: the definition of Λ_t as the aggregate labor wedge, the resource constraint (16), as well as the three aggregate optimality conditions (33), (34), and (35). Together, they solve for the Ramsey plan as a function of shocks, i.e., $\boldsymbol{X}(\boldsymbol{Z})$, $\boldsymbol{M}(\boldsymbol{Z})$, and $\boldsymbol{i}(\boldsymbol{Z})$, taking as given an initial cross-sectional distribution $g_0(a,z)$, as well as initial penalties $\phi(a,z)$ and θ.

5.1.1 Non-Linear Optimal Policy

The sequence-space representation of Ramsey plans in Proposition 10 is a system of non-linear equations. We can directly solve (52) non-linearly to compute optimal stabilization policy around the stationary Ramsey plan for any sequence of shocks $\boldsymbol{Z}$ that reverts back to $\boldsymbol{Z}_{\text{ss}}$. Computing the timeless Ramsey plan non-linearly is tractable and fast in our baseline HANK economy. Using an efficient quasi-Newton algorithm, we can solve (52) non-linearly in less than 10 seconds.[45]

However, computing non-linear transition paths in more complex HANK economies with richer cross-sectional heterogeneity can become cumbersome. Local perturbation methods, on the other hand, are fast and oftentimes very accurate in the context of canonical HANK environments.[46] In the remainder of this section, we develop sequence-space perturbation methods to approximate optimal policy in a neighborhood around the stationary Ramsey plan. In principle, we can take either the primal or the dual representation of our Ramsey problem as a starting point to approximate optimal policy. In Sections 5.1.2 and 5.1.3, we present both approaches and argue that they have distinct advantages and disadvantages in different contexts.

[45] We use the quasi-Newton algorithm developed by Schaab and Zhang (2022) and Schaab (2020) to compute non-linear transition paths in heterogenous-agent economies. The code is available at `https://github.com/schaab-lab/SparseEcon`. Using this solver, computing the non-linear Ramsey plan of our model takes less than 10 seconds on a personal computer for discretized time grids with 150 nodes, using a 2020 13-inch MacBook Pro with an M1 chip and 16 GB memory.

[46] When computing Ramsey plans non-linearly, we use quasi-Newton rather than standard Newton methods. This means that we compute the Jacobians involved in the algorithm once and subsequently use a recursive approximation. In practice, the algorithm never has to recompute the Jacobians and converges quickly, precisely because first-order perturbation solutions are typically very accurate approximations in canonical HANK economies. Therefore, the objects we need are precisely those we also compute below in Section 5.1.2, i.e., $\mathcal{R}_{\boldsymbol{R}}$ and $\mathcal{R}_{\boldsymbol{Z}}$ evaluated around the stationary Ramsey plan, using a fake-news algorithm. As long as the quasi-Newton algorithm does not require that we recompute the Jacobian matrix, computing the non-linear solution is just as fast as the fake-news algorithm for the perturbation approach, requiring the computation of only a single column of the Jacobians $\mathcal{R}_{\boldsymbol{R}}$ and $\mathcal{R}_{\boldsymbol{Z}}$.

5.1.2 Optimal Policy Perturbations in the Primal

To approximate optimal policy in the primal representation of the Ramsey problem, we take as our starting point the system of equations (52).

Proposition 11. (**Optimal Policy Perturbations in the Primal**) *Consider the primal Ramsey problem and the associated Ramsey plan, which is characterized by* (52) *and solves* $\mathcal{R}(\cdot) = 0$*. Suppose we initialize the Ramsey plan at the cross-sectional distribution* $g_0(a,z) = g_{ss}(a,z)$ *and with initial timeless penalties* $\phi(a,z) = \phi_{ss}(a,z)$ *and* $\theta = \theta_{ss}$*. To first order, optimal stabilization policy is then characterized as part of the timeless Ramsey plan by*

$$d\boldsymbol{R} = -\mathcal{R}_{\boldsymbol{R}}^{-1}\mathcal{R}_{\boldsymbol{Z}} d\boldsymbol{Z} \tag{53}$$

where $d\boldsymbol{Z} = \boldsymbol{Z} - \boldsymbol{Z}_{ss}$ *is the exogenous shock,* $d\boldsymbol{R} = (d\boldsymbol{X}, d\boldsymbol{M}, d\boldsymbol{i})$ *denotes the response of the Ramsey plan, and* $\mathcal{R}_{\boldsymbol{R}}$ *and* $\mathcal{R}_{\boldsymbol{Z}}$ *are Jacobians of the Ramsey plan map.*

We prove Proposition 11 in Appendix D.3.

It is critical to note that the validity of the sequence-space perturbation method in Proposition 11 relies on initializing the Ramsey problem with the timeless penalties, so that $\mathcal{R}(\cdot)$ characterizes a timeless Ramsey plan. With the timeless penalties, $d\boldsymbol{R}$ only captures the planner's stabilization motive in response to shocks $d\boldsymbol{Z}$. Without them, $d\boldsymbol{R}$ conflates the stabilization motive with the time-0 problem and is consequently no valid solution of optimal stabilization policy to first order. Our timeless Ramsey approach is therefore the critical foundation that allows us to leverage perturbation methods to compute optimal stabilization policy.

To approximate Ramsey plans to first order in the primal, we have to compute two first-order derivative matrices, $\mathcal{R}_{\boldsymbol{R}}$ and $\mathcal{R}_{\boldsymbol{Z}}$. These matrices can in turn be constructed from sequence-space Jacobians, which allows us to leverage the power of sequence-space perturbation methods. In Appendix D, we extend the fake-news algorithm developed by Auclert et al. (2021) for sequence-space Jacobians to compute optimal policy via the Ramsey map Jacobians $\mathcal{R}_{\boldsymbol{R}}$ and $\mathcal{R}_{\boldsymbol{Z}}$.[47]

5.1.3 Optimal Policy Perturbations in the Dual

Appendix B.6 formally introduces the dual form of our timeless Ramsey problem. While the previous subsection directly uses the primal representation of Ramsey plans, an alternative sequence-

[47] Auclert et al. (2021) show how to use the equilibrium map (51) to efficiently compute transition dynamics for a given path of policy to first order. They develop a general model representation of the standard micro block of competitive equilibria in heterogeneous-agent economies, i.e., the set of equations that characterize the allocations and behavior of individual agents. We show in Appendix D that computing optimal policy using the sequence-space Ramsey plan representation (52) requires a second "micro block," namely the set of individual multiplier equations. We develop a general sequence-space representation for this multiplier block and show that the same principles underlying Auclert et al. (2021)'s fake-news algorithm can be used to efficiently compute sequence-space Jacobians for multipliers.

space perturbation method can be developed by using the dual form as a starting point. We do so in Appendix D.4.

The key advantage of the dual approach is that multipliers do not explicitly have to be computed as part of the Ramsey plan solution. When the multiplier equations are particularly complex and computationally intensive, this can be an important advantage. The main disadvantage of the dual approach is that it relies on second-order derivatives, whereas the primal approach relies on first-order derivatives. In Section D, we therefore introduce *sequence-space Hessians* as the natural second-order generalization of sequence-space Jacobians. Finally, Appendix D.4 offers a detailed discussion on the advantages and disadvantages of the dual approach relative to the primal approach of Section 5.1.2.

5.2 Optimal Stabilization Policy: Quantitative Analysis

We now compute optimal monetary stabilization policy in response to demand shocks (this section), as well as TFP (Appendix F.1) and cost-push (Appendix F.2) shocks.

Calibration. Adopting isoelastic preferences, we set the discount rate to a quarterly $\rho = 0.02$, the elasticity of intertemporal substitution to $\gamma = 2$, and the inverse Frisch elasticity to $\eta = 2$. We set the elasticity of substitution between labor varieties to $\epsilon = 10$ and the nominal wage adjustment cost to $\delta = 100$, following standard practice in the wage rigidity literature (Auclert et al., 2020). Finally, we allow for an employment subsidy $(1 + \tau^L)\frac{\epsilon-1}{\epsilon} = 1$ that offsets the wage-markup distortion in stationary equilibrium.

In our HANK model, we model households' earnings risk as a two-state Markov chain with $z_t \in \{\underline{z}, \overline{z}\}$, where $\underline{z} = 0.8$ and $\overline{z} = 1.2$. We set the quarterly Poisson transition rate out of both states to 0.33. Our RANK benchmark can be seen as the limit as $\underline{z}, \overline{z} \to 1$, using as initial condition for the cross-sectional distribution a Dirac mass at $(a, z) = (0, 1)$.

Finally, we model the demand shock as a mean-reverting AR(1) process. In continuous time, this implies that $\dot{\rho}_t = \xi_\rho(\rho - \rho_t)$, where ρ denotes the steady-state level. We study a one-time, unanticipated ("MIT") shock at time $t = 0$, initializing the shock at $\rho_0 = 1.5\rho$ and calibrating its persistence to a half-life of one quarter.

Optimal monetary stabilization of demand shocks. Figure 3 plots the optimal transition dynamics under the timeless Ramsey plan in response to a demand shock.

Divine Coincidence obtains in RANK in the face of demand and productivity shocks: the planner perfectly stabilizes both the output and inflation gaps. This benchmark result requires the appropriate employment subsidy, which we assume here. To support this desired allocation, the

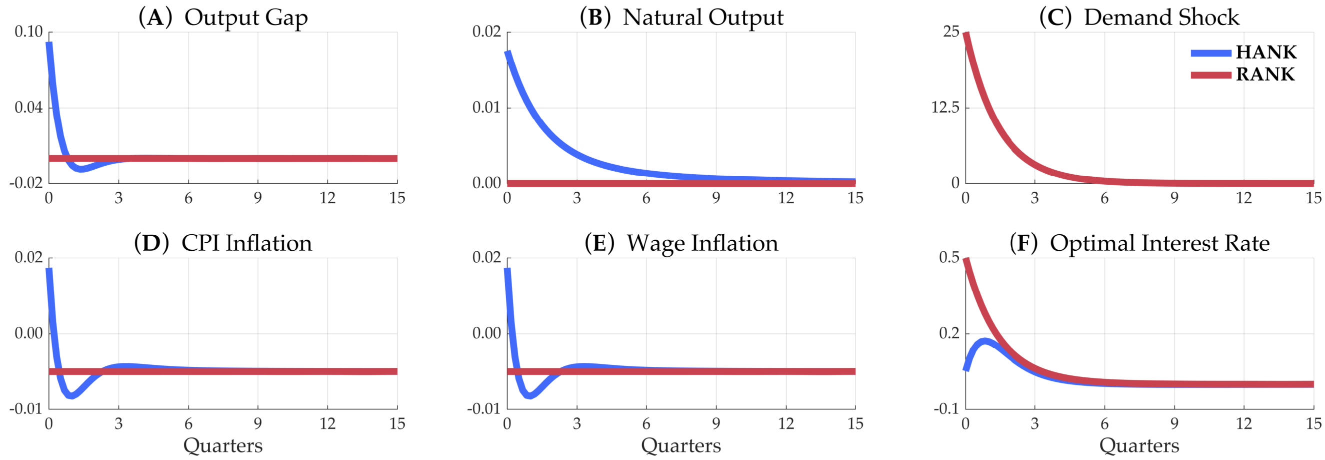


Figure 3. Optimal Policy Transition Dynamics: Demand Shock

Note. Transition dynamics after a positive discount rate shock in RANK (red) and HANK (blue) under optimal monetary stabilization policy. The discount rate shock is initialized at $\rho_0 = 0.025$ and mean-reverts to its steady state value $\rho = 0.02$, with a half-life of 1 quarter. Panels (**A**) through (**C**) report the dynamics of the output gap, $\frac{Y_t - \tilde{Y}_t}{\tilde{Y}_t}$, natural output, and the shock, all in percent deviations from the stationary Ramsey plan. Panels (**D**) through (**F**) report CPI inflation, wage inflation, and the optimal interest rate, all in percentage point deviations from the stationary Ramsey plan.

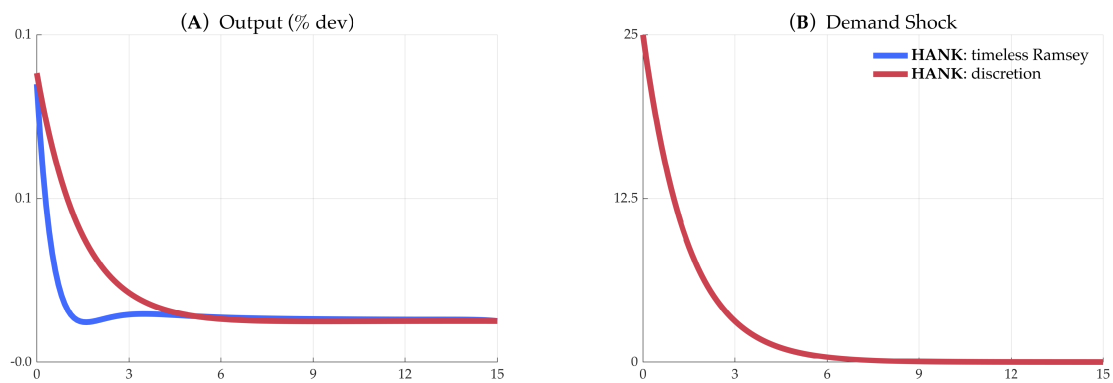


Figure 4. Optimal Policy under Discretion: Demand Shock

Note. Transition dynamics after a positive discount rate shock, comparing optimal policy in HANK with commitment (blue), i.e., under the timeless Ramsey problem, and under discretion (red). Discount rate shock is initialized at $\rho_0 = 0.025$ and mean-reverts to its steady state value $\rho = 0.02$, with a half-life of 1 quarter. Panel (**A**) plots output in percent deviation from the 0-inflation steady state for commitment (blue) and in percent deviation from the Markov perfect equilibrium with inflationary bias for discretion (red). Panel (**B**) plots the underlying shock in percent deviations.

planner raises the interest rate by about 50 basis poins to lean against the 50 basis point discount rate shock.

In HANK, the planner again leans against the demand shock, stabilizing output and inflation gaps, but not as strongly as in RANK. Especially the output gap is allowed to open up meaningfully. The on-impact output gap response under optimal policy is only dampened by 50% relative to the Taylor rule case. The inflation gap, on the other hand, is stabilized almost entirely. Unlike in RANK, the path of interest rates that supports this allocation features a hump, where the planner only gradually increases the nominal rate. The hump-shaped paths of the output and inflation gaps in Figure 3 are due to commitment—see our discussion in Section 4.6. In Figure 4, we plot the relative output paths under discretion (red) and under the timeless Ramsey problem (blue). Under discretion, the planner takes future policy as given and does not benefit from promising an over-shooting. With commitment, the planner finds it optimal to promise over-shooting to improve contemporaneous tradeoffs. This is reflected in the hump-shaped optimal interest rate path in Figure 3.

To pinpoint the source of departure from Divine Coincidence in HANK, we rely on decomposition (65). The markup wedge does not respond to a demand shock. Similarly, Figure 4 highlights that the two penalty wedges, which vanish under discretion, contribute little to the optimal on-impact response of the output gap, which is nearly identical when policy is set under discretion. Therefore, the departure from Divine Coincidence in response to a demand shock is quantitatively driven—at least on impact—by the redistribution wedge.

6 Conclusion and Broader Insights

This paper draws three main conclusions for the design of optimal monetary policy in the presence of heterogeneous households. First, a utilitarian planner under discretion trades off aggregate stabilization against a novel redistribution motive. This redistribution motive is a new source of time inconsistency that substantially exacerbates inflationary bias. In HANK, policy under discretion consequently leads to inflationary bias even with the appropriate employment subsidy to correct the markup distortion in steady state. Under commitment, the utilitarian planner recognizes that monetary policy is an inappropriate instrument to address this new source of perceived suboptimality by promising zero inflation in the long run. Second, two sets of penalties are necessary for monetary policy to be time consistent: the standard inflation penalty, which must be augmented by distributional considerations, and new distributional penalties, which penalize those individuals who benefit from discretionary policy. Abiding by these penalties allows the planner to achieve zero inflation in the long-run. The commitment solution can still be implemented by a planner under discretion, as long as she is confronted with the appropriate penalties, or through an appropriate inflation targeting framework. Third, Divine Coincidence breaks down, even in the absence of cost-push shocks, and optimal monetary stabilization policy will account for the distributional impact of policies, trading off aggregate stabilization against distributional considerations.

Three broader insights emerge from our study of optimal policy in HANK economies.

1. **HANK vs. RANK.** Household heterogeneity has stark implications for optimal monetary policy under discretion, where a new source of time inconsistency exacerbates inflationary bias. New penalties are required to make monetary policy time consistent. On the other hand, household heterogeneity in our model does not alter the optimality of 0 inflation in the long run. And while it has implications for optimal stabilization policy, departures from RANK are quantitatively small.

2. **Joint aggregate and distributional impact of policy.** Optimal policy is shaped by its joint aggregate and distributional impact. In the stylized model of this paper, lowering rates stimulates the economy and improves redistribution. However, this pattern may be different in richer environments. While our approach and the logic of our results will extend to these cases, the exact conclusions may not, which opens the door to future research. For instance, lowering interest rates may benefit wealthy, low marginal utility households, plausibly through labor market effects, credit market access, or differential inflation. Alternatively, bailouts or unconventional monetary policy may stimulate the economy but harm redistribution by favoring wealthy, low marginal utility households. Through the logic developed in this paper, a planner will have an incentive to run an underheated economy in both of these scenarios, inducing deflationary bias under discretion and requiring different inflation and distributional penalties to implement optimal policy under commitment.

3. **Role of mandates.** The commitment solution is one possible way of addressing the new source of inflationary bias identified in this paper. In RANK economies, Rogoff (1985) argues that increasing the weight on inflation in the central bank's loss function (mandate) may be valuable to reduce inflationary bias. In HANK economies, the analog to Rogoff's solution is to lower the weight on redistribution in the central bank's loss function (mandate). We study the design of central bank mandates when society values distributional considerations in ongoing and future work. In Dávila and Schaab (2022), we develop a methodology to decompose the normative considerations that determine the aggregate and redistribution consequences of welfare assessments in general economies. In Dávila and Schaab (2023), we leverage that methodology to explore the role of central bank mandates and how they impact optimal monetary under discretion and commitment.

References

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

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2022. Income and wealth distribution in macroeconomics: A continuous-time approach. *The review of economic studies*, 89(1):45–86.

Ampudia, Miguel, Dimitris Georgarakos, Jiri Slacalek, Oreste Tristani, Philip Vermeulen, and Giovanni Violante. 2018. Monetary policy and household inequality.

Auclert, Adrien. 2019. Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–2367.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model.

—. 2023. The intertemporal keynesian cross.

Barro, Robert J and David B Gordon. 1983. Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, 12(1):101–121.

Benigno, Pierpaolo, Gauti B Eggertsson, and Federica Romei. 2020. Dynamic debt deleveraging and optimal monetary policy. *American Economic Journal: Macroeconomics*, 12(2):310–350.

Benigno, Pierpaolo and Michael Woodford. 2012. Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory*, 147(1):1–42.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J Sargent. 2021. Inequality, Business Cycles, and Monetary-Fiscal Policy.

Bilbiie, Florin O. 2008. Limited asset markets participation, monetary policy and (inverted) aggregate demand logic. *Journal of economic theory*, 140(1):162–196.

Bilbiie, Florin Ovidiu. 2018. Monetary policy and heterogeneity: An analytical framework.

Blanchard, Olivier and Jordi Galí. 2007. Real wage rigidities and the New Keynesian model. *Journal of money, credit and banking*, 39:35–65.

Boppart, Timo, Per Krusell, and Kurt Mitman. 2018. Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, 89:68–92.

Chari, Varadarajan V and Patrick J Kehoe. 1999. Optimal fiscal and monetary policy. *Handbook of macroeconomics*, 1:1671–1745.

Chari, Varadarajan V, Juan Pablo Nicolini, and Pedro Teles. 2020. Optimal capital taxation revisited. *Journal of Monetary Economics*, 116:147–165.

Clarida, Richard, Jordi Galí, and Mark Gertler. 1999. The science of monetary policy: a new Keynesian perspective. *Journal of economic literature*, 37(4):1661–1707.

Clayton, Christopher and Andreas Schaab. 2022. A Theory of Dynamic Inflation Targets.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. Innocent Bystanders? Monetary policy and inequality. *Journal of Monetary Economics*, 88:70–89.

Cúrdia, Vasco and Michael Woodford. 2016. Credit frictions and optimal monetary policy. *Journal of Monetary Economics*, 84:30–65.

Dávila, Eduardo and Anton Korinek. 2018. Pecuniary Externalities in Economies with Financial Frictions. *The Review of Economic Studies*, 85(1):352–395.

Dávila, Eduardo and Andreas Schaab. 2022. Welfare Assessments with Heterogeneous Individuals.

—. 2023. Central Bank Mandates with Distributional Considerations.

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

Erceg, Christopher J, Dale W Henderson, and Andrew T Levin. 2000. Optimal monetary policy with staggered wage and price contracts. *Journal of monetary Economics*, 46(2):281–313.

Farhi, Emmanuel and Iván Werning. 2016. A theory of macroprudential policies in the presence of nominal rigidities. *Econometrica*, 84(5):1645–1704.

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

González, Beatriz, Galo Nuño, Dominik Thaler, and Silvia Albrizio. 2021. Optimal monetary policy with heterogeneous firms.

Harris, Christopher and David Laibson. 2013. Instantaneous gratification. *The Quarterly Journal of Economics*, 128(1):205–248.

Kamien, Morton I and Nancy Lou Schwartz. 2012. *Dynamic optimization: the calculus of variations and optimal control in economics and management*. courier corporation.

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

Khan, Aubhik, Robert G King, and Alexander L Wolman. 2003. Optimal monetary policy. *The Review of Economic Studies*, 70(4):825–860.

Kydland, Finn E and Edward C Prescott. 1977. Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, 85(3):473–491.

—. 1980. Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and control*, 2:79–91.

Le Grand, François, Alaïs Martin-Baillon, and Xavier Ragot. 2021. Should monetary policy care about redistribution? Optimal fiscal and monetary policy with heterogeneous agents.

Ljungqvist, Lars and Thomas J Sargent. 2018. *Recursive macroeconomic theory*. MIT press.

Marcet, Albert and Ramon Marimon. 2019. Recursive contracts. *Econometrica*, 87(5):1589–1631.

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

McKay, Alisdair and Christian K Wolf. 2022. Optimal Policy Rules in HANK.

Nuño, Galo and Benjamin Moll. 2018. Social optima in economies with heterogeneous agents. *Review of Economic Dynamics*, 28:150–180.

Nuño, Galo and Carlos Thomas. 2020. Optimal monetary policy with heterogeneous agents.

Rogoff, Kenneth. 1985. The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics*, 100(4):1169–1189.

Rotemberg, Julio J. 1982. Sticky prices in the United States. *Journal of Political Economy*, 90(6):1187–1211.

Sannikov, Yuliy. 2008. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984.

Schaab, Andreas. 2020. Micro and Macro Uncertainty.

Schaab, Andreas and Allen Tianlun Zhang. 2022. Dynamic Programming in Continuous Time with Adaptive Sparse Grids.

Schaumburg, Ernst and Andrea Tambalotti. 2007. An investigation of the gains from commitment in monetary policy. *Journal of monetary economics*, 54(2):302–324.

Schmitt-Grohé, Stephanie and Martín Uribe. 2004a. Optimal fiscal and monetary policy under sticky prices. *Journal of economic Theory*, 114(2):198–230.

—. 2004b. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of economic dynamics and control*, 28(4):755–775.

—. 2005. Optimal fiscal and monetary policy in a medium-scale macroeconomic model. *NBER Macroeconomics Annual*, 20:383–425.

—. 2007. Optimal simple and implementable monetary and fiscal rules. *Journal of monetary Economics*, 54(6):1702–1725.

—. 2010. The optimal rate of inflation. In *Handbook of monetary economics*, volume 3, pp. 653–722.

Smirnov, Danila. 2022. Optimal Monetary Policy in HANK.

Svensson, Lars E O. 1997. Optimal Inflation Targets, "Conservative" Central Banks, and Linear Inflation Contracts. *The American Economic Review*, 87(1):98–114.

Werning, Iván. 2011. Managing a liquidity trap: Monetary and fiscal policy.

Williams, Noah. 2011. Persistent private information. *Econometrica*, 79(4):1233–1275.

Woodford, Michael. 1999. Commentary: How should monetary policy be conducted in an era of price stability? *New challenges for monetary policy*, 277316.

—. 2003. *Interest and prices*. Princeton University Press.

—. 2010. Optimal monetary stabilization policy. *Handbook of monetary economics*, 3:723–828.

Online Appendix (not for publication)

A Additional Model Details and Derivations

A.1 Labor Unions and Wage Rigidity

Each household supplies labor to all of $k \in [0,1]$ unions. We denote a household's total hours of work by $n_t = \int_0^1 n_{k,t} dk$. Each union pays the household a nominal wage $W_{k,t}$. The household budget constraint therefore corresponds to

$$\dot{a}_t = r_t a_t + z_t \frac{1}{P_t} \int_0^1 W_{k,t} n_{k,t} dk + \tau_t(z_t) - c_t. \tag{54}$$

Each union $k \in [0,1]$ transforms hours supplied by households into a differentiated labor service according to the linear aggregation technology

$$N_{k,t} = \iint z n_{k,t} g_t(a,z)\, da\, dz,$$

where $N_{k,t}$ is expressed in units of effective labor. Each union also rations labor, so that all households work the same hours. In particular, this implies $N_{k,t} = n_{k,t} \iint z g_t(a,z)\, da\, dz = n_{k,t}$, after normalizing cross-sectional average labor productivity to 1.

Labor packer. Unions sell their differentiated labor services to an aggregate labor packer. The packer operates the CES aggregation technology

$$N_t = \left(\int_0^1 N_{k,t}^{\frac{\epsilon_t - 1}{\epsilon_t}} dk \right)^{\frac{\epsilon_t}{\epsilon_t - 1}},$$

where the elasticity of substitution ϵ_t is potentially time-varying. We interpret time variation in the desired wage mark-up of unions as a source of cost-push shocks, following standard practice (see, e.g., Galí, 2015). The packer sells the aggregate labor bundle to firms at nominal wage rate W_t. The labor packer's cost-minimization problem is standard and yields the demand function and wage index

$$N_{k,t} = \left(\frac{W_{k,t}}{W_t}\right)^{-\epsilon_t} N_t \tag{55}$$

$$W_t = \left(\int_0^1 W_{k,t}^{1-\epsilon_t} dk \right)^{\frac{1}{1-\epsilon_t}} \tag{56}$$

where $W_{k,t}$ is the nominal wage rate charged by union k.

Wage rigidity. Nominal wages are sticky in our model. Each union k faces an adjustment cost to change its wage. Formally, the union takes $W_{k,t}$ as a state variable and controls how the wage evolves by setting wage inflation $\pi^w_{k,t}$, with

$$\pi^w_{k,t} = \frac{\dot{W}_{k,t}}{W_{k,t}}. \tag{57}$$

The union's adjustment cost is directly passed to union members as a quadratic utility cost, so households' instantaneous flow utility is formally given by

$$U_t\left(c_t, \left\{n_{k,t}, \pi^w_{k,t}\right\}_{k\in[0,1]}\right) = u(c_t) - v\left(\int_0^1 n_{k,t}\, dk\right) + \frac{\delta}{2}\int_0^1 (\pi^w_{k,t})^2 dk,$$

where $v(\cdot)$ captures pure disutility from working and δ modulates the strength of the wage rigidity.[48] The representation in the main text, i.e., equation (4), is valid in any equilibrium that features symmetric unions, which we assume.

We now formalize the union's wage setting problem to derive a New Keynesian wage Phillips curve. We assume that the union chooses wages in order to maximize stakeholder value—the sum of stakeholders', i.e., union members', utilities. That is, union k solves

$$\max_{\pi^w_{k,t}} \int_0^\infty e^{-\int_0^t \rho_s ds}\left(\iint \left[u(c_t(a,z;W_{k,t})) - v\left(\int_0^1 n_{k,t}\, dk\right) - \frac{\delta}{2}\int_0^1 (\pi^w_{k,t})^2\, dk\right] g_t(a,z)\, da\, dz\right) dt, \tag{58}$$

subject to equations (55) and (57). The union further internalizes the effect of its wage policy on its members' consumption—hence the explicit dependence of c_t on $W_{k,t}$ in equation (58). However, since union k is small, it takes as given all macroeconomic aggregates, including the cross-sectional household distribution.

