Monetary Policy without Redistributive Concerns*

David Baqaee
UCLA

Ariel Burstein
UCLA

Joao Guerreiro
UCLA
FRB Minneapolis

March 26, 2026

Preliminary and Incomplete
Please do not circulate

Abstract

We study optimal monetary policy in a general heterogeneous-agent New Keynesian (HANK) model under a strict efficiency mandate, evaluated using the aggregate-productivity measure of Baqaee and Burstein (2025a). We derive a second-order approximation to the objective that can be expressed as a sum of Harberger triangles. This objective nests the standard New Keynesian loss function and augments it with terms reflecting deadweight losses due to incomplete markets, person-specific labor wedges, and cyclical income risk. We show that, in sticky-price and sticky-wage HANK models with acyclical income risk, incomplete-market losses are locally orthogonal to monetary policy, so optimal policy coincides with the representative-agent New Keynesian (RANK) benchmark. In richer environments with cyclical income risk, however, optimal policy departs from RANK in the short run, but in the long run converges to a dual-mandate policy that places greater weight on output-gap stabilization than the RANK benchmark does.

Keywords: Monetary Policy, Heterogeneous Agents, New Keynesian Models, Efficiency, Aggregate Productivity.

JEL Codes: E3, E52, E61.

*The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

Recent work on HANK models has shown that monetary policy affects not only aggregate demand and inflation, but also the cross-sectional distribution of wealth, income, and consumption (e.g., Kaplan et al., 2018). This has led a growing literature to study optimal monetary policy in HANK economies using explicit social welfare functions. A central conclusion in that literature is that optimal policy can differ materially from the familiar inflation-output stabilization trade-off in RANK models, because monetary policy operates through distributional as well as aggregate channels (see the related literature section).

This paper asks the following question: how much does household heterogeneity change optimal monetary policy when the central bank is concerned only with economic efficiency, not redistribution? In heterogeneous-agent environments, social-welfare-based analyses generally combine two distinct motives: improving aggregate efficiency and redistributing resources across households. For example, under a "behind-the-veil-of-ignorance" utilitarian social welfare function, the policymaker values policies partly because they compress cross-sectional differences in marginal utility and, when consumption and leisure are separable, this creates pressure toward equalizing consumption across households. Monetary policy may therefore be used to redistribute across ex-post heterogeneous households even when the underlying inequality is not itself an inefficiency that monetary policy can correct. Optimal policy then depends not only on stabilization considerations, but also on subjective normative judgments about the desirable distribution of resources across households.

To isolate the efficiency component of the problem, we evaluate monetary policy using a form of cost-benefit analysis. Cost-benefit analysis, which is commonly used outside of monetary economics, measures the dollar-equivalent gains and losses from a policy for each agent and sums them across agents (Harberger, 1971). The resulting object is the surplus created by the policy: the gains net of the compensation required to make losers whole. Since monetary policy is inherently a general-equilibrium problem, we apply the general-equilibrium formulation developed by Baqaee and Burstein (2025a) and ask: if a central bank is given a strict *Efficiency Mandate*—tasked solely with maximizing aggregate surplus—how does household heterogeneity change optimal monetary policy relative to the RANK benchmark?

Our criterion is *Aggregate Productivity*, a money-metric-style measure of surplus in general equilibrium. Formally, we ask by how much aggregate factor endowments could be proportionally reduced after a monetary reform, while still allowing winners to com-

pensate losers through one-time lump-sum transfers that leave every household as well off as under the status quo. A policy with higher aggregate productivity creates a larger surplus. This criterion separates efficiency from redistribution by construction: it asks how large the pie can be made, not how it should be divided. It is also invariant to monotone transformations of individual utility functions, so it does not require cardinal interpersonal comparisons. Although related to social-welfare-based evaluations with transfers, it is generally distinct from them in second-best environments, where the policymaker cannot close all wedges and transfers cannot implement the technological frontier (Baqaee et al., 2026).

Our main result is that, under this criterion, household heterogeneity matters less for optimal monetary policy than existing HANK welfare analyses might suggest. Our central results can be summarized as follows. First, we derive a second-order approximation of aggregate productivity that decomposes efficiency losses into a sum of generalized Harberger (1964) triangles, one for each source of inefficiency. This approximation nests the standard New Keynesian quadratic loss function and augments it with additional terms capturing person-specific labor wedges, incomplete-market distortions, and their interaction under cyclical income risk.

Second, we show that minimizing this quadratic loss function subject to linearized equilibrium equations yields a first-order approximation to optimal policy. In canonical sticky-price and sticky-wage HANK models, with acyclical income risk, incomplete-market losses are locally orthogonal to monetary policy, so the optimal policy coincides with the RANK benchmark. Heterogeneity alters the efficient stabilization problem only when aggregate fluctuations have heterogeneous labor-income incidence. In these richer sticky-wage environments with cyclical income risk, optimal policy deviates from RANK in the short run. However, asymptotically, the optimal policy converges to a dual-mandate rule that is isomorphic to RANK but with a larger coefficient on the output gap. This larger coefficient reflects the fact that output gap fluctuations can generate additional efficiency losses through dispersed labor wedges.

We establish these results in a general HANK environment with nominal price and wage rigidities, incomplete financial markets, and cyclical labor-income risk—rich enough to nest the canonical sticky-price and sticky-wage HANK models as special cases. Aggregate demand depends on the distribution of household balance sheets, income expectations, and marginal propensities to consume, which we represent using the sequence-space Jacobian methods of Auclert et al. (2021, 2024). On the labor side, a flexible rationing rule governs how aggregate labor fluctuations are distributed across households. This rule is parameterized by a household-specific incidence elasticity, γ_h. When $\gamma_h = 0$

for all households, income risk is acyclical and our framework nests the canonical HANK specifications of Auclert et al. (2024, 2025). When γ_h varies across the income distribution, the model generates cyclical income risk consistent with the evidence in Guvenen et al. (2017). Throughout, we assume the policymaker has access to *time-0 transfers* only. This restriction prevents compensating transfers from undoing incomplete-market wedges period by period and allows us to isolate how market incompleteness shapes efficient stabilization.[1]

The paper makes three contributions.

The Efficiency Mandate. We adapt the general-equilibrium cost-benefit framework of Baqaee and Burstein (2025a) to monetary policy in HANK economies. The resulting objective—aggregate productivity—is the general-equilibrium analogue of the classic Kaldor-Hicks compensation test (Kaldor, 1939; Hicks, 1939). Two properties make it especially attractive as a policy criterion. First, the ranking of policies is invariant to monotone transformations of individual utility functions, so it does not require cardinal comparisons of well-being across households. Second, it is a pure efficiency measure: any policy that raises aggregate productivity generates a surplus large enough to make everyone better off, without taking a stance on how that surplus should ultimately be distributed. The central bank's problem under this mandate is simply to choose monetary policy to maximize aggregate productivity—to make the pie as large as possible. In a companion paper, Baqaee et al. (2026), we develop a general approximation theory for optimal policy in distorted heterogeneous-agent economies. Here we apply that framework to a monetary economy with nominal rigidities and incomplete markets.

Harberger-triangle representation of aggregate productivity. We derive a second-order approximation of aggregate productivity around an efficient benchmark allocation: a zero-inflation steady state with no aggregate or idiosyncratic shocks, so that all wedges vanish. We show that approximate aggregate productivity can be written as the distance to the technological frontier minus a sum of generalized Harberger triangles, one for each source of inefficiency in the economy. In our environment these inefficiencies come from price-markup wedges, wage-markup wedges, person-specific labor wedges induced by heterogeneous labor-income incidence, and incomplete-markets wedges. This representation nests the standard New Keynesian quadratic loss function and augments it with additional terms capturing labor-income incidence heterogeneity, incomplete markets,

[1] Appendix C considers *state-contingent transfers*. With a sufficiently rich transfer system, the planner can compensate households in every period and state and thereby eliminate the incomplete-market distortions relevant for our analysis. In that case, the optimal policy coincides with the RANK benchmark.

and their interaction under cyclical income risk. When labor-income incidence is homogeneous, the approximation collapses to the standard RANK loss function plus an additive incomplete-market term which is unaffected by policy. When incidence is heterogeneous, output-gap fluctuations become more costly because they induce inefficient dispersion in household labor wedges and interact with incomplete-market distortions. Because the approximation is quadratic in the usual linearized equilibrium variables, the optimal policy can be characterized with standard linear-quadratic methods while retaining a clear efficiency interpretation. This ensures that our policy conclusions inherit the transparency and tractability of the RANK optimal-policy framework while correctly accounting for the efficiency costs of heterogeneity and market incompleteness.

Optimal policy under the Efficiency Mandate. We characterize optimal monetary policy under this criterion. We obtain three substantive results. First, in the sticky-price HANK model and in the canonical sticky-wage HANK model with acyclical income risk, incomplete-market losses are locally orthogonal to monetary policy around the efficient allocation. Monetary policy therefore cannot reduce these losses at first order, and the efficient-mandate optimal policy coincides with the optimal policy in the RANK benchmark.

Second, this orthogonality result breaks down in sticky-wage HANK models with cyclical income risk. We derive the optimal targeting criterion in this environment. When aggregate labor fluctuations are borne disproportionately by some households, output-gap fluctuations create additional efficiency losses through the dispersion of person-specific labor wedges and through their interaction with incomplete-market distortions, which lead to two changes in the optimal targeting criterion: (i) due to the dispersion of labor wedges, the optimal policy places relatively more weight on output-gap stabilization than in RANK; and, (ii) in the short run, the optimal policy also deviates from RANK because it internalizes the interaction between cyclical income risk and incomplete-markets wedges. This term depends on the covariance between the incidence γ_{lt} and the incomplete-market wedges. When the covariance is positive, the optimal policy is to expand output. When the covariance is negative, the optimal policy is to contract output.

Our final result establishes a long-run equivalence: the efficiency-mandate optimal policy converges asymptotically to a modified-RANK dual-mandate problem with the standard loadings on price and wage inflation but a higher effective loading on the output gap. Formally, we establish that the incentives to deviate from a zero output gap due to the covariance between incomplete-markets wedges and labor-market incidence vanish in the long-run. This result holds in the model with both nominal price and wage

stickiness. So, the efficiency-mandate problem delivers policy prescriptions that converge to the modified-RANK dual-mandate problem.

We quantify these forces in a sticky-wage HANK model calibrated to the U.S. economy. We use the worker-level earnings betas from Guvenen et al. (2017) to discipline the cyclicality of labor-income risk. In this calibration, the covariance between the incidence elasticity γ_h and the incomplete-market wedges is positive. So, the optimal policy generates a quantitatively small expansion of output initially, which asymptotically converges to zero. Furthermore, the dispersion in labor-income incidence raises the effective targeting-criterion coefficient on output stabilization by 45 percent relative to RANK. This force for greater output stabilization is quantitatively important: following a cost-push shock, the efficiency-mandate HANK policy delivers 10.7 percent more cumulative output stabilization than the RANK-optimal policy, at the cost of higher inflation. We also derive closed-form solutions for simple rules in the spirit of Woodford (1999); these rules imply the same qualitative lesson, namely that cyclical labor-income incidence strengthens the efficient case for stabilizing the output gap. Under the optimal simple rules response to a cost-push shock, the output-gap response is 17% smaller and the wage-inflation response is 22% larger in HANK than under the corresponding RANK rule.

1.1 Related literature

Our analysis of optimal policy is grounded in the standard New Keynesian tradition. In that approach, one studies policy in a linear-quadratic environment: equilibrium conditions are linearized around an efficient equilibrium and the objective function is approximated to second order. This framework delivers transparent welfare-based targets and has become the workhorse for characterizing the inflation-activity trade-off in standard sticky-price RANK environments, e.g., Clarida et al. (1999), Benigno and Woodford (2003), Woodford (2003), Schmitt-Grohé and Uribe (2004), and Galí (2015); in RANK models with both sticky prices and wages, e.g., Erceg et al. (2000); in multi-sector New Keynesian environments, e.g., Aoki (2001), Rubbo (2023), and La'O and Tahbaz-Salehi (2022); in open-economy NK models, e.g., Gali and Monacelli (2005, 2008); in models with unemployment, e.g., Blanchard and Galí (2010); in models with credit market frictions, e.g., Monacelli (2008) and Cúrdia and Woodford (2016); in limited information/behavioral extensions of the NK model, e.g., Ball et al. (2005), Gabaix (2020), and Woodford and Xie (2022); and in HANK models, e.g., Bilbiie (2008, 2025), Bilbiie and Ragot (2021), and McKay and Wolf (2022a). We adopt this same discipline and extend it to heterogeneous-agent settings under an Efficiency Mandate.

A large body of work studies optimal monetary policy in HANK economies under a social welfare criterion. These contributions show that the planner internalizes the effects of inflation and real interest rates on risk sharing, consumption dispersion, and borrowers–savers redistribution, generating policy prescriptions that differ materially from conventional inflation-output stabilization trade-offs (Smirnov, 2022; Acharya et al., 2023; Bhandari et al., 2021; Dávila and Schaab, 2023; McKay and Wolf, 2022b; Nuño and Thomas, 2020; Bilbiie, 2008; Bilbiie and Ragot, 2021; Bilbiie, 2025; La'O and Morrison, 2024; Le Grand et al., 2022). Depending on commitment versus discretion and the availability of fiscal instruments, these distributional channels can imply an inflationary bias, larger state-contingent policy responses, or a sizable redistributive component of optimal policy. Our paper is deliberately different in scope from this body of work: we shut down redistribution concerns by construction and focus exclusively on efficiency.