Solving the union problem. To solve the union's problem we associate it with the Lagrangian

$$L = \int_0^\infty e^{-\rho t}\int\left[u\left(c_t\,(a,z;W_{k,t})\right) - v\left(\int_0^1 \left(\frac{W_{k,t}}{W_t}\right)^{-\epsilon} N_t dk\right) - \frac{\delta}{2}\int_0^1 \left(\pi^w_{k,t}\right)^2 dk\right] g_t\,(a,z)\, d\,(a,z)\, dt$$
$$+ \int_0^\infty e^{-\rho t}\left[\mu_t \pi^w_{k,t} W_{k,t} - \rho\mu_t W_{k,t} + W_{k,t}\dot{\mu}_t\right] dt + \mu_0 W_{k,0},$$

[48] There are three natural ways to model wage adjustment costs: as an explicit resource cost that is passed on to households, as labor productivity distortions, or as a direct utility cost. In the main text, we adopt the utility cost specification largely because it is most tractable.

where in the second line we already integrated by parts. Thus, the two first-order conditions are given by

$$
\begin{aligned}
0 &= \int u'(c_t)\frac{\partial c_t(a,z;W_{k,t})}{\partial W_{k,t}} g_t\left(a,z\right) d\left(a,z\right) + \epsilon v'(N_t)\frac{N_t}{W_t} + \mu_t \pi^w_{k,t} - \rho\mu_t + \dot{\mu}_t \\
0 &= -\delta\pi^w_{k,t} + \mu_t W_{k,t},
\end{aligned}
$$

as well as the initial condition $\mu_0 = 0$. By the envelope theorem, we have

$$
\frac{\partial c_t(a,z;W_{k,t})}{\partial W_{k,t}} = \frac{1}{P_t}(1+\tau^L)(1-\epsilon)z_t N_t.
$$

Defining

$$
\Lambda_t = \int z u'(c_t\left(a,z\right)) g_t\left(a,z\right) d\left(a,z\right),
$$

the first FOC becomes

$$
0 = (1+\tau^L)(1-\epsilon)w_t N_t \Lambda_t + \epsilon v'(N_t)N_t + \mu_t \dot{W}_t - \rho W_t \mu_t + W_t \dot{\mu}_t.
$$

Differentiating the second FOC yields

$$
\mu_t \dot{W}_t + W_t \dot{\mu}_t = \delta \dot{\pi}^w_t.
$$

Plugging back into the first FOC, we arrive at

$$
0 = (1+\tau^L)(1-\epsilon)w_t N_t \Lambda_t + \epsilon v'(N_t)N_t - \rho\delta\pi^w_t + \delta\dot{\pi}^w_t,
$$

which yields the result after rearranging. In particular, when equilibrium is initialized at a symmetric nominal wage distribution $\{W_{k,0}\}$, then the wage policies that result from the union's problem maintain symmetry of equilibrium. That is, wages and labor allocations remain equalized across unions, with $W_{k,t} = W_t$ and $N_{k,t} = N_t$. In such a symmetric equilibrium, the non-linear New Keynesian wage Phillips curve is then as in the main text.

A.2 Fiscal Rebates

Given union wage receipts $z_t W_{k,t} n_{k,t}$ to a household with labor productivity z_t, the government pays the household a proportional income subsidy $\tau^L z_t W_{k,t} n_{k,t}$, which the union internalizes when setting wages. Running a balanced budget, it pays for these outlays with a lump-sum tax based on aggregate employment. We assume that both the subsidy and the tax are proportional to a household's labor productivity. That is, the net fiscal rebate that a household with idiosyncratic labor productivity z receives is zero, with

$$P_t \tau_t(z) = \int_0^1 \tau^L z W_{k,t} n_{k,t} dk - \tau^L z W_t N_t = 0.$$

A.3 Natural Output and the Flexible Wage Limit

We define natural output as the output that obtains in the limit of flexible wages, i.e., as $\delta \to 0$. With isoelastic (CRRA) preferences $u(c) = \frac{1}{1-\gamma} c^{1-\gamma}$ and $v(n) = \frac{1}{1+\eta} n^{1+\eta}$, natural output in HANK is given by

$$\tilde{Y}_t = \left(\frac{\epsilon_t - 1}{\epsilon_t} (1 + \tau^L) A_t^{1+\eta} \iint \frac{z u'(c_t(a,z))}{u'(C_t)} g_t(a,z) \, da \, dz \right)^{\frac{1}{\gamma+\eta}}, \tag{59}$$

where the integral term reflects labor rationing. In the RANK limit, where this integral term vanishes, natural output is simply given by $\tilde{Y}_t^{\text{RA}} = \left(\frac{\epsilon_t - 1}{\epsilon_t} (1 + \tau^L) A_t^{1+\eta} \right)^{\frac{1}{\gamma+\eta}}$.

As $\delta \to 0$, equilibrium requires that $0 = \frac{\epsilon_t - 1}{\epsilon_t} (1 + \tau^L) w_t \Lambda_t - v'(N_t)$. That is, the augmented labor wedge is 0 in the flexible wage allocation, and we obtain natural output from this equation.

A.4 Competitive Equilibrium and Implementability

To conclude our discussion of the model details, we now state formally the implementability conditions for the Ramsey problem in continuous time. As part of our discussion, we also provide additional details on the generator $\mathcal{A}_t$ and its adjoint $\mathcal{A}_t^*$, which we use in the main text. Finally, in Section A.5, we develop a discretized representation of these implementability conditions, which we leverage in our proofs below.

A competitive equilibrium of our baseline HANK model can be characterized by three blocks of equations. First, there is an individual block, explained in the text, which corresponds to the households' HJB, their optimality condition for consumption, and the Kolmogorov forward equation:

$$\begin{aligned} \rho V_t(a,z) &= u(c_t(a,z)) - v(N_t) - \frac{\delta}{2} (\pi_t^w)^2 + \partial_t V_t(a,z) + \mathcal{A}_t V_t(a,z) \\ u'(c_t(a,z)) &= \partial_a V_t(a,z) \\ \partial_t g_t(a,z) &= \mathcal{A}_t^* g_t(a,z), \end{aligned}$$

where $\mathcal{A}_t$ is the infinitesimal generator of the process (a_t, z_t). Intuitively, it captures an agent's perceived law of motion of the process $d(a_t, z_t)$. It is analogous to a transition matrix in discrete time, and it is defined by

$$\mathcal{A}_t f_t(a,z) = \Big(r_t a + z w_t N_t - c_t(a,z) \Big) \partial_a f_t(a,z) + \mathcal{A}_z f_t(a,z), \tag{60}$$

for any function $f_t(a,z) : \mathbb{R}^3 \to \mathbb{R}$, where $\mathcal{A}_z$ is an additively separable component that captures perceived transition dynamics of earnings risk. We leave the structure of $\mathcal{A}_z$ fully general in our derivations, except that we assume it to be independent from policy. Our baseline results currently do not apply to the case of counter-cyclical earnings risk that responds to monetary policy, for example, but extending our approach to this more general case is straightforward.

We denote the adjoint of the infinitesimal generator by $\mathcal{A}_t^*$. The adjoint is defined by

$$\mathcal{A}_t^* f_t(a,z) = -\partial_a \Big[\Big(r_t a + z w_t N_t - c_t(a,z) \Big) f_t(a,z) \Big] + \mathcal{A}_z^* f_t(a,z), \tag{61}$$

where $\mathcal{A}_z^*$ is the adjoint of $\mathcal{A}_z$.

Second, there is an aggregate block, which includes the New Keynesian wage Phillips curve, the production technology, the wage equation, the Fisher equation, and and equation that relates price and wage inflation:

$$\begin{aligned}
\dot{\pi}_t^w &= \rho_t \pi_t^w + \frac{\epsilon_t}{\delta} \iint n_t \left(\frac{\epsilon_t - 1}{\epsilon_t} (1+\tau^L) w_t z u'(c_t) - v'(n_t) \right) g_t(a,z)\, da\, dz \\
Y_t &= A_t N_t \\
w_t &= A_t \\
r_t &= i_t - \pi_t \\
\pi_t &= \pi_t^w - \frac{\dot{A}_t}{A_t}.
\end{aligned}$$

Finally, we have the market clearing conditions in the goods and bond markets, given by

$$\begin{aligned}
Y_t = C_t &= \iint c_t\,(a,z)\, g_t\,(a,z)\, da\, dz \\
0 = B_t &= \iint a g_t\,(a,z)\, da\, dz.
\end{aligned}$$

The following Lemma defines the set of implementability conditions that act as constraints for a Ramsey planner.

Lemma 12. (**Implementability conditions**) *The set of equations that define an equilibrium can be* *expressed as implementability conditions for a standard primal Ramsey problem as follows:*

$$\begin{aligned}
\rho V_t(a,z) &= u(c_t(a,z)) - v(N_t) - \frac{\delta}{2}(\pi_t^w)^2 + \partial_t V_t(a,z) + \mathcal{A}_t V_t(a,z) \\
u'(c_t(a,z)) &= \partial_a V_t(a,z) \\
\partial_t g_t(a,z) &= \mathcal{A}_t^* g_t(a,z) \\
0 &= A_t N_t - \iint c_t(a,z) g_t(a,z)\, da\, dz \\
\dot{\pi}_t^w &= \rho \pi_t^w + \frac{\epsilon}{\delta}\left[\frac{\epsilon-1}{\epsilon}(1+\tau^L) A_t \iint z u'(c_t(a,z)) g_t(a,z) da dz - v'(N_t)\right] N_t.
\end{aligned}$$

We conclude this subsection by characterizing the operator $\mathcal{M}_t(a,z)$, which is an important input in the targeting rules we present in the main text (and in the proofs below). In particular, the operator admits the representation

$$\mathcal{M}_t(a,z) = (\rho - r_t + \partial_a c_t(a,z) - \partial_t - \mathcal{A}_t)^{-1}\, \partial_a c_t(a,z),$$

where the term $\rho - r_t + \partial_a c_t(a,z) = \rho - \partial_a s_t(a,z)$ captures time discounting net of the interest rate on the assets not consumed.

The terms ∂_t and $\mathcal{A}_t$ account for changes in aggregate conditions over time, ∂_t, and for the expected transition of the household across states, $\mathcal{A}_t$. Finally, $\partial_a c_t(a,z)$ is simply the instantaneous marginal propensity to consume (MPC).

The difference between the private and the social marginal of wealth, $\mathcal{M}_t(a,z)\mu_t$, can be interpreted as the present discounted value of the contribution of future consumption to aggregate excess demand induced by an increase in the household's wealth at time t. Intuitively, a marginal increase in wealth translates into higher aggregate demand at time t and in the future, depending on the household's propensities to consume and save out of wealth. Such spending is socially beneficial when $\mu_t > 0$ or costly when $\mu_t < 0$—an effect that only the planner internalizes. Note that a planner under discretion accounts for the social impact of future consumption via the path of future multipliers μ_t, despite taking future policy and expectations as given.

A.5 Discretized Competitive Equilibrium Conditions

We now develop a discretized representation of competitive equilibrium and the associated implementability conditions for the Ramsey problem. This discretized representation will elucidate how boundary conditions are treated formally by the Ramsey planner. In particular, we leverage this representation to explicitly account for households' borrowing constraint when deriving our proofs below.

For any function $c_t(a,z)$, we discretize both in the individual state space (a,z) and in time t. We denote this discretization by $\boldsymbol{c}_n$ for $n = 0, \ldots, N$. In particular, $\boldsymbol{c}_n$ is a $J \times 1$ vector, so that $c_{i,n} = c_{t_n}(a_i, z_i)$ associated with grid point i and date t_n. We also use notation $\boldsymbol{c}_{n,[2:J]}$, for example, to denote the $(J-1) \times 1$ vector consisting of elements 2 through J in $\boldsymbol{c}_n$.

We follow Achdou et al. (2022) and work with a consistent finite-difference discretization of our continuous-time heterogeneous-agent equations, which of course converge in the limit to our baseline HANK economy. We follow this approach in the remainder of this appendix. The following Lemma summarizes the discretized competitive equilibrium conditions of our model, using a finite-difference discretization given a policy path $\boldsymbol{i} = \{i_n\}_{n \geq 0}$. The proof follows along the lines of Achdou et al. (2022) and Schaab and Zhang (2022), and we refer the interested reader to those papers. This characterization will justify setting up the Ramsey problem using the following discretized equations as implementability conditions.

Lemma 13. *A consistent finite-difference discretization of the implementability conditions of our baseline HANK model is as follows. For the Hamilton-Jacobi-Bellman equation, we have*

$$\rho \boldsymbol{V}_n = \frac{\boldsymbol{V}_{n+1} - \boldsymbol{V}_n}{dt} + u \begin{pmatrix} i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} - v(N_n) - \frac{\delta}{2} (\pi^w_n)^2$$
$$+ \begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi^w_n \boldsymbol{a}_{[2:J]} + \frac{A_{n+1} - A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_n + \boldsymbol{A}^z \boldsymbol{V}_n$$

For the consumption first-order condition of the household, we simply have

$$u'(\boldsymbol{c}_{n,[2:J]}) = \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{n+1} \right)_{[2:J]}$$

For the Kolmogorov forward equation, we have

$$\frac{\boldsymbol{g}_{n+1} - \boldsymbol{g}_n}{dt} = (\boldsymbol{A}^z)' \boldsymbol{g}_n + \frac{\boldsymbol{D}_a'}{da} \left[\begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi^w_n \boldsymbol{a}_{[2:J]} + \frac{A_{n+1} - A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \boldsymbol{g}_n \right]$$

Finally, for the resource constraint we simply have

$$A_n N_n = \boldsymbol{c}_n' \boldsymbol{g}_n d\boldsymbol{x}$$

and for the Phillips curve

$$\frac{\pi^w_{n+1} - \pi^w_n}{dt} = \rho \pi^w_n + \frac{\epsilon}{\delta} \left[\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right] N_n$$

and we have already used $c_{n,1} = i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n$.

In this Lemma, we denote by $\boldsymbol{D}_a$ the finite-difference matrix that discretizes the partial derivative operator ∂_a. We also denote by $\boldsymbol{A}^z$ the (finite-difference) matrix that discretizes the operator $\mathcal{A}^z$ associated with the earnings process. Finally, $d\boldsymbol{x}$ denotes the integration measure of households. See Schaab and Zhang (2022) for details.

Crucially, the discretized system of equations in the above Lemma properly accounts for the household borrowing constraint, leveraging results from Achdou et al. (2022). In particular, they prove that in the simple Huggett economy with two earnings states the only point in the state space where the borrowing constraint binds is $(\underline{a}, z^L)$. We use this result here to plug in the borrowing constraint directly at that discretized point. While we have not formally proven that their representation extends to our HANK economy, we verify its validity numerically ex-post. And since the stationary equilibrium of our model is almost identical to theirs, there is little reason to expect any sharp discrepancies.

B Appendix for Section 4

We invert our presentation of proofs and formal derivations for Sections 3 and 4. In this Appendix, we start by setting up and characterizing the standard Ramsey problem, which is an instructive building block for the proofs that follow.

In particular, we state the continuous-time Ramsey problem in Section B.1 and present an illustrative but heuristic derivation of its first-order conditions. To formally account for boundary conditions, we then introduce and characterize the discretized standard Ramsey problem in Section B.2, leveraging the discretized representation of equilibrium from Appendix A.5.[49]

B.1 Standard Ramsey Problem in Continuous Time

In this section, we restate for convenience the standard Ramsey problem in continuous time and develop a heuristic derivation of its optimality conditions. We defer a formal treatment of boundary conditions to Appendix B.2. To adopt more compact notation, we drop time subscripts and make implicit the dependence of individual variables on states, so that $c_t(a,z)$ simply becomes c. Furthermore, we now reserve subscripts to denote partial derivatives, so that $\partial_t c_t(a,z)$ becomes c_t.

The functional Lagrangian associated with the standard primal Ramsey problem is given by

$$
\begin{aligned}
L^{\mathrm{SP}}(g_0) = \int_0^\infty e^{-\rho t} \Bigg\{ \iint \Bigg\{ & \Big[u(c) - v(N) - \frac{\delta}{2}(\pi^w)^2 \Big] g \\
& + \phi \Big[-\rho V + V_t + u(c) - v(N) - \frac{\delta}{2}(\pi^w)^2 + \mathcal{A}V \Big] \\
& + \chi \Big[u'(c) - V_a \Big] \\
& + \lambda \Big[-g_t + \mathcal{A}^* g \Big] \Bigg\} dadz \\
& - \mu \Big[\int\int cgdadz - AN \Big] \\
& + \theta \Big[-\dot{\pi}^w + \rho\pi^w + \frac{\epsilon}{\delta}\Big(\frac{\epsilon-1}{\epsilon}(1+\tau^L)A\Lambda - v'(N) \Big) N \Big] \Bigg\} dt
\end{aligned}
$$

[49] In recent work, González et al. (2021) follow a similar approach, first casting the optimal policy problem in continuous time, and then discretizing the resulting Ramsey plan conditions. The main difference between our paper and theirs is that they directly take their discretized system of equations to Dynare to obtain a numerical characterization of the Ramsey plan. We leverage the discretized equations to prove the main results of our paper. Our primary interest in discretizing the Ramsey plan conditions is to properly take into account the borrowing constraint faced by households, as well as the distribution mass point that emerges at the borrowing constraint.

Heuristic derivation. We provide an illustrative but heuristic derivation of the optimality conditions, abstracting from formally taking into account boundary conditions. This derivation is valuable because it is relatively brief and accessible. The proofs that follow become more complex only insofar as they formally take into account boundary conditions.

The following auxilliary results will be helpful. Integrating various partial derivatives in the above Lagrangian by parts, we have

$$\int_0^\infty \iint \Big[e^{-\rho t}\phi V_t\Big] da\, dz\, dt = \int \Big[-\phi(0,a,z)V(0,a,z) + \rho\int_0^\infty e^{-\rho t}\phi V dt - \int_0^\infty e^{-\rho t}\phi_t V dt\Big] da\, dz$$

$$\int_0^\infty \iint \Big[e^{-\rho t}\lambda g_t\Big] da\, dz\, dt = \int \Big[-\lambda(0,a,z)g(0,a,z) + \rho\int_0^\infty e^{-\rho t}\lambda g dt - \int_0^\infty e^{-\rho t}\lambda_t g dt\Big] da\, dz$$

$$\int_0^\infty \Big[e^{-\rho t}\theta\pi_t\Big] dt = -\theta(0)\pi(0) + \rho\int_0^\infty e^{-\rho t}\theta\pi dt - \int_0^\infty e^{-\rho t}\theta_t\pi dt.$$

Next, for the adjoint, we have

$$-\int_0^\infty e^{-\rho t}\iint \lambda\mathcal{A}^* g da\, dz\, dt = -\int_0^\infty e^{-\rho t}\iint (\mathcal{A}\lambda) g d(a,z) dt,$$

where we drop boundary terms, which we consider formally in the following subsections. And for the generator, we have

$$\int_0^\infty e^{-\rho t}\int \phi\mathcal{A}V da\, dz\, dt = \int_0^\infty e^{-\rho t}\iint V\mathcal{A}^*\phi da\, dz\, dt.$$

Finally, for the consumption FOC, we simply have

$$-\int_0^\infty e^{-\rho t}\iint \chi V_a da\, dz\, dt = \int_0^\infty e^{-\rho t}\iint \chi_a V da\, dz\, dt,$$

where we also drop boundary terms.

The functional Lagrangian can thus be rewritten as

$$
\begin{aligned}
L^{\text{SP}}(g_0) = \int_0^\infty e^{-\rho t} \Bigg\{ \iint \Bigg\{ & \left[u(c) - \mu c - v(N) - \frac{\delta}{2}(\pi^w)^2 \right] g \\
& - V\phi_t + V\mathcal{A}^*\phi + \phi\left[u(c) - v(N) - \frac{\delta}{2}(\pi^w)^2 \right] \\
& + \chi u'(c) + \chi_a V \\
& + g\lambda_t - \rho\lambda g + g\mathcal{A}\lambda \Bigg\} dadz \\
& + \mu AN \\
& + \theta_t \pi^w + \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon}(1 + \tau^L)A\Lambda - v'(N) \right) N \Bigg\} dt.
\end{aligned}
$$

We now consider a general functional perturbation around a candidate optimal Ramsey plan, and parametrize this perturbation by $\alpha \in \mathbb{R}$. Since α is a scalar, the maximum principle then implies that our candidate plan can only be optimal if $L^{\text{SP}}_\alpha(g_0, \alpha) \mid_{\alpha=0} = 0$.

We have

$$
\begin{aligned}
L^{\text{SP}}(g_0, \alpha) = \int_0^\infty e^{-\rho t} \Bigg\{ \iint \Bigg\{ & \left[u(c + \alpha h_c) - \mu(c + \alpha h_c) - v(N + \alpha h_N) - \frac{\delta}{2}(\pi^w + \alpha h_\pi)^2 \right] (g + \alpha h_g) \\
& - (V + \alpha h_V)\phi_t + (V + \alpha h_V)\mathcal{A}^*(\alpha)\phi \\
& + \phi\left[u(c + \alpha h_c) - v(N + \alpha h_N) - \frac{\delta}{2}(\pi^w + \alpha h_\pi)^2 \right] \\
& + \chi u'(c + \alpha h_c) + \chi_a(V + \alpha h_V) \\
& + (g + \alpha h_g)\lambda_t - \rho\lambda(g + \alpha h_g) + (g + \alpha h_g)\mathcal{A}(\alpha)\lambda \Bigg\} dadz \\
& + \mu A(N + \alpha h_N) + \theta_t(\pi^w + \alpha h_\pi) \\
& + \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon}(1 + \tau^L)A \iint u'(c + \alpha h_c)(g + \alpha h_g) dadz - v'(N + \alpha h_N) \right) (N + \alpha h_N) \Bigg\} dt.
\end{aligned}
$$

We now differentiate and take the limit $\alpha \to 0$. Setting the resulting expression to 0, we have the following first-order necessary condition for optimality:

$$
\begin{aligned}
0 = \int_0^\infty e^{-\rho t} \Bigg\{ \iint \Bigg\{ & \left[u'(c) h_c - \mu h_c - v'(N) h_N - \delta \pi^w h_\pi \right] g + h_g \left[u(c) - \mu c - v(N) - \frac{\delta}{2} (\pi^w)^2 \right] \\
& - h_V \phi_t + h_V \mathcal{A}^*(0) \phi + V \frac{d}{d\alpha} \mathcal{A}^*(0) \phi + \phi \left[u'(c) h_c - v'(N) h_N - \delta \pi^w h_\pi \right] \\
& + \chi u''(c) h_c + \chi_a h_V \\
& + h_g \lambda_t - \rho \lambda h_g + h_g \mathcal{A}(0) \lambda + g \frac{d}{d\alpha} \mathcal{A}(0) \lambda \Bigg\} da dz \\
& + \mu A h_N + \theta_t h_\pi \\
& + \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A \iint [u'(c) h_g + u''(c) g h_c] da dz - v'(N) h_N \right) N \\
& + h_N \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A \Lambda - v'(N) \right) \Bigg\} dt,
\end{aligned}
$$

where we have

$$
\frac{d}{d\alpha} \mathcal{A}(0) = (a h_r + z h_w N + z w h_N - h_c) \partial_a
$$

and, again dropping boundary terms,

$$
\begin{aligned}
V \frac{d}{d\alpha} \mathcal{A}^*(0) \phi &= \phi \frac{d}{d\alpha} \mathcal{A}(0) V \\
&= \phi (a h_r + z h_w N + z w h_N - h_c) V_a.
\end{aligned}
$$

Finally, we group terms by h_c, h_g, etc., and invoke the fundamental lemma of the calculus of variations. We directly obtain the optimality conditions that characterize the optimal Ramsey plan of Proposition 4 in the interior of the state space, i.e., abstracting from boundary conditions.

B.2 Discretized Standard Ramsey Problem

A key challenge in solving Ramsey problems with heterogeneous agents is to formally account for boundary conditions, in particular the borrowing constraint at $\underline{a}$. We find it convenient to derive all proofs that explicitly account for the boundary of the state space in a discretized version of our model. To that end, we work with the discretized representation of equilibrium developed in Appendix A.5.

The standard primal Ramsey problem in our baseline HANK model is associated with the discretized Lagrangian

$$
\begin{aligned}
L^{\mathrm{SP}}(\boldsymbol{g}_0) = & \min_{\{\boldsymbol{\phi}_n, \boldsymbol{\chi}_n, \boldsymbol{\lambda}_n, \mu_n, \theta_n\}} \max_{\{\boldsymbol{V}_n, \boldsymbol{c}_{n,[2:J]}, \boldsymbol{g}_n, \pi^w_n, N_n, i_n\}} \sum_{n=0}^{N-1} e^{-\rho t_n} \Bigg\{ \Bigg\{ \\
& + u \begin{pmatrix} i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_t - v(N_n) \mathbf{1}' \boldsymbol{g}_t - \frac{\delta}{2} (\pi^w_n)^2 \mathbf{1}' \boldsymbol{g}_t \\
& + \boldsymbol{\phi}_n' \left[- \rho \boldsymbol{V}_n + \frac{\boldsymbol{V}_{n+1} - \boldsymbol{V}_n}{dt} + u \begin{pmatrix} i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} - v(N_n) - \frac{\delta}{2} (\pi^w_n)^2 \right] \\
& + \boldsymbol{\phi}_n' \boldsymbol{A}^z \boldsymbol{V}_n + \sum_{i \geq 2} \phi_{i,n} \left(i_n a_i - \pi^w_n a_i + \frac{A_{n+1} - A_n}{dt A_n} a_i + z_i A_n N_n - c_{n,i} \right) \frac{\boldsymbol{D}_{a,[i,:]}}{da} \boldsymbol{V}_n \\
& + \boldsymbol{\chi}_{n,[2:J]}' \left[u'(\boldsymbol{c}_{n,[2:J]}) - \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{n+1} \right)_{[2:J]} \right] \\
& - \boldsymbol{\lambda}_n' \frac{\boldsymbol{g}_{n+1} - \boldsymbol{g}_n}{dt} + \boldsymbol{\lambda}_n' (\boldsymbol{A}^z)' \boldsymbol{g}_n \\
& + \sum_{i \geq 2} \lambda_{n,i} \frac{\boldsymbol{D}'_{a,[i,:]}}{da} \left[\begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi^w_n \boldsymbol{a}_{[2:J]} + \frac{A_{n+1}-A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \boldsymbol{g}_n \right] \Bigg\} d\boldsymbol{x} \\
& + \mu_n \left[\boldsymbol{c}_n' \boldsymbol{g}_n d\boldsymbol{x} - A_n N_n \right] \\
& + \theta_n \left[- \frac{\pi^w_{n+1} - \pi^w_n}{dt} + \rho \pi^w_n + \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right) N_n \right] \Bigg\} dt,
\end{aligned}
$$

where the planner takes as given an initial condition for the cross-sectional distribution, $\boldsymbol{g}_0$.

As in Appendix A.5, we fix from the beginning that unemployed households at the borrowing constraint always consume their income, that is

$$
c_{n,1} = i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n
$$

for all n. The planner takes this as given and does not get to consider perturbations in $c_{n,1}$ for any n.

We want to emphasize at this point how important it is exactly which finite-difference stencils are used for the discretization. For discretization in the time dimension, for example, the above Lagrangian assumes a *semi-implicit backwards* discretization of $\partial_t \boldsymbol{V}_t$ in the HJB. And it assumes an *explicit forwards* discretization of $\partial_t \boldsymbol{g}_t$ in the KF equation. For the aggregates, it assumes an *explicit forwards* discretization for $\dot{A}_t$ and also an *explicit forwards* discretization for $\dot{\pi}^w_t$. These assumptions also correspond to the appropriate stencils we use numerically to implement our results.

We also want to echo Achdou et al. (2022) at this point, recalling that the correct discretization stencil for the KF equation in the wealth dimension is given by

$$(\boldsymbol{A}^a)'\boldsymbol{g} = \frac{1}{da}(\boldsymbol{s} \cdot \boldsymbol{D}_a)'\boldsymbol{g} = \frac{1}{da}\boldsymbol{D}_a'(\boldsymbol{s} \cdot \boldsymbol{g}).$$

That is, the correct stencil uses the tranpose $\boldsymbol{D}_a'$ rather than, as one might have expected, $-\boldsymbol{D}_a(\boldsymbol{s} \cdot \boldsymbol{g})$.

B.3 Auxilliary Results

Before tackling the main proof of this appendix, we state several auxilliary results that will be helpful below. Most of these results follow trivially by applying well-known properties of matrix algebra. We consequently provide only some of the proofs explicitly.

Lemma 14. *The following matrix algebra tricks will be useful. Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ be $J \times 1$ vectors and $\boldsymbol{A}$ a $J \times J$ matrix. Transposition satisfies*

$$(\boldsymbol{A}\boldsymbol{x})' = \boldsymbol{x}'\boldsymbol{A}'.$$

We also have

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y} = \sum_i x_i \boldsymbol{A}_{[i,:]}\boldsymbol{y} = \sum_i x_i \sum_j \boldsymbol{A}_{[i,j]} y_j = \sum_j y_j \sum_i \boldsymbol{A}'_{[j,i]} x_i = \boldsymbol{y}'\boldsymbol{A}'\boldsymbol{x}.$$

We also have

$$\boldsymbol{x}'(\boldsymbol{y} \cdot \boldsymbol{A})z = \boldsymbol{x}'(\boldsymbol{y} \cdot (\boldsymbol{A}\boldsymbol{z})) = (\boldsymbol{x} \cdot \boldsymbol{y})'\boldsymbol{A}z = (\boldsymbol{A}\boldsymbol{z})'(\boldsymbol{x} \cdot \boldsymbol{y}) = \boldsymbol{z}'\boldsymbol{A}'(\boldsymbol{x} \cdot \boldsymbol{y}) = \boldsymbol{z}'(\boldsymbol{y} \cdot \boldsymbol{A})'\boldsymbol{x}.$$

Taking derivatives, we have

$$\begin{aligned}
\frac{d}{d\boldsymbol{x}}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y} &= \boldsymbol{A}\boldsymbol{y} \\
\frac{d}{d\boldsymbol{x}}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{x} &= (\boldsymbol{y}'\boldsymbol{A})' = \boldsymbol{A}'\boldsymbol{y}.
\end{aligned}$$

Lemma 15. *In the Lagrangian, the HJB term can be rearranged as follows:*

$$\begin{aligned}\frac{1}{da}\sum_{i\geq 2}\phi_i s_i \boldsymbol{D}_{a,[i,:]}\boldsymbol{V} &= \frac{1}{da}\sum_{i\geq 1}\phi_i s_i \boldsymbol{D}_{a,[i,:]}\boldsymbol{V} \\ &= \frac{1}{da}\boldsymbol{\phi}'(\boldsymbol{s}\cdot\boldsymbol{D}_a)\boldsymbol{V} \\ &= \frac{1}{da}\boldsymbol{V}'(\boldsymbol{s}\cdot\boldsymbol{D}_a)'\boldsymbol{\phi} \\ &= \frac{1}{da}\boldsymbol{V}'\boldsymbol{D}_a'(\boldsymbol{s}\cdot\boldsymbol{\phi}),\end{aligned}$$

where $\boldsymbol{D}_a$ is the upwind finite-difference matrix in the a dimension. We sometimes use $\boldsymbol{s}\cdot\boldsymbol{D}_a = \boldsymbol{A}^a$.

Proof. We have

$$\begin{aligned}\sum_{i\geq 2}\phi_i s_i \boldsymbol{D}_{a,[i,\,:]}\boldsymbol{V} &= \sum_{i\geq 2}\phi_i s_i \sum_{j\geq 1}\boldsymbol{D}_{a,[i,\,j]}\boldsymbol{V}_j \\ &= \sum_{i\geq 2}\phi_i s_i \sum_{j\geq 1}\boldsymbol{D}'_{a,[j,\,i]}\boldsymbol{V}_j \\ &= \sum_{j\geq 1}\boldsymbol{V}_j \sum_{i\geq 2}\boldsymbol{D}'_{a,[j,\,i]}\phi_i s_i \\ &= \sum_{j\geq 1}\boldsymbol{V}_j \sum_{i\geq 1}\boldsymbol{D}'_{a,[j,\,i]}\phi_i s_i \\ &= \sum_{j\geq 1}\boldsymbol{V}_j \boldsymbol{D}'_{a,[j,\,:]}(\boldsymbol{s}\cdot\boldsymbol{\phi}) \\ &= \boldsymbol{V}'\boldsymbol{D}_a'(\boldsymbol{s}\cdot\boldsymbol{\phi}),\end{aligned}$$

where $D'_{a,[j,\,:]}$ denotes the jth row of the matrix $\boldsymbol{D}'_a$. ■

Lemma 16. *The correct adjoint operation, i.e., the one we use to define $\mathcal{A}^* \approx \boldsymbol{A}'$, is given by*

$$\boldsymbol{D}_a'(\boldsymbol{s}\cdot\boldsymbol{\phi}) = (\boldsymbol{A}^a)'\boldsymbol{\phi}.$$

In particular, we have

$$\begin{aligned}\boldsymbol{\lambda}'(\boldsymbol{A}^a)'\boldsymbol{g} &= \boldsymbol{\lambda}'\boldsymbol{D}_a'(\boldsymbol{s}\cdot\boldsymbol{g}) \\ &= (\boldsymbol{s}\cdot\boldsymbol{g})'\boldsymbol{D}_a\boldsymbol{\lambda} \\ &= \boldsymbol{g}'(\boldsymbol{s}\cdot\boldsymbol{D}_a)\boldsymbol{\lambda} \\ &= (\boldsymbol{D}_a\boldsymbol{\lambda})'(\boldsymbol{s}\cdot\boldsymbol{g}).\end{aligned}$$

Lemma 17. *We can "integrate by parts" the FOC term in the Lagrangian to arrive at*

$$\frac{1}{da}\boldsymbol{\chi}'_{t(n),[2:J]}\left(\boldsymbol{D}_a\boldsymbol{V}_{t(n+1)}\right)_{[2:J]} = \frac{1}{da}\boldsymbol{V}'_{t(n+1)}\boldsymbol{D}_a'\begin{pmatrix}0\\ \boldsymbol{\chi}_{[2:J],t(n)}\end{pmatrix}.$$

Proof. We have

$$\begin{aligned}\frac{1}{da}\sum_{i\geq 2}\chi_{i,t(n)}\boldsymbol{D}_{a,[i,:]}\boldsymbol{V}_{t(n+1)} &= \frac{1}{da}\sum_{i\geq 2}\chi_{i,t(n)}\sum_{j\geq 1}\boldsymbol{D}_{a,[i,j]}\boldsymbol{V}_{j,t(n+1)} \\ &= \frac{1}{da}\sum_{j\geq 1}\boldsymbol{V}_{j,t(n+1)}\sum_{i\geq 2}\boldsymbol{D}'_{a,[j,i]}\chi_{i,t(n)} \\ &= \frac{1}{da}\sum_{j\geq 1}\boldsymbol{V}_{j,t(n+1)}\boldsymbol{D}'_{a,[j,:]}\begin{pmatrix}0\\ \boldsymbol{\chi}_{[2:J],t(n)}\end{pmatrix} \\ &= \frac{1}{da}\boldsymbol{V}'_{t(n+1)}\boldsymbol{D}_a'\begin{pmatrix}0\\ \boldsymbol{\chi}_{[2:J],t(n)}\end{pmatrix}.\end{aligned}$$

It is important to note that we *cannot* roll the sum $\sum_{i\geq 2}$ forward to simply read $\sum_{i\geq 1}$. This is only possible for the terms that include savings, using the fact that $s_1 = 0$. ■

Lemma 18. *We can "integrate by parts" in the time dimension as follows. For any $\boldsymbol{x}_n$, we have*

$$\sum_{n=0}^{N-1} e^{-\rho t_n}\boldsymbol{x}_{n+1} = e^{\rho dt}\sum_{n=0}^{N-1} e^{-\rho t_n}\boldsymbol{x}_n - e^{\rho dt}\boldsymbol{x}_0 + e^{\rho dt}e^{-\rho t_N}\boldsymbol{x}_N.$$

We prove the following results below. In particular, this implies

$$\sum_{n=0}^{N-1} e^{-\rho t_n}\boldsymbol{\phi}_n'\boldsymbol{V}_{n+1} = \sum_{n=0}^{N-1} e^{-\rho t_n}e^{\rho dt}\boldsymbol{\phi}'_{n-1}\boldsymbol{V}_n - e^{\rho dt}\boldsymbol{\phi}'_{-1}\boldsymbol{V}_0 + e^{\rho dt}e^{-\rho t_N}\boldsymbol{\phi}_{N-1}\boldsymbol{V}_N,$$

as well as

$$-\sum_{n=0}^{N-1} e^{-\rho t_n}\boldsymbol{\lambda}_n' \frac{\boldsymbol{g}_{n+1}-\boldsymbol{g}_n}{dt} = \sum_{n=0}^{N-1} e^{-\rho t_n}\frac{\boldsymbol{\lambda}_n' - e^{\rho dt}\boldsymbol{\lambda}_{n-1}'}{dt}\boldsymbol{g}_n + \frac{1}{dt}e^{\rho dt}\boldsymbol{\lambda}_{-1}'\boldsymbol{g}_0 - \frac{1}{dt}e^{\rho dt}e^{-\rho t_N}\boldsymbol{\lambda}_{N-1}'\boldsymbol{g}_N,$$

and

$$\sum_{n=0}^{N-1} e^{-\rho t_n}\boldsymbol{\chi}_{n,[2:J]}'\left(\frac{\boldsymbol{D}_a}{da}\boldsymbol{V}_{n+1}\right)_{[2:J]} = \sum_{n=0}^{N-1} e^{-\rho t_n}e^{\rho dt}\boldsymbol{\chi}_{n-1,[2:J]}'\left(\frac{\boldsymbol{D}_a}{da}\boldsymbol{V}_n\right)_{[2:J]} - e^{\rho dt}\boldsymbol{\chi}_{-1,[2:J]}'\left(\frac{\boldsymbol{D}_a}{da}\boldsymbol{V}_0\right)_{[2:J]} + e^{\rho dt}e^{-\rho t_N}\boldsymbol{\chi}_{N-1,[2:J]}'\left(\frac{\boldsymbol{D}_a}{da}\boldsymbol{V}_N\right)_{[2:J]}.$$

Finally, we have

$$-\sum_{n=0}^{N-1} e^{-\rho t_n}\theta_n \frac{\pi_{n+1}^w - \pi_n^w}{dt} = \sum_{n=0}^{N-1} e^{-\rho t_n}\frac{\theta_n - e^{\rho dt}\theta_{n-1}}{dt}\pi_n^w + \frac{1}{dt}e^{\rho dt}\theta_{-1}\pi_0^w - \frac{1}{dt}e^{\rho dt}e^{-\rho t_N}\theta_{N-1}\pi_N^w.$$

Proof. We have

$$\begin{aligned}
\sum_{n=0}^{\infty} e^{-\rho t(n)}\boldsymbol{\phi}_{t(n)}'\frac{1}{dt}\boldsymbol{V}_{t(n+1)} &= \sum_{n=0}^{\infty} e^{-\rho t(n)}e^{\rho t(n+1)}e^{-\rho t(n+1)}\boldsymbol{\phi}_{t(n)}'\frac{1}{dt}\boldsymbol{V}_{t(n+1)} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n+1)}e^{\rho t(n+1)-\rho t(n)}\boldsymbol{\phi}_{t(n)}'\frac{1}{dt}\boldsymbol{V}_{t(n+1)} \\
&= \sum_{n=1}^{\infty} e^{-\rho t(n)}e^{\rho dt}\boldsymbol{\phi}_{t(n-1)}'\frac{1}{dt}\boldsymbol{V}_{t(n)} \\
&= \sum_{n=1}^{\infty} e^{-\rho t(n)}e^{\rho dt}\boldsymbol{\phi}_{t(n-1)}'\frac{1}{dt}\boldsymbol{V}_{t(n)} + e^{-\rho t(0)}e^{\rho dt}\boldsymbol{\phi}_{-1}'\frac{1}{dt}\boldsymbol{V}_{t(0)} - e^{-\rho t(0)}e^{\rho dt}\boldsymbol{\phi}_{-1}'\frac{1}{dt}\boldsymbol{V}_{t(0)} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n)}e^{\rho dt}\boldsymbol{\phi}_{t(n-1)}'\frac{1}{dt}\boldsymbol{V}_{t(n)} - e^{-\rho t(0)}e^{\rho dt}\boldsymbol{\phi}_{-1}'\frac{1}{dt}\boldsymbol{V}_{t(0)}.
\end{aligned}$$

Similarly, we can rearrange

$$\begin{aligned}
\sum_{n=0}^{\infty} e^{-\rho t(n)} \boldsymbol{\chi}'_{t(n),[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(n+1)} \right)_{[2:J]} &= \sum_{n=0}^{\infty} e^{-\rho t(n)} e^{\rho t(n+1)} e^{-\rho t(n+1)} \boldsymbol{\chi}'_{t(n),[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(n+1)} \right)_{[2:J]} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n+1)} e^{\rho t(n+1) - \rho t(n)} \boldsymbol{\chi}'_{t(n),[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(n+1)} \right)_{[2:J]} \\
&= \sum_{n=1}^{\infty} e^{-\rho t(n)} e^{\rho dt} \boldsymbol{\chi}'_{t(n-1),[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(n)} \right)_{[2:J]} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n)} e^{\rho dt} \boldsymbol{\chi}'_{t(n-1),[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(n)} \right)_{[2:J]} - e^{-\rho t(0)} e^{\rho dt} \boldsymbol{\chi}'_{-1,[2:J]} \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{t(0)} \right)_{[2:J]}.
\end{aligned}$$

Finally, notice that

$$\begin{aligned}
e^{\rho dt} \boldsymbol{\phi}'_{t(n-1)} \frac{1}{dt} \boldsymbol{V}_{t(n)} &= (1 + \rho dt) \boldsymbol{\phi}'_{t(n-1)} \frac{1}{dt} \boldsymbol{V}_{t(n)} \\
&= \boldsymbol{\phi}'_{t(n-1)} \frac{1}{dt} \boldsymbol{V}_{t(n)} + \rho \boldsymbol{\phi}'_{t(n-1)} \boldsymbol{V}_{t(n)}.
\end{aligned}$$

Lastly,

$$\begin{aligned}
-\sum_{n=0}^{\infty} e^{-\rho t(n)} \boldsymbol{\lambda}'_{t(n)} \frac{\boldsymbol{g}_{t(n+1)} - \boldsymbol{g}_{t(n)}}{dt(n)} &= -\sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n+1)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)} \\
&= -\frac{1}{dt} \sum_{n=0}^{\infty} e^{-\rho t(n)} e^{\rho t(n+1)} e^{-\rho t(n+1)} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n+1)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)} \\
&= -\frac{1}{dt} e^{\rho dt} \sum_{n=0}^{\infty} e^{-\rho t(n+1)} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n+1)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)} \\
&= -\frac{1}{dt} e^{\rho dt} \sum_{n=1}^{\infty} e^{-\rho t(n)} \boldsymbol{\lambda}'_{t(n-1)} \boldsymbol{g}_{t(n)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)}.
\end{aligned}$$

And so we get

$$\begin{aligned}
&= -\frac{1}{dt} e^{\rho dt} \sum_{n=1}^{\infty} e^{-\rho t(n)} \boldsymbol{\lambda}'_{t(n-1)} \boldsymbol{g}_{t(n)} + \frac{1}{dt} e^{\rho dt} e^{-\rho t(0)} \boldsymbol{\lambda}'_{t(-1)} \boldsymbol{g}_{t(0)} - \frac{1}{dt} e^{\rho dt} e^{-\rho t(0)} \boldsymbol{\lambda}'_{t(-1)} \boldsymbol{g}_{t(0)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)} \\
&= -\frac{1}{dt} e^{\rho dt} \sum_{n=0}^{\infty} e^{-\rho t(n)} \boldsymbol{\lambda}'_{t(n-1)} \boldsymbol{g}_{t(n)} + \frac{1}{dt} e^{\rho dt} e^{-\rho t(0)} \boldsymbol{\lambda}'_{t(-1)} \boldsymbol{g}_{t(0)} + \sum_{n=0}^{\infty} e^{-\rho t(n)} \frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} \boldsymbol{g}_{t(n)} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n)} \left(\frac{1}{dt} \boldsymbol{\lambda}'_{t(n)} - \frac{1}{dt} e^{\rho dt} \boldsymbol{\lambda}'_{t(n-1)} \right) \boldsymbol{g}_{t(n)} + \frac{1}{dt} e^{\rho dt} e^{-\rho t(0)} \boldsymbol{\lambda}'_{t(-1)} \boldsymbol{g}_{t(0)} \\
&= \sum_{n=0}^{\infty} e^{-\rho t(n)} \left(\frac{\boldsymbol{\lambda}'_{t(n)} - \boldsymbol{\lambda}'_{t(n-1)}}{dt} - \rho \boldsymbol{\lambda}'_{t(n-1)} \right) \boldsymbol{g}_{t(n)} + \frac{1}{dt} e^{\rho dt} e^{-\rho t(0)} \boldsymbol{\lambda}'_{t(-1)} \boldsymbol{g}_{t(0)}.
\end{aligned}$$

Finally, we drop the second term on the RHS because $\boldsymbol{g}_{t(0)}$ is fixed as an initial condition and so it does not respond to $\frac{d}{d\boldsymbol{\theta}}$, which is precisely why the KFE is not a forward-looking constraint. ■

Lemma 19. *In the continuous time limit as $dt \to 0$, we have*

$$e^{\rho dt} \approx 1 + \rho dt.$$

B.4 Proof of Proposition 4

We are now ready to present our main proof. We use the auxilliary results above to rewrite the discretized Lagrangian that corresponds to the standard primal Ramsey problem of Section 4 as

$$
\begin{aligned}
L^{\text{SP}}(\boldsymbol{g}_0) = \min_{\{\boldsymbol{\phi}_n,\boldsymbol{\chi}_n,\boldsymbol{\lambda}_n,\mu_n,\theta_n\}} \max_{\{\boldsymbol{V}_n,\boldsymbol{c}_{n,[2:J]},\boldsymbol{g}_n,\pi_n^w,N_n,i_n\}} \sum_{n=0}^{N-1} e^{-\rho t_n} \Bigg\{ \Bigg\{ & u\begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_n \\
& + \mu_n \begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_n - v(N_n)\mathbf{1}'\boldsymbol{g}_n - \frac{\delta}{2}(\pi_n^w)^2 \mathbf{1}'\boldsymbol{g}_n \\
& - \frac{\boldsymbol{\phi}_n' - e^{\rho dt}\boldsymbol{\phi}_{n-1}'}{dt}\boldsymbol{V}_n + \boldsymbol{\phi}_n'\Bigg[-\rho \boldsymbol{V}_n + u\begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} - v(N_n) - \frac{\delta}{2}(\pi_n^w)^2 \Bigg] \\
& + \boldsymbol{\phi}_n' \boldsymbol{A}^z \boldsymbol{V}_n + \frac{1}{da}\boldsymbol{V}_n'(\boldsymbol{\phi}_n \cdot \boldsymbol{D}_a)' \begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi_n^w \boldsymbol{a}_{[2:J]} + \frac{A_{n+1}-A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \\
& + \boldsymbol{\chi}_{n,[2:J]}' u'(\boldsymbol{c}_{n,[2:J]}) - e^{\rho dt}\frac{1}{da}\boldsymbol{V}_n'\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{n-1,[2:J]} \end{pmatrix} \\
& + \frac{\boldsymbol{\lambda}_n' - e^{\rho dt}\boldsymbol{\lambda}_{n-1}'}{dt}\boldsymbol{g}_n + \boldsymbol{\lambda}_n'(\boldsymbol{A}^z)'\boldsymbol{g}_n \\
& + \frac{1}{da}(\boldsymbol{D}_a \boldsymbol{\lambda}_t)' \Bigg[\begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi_n^w \boldsymbol{a}_{[2:J]} + \frac{A_{n+1}-A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \boldsymbol{g}_n \Bigg] \Bigg\} d\boldsymbol{x} \\
& - \mu_n A_n N_n \\
& + \frac{\theta_n - e^{\rho dt}\theta_{n-1}}{dt}\pi_n^w + \theta_n \rho \pi_n^w - \theta_n \frac{\epsilon}{\delta} v'(N_n) N_n \\
& + \theta_n \frac{\epsilon}{\delta} N_n \frac{\epsilon - 1}{\epsilon}(1+\tau^L) A_n (\boldsymbol{z}\cdot \boldsymbol{g}_n)' u' \begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1}-A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} d\boldsymbol{x} \Bigg\} dt \\
& - e^{\rho dt}\boldsymbol{\phi}_{-1}'\boldsymbol{V}_0 d\boldsymbol{x} + e^{\rho dt} e^{-\rho t_N}\boldsymbol{\phi}_{N-1}\boldsymbol{V}_N d\boldsymbol{x} \\
& + e^{\rho dt}\frac{1}{da}\boldsymbol{V}_0'\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{-1,[2:J]} \end{pmatrix} d\boldsymbol{x}dt - e^{\rho dt} e^{-\rho t_N} \boldsymbol{V}_N' \boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{N-1,[2:J]} \end{pmatrix} d\boldsymbol{x}dt \\
& + e^{\rho dt}\boldsymbol{\lambda}_{-1}'\boldsymbol{g}_0 d\boldsymbol{x} - e^{\rho dt}e^{-\rho t_N}\boldsymbol{\lambda}_{N-1}'\boldsymbol{g}_N d\boldsymbol{x} \\
& + e^{\rho dt}\theta_{-1}\pi_0^w - e^{\rho dt}e^{-\rho t_N}\theta_{N-1}\pi_N^w.
\end{aligned}
$$

In the spirit of Marcet and Marimon (2019), we have reordered the forward-looking constraints—this corresponds to summation (integration) by parts. The resulting "boundary" terms in the last few lines of the above Lagrangian are the key objects at the heart of the time-0 problem, which we discuss in Section 4.

To conclude our proof of Proposition 4, we now take derivatives and characterize the necessary first-order conditions. In particular, we do so for $n \geq 1$ precisely in order to avoid the boundary terms that give rise to the time-0 problem. In the continuous time limit as $dt \to 0$, these terms give rise to the initial conditions $\phi_0(a, z) = 0$ and $\theta_0 = 0$, as we explain in our statement of Proposition 4 and the discussion in the main text. This follows straightforwardly from basic calculus of variations (see e.g. Kamien and Schwartz, 2012). We revisit these boundary terms below when we prove the timeless property of the timeless Ramsey plans.

Derivative $\boldsymbol{V}_n$. We have

$$0 = -\frac{\boldsymbol{\phi}_n' - e^{\rho dt}\boldsymbol{\phi}_{n-1}'}{dt} - \rho\boldsymbol{\phi}_n + (\boldsymbol{A}^z)'\boldsymbol{\phi}_n + \frac{1}{da}(\boldsymbol{\phi}_n \cdot \boldsymbol{D}_a)'\boldsymbol{s}_n - e^{\rho dt}\frac{1}{da}\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{n-1,[2:J]} \end{pmatrix}.$$

Using our auxilliary results, we have $(\boldsymbol{\phi}_n \cdot \boldsymbol{D}_a)'\boldsymbol{s}_n = (\boldsymbol{s}_n \cdot \boldsymbol{D}_a)'\boldsymbol{\phi}_n = (\boldsymbol{A}^a)'\boldsymbol{\phi}_n$, and so

$$0 = -\frac{\boldsymbol{\phi}_n' - e^{\rho dt}\boldsymbol{\phi}_{n-1}'}{dt} - \rho\boldsymbol{\phi}_n + \boldsymbol{A}'\boldsymbol{\phi}_n - e^{\rho dt}\frac{1}{da}\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{n-1,[2:J]} \end{pmatrix}.$$

Derivative $\boldsymbol{g}_n$. We have

$$\begin{aligned} 0 =& u(\boldsymbol{c}_n) + \mu_n \boldsymbol{c}_n - v(N_n)\mathbf{1} - \frac{\delta}{2}(\pi_n^w)^2\mathbf{1} + \frac{\boldsymbol{\lambda}_n' - e^{\rho dt}\boldsymbol{\lambda}_{n-1}'}{dt} + (\boldsymbol{\lambda}_n'(\boldsymbol{A}^z)')' \\ &+ \frac{d}{d\boldsymbol{g}_n}\left[\frac{1}{da}(\boldsymbol{D}_a\boldsymbol{\lambda}_n)'[\boldsymbol{s}_n \cdot \boldsymbol{g}_n]\right] + \theta_n N_t \frac{\epsilon}{\delta}\frac{\epsilon - 1}{\epsilon}(1 + \tau^L)A_n \boldsymbol{z} \cdot u'(\boldsymbol{c}_n). \end{aligned}$$

Now we work out the remaining derivative,

$$\begin{aligned} \frac{d}{d\boldsymbol{g}_n}\left[\frac{1}{da}(\boldsymbol{D}_a\boldsymbol{\lambda}_n)'[\boldsymbol{s}_n \cdot \boldsymbol{g}_n]\right] &= \frac{1}{da}\frac{d}{d\boldsymbol{g}_n}\left[\left(\boldsymbol{s}_n' \cdot (\boldsymbol{D}_a\boldsymbol{\lambda}_n)'\right)\boldsymbol{g}_n\right] \\ &= \frac{1}{da}\frac{d}{d\boldsymbol{g}_n}\left[\boldsymbol{g}_n'\left(\boldsymbol{s}_n \cdot (\boldsymbol{D}_a\boldsymbol{\lambda}_n)\right)\right] \\ &= \frac{1}{da}\frac{d}{d\boldsymbol{g}_n}\left[\boldsymbol{g}_n'\left((\boldsymbol{s}_n \cdot \boldsymbol{D}_a)\boldsymbol{\lambda}_n\right)\right] \\ &= \frac{1}{da}(\boldsymbol{s}_n \cdot \boldsymbol{D}_a)\boldsymbol{\lambda}_n \\ &= \boldsymbol{A}^a\boldsymbol{\lambda}_n. \end{aligned}$$

Thus, we have

$$0 = u(\boldsymbol{c}_n) + \mu_n \boldsymbol{c}_n - v(N_n)\mathbf{1} - \frac{\delta}{2}(\pi_n^w)^2\mathbf{1} + \frac{\boldsymbol{\lambda}_n' - e^{\rho dt}\boldsymbol{\lambda}_{n-1}'}{dt} + \boldsymbol{A}\boldsymbol{\lambda}_n + \theta_n N_t \frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n \boldsymbol{z}\cdot u'(\boldsymbol{c}_n).$$

Derivative $\boldsymbol{c}_{n,[2:J]}$. We now take the derivative with respect to $c_{n,i}$ for $i \geq 2$. We have

$$\begin{aligned} 0 =& u'(c_{n,i})g_{n,i} + \mu_n g_{n,i} + u'(c_{n,i})\phi_{n,i} + u''(c_{n,i})\chi_{n,i} + \theta_n N_n \frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n z_i u''(c_{n,i})g_{n,i} \\ &+ \frac{d}{dc_{n,i}}\left[\frac{1}{da}\boldsymbol{s}_n'(\boldsymbol{\phi}_n\cdot\boldsymbol{D}_a)\boldsymbol{V}_n\right] + \frac{d}{dc_{n,i}}\left[\frac{1}{da}\boldsymbol{s}_n'\cdot(\boldsymbol{D}_a\boldsymbol{\lambda}_n)'\boldsymbol{g}_n\right]. \end{aligned}$$

Working out the remaining derivatives, we have

$$\begin{aligned} \frac{d}{dc_{n,i}}\left[\frac{1}{da}\boldsymbol{s}_n'(\boldsymbol{\phi}_n\cdot\boldsymbol{D}_a)\boldsymbol{V}_n\right] &= \frac{1}{da}\Big((\boldsymbol{\phi}_n\cdot\boldsymbol{D}_a)\boldsymbol{V}_n\Big)_{[i]}\frac{ds_{n,i}}{dc_{n,i}} \\ &= -\frac{1}{da}\Big((\boldsymbol{\phi}_n\cdot\boldsymbol{D}_a)\boldsymbol{V}_n\Big)_{[i]} \\ &= -\frac{1}{da}\phi_{n,i}\Big(\boldsymbol{D}_a\boldsymbol{V}_n\Big)_{[i]} \\ &= -\frac{1}{da}\phi_{n,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{V}_n. \end{aligned}$$

And similarly,

$$\begin{aligned} \frac{d}{dc_{n,i}}\left[\frac{1}{da}\boldsymbol{s}_n'\cdot(\boldsymbol{D}_a\boldsymbol{\lambda}_n)'\boldsymbol{g}_n\right] &= \frac{d}{dc_{n,i}}\left[\frac{1}{da}\boldsymbol{g}_n'\Big(\boldsymbol{s}_n\cdot(\boldsymbol{D}_a\boldsymbol{\lambda}_n)\Big)\right] \\ &= \frac{ds_{n,it}}{dc_{n,i}}\frac{1}{da}g_{n,i}(\boldsymbol{D}_a\boldsymbol{\lambda}_n)_{[i]} \\ &= -\frac{1}{da}g_{n,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{\lambda}_n. \end{aligned}$$