Our approach to measuring aggregate productivity is developed in Baqaee and Burstein (2025a), who build on ideas from Allais (1979), Debreu (1951), and Luenberger (1996). Baqaee and Burstein (2025b) also use this approach to quantify the efficiency costs of incomplete markets and Baqaee and Burstein (2026) study aggregate productivity in efficient economies with discrete choice, such as location and occupation choices.

Paper Outline. In Section 2, we introduce our general HANK environment. In Section 3, we characterize the Efficiency Mandate and in Section 4 we derive a second-order approximation of the policymaker's objective function. In Section 5, we characterize the optimal policy in our general HANK environment and in the sticky-price and wage only limits. In Section 6 we quantify the optimal policy in a calibrated sticky-wage HANK model. Finally, Section 7 concludes.

2 A New Keynesian Model with Heterogeneous Agents

In this section, we introduce a class of models integrating heterogeneous agents, nominal rigidities, and incomplete markets. Our class allows both for nominal price and wage rigidities, general financial constraints limiting households' ability to smooth consumption, and endogenous cyclical income risk.

Environment. Time is discrete and infinite, indexed by $t = 0, 1, 2, ...$ There is a continuum of households indexed by $h \in [0, 1]$, final goods producers, intermediate-goods firms indexed by $i \in [0, 1]$, a continuum of unions indexed by $u \in [0, 1]$, and a central bank.

In each period, households draw idiosyncratic productivity shocks $z_{h,t}$ from a finite-state Markov process with transition matrix Π, where $\Pi(z, z')$ denotes the probability of transitioning between productivity state z today and z' tomorrow. The idiosyncratic shocks are independent across households. We let $\pi(z)$ denote the stationary distribution of the idiosyncratic shock, and we normalize the productivity shocks such that $\sum \pi(z)z = 1$.

Households are infinitely lived and have preferences over consumption $c_{h,t}$ and leisure $l_{h,t}$. Let $\mathbf{x}_h = \{c_{h,t}, l_{h,t}\}$ denote the market-goods and leisure consumption bundle for household h. The household preferences are given by

$$U(\mathbf{x}_h) \equiv (1-\beta)\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1+\chi} u(c_{h,t}) + \frac{\chi}{1+\chi} u(l_{h,t}) \right], \tag{1}$$

where $u(c) = \log c$, $\beta \in (0,1)$ is the discount factor, and $\chi > 0$ is a parameter governing the relative weight of leisure in household preferences.[2] Each household is endowed with 1 unit of time in each period, so they supply $z_{h,t}(1 - l_{h,t})$ units of labor services.

The final good is produced by a perfectly competitive final goods producer who combines intermediate goods according to a constant elasticity of substitution (CES) production function:

$$Y_t = \left(\int y_{i,t}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{2}$$

where $\varepsilon > 0$ is the elasticity of substitution between intermediate varieties.

Intermediate goods are produced by monopolistically competitive firms that combine labor from the unions according to a CES technology:

$$y_{i,t} = Z_t \left(\int n_{i,u,t}^{\frac{\psi-1}{\psi}} du \right)^{\frac{\psi}{\psi-1}}, \tag{3}$$

where Z_t denotes an aggregate productivity shock, $n_{i,u,t}$ denotes the quantity of labor services demanded by firm i from union u, and $\psi > 0$ is the elasticity of substitution between different types of labor. In the tradition of the sticky wage NK literature, we assume that each labor variety is supplied by a monopolistically competitive union who demands labor hours from each household (e.g., Erceg et al., 2000, Schmitt-Grohé and Uribe, 2005, and Auclert et al., 2024). The demand for labor services from union u is given by $n_{u,t}^d \equiv \int n_{i,u,t} di$. Both the final goods producer and the intermediate goods firms are

[2] In the baseline model, we assume that households have log utility over consumption and leisure. This assumption simplifies the exposition, but is not necessary for our main results. We extend our analysis beyond log-utility in Appendix D.

subject to nominal rigidities. We model nominal rigidities using a Calvo (1983)-pricing structure: in each period, each intermediate goods firm can only adjust its price with probability $1-\theta_p$, and each union can only adjust its wage with probability $1-\theta_w$.

In equilibrium, the total supply of labor services by households must equal the total demand for labor services by intermediate goods firms

$$\int \int n_{i,u,t} di du = 1 - L_t, \tag{4}$$

where $1 - L_t \equiv \int z_{h,t}(1 - l_{h,t})dh$ denotes aggregate labor supply in efficiency units. Furthermore, total output must equal total consumption

$$Y_t = C_t, \tag{5}$$

where $C_t \equiv \int c_{h,t} dh$ is aggregate consumption.

We now describe the problem of each agent in turn.

Final Goods Producer. The final goods producer takes the price of the final good as given and chooses how much of each intermediate good to demand. The problem of the final goods producer is given by

$$\max_{y_{i,t}} P_t Y_t - \int p_{i,t} y_{i,t} di, \tag{6}$$

where P_t is the price of the final good and $p_{i,t}$ is the price of intermediate good i. The solution implies that the demand for each intermediate good is given by

$$y_{i,t} = \left(\frac{p_{i,t}}{P_t}\right)^{-\varepsilon} Y_t, \tag{7}$$

where

$$P_t = \left(\int p_{i,t}^{1-\varepsilon} di\right)^{\frac{1}{1-\varepsilon}}. \tag{8}$$

Intermediate Goods Firms. Intermediate goods firms are monopolistically competitive in the product market and take the wage of each labor variety as given. The monopolist of variety i chooses how much of each labor variety to demand to minimize the cost of

producing. The cost minimization problem is

$$\min_{n_{i,u,t}} \int w_{u,t} n_{i,u,t} du \quad \text{subject to } y_{i,t} = Z_t \left(\int n_{i,u,t}^{\frac{\psi-1}{\psi}} du \right)^{\frac{\psi}{\psi-1}},$$

where $w_{u,t}$ is the wage of labor variety u. The solution to this problem implies that the demand for each labor variety is given by

$$n_{i,u,t} = \left(\frac{w_{u,t}}{W_t} \right)^{-\psi} \frac{y_{i,t}}{Z_t}, \tag{9}$$

where

$$W_t = \left(\int w_{u,t}^{1-\psi} du \right)^{\frac{1}{1-\psi}}. \tag{10}$$

denotes the wage index of labor services.

In each period t, a random θ_p fraction of intermediate goods firms are not allowed to adjust their price. Each firm that can adjust its price chooses the price that maximizes the expected discounted stream of profits until the next time it can adjust its price. The objective function of a firm that can adjust its price in period t is given by

$$\max_{p_{i,t}} \mathbb{E}_t \sum_{k=0}^{\infty} \tilde{Q}_{t,t+k} \theta_p^k \left[\Big((1 + \tau_{p,t+k}) p_{i,t} - MC_{t+k} \Big) y_{i,t+k} \right], \tag{11}$$

where $y_{i,t+k} = \left(\frac{p_{i,t}}{P_{t+k}} \right)^{-\varepsilon} Y_{t+k}$ is the demand for good i, $MC_t \equiv \frac{W_t}{Z_t}$ is the nominal marginal cost of producing good in period t, and $\tilde{Q}_{t,t+k}$ denotes the market discount factor. Following the standard approach in the NK literature, we introduce a subsidy to sales $\tau_{p,t}$, which allows us both to undo the monopoly distortion in steady state and to model cost-push shocks. We assume that the subsidy is contemporaneously financed by a lump-sum tax of the same value as the subsidy: $\tau_{p,t} p_{i,t} y_{i,t}$ to ensure the budget neutrality in each period. We assume that $\tau_p = \frac{1}{\varepsilon - 1}$ in steady state, so that the steady state allocation is efficient.

The first-order condition of this problem implies that the optimal price of a resetting firm is given by:

$$p_t^* = \frac{\varepsilon}{\varepsilon - 1} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} \tilde{Q}_{t,t+k} \theta_p^k P_{t+k}^{1-\varepsilon} Y_{t+k} \frac{W_{t+k}}{Z_{t+k} P_{t+k}}}{\mathbb{E}_t \sum_{k=0}^{\infty} \tilde{Q}_{t,t+k} (1 + \tau_{p,t+k}) \theta_p^k P_{t+k}^{\varepsilon} Y_{t+k}}. \tag{12}$$

Since final-goods producers make zero profits, total profits in the economy are given

by

$$\Xi_t \equiv \int \left[p_{i,t} - MC_{i,t}\right] y_{i,t} di. \tag{13}$$

Household Problem. Households begin the period with financial assets inherited from the previous period and draw a new idiosyncratic productivity. We take a flexible model of financial markets, allowing households to trade risk-free nominal bonds and a set of additional assets that allow them to insure against aggregate risks.

We let i_t denote the nominal interest rate on the risk-free bond, and $q_{k,t}$ denote the price of one unit of asset k in period t, which has returns $R_{k,t+1}$. We let $a_{h,t}$ denote the quantity of risk-free bonds held by household h at the end of period t, and $\tilde{a}_{k,h,t}$ denote the quantity of asset k held by household h at the end of period t. The budget constraint of household h is given by

$$P_t c_{h,t} + \frac{a_{h,t}}{(1+i_t)(1+\delta_t)} + \sum_k q_{k,t}\tilde{a}_{k,h,t} = y_t(\zeta_{h,t}) + z_{h,t}\Xi_t + a_{h,t-1} + \sum_k R_{k,t}\tilde{a}_{k,h,t-1} - \tilde{t}_{h,t} + T_{h,t}, \tag{14}$$

where $y_t(\zeta_{h,t})$ is the labor income of household h, which depends on the household idiosyncratic characteristics $\zeta_{h,t} \equiv (z_{h,t}, a_{h,t-1}, \{\tilde{a}_{k,h,t-1}\}_k, T_{h,t})$, and Ξ_t denotes aggregate profits, which we assume are distributed to households in proportion to their productivity, so that the profit income of household h is given by $z_{h,t}\Xi_t$. Under our normalization $\int z_{h,t} dh = 1$, this distribution rule is budget-consistent in aggregate. Following the NK tradition, we introduce a "financial/risk shock" δ_t, modelled as an additional benefit of holding riskless claims (see, e.g., Smets and Wouters, 2007). To preserve budget neutrality, we assume that this shock is offset by a lump-sum tax $\tilde{t}_{h,t} = \frac{\delta_t}{1+\delta_t}\frac{a_{h,t}}{(1+i_t)}$ in each period.

The variable $T_{h,t}$ denotes a transfer to household h in period t, which we will use to implement compensatory transfers under alternative monetary-policy regimes. These transfers satisfy $\int T_{h,t} dh = 0$ for all t and are zero in the baseline equilibrium. These transfers allow us to characterize the set of feasible allocations under alternative monetary policies, which is a key step in defining the problem of the efficiency-maximizing central bank. Our baseline case is one in which the planner only has access to time-0 transfers, so that $T_{h,t} = 0$ for all $t > 0$. In Appendix C, we also consider the case where the planner has access to state-contingent transfers, so that $T_{h,t}$ can be non-zero for all t and can depend on the history of shocks up to period t.

In addition, we allow for arbitrary constraints on the household's choice of financial portfolios

$$\Omega(a_{h,t}, \{\tilde{a}_{k,h,t}\}_k) \geq 0, \tag{15}$$

for some vector-valued function $\Omega(\cdot)$. This allows us to capture a wide range of financial constraints, including borrowing constraints, collateral constraints, and portfolio constraints. We assume that $\Omega(\cdot)$ is continuously differentiable and that $\Omega(0, \{0\}_k) > 0$.[3]

Labor Income. Household h has labor income $y_t(\zeta_{h,t})$. Due to the presence of nominal wage rigidities, the household does not optimally choose its labor supply and takes labor income as exogenous. The household receives the wage $w_{u,t}$ from union u for the labor services it supplies to that union. The labor income of household h is given by

$$y_t(\zeta_{h,t}) = \int w_{u,t} z_{h,t} n^s_{h,u,t}(\zeta_{h,t}) du, \tag{16}$$

where $w_{u,t}$ is the wage of union u and $n^s_{h,u,t}$ is the number of hours supplied by household h to union u.

We assume that the labor hours of household h are given by

$$n^s_{h,u,t} = \Gamma(1 - L_t; \zeta_{h,t}) \int n_{i,u,t} di. \tag{17}$$

This specification allows us to capture a wide range of labor supply responses to aggregate shocks and to model the cyclical properties of income risk (see, e.g., Werning, 2015, Auclert and Rognlie, 2018, Alves et al., 2020, and McKay and Wolf, 2022a). The function Γ captures the differential response of labor supply of household h to aggregate shocks, which can be used to capture the empirical fact that lower income households are more exposed to aggregate risk, suffering larger income drops during recessions and larger income increases during booms (see Guvenen et al., 2017).