Thus, we have

$$\begin{aligned} 0 =& u'(c_{n,i})g_{n,i} + \mu_n g_{n,i} + u'(c_{n,i})\phi_{n,i} + u''(c_{n,i})\chi_{n,i} + \theta_n N_n \frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n z_i u''(c_{n,i})g_{n,i} \\ &- \frac{1}{da}\phi_{n,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{V}_n - \frac{1}{da}g_{n,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{\lambda}_n. \end{aligned}$$

Derivative π_n^w. We have

$$0 = \Big[-u'(c_{n,1})g_{n,1}a_1 - \mu_n g_{n,1}a_1 - \delta\pi_n^w \mathbf{1}'\boldsymbol{g}_n - \phi_{n,1}u'(c_{n,1})a_1 - \delta\pi_n^w \boldsymbol{\phi}_n'\mathbf{1}\Big] d\boldsymbol{x}$$
$$- \theta_n \frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n N_n z_1 u''(c_{n,1})g_{n,1}a_1 d\boldsymbol{x}$$
$$+ \Big[-\sum_{i\geq 2}\phi_{i,n}a_i \frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_n + \sum_{i\geq 2}\lambda_{n,i}\frac{\boldsymbol{D}'_{a,[i,:]}}{da}\Big[\begin{pmatrix} 0 \\ -\boldsymbol{a}_{[2:J]}\end{pmatrix} \cdot \boldsymbol{g}_n\Big] d\boldsymbol{x}$$
$$+ \frac{\theta_n - e^{\rho dt}\theta_{n-1}}{dt} + \rho\theta_n.$$

Alternatively, we have

$$\begin{aligned}
\frac{d}{d\pi_n^w}\frac{1}{da}(\boldsymbol{D}_a\boldsymbol{\lambda}_n)'[\boldsymbol{s}_n \cdot \boldsymbol{g}_n] &= \frac{d}{d\pi_n^w}\frac{1}{da}\Big(\boldsymbol{s}_n \cdot \boldsymbol{D}_a\boldsymbol{\lambda}_n\Big)'\boldsymbol{g}_n \\
&= \frac{d}{d\pi_n^w}\frac{1}{da}\boldsymbol{g}_n'\Big(\boldsymbol{s}_n \cdot \boldsymbol{D}_a\boldsymbol{\lambda}_n\Big) \\
&= \frac{1}{da}\boldsymbol{g}_n'\Big(\frac{d\boldsymbol{s}_n}{d\pi_n^w} \cdot \boldsymbol{D}_a\boldsymbol{\lambda}_n\Big) \\
&= \frac{1}{da}\boldsymbol{g}_n'\left(\begin{pmatrix} 0 \\ -\boldsymbol{a}_{[2:J]}\end{pmatrix} \cdot \boldsymbol{D}_a\boldsymbol{\lambda}_n\right) \\
&= \sum_{i\geq 1} g_{n,i}\left(\begin{pmatrix} 0 \\ -\boldsymbol{a}_{[2:J]}\end{pmatrix} \cdot \frac{\boldsymbol{D}_a}{da}\boldsymbol{\lambda}_n\right)_{[i]} \\
&= -\sum_{i\geq 2} g_{n,i}a_i \frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_n.
\end{aligned}$$

Thus, we have

$$0 = \Big[-u'(c_{n,1})g_{n,1}a_1 - \mu_n g_{n,1}a_1 - \delta\pi_n^w \mathbf{1}'\boldsymbol{g}_n - \phi_{n,1}u'(c_{n,1})a_1 - \delta\pi_n^w \boldsymbol{\phi}_n'\mathbf{1}\Big] d\boldsymbol{x}$$
$$- \theta_n \frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n N_n z_1 u''(c_{n,1})g_{n,1}a_1 d\boldsymbol{x}$$
$$- \sum_{i\geq 2}\phi_{i,n}a_i \frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_n d\boldsymbol{x} - \sum_{i\geq 2} g_{n,i}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_n d\boldsymbol{x} + \frac{\theta_n - e^{\rho dt}\theta_{n-1}}{dt} + \rho\theta_n.$$

Derivative i_n. The nominal interest rate derivative is very easy because it's parallel to wage inflation, except in the Phillips curve. That is, we have

$$\begin{aligned}0 =& u'(c_{n,1})g_{n,1}a_1 + \mu_n g_{n,1}a_1 + \phi_{n,1}u'(c_{n,1})a_1 + \sum_{i\geq 2}\phi_{i,n}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_n + \sum_{i\geq 2}g_{n,i}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_n \\ &+ \theta_n\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_nN_nz_1u''(c_{n,1})g_{n,1}a_1.\end{aligned}$$

Derivative N_n. Finally, we take the derivative for aggregate labor. This yields

$$\begin{aligned}0 =& \left[u'(c_{n,1})g_{n,1}z_1A_n + \mu_n g_{n,1}z_1A_n + \phi_{n,1}u'(c_{n,1})z_1A_n + \sum_{i\geq 2}\phi_{i,n}z_iA_n\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_n + \sum_{i\geq 2}g_{n,i}z_iA_n\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_n\right]d\boldsymbol{x} \\ &+ \theta_n\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_nN_nz_1u''(c_{n,1})g_{n,1}z_1A_nd\boldsymbol{x} \\ &- v'(N_n)\boldsymbol{1}'\boldsymbol{g}_nd\boldsymbol{x} - v'(N_n)\boldsymbol{\phi}_n'\boldsymbol{1}d\boldsymbol{x} \\ &- \mu_nA_n + \theta_n\frac{\epsilon}{\delta}\left(\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_n(\boldsymbol{z}\cdot u'(\boldsymbol{c}_n))'\boldsymbol{g}_nd\boldsymbol{x} - v'(N_n)\right) - \frac{\epsilon}{\delta}\theta_nv''(N_n)N_n.\end{aligned}$$

These derivations conclude our proof. In particular, the first-order conditions we have now derived are the exact, discretized analogs of the conditions we present in Proposition 4.

B.5 Stationary Ramsey Plan and Proof of Proposition 5

We now formally state the discretized characterization of the stationary Ramsey plan. We use the fact that, in any stationary equilibrium, we simply have

$$u'(\boldsymbol{c}_i) = \frac{1}{da}\boldsymbol{D}_{a,[i,:]}\boldsymbol{V},$$

for $i \geq 2$.

Lemma 20. **(Discretized Stationary Ramsey Plan)** *A consistent discretization of the stationary Ramsey plan, with $A_{ss} = 1$, is given by the following equations. For the value function, we have*

$$0 = -\frac{1-e^{\rho dt}}{dt}\boldsymbol{\phi} - \rho\boldsymbol{\phi} + \boldsymbol{A}'\boldsymbol{\phi} - e^{\rho dt}\frac{1}{da}\boldsymbol{D}_a'\begin{pmatrix}0\\ \boldsymbol{\chi}_{[2:J]}\end{pmatrix}$$

and for the distribution

$$0 = \frac{1 - e^{\rho dt}}{dt}\boldsymbol{\lambda} + \boldsymbol{A}\boldsymbol{\lambda} + u(\boldsymbol{c}) + \mu \boldsymbol{c} - v(N) - \frac{\delta}{2}(\pi^w)^2 + \theta N \frac{\epsilon}{\delta}\frac{\epsilon - 1}{\epsilon}(1 + \tau^L)\boldsymbol{z} \cdot u'(\boldsymbol{c}).$$

For consumption, for $i \geq 2$, we have

$$-u''(c_i)\chi_i = \left[u'(c_i) + \mu + \theta N \frac{\epsilon}{\delta}\frac{\epsilon - 1}{\epsilon}(1 + \tau^L) z_i u''(c_i) - \frac{1}{da}\boldsymbol{D}_{a,[i,:]}\boldsymbol{\lambda}\right] g_i.$$

The optimality condition for monetary policy, i.e., the nominal interest rate, is given by

$$0 = \left(u'(c_1) + \mu - \frac{1}{da}\boldsymbol{D}_{a,[1,:]}\boldsymbol{\lambda} + \theta \frac{\epsilon}{\delta}\frac{\epsilon - 1}{\epsilon}(1 + \tau^L) N z_1 u''(c_1)\right) g_1 a_1 + \sum_{i \geq 1} \phi_i a_i u'(c_i) + \sum_{i \geq 1} g_i a_i \frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}.$$

We see here nicely how we need a boundary correction at the borrowing constraint. For inflation, we have

$$0 = -\delta \pi^w - \delta \pi^w \boldsymbol{\phi}' \boldsymbol{1} d\boldsymbol{x} + \frac{1 - e^{\rho dt}}{dt}\theta + \rho \theta$$

where we used the optimality condition for monetary policy to drop terms. Finally, the optimality condition for aggregate labor, i.e., aggregate economic activity, is given by

$$0 = \left[\left(u'(c_1) + \mu - \frac{1}{da}\boldsymbol{D}_{a,[1,:]}\boldsymbol{\lambda} + \theta \frac{\epsilon}{\delta}\frac{\epsilon - 1}{\epsilon}(1 + \tau^L) N z_1 u''(c_1)\right) g_1 z_1 + \sum_{i \geq 1} \phi_i z_i u'(c_i) + \sum_{i \geq 1} g_i z_i \frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}\right] d\boldsymbol{x}$$
$$- v'(N) - v'(N)\boldsymbol{\phi}'\boldsymbol{1} d\boldsymbol{x} - \mu + \theta \frac{\epsilon}{\delta}\left(\frac{\epsilon - 1}{\epsilon}(1 + \tau^L)(\boldsymbol{z} \cdot u'(\boldsymbol{c}))' \boldsymbol{g} d\boldsymbol{x} - v'(N)\right) - \frac{\epsilon}{\delta}\theta v''(N) N.$$

This representation follows from setting all equilibrium objects to constants, e.g., $\theta_n = \theta$. This Lemma states necessary conditions that any stationary Ramsey plan must satisfy. It does not speak to convergence to a stationary Ramsey plan. Crucially, this discretized representation of the Ramsey plan provides a formal treatment of boundary conditions. We see exactly how the planner takes into account the borrowing constraint that households face. And we see exactly where the corresponding boundary terms enter the optimality conditions and targeting rules for optimal monetary policy.

From the stationary counterpart of equation (30), it immediately follows that there is a key necessary condition for the existence of a stationary Ramsey plan, given by $\iint \partial_a \chi_{\text{ss}}(a, z)\, da\, dz = 0$. We highlight this condition because it has an important economic interpretation in the context of equation (30). It implies that the "births" and "deaths" of distributional penalties must average out to zero in a stationary Ramsey plan. In additional derivations available upon request, we show that this condition is satisfied in our baseline HANK model. This result has the interpretation that a planner does not want to over- or under-promise in the aggregate in terms of lifetime utilities.

Proof of Proposition 5. Note that in any stationary equilibrium we must have $\boldsymbol{\phi}'\mathbf{1}d\boldsymbol{x} = 1$, which we show below. Now notice that

$$\frac{1-e^{\rho dt}}{dt}\theta + \rho\theta \to \frac{1-1-\rho dt}{dt}\theta + \rho\theta = 0$$

in the limit as $dt \to 0$. Therefore, we must have $\pi^w = \pi = 0$ at the stationary Ramsey plan.

B.6 The Timeless Ramsey Problem in Dual Form

The timeless Ramsey problem also admits a dual representation, which we introduce next. The distinction between the primal and dual problems lies in the treatment of the constraints that a planner faces. In the primal approach, the planner optimizes over allocations, prices, and instruments given a set of constraints or implementability conditions. In the dual approach, the planner explicitly optimizes over the policy instrument, in this case, interest rates, using the implementability conditions to characterize the comparative statics of endogenous variables to policy.[50] The primal and dual representations of the timeless Ramsey problem have their distinct advantages and we leverage both in our analysis.

Definition. (**Timeless Dual Ramsey Problem**) *A timeless dual Ramsey problem solves*

$$\max_{\{i_t\}} L^{\mathrm{TD}}(g_{ss}, \phi_{ss}, \theta_{ss}),$$

where $L^{TD}(g_{ss}, \phi_{ss}, \theta_{ss})$ denotes the timeless dual Lagrangian, given an initial distribution g_{ss} as well as initial promises ϕ_{ss} and θ_{ss}. The Lagrangian is defined as

$$L^{\mathrm{TD}}(g_{ss}, \phi_{ss}, \theta_{ss}) = \int_0^\infty e^{-\rho t} \iint \left[u(c_t(a,z)) - v(N_t) - \frac{\delta}{2}(\pi_t^w)^2\right] g_t(a,z)\, da\, dz\, dt + \mathcal{T}(\phi_{ss}, \theta_{ss}), \quad (62)$$

where all endogenous variables are understood as functions of the policy path $\{i_t\}$.

Proposition 21. (**Timeless Ramsey Problem Resolves Time-0 Problem under Primal and Dual**) *Optimal policy under the timeless primal and dual Lagrangians resolves the time-0 problem. That is,*

$$\frac{d}{d\boldsymbol{i}} L^{\mathrm{TD}}\Big(g_{ss}, \phi_{ss}, \theta_{ss}, \boldsymbol{i}_{ss}, \boldsymbol{Z}_{ss}\Big) = \frac{d}{d\boldsymbol{i}} L^{\mathrm{TP}}\Big(g_{ss}, \phi_{ss}, \theta_{ss}, \boldsymbol{i}_{ss}, \boldsymbol{Z}_{ss}\Big) = 0. \quad (63)$$

[50] In simple terms, a useful analogy may be to interpret the dual approach as substituting constraints into the objective of an optimization problem, and the primal approach as accounting for constraints as additional terms in a Lagrangian.

Proposition 21 is a slightly more general variant of Proposition 6 in the main text. In the main text, we exclusively discuss the primal form of the Ramsey problem. In this Appendix, we leverage the duality between primal and dual because some results and proofs are easier to derive in one or the other. Proposition 21 establishes that both the primal and dual timeless Ramsey problems resolve the time-0 problem. We now prove this result.

B.7 Proof of Proposition 6

Our goal is to show that

$$\frac{dL^{\text{TP}}(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})}{d\boldsymbol{i}} = F\Big(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}\Big) = 0. \tag{64}$$

We proceed as follows: First, instead of working with the timeless primal Lagrangian, we leverage the observation that $\frac{d}{d\boldsymbol{i}}L^{\text{TP}} = 0$ if and only if the analogous perturbation for the timeless *dual* Lagrangian is 0, i.e., $\frac{d}{d\boldsymbol{i}}L^{\text{TD}} = 0$. We make this point in our discussion of the dual approach below, noting the generic duality between primal and dual representations of the Ramsey problem. Second, we will prove that this perturbation is 0 for a given $\frac{d}{di_k}$, i.e., for a one-time perturbation in the interest rate at time k. We can then simply "stack" up to arrive at any perturbation $\frac{d}{d\boldsymbol{i}}$.

For our baseline HANK model, $\frac{dL^{\text{TD}}}{di_k}$ takes the form

$$0 = \frac{d}{di_k}\left\{ \sum_{n=0}^{\infty} e^{-\rho t}\left\{ u\begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1}-A_n}{dtA_n}a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_n - v(N_n)\mathbf{1}'\boldsymbol{g}_n - \frac{\delta}{2}(\pi_n^w)^2\mathbf{1}'\boldsymbol{g}_n \right\} d\boldsymbol{x}dt \right.$$
$$\left.\left. \underbrace{+\frac{1}{dt}e^{\rho dt}\boldsymbol{\phi}'\boldsymbol{V}_0 d\boldsymbol{x} - \frac{1}{dt}e^{\rho dt}\theta\pi_0^w}_{\text{Timeless Penalties}} \right\}\right|_{\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}}$$

for all $k \geq 0$. We start by evaluating the derivative for any arbitrary set of inputs to $F(\cdot)$. This yields

$$0 = \sum_{n=0}^{\infty} e^{-\rho t}\left\{ \left[u(\boldsymbol{c}_n) - v(N_n) - \frac{\delta}{2}(\pi_n^w)^2\right]' \frac{d\boldsymbol{g}_t}{di_k} + (\boldsymbol{g}_n \cdot u'(\boldsymbol{c}_n))' \frac{d\boldsymbol{c}_n}{di_k} - (v'(N_n)\mathbf{1} + \delta\pi_n^w\mathbf{1})'\boldsymbol{g}_n \frac{dN_n}{di_k} \right\} dt$$
$$+ \frac{1}{dt}e^{\rho dt}\boldsymbol{\phi}'\frac{d\boldsymbol{V}_0}{di_k} - \frac{1}{dt}e^{\rho dt}\theta\frac{d\pi_0^w}{di_k}\frac{1}{d\boldsymbol{x}},$$

where we note that we always have $\frac{dc_{1,n}}{di_k} = 0$ because the planner is constrained by the same boundary condition that the household faces when considering policy perturbations.

Our proof strategy will be to add five sets of auxilliary terms to this equation, each of which evaluates to 0, and then use these additional terms to rearrange. In particular, the expressions we add correspond to the discretized competitive equilibrium conditions. And our goal will be to then evaluate the corresponding expression at the stationary equilibrium, group terms, and show that everything evaluates to 0.

Equation 1. We have

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \boldsymbol{\phi}' \left[-\rho \boldsymbol{V}_n + \frac{\boldsymbol{V}_{n+1} - \boldsymbol{V}_n}{dt} + u(\boldsymbol{c}_n) - v(N_n) - \frac{\delta}{2}(\pi_n^w)^2 + \boldsymbol{A}^z \boldsymbol{V}_n + \frac{1}{da}(\boldsymbol{s}_n \cdot \boldsymbol{D}_a \boldsymbol{V}_n) \right],$$

where we use $\boldsymbol{\phi} = \boldsymbol{\phi}_{\text{ss}}$. We now use auxilliary results and derivations from before to rewrite this equation as

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \boldsymbol{\phi}' \left[u(\boldsymbol{c}_n) - v(N_n) - \frac{\delta}{2}(\pi_n^w)^2 + \boldsymbol{A}^z \boldsymbol{V}_n + \frac{1}{da}(\boldsymbol{s}_n \cdot \boldsymbol{D}_a \boldsymbol{V}_n) \right] - e^{\rho dt} \boldsymbol{\phi}' \frac{1}{dt} \boldsymbol{V}_0.$$

Differentiating, we obtain

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \left[(\boldsymbol{\phi} \cdot u'(\boldsymbol{c}_n)) \frac{d\boldsymbol{c}_n}{di_k} - \boldsymbol{\phi}' \mathbf{1} \left(v'(N_t) \frac{dN_n}{di_k} + \delta \pi_n^w \frac{d\pi_n^w}{di_k} \right) \right.$$
$$\left. + \boldsymbol{\phi}' \boldsymbol{A}^z \frac{d\boldsymbol{V}_n}{di_k} + \frac{1}{da} (\boldsymbol{\phi} \cdot \boldsymbol{D}_a \boldsymbol{V}_n)' \frac{d\boldsymbol{s}_n}{di_k} + \frac{1}{da} \boldsymbol{\phi}' \boldsymbol{s}_n \cdot \boldsymbol{D}_a \frac{d\boldsymbol{V}_n}{di_k} \right] - e^{\rho dt} \boldsymbol{\phi}' \frac{1}{dt} \frac{d\boldsymbol{V}_0}{di_k}.$$

This is the first auxilliary equation that we will add to our desired expression.

Equation 2. We obtain the second auxilliary condition by simply differentiating the consumption first-order condition. We rewrite the equation as

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \left[\boldsymbol{\chi}_{[2:J]}' u''(\boldsymbol{c}_{n,[2:J]}) - e^{\rho dt} \frac{1}{da} \boldsymbol{V}_n' \boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix} \right] + e^{\rho dt} \frac{1}{da} \boldsymbol{V}_0' \boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix},$$

where we use $\boldsymbol{\chi} = \boldsymbol{\chi}_{\text{ss}}$, and then differentiate to obtain

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \left[\left(\boldsymbol{\chi}_{[2:J]} \cdot u''(\boldsymbol{c}_{n,[2:J]}) \right)' \frac{d\boldsymbol{c}_{n,[2:J]}}{di_k} - e^{\rho dt} \frac{1}{da} \left[\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix} \right]' \frac{d\boldsymbol{V}_n}{di_k} \right] + e^{\rho dt} \frac{1}{da} \left[\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix} \right]' \frac{d\boldsymbol{V}_0}{di_k}.$$

Equation 3. For our third auxilliary equation, we differentiate the discretized Kolmogorov forward equation. From before, we have

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \left[-\rho \boldsymbol{\lambda}' \boldsymbol{g}_n + \boldsymbol{\lambda}' (\boldsymbol{A}^z)' \boldsymbol{g}_n - \frac{1}{da} (\boldsymbol{s}_n \cdot \boldsymbol{g}_n)' \boldsymbol{D}_a' \boldsymbol{\lambda} \right] + \frac{1}{dt} e^{\rho dt} \boldsymbol{\lambda}' \boldsymbol{g}_0.$$

Differentiating with respect to i_k, we obtain

$$0 = = \sum_{n=0}^{\infty} e^{-\rho t_n} \left[-\rho \boldsymbol{\lambda}' \frac{d\boldsymbol{g}_n}{di_k} + \boldsymbol{\lambda}'(\boldsymbol{A}^z)' \frac{d\boldsymbol{g}_n}{di_k} + (\boldsymbol{g}_n \cdot \boldsymbol{D}_a \boldsymbol{\lambda})' \frac{d\boldsymbol{s}_n}{di_k} + (\boldsymbol{s}_n \cdot \boldsymbol{D}_a \boldsymbol{\lambda})' \frac{d\boldsymbol{g}_n}{di_k} \right] + \frac{1}{dt} e^{\rho dt} \boldsymbol{\lambda}' \frac{d\boldsymbol{g}_0}{di_k},$$

where we again use $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{\text{ss}}$.

Equation 4. We have the aggregate resource constraint with

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \mu \left[\frac{1}{d\boldsymbol{x}} A_n N_n - \boldsymbol{c}_n' \boldsymbol{g}_n \right],$$

where we use $\mu = \mu_{\text{ss}}$. Differentiating, we have

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \mu_{ss} \left[\frac{1}{d\boldsymbol{x}} A_n \frac{dN_n}{di_k} - \boldsymbol{c}_n' \frac{d\boldsymbol{g}_n}{di_k} - \boldsymbol{g}_n' \frac{d\boldsymbol{c}_n}{di_k} \right].$$

Equation 5. And finally, we use the Phillips curve, which we rewrite using previous results as

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \theta_{ss} \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right) N_n + \frac{1}{dt} e^{\rho dt} \theta \pi_0^w.$$

Differentiating, we obtain

$$0 = \sum_{n=0}^{\infty} e^{-\rho t_n} \theta_{ss} \frac{\epsilon}{\delta} \left[\left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n \left((\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \frac{d\boldsymbol{g}_n}{di_k} + (\boldsymbol{z} \cdot u''(\boldsymbol{c}_n) \cdot \boldsymbol{g}_n)' \frac{d\boldsymbol{c}_n}{di_k} \right) d\boldsymbol{x} - v''(N_n) \frac{dN_n}{di_k} \right) N_n \right.$$
$$\left. + \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right) \frac{dN_n}{di_k} + \frac{1}{dt} e^{\rho dt} \theta \frac{d\pi_0^w}{di_k}. \right.$$

Evaluate at stationary Ramsey plan. Crucially, each of our five auxilliary equations must necessarily also hold when evaluated around the stationary Ramsey plan. The key step now, is to evaluate each of the first-order derivatives we taken at the stationary Ramsey plan.

Putting everything together. Having evaluated all derivatives around the stationary Ramsey plan, we add the five auxilliary equations we have derived to the expression for $\frac{dL^{\text{TD}}}{di_k}$ which we started out with, where we now also evaluate the latter at the stationary Ramsey plan. This yields

$$
\begin{aligned}
0 = \sum_{n=0}^{\infty} e^{-\rho t} \Bigg\{ & \left[u(\boldsymbol{c}) - v(N) - \frac{\delta}{2}(\pi^w)^2 \right]' \frac{d\boldsymbol{g}_t}{di_k} + (\boldsymbol{g} \cdot u'(\boldsymbol{c}))' \frac{d\boldsymbol{c}_n}{di_k} - (v'(N)\mathbf{1} + \delta \pi^w \mathbf{1})' \boldsymbol{g} \frac{dN_n}{di_k} \\
& + (\boldsymbol{\phi} \cdot u'(\boldsymbol{c})) \frac{d\boldsymbol{c}_n}{di_k} - \boldsymbol{\phi}' \mathbf{1} \left(v'(N) \frac{dN_n}{di_k} + \delta \pi_n^w \frac{d\pi_n^w}{di_k} \right) \\
& + \boldsymbol{\phi}' \boldsymbol{A}^z \frac{d\boldsymbol{V}_n}{di_k} + \frac{1}{da} (\boldsymbol{\phi} \cdot \boldsymbol{D}_a \boldsymbol{V})' \frac{d\boldsymbol{s}_n}{di_k} + \frac{1}{da} \boldsymbol{\phi}' \boldsymbol{s} \cdot \boldsymbol{D}_a \frac{d\boldsymbol{V}_n}{di_k} \\
& + \left(\boldsymbol{\chi}_{[2:J]} \cdot u''(\boldsymbol{c}_{[2:J]}) \right)' \frac{d\boldsymbol{c}_{n,[2:J]}}{di_k} - e^{\rho dt} \frac{1}{da} \left[\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix} \right]' \frac{d\boldsymbol{V}_n}{di_k} \\
& - \rho \boldsymbol{\lambda}' \frac{d\boldsymbol{g}_n}{di_k} + \boldsymbol{\lambda}' (\boldsymbol{A}^z)' \frac{d\boldsymbol{g}_n}{di_k} + (\boldsymbol{g} \cdot \boldsymbol{D}_a \boldsymbol{\lambda})' \frac{d\boldsymbol{s}_n}{di_k} + (\boldsymbol{s} \cdot \boldsymbol{D}_a \boldsymbol{\lambda})' \frac{d\boldsymbol{g}_n}{di_k} \\
& + \mu \frac{1}{d\boldsymbol{x}} A \frac{dN_n}{di_k} - \mu \boldsymbol{c}' \frac{d\boldsymbol{g}_n}{di_k} - \mu \boldsymbol{g}' \frac{d\boldsymbol{c}_n}{di_k} \\
& + \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A \left((\boldsymbol{z} \cdot u'(\boldsymbol{c}))' \frac{d\boldsymbol{g}_n}{di_k} + (\boldsymbol{z} \cdot u''(\boldsymbol{c}) \cdot \boldsymbol{g})' \frac{d\boldsymbol{c}_n}{di_k} \right) d\boldsymbol{x} - v''(N) \frac{dN_n}{di_k} \right) N_n \\
& + \theta \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A (\boldsymbol{z} \cdot u'(\boldsymbol{c}))' \boldsymbol{g} d\boldsymbol{x} - v'(N) \right) \frac{dN_n}{di_k} \Bigg\} dt \\
& + e^{\rho dt} \boldsymbol{\phi}' \frac{d\boldsymbol{V}_0}{di_k} - e^{\rho dt} \theta \frac{d\pi_0^w}{di_k} \frac{1}{d\boldsymbol{x}} \\
& - e^{\rho dt} \boldsymbol{\phi}' \frac{d\boldsymbol{V}_0}{di_k} + e^{\rho dt} \boldsymbol{\lambda}' \frac{d\boldsymbol{g}_0}{di_k} + e^{\rho dt} \theta \frac{d\pi_0^w}{di_k} \frac{1}{d\boldsymbol{x}} \\
& + dt e^{\rho dt} \frac{1}{da} \left[\boldsymbol{D}_a' \begin{pmatrix} 0 \\ \boldsymbol{\chi}_{[2:J]} \end{pmatrix} \right]' \frac{d\boldsymbol{V}_0}{di_k},
\end{aligned}
$$

where every term that does not have a time step subscript n is understood to have been evaluated at the stationary Ramsey plan.

Our proof is now almost complete. First, note how the timeless penalties *exactly offset* the "boundary terms" that resulted from rearranging the forward looking implementability conditions. In particular, notice that $\frac{dg_0}{di_k} = 0$ and the term in the very last line goes to 0 as $dt \to 0$. The remaining boundary (or initial condition) terms exactly cancel out.

Second, we plug in for

$$\frac{d\boldsymbol{s}_n}{di_k} = \frac{dr_n}{di_k} a + zw \frac{dN_n}{di_k} + zN \frac{dw_n}{di_k} - \frac{d\boldsymbol{c}_n}{di_k}$$

when evaluated at the stationary Ramsey plan.

Third and finally, we group all terms by *derivatives*. After this last step, we see that the grouped expressions correspond *exactly* to the optimality conditions that define the stationary Ramsey plan. Consequently, they must be 0. This concludes the proof: We started with an expression for $\frac{dL^{\mathrm{TD}}}{d\theta_k}$, and added five auxilliary expressions, each of which itself evaluated to 0. Then we evaluated the resulting expression around the stationary Ramsey plan and showed that it was 0. Consequently, we have shown that

$$\frac{dL^{\mathrm{TD}}}{di_k} = 0$$

when evaluated at the stationary Ramsey plan. And since k was arbitrary, we have our desired result for any policy perturbation around the stationary Ramsey plan. We have thus shown that Ramsey policy according to the timeless dual Lagrangian L^{TD}—and consequently also the timeless primal Lagrangian L^{TP}—indeed resolves the time-0 problem. This proof demonstrates that our timeless Ramsey approach formalizes Woodford (1999)'s timeless perspective in our HANK economy.

B.8 Proof of Proposition 7

From equation (33), we have that

$$\begin{aligned} 0 = & \iint z\partial_a \lambda_t(a,z)\, g_t(a,z)\, dadz + \underline{z}\xi_t^{HTM} g_t(\underline{a},\underline{z})\, dadz - \mu_t - \frac{v'(N_t)}{A_t} \\ & + \iint \phi_t(a,z)\left(zu'(c_t(a,z)) - \frac{v'(N_t)}{A_t}\right) g_t(a,z)\, dadz + \theta_t \frac{\epsilon_t}{\delta}\frac{1}{A_t}\iint z\frac{d\tau_t(a,z)}{dn_t(a,z)} g_t(a,z)\, dadz. \end{aligned}$$

Combining equation (32) and the definition of ξ_t^{HTM}, we have

$$\begin{aligned} 0 = & \iint zu'(c_t(a,z))\, g_t(a,z)\, dadz - \iint z\partial_a\lambda_t(a,z)\, g_t(a,z)\, dadz + \iint z\mu_t g_t(a,z)\, dadz \\ & + \theta_t\frac{\epsilon_t}{\delta}\iint z\frac{d\tau_t^L(a,z)}{dc_t(a,z)} g_t(a,z)\, dadz - \iint z\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz - \underline{z}\xi_t^{HTM} g_t(\underline{a},\underline{z}), \end{aligned}$$

where $\iint z\mu_t g_t(a,z)\, da\, dz = \mu_t$.

Combining both of these equations, we obtain

$$\begin{aligned} 0 = & \iint \left(zu'(c_t(a,z)) - \frac{v'(n_t(a,z))}{A_t}\right)\left(1 + \frac{\phi_t(a,z)}{g_t(a,z)}\right) g_t(a,z)\, dadz \\ & + \theta_t\frac{\epsilon_t}{\delta}\iint z\left(\frac{d\tau_t^L(a,z)}{dc_t(a,z)} + \frac{1}{A_t}\frac{d\tau_t^L(a,z)}{dn_t(a,z)}\right) g_t(a,z)\, dadz - \iint z\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz. \end{aligned}$$

Solving for θ_{ss} after imposing that all variables have reached the steady state immediately recovers equation (46) in the text.

B.9 Proof of Proposition 8

The stationary version of the promise-keeping Kolmogorov forward equation (47) follows trivially from the time-varying equation (30), setting $\partial_t \phi_t(a,z) = 0$ and evaluating the RHS at the stationary Ramsey plan.

B.10 Proof of Proposition 9

The optimality conditions of the Ramsey problem imply that

$$\iint a\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz = \iint au'(c_t(a,z))\, g_t(a,z)\, dadz + \theta_t \frac{\epsilon_t}{\delta} \iint a \frac{d\tau_t^L(a,z)}{dc_t(a,z)} g_t(a,z)\, dadz$$
$$- \iint a\partial_a \lambda_t(a,z)\, g_t(a,z)\, dadz - \underline{a}\zeta_t^{HTM} g_t(\underline{a},\underline{z})$$

and

$$\iint z\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz = \mu_t + \iint zu'(c_t(a,z))\, g_t(a,z)\, dadz + \theta_t \frac{\epsilon_t}{\delta} \iint z \frac{d\tau_t^L(a,z)}{dc_t(a,z)} g_t(a,z)\, dadz$$
$$- \iint z\partial_a \lambda_t(a,z)\, g_t(a,z)\, dadz - \underline{z}\zeta_t^{HTM} g_t(\underline{a},\underline{z})$$

Combining these equations with FOCs (33) and (35), we can write, respectively

$$\iint z\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz = \iint \left(zu'(c_t(a,z)) - \frac{v'(n_t(a,z))}{A_t}\right)\left(1 + \frac{\phi_t(a,z)}{g_t(a,z)}\right) g_t(a,z)\, dadz$$
$$+ \theta_t \frac{\epsilon_t}{\delta} \iint z \left(\frac{d\tau_t(a,z)}{dc_t(a,z)} + \frac{1}{A_t}\frac{d\tau_t(a,z)}{dn_t(a,z)}\right) g_t(a,z)\, dadz$$

and

$$\iint a\tilde{\chi}_t(a,z)\, g_t(a,z)\, dadz = \iint a\left(1 + \frac{\phi_t(a,z)}{g_t(a,z)}\right) u'(c_t(a,z))\, g_t(a,z)\, dadz$$
$$+ \theta_t \frac{\epsilon_t}{\delta} \iint a \frac{d\tau_t(a,z)}{dc_t(a,z)} g_t(a,z)\, dadz$$

Finally, expressing $\tilde{\chi}_t(a,z)$ as $\tilde{\chi}_t(a,z) = (1 - \mathcal{M}_t(a,z))\left(\mu_t + \theta_t \frac{\epsilon_t}{\delta}\frac{d\tau_t(a,z)}{dc_t(a,z)}\right)$, we can write

$$\mu_t = \frac{\iint \left(zu'(c_t(a,z)) - \frac{v'(n_t(a,z))}{A_t}\right)\left(1 + \frac{\phi_t(a,z)}{g_t(a,z)}\right) g_t(a,z)\, dadz + \theta_t \frac{\epsilon_t}{\delta} \iint z\left(\frac{1}{A_t}\frac{d\tau_t(a,z)}{dn_t(a,z)} + \mathcal{M}_t(a,z)\frac{d\tau_t(a,z)}{dc_t(a,z)}\right) g_t(a,z)\, dadz}{\iint z(1 - \mathcal{M}_t(a,z))\, g_t(a,z)\, dadz}$$

and

$$\mu_t = \frac{\iint a\left(1+\frac{\phi_t(a,z)}{g_t(a,z)}\right) u'\left(c_t\left(a,z\right)\right) g_t\left(a,z\right) dadz + \theta_t \frac{\epsilon_t}{\delta} \iint a \mathcal{M}_t\left(a,z\right) \frac{d\tau_t(a,z)}{dc_t(a,z)} g_t\left(a,z\right) dadz}{\iint a\left(1-\mathcal{M}_t\left(a,z\right)\right) g_t\left(a,z\right) dadz}$$

Equation (49) follows from combining these two conditions after collecting terms. The logic behind these perturbations is identical to the discretion case, explained in detail in the Proof of Proposition 2.

Output gap targeting rule under isoelastic preferences. Under isoelastic preferences, the targeting rule admits the alternative representation

$$\begin{aligned} Y_t = \tilde{Y}_t \times \left(\frac{\epsilon_t}{\epsilon_t - 1}\frac{1}{1+\tau^L}\right)^{\frac{1}{\gamma+\eta}} \times \Bigg\{ 1 - \Omega_t^D \frac{\iint au'(c_t(a,z))g_t(a,z)\,da\,dz}{\iint zu'(c_t(a,z))g_t(a,z)\,da\,dz} \\ + \frac{\iint (z-\Omega_t^D)u'(c_t(a,z))\phi_t(a,z)\,da\,dz}{\iint zu'(c_t(a,z))g_t(a,z)\,da\,dz} \\ + \theta_t \frac{\epsilon_t}{\delta} \frac{\iint \left(\frac{z}{A_t}\frac{d\tau_t(a,z)}{dN_t} + (z-\Omega_t^D a)\mathcal{M}_t(a,z)\frac{d\tau_t(a,z)}{dc_t(a,z)}\right) g_t(a,z)\,da\,dz}{\iint zu'(c_t(a,z))g_t(a,z)\,da\,dz} \Bigg\}^{\frac{1}{\gamma+\eta}} \end{aligned} \tag{65}$$

in terms of output gaps, where $\tilde{Y}_t$ is natural output in HANK.

C Appendix for Section 3

C.1 Proof of Proposition 1

Having introduced the discretization of the implementability conditions and the Ramsey problem, it is now straightforward to derive the optimality conditions for policy under discretion. We again discretize the planning problem in both individual state variables (a, z) and in time. In particular, we discretize the time dimension over a finite horizon, $t \in [0, T]$, where T can be arbitrarily large, using N discrete time steps, which we denote by $n \in \{1, \dots, N\}$. With a step size $dt = \frac{T}{N-1}$, we have $t_n = dt(n-1)$. We assume that a planner under discretion controls policy at time step s and takes as given policy from $s+1$ onwards.

The planning problem under discretion at time s is associated with the Lagrangian

$$
\begin{aligned}
L^{\mathrm{D}}(\boldsymbol{g}_s) = \min_{\boldsymbol{\phi}_s, \boldsymbol{\chi}_s, \boldsymbol{\lambda}_s, \mu_s, \vartheta_s} \max_{\boldsymbol{V}_s, \boldsymbol{c}_{s,[2:J]}, \boldsymbol{g}_{s+1}, \pi^w_s, N_s, i_s} \sum_{n=s}^{N-1} e^{-\rho t_n} \Bigg\{ \Bigg\{ & \\
+ u \begin{pmatrix} i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_t - v(N_n) \mathbf{1}' \boldsymbol{g}_t - \frac{\delta}{2} (\pi^w_n)^2 \mathbf{1}' \boldsymbol{g}_t & \\
+ \boldsymbol{\phi}'_n \left[-\rho \boldsymbol{V}_n + \frac{\boldsymbol{V}_{n+1} - \boldsymbol{V}_n}{dt} + u \begin{pmatrix} i_n a_1 - \pi^w_n a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} - v(N_n) - \frac{\delta}{2} (\pi^w_n)^2 \right] & \\
+ \boldsymbol{\phi}'_n \boldsymbol{A}^z \boldsymbol{V}_n + \sum_{i \geq 2} \phi_{i,n} \left(i_n a_i - \pi^w_n a_i + \frac{A_{n+1} - A_n}{dt A_n} a_i + z_i A_n N_n - c_{n,i} \right) \frac{\boldsymbol{D}_{a,[i,:]}}{da} \boldsymbol{V}_n & \\
+ \boldsymbol{\chi}'_{n,[2:J]} \left[u'(\boldsymbol{c}_{n,[2:J]}) - \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{n+1} \right)_{[2:J]} \right] & \\
- \boldsymbol{\lambda}'_n \frac{\boldsymbol{g}_{n+1} - \boldsymbol{g}_n}{dt} + \boldsymbol{\lambda}'_n (\boldsymbol{A}^z)' \boldsymbol{g}_n & \\
+ \sum_{i \geq 2} \lambda_{n,i} \frac{\boldsymbol{D}'_{a,[i,:]}}{da} \left[\begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi^w_n \boldsymbol{a}_{[2:J]} + \frac{A_{n+1} - A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \boldsymbol{g}_n \right] \Bigg\} d\boldsymbol{x} & \\
+ \mu_n \left[\boldsymbol{c}'_n \boldsymbol{g}_n d\boldsymbol{x} - A_n N_n \right] & \\
+ \theta_n \left[-\frac{\pi^w_{n+1} - \pi^w_n}{dt} + \rho \pi^w_n + \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right) N_n \right] \Bigg\} dt, &
\end{aligned}
$$

where the superscript D denotes the planning problem under discretion. The planner takes as given an initial condition for the cross-sectional distribution, $\boldsymbol{g}_s$.

Unlike in the Ramsey problem with commitment, we only sum (integrate) by parts the *state* *variables* of the problem, and not those terms associated with forward-looking constraints. That is, we use

$$-\sum_{n=s}^{N-1} e^{-\rho t_n} \boldsymbol{\lambda}_n' \frac{\boldsymbol{g}_{n+1} - \boldsymbol{g}_n}{dt} = \sum_{n=s}^{N-1} e^{-\rho t_n} \frac{\boldsymbol{\lambda}_n' - e^{\rho dt} \boldsymbol{\lambda}_{n-1}'}{dt} \boldsymbol{g}_n + \frac{1}{dt} e^{\rho dt} \boldsymbol{\lambda}_{s-1}' \boldsymbol{g}_s - \frac{1}{dt} e^{\rho dt} e^{-\rho t_N} \boldsymbol{\lambda}_{N-1}' \boldsymbol{g}_N$$

and rewrite the Lagrangian as

$$\begin{aligned}
L^{\mathrm{D}}(\boldsymbol{g}_s) = \min_{\boldsymbol{\phi}_s, \boldsymbol{\chi}_s, \boldsymbol{\lambda}_s, \mu_s, \vartheta_s} \max_{\boldsymbol{V}_s, \boldsymbol{c}_{s,[2:J]}, \boldsymbol{g}_{s+1}, \pi_s^w, N_s, i_s} \sum_{n=s}^{N-1} e^{-\rho t_n} \Bigg\{ \Bigg\{ \\
&+ u \begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix}' \boldsymbol{g}_t - v(N_n) \mathbf{1}' \boldsymbol{g}_t - \frac{\delta}{2} (\pi_n^w)^2 \mathbf{1}' \boldsymbol{g}_t \\
&+ \boldsymbol{\phi}_n' \left[-\rho \boldsymbol{V}_n + \frac{\boldsymbol{V}_{n+1} - \boldsymbol{V}_n}{dt} + u \begin{pmatrix} i_n a_1 - \pi_n^w a_1 + \frac{A_{n+1} - A_n}{dt A_n} a_1 + z_1 A_n N_n \\ \boldsymbol{c}_{n,[2:J]} \end{pmatrix} - v(N_n) - \frac{\delta}{2} (\pi_n^w)^2 \right] \\
&+ \boldsymbol{\phi}_n' \boldsymbol{A}^z \boldsymbol{V}_n + \sum_{i \geq 2} \phi_{i,n} \left(i_n a_i - \pi_n^w a_i + \frac{A_{n+1} - A_n}{dt A_n} a_i + z_i A_n N_n - c_{n,i} \right) \frac{\boldsymbol{D}_{a,[i,:]}}{da} \boldsymbol{V}_n \\
&+ \boldsymbol{\chi}_{n,[2:J]}' \left[u'(\boldsymbol{c}_{n,[2:J]}) - \left(\frac{\boldsymbol{D}_a}{da} \boldsymbol{V}_{n+1} \right)_{[2:J]} \right] \\
&+ e^{-\rho t_n} \frac{\boldsymbol{\lambda}_n' - e^{\rho dt} \boldsymbol{\lambda}_{n-1}'}{dt} \boldsymbol{g}_n + \boldsymbol{\lambda}_n' (\boldsymbol{A}^z)' \boldsymbol{g}_n \\
&+ \sum_{i \geq 2} \lambda_{n,i} \frac{\boldsymbol{D}_{a,[i,:]}'}{da} \left[\begin{pmatrix} 0 \\ i_n \boldsymbol{a}_{[2:J]} - \pi_n^w \boldsymbol{a}_{[2:J]} + \frac{A_{n+1} - A_n}{dt A_n} \boldsymbol{a}_{[2:J]} + \boldsymbol{z}_{[2:J]} A_n N_n - \boldsymbol{c}_{n,[2:J]} \end{pmatrix} \cdot \boldsymbol{g}_n \right] \Bigg\} d\boldsymbol{x} \\
&+ \mu_n \left[\boldsymbol{c}_n' \boldsymbol{g}_n d\boldsymbol{x} - A_n N_n \right] \\
&+ \theta_n \left[-\frac{\pi_{n+1}^w - \pi_n^w}{dt} + \rho \pi_n^w + \frac{\epsilon}{\delta} \left(\frac{\epsilon - 1}{\epsilon} (1 + \tau^L) A_n (\boldsymbol{z} \cdot u'(\boldsymbol{c}_n))' \boldsymbol{g}_n d\boldsymbol{x} - v'(N_n) \right) N_n \right] \Bigg\} dt \\
&+ e^{\rho dt} \boldsymbol{\lambda}_{s-1}' \boldsymbol{g}_s d\boldsymbol{x} - e^{\rho dt} e^{-\rho t_N} \boldsymbol{\lambda}_{N-1}' \boldsymbol{g}_N d\boldsymbol{x}.
\end{aligned}$$

Crucially, the Markov planner at time step s does realize that her policy decisions affect the evolution of state variables, i.e., the distribution $\boldsymbol{g}_{s+1}$ that the "future planner" at time step $s+1$ takes as her initial condition. She does not internalize, however, that her policy decisions also determine the terminal conditions on forward-looking equations, i.e., inflation and lifetime values, that "past planners" take as given.

We now characterize the first-order optimality conditions associated with the planning problem under discretion.

Derivative for $\boldsymbol{V}_s$. We have

$$0 = -\rho\boldsymbol{\phi}_s - \frac{1}{dt}\boldsymbol{\phi}_s + (\boldsymbol{A}^z)'\boldsymbol{\phi}_s + \frac{1}{da}(\boldsymbol{\phi}_s\boldsymbol{D}_a)'\boldsymbol{s}_s - e^{\rho dt}\frac{\boldsymbol{D}_a'}{da}\begin{pmatrix} 0 \\ \boldsymbol{\chi}_{s-1,[2:J]} \end{pmatrix}$$

or simply

$$0 = -\rho\boldsymbol{\phi}_s - \frac{1}{dt}\boldsymbol{\phi}_s + \boldsymbol{A}'\boldsymbol{\phi}_s - e^{\rho dt}\frac{\boldsymbol{D}_a'}{da}\begin{pmatrix} 0 \\ \boldsymbol{\chi}_{s-1,[2:J]} \end{pmatrix}.$$

Consider the last term in this equation. The household's consumption FOC says that consumption today is a function of "expected" future value, which therefore uses $\boldsymbol{V}_{s+1}$. The planner under discretion takes the future value $\boldsymbol{V}_{s+1}$ as given. And the planner is constrained by the competitive equilibrium condition that households make consumption decisions *purely* in terms of $\boldsymbol{V}_{s+1}$. By the household's first-order condition, then, $\boldsymbol{c}_s$ is pinned down as a function of $\boldsymbol{V}_{s+1}$.

We now see from this that, in the continuous-time limit with $dt \to 0$, we must have

$$\boldsymbol{\phi}_s = 0.$$

This is the proper boundary condition for the formal continuous-time problem under discretion. Moreover, from the consumption FOC in the Lagrangian, we also have

$$0 = \frac{\boldsymbol{D}_a'}{da}\begin{pmatrix} 0 \\ \boldsymbol{\chi}_{s-1,[2:J]} \end{pmatrix}$$

for all s.

Derivative for $\boldsymbol{g}_{s+1}$. We have

$$\begin{aligned} 0 =& u(\boldsymbol{c}_{s+1}) + \mu_{s+1}\boldsymbol{c}_{s+1} - v(N_{s+1})\mathbf{1} - \frac{\delta}{2}(\pi^w_{s+1})^2\mathbf{1} + \frac{\boldsymbol{\lambda}_{s+1}' - e^{\rho dt}\boldsymbol{\lambda}_s'}{dt} + (\boldsymbol{\lambda}_{s+1}'(\boldsymbol{A}^z)')' \\ &+ \frac{d}{d\boldsymbol{g}_{s+1}}\left[\frac{1}{da}(\boldsymbol{D}_a\boldsymbol{\lambda}_{s+1})'[\boldsymbol{s}_{s+1}\cdot\boldsymbol{g}_{s+1}]\right] + \theta_{s+1}N_{s+1}\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_{s+1}\boldsymbol{z}\cdot u'(\boldsymbol{c}_{s+1}). \end{aligned}$$

Now we work out the remaining derivative,

$$\frac{d}{d\boldsymbol{g}_{s+1}}\left[\frac{1}{da}(\boldsymbol{D}_a\boldsymbol{\lambda}_{s+1})'[\boldsymbol{s}_{s+1}\cdot\boldsymbol{g}_{s+1}]\right] = \boldsymbol{A}^a\boldsymbol{\lambda}_{s+1}.$$

Thus, we have

$$
\begin{aligned}
0 =& u(\boldsymbol{c}_{s+1}) + \mu_{s+1}\boldsymbol{c}_{s+1} - v(N_{s+1})\mathbf{1} - \frac{\delta}{2}(\pi^w_{s+1})^2\mathbf{1} + \frac{\boldsymbol{\lambda}'_{s+1} - e^{\rho dt}\boldsymbol{\lambda}'_s}{dt} \\
&+ \boldsymbol{A}\boldsymbol{\lambda}_{s+1} + \theta_{s+1}N_{s+1}\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_{s+1}\boldsymbol{z}\cdot u'(\boldsymbol{c}_{s+1}).
\end{aligned}
$$

Derivative $\boldsymbol{c}_{s,[2:J]}$. Notice that the planner under commitment also just solves a static problem for consumption at every time step. In other words, the choice of consumption today doesn't "bind" the planner tomorrow in any way under commitment. Therefore, we again have

$$
\begin{aligned}
0 =& u'(c_{s,i})g_{s,i} + \mu_s g_{s,i} + u'(c_{s,i})\phi_{s,i} + u''(c_{s,i})\chi_{s,i} + \theta_s N_s\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_s z_i u''(c_{s,i})g_{s,i} \\
&- \frac{1}{da}\phi_{s,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{V}_s - \frac{1}{da}g_{s,i}\boldsymbol{D}_{a,[i,:]}\boldsymbol{\lambda}_s.
\end{aligned}
$$

Derivative π^w_n. We have

$$
\begin{aligned}
0 =& \Big[-u'(c_{s,1})g_{s,1}a_1 - \mu_s g_{s,1}a_1 - \delta\pi^w_s\mathbf{1}'\boldsymbol{g}_s - \phi_{s,1}u'(c_{s,1})a_1 - \delta\pi^w_s\boldsymbol{\phi}'_s\mathbf{1}\Big]d\boldsymbol{x} \\
&- \theta_s\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_s N_s z_1 u''(c_{s,1})g_{s,1}a_1 d\boldsymbol{x} \\
&+ \Big[-\sum_{i\geq 2}\phi_{i,s}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_s + \sum_{i\geq 2}\lambda_{s,i}\frac{\boldsymbol{D}'_{a,[i,:]}}{da}\Big[\begin{pmatrix} 0 \\ -\boldsymbol{a}_{[2:J]}\end{pmatrix}\cdot\boldsymbol{g}_s\Big]d\boldsymbol{x} \\
&+ \frac{1}{dt}\theta_s.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
0 =& \Big[-u'(c_{s,1})g_{s,1}a_1 - \mu_s g_{s,1}a_1 - \delta\pi^w_s\mathbf{1}'\boldsymbol{g}_s - \phi_{s,1}u'(c_{s,1})a_1 - \delta\pi^w_s\boldsymbol{\phi}'_s\mathbf{1}\Big]d\boldsymbol{x} \\
&- \theta_s\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_s N_s z_1 u''(c_{s,1})g_{s,1}a_1 d\boldsymbol{x} \\
&- \sum_{i\geq 2}\phi_{i,s}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_s d\boldsymbol{x} - \sum_{i\geq 2}g_{s,i}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_s d\boldsymbol{x} + \frac{1}{dt}\theta_s.
\end{aligned}
$$

Derivative i_n. The nominal interest rate derivative is parallel to that for wage inflation, except in the Phillips curve. In particular, the choice of the nominal interest rate is again a fundamentally static problem, even in the case with commitment. We have

$$0 = u'(c_{s,1})g_{s,1}a_1 + \mu_s g_{s,1}a_1 + \phi_{s,1}u'(c_{s,1})a_1 + \sum_{i\geq 2}\phi_{i,s}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_s + \sum_{i\geq 2}g_{s,i}a_i\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_s$$
$$+ \theta_s\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_sN_sz_1u''(c_{s,1})g_{s,1}a_1.$$

Derivative N_n. Finally, we take the derivative for aggregate labor. This is again a static problem. We have

$$0 = \left[u'(c_{s,1})g_{s,1}z_1A_s + \mu_s g_{s,1}z_1A_s + \phi_{s,1}u'(c_{s,1})z_1A_s + \sum_{i\geq 2}\phi_{i,s}z_iA_s\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{V}_s + \sum_{i\geq 2}g_{s,i}z_iA_s\frac{\boldsymbol{D}_{a,[i,:]}}{da}\boldsymbol{\lambda}_s\right]d\boldsymbol{x}$$
$$+ \theta_s\frac{\epsilon}{\delta}\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_sN_sz_1u''(c_{s,1})g_{s,1}z_1A_sd\boldsymbol{x}$$
$$- v'(N_s)\mathbf{1}'\boldsymbol{g}_sd\boldsymbol{x} - v'(N_s)\boldsymbol{\phi}_s'\mathbf{1}d\boldsymbol{x}$$
$$- \mu_sA_s + \theta_s\frac{\epsilon}{\delta}\left(\frac{\epsilon-1}{\epsilon}(1+\tau^L)A_s(\boldsymbol{z}\cdot u'(\boldsymbol{c}_s))'\boldsymbol{g}_sd\boldsymbol{x} - v'(N_s)\right) - \frac{\epsilon}{\delta}\theta_sv''(N_s)N_s.$$

We now summarize the resulting optimality conditions for the problem under discretion. We state these optimality conditions here for the fully discretized problem, which we have worked with thus far. For the main text, we bring these equations back to the continuous case.

We see immediately that

$$\theta_s = 0$$
$$\boldsymbol{\phi}_s = 0$$

because the planner does not respect promises from the past. These two conditions signify the lack of commitment. The optimality condition for the cross-sectional distribution still characterizes the evolution of the social lifetime value. Using $\theta_s = 0$, we have

$$0 = u(\boldsymbol{c}_{s+1}) + \mu_{s+1}\boldsymbol{c}_{s+1} - v(N_{s+1})\mathbf{1} - \frac{\delta}{2}(\pi^w_{s+1})^2\mathbf{1} + \frac{\boldsymbol{\lambda}'_{s+1} - e^{\rho dt}\boldsymbol{\lambda}'_s}{dt} + \boldsymbol{A}\boldsymbol{\lambda}_{s+1}.$$

The optimality condition for consumption becomes

$$\tilde{\chi}_s = u'(c_s) + \mu_s - \lambda_{a,s},$$

where $\tilde{\chi}_s = -\chi_s \frac{u''(c_s)}{g_s}$. The optimality condition for monetary policy now becomes

$$0 = \Big[u'(c_{s,1}) + \mu_s \Big] g_{s,1} a_1 + \sum_{i \geq 2} g_{s,i} a_i \frac{\boldsymbol{D}_{a,[i,:]}}{da} \boldsymbol{\lambda}_s$$

And finally, the optimality condition for aggregate economic activity becomes

$$0 = \Big[u'(c_{s,1}) g_{s,1} z_1 A_s + \mu_s g_{s,1} z_1 A_s + \sum_{i \geq 2} g_{s,i} z_i A_s \frac{\boldsymbol{D}_{a,[i,:]}}{da} \boldsymbol{\lambda}_s \Big] d\boldsymbol{x} - v'(N_s) - \mu_s A_s.$$

C.2 Proof of Proposition 2

Proving Proposition 2 amounts to a judicious combination of the optimality conditions for policy under discretion (Proposition 1). In particular, we introduce two useful policy perturbations that are purposefully designed so that they have a neutral impact on aggregate excess demand. The first perturbation combines equations (20) and (21), yielding an *aggregate activity condition*. This perturbation entails making households work an extra hour while forcing them to consume the proceeds of any additional income. At an optimum, the marginal value of this perturbation for a planner must satisfy

$$\underbrace{\iint \left(z u'(c_t(a,z)) - \frac{v'(N_t)}{A_t} \right) g_t(a,z)\, da\, dz}_{\text{Aggregate Labor Wedge}} - \iint z \tilde{\chi}_t(a,z) g_t(a,z)\, da\, dz = 0.$$

The aggregate labor wedge captures the social marginal benefit of increasing aggregate activity. And if the planner had the ability to control households consumption-savings decisions (i.e., if $\chi_t(a,z) = 0$), the aggregate activity condition shows that a planner would set the labor wedge to zero. However, the planner must account for the fact that increasing consumption impacts households' savings decisions, which counterbalacnes the desire to set the aggregate labor wedge to zero.[51]

The second perturbation combines equations (20) and (22) and yields an *interest rate condition*. This perturbation entails a unit increase in interest rates while making households directly consume the resulting pecuniary gains. At an optimum, the marginal value of this perturbation for a planner

[51] The aggregate activity condition also connects μ_t directly to the aggregate labor wedge when policy is set with discretion. When substituting in for $\tilde{\chi}_t(a,z)$ from equation (20), we obtain a condition that defines μ_t as a weighted sum of future labor wedges:

$$\mu_t = \frac{\iint \left(z u'\left(c_t\left(a,z\right)\right) - \frac{v'(N_t)}{A_t} \right) g_t\left(a,z\right)\, da\, dz}{\iint z\left(1 - \mathcal{M}_t\left(a,z\right)\right) g_t\left(a,z\right)\, da\, dz}$$

must satisfy

$$\underbrace{\iint au'(c_t(a,z))g_t(a,z)\,da\,dz}_{\text{Distributive Pecuniary Effect}} - \iint a\tilde{\chi}_t(a,z)g_t(a,z)\,da\,dz = 0.$$

In this case, the social marginal benefit of increasing interest rates is captured by its distributive pecuniary effect, which is negative.