We assume that the function Γ satisfies $\int \Gamma(1 - L_t; \zeta_{h,t}) dh = 1$, so that the labor supply of each household aggregates up to the total labor demand by firms. Furthermore, we assume that, in the absence of aggregate shocks, Γ is such that the labor supply of each household is optimal.

Note that under these assumptions and using the conditions for labor demand, we can write household labor income as follows:

$$y_t(\zeta_{h,t}) = \tilde{W}_t \cdot (1 - L_t) \cdot z_{h,t} \cdot \Gamma\left(1 - L_t; \zeta_{h,t}\right), \tag{18}$$

where $\tilde{W}_t \equiv W_t^{1-\psi} \cdot \left(\int w_{u,t}^{-\psi} du \right)^{-1}$.

[3]This assumption ensures that the household can hold some financial debt, which rules out the zero liquidity limit.

Aggregate Demand The aggregate demand of the economy is given by

$$C_t = \int c_{h,t} dh. \tag{19}$$

As in Angeletos et al. (2025), we can represent the aggregate consumption function by a functional $\mathcal{C}$ such that

$$C_t = \mathcal{C}\left(\{\{\mathbb{E}_s[\Theta_{s+h}]\}_{h=0}^{\infty}\}_{s=-\infty}^{t}\right), \tag{20}$$

where $\Theta_t \equiv \{\tilde{W}_t/P_t, L_t, 1+i_t, 1+\delta_t, P_t/P_{t-1}, \Xi_t/P_t, \{q_k\}, \{T_{h,t}\}\}$ are all the variables that enter the consumption block of the model. Importantly, $\mathbb{E}_s[\Theta_{s+h}] = \Theta$ are constant at steady state for $s < 0$.

The dependence of aggregate demand on current fundamentals and current expectations is immediate from the dependence of current decisions on people's current expectations. The additional past expectations capture the influence of the endogenous state variables (e.g., asset distribution) on past behavior and therefore on past fundamentals and past expectations.

Note that we write the consumption function as a functional of the point expectations directly, disregarding uncertainty. Since we focus on linearized equilibrium conditions, where certainty equivalence holds, this is without loss of generality.

Unions. Each union u is a monopolist in the labor market for labor variety u. In each period, a random θ_w fraction of unions are not allowed to adjust their wage. Unions face the demand $n_{u,t}^d \equiv \int n_{i,u,t} di = \left(\frac{w_{u,t}}{W_t}\right)^{-\psi} \frac{\int y_{i,t} di}{Z_t}$ for their labor variety. As in McKay and Wolf (2022a), we assume that unions evaluate the benefits of a wage according to the marginal utility of average consumption $u'(C_t)$ and of the utility of aggregate leisure $u'(L_t)$. This assumption guarantees that the supply side of our economy remains close to the standard RANK model and leads to the same wage NK Phillips curve under our approximation. Alternatively, we could assume that unions evaluate the benefits according to a weighted average of the marginal utilities of their members. This alternative assumption would lead to a more complex model, but it would not change the qualitative results of our analysis.

Resetting unions choose $w_{u,t}$ to maximize

$$\mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k \left[u'\left(C_{t+k}\right)\left(1+\tau_{w,t+k}\right) \frac{w_{u,t} n_{u,t+k}^d}{P_{t+k}} - \chi u'\left(L_{t+k}\right) n_{u,t+k}^d \right].$$

As in the case of the intermediate goods producer, we assume that there is a transfer $\tau_{w,t}$

to undo the monopoly distortion in the labor market and that this is financed with income of the household, such that household only receive the pre-transfer income. We assume that $\tau_w = \frac{1}{\psi - 1}$ in steady state, so that the steady state allocation is efficient.

The first-order condition of this problem implies that the optimal reset wage is given by

$$w_t^* = \frac{\psi}{\psi - 1} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k \chi u' \left(L_{t+k}\right) \left(\frac{1}{W_{t+k}}\right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k u' \left(C_{t+k}\right) \frac{\left(1+\tau_{w,t+k}\right)}{P_{t+k}} \left(\frac{1}{W_{t+k}}\right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}. \tag{21}$$

Central Bank. We assume that the central bank has direct control of the nominal interest rate, $\mathcal{P} \equiv \{i_t\}_{t=0}^{\infty}$. In the next sections, we describe how the central bank chooses the nominal interest rate.

Aggregate shocks. The economy described features multiple standard sources of aggregate disturbances that can drive business cycles: (i) aggregate productivity Z_t; (ii) the cost-push shock to prices and wages, $\tau_{p,t}$ and $\tau_{w,t}$, respectively; and (iii) the financial shock δ_t.

Log-linearized equilibrium We now briefly characterize the log-linearized equilibrium conditions in the economy.

The linearized Phillips curves for price and wage inflation are standard and given by

$$\pi_t = \lambda_p \left(\hat{w}_t - \hat{z}_t\right) + \beta \mathbb{E}_t[\pi_{t+1}] + u_{p,t} \tag{22}$$

$$\pi_{w,t} = \lambda_w \left(\left\{ 1 + \nu^{-1} \right\} \left(\hat{y}_t - \hat{y}_t^*\right) - \left(\hat{w}_t - \hat{z}_t\right) \right) + \beta \mathbb{E}_t[\pi_{w,t+1}] + u_{w,t}, \tag{23}$$

where $\hat{y}_t^*$ is the efficient level of output, $\hat{z}_t$ is the aggregate productivity shock, and $u_{p,t}$ and $u_{w,t}$ are the cost-push shocks to prices and wages, respectively. The parameters λ_p and λ_w capture the slope of the price and wage Phillips curves, respectively, and depend on the degree of nominal rigidity in prices and wages.

As is standard, the combination of price and wage stickiness makes the real wage a state variable, with dynamics given by

$$\hat{w}_t = \hat{w}_{t-1} + \pi_{w,t} - \pi_t. \tag{24}$$

Finally, we can represent the aggregate demand block of the economy as a functional of the sequence of expectations of the fundamentals and policy variables. We leverage the sequence-space Jacobian approach of Auclert et al. (2021, 2024) to write the aggregate

demand block as a linear combination of the sequence of expectations of the fundamentals and policy variables, as follows. Let $d\mathbf{C}_{t:\infty} = [dC_t, dC_{t+1}, \ldots]$ be the vector of changes in consumption over time. Then,

$$\mathbb{E}_t[d\mathbf{C}_{t:\infty}] = \mathbb{E}_{t-1}[d\mathbf{C}_{t:\infty}] + \sum_{\Theta} \mathbf{M}^{\Theta} \left(\mathbb{E}_t \left[d\mathbf{\Theta}_{t:\infty} \right] - \mathbb{E}_{t-1} \left[d\mathbf{\Theta}_{t:\infty} \right] \right), \tag{25}$$

given $\mathbb{E}_{-1}[d\mathbf{C}_{t:\infty}] = \mathbb{E}_{-1}\left[d\mathbf{\Theta}_{t:\infty}\right] = 0$. Here, $\mathbf{M}^{\Theta}$ is the sequence-space Jacobian associated with variable Θ. The sequence-space Jacobian captures the effect of a change in the time-t expectation of variable Θ on the entire future path of consumption.

Finally, in equilibrium, market clearing in the goods market implies that

$$dC_t = dY_t. \tag{26}$$

3 The Efficiency-Mandate Central Bank

In this section, we describe the problem of a central bank with an efficiency-maximizing mandate. To do so, we first introduce the notion of aggregate productivity in our economy, and we then describe the central bank's problem.

Aggregate Productivity The key concept in our analysis is a definition of aggregate productivity for an economy with heterogeneous agents, which serves as the objective of the efficiency-maximizing central bank. In defining aggregate productivity, we follow recent developments in Baqaee and Burstein (2025a). This measure of aggregate productivity is defined as the largest shift in total factor-augmenting productivity that makes it feasible to leave every household at least as well off as under the status quo allocation. This construction maps arbitrary shocks to the economy into a TFP-equivalent change, and it can be interpreted as a general-equilibrium analogue of cost-benefit analysis, increasing whenever the gains to winners are more than sufficient to compensate the losers.

In order to define aggregate productivity, we assume that there is a factor-augmenting productivity level H that shifts the total amount of time available for market production and leisure. In the baseline economy, $H = 1$, but we will use H to measure the efficiency gains of different policies.

First, we fix a status-quo allocation $\mathbf{x}^0 \equiv \{c^0_{h,t}, l^0_{h,t}\}$, which is the result of the competitive equilibrium under the status quo monetary policy, $\mathcal{P}^0$. We then consider how the change in policy $\mathcal{P}$ affects the set of feasible allocations in the economy. The set of feasible allocations comprises the set of allocations that are a competitive equilibrium under a

new policy $\mathcal{P}$ and some transfers.

In general, the set of feasible allocations depends on the set of available transfers. We consider two alternatives.

Definition 1. Let $\mathcal{X}(H;\mathcal{P})$ denote the set of feasible consumption and leisure allocations, supported by an equilibrium with policy $\mathcal{P}$, factor-augmenting productivity level H, and for some possible time-0 transfers $\{T_{h,0}\}$ such that $\int T_{h,0}dh = 0$.[4]

Using the definition of the set of feasible allocations, we can now define aggregate productivity as follows.

Definition 2. Fix a status-quo allocation $\mathbf{x}^0 \equiv \{c^0_{h,t}, l^0_{h,t}\}$. Aggregate productivity, A, is measured by the maximum contraction of H such that every agent can be kept at least indifferent to $\mathbf{x}^0$. Formally,

$$A(\mathcal{P}) \equiv \max\left\{H > 0 : \text{there exists } \mathbf{x} \in \mathcal{X}(1/H;\mathcal{P}) \text{ with } U(\mathbf{x}_h) \geq U(\mathbf{x}^0_h), \text{ for all } h\right\}. \tag{27}$$

Note that H can be equivalently interpreted as a change in the number of available hours for market production and leisure, which gives A a natural interpretation as a measure of aggregate productivity. If $A = 1.1$, then under the new policy, the economy can achieve the same level of welfare for all households as under the status quo allocation, while only using 90% of the total time available.

By construction, this measure of efficiency does not require making any interpersonal comparisons of utility. It follows that $A(\mathcal{P})$ is a measure of the efficiency of the policy $\mathcal{P}$ relative to the status quo policy $\mathcal{P}^0$. If $A(\mathcal{P}) > 1$, then the policy $\mathcal{P}$ is more efficient than the status quo policy, since it can achieve the same level of welfare for all households while using fewer resources. If $A(\mathcal{P}) < 1$, then the policy $\mathcal{P}$ is less efficient than the status quo policy, since it requires more resources to achieve the same level of welfare for all households.

The Efficiency Mandate We now use the concept of aggregate productivity to define the problem of an efficiency-maximizing central bank. The central bank chooses a monetary policy $\mathcal{P}$ to maximize aggregate productivity relative to the status quo policy $\mathcal{P}^0$. Formally, the problem of the central bank is given by

$$\max_{\mathcal{P}} A(\mathcal{P}). \tag{28}$$

[4]Recall that we set $T_{h,t} = 0$ for all $t > 0$ in our baseline analysis. In Appendix C, we also consider the case in which the planner has access to state-contingent transfers.

Note that, as in Baqaee and Burstein (2025a), $A(\mathcal{P})$ can be computed as the solution to a compensated representative agent's equilibrium in this economy. In particular, let $\tilde{U}_h(\mathbf{x}_h)$ denote the homothetized utility of household h, defined as

$$U\left(\left\{\frac{c_{h,t}}{\tilde{U}_h(\mathbf{x}_h)}, \frac{l_{h,t}}{\tilde{U}_h(\mathbf{x}_h)}\right\}\right) = U(\mathbf{x}_h^0) \Leftrightarrow \tilde{U}_h(\mathbf{x}_h) = \frac{\exp\{U(\mathbf{x}_h)\}}{\exp\{U(\mathbf{x}_h^0)\}}. \tag{29}$$

The problem of the compensated agent is to maximize utility given by $\min_h \tilde{U}_h(\mathbf{x}_h)$, subject to the equilibrium conditions of the economy under policy $\mathcal{P}$ and for some set of transfers. The solution to this problem gives us the value of aggregate productivity as the equilibrium value of utility of the compensated agent

$$A(\mathcal{P}) = \max_{x \in \mathcal{X}(\mathcal{P})} \min_h \tilde{U}_h(\mathbf{x}_h).$$

This construction allows us to compute the efficiency-mandate optimal policy by solving a ficticious representative agent problem with the compensated agent's preferences. We now illustrate this construction in a simple representative-agent NK model (RANK).

Example 1 (Aggregate Efficiency in a Representative-Agent New Keynesian Model). To build intuition for the concept of aggregate productivity, we can consider a simple RANK model. In this model, there is a single household that supplies labor and consumes the final good, and no idiosyncratic productivity. Financial markets are complete and the household is not subject to borrowing constraints. In this case, the set of feasible allocations under a policy $\mathcal{P}$ is a singleton, since there is only one equilibrium allocations–no transfers with a single agent.