The planner understands that a change in rates simply redistributes resources across savers and borrowers, since distributive pecuniary effects are always zero-sum in aggregate in dollar terms. However, since borrowers typically have a higher marginal utility of consumption than savers, a utilitarian planner perceives that an increase in rates decreases social welfare through this channel. This desire to redistribute towards high marginal utility households by reducing interest rates—a motive that is absent in representative-agent economies—is a central determinant of optimal monetary policy in our environment. As in the case of the aggregate activity perturbation, the planner must account for the fact that change interest rates impacts households' savings decisions.

While both policy perturbations are neutral in terms of aggregate excess demand, they are not neutral intertemporally in terms of their impact on households' savings decisions. However, we can scale and combine both perturbations to neutralize the intertemporal effect, obtaining the targeting rule of Proposition 2. Formally, we use the fact that we can write $\chi_t(a,z)$ as

$$\tilde{\chi}_t(a,z) = (1 - \mathcal{M}_t(a,z))\,\mu_t. \tag{66}$$

Hence, by substituting for $\chi_t(a,z)$ in both the aggregate activity condition and the interest rate condition, and equalizing μ_t, we recover equation (25) in the text. This targeting rule shows that, under discretion, a utilitarian planner in a heterogeneous-agent environment trades off aggregate stabilization against redistribution.

Equation (25) shows that the welfare impact of a perturbation that jointly increases hours worked by all households and reduces the interest rate by Ω_t^D basis points—forcing households to consume the resulting proceeds—must be zero at an optimum.[52] Equation (25) implies that, at an optimum, the planner sets policy trading off aggregate stabilization and redistribution motives.

Derivation of output gap targeting rule for isoelastic preferences. Under isoelastic (CRRA) preferences, we can represent the targeting rule of Proposition 2 in terms of output gaps. We use

$$\iint \frac{zu'(c_t(a,z))}{u'(C_t)}g_t(a,z)\,da\,dz - Y_t^{\gamma+\eta}A_t^{-(1+\eta)} = \Omega_t^{\mathrm{D}} \iint \frac{au'(c_t(a,z))}{u'(C_t)}g_t(a,z)\,da\,dz.$$

[52] And since the net present value MPC, $\mathcal{M}_t$, is bounded between 0 and 1, we have $\Omega_t^{\mathrm{D}} > 0$.

Using natural output definitions

$$\tilde{Y}_t^{\mathrm{RA}} = \left(\frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L)A_t^{1+\eta}\right)^{\frac{1}{\gamma+\eta}}$$

$$\tilde{Y}_t^{\mathrm{HA}} = \left(\frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L)A_t^{1+\eta}\iint \frac{zu'(c_t(a,z))}{u'(C_t)} g_t(a,z)\,da\,dz\right)^{\frac{1}{\gamma+\eta}}$$

we rearrange and obtain

$$\frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L)Y_t^{\gamma+\eta} = \frac{\epsilon_t - 1}{\epsilon_t}(1+\tau^L)A_t^{1+\eta}\left(\iint \frac{zu'(c_t(a,z))}{u'(C_t)} g_t(a,z)\,da\,dz - \Omega_t^{\mathrm{D}}\iint \frac{au'(c_t(a,z))}{u'(C_t)} g_t(a,z)\,da\,dz\right)$$

or simply

$$Y_t = \tilde{Y}_t \times \underbrace{\left(\frac{\epsilon_t}{\epsilon_t - 1}\frac{1}{1+\tau^L}\right)^{\frac{1}{\gamma+\eta}}}_{\text{Cost-Push Wedge}} \times \underbrace{\left(1 - \Omega_t^{\mathrm{D}}\frac{\iint \frac{au'(c_t(a,z))}{u'(C_t)} g_t(a,z)\,da\,dz}{\iint \frac{zu'(c_t(a,z))}{u'(C_t)} g_t(a,z)\,da\,dz}\right)^{\frac{1}{\gamma+\eta}}}_{\text{Redistribution}}$$

C.3 Proof of Proposition 3

Our expression for inflationary bias follows by plugging in the targeting rule of Proposition 2, rewritten as an expression for the aggregate labor wedge, into the Phillips curve (6). Setting $\dot{\pi}_{\mathrm{ss}}^w = 0$, this yields

$$\pi_{\mathrm{ss}}^w = -\frac{\epsilon}{\delta}A_{\mathrm{ss}}N_{\mathrm{ss}}\iint\left(\frac{\epsilon - 1}{\epsilon}(1+\tau^L)zu'(c_{\mathrm{ss}}(a,z)) - \frac{v'(N_{\mathrm{ss}})}{A_{\mathrm{ss}}}\right)g_{\mathrm{ss}}(a,z)\,da\,dz.$$

We now separate terms into a markup component and a redistribution component, analogous to Proposition 2,

$$\pi_{\mathrm{ss}}^w = -\frac{\epsilon}{\delta}A_{\mathrm{ss}}N_{\mathrm{ss}}\iint\left(\frac{\epsilon - 1}{\epsilon}(1+\tau^L)zu'(c_{\mathrm{ss}}(a,z)) - zu'(c_{\mathrm{ss}}(a,z)) + \Omega_{\mathrm{ss}}^{\mathrm{D}}au'(c_{\mathrm{ss}}(a,z))\right)g_{\mathrm{ss}}(a,z)\,da\,dz$$

or simply

$$\pi_{\mathrm{ss}}^w = \frac{\epsilon}{\delta}A_{\mathrm{ss}}N_{\mathrm{ss}}\left[\left(1 - \frac{\epsilon - 1}{\epsilon}(1+\tau^L)\right)\iint zu'(c_{\mathrm{ss}}(a,z))g_{\mathrm{ss}}(a,z)\,da\,dz - \Omega_{\mathrm{ss}}^{\mathrm{D}}\iint au'(c_{\mathrm{ss}}(a,z))g_{\mathrm{ss}}(a,z)\,da\,dz\right],$$

which concludes the proof.

D Optimal Policy and Ramsey Plans in Sequence Space

In this Appendix, we discuss how to operationalize our method and compute optimal policy numerically. Following much of the recent literature on computational methods in heterogeneous-agent economies, we work with a sequence-space representation of our model. In the interest of accessibility, we follow the notation and conventions of Auclert et al. (2021) as closely as possible, extending their work on sequence-space Jacobians to Ramsey problems and welfare analysis. While they work in discrete time, we show below that continuous-time heterogeneous-agent models are nested by the same general model representation they propose. To establish this relationship, we first discretize our model following the same steps that would also be required for numerical implementation.

Discretization. We first discretize the equations that characterize competitive equilibrium and optimal policy in both time and space. We use a finite-difference discretization scheme building on Achdou et al. (2022).[53] In particular, we discretize the time dimension over a finite horizon, $t \in [0, T]$ where T can be arbitrarily large, using N discrete time steps, which we denote by $n = 1, \ldots, N$. With a step size $dt = \frac{T}{N-1}$, we have $t_n = dt(n-1)$. We similarly discretize the idiosyncratic state space over (a, z) using J grid points. Using bold-faced notation, we denote the discretized consumption policy function of the household at time t_n as the $J \times 1$ vector $\boldsymbol{c}_n$, where the ith element corresponds to $c_{t_n}(a_i, z_i)$.

D.1 Sequence-Space Representation of Equilibrium

After discretizing our model, the resulting equations satisfy the general model representation of heterogeneous-agent economies presented in Auclert et al. (2021). To facilitate comparison, we follow their notation in this Appendix. We consider a general representation of a heterogeneous-agent problem as a mapping from time paths of aggregate inputs $(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z})$ to time paths of aggregate outputs $\boldsymbol{Y}$. We use bold-faced notation here to indicate time paths, with $\boldsymbol{i} = \{i_n\}_{n=1}^{N}$. It will be useful to explicitly distinguish between the time paths for policy $\boldsymbol{i}$ and the exogenous shock $\boldsymbol{Z}$ on the one hand, and the time paths for other aggregate inputs $\boldsymbol{X}$ on the other hand. To simplify the exposition, we assume that there is only one aggregate input variable other than policy and the shock, so that $X_n \in \mathbb{R}$.

Denoting the discretized cross-sectional distribution by the $J \times 1$ vector $\boldsymbol{g}_n$, our main focus will be on outcome variables that take the form $Y_n = \boldsymbol{y}_n' \boldsymbol{g}_n$, where $\boldsymbol{y}_n$ is a $J \times 1$ vector that represents

[53] For a detailed description of the discretization procedure, see Achdou et al. (2022) or Schaab and Zhang (2022). We also leverage the adaptive sparse grid method developed by Schaab and Zhang (2022) and Schaab (2020) to solve dynamic programming problems in continuous time.

an individual outcome.[54] For example, aggregate consumption takes the form $C_n = \boldsymbol{c}_n' \boldsymbol{g}$. Given an initial distribution $\boldsymbol{g}_0$, aggregate outcomes $\boldsymbol{Y}$ then solve the system of equations

$$\boldsymbol{V}_n = v(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n) \tag{67}$$

$$\boldsymbol{g}_{n+1} = \Lambda(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n)\boldsymbol{g}_n \tag{68}$$

$$Y_n = y(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n)'\boldsymbol{g}_n. \tag{69}$$

The implementability conditions of our baseline HANK economy can be expressed in terms of the time paths of macroeconomic aggregates as well as those of aggregate outcomes $\boldsymbol{Y}$. Using the above representation, aggregate outcomes $\boldsymbol{Y}$ can in turn be expressed in terms of the time paths of aggregate allocations and prices $\boldsymbol{X}$, policy $\boldsymbol{i}$, and shocks $\boldsymbol{Z}$. In summary, equilibria given policy and shocks can be expressed in terms of the *equilibrium map* $\mathcal{H}(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z}) = 0$.

D.2 Sequence-Space Representation of Ramsey Plans

We now show how to express the Ramsey plan optimality conditions, which characterize the multipliers and optimal policy, in a general model representation akin to equations (67) through (69). In general, Ramsey plans in heterogeneous-agent economies feature three types of multipliers: aggregate multipliers, individual forward-looking multipliers, and individual backward-looking multipliers. In our baseline environment, the aggregate multipliers are θ_t and μ_t, the individual forward-looking multiplier is $\lambda_t(a, z)$, and the (system of) individual backward-looking multipliers is $\phi_t(a, z)$ and $\chi_t(a, z)$.

The Ramsey plan representation (52) summarizes the optimality conditions of the timeless Ramsey plan in sequence-space form. In particular, the Ramsey map $\mathcal{R}(\cdot)$ takes the time paths of aggregate multipliers $\boldsymbol{M}$ as explicit inputs. Our goal now is to show that the optimality conditions of the timeless Ramsey plan can be written in terms of $\boldsymbol{R} = (\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{i})$ and $\boldsymbol{Z}$.

Forward-looking individual multipliers take the form

$$\boldsymbol{\lambda}_n = f(\boldsymbol{\lambda}_{n+1}, \boldsymbol{V}_n, X_n, M_n, i_n, Z_n), \tag{70}$$

which is analogous to equation (67), which characterizes individual forward-looking behavior. For example, it is straightforward to verify that equation (31) satisfies this form: it expresses today's multiplier $\lambda_n(a, z)$ in terms of today's aggregate multipliers, individual allocations, aggregate allocations and prices, as well as the future multiplier $\lambda_{n+1}(a, z)$.

Analogously to equation (67), the recursive structure of forward-looking individual multipliers

[54] We normalize the discretized distribution representation so that g_n sums to 1, i.e., $\mathbf{1}' g_n = 1$, where $\mathbf{1}$ is a $J \times 1$ vector of 1s.

allows us to efficiently compute their first-order derivatives. We summarize this observation in the following Lemma.

Lemma 22. *For any $k \geq 1$, we have*

$$\frac{\partial \boldsymbol{\lambda}_n}{\partial i_k} = \begin{cases} 0 & \text{if} \quad n > k \\ \frac{\partial \boldsymbol{\lambda}_{n-s}}{\partial i_{k-s}} & \text{else for } s < n \end{cases}$$

and likewise for first-order partial derivatives in X_k, M_n, and Z_n.

Lemma 22 represents an extension of Auclert et al. (2021)'s fake-news result to the multipliers that emerge from Ramsey problems. The unifying insight here is that, just like competitive equilibrium in heterogeneous-agent economies comprises a forward-backward system of dynamic equations, so do Ramsey plans. In other words, equation (31), which defines the social lifetime value $\lambda_t(a,z)$ still satisfies a (Hamilton-Jacobi-) Bellman equation. The "fake-news property" identified by Auclert et al. (2021) applies to any backward equation, including those satisfied by multipliers.

Backward-looking individual multipliers typically correspond to promises that the Ramsey planner makes to individuals. They are characterized by a particular kind of Kolmogorov forward equation. In particular, promise-keeping Kolmogorov forward equations feature a forcing term that captures the "births" and "deaths" of promises, captured by $\partial_a \chi_t(a,z)$ in equation (30). Consequently, backward-looking multipliers can be represented as

$$\boldsymbol{\phi}_{n+1} = \Lambda(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n)\boldsymbol{\phi}_n + b(\boldsymbol{\phi}_n, \boldsymbol{\lambda}_n, \boldsymbol{V}_n, \boldsymbol{g}_n, X_n, M_n, i_n, Z_n). \tag{71}$$

The same arguments developed by Auclert et al. (2021) for efficiently computing sequence-space derivatives of the cross-sectional distribution also apply to multipliers that satisfy (Kolmogorov) forward equations.

The multiplier representations (70) and (71), together with equations (67) through (69) and the equilibrium map (51), let us conclude that timeless Ramsey plans admit the sequence-space representation (52).

D.3 Sequence-Space Perturbations in the Primal: Proof of Proposition 11

Denote the Ramsey plan by $\boldsymbol{R} = (\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{i})$. Using a first-order Taylor expansion around the stationary Ramsey plan, we have

$$\mathcal{R}(\boldsymbol{R}, \boldsymbol{Z}) \approx \mathcal{R}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) + \mathcal{R}_{\boldsymbol{R}}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})d\boldsymbol{R} + \mathcal{R}_{\boldsymbol{Z}}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})d\boldsymbol{Z}.$$

Notice that we have $\mathcal{R}(\boldsymbol{R}, \boldsymbol{Z}) = 0$ by the definition of $\boldsymbol{R}$ as a Ramsey plan, i.e., as solving $\mathcal{R}(\cdot) = 0$ for a given $\boldsymbol{Z}$ as in (52).

We now show that $\mathcal{R}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) = 0$ as well, assuming, as we do in Proposition 11, that the Ramsey problem is initialized at $(g_0, \phi, \theta) = (g_{\text{ss}}, \phi_{\text{ss}}, \theta_{\text{ss}})$. The Ramsey plan map $\mathcal{R}(\cdot)$ is a system of equations that comprises two sets of conditions, those for competitive equilibrium as well as the first-order conditions associated with the timeless Ramsey problem in the primal representation. By definition, the stationary Ramsey plan comprises a feasible competitive equilibrium. Consequently, when evaluated at $(\boldsymbol{R}, \boldsymbol{Z}) = (\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})$, those conditions in $\mathcal{R}(\cdot)$ associated with competitive equilibrium are 0.

That leaves the first-order conditions associated with the timeless primal Ramsey problem. It follows from Proposition 6 that the timeless primal Ramsey problem is time-consistent, so that the planner does not want to deviate from the stationary Ramsey plan when $\boldsymbol{Z} = \boldsymbol{Z}_{\text{ss}}$. It also follows from our duality proof for the primal and dual representations that $\frac{d}{d\boldsymbol{i}} L^{\text{TP}}(\boldsymbol{\theta}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) = 0$ also implies that each of the associated first-order conditions of the timeless primal problem are 0 when evaluated at $(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})$. Putting these observations together implies $\mathcal{R}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) = 0$.

We are then simply left with

$$0 \approx \mathcal{R}_{\boldsymbol{R}}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) d\boldsymbol{R} + \mathcal{R}_{\boldsymbol{Z}}(\boldsymbol{R}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) d\boldsymbol{Z}.$$

Rearranging and inverting $\mathcal{R}_{\boldsymbol{R}}$ yields the desired result.

D.4 Sequence-Space Perturbations in the Dual

In this section, we develop a sequence-space perturbation approach to solve optimal stabilization policy in the dual. We take as our starting point not equation (52) but a sequence-space representation of the timeless dual Lagrangian, which we now introduce.

The timeless dual Lagrangian defined in Appendix B.6 takes as its inputs (i) the time paths of allocations and prices, (ii) an initial distribution, and (iii) initial timeless penalties. Unlike in the primal form, it does not explicitly feature the time paths of multipliers. For a given path of policy $\boldsymbol{i}$ and shocks $\boldsymbol{Z}$, we can directly use the equilibrium map (51) to solve out for allocations and prices, i.e., $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z})$. A sequence-space representation of the timeless dual Lagrangian is then given by

$$L^{\text{TD}}(\boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z}), \boldsymbol{i}, \boldsymbol{Z}), \tag{72}$$

where we again leave implicit the dependence of $L^{\text{TD}}(\cdot)$ on $g_0(a, z)$ as well as $\phi(a, z)$ and θ.

The timeless dual Lagrangian (72) implies the local efficiency criterion for optimal policy

$$\mathcal{F}(\boldsymbol{i}, \boldsymbol{Z}) = \frac{d}{d\boldsymbol{i}} L^{\text{TD}}(\boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z}), \boldsymbol{i}, \boldsymbol{Z}) = 0, \tag{73}$$

where $\mathcal{F}(\cdot)$ implicitly takes as given an initial distribution as well as initial promises. Equation (73) represents the planner's necessary first-order optimality condition in sequence-space form. We can use it to directly characterize optimal policy in the dual in terms of exogenous shocks, i.e.,

$$\boldsymbol{i} = \boldsymbol{i}(\boldsymbol{Z}).$$

Importantly, given policy $\boldsymbol{i}$, the timeless penalty (ϕ, θ) does not affect competitive equilibrium, summarized by (51). It only influences the planner's assessment of optimal policy.

Proposition 23. (**Optimal Policy Perturbations in the Dual**) *Consider the dual Ramsey problem, under which a locally efficient policy is characterized by $\mathcal{F}(\cdot) = 0$. Suppose we initialize the Ramsey problem at the stationary Ramsey plan, with $g_0(a,z) = g_{\text{ss}}(a,z)$, and with initial timeless penalties $\phi(a,z) = \phi_{\text{ss}}(a,z)$ and $\theta = \theta_{\text{ss}}$. To first order, optimal stabilization policy is then characterized by*

$$d\boldsymbol{i} = -\mathcal{F}_{\boldsymbol{i}}^{-1}\mathcal{F}_{\boldsymbol{Z}}d\boldsymbol{Z}, \tag{74}$$

where $d\boldsymbol{Z} = \boldsymbol{Z} - \boldsymbol{Z}_{\text{ss}}$ is the exogenous shock, and where $\mathcal{F}_{\boldsymbol{i}}$ and $\mathcal{F}_{\boldsymbol{Z}}$ denote Hessians of the timeless dual Lagrangian.

We prove Proposition 23 at the end of this subsection.

In the dual, approximating optimal policy using sequence-space perturbation methods requires computing second-order total derivatives of the timeless dual Lagrangian. These are in turn given by the Jacobians of the planner's first-order condition, i.e., $\mathcal{F}_{\boldsymbol{i}}$ and $\mathcal{F}_{\boldsymbol{Z}}$. The recent literature on perturbation methods in heterogeneous-agent economies has shown that sequence-space Jacobians, i.e., first-order derivatives of model objects in sequence-space representation, are sufficient to characterize transition dynamics to first order. Similarly, we have shown in Section 5.1.2 that computing optimal policy and Ramsey plans in the primal also only requires sequence-space Jacobians. Computing optimal policy and welfare in the dual representation of our Ramsey problem requires a second-order analysis, however.

To that end, we introduce *sequence-space Hessians* as the natural, second-order generalization of sequence-space Jacobians. In Appendix D.5, we formally define sequence-space Hessians, and we show both how to efficiently compute and leverage them to characterize optimal policy in the dual. We extend the methodology developed by Auclert et al. (2021) to problems that require second-order derivatives, i.e., sequence-space Hessians. While we focus on their use in the context of computing Ramsey plans in the dual, we argue that sequence-space Hessians are useful more broadly whenever a second-order analysis is required.

Advantages and disadvantages of primal and dual formulations. Our approach allows for a representation of Ramsey problems in either the primal or the dual form, and we show how to characterize and compute optimal policy in both cases. We conclude this section with a discussion of the advantages and disadvantages of both approaches.

The primal form and, in particular, the associated Ramsey plan representation (52) are more conducive to computing optimal policy non-linearly. Computing the Ramsey plan in the primal requires solving a system of non-linear equations. Fully optimized quasi-Newton methods and other non-linear equation solvers can be leveraged for this task.

In the primal approach, however, we have to compute multipliers and their transition dynamics. In the Ramsey plan representation (52), multipliers $\boldsymbol{M}$ enter as an explicit argument. It is not generically possible to characterize Ramsey plans in the spirit of (52) without multipliers. The dual approach, on the other hand, takes as its starting point the timeless dual Lagrangian, which does not explicitly depend on multipliers. Consequently, it is not necessary to compute the time paths of multipliers to characterize optimal policy in the dual.

The relative disadvantage of the dual approach, however, is that a first-order approximation of optimal policy requires second-order total derivatives of the timeless dual Lagrangian. Unlike in the primal approach, which only requires computing sequence-space Jacobians, in the dual we have to compute sequence-space Hessians. Computationally, this is a more complex task both in terms of compute time and memory demands. In summary, therefore, the main tradeoff between the primal and dual approaches is that the former requires computing the time paths of multipliers, while the latter requires sequence-space Hessians instead of only Jacobians.

Another advantage of the dual representation is that it provides an easily implementable local efficiency criterion. In particular, assessing whether a policy $\boldsymbol{i}$ is locally efficient in the dual only requires computing first-order derivatives of the timeless dual Lagrangian, which is possible in terms of sequence-space Jacobians. Unlike in the primal, however, efficiency assessments in the dual do not require computing the derivatives of multipliers. In practice, even if optimal policy is computed in the primal, verifying local efficiency in the dual is a cheap but helpful exercise.

Lastly, sequence-space Hessians and, more broadly, second-order sequence-space perturbation methods likely have many useful applications beyond computing Ramsey plans in the dual. We therefore view our treatment of sequence-space Hessians as a standalone contribution of this paper.

Proof of Proposition 23. We proceed as follows. A first-order Taylor approximation of $F(\cdot)$ (in $\boldsymbol{i}$ and $\boldsymbol{Z}$) around the stationary Ramsey plan yields

$$\begin{aligned} F(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}, \boldsymbol{Z}) = & F(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) \\ & + F_{\boldsymbol{i}}(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})(\boldsymbol{i} - \boldsymbol{i}_{\text{ss}}) + F_{\boldsymbol{Z}}(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}})(\boldsymbol{Z} - \boldsymbol{Z}_{\text{ss}}). \end{aligned}$$

First, we must have

$$F(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}, \boldsymbol{Z}) = 0$$

by construction because this is our definition for optimal policy $\boldsymbol{i}(\boldsymbol{Z})$. Second, we also have

$$F(\boldsymbol{g}_{\text{ss}}, \boldsymbol{\phi}_{\text{ss}}, \theta_{\text{ss}}, \boldsymbol{i}_{\text{ss}}, \boldsymbol{Z}_{\text{ss}}) = 0,$$

which is the main result of Section 4.3 and whose proof is in Appendix B.7. Denoting $d\boldsymbol{i} = \boldsymbol{i} - \boldsymbol{i}_{\text{ss}}$ and $d\boldsymbol{Z} = \boldsymbol{Z} - \boldsymbol{Z}_{\text{ss}}$, we thus have

$$0 = F_{\boldsymbol{i}} d\boldsymbol{i} + F_{\boldsymbol{Z}} d\boldsymbol{Z},$$

where the Jacobians of $F(\cdot)$ are evaluated at the stationary Ramsey plan.

D.5 Sequence-Space Hessians

To compute optimal policy in the dual using Proposition 23, we effectively need to differentiate $L^{\text{TD}}(\cdot)$ twice. In particular, $F(\cdot) = \frac{d}{d\boldsymbol{i}} L^{\text{TD}}$ features first-order derivative terms, which can be cast as sequence-space Jacobians (Auclert et al., 2021). Therefore, computing the total derivatives $\frac{d}{d\boldsymbol{i}} F(\cdot)$ and $\frac{d}{d\boldsymbol{Z}} F(\cdot)$, which are used in equation (74) to characterize optimal policy $d\boldsymbol{i}$, we require second-order derivatives. Consequently, computing optimal stabilization policy using our approach requires that we compute both first- and second-order total derivatives of all objects that feature in the timeless dual Lagrangian.

In a sequence-space representation of our model, these objects are all functions of the time paths of aggregate inputs, i.e., $(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z})$, where $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z})$. Moreover, the timeless dual Lagrangian itself can be represented in terms of aggregate outcomes $\boldsymbol{Y}$, using the general model representation above. Consequently, computing the matrices $\frac{d}{d\boldsymbol{i}} F$ and $\frac{d}{d\boldsymbol{Z}} F$ requires taking total derivatives of specific aggregate outcomes $\boldsymbol{Y}$.

We define *sequence-space Hessians* as the matrices of mixed partial derivatives of model objects that can be represented as functions of aggregate sequences around the stationary Ramsey plan. We discuss these mixed partial derivative matrices in detail in Section D.5.1. Subsequently, in Section D.5.2, we show how to build up the second-order total derivatives of the timeless dual Lagrangian, i.e., $\frac{d}{d\boldsymbol{i}} F$ and $\frac{d}{d\boldsymbol{Z}} F$, from sequence-space Hessians.

D.5.1 A Fake-News Algorithm to Compute Sequence-Space Hessians

We now extend the fake-news algorithm of Auclert et al. (2021) to compute sequence-space Hessians, i.e., the matrices of mixed partial derivatives $\frac{\partial^2}{\partial i_k \partial i_l} Y_n$ in a sequence-space representation of the model. The results we present below hold for any mixed partial derivative of $Y_n(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z})$, but to ease notation we focus specifically on the mixed derivative $\frac{\partial^2}{\partial i_k \partial i_l}$ for some given $k, l \in \{1, \ldots, N\}$.

Using equation (69), we can rewrite the mixed derivative of aggregate outcome Y_n at time t_n as

$$\frac{\partial^2 Y_n}{\partial i_k \partial i_l} = \frac{\partial}{\partial i_l}\left(\boldsymbol{y}_n' \frac{\partial \boldsymbol{g}_n}{\partial i_k} + \boldsymbol{g}_n' \frac{\partial \boldsymbol{y}_n}{\partial i_k} \right) = \boldsymbol{y}_n' \frac{\partial^2 \boldsymbol{g}_n}{\partial i_k \partial i_l} + \frac{\partial \boldsymbol{g}_n}{\partial i_k}' \frac{\partial \boldsymbol{y}_n}{\partial i_l} + \boldsymbol{g}_n' \frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial i_l} + \frac{\partial \boldsymbol{y}_n}{\partial i_k}' \frac{\partial \boldsymbol{g}_n}{\partial i_l}$$

This derivation underscores that we generally need both the first-order and second-order mixed partial derivatives of individual outcomes $\boldsymbol{y}_n$ and the distribution $\boldsymbol{g}_n$ to compute aggregate sequence-space Hessians $\partial^2 \boldsymbol{Y}$. Our method leverages several useful properties of these first- and second-order derivatives. In the following, we prove key properties of the second-order mixed derivatives $\frac{\partial^2}{\partial i_k \partial i_l} \boldsymbol{y}_n$ and $\frac{\partial^2}{\partial i_k \partial i_l} \boldsymbol{g}_n$, and we refer the reader to Auclert et al. (2021) for the properties of the first-order partial derivatives.

First, notice that mixed partial derivatives are symmetric, or interchangeable, by the standard continuity argument. That is

$$\frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial i_l} = \frac{\partial^2 \boldsymbol{y}_n}{\partial i_l \partial i_k} \qquad \text{and} \qquad \frac{\partial^2 \boldsymbol{g}_n}{\partial i_k \partial i_l} = \frac{\partial^2 \boldsymbol{g}_n}{\partial i_l \partial i_k}.$$

Second, the recursive structure of the system (67) - (69) gives rise to the following key property of mixed partial derivatives in sequence space.

Lemma 24. *We have*

$$\frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial i_l} = \begin{cases} 0 & \text{if} \quad n > \min\{k, l\} \\ \frac{\partial^2 \boldsymbol{y}_{n-s}}{\partial i_{k-s} \partial i_{l-s}} & \text{else for } s < n \end{cases}$$

Leveraging these first two properties of mixed derivatives of individual outcomes in sequence space, we can construct sequence-space Hessian matrices using the following shortcut: Instead of computing all N^2 numerical derivatives, we simply compute

$$\frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial i_N}$$

for $1 \le k \le N$, which requires only N numerical derivative evaluations.[55]

Third, we exploit the fact that the transition matrix Λ, which describes the law of motion of the cross-sectional distribution in equation (69), has a particular structure in continuous time. In particular, we have

$$\Lambda(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n) = 1 + dt\boldsymbol{A}(\boldsymbol{V}_{n+1}, X_n, i_n, Z_n)',$$

where $\boldsymbol{A}$ is the $J \times J$ matrix that discretizes the HJB operator $\mathcal{A}$, and $\boldsymbol{A}'$, its transpose, is the analog

[55] For other mixed derivatives, such as $\frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial Z_l}$, we require $2N$ evaluations, i.e., both $\frac{\partial^2 \boldsymbol{y}_n}{\partial i_k \partial Z_N}$ and $\frac{\partial^2 \boldsymbol{y}_n}{\partial i_N \partial Z_l}$.

for the adjoint $\mathcal{A}^*$. In our baseline HANK model, the discretized transition matrix takes the form

$$\boldsymbol{A}_n = \boldsymbol{s}_n \cdot \boldsymbol{D}_a + \boldsymbol{A}^z \tag{75}$$

where $\boldsymbol{A}^z$ is given exogenously, and its derivatives with respect to θ_k, X_k and Z_k are therefore 0. The matrix $\boldsymbol{D}_a$ discretizes the partial derivative ∂_a, and it is also invariant to perturbations in aggregate inputs as long as the step size used for the numerical derivative is fine enough. In particular, the key insight here is that taking derivatives of the general transition matrix Λ in equation (69) simply amounts to differentiating $\boldsymbol{s}_n$ in equation (75).[56] We record this observation in the following Lemma.

Lemma 25. *The first- and second-order mixed partial derivatives of the transition matrix Λ_n in our setting are given by*

$$\frac{\partial \Lambda_n}{\partial i_k} = dt \frac{\partial \boldsymbol{s}_n}{\partial i_k} \cdot \boldsymbol{D}_a \quad \text{and} \quad \frac{\partial^2 \Lambda_n}{\partial i_k \partial i_l} = dt \frac{\partial^2 \boldsymbol{s}_n}{\partial i_k \partial i_l} \cdot \boldsymbol{D}_a.$$

Fourth, we characterize the properties of the mixed derivatives of the cross-sectional distribution. We assume for simplicity that the economy is initialized at the cross-sectional distribution that corresponds to the stationary Ramsey plan, that is, $\boldsymbol{g}_1 = \boldsymbol{g}_{\text{ss}}$, where we recall that n starts at 1 and $t_1 = 0$. The initial distribution is given exogenously and does not adjust on impact. That is, $\frac{\partial^2 \boldsymbol{g}_1}{\partial i_k \partial i_l} = 0$. Using equation (69) and Lemma 25, the response of the cross-sectional distribution at time step $n = 2$ is thus

$$\frac{\partial^2 \boldsymbol{g}_2}{\partial i_k \partial i_l} = \frac{\partial^2 \Lambda_1}{\partial i_k \partial i_l} \boldsymbol{g}_{\text{ss}} = dt \left(\frac{\partial^2 \boldsymbol{s}_1}{\partial i_k \partial i_l} \cdot \boldsymbol{D}_a \right) \boldsymbol{g}_{\text{ss}}$$

We now exploit the recursive structure of equation (69) to derive two alternative expressions for the mixed derivatives $\frac{\partial^2 \boldsymbol{g}_n}{\partial i_k \partial i_l}$, for $n \geq 3$. We summarize in the next Lemma.

Lemma 26. *The mixed partial derivatives of the cross-sectional distribution $\boldsymbol{g}_n$ at time steps $n \geq 3$ can be computed recursively using*

$$\begin{aligned} \frac{\partial^2 \boldsymbol{g}_n}{\partial i_k \partial i_l} =& \Lambda_{ss} \frac{\partial^2 \boldsymbol{g}_{n-1}}{\partial i_k \partial i_l} + \frac{\partial^2 \boldsymbol{g}_2}{\partial i_{k-(n-2)} \partial i_{l-(n-2)}} \mathbb{1}_{\min\{k-(n-2), l-(n-2)\} \geq 1} \\ &+ dt \left(\frac{\partial \boldsymbol{s}_1}{\partial i_{l-(n-2)}} \cdot \boldsymbol{D}_a \right) \frac{\partial \boldsymbol{g}_{n-1}}{\partial i_k} \mathbb{1}_{l-(n-2) \geq 1} + dt \left(\frac{\partial \boldsymbol{s}_1}{\partial i_{k-(n-2)}} \cdot \boldsymbol{D}_a \right) \frac{\partial \boldsymbol{g}_{n-1}}{\partial i_l} \mathbb{1}_{k-(n-2) \geq 1} \end{aligned}$$

[56] Notice that equation (75) is specific to our baseline model and consequently breaks with the spirit of generality otherwise adopted in this section. However, an equation *like* (75) will generally hold in any continuous-time heterogeneous-agent model. We think there is some value to highlighting how to leverage this equation when constructing sequence-space Jacobians and Hessians, and we therefore use equation (75) in the following while otherwise maintaining our general notation.

or non-recursively using

$$\frac{\partial^2 g_n}{\partial i_k \partial i_l} = \sum_{r=1}^{R_1} (\Lambda_{ss})^{n-r-1} \frac{\partial^2 g_2}{\partial i_k \partial i_l} + dt \sum_{r=1}^{R_2} (\Lambda_{ss})^{n-r-2} \left(\frac{\partial s_1}{\partial i_{k-r}} \cdot D_a \right) \frac{\partial g_{1+r}}{\partial i_l} + dt \sum_{r=1}^{R_3} (\Lambda_{ss})^{n-r-2} \left(\frac{\partial s_1}{\partial i_{l-r}} \cdot D_a \right) \frac{\partial g_{1+r}}{\partial i_k}$$

where $R_1 = \min\{k, l, n-1\}$, $R_2 = \min\{k-1, n-2\}$, *and* $R_3 = \min\{l-1, n-2\}$.

Fifth and finally, we discuss how to efficiently compute a given mixed partial derivative numerically. The most popular finite-difference stencil to compute second-order mixed derivatives is given by

$$\frac{\partial^2 y_n}{\partial i_k \partial i_l} = \frac{y_n^{++} - y_n^{+-} - y_n^{-+} + y_n^{--}}{4h^2} \tag{76}$$

where $y_n^{++} = y_n(\ldots, i_k + h, \ldots, i_l + h, \ldots)$, $y_n^{+-} = y_n(\ldots, i_k + h, \ldots, i_l - h, \ldots)$, $y_n^{-+} = y_n(\ldots, i_k - h, \ldots, i_l + h, \ldots)$, and $y_n^{--} = y_n(\ldots, i_k - h, \ldots, i_l - h, \ldots)$. This stencil requires 4 function evaluations for every mixed derivative and is therefore very costly.

An alternative and, in our case, substantially more efficient stencil is

$$\frac{\partial^2 y_n}{\partial i_k \partial i_l} = \frac{y_n^{++} - y_n^{+\cdot} - y_n^{\cdot+} + 2y_n - y_n^{-\cdot} - y_n^{\cdot-} + y_n^{--}}{2h^2} \tag{77}$$

where $y_n^{+\cdot} = y_n(\ldots, i_k + h, \ldots, i_l, \ldots)$, $y_n^{\cdot+} = y_n(\ldots, i_k, \ldots, i_l + h, \ldots)$, $y_n^{-\cdot} = y_n(\ldots, i_k - h, \ldots, i_l, \ldots)$, and $y_n^{\cdot-} = y_n(\ldots, i_k, \ldots, i_l - h, \ldots)$. Stencil (77) requires only 2 new function evaluations compared to stencil (76)'s 4. The additional terms y_n, $y_n^{+\cdot}$, and $y_n^{\cdot+}$ are already available from constructing the first-order sequence-space Jacobians. And the terms $y_n^{-\cdot}$ and $y_n^{\cdot-}$ can be computed very cheaply using the standard fake-news algorithm for first-order derivatives.

Comparison to fake-news algorithm of Auclert et al. (2021). In their seminal contribution, Auclert et al. (2021) develop a highly efficient algorithm to compute sequence-space Jacobians, showing that computing a single column of the Jacobian suffices to derive all other columns from it. For sequence-space Hessians, on the other hand, we need to evaluate one "block" of the Hessian, which requires N numerical derivatives, and is consequently substantially more expensive than computing a sequence-space Jacobian.

Why does the Hessian matrix have a higher information requirement? For Jacobians, Auclert et al. (2021) show that we only require a single piece of information to evaluate the impact of shocks on household behavior: How far in the future is the shock, i.e., what is the distance from the present to the shock. For Hessians, on the other hand, we need two pieces of information: How far in the future is the (later of the two) shock(s), and, in addition, what is the relative distance between the two shocks. We therefore cannot obtain all required information with a single numerical derivative as in the case of the Jacobian.

Nonetheless, our fake-news algorithm for sequence-space Hessians represents a substantial improvement over computing the Hessian matrices directly, which would require the evaluation of N^2 numerical derivatives.

D.5.2 Total Derivatives and General Equilibrium

Our perturbation approach to optimal stabilization policy in the dual requires the two total derivatives $\frac{d}{d\boldsymbol{i}}F$ and $\frac{d}{d\boldsymbol{Z}}F$. In particular, the $[k,l]$th entry of the $N \times N$ matrix $F_{\boldsymbol{\theta}}$ is given by

$$(F_{\boldsymbol{i}})_{[k,l]} = \sum_{n=1}^{N} e^{-\rho t_n} \frac{d^2 U_n}{di_k di_l} dt + (\boldsymbol{\phi}_{\text{ss}})' \frac{d^2 \boldsymbol{V}_1}{di_k di_l} + \theta_{\text{ss}} \frac{d^2 \pi_1^w}{di_k di_l}, \tag{78}$$

where $U_n = (u(\boldsymbol{c}_n) - v(N_n) - \frac{\delta}{2}(\pi_n^w)^2)'\boldsymbol{g}_n$. The first term in equation (78) thus captures the present discounted sum of future aggregate social welfare flows, and the second and third terms capture the timeless penalties.

So far, we have discussed how to construct the first- and second-order partial derivatives of the economic variables that comprise F. To compute total derivatives, we start with a discussion of general equilibrium.

General equilibrium. General equilibrium considerations in our model can be summarized in terms of the equilibrium map (51), which is a system of N equation, assuming for now that $X_n \in \mathbb{R}$. Given paths for policy $\boldsymbol{i}$ and the exogenous shock $\boldsymbol{Z}$, we can solve equation (51) for $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{i}, \boldsymbol{Z})$.

To compute the total derivative $F_{\boldsymbol{i}}$, i.e., the response in the planner's first-order condition to a perturbation in the policy path, we must take into account both the direct effect of the policy via its partial derivative and the indirect general equilibrium effects. We use the first-order derivatives $\boldsymbol{X}_{i_k} = -\boldsymbol{H}_{\boldsymbol{X}}^{-1}\boldsymbol{H}_{i_k}$. Likewise, the mixed partial derivatives are given by

$$\boldsymbol{X}_{i_k i_l} = -\boldsymbol{H}_{\boldsymbol{X}}^{-1}\boldsymbol{H}_{i_k i_l} + \boldsymbol{H}_{\boldsymbol{X}}^{-1}\boldsymbol{H}_{\boldsymbol{X}i_l}\boldsymbol{H}_{\boldsymbol{X}}^{-1}\boldsymbol{H}_{i_k}. \tag{79}$$

Total derivatives. We now summarize how total derivatives of Y_n relate to the partial derivatives we have discussed so far. For notational convenience, we drop the n subscript and instead use subscripts to denote partial derivatives. Recall that Y depends on the time paths of all aggregate inputs, $Y(\boldsymbol{X}, \boldsymbol{i}, \boldsymbol{Z})$.

Lemma 27. *The total derivatives of Y are given by*

$$\frac{d^2Y}{di_k di_l} = \begin{pmatrix} Y_{X_1\boldsymbol{X}}\boldsymbol{X}_{i_l} & \dots & Y_{X_N\boldsymbol{X}}\boldsymbol{X}_{i_l} \end{pmatrix} \boldsymbol{X}_{i_k} + Y_{\boldsymbol{X}i_l}\boldsymbol{X}_{i_k} + Y_{\boldsymbol{X}}\boldsymbol{X}_{i_k i_l} + Y_{i_k\boldsymbol{X}}\boldsymbol{X}_{i_l} + Y_{i_k i_l} \tag{80}$$

where subscripts denote partial derivatives, and likewise for the total derivatives $\frac{d^2Y}{di_k dZ_l}$.

The total derivatives for $\boldsymbol{V}_1$ and π_1^w take the same form and can be computed via their second-order partial derivatives together with the general equilibrium maps, i.e., the partial derivatives of $\boldsymbol{X}$. We now have all the objects we need to implement our perturbation approach in the dual and compute optimal stabilization policy numerically.

D.5.3 Algorithm to Compute Optimal Policy in the Dual

We summarize in Algorithm 1 our fake-news algorithm to compute sequence-space Hessians and, with them, optimal stabilization policy to first order in the dual representation of the timeless Ramsey problem.

Algorithm 1 Optimal Stabilization Policy using Sequence-Space Hessians

1: Compute stationary Ramsey plan

2: Compute sequence-space Jacobians around the stationary Ramsey plan using fake-news algorithm of Auclert et al. (2021)

3: Compute N numerical mixed partial derivatives, and ▷ use stencil (77)
  a: construct policy Hessians ▷ use Lemma 24
  b: construct distribution Hessians ▷ use Lemmas 25 and 26

4: Use Hessians to compute mixed derivatives of $\boldsymbol{H}$ and $\boldsymbol{X}$ ▷ use equations (51) and (79)

5: Compute total derivatives for $F_{\boldsymbol{i}}$ and $F_{\boldsymbol{Z}}$ ▷ use equations (78) and (80)

6: Compute optimal stabilization policy as $d\boldsymbol{i} = -F_{\boldsymbol{i}}^{-1}F_{\boldsymbol{Z}}d\boldsymbol{Z}$

D.5.4 Accuracy and Performance

We test the accuracy of our method in Appendix F.3. We show there that the numerical solution of optimal policy in RANK using our perturbation method based on sequence-space Hessians is highly accurate. In RANK, we can compute optimal policy analytically. We compare this exact analytical solution to the first-order approximation of optimal policy given by $d\boldsymbol{i} = -F_{\boldsymbol{i}}^{-1}F_{\boldsymbol{Z}}d\boldsymbol{Z}$. For demand shocks, we show that the difference in optimal CPI inflation, for example, is on the order of 10^{-6}. In the case of TFP shocks, the remaining discrepancy is slightly larger, with the two optimal interest rate paths differing by about 1 basis point.

E RANK with Wage Rigidity

In this Appendix, we present a self-contained treatment of optimal monetary policy in the RANK limit of our model. The RANK limit obtains when *i*) households' idiosyncratic labor productivity converges to a constant value, that is, $z_t \to \bar{z}$ for all t, and *ii*) the economy is initialized with a cross-sectional distribution of bond holdings and productivities that is degenerate at $a = 0$ and $z = \bar{z}$, that is, $g_0^{\mathrm{RA}}(a, z) = \delta(a = 0, z = \bar{z})$, where $\delta(\cdot)$ denotes a two-dimensional Dirac delta function.

It is well known that the Ramsey problem in the standard New Keynesian model can be represented with a single implementability condition, the Phillips curve. The goal of this Appendix is to parallel our derivations for HANK and facilitate comparison where possible. We therefore represent the Ramsey problem in terms of the same set of implementability conditions we use in the main text. The following Lemma summarizes these for the RANK limit.

Lemma 28. *The implementability conditions that a Ramsey planner faces in RANK can be summarized as*

$$\dot{C}_t = \frac{1}{\gamma}\left(i_t - \pi_t^w + \frac{\dot{A}_t}{A_t} - \rho_t\right)C_t$$

$$C_t = A_t N_t$$

$$\dot{\pi}_t^w = \rho_t \pi_t^w + \frac{\epsilon_t}{\delta}\left[(1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t u'(C_t) - v'(N_t)\right] N_t.$$

E.1 Standard Ramsey Problem

We associate the standard Ramsey problem in primal form with the following Lagrangian, where we drop time subscripts for convenience,

$$\begin{aligned} L = \int_0^\infty e^{-\rho t}\Bigg\{ & \frac{1}{1-\gamma} C^{1-\gamma} - v(N) - \frac{\delta}{2}(\pi^w)^2 \\ & + \phi\left[\frac{1}{\gamma}\left(i - \pi^w + \frac{\dot{A}}{A} - \rho\right)C - \dot{C}\right] \\ & + \mu\Big[AN - C\Big] \\ & + \vartheta\left[\rho\pi^w + \frac{\epsilon}{\delta}\left((1+\tau^L)\frac{\epsilon - 1}{\epsilon} A u'(C) - v'(N)\right)N - \dot{\pi}^w\right]\Bigg\} dt \end{aligned}$$

Crucially, both C_0 and π_0^w are *free* from the planner's perspective.

Proposition 29. *The first-order conditions for optimal monetary policy in RANK are given by*

$$0 = C^{-\gamma} - \mu + \phi \frac{1}{\gamma}\left(i - \pi^w + \frac{\dot{A}}{A} - \rho\right) - \rho\phi + \dot{\phi} + \theta \frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon - 1}{\epsilon} A u''(C) N \tag{81}$$

$$0 = -\delta \pi^w - \phi \frac{1}{\gamma} C + \vartheta \rho - \rho \vartheta + \dot{\vartheta} \tag{82}$$

$$0 = -v'(N) + \mu A + \vartheta \frac{\epsilon}{\delta}\left((1+\tau^L)\frac{\epsilon-1}{\epsilon} A u'(C) - v'(N)\right) - \theta \frac{\epsilon}{\delta} v''(N) N \tag{83}$$

$$0 = \phi \frac{1}{\gamma} C, \tag{84}$$

with initial conditions

$$0 = \phi_0$$
$$0 = \vartheta_0.$$

We see that we must have $\phi = 0$ for all t. This allows us to simplify the first-order conditions and arrive at

$$0 = C^{-\gamma} - \mu + \theta \frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon} A u''(C) N$$
$$\dot{\theta} = -\delta \pi^w$$
$$0 = -v'(N) + \mu A + \theta \frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon} A u'(C) - \theta \frac{\epsilon}{\delta} v'(N) - \theta \frac{\epsilon}{\delta} v''(N) N$$

with initial condition $\theta_0 = 0$. The stationary Ramsey plan satisfies

$$\pi^w_{ss} = 0$$
$$i_{ss} = \rho$$
$$N_{ss} = \left[(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]^{\frac{1}{\gamma+\eta}}$$
$$C_{ss} = N_{ss}$$
$$\theta_{ss} = \frac{1 - (1+\tau^L)\frac{\epsilon-1}{\epsilon}}{(\gamma+\eta)\frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}}$$
$$\mu_{ss} = \frac{\gamma}{\gamma+\eta}(N_{ss})^{\eta} + \frac{\eta}{\gamma+\eta}(C_{ss})^{-\gamma}.$$

We see that $\theta_{ss} = 0$ if and only if an appropriate employment subsidy is in place, so that $(1+\tau^L)\frac{\epsilon-1}{\epsilon} = 1$.

Proof of Proposition 29. We now integrate by parts and consider a general functional perturbation, yielding

$$\begin{aligned}
L = \int_0^\infty e^{-\rho t} \Bigg\{ & \frac{1}{1-\gamma}(C+\alpha h_C)^{1-\gamma} - v(N+\alpha h_N) - \frac{\delta}{2}(\pi^w + \alpha h_\pi)^2 \\
& + \phi\left[\frac{1}{\gamma}\left(i + \alpha h_i - \pi^w - \alpha h_\pi + \frac{\dot{A}}{A} - \rho\right)(C+\alpha h_C)\right] \\
& + \mu\left[A(N+\alpha h_N) - C - \alpha h_C\right] \\
& + \theta\left[\rho(\pi^w + \alpha h_\pi) + \frac{\epsilon}{\delta}\left((1+\tau^L)\frac{\epsilon-1}{\epsilon}Au'(C+\alpha h_C) - v'(N+\alpha h_N)\right)(N+\alpha h_N)\right] \\
& - \rho\phi(C+\alpha h_C) + (C+\alpha h_C)\dot{\phi} - \rho\theta(\pi^w+\alpha h_\pi) + (\pi^w + \alpha h_\pi)\dot{\theta}\Bigg\} dt \\
& + \phi_0(C_0 + \alpha h_{C,0}) + \vartheta_0(\pi_0^w + \alpha h_{\pi,0})
\end{aligned}$$

Working out the Gateaux derivatives and employing the fundamental lemma of the calculus of variations, we arrive at the following

$$\begin{aligned}
0 = \int_0^\infty e^{-\rho t} \Bigg\{ & C^{-\gamma}h_C - v'(N)h_N - \delta\pi^w h_\pi \\
& + \phi\left[\frac{1}{\gamma}\left(h_i - h_\pi\right)C + \frac{1}{\gamma}\left(i - \pi^w + \frac{\dot{A}}{A} - \rho\right)h_C\right] \\
& + \mu\left[Ah_N - h_C\right] \\
& + \theta\Bigg[\rho h_\pi + \frac{\epsilon}{\delta}\left((1+\tau^L)\frac{\epsilon-1}{\epsilon}Au''(C)h_C - v''(N)h_N\right)N \\
& \qquad + \frac{\epsilon}{\delta}\left((1+\tau^L)\frac{\epsilon-1}{\epsilon}Au'(C) - v'(N)\right)h_N\Bigg] \\
& - \rho\phi h_C + h_C\dot{\phi} - \rho\theta h_\pi + h_\pi\dot{\theta}\Bigg\} dt + \phi_0 h_{C,0} + \theta_0 h_{\pi,0}
\end{aligned}$$

Grouping terms,

$$0 = \int_0^\infty e^{-\rho t} \Bigg\{ \bigg[C^{-\gamma} - \mu + \phi \frac{1}{\gamma} \Big(i - \pi^w + \frac{\dot{A}}{A} - \rho \Big) - \rho\phi + \dot{\phi} + \theta \frac{\epsilon}{\delta} (1+\tau^L) \frac{\epsilon - 1}{\epsilon} A u''(C) N \bigg] h_C$$

$$+ \bigg[-\delta \pi^w - \phi \frac{1}{\gamma} C + \theta\rho - \rho\theta + \dot{\vartheta} \bigg] h_\pi$$

$$+ \bigg[-v'(N) + \mu A + \vartheta \frac{\epsilon}{\delta} \Big((1+\tau^L) \frac{\epsilon-1}{\epsilon} A u'(C) - v'(N) \Big) - \theta \frac{\epsilon}{\delta} v''(N) N \bigg] h_N$$

$$+ \bigg[\phi \frac{1}{\gamma} C \bigg] h_i \Bigg\} dt + \phi_0 h_{C,0} + \theta_0 h_{\pi,0}$$

The fundamental lemma of the calculus of variations yields the desired result. Since C_0 and π_0 are both free, it follows that optimality requires $\phi_0 = \theta_0 = 0$. Finally, it follows directly from $0 = \frac{1}{\gamma}\phi_t C_t$ that we must have $\phi_t = 0$ for all t.

E.2 Timeless Ramsey Problem

In the following, we leverage our timeless Ramsey approach to give a novel, non-linear characterization of optimal monetary policy in RANK. We associate the timeless Ramsey problem in the dual with the Lagrangian

$$L^{\mathrm{TD}}(\theta) = \int_0^\infty e^{-\rho t} \left\{ \frac{1}{1-\gamma} C_t^{1-\gamma} - v(N_t) - \frac{\delta}{2} (\pi_t^w)^2 \right\} dt \underbrace{-\theta \pi_0^w}_{\text{Inflation Penalty}}$$

Lemma 30. *The timeless dual Ramsey problem in RANK is time consistent. In the absence of shocks, the Ramsey planner has no incentive to deviate from the stationary Ramsey plan. That is,*

$$\frac{d}{d\boldsymbol{i}} L^{TD}(\theta) \bigg|_{ss} = 0.$$

Proof. Suppose we differentiate

$$\frac{d}{d\boldsymbol{i}} L^{\mathrm{TD}}(\theta) = \int_0^\infty e^{-\rho t} \left\{ C_t^{-\gamma} \frac{d}{d\boldsymbol{i}} - N_t^\eta \frac{dN_t}{d\boldsymbol{i}} - \delta \pi_t^w \frac{d\pi_t^w}{d\boldsymbol{i}} \right\} dt - \theta \frac{d\pi_0^w}{d\boldsymbol{i}}$$

Next, we evaluate at the stationary Ramsey plan. This yields

$$\begin{aligned}\frac{d}{d\boldsymbol{i}}L^{\mathrm{TD}}(\theta) &= \int_0^\infty e^{-\rho t}\left\{C^{-\gamma}\frac{dC_t}{d\boldsymbol{i}} - N^\eta \frac{dN_t}{d\boldsymbol{i}}\right\}dt - \theta\frac{d\pi_0^w}{d\boldsymbol{i}} \\ &= \int_0^\infty e^{-\rho t}\left\{C^{-\gamma}\left[\frac{dC_t}{d\boldsymbol{i}} - (1+\tau^L)\frac{\epsilon-1}{\epsilon}\frac{dN_t}{d\boldsymbol{i}}\right]\right\}dt - \theta\frac{d\pi_0^w}{d\boldsymbol{i}}\end{aligned}$$

Next, from $C_t = A_t N_t$, we have when evaluated at the stationary Ramsey plan that

$$\frac{dC_t}{d\boldsymbol{i}} = A\frac{dN_t}{d\boldsymbol{i}}.$$

Thus,

$$\begin{aligned}\frac{d}{d\boldsymbol{i}}L^{\mathrm{TD}}(\theta) &= \int_0^\infty e^{-\rho t}\left\{C^{-\gamma}\frac{dC_t}{d\boldsymbol{i}} - N^\eta \frac{dN_t}{d\boldsymbol{i}}\right\}dt - \theta\frac{d\pi_0^w}{d\boldsymbol{i}} \\ &= C^{-\gamma}\left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]\int_0^\infty e^{-\rho t}\frac{dC_t}{d\boldsymbol{i}}dt - \theta\frac{d\pi_0^w}{d\boldsymbol{i}}.\end{aligned}$$

Next, we use the Phillips curve. With $\lim_{T\to\infty}\pi_T^w = 0$, we have in integral form

$$\dot{\pi}_t^w = \rho\pi_t^w + \frac{\epsilon}{\delta}\left[(1+\tau^L)\frac{\epsilon-1}{\epsilon}A_t u'(C_t) - v'(N_t)\right]N_t$$

$$\pi_t^w = -\int_t^\infty e^{-\rho(s-t)}\frac{\epsilon}{\delta}\left[(1+\tau^L)\frac{\epsilon-1}{\epsilon}A_s u'(C_s) - v'(N_s)\right]N_s$$

Thus, we have

$$\begin{aligned}\frac{d\pi_0^w}{d\boldsymbol{i}} &= -\int_0^\infty e^{-\rho(s-0)}\frac{\epsilon}{\delta}\left[(1+\tau^L)\frac{\epsilon-1}{\epsilon}(1-\gamma)C^{-\gamma}\frac{dC_s}{d\boldsymbol{i}} - (1+\eta)N^\eta\frac{dN_s}{d\boldsymbol{i}}\right]ds \\ &= -\frac{\epsilon}{\delta}\left[(1+\tau^L)\frac{\epsilon-1}{\epsilon}(1-\gamma)C^{-\gamma} - (1+\eta)N^\eta\right]\int_0^\infty e^{-\rho t}\frac{dC_t}{d\boldsymbol{i}}dt \\ &= \frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}C^{-\gamma}(\gamma+\eta)\int_0^\infty e^{-\rho t}\frac{dC_t}{d\boldsymbol{i}}dt\end{aligned}$$

Thus, we have

$$\frac{d}{d\boldsymbol{i}}L^{\text{TD}}(\theta) = \overbrace{C^{-\gamma}\left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]\int_0^\infty e^{-\rho t}\frac{dC_t}{d\boldsymbol{i}}dt}^{\text{Marginal benefit from time-inconsistent deviations}}$$
$$\underbrace{-\theta\frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}C^{-\gamma}(\gamma+\eta)\int_0^\infty e^{-\rho t}\frac{dC_t}{d\boldsymbol{i}}dt}_{\text{Marginal cost of time-inconsistent deviations under timeless penalty}}$$

Finally, we now have

$$\begin{aligned}
0 &= C^{-\gamma}\left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]-\theta\frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}C^{-\gamma}(\gamma+\eta)\\
&= \left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]-\frac{1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}}{(\gamma+\eta)\frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}}\frac{\epsilon}{\delta}(1+\tau^L)\frac{\epsilon-1}{\epsilon}(\gamma+\eta)\\
&= \left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right]-\left[1-(1+\tau^L)\frac{\epsilon-1}{\epsilon}\right],
\end{aligned}$$

which concludes the proof.