Following Baqaee and Burstein (2025a), we begin by constructing the compensated agent who has preferences given by the representative agent's homothetized utility. Let $\tilde{U}$ denote the homothetized utility:

$$U\left(\left\{\frac{C_t}{\tilde{U}\left(\{C_t, L_t\}\right)}, \frac{L_t}{\tilde{U}\left(\{C_t, L_t\}\right)}\right\}\right) = U\left(\left\{C_t^0, L_t^0\right\}\right) \Leftrightarrow \tilde{U}\left(\{C_t, L_t\}\right) = \frac{\exp\{U\left(\{C_t, L_t\}\right)\}}{\exp\{U\left(\left\{C_t^0, L_t^0\right\}\right)\}}.$$

The compensated equilibrium coincides with the competitive equilibrium under policy $\mathcal{P}$ and where there is a representative household with preferences given by $\tilde{U}$. Note that, in this case, the compensated agent's decisions coincide with those of the representative agent. Let $\{C_t(\mathcal{P}), L_t(\mathcal{P})\}$ denote the equilibrium under policy $\mathcal{P}$, then aggregate efficiency is given by

$$A(\mathcal{P}) = \frac{\exp\{U\left(\{C_t, L_t\}\right)\}}{\exp\{U\left(\left\{C_t^0, L_t^0\right\}\right)\}}.$$

Note that in this representative-agent model, the policy that maximizes aggregate efficiency

$$\arg\max_{\mathcal{P}} A(\mathcal{P}) = \arg\max_{\mathcal{P}} U\left(\{C_t(\mathcal{P}), L_t(\mathcal{P})\}\right)$$

is the same as the policy that maximizes the representative agent's utility. So, in the representative agent model, the efficiency-mandate central bank is equivalent to the textbook Ramsey approach, as in Woodford (2003) and Galí (2015).

Wedges and the Sources of Inefficiency. Before turning to the analysis of optimal monetary policy, it is useful to emphasize the sources of inefficiency in the economy. To identify the sources of inefficiencies, we follow Baqaee and Burstein (2025a), and consider an alternative decentralized economy with wedges that is equivalent to the compensated equilibrium described above. This step is also key in analyzing optimal monetary policy, since it allows us to translate the problem into the general setting developed in Baqaee et al. (2026).

The model features four key sources of inefficiency: (1) nominal price rigidity, (2) nominal wage rigidity, (3) the rationing rule Γ, and (4) the presence of incomplete markets and financial frictions.

First, the presence of sticky prices introduces heterogeneity in markups across firms due to the differences in the time since their last price adjustment. This heterogeneity leads to an inefficient allocation of production across firms. We denote the wedge markups by $\mu^p_{i,t}$, and they satisfy

$$\log \mu^p_{i,t} \equiv \log p_{i,t} - \log \frac{W_t}{Z_t}. \tag{30}$$

Second, the presence of sticky wages introduces heterogeneity in wages across unions also due to differences in the time since their last wage adjustment. These wedges create an inefficient allocation of labor across unions. We denote the wedge in wages by $\mu^w_{u,t}$, and they satisfy

$$\log \mu^w_{u,t} \equiv \log w_{u,t} - \log \tilde{W}_t. \tag{31}$$

Third, the rationing rule Γ can also create an inefficient allocation of labor across households, leading to heterogeneous labor wedges across households. We denote household h's labor wedge by $\mu^L_{h,t}$, and they satisfy

$$\log \mu^L_{h,t} \equiv \log \left(\frac{\chi u'(l_{h,t})}{u'\left(c_{h,t}\right)} \right) - \log \frac{\tilde{W}_t}{P_t}. \tag{32}$$

Jointly, the price, wage, and individual labor wedges also imply the standard aggregate labor wedge present in NK models.

Finally, as in Baqaee and Burstein (2025b), the presence of incomplete markets and financial frictions introduces inefficiencies in households allocation of consumption over time and across idiosyncratic states. We denote the consumption wedge of household h by $\mu_{h,t}^{IM}$, and they satisfy

$$\log \mu_{h,t}^{IM} \equiv \log \left(\frac{u'(c_{h,t})}{u'(C_t)} \right) - \log \left(\frac{u'(c_{h,0})}{u'(C_0)} \right). \tag{33}$$

Together, these four sets of wedges capture the sources of inefficiency in the economy. These wedges are endogenous and a result both of the underlying fundamentals and monetary policy. They are the key objects that the central bank needs to address to maximize aggregate productivity. For instance, pursuing a policy of inflation targeting reduces price-markup heterogeneity across firms, which is a key source of inefficiency in the product market. Pursuing a policy of wage inflation targeting reduces wage-markup heterogeneity, which is a key source of inefficiency in the labor market. Pursuing a policy that stabilizes output can affect the distribution of individual labor wedges across households, which are a key source of inefficiency in the allocation of labor across households. Finally, pursuing a policy that affects the distribution of income across households can affect the consumption wedges, which is a key source of inefficiency due to incomplete financial markets.

A Comparison to Distance to the Frontier Aggregate productivity, as defined above, is a measure of the efficiency of a policy relative to a given status quo policy. However, to the extent that the policy cannot close all of the wedges described above, the efficiency-mandate optimal policy will not be able to close the gap to the frontier, and there will be some residual inefficiency.

Definition 3 (Distance to the Frontier)**.** Let $\mathcal{X}^*(H)$ denote the set of allocations that are satisfy the resource constraints in the economy with factor-augmenting productivity level H. The distance to the frontier is given by the maximum aggregate productivity that can be achieved by any policy, i.e.,

$$\mathcal{D} \equiv \max \left\{ H > 0 : \text{there exists } \mathbf{x} \in \mathcal{X}^*(1/H) \text{ with } U(\mathbf{x}_h) \geq U(\mathbf{x}_h^0), \text{ for all } h \right\}. \tag{34}$$

The following proposition follows directly from Proposition 1 in Baqaee et al. (2026).

Proposition 1. *For any policy $\mathcal{P}$, aggregate productivity is bounded above by the distance to the frontier, i.e., $A(\mathcal{P}) \leq \mathcal{D}$. If there exists a policy $\mathcal{P}$ such that $A(\mathcal{P}) = \mathcal{D}$, then $\mathcal{P}$ is an optimal policy.*

For any status-quo allocation, the best possible solution is one which eliminates all sources of inefficiency, so that the economy operates at the frontier. However, in general, there is no guarantee that such a policy exists, and the efficiency-mandate optimal policy will be one that closes as much of the gap to the frontier as possible.

The compensated agent problem can also be used to characterize the distance to the frontier of any status-quo allocation. In particular, the distance to the frontier is given by the solution to the following problem:

$$\mathcal{D} = \max_{x \in \mathcal{X}^*} \min_{h} \tilde{U}_h(\mathbf{x}_h). \tag{35}$$

The next proposition establishes a key property of the frontier allocations, which are the solution to the problem of the compensated agent. For any status-quo allocation $\mathbf{x}^0$, the frontier allocations satisfy

$$Y_t^* = C_t^* = \frac{Z_t}{1+\chi}, \qquad L_t^* = \frac{\chi}{1+\chi}. \tag{36}$$

Furthermore, the distance to the frontier is given by

$$\mathcal{D} = \frac{\exp\{U(\{C_t^*, L_t^*\})\}}{\int \exp\{U(\mathbf{x}_h^0)\} dh}. \tag{37}$$

We denote the frontier aggregate allocations with superscript $*$ to highlight the fact that they are the "potential levels" in our economy, i.e., they are the levels that would obtain in the absence of any distortions.

The notion of distance to the frontier is also useful in characterizing the second-order approximation of aggregate productivity. This follows from the fact that it is useful to decompose aggregate productivity into the distance to the frontier and the gap that a given policy $\mathcal{P}$ still remains open relative to the frontier $A(\mathcal{P})/\mathcal{D} \in [0, 1]$. To see this fact in practice, we return to the RANK example above.

Example 2 (Approximate Aggregate Productivity in a RANK Model). In the tradition of the NK literature, it is common to consider an approximated version of the model, where the equilibrium conditions are approximated around an efficient allocation to first-order, and the objective is approximated to second order. In this RANK model, we consider

approximating $A(\mathcal{P})$ to second order around a zero-inflation steady state. In this case, we can show that the approximated version of $A(\mathcal{P})$ is given by

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1+\nu^{-1})(\hat{y}_t - \hat{y}_t^*)^2 + \frac{\varepsilon}{\lambda_p}(\pi_t)^2 + \frac{\psi}{\lambda_w}(\pi_{w,t})^2 \right],$$

where $\hat{x}_t = d\log(X_t)$ denote log-deviations of variable X from its steady-state value, $\pi_t = \hat{p}_t - \hat{p}_{t-1}$ and $\pi_{w,t} \equiv \hat{w}_t - \hat{w}_{t-1}$ denote the inflation rate and wage inflation rate, respectively, $\hat{y}_t^* \equiv \hat{z}_t$ is the natural level of output. The parameter $\nu = \frac{L}{1-L}$ is the Frisch elasticity of labor supply, $\omega_C \equiv C/Z$ denotes the consumption share, and $\lambda_p \equiv \frac{(1-\beta\theta_p)(1-\theta_p)}{\theta_p}$ and $\lambda_w \equiv \frac{(1-\beta\theta_w)(1-\theta_w)}{\theta_w}$ are the standard NK price and wage Phillips curve slopes, respectively.[5]

In this case, aggregate productivity is equal to the distance to the frontier minus the standard quadratic loss function of the NK model, with structural weights pinned down by preferences and the price and wage Phillips curve slopes. A non-zero output gap is associated with a non-zero labor wedge, which reduces aggregate productivity. Price inflation is associated with price markup dispersion, which reduces aggregate productivity. Wage inflation is associated with wage markup dispersion, which reduces aggregate productivity. The resulting objective function of the efficiency-mandate central bank is isomorphic to the canonical Woodford (2003) loss function.

---

4 Harberger (1964)-Triangles Representation of Aggregate Productivity

Aggregate productivity is a complex object to characterize, since it involves characterizing the solution to an incomplete-markets model under alternative policies $\mathcal{P}$. To make progress, we use the general results in Baqaee et al. (2026), who show how to characterize the "approximate optimal policy" around an efficient allocation. Effectively, this amounts to characterizing the optimal policy in a second-order approximation of the model around

[5]Note that defining the output gap, x_t as the difference between log-output and the flexible price level of output $\hat{y}_t^n$, we can write $x_t \equiv \hat{y}_t - \hat{y}_t^n$. We allow for the possibility that the steady state of the economy is distorted so $\hat{y}_t^n + x^* = \hat{y}_t^*$ (equivalently, a permanent cost-push shock). So, we can write aggregate productivity as:

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1+\nu^{-1})(x_t - x^*)^2 + \frac{\varepsilon}{\lambda_p}(\pi_t)^2 + \frac{\psi}{\lambda_w}(\pi_{w,t})^2 \right],$$

which is the same form as in Chapter 6 of Woodford (2003).

an efficient allocation, which is a standard approach in the NK literature.

In particular, we consider a second-order approximation of aggregate productivity around a steady-state allocation with zero inflation, no idiosyncratic productivity risk, and no aggregate shocks. This approximation ensures that, in the limiting economy, all wedges are zero and the allocation is efficient. The next proposition characterizes the second-order approximation of aggregate productivity in terms of the wedges defined above.

We begin by stating the key result in this section in proposition 2. We then provide a constructive proof of this key result. The proof shows how to obtain each term of the second-order approximation as a consequence of the specific wedges present in the economy, which allows us to give a clear economic interpretation to each term in the second-order approximation. In particular, we show that each term can be interpreted as a general-equilibrium analogue of a collection of Harberger triangles.

Proposition 2 (Second-Order Approximation of Aggregate Productivity)**.** *To a second-order approximation, aggregate productivity can be written as*

$$\begin{aligned}\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\Bigg\{ \sum_{t=0}^{\infty}\beta^t\Bigg\{ \left(1+\nu^{-1}\right)\left(\hat{y}_t-\hat{y}_t^*\right)^2 + \frac{\varepsilon}{\lambda_p}\left(\pi_t\right)^2 + \frac{\psi}{\lambda_w}\left(\pi_{w,t}\right)^2 \Bigg\} \\ + \nu^{-1} Var_{\chi_h}\left(\gamma_h\right)\sum_{t=0}^{\infty}\beta^t\left(\hat{y}_t-\hat{y}_t^*\right)^2 + \left(1-\beta\right) Var_{\chi_h}\left[\gamma_h\right]\left[\sum_t \beta^t(\hat{y}_t-\hat{y}_t^*)\right]^2 \\ + \sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right) - \sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\frac{c_{h,t}}{\bar{c}_h}, \gamma_h\right)\left(\hat{y}_t-\hat{y}_t^*\right)\Bigg\}.\end{aligned} \tag{38}$$

As we show next, the first line of the second-order approximation captures the losses from price and wage markups as well as the losses associated with the aggregate labor wedge. These losses coincide exactly with those in the RANK model. The second line captures the additional losses from the dispersion of labor wedges across households. These terms take the form of additional costs associated with a non-zero output gap. The third line captures the losses from incomplete markets. The first term in this expression coincides with the costs of incomplete markets in Baqaee and Burstein (2025b), while the second term captures the additional costs of incomplete markets in the presence of an heterogeneous distribution of labor wedges across households.

In the RANK model, there are complete markets and no heterogeneity across households. So, $Var_{\chi_h}(\gamma_h) = 0$ and $Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right) = 0$. In this case, the

second-order approximation of aggregate productivity reduces to the first line of the expression above, which coincides with the standard quadratic loss function derived in the RANK model in Example 1.

In the canonical HANK model, with acyclical income risk (see Auclert et al., 2024), $\gamma_h = 0$ for all h. In this case, the second-order approximation of aggregate productivity reduces to the distance to the frontier subtracting the RANK losses and the additional costs of incomplete markets, which coincide with those in Baqaee and Burstein (2025b).

More generally, in the presence of cyclical income risk, the second-order approximation also includes the additional terms capturing the losses from the dispersion of labor wedges across households, and the interaction between incomplete markets and the distribution of labor wedges. These terms capture the additional costs of cyclical income risk in a HANK model, which are not present in a RANK model or the canonical HANK model.

Deriving the Approximate Aggregate Productivity. To derive the second-order approximation of aggregate productivity, we use the general results in Baqaee et al. (2026), who show how to characterize the approximate optimal policy in a general equilibrium model with endogenous wedges. For this, it is useful to first characterize the approximated aggregate productivity in terms of the wedges defined above, and then to rewrite the resulting expression in a more interpretable way. The next proposition characterizes the second-order approximation of aggregate productivity in terms of the Harberger (1964) triangles associated with each wedge.

Lemma 1 (Aggregate Productivity as Harberger Triangles)**.** *To a second-order approximation, aggregate productivity can be written as*

$$\log A(\mathcal{P}) \approx \log \mathcal{D} + \frac{1}{2}\Big\{\Delta^{IM} + \Delta^{p} + \Delta^{w} + \Delta^{L}\Big\}, \tag{39}$$

where

i. The losses from incomplete markets are given by

$$\Delta^{IM} \equiv \mathbb{E}_0 \sum_t \beta^t \int_0^1 \chi_h \cdot d\log \mu_{h,t}^{IM} \cdot d\log x_{h,t}\, dh, \tag{40}$$

where $\log x_{h,t} \equiv \left[\frac{1}{1+\chi}\log c_{h,t} + \frac{\chi}{1+\chi}\log l_{h,t}\right]$ *is the effective consumption of household h in period t and χ_h denotes the consumption share of household h.*

ii. The losses from price markups are given by

$$\Delta^p \equiv \omega_C \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 d \log \mu_{i,t} \cdot \partial \log y_{i,t} \, di, \tag{41}$$

where $\omega_C \equiv C/Z$ is the consumption share.

iii. The losses from wage markups are given by

$$\Delta^w \equiv \omega_C \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 d \log \mu^w_{u,t} \cdot \partial \log n^d_{u,t} \, du. \tag{42}$$

iv. The losses from the person-specific labor wedges are given by

$$\Delta^L \equiv (1 - \omega_C) \cdot \mathbb{E}_0 \sum_{t} \beta^t \int_0^1 \chi_h \cdot d \log \mu^L_{h,t} \cdot \partial \log l_{h,t} \, dh. \tag{43}$$

where $\partial \log x_{h,t}$, $\partial \log y_{i,t}$, $\partial \log n^d_{u,t}$, and $\partial \log l_{h,t}$ denote the changes in the corresponding quantities caused by the wedges in the general-equilibrium with the compensated representative agent around the efficient allocation.

Lemma 1 shows that, to a second-order approximation, aggregate productivity can be written as the sum of four terms, each of which captures the losses from a different source of inefficiency in the economy. Each term is given by the expected discounted sum of the product of the change in the corresponding wedge and the change in the corresponding quantity. Hence, each term can be interpreted as a general-equilibrium analogue of a Harberger (1964) triangle, which captures the efficiency losses from a wedge in a partial equilibrium setting.

Lemma 1 relies on translating the results in Baqaee and Farhi (2024) and Baqaee and Burstein (2025a) to our setting. This result is the key step in characterizing the optimal policy, since it allows us to express aggregate productivity in terms of a set of interpretable wedges. In particular, by setting monetary policy, the central bank can affect the Harberger triangles. In this way, the problem of the efficiency-mandate central bank is to choose policy to minimize the losses associated with the Harberger triangles.

Our next set of results characterize how the different wedges respond to changes in aggregate variables, and ultimately monetary policy.

In the compensated equilibrium, the change in $y_{i,t}$ and $n^d_{u,t}$ caused by the wedges

satisfy

$$d\log y_{i,t} = -\varepsilon\left(d\log\mu^p_{i,t} - \int d\log\mu^p_{i',t}di'\right) + \partial\log Y_t,$$
$$d\log n^d_{u,t} = -\psi\left(d\log\mu^w_{u,t} - \int d\log\mu^w_{u',t}du'\right) + \partial\log Y_t.$$

Using these expressions, we can rewrite the Harberger triangles for price and wage markups using the fact that

$$\int_0^1 d\log\mu^p_{i,t}d\log y_{i,t}di = -\varepsilon Var_i(d\log\mu^p_{i,t}) + E_i[d\log\mu^p_{i,t}]\partial\log Y_t,$$
$$\int_0^1 d\log\mu^w_{u,t}d\log n^d_{u,t}du = -\psi Var_u(d\log\mu^w_{u,t}) + E_u[d\log\mu^w_{u,t}]\partial\log Y_t,$$

where $Var_i(d\log\mu^p_{i,t})$ and $Var_u(d\log\mu^w_{u,t})$ denote the cross-sectional variance of price and wage markups across firms and unions, respectively, $E_i[d\log\mu^p_{i,t}]$ and $E_u[d\log\mu^w_{u,t}]$ denote the cross-sectional average of price and wage markups across firms and unions, respectively, and $\partial\log Y_t$ represents the change in aggregate output relative to the efficient allocation. Using these expressions, we obtain the following result.

Lemma 2 (Characterizing the Harberger Triangles for Price and Wage Markups)**.** *The Harberger triangles for price and wage markups can be written as*

$$\Delta^p = -\omega_C\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\frac{\varepsilon}{\lambda_p}(\pi_t)^2 + \omega_C\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t E_i[d\log\mu^p_{i,t}]\partial\log Y_t$$

and

$$\Delta^w = -\omega_C\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\frac{\psi}{\lambda_w}(\pi_{w,t})^2 + \omega_C\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t E_u[d\log\mu^w_{u,t}]\partial\log Y_t$$

respectively, where π_t and $\pi_{w,t}$ are the price and wage inflation rates, respectively, and $\lambda_p \equiv (1-\beta\theta_p)(1-\theta_p)/\theta_p$ and $\lambda_w \equiv (1-\beta\theta_w)(1-\theta_w)/\theta_w$ are the price and wage Phillips curve slopes, respectively.

The proof for this result uses the well known fact that, with Calvo pricing rigidities, the variance of markups across firms (unions) is proportional to the square of accumulated inflation (wage inflation) appropriately weighed by the probability of price (wage) adjustment. Hence, the first term in the expression for Δ^p and Δ^w captures the familiar losses from price and wage dispersion caused by inflation, while the second term captures the losses from the average level of markups across firms and unions, which also make up the average labor wedge.

Turning to the person-specific labor wedges, we know that, in equilibrium, the change in $l_{h,t}$ is given by

$$\partial \log l_{h,t} = (1+\gamma_h)\partial \log L_t = -(1+\gamma_h)\frac{\omega_C}{1-\omega_C}\partial \log Y_t.$$

where $\gamma_h \equiv \frac{\partial \log \Gamma(1-L,\zeta_{h,0})}{d\log(1-L)}$ is the elasticity of the rationing rule with respect to aggregate labor supply. Using this expression, we can rewrite the Harberger triangle for person-specific labor wedges as

$$\Delta^L \equiv -\omega_C \cdot \mathbb{E}_0 \sum_t \beta^t \left\{ E_{\chi_h}\left[d\log \mu^L_{h,t}\right]\partial \log Y_t + Cov_{\chi_h}\left(d\log \mu^L_{h,t}, \gamma_h\right)\partial \log Y_t \right\}$$

where $E_{\chi_h}\left[d\log \mu^L_{h,t}\right]$ and $Cov_{\chi_h}\left(d\log \mu^L_{h,t}, \gamma_h\right)$ denote the cross-sectional average and covariance of the labor wedges across households weighted by the consumption share χ_h, respectively. The first term captures the losses from the average level of labor wedges across households, which also make up the average labor wedge, while the second term captures the losses from the heterogeneity in labor wedges across households, which is a key feature of HANK models.

Lemma 3 (Characterizing the Harberger Triangles for Person-Specific Labor Wedges)**.** *The Harberger triangle for person-specific labor wedges can be written as*

$$\begin{aligned}\Delta^L = -\omega_C \cdot \mathbb{E}_0 \sum_t \beta^t \Bigg[&\left[\left(1+\nu^{-1}\right)\partial \log Y_t + E_i\left[d\log \mu_{i,t}\right] + E_u\left[d\log \mu_{u,t}\right]\right]\cdot \partial \log Y_t \\ &- \nu^{-1} Var_{\chi_h}\left(\gamma_h\right)\left(\partial \log Y_t\right)^2\Bigg] - \omega_C \cdot Var_{\chi_h}\left[\gamma_h\right]\cdot(1-\beta)\,\mathbb{E}_0\left(\sum_t \beta^t \partial \log Y_t\right)^2. \quad (44)\end{aligned}$$

Note that the terms in the first row capture the effect of the average labor wedge, $E_{\chi_h}\left[d\log \mu^L_{h,t}\right]$, on aggregate productivity, while the terms in the second row capture how the rationing rule induces heterogeneity in the labor wedges across households, $Cov_{\chi_h}\left(d\log \mu^L_{h,t}, \gamma_h\right)$, which also lowers aggregate productivity. If $Var_{\chi_h}\left(\gamma_h\right) = 0$, then the second row is zero and the losses from labor wedges are only due to the average level of labor wedges across households. However, if $Var_{\chi_h}\left(\gamma_h\right) > 0$, then the heterogeneity in labor wedges across households also creates losses in aggregate productivity.

Note also that, ultimately, the productivity effects of average labor wedge are a consequence of the sum of the price and wage markups, minus the average person-specific labor wedge: $E_i[d\log \mu^p_{i,t}] + E_u[d\log \mu^w_{u,t}] - E_{\chi_h}[d\log \mu^L_{h,t}]$. These results also show that the

average labor wedge term is

$$E_i[d\log\mu^p_{i,t}] + E_u[d\log\mu^w_{u,t}] - E_{\chi_h}[d\log\mu^L_{h,t}] = -(1+\nu^{-1})(\partial\log Y_t)^2, \tag{45}$$

which is the familiar expression for the labor wedge in RANK models.

Let the lifetime average incomplete market and labor wedge for person h be given by

$$d\log\overline{\mu}^{IM}_h \equiv (1-\beta)\mathbb{E}_0\sum_t \beta^t d\log\mu^{IM}_{h,t}$$

and

$$d\log\overline{\mu}^{L}_h \equiv (1-\beta)\mathbb{E}_0\sum_t \beta^t d\log\mu^{L}_{h,t},$$

respectively. Using these expressions, we can rewrite the Harberger triangle for incomplete markets, solving out for $\partial\log x_{h,t}$, as

$$\Delta^{IM} = -\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\mu^{IM}_{h,t}, d\log\left(\frac{\mu^{IM}_{h,t}}{\overline{\mu}^{IM}_h}\right)\right) - (1-\omega_C)\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\mu^{IM}_{h,t}, d\log\left(\frac{\mu^{L}_{h,t}}{\overline{\mu}^{L}_h}\right)\right).$$

We now use this expression to characterize the aggregate productivity losses from incomplete markets and how they depend on monetary policy. We now use the fact that $d\log\mu^{IM}_{h,t} = -\log\left(\frac{c_{h,t}}{c_{h,0}}\right) + (\text{Terms independent of } h)$ to establish the following result.

Lemma 4. *The Harberger triangles for incomplete markets can be written as*

$$\Delta^{IM} = -\omega_C\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\overline{c}_h}\right)\right) + \omega_C\cdot\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\frac{c_{h,t}}{\overline{c}_h}, \gamma_h\right)\partial\log Y_t.$$

where $d\log\overline{c}_h = (1-\beta)\mathbb{E}_0\sum_t\beta^t d\log c_{h,t}$ *is the average lifetime consumption of household* h *and* $\gamma_h \equiv \frac{\partial\log\Gamma(1-L,\zeta_{h,0})}{d\log(1-L)}$ *is the elasticity of the rationing rule with respect to aggregate labor supply.*

This lemma decomposes the losses from incomplete markets into two components. The first component captures the losses from incomplete markets driven by idiosyncratic income risk, which is given by the covariance between the consumption of household h relative to its average lifetime consumption, $d\log\left(\frac{c_{h,t}}{\overline{c}_h}\right)$, and the consumption of house-

hold h relative to its initial consumption, $d \log \left(\frac{c_{h,t}}{c_{h,0}} \right)$. The second component captures the losses from incomplete markets driven by aggregate shocks, which is given by the covariance between the consumption of household h relative to its average lifetime consumption, $d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right)$, and the elasticity of the rationing rule with respect to aggregate labor supply, γ_h, weighted by the change in aggregate output relative to the efficient allocation, $\partial \log Y_t = (\hat{y}_t - \hat{y}_t^*)$.