■

Our constructive proof of Lemma 30 characterizes clearly the *marginal benefit* from time-inconsistent deviations from the stationary Ramsey plan. And it also shows clearly how the timeless penalty, the *marginal cost* of deviations, exactly offsets the marginal benefit. Importantly, we see here in closed-form what the economic determinants are of the marginal benefit and the timeless penalty.

E.3 Retracing Classical RANK Results

We are now ready to use our apparatus to retrace the classical analysis of optimal monetary stabilization policy in RANK. In this subsection, we restate several of the classical results in an exact, non-linear form. In much of the standard RANK literature, e.g., Clarida et al. (1999), optimal policy analysis drops the IS equation as an implementability condition and then proceeds to derive *targeting rules* for inflation and output (gaps). In the following, our goal is to retrace this classical analysis in our setting. We leverage the results we derive here in Section 4.6 of the main text to compare optimal policy and targeting rules across RANK and HANK.

Following Galí (2015), we define the *natural level of output*, denoted $\tilde{Y}_t$, as the equilibrium level of output under flexible prices. From the Phillips curve, which is in our setting given by

$$\tilde{Y}_t = \left[(1+\tau^L)\frac{\epsilon_t-1}{\epsilon_t}A_t^{1+\eta}\right]^{\frac{1}{\gamma+\eta}}. \tag{85}$$

Going back to the Phillips curve and using the resource constraint with $Y_t = A_t N_t = C_t$, we have

$$\dot{\pi}_t^w = \rho_t \pi_t^w + \frac{\epsilon_t}{\delta}\left[\tilde{Y}_t^{\gamma+\eta} - Y_t^{\gamma+\eta}\right] Y_t^{1-\gamma} A_t^{-1-\eta} \tag{86}$$

which is our sole remaining implementability condition and features all three shocks: A_t, ϵ_t, and ρ_t.

The planner's Ramsey problem can now be associated with the Lagrangian

$$L = \int_0^\infty e^{-\int_0^t \rho_s ds} \Bigg\{ u(Y_t) - v\left(\frac{Y_t}{A_t}\right) - \frac{\delta}{2}(\pi_t^w)^2 + \theta_t \left[\rho_t \pi^w + \frac{\epsilon_t}{\delta}\left((1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t u'(Y_t) - v'\left(\frac{Y_t}{A_t}\right)\right)\frac{Y_t}{A_t} - \dot{\pi}_t^w\right] \Bigg\} dt$$

We now state the main result of this appendix: a non-linear targeting rule for optimal monetary policy in RANK under demand, TFP, and cost-push shocks.

Proposition 31. (Optimal Policy Targeting Rules / Divine Coincidence in RANK)
a) (Targeting Rule) Optimal monetary policy in RANK is fully characterized by the non-linear targeting rule

$$Y_t = \tilde{Y}_t \left(\frac{\frac{1}{1+\tau^L}\frac{\epsilon_t}{\epsilon_t - 1} + \frac{\epsilon_t}{\delta}\theta_t(1-\gamma)}{1 + \frac{\epsilon_t}{\delta}\theta_t(1+\eta)} \right)^{\frac{1}{\gamma+\eta}} \tag{87}$$

b) (Divine Coincidence) Suppose there are no cost-push shocks, i.e., $\epsilon_t = \epsilon$, and we implement an employment subsidy so that $(1+\tau^L)\frac{\epsilon-1}{\epsilon} = 1$. We have

$$Y_t = \tilde{Y}_t \left(\frac{1 + \frac{\epsilon}{\delta}\theta_t(1-\gamma)}{1 + \frac{\epsilon}{\delta}\theta_t(1+\eta)} \right)^{\frac{1}{\gamma+\eta}}. \tag{88}$$

A solution to the non-linear Ramsey plan is then given by $Y_t = \tilde{Y}_t$, $\theta_t = 0$, and $\pi_t^w = 0$.

Proof. Crucially, both Y_0 and π_0^w are *free* from the planner's perspective. We start by integrating by parts, yielding

$$
\begin{aligned}
L = \int_0^\infty e^{-\int_0^t \rho_s ds} \Bigg\{ & u(Y_t) - v\left(\frac{Y_t}{A_t}\right) - \frac{\delta}{2}(\pi_t^w)^2 \\
& + \theta_t \left[\rho_t \pi^w + \frac{\epsilon_t}{\delta}\left((1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t u'(Y_t) - v'\left(\frac{Y_t}{A_t}\right)\right)\frac{Y_t}{A_t}\right] \\
& - \rho_t \theta_t \pi_t^w + \pi_t^w \dot{\theta}_t \Bigg] \Bigg\} dt + \theta_0 \pi_0^w
\end{aligned}
$$

The two first-order conditions are then given by

$$
\begin{aligned}
0 =& u'(Y_t) - v'\left(\frac{Y_t}{A_t}\right)\frac{1}{A_t} + \frac{\epsilon_t}{\delta}\theta_t \left[(1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t u''(Y_t) - v''\left(\frac{Y_t}{A_t}\right)\frac{1}{A_t}\right]\frac{Y_t}{A_t} \\
& + \frac{\epsilon_t}{\delta}\theta_t \left((1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t u'(Y_t) - v'\left(\frac{Y_t}{A_t}\right)\right)\frac{1}{A_t}
\end{aligned}
$$

for output and $\dot{\vartheta}_t = \delta \pi_t^w$ for the multiplier.

We now simplify the first condition, which will take the form of a targeting rule, as discussed in much of the classical optimal policy analysis in RANK. With isoelastic preferences, we have

$$
0 = Y_t^{-\gamma} - Y_t^\eta A_t^{-\eta-1} + \frac{\epsilon_t}{\delta}\theta_t \left((1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t}(1-\gamma) Y_t^{-\gamma} - (1+\eta) Y_t^\eta A_t^{-\eta-1}\right)
$$

Further rearranging yields

$$
0 = A_t^{1+\eta} - Y_t^{\gamma+\eta} + \frac{\epsilon_t}{\delta}\theta_t (1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t}(1-\gamma) A_t^{1+\eta} - \frac{\epsilon_t}{\delta}\vartheta_t (1+\eta) Y_t^{\gamma+\eta}
$$

or simply

$$
\left[1 + \frac{\epsilon_t}{\delta}\vartheta_t (1+\eta)\right]^{\frac{1}{\gamma+\eta}} Y_t = \left[\left(\frac{1}{1+\tau^L}\frac{\epsilon_t}{\epsilon_t - 1} + \frac{\epsilon_t}{\delta}\theta_t (1-\gamma)\right)(1+\tau^L)\frac{\epsilon_t - 1}{\epsilon_t} A_t^{1+\eta}\right]^{\frac{1}{\gamma+\eta}}
$$

Using the definition of natural output, we therefore have

$$
\left[1 + \frac{\epsilon_t}{\delta}\theta_t (1+\eta)\right]^{\frac{1}{\gamma+\eta}} Y_t = \left(\frac{1}{1+\tau^L}\frac{\epsilon_t}{\epsilon_t - 1} + \frac{\epsilon_t}{\delta}\theta_t (1-\gamma)\right)^{\frac{1}{\gamma+\eta}} \tilde{Y}_t
$$

which concludes the proof.

■

Importantly, the targeting rule of Proposition 31 echoes the result of the standard New Keynesian framework, that Divine Coincidence obtains unless there are cost-push shocks. In the presence of only productivity and demand shocks, the planner perceives no tradeoff between inflation and output.

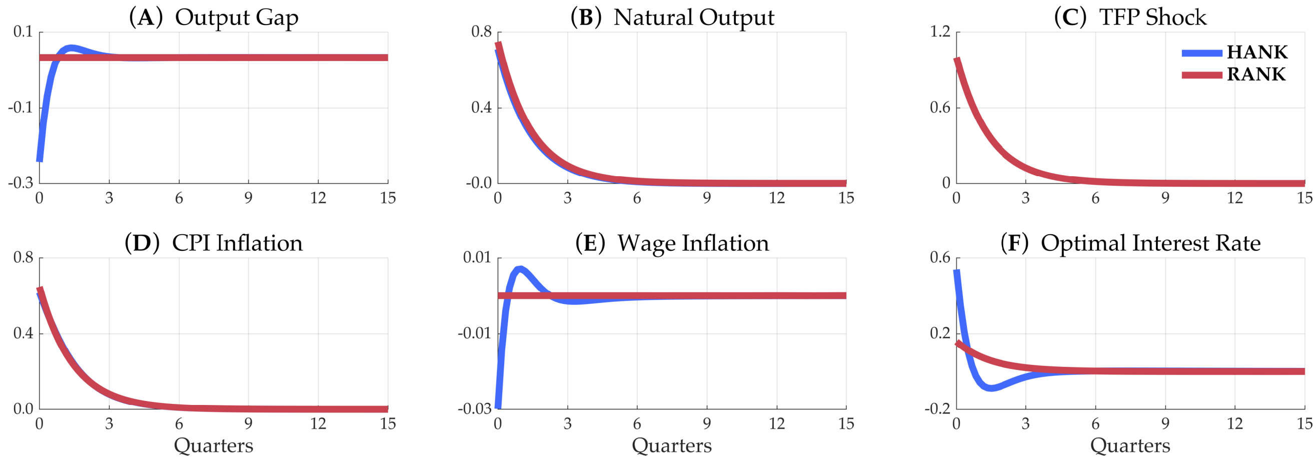


Figure 5. Optimal Policy Transition Dynamics: TFP Shock

Note. Transition dynamics after a positive TFP shock in both RANK (red) and HANK (blue) models under optimal monetary stabilization policy. Initial shock is 1% of steady state TFP and mean-reverts with a half-life of 1 quarter. Panels (**A**) through (**C**) report the dynamics of the output gap, $\frac{Y_t - \tilde{Y}_t}{\tilde{Y}_t}$, natural output, and the shock, all in percent deviations from the stationary Ramsey plan. Panels (**D**) through (**F**) report CPI inflation, wage inflation, and the optimal interest rate, all in percentage point deviations from the stationary Ramsey plan.

F Quantitative Analysis: Additional Results and Robustness

F.1 Productivity Shocks

We next turn to optimal stabilization policy in response to a TFP shock. Figure 5 reports the transition dynamics of the economy under optimal policy, while Figure 9 in Appendix F.4 reports those under a Taylor rule for comparison.

In both model benchmarks, natural output increases in response to a positive productivity shock. Natural output increases less than one-for-one, primarily due to diminishing marginal utility from consumption and convex disutility from labor. In HANK, natural output increases slightly less than in RANK as a result of union wage bargaining, which now features a distributional consideration.

Optimal stabilization policy in HANK follows the same principles as in RANK, with minor quantitative departures. The planner largely stabilizes both output and (wage) inflation gaps, but not fully. The planner allows both to become briefly negative on impact, before becoming positive and overshooting, yielding a hump-shaped response. The wage inflation gap on impact is small, reaching only -0.02%, and consequently not meaningfully different from 0. Compared to the response of wage inflation under a Taylor rule, where the wage inflation gap opens up to 0.4% under the same shock, this deviation from the Divine Coincidence benchmark of RANK should be

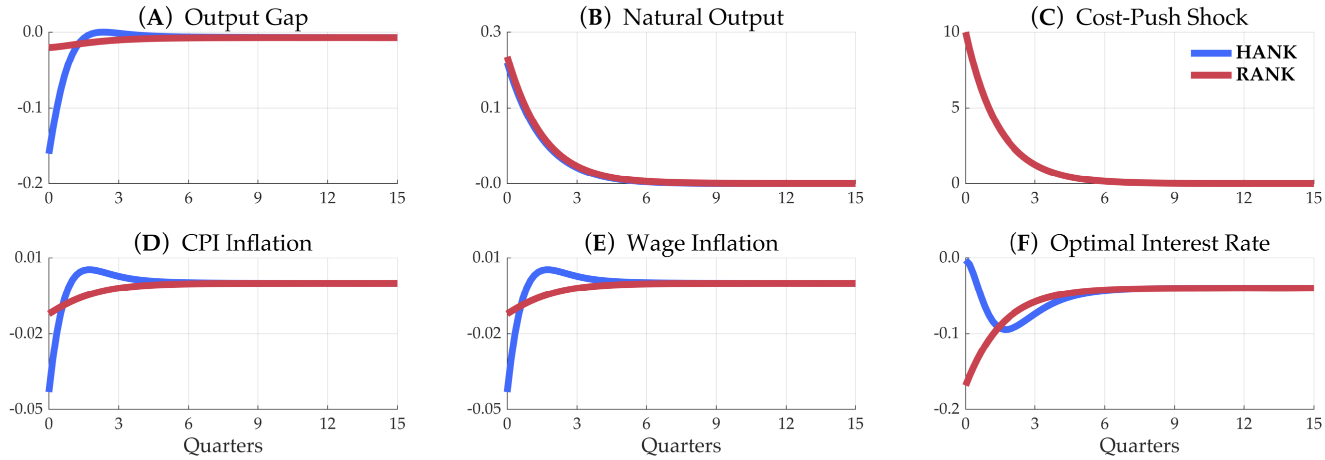


Figure 6. Optimal Policy Transition Dynamics: Cost-Push Shock

Note. Transition dynamics after a positive cost-push shock in both RANK (red) and HANK (blue) models under optimal monetary stabilization policy. The cost-push shock is modeled as an increase in labor union's desired wage mark-up. The shock is initialized at $\epsilon_0 = 11$ and mean-reverts to its steady state value $\epsilon = 10$, with a half-life of 1 quarter. Panels (**A**) through (**C**) report the dynamics of the output gap, $\frac{Y_t - \tilde{Y}_t}{\tilde{Y}_t}$, natural output, and the shock, all in percent deviations from the stationary Ramsey plan. Panels (**D**) through (**F**) report CPI inflation, wage inflation, and the optimal interest rate, all in percentage point deviations from the stationary Ramsey plan.

viewed as minimal. Similarly, while optimal policy stabilizes the output gap substantially relative to policy under the Taylor rule, the planner allows a small negative output gap to open up. The on-impact negative output gap under optimal policy is less than 20% of the size of the output gap under the Taylor rule.

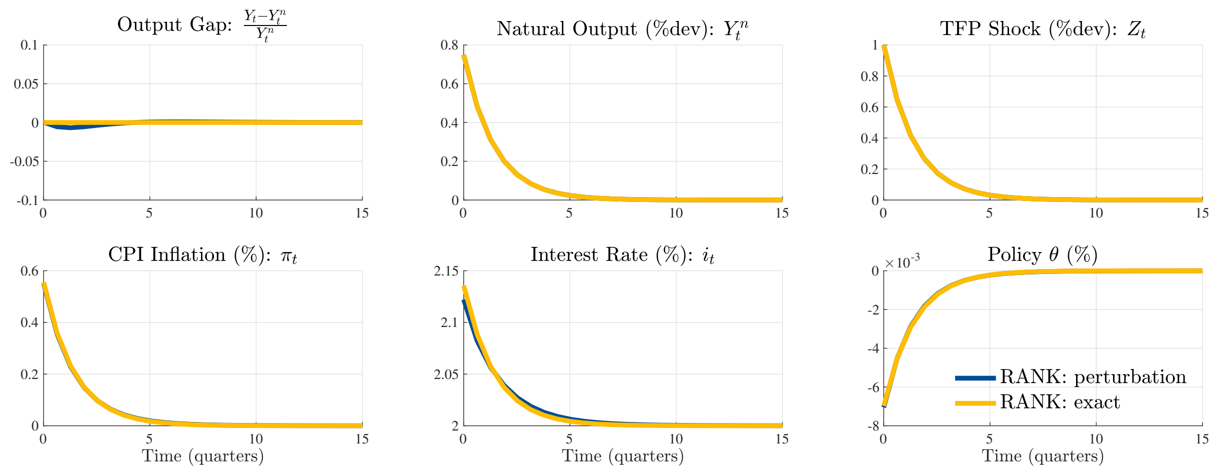
F.2 Cost-Push Shocks

Finally, we consider a cost-push shock under which the desired wage mark-up of labor unions changes and natural output increases by 0.25%. We report the transition dynamics under optimal policy in Figure 6, and also report the analogous transition dynamics under a Taylor rule in Figure 11 in Appendix F for comparison.

In RANK, Divine Coincidence fails in the presence of cost-push shocks and the planner now faces a tradeoff between inflation and output. Optimal stabilization policy is accommodative, lowering the nominal interest rate, but a small negative output gap still opens up.

In HANK, natural output again increase but slightly less due to distributional concerns in union bargaining. Monetary policy eases substantially less than in RANK, allowing a sizable negative output gap to open up. However, there is still substantial stabilization relative to the Taylor rule case. Especially inflation is again stabilized substantially.

Figure 7. Transition Dynamics with Optimal Policy: Perturbation vs. Exact Solution



Note. Impulse responses to positive TFP under optimal monetary policy in RANK. The Figure compares the exact analytical solution of optimal policy (yellow) against our numerical perturbation approach using sequence-space Hessians (blue).

F.3 Accuracy

In this section, we report a series of numerical tests to benchmark the accuracy of our perturbation method using sequence-space Hessians. In Figure 7, we compute the transition dynamics under optimal policy in RANK in response to a TFP shock using both our perturbation method and the exact analytical solution. The Figure underscores that our first-order perturbation method is highly accurate in the case of the baseline RANK model. The remaining error in the two solutions amounts to 0.01% in the output gap or, conversely, 1bps in the optimal interest rate response.

Likewise, Figure 8 reports the analogous comparison exercise for optimal policy in response to a demand shock in RANK. Here, the numerical error is even smaller. The discrepancy in optimal CPI inflation, for example, is on the order of 10^{-6}.

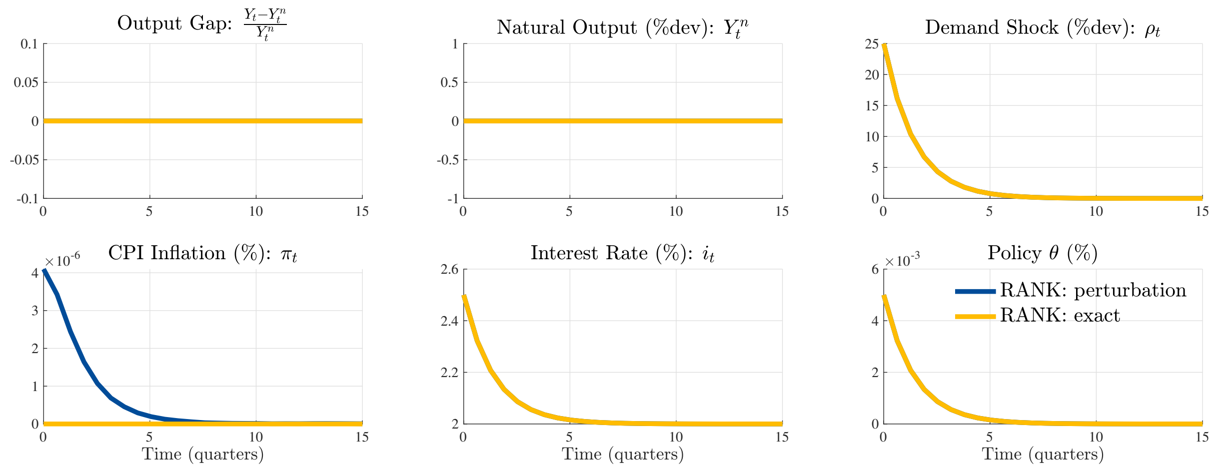
F.4 Transition Dynamics without Optimal Policy

In this section, we present impulse response plots that display the transition dynamics of both RANK and HANK economies in response to TFP, demand, and cost-push shocks without optimal policy interventions. We model monetary policy instead as following a Taylor rule, with

$$i_t = r^{ss} + \lambda_\pi \pi_t, \tag{89}$$

where we calibrate $\lambda_\pi = 1.5$.

Figure 8. Transition Dynamics with Optimal Policy: Perturbation vs. Exact Solution



Note. Impulse responses to positive demand under optimal monetary policy in RANK. The Figure compares the exact analytical solution of optimal policy (yellow) against our numerical perturbation approach using sequence-space Hessians (blue).

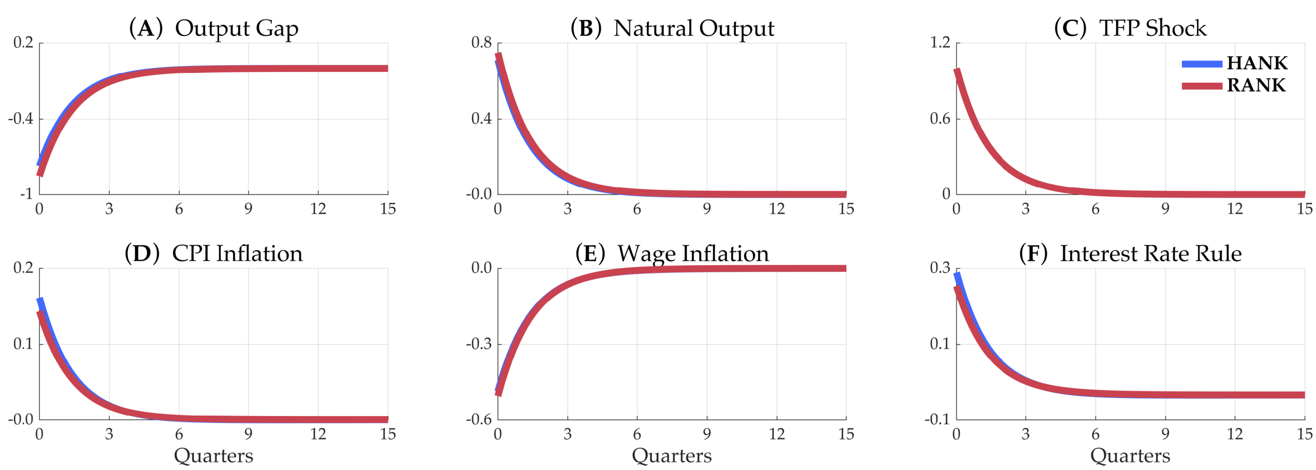


Figure 9. Transition Dynamics under Taylor Rule: TFP Shock

Note. Impulse responses to positive cost-push shock in both RANK (yellow) and HANK (blue) models. The nominal interest rate follows the Taylor rule (89) and is not set optimally. The cost-push shock is modeled as an increase in labor union's desired wage mark-up. The shock is initialized at $\epsilon_0 = 11$ and mean-reverts to its steady state value $\epsilon = 10$, with a half-life of 2 quarters.

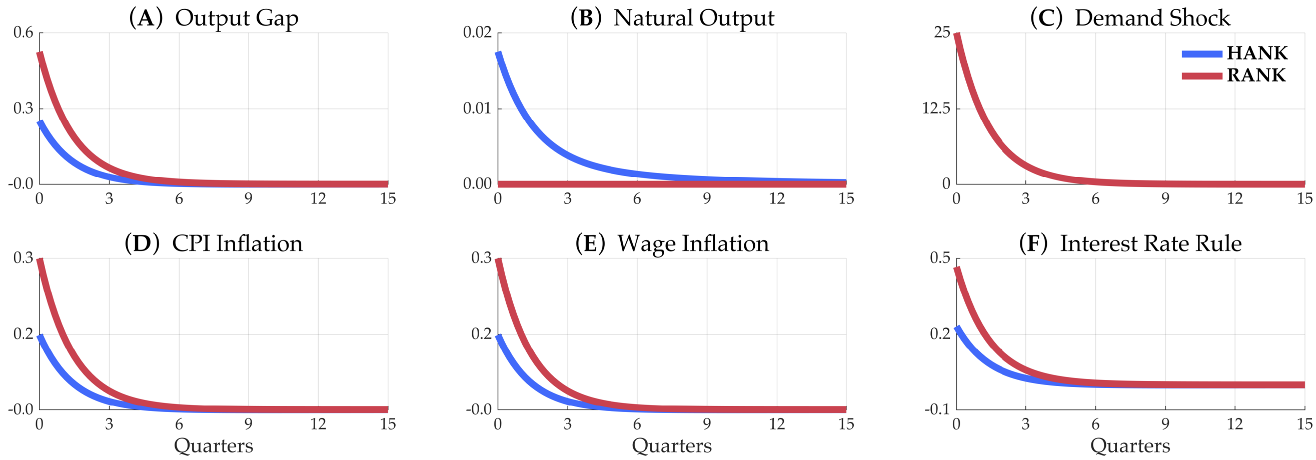


Figure 10. Transition Dynamics under Taylor Rule: Demand Shock

Note. Impulse responses to positive cost-push shock in both RANK (yellow) and HANK (blue) models. The nominal interest rate follows the Taylor rule (89) and is not set optimally. The cost-push shock is modeled as an increase in labor union's desired wage mark-up. The shock is initialized at $\epsilon_0 = 11$ and mean-reverts to its steady state value $\epsilon = 10$, with a half-life of 2 quarters.

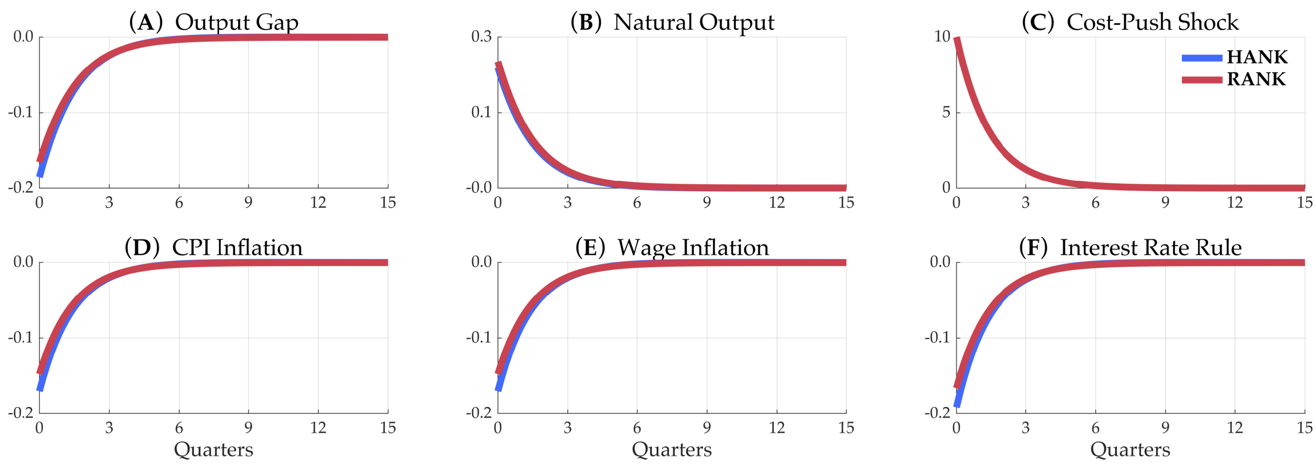


Figure 11. Transition Dynamics under Taylor Rule: Cost-Push Shock

Note. Impulse responses to positive cost-push shock in both RANK (yellow) and HANK (blue) models. The nominal interest rate follows the Taylor rule (89) and is not set optimally. The cost-push shock is modeled as an increase in labor union's desired wage mark-up. The shock is initialized at $\epsilon_0 = 11$ and mean-reverts to its steady state value $\epsilon = 10$, with a half-life of 2 quarters.