Combining these results obtains Proposition 2.

5 Optimal Monetary Policy

We now use the results formally established in Baqaee et al. (2026) to characterize the optimal policy in this setting. The optimal policy maximizes the second-order approximation of aggregate productivity, which is given by the expression in Proposition 2, and so it is a first-order approximation of the optimal policy in the non-linear problem. The following result formalizes this statement.

Theorem 1 (Approximate Optimal Policy with Time-0 Transfers)**.** *Let $\mathcal{P}^*$ be the policy that maximizes the quadratic-approximation of aggregate productivity* (38)*, subject to the equilibrium conditions* (22)–(26)*. Then, $\mathcal{P}^*$ is a first-order approximation of the optimal policy in the non-linear problem in* (28)*.*

Our first result in characterizing optimal monetary policy is that it is always optimal to fully accommodate demand shocks.

Proposition 3 (Full accomodation of demand shocks)**.** *In the optimal policy, the allocations $\{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$ are independent of $\{\delta_t\}$. The optimal monetary policy satisfies that $i_t + \delta_t = \hat{i}_t$, where $\hat{i}_t$ is the optimal policy in the economy in the absence of demand shocks.*

This proposition establishes that it is optimal to fully accomodate demand shocks, i.e., to set the nominal interest rate in a way that the real interest rate is the same as in the economy without demand shocks.

To make progress in characterizing the optimal policy, we now assume that (25) can be satisfied by an appropriate choice of interest rates. This is a standard assumption in the NK literature, which allows us to drop that equation from the set of constraints in the optimization problem and to directly choose the paths of $\{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$, which maximize the second-order approximation of aggregate productivity, subject to the equilibrium conditions (22)–(24). McKay and Wolf (2022a) provide conditions under which the optimal policy can be characterized in this way.

In the presence of both nominal price and wage rigidities, the optimal policy problem is technically complex, since the real wage becomes a state variable. To connect with the existing literature, we first characterize the optimal policy in the presence of only nominal price rigidities, and then we characterize the optimal policy in the presence of only wage rigidities. We then discuss the optimal policy in the presence of both nominal price and wage rigidities.

5.1 Optimal Policy in the Sticky-Price HANK Model

Definition 4 (Sticky-Price Models)**.** The *Sticky-Price HANK* model is the HANK model with $\lambda_w \to \infty$, $u_{w,t} = 0$, and $\gamma_h = 0$ for all h.

In the sticky-price models, the approximate aggregate efficiency is given by:

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\left\{\sum_{t=0}^{\infty}\beta^t\left\{\left(1+\nu^{-1}\right)(\hat{y}_t-\hat{y}_t^*)^2+\frac{\varepsilon}{\lambda_p}(\pi_t)^2\right\}\right.$$
$$\left.+\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right),d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right)\right\}.$$

In this model, the losses relative to the frontier, $\log \mathcal{D} - \log A(\mathcal{P})$, are equal to the losses in the sticky-price RANK model, up to the losses from incomplete markets

$$\frac{(1-\beta)\omega_C}{2}\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right),d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right).$$

We now by providing a surprising but essential result. To a first-order approximation around the efficient allocation, the losses from incomplete markets are orthogonal to aggregate conditions, and so they are unaffected by changes in monetary policy.

Proposition 4 (Optimal Policy in a Sticky Price HANK Model)**.** *Consider the Sticky Price HANK model. In the second-order approximation of aggregate productivity, the losses from incomplete markets*

$$-\omega_C\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right),d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right), \tag{46}$$

are orthogonal to aggregate conditions around the efficient allocation. So the losses from incomplete markets are unaffected by changes in monetary policy.

It follows that, the optimal monetary policy coincides with the optimal policy in the Sticky-Price RANK model.

This result is key to understanding the optimal policy in a HANK model. While aggregate efficiency is affected by the presence of wedges associated with market incompleteness, there is nothing that monetary policy can do to affect these wedges.

So, the optimal policy in a Sticky-Price HANK model coincides with the optimal policy in the Sticky-Price RANK model, even though the losses from incomplete markets are still present and affect aggregate productivity under any policy $\mathcal{P}$.

5.2 Optimal Policy with Sticky Wages Only

Definition 5 (Sticky-Wage HANK and RANK Models)**.** The *Sticky-Wage HANK* model is the HANK model with $\lambda_p \to \infty$ and $u_{p,t} = 0$. The *canonical HANK* model (Auclert et al., 2025) is the Sticky-Wage HANK model with $\gamma_h = 0$ for all h.

For the special case of the canonical HANK model, the second order approximation of aggregate productivity reduces to

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\left\{ \sum_{t=0}^{\infty} \beta^t \left\{ \left(1+\nu^{-1}\right)(\hat{y}_t - \hat{y}_t^*)^2 + \frac{\psi}{\lambda_w}(\pi_{w,t})^2 \right\} \right.$$
$$\left. + \sum_{t=0}^{\infty} \beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right) \right\}.$$

As in the case of the Sticky-Price HANK model, the losses from incomplete markets are orthogonal to monetary policy, and so the optimal policy in the canonical HANK model coincides with the optimal policy in the Sticky-Wage RANK model, even though the losses from incomplete markets are still present and affect aggregate productivity under any policy $\mathcal{P}$.

Proposition 5 (Optimal Policy in a Canonical HANK Model)**.** *Consider the Canonical HANK model. In the second-order approximation of aggregate productivity, the losses from incomplete markets*

$$-\omega_C \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right), \tag{47}$$

are orthogonal to aggregate conditions around the efficient allocation. So the losses from incomplete markets are unaffected by changes in monetary policy.

It follows that the optimal monetary policy coincides with the optimal policy in the Sticky-Wage RANK model.

However, in the more general sticky-wage HANK model, with $\gamma_h \neq 0$, the approximation of aggregate productivity includes additional terms that capture the losses from the dispersion of labor wedges across households, and the interaction between incomplete markets and the distribution of labor wedges. These terms are not orthogonal to aggregate conditions around the efficient allocation, and so they can be affected by changes in monetary policy. Hence, in this more general model, the optimal policy does not coincide with the optimal policy in the Sticky-Wage RANK model, since the optimal policy in the Sticky-Wage RANK model does not take into account these additional losses. In the sticky-wage models, the approximate aggregate efficiency is given by:

$$\begin{aligned}\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\Bigg\{ \sum_{t=0}^{\infty}\beta^t\Bigg\{ \left(1+\nu^{-1}\right)(\hat{y}_t-\hat{y}_t^*)^2 + \frac{\psi}{\lambda_w}(\pi_{w,t})^2 \Bigg\} \\ + \nu^{-1} Var_{\chi_h}(\gamma_h)\sum_{t=0}^{\infty}\beta^t(\hat{y}_t-\hat{y}_t^*)^2 + (1-\beta)\,Var_{\chi_h}[\gamma_h]\left[\sum_t \beta^t(\hat{y}_t-\hat{y}_t^*)\right]^2 \\ + \sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\left(\frac{c_{h,t}}{c_{h,0}}\right), d\log\left(\frac{c_{h,t}}{\bar{c}_h}\right)\right) - \sum_{t=0}^{\infty}\beta^t Cov_{\chi_h}\left(d\log\frac{c_{h,t}}{\bar{c}_h}, \gamma_h\right)(\hat{y}_t-\hat{y}_t^*)\Bigg\}.\end{aligned} \tag{48}$$

Proposition 6 (Targeting Criterion)**.** *Consider the Sticky-Wage HANK model. The optimal policy satisfies the following targeting criterion*

$$\pi_{w,t} + \hat{\psi}^{-1}(\tilde{y}_t - \tilde{y}_{t-1}) = \zeta_{IM}\cdot\frac{Cov_{\chi_h}\left(d\log\frac{c_{h,t}}{c_{ht-1}}, \gamma_h\right)}{Var_{\chi_h}(\gamma_h)} \tag{49}$$

and

$$\pi_{w,0} + \hat{\psi}^{-1}\tilde{y}_0 = \zeta_{IM}\cdot\frac{Cov_{\chi_h}(d\log c_{h,0}, \gamma_h)}{Var_{\chi_h}(\gamma_h)} \tag{50}$$

where $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$ *is the output gap,* $\hat{\psi} \equiv \frac{\psi}{1+\frac{\nu^{-1}}{1+\nu^{-1}}Var_{\chi_h}(\gamma_h)}$*, and* $\zeta_{IM} \equiv \frac{\hat{\psi}^{-1}-\psi^{-1}}{2\nu^{-1}}$

The Targeting Criterion in Proposition 6 shows that the optimal policy in a Sticky-Wage HANK model is a combination of two terms. The first term captures the optimal response to the interaction between incomplete markets and the distribution of labor wedges across households, while the second term captures the optimal response to the output gap, which is similar to the optimal policy in a Sticky-Wage RANK model but with an additional term that captures the effect of the dispersion of labor wedges across households.

This targeting criterion generalizes the optimal policy in a Sticky-Wage RANK model. First, note that the presence of heterogeneity in labor market outcomes, $Var_{\chi_h}(\gamma_h) > 0$, reduces the coefficient trading off inflation and output stabilization $\hat{\psi} < \psi$. This result follows from the fact that, in the presence of heterogeneity in labor market outcomes, deviations of output from its natural level become more costly due to the dispersion in individual labor wedges. It follows that the optimal policy puts relatively more emphasis on output stabilization, relative to inflation stabilization than in the RANK version of the model.

Second, note that the central bank also understands how affecting the output gap affects the losses from incomplete markets, through the second term. In fact, if the covariance is positive, the central bank can reduce the losses from incomplete markets by adjusting the output gap. If the covariance is negative, then the central bank can reduce the losses from incomplete markets by engineering an expansion at time $t-1$ relative to t.

Corollary 1 (Long-Run Targeting Criterion)**.** *Consider the Sticky-Wage HANK model. In the long run, $t \to \infty$, the optimal targeting rule converges to*

$$\tilde{y}_t - \tilde{y}_{t-1} = -\hat{\psi}\pi_{w,t}. \tag{51}$$

In the long-run, the targeting criterion characterizing the optimal policy trade-off between output and inflation stabilization becomes identical to that of a Sticky-Wage RANK model, but with a modified coefficient $\hat{\psi}$ that puts more weight on output stabilization relative to inflation stabilization than in the RANK version of the model. This larger coefficient on output stabilization reflects the fact that, in the presence of heterogeneity in labor market outcomes, deviations of output from its natural level become more costly due to the dispersion in individual labor wedges. Concerns with incomplete markets vanish in the long-run, since the covariance term converges.

This result implies that the long-run optimal policy in the HANK model is the same as in a Modified RANK, that puts more weight on output stabilization relative to inflation stabilization than in the standard RANK model. This insight, however, is more general and we now establish it for the general model.

5.3 Optimal Policy with Sticky Prices and Sticky Wages

Definition 6 (The Dual-Mandate Central Bank)**.** We define the problem of the Dual-Mandate Central Bank as the problem of minimizing the ad-hoc loss function:

$$\mathcal{L} = \frac{1}{2}\mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\left(\omega_y(\hat{y}_t - \hat{y}_t^*)^2 + \omega_p\pi_t^2 + \omega_w\pi_{w,t}^2\right)\right] \tag{52}$$

subject to the equilibrium conditions in Section 2. Here $\omega_y, \omega_p, \omega_w \geq 0$ are weights on output stabilization, price inflation stabilization, and wage inflation stabilization respectively. We denote the allocations implemented by the Dual-Mandate CB as $S_t^* \equiv \{\hat{y}_t^*, \pi_t^*, \pi_{w,t}^*, \hat{w}_t^*\}$.

The objective of the Dual-Mandate CB is an ad-hoc loss function that puts weight on output-gap stabilization and overall price stabilization (including both goods prices and wages). When $\omega_y = (1+\nu^{-1})$, $\omega_p = \frac{\varepsilon}{\lambda_p}$, and $\omega_w = \frac{\psi}{\lambda_w}$, the Dual-Mandate CB coincides with the Efficiency Mandate optimal policy in the RANK model. The Dual-Mandate central bank therefore does not consider the losses from incomplete markets when choosing its policy.

Proposition 7 (Asymptotic Equivalence to the Dual-Mandate CB)**.** *Consider the Sticky-Price and Sticky-Wage HANK model. Suppose that there is a unique bounded Efficiency Mandate solution. Let $S_t \equiv \{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$ denote the allocations implemented by the optimal Efficiency Mandate policy. Let S_t^* denote the allocations implemented by the Dual-Mandate CB with weights $\omega_y = \left[1+\nu^{-1}(1+Var_{\chi_h}(\gamma_h))\right]$, $\omega_p = \frac{\varepsilon}{\lambda_p}$, and $\omega_w = \frac{\psi}{\lambda_w}$. Then, there exists a deterministic sequence D_t, such that*

$$S_t = S_t^* + D_t, \quad \forall t \geq 0 \tag{53}$$

and $D_t \to 0$ as $t \to \infty$.

In the general HANK model with both sticky prices and sticky wages, the optimal policy does not coincide with the optimal policy in the RANK model, since the optimal policy in the RANK model does not take into account the additional losses from the dispersion of labor wedges across households, and the interaction between incomplete markets and the distribution of labor wedges. However, the asymptotic equivalence result shows that in the long-run, the optimal policy in the HANK model converges to the policy of a RANK model that puts more weight on output stabilization relative to inflation stabilization than in the standard RANK model.

Note that this result applies also in the cases of only sticky prices or only sticky wages,

since these are special cases of the general model. In the case of only sticky prices, $D_t = 0$ for all t, and the optimal policy is always as if RANK.

5.4 Optimal Simple Policy Rules

As it is well known, optimal monetary policy in NK models may involve complex time- and state-dependent policies that depend on the entire history of shocks. In this section, we characterize the optimal simple policy rules, as in Woodford (1999). These optimal policy rules constrain the policymaker to implement allocations that are only a function of the current state of the economy in a time-invariant way. We also assume that each shock follows a first-order autoregressive process with persistence parameter ρ_ϑ, for $\vartheta \in \{\delta, z, u\}$, and that the initial conditions for the shocks are given by their unconditional distribution.

We now characterize the optimal policy rules in the Sticky-Price and Sticky-Wage HANK models, separately.[6]

Not surprisingly, in the Sticky-Price HANK model, the optimal policy rules are the same as the optimal policy rules in the Sticky-Price RANK model, since the optimal policy in the Sticky-Price HANK model coincides with the optimal policy in the Sticky-Price RANK model.

Proposition 8. *Consider the Sticky-Price models with policy rules of the form:*

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{y}_t^* + \phi_{y,u}u_{p,t} \tag{54}$$

$$\hat{\pi}_t = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{y}_t^* + \phi_{\pi,u}u_{p,t}. \tag{55}$$

Then, the optimal policy rules in HANK coincide with those in the RANK model and satisfy:

i. Zero average inflation and output gap: $\phi_y = \phi_y^{RANK} \equiv 0$ *and* $\phi_\pi = \phi_\pi^{RANK} \equiv 0$.

ii. Full accommodation of demand shocks: $\phi_{y,\delta} = \phi_{y,\delta}^{RANK} \equiv 0$ *and* $\phi_{\pi,\delta} = \phi_{\pi,\delta}^{RANK} \equiv 0$.

iii. Optimal response to natural output shocks: $\phi_{y,z} = \phi_{y,z}^{RANK} \equiv 1$ *and* $\phi_{\pi,z} = \phi_{\pi,z}^{RANK} \equiv 0$.

iv. Leaning against cost-push shocks:

$$\phi_{y,u} = \phi_{y,u}^{RANK} \equiv -\frac{\varepsilon}{\kappa_p\varepsilon + (1-\beta\rho_u)^2} \quad \text{and} \quad \phi_{\pi,u} = \phi_{\pi,u}^{RANK} \equiv \frac{(1-\beta\rho_u)}{\kappa_p\varepsilon + (1-\beta\rho_u)^2},$$

[6]Focusing on one source of rigidity at a time allows us to derive closed form expressions for the optimal policy.

where $\kappa_p \equiv \lambda_p \left\{1 + \nu^{-1}\right\}$.

The optimal policy rules in the Sticky-Wage HANK model are more complex, since in general optimal policy will not coincide with that in its RANK counterpart. However, we can still characterize these optimal policies simply. The next proposition shows that the optimal policy rules in the Sticky-Wage HANK model have the characteristic that they are analogous to those in the RANK model, except that the Central Bank puts more weight on output stabilization relative to inflation stabilization. A surprising fact is that the interaction between incomplete markets and the policy no longer affects the ex-ante optimal policies.

Proposition 9 (Simple Rules in a Sticky-Wage HANK Model)**.** *Consider the Sticky-Wage model with policy rules of the form:*

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{y}_t^* + \phi_{y,u}u_{w,t} \tag{56}$$

$$\hat{\pi}_{w,t} = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{y}_t^* + \phi_{\pi,u}u_{w,t}. \tag{57}$$

Then, the optimal policy rules in HANK satisfy:

i. Zero average inflation and output gap: $\phi_y = \phi_y^{RANK} \equiv 0$ *and* $\phi_\pi = \phi_\pi^{RANK} \equiv 0$.

ii. Full accommodation of demand shocks: $\phi_{y,\delta} = \phi_{y,\delta}^{RANK} \equiv 0$ *and* $\phi_{\pi,\delta} = \phi_{\pi,\delta}^{RANK} \equiv 0$.

iii. Optimal response to natural output shocks: $\phi_{y,z} = \phi_{y,z}^{RANK} \equiv 1$ *and* $\phi_{\pi,z} = \phi_{\pi,z}^{RANK} \equiv 0$.

iv. Weaker leaning against cost-push shocks:

$$\phi_{y,u} = -\frac{\tilde{\psi}}{\kappa_w\tilde{\psi} + (1-\beta\rho_u)^2} > -\frac{\psi}{\kappa_w\psi + (1-\beta\rho_u)^2} \equiv \phi_{y,u}^{RANK},$$

$$\phi_{\pi,u} = \frac{(1-\beta\rho_u)}{\kappa_w\tilde{\psi} + (1-\beta\rho_u)^2} > \frac{(1-\beta\rho_u)}{\kappa_w\psi + (1-\beta\rho_u)^2} \equiv \phi_{\pi,u}^{RANK}.$$

where $\kappa_w \equiv \lambda_w \left\{1 + \nu^{-1}\right\} > 0$ *and* $\tilde{\psi} \equiv \frac{\psi}{1 + \frac{Var_{\chi_h}[\gamma_h]}{(1+\nu^{-1})}\left[\nu^{-1} + \frac{(1-\beta)}{(1-\beta\rho_u)^2(1+\beta)}\right]} > 0$.

The optimal rules in the Sticky-Wage HANK model are as if it is a RANK model but with a modified coefficient on the output gap. In particular, the planner focuses more heavily on output stabilization than in the RANK counterpart. The optimal policy is as if RANK but with a lower elasticity of substitution across labor varieties $\tilde{\psi} < \psi$, where the reduction is a consequence of the fact that $Var_{\chi_h}(\gamma_h) > 0$ increases the losses from output gaps, and so the planner puts more weight on output stabilization relative to inflation stabilization.

6 Numerical Illustration in a Sticky-Wage HANK Model

In this section, we quantify the optimal policy in the HANK model. Our key question is to understand how the optimal policy in the HANK model differs from the optimal policy in the RANK model, and how important these differences are quantitatively. To this end, we focus on the Sticky-Wage HANK model, since in the Sticky-Price HANK model the optimal policy coincides with that in the RANK model. We also focus on the case with only time-0 transfers, where the optimal policy deviates from RANK due to losses from incomplete markets. We calibrate the model to match key features of the US economy, and we compute the optimal policy using the second-order perturbation method.

Calibration. We calibrate the model to a period length of one quarter. The elasticity of intertemporal substitution is set to 1, and we set χ so that in steady state the Frisch elasticity $\nu = \frac{L}{1-L} = 0.75$, which are standard parameter values. We set productivity $Z = 1.75$ to normalize output to 1 in steady state. We assume that the productivity process follows a log AR(1) process with persistence $\rho_z = 0.95$ and standard deviation $\sigma_z = 0.75$. We assume that the households face a simple borrowing limit, $\underline{a}$ on the total value of their assets. We choose the discount factor $\beta = 0.97$ and the borrowing limit $\underline{a} = -4.82$ to match a real interest rate of 2% a year and an MPC of 0.21 at quarterly frequency. We assume that prices are fully flexible $\lambda_p \to \infty$, and we set the slope of the Phillips Curve $\lambda_w = 0.09$ and the elasticity of substitution across labor varieties $\psi = 4$, consistent with Galí (2015).

Table 1: Calibrated Parameters

Parameter	Description	Value	Parameter	Description	Value
β	Discount factor	0.97	λ_w	Wage Phillips slope	0.09
r	Real interest rate	2%	ϕ_π	Taylor coefficient	1.5
ρ_z	Persistence z	0.95	Z	Productivity	1.75
σ_z	Std. dev. z	0.75	$\underline{a}$	Borrowing limit	-4.82
χ	Labor disutility	0.75	ψ	Elast. of Substitution n_u	4

Notes: This table reports the parameter values in the baseline calibrated model. The borrowing limit $\underline{a}$ and discount factor β are chosen to match an MPC of 0.21 at quarterly frequency and a 2% yearly real interest rate. Idiosyncratic productivity follows a log AR(1) process with persistence ρ_z and standard deviation σ_z.

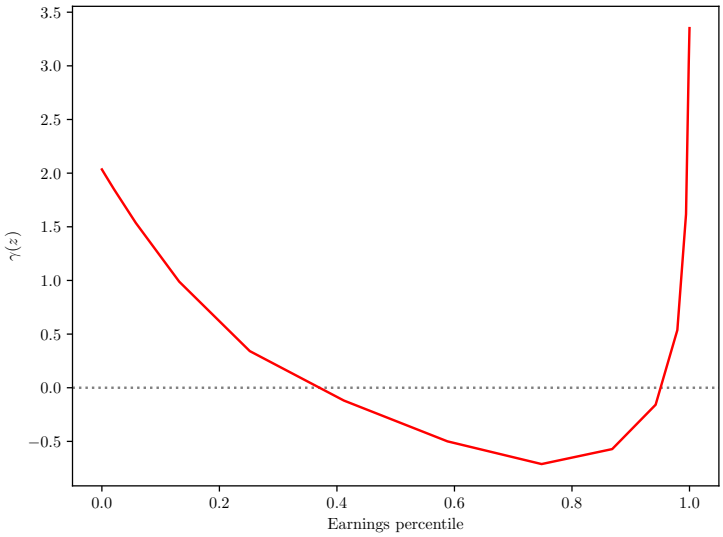
To model the cyclicality of income risk, we follow Auclert and Rognlie (2018) and assume that the incidence elasticity is only a function of the household's current produc-

tivity level:

$$\gamma_h \equiv \frac{\partial \log \Gamma(1 - L_t; \zeta_{h,t})}{\partial \log(1 - L_t)} = \gamma(z_{h,t}).$$

We then use data on worker betas from Guvenen et al. (2017), to calibrate the function $\gamma(z)$, see Appendix B.1 for details. We plot the implied $\gamma(z)$ in Figure 1.

Figure 1: Incidence Elasticity $\gamma(z)$



Notes: This figure plots the incidence elasticity $\gamma(z)$ as a function of household idiosyncratic productivity z (mapped to their earnings percentile). The incidence elasticity measures the percent change in a household's labor income share in response to a one percent change in aggregate labor supply. The function is calibrated using worker-level earnings betas from Guvenen et al. (2017); see Appendix B.1 for details. Higher values of $\gamma(z)$ indicate that the corresponding productivity group bears a disproportionately large share of aggregate labor-supply fluctuations.

Targeting Criterion. The targeting criterion in Proposition 6 shows that the optimal policy in the Sticky-Wage HANK model satisfies the conditions in Proposition 6. In our parametrization, the variance of the incidence elasticity is given by $Var_{\chi_h}(\gamma_h) = 0.78$. This value implies that $\hat{\psi} = 2.76$ (see Table 2), which represents an increase in the weight attributed to output stabilization relative to inflation stabilization of 44.78%.

The second component of the targeting criterion captures the optimal response to the interaction between incomplete markets and the distribution of labor wedges across households. The covariance term depends on the covariance of people's expected consumption at time t as a function of their initial productivity $\mathbb{E}[d \log c_t | z_0]$ and the cyclicality of their labor income $\gamma(z_0)$. The first panel of Figure 2 plots the expected consumption

Table 2: Targeting Criterion Coefficients

Coefficient	Description	Value
ψ	Elast. of Substitution n_u	4
ν	Frisch	0.75
$Var_{\chi_h}(\gamma_h)$	Variance of Incidence Elasticity	0.78
$\hat{\psi}$	Targeting Parameter	2.76
ζ_{IM}	Incomplete Markets Coefficient	0.04

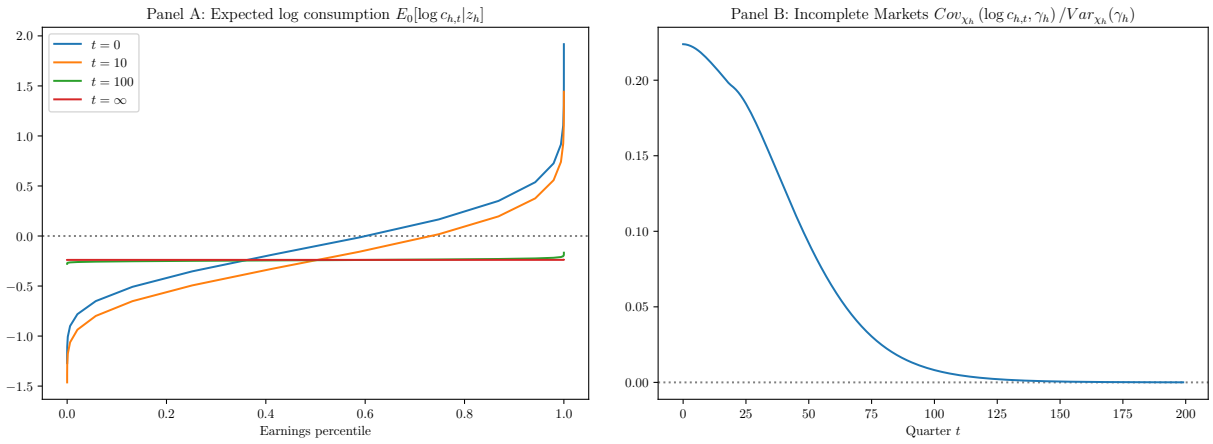
Notes: This table reports the key coefficients entering the targeting criterion in Proposition 6. The upper panel lists the primitive parameters: the elasticity of substitution across labor varieties (ψ), the Frisch elasticity (ν), and the variance of the incidence elasticity $Var_{\chi_h}(\gamma_h)$. The middle panel reports the implied targeting parameter $\hat{\psi}$, which governs the relative weight on output-gap stabilization versus wage-inflation stabilization in the HANK targeting criterion. The bottom panel reports the incomplete-markets coefficient ζ_{IM}, which captures the effect of the interaction between cyclical income risk and incomplete markets on optimal policy.

at time t as a function of the initial productivity level, $\mathbb{E}[d \log c_t | z_0]$. The figure shows that expected consumption growth is decreasing in the initial productivity level, which implies that the covariance term is negative. Intuitively, consumption in earlier dates is heavily influenced by the initial productivity level, since households with higher initial productivity have higher income, and so they consume more. As time passes, households become exposed to uninsurable idiosyncratic shocks, and so the influence of the initial productivity level on consumption weakens. In the limit, consumption is independent of the initial productivity level.

The second panel of Figure 2 plots the covariance term, $Cov_{\chi_h}(\log c_{h,t}, \gamma_h)$, over time t. The covariance is highest at $t = 0$ and falls over time. As expected, the covariance is initially positive and falls over time, eventually reaching 0 in the long-run. The fact that the covariance is positive reflects the fact that the cyclicality of labor income rises particularly fast at the top of the income distribution, where households consume more and have a higher χ_h.

The Optimal Policy. We use these quantitative results to compute the optimal policy in the Sticky-Wage HANK model, and compare them to the equivalent Sticky-Wage RANK model, with the standard weights. Using the result of Proposition 7, we decompose the response of optimal policy into a deterministic component, D_t, that captures the deviation of optimal policy from the Dual-Mandate policy, and a stochastic component that captures the response of optimal policy to shocks.

Figure 2: Targeting Criterion: Incomplete Markets



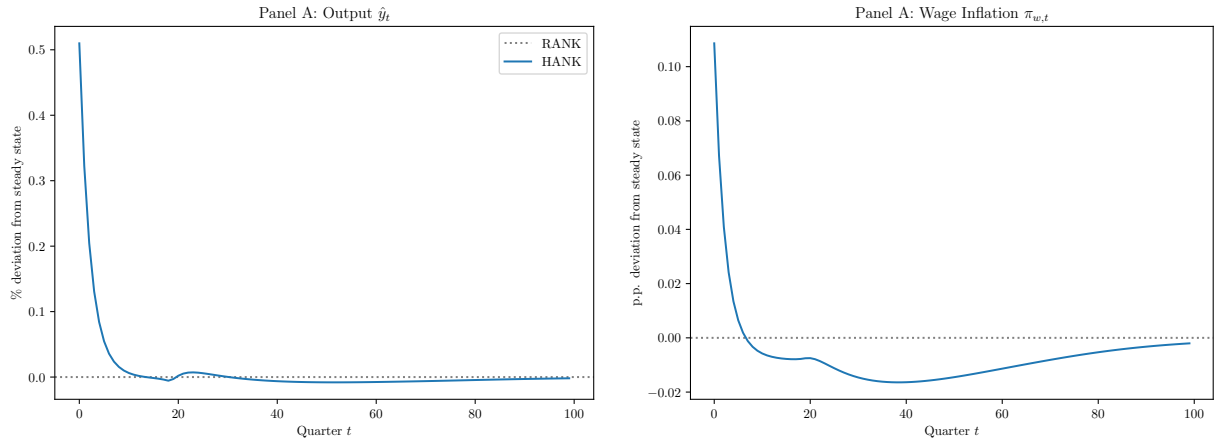
Notes: The left panel plots the expected log consumption at horizon t conditional on initial productivity z_0, i.e., $\mathbb{E}[d \log c_t | z_0]$. The right panel plots the covariance $Cov_{\chi_h}(d \log c_{h,t}, \gamma_h)$ between expected consumption and the incidence elasticity as a function of the horizon t. This covariance enters the incomplete-markets term of the targeting criterion in Proposition 6. Both panels are computed under the baseline calibration reported in Table 1.

Deterministic Component. Figure 3 plots the deterministic component, D_t. The figure shows that the optimal policy implements a small expansion initially.

The intuition for this result is as follows. In the targeting criterion (49), the relevant object is the change in the covariance across two consecutive periods, $Cov_{\chi_h}\left(d \log \frac{c_{h,t}}{c_{h,t-1}}, \gamma_h\right)$. Figure 2 indicates that this difference is uniformly positive. This pattern reflects the implications of incomplete markets: individuals are unable to fully smooth consumption across time and states. In particular, households with low productivity consume less than their average lifetime consumption because of borrowing constraints and income risk, whereas households with high productivity consume more than their average lifetime consumption because the interest rate is depressed relative to its complete-markets counterpart. This pattern of under- and over-consumption arises across all goods, including both consumption and leisure.

It follows that economic efficiency can be improved by a policy that effectively taxes leisure for relatively high-productivity households and subsidizes leisure for relatively low-productivity households. In our calibration, the optimal policy implements this allocation through a small initial expansion. Given the empirical evidence in Guvenen et al. (2017), this expansion generates a relative increase in labor supply—equivalently, a higher tax on leisure—at both the top and the bottom of the income distribution. However, because the covariance is income-weighted, the increase in labor supply at the top

Figure 3: Optimal Policy – Deterministic Component



Notes: This figure plots the deterministic component D_t of optimal policy in the Sticky-Wage HANK model, which captures the deviation from the RANK* policy (Proposition 7). The left panel shows the output gap and the right panel shows wage inflation. The deterministic component reflects the incomplete-markets covariance term in the targeting criterion and converges to zero as $t \to \infty$. Baseline calibration; see Table 1 for parameter values.

of the distribution receives greater weight, implying a positive covariance. Hence, the optimal policy is to engineer a small initial expansion to tax the leisure of high-productivity households.

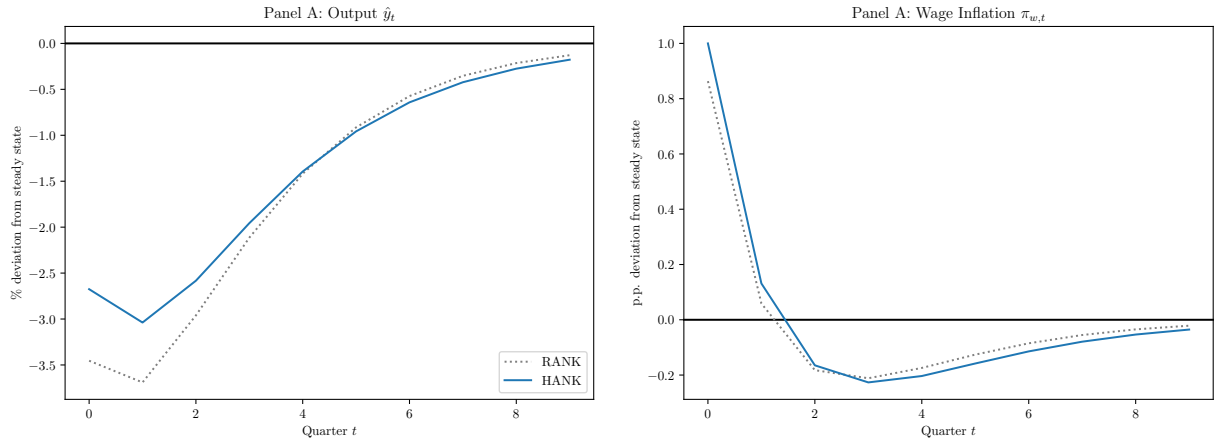
Response to aggregate shocks. We now turn to the response of optimal policy to aggregate shocks. We have already established that the optimal policy in the general HANK model effectively offsets the demand shock δ_t, so that output and inflation are unaffected by this shock. In this model with only sticky wages, the optimal response to a productivity shock is also to close the output gap and keep inflation constant, as in the RANK model. For this reason, we instead focus on the response to a cost-push shock, which is the only shock that generates a non-trivial response in the optimal policy.

We calibrate the shock to be an AR(1) with persistence 0.5 (a standard value; e.g., Galí, 2015). We normalize the size of the innovation such that it leads to a 1 p.p. increase in inflation in the HANK model. Figure 4 plots the response of optimal policy to a cost-push shock. Relative to the RANK model, the optimal policy implements a weaker response to the inflationary pressures of the cost push shock due to the higher weight put on output stabilization.

The higher weight on output stabilization implies that the present value of output gaps in HANK is 10.70% higher than in RANK. This stronger stabilization of the output

gap comes at the cost of higher inflation initially.

Figure 4: Optimal Policy – Response to Cost-Push Shock



Notes: This figure plots the impulse responses of optimal policy to a cost-push wage shock in the Sticky-Wage model under the HANK (solid) and RANK (dashed) Efficiency Mandates. The left panel shows the output gap and the right panel shows wage inflation. The cost-push shock follows an AR(1) process with persistence 0.5; the innovation is normalized so that the initial impact on wage inflation equals 1 p.p. in the HANK model. Relative to RANK, the HANK-optimal policy tolerates higher cumulative inflation in exchange for a smaller output-gap contraction, reflecting the greater weight on output stabilization implied by cyclical income-risk heterogeneity. Baseline calibration; see Table 1 for parameter values.

Optimal Simple Rules The Ramsey problem under commitment can be complex to implement in practice. A natural question is whether the key insights carry over to simple policy rules that a central bank could more easily communicate and follow. We therefore consider optimal simple rules that express the policy response to a cost-push shock as a linear function of the output gap and wage inflation. When the cost-push shock follows an AR(1) process with persistence ρ_u, these rules take the form in Proposition 9.

The HANK simple rule responds less aggressively to the output gap (in absolute value) and more aggressively to wage inflation than the RANK rule, consistent with the tilting toward output stabilization documented above. In this calibration, we find that

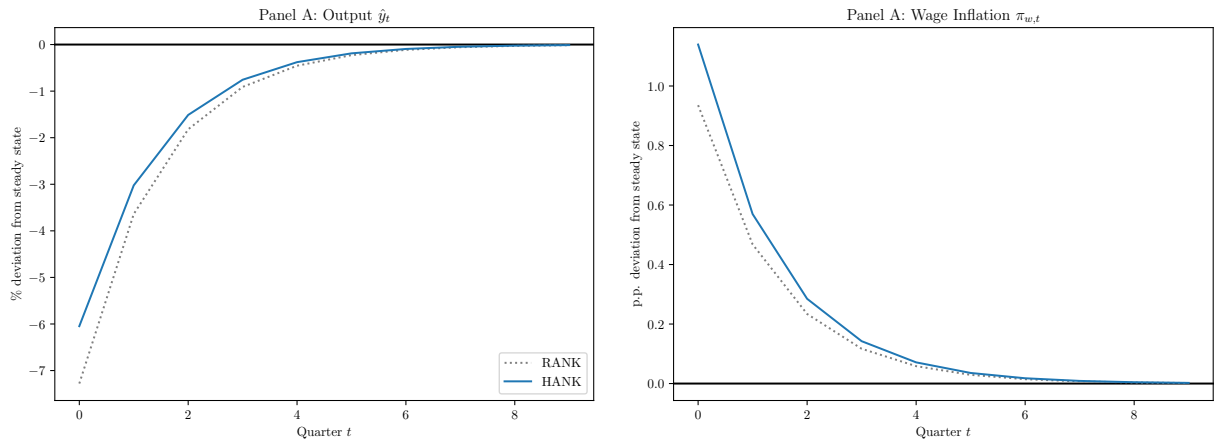
$$\frac{\phi_{u,y}}{\phi_{u,y}^{RANK}} = 0.83 \quad \text{and} \quad \frac{\phi_{u,\pi}}{\phi_{u,\pi}^{RANK}} = 1.22,$$

which imply that, in response to the same cost-push shock, the HANK simple rule produces a 17% smaller output-gap contraction and a 22% larger wage-inflation response

than the RANK simple rule. These differences are quantitatively significant and mirror the pattern in the full Ramsey solution.

Figure 5 illustrates these differences by plotting the impulse responses to a cost-push shock under the optimal simple rules and the same magnitude of cost-push shock that is used in the full Ramsey solution. Relative to the full solution, the simple rule produces more inflation and a sharper contraction of output, both in HANK and in RANK. The qualitative features that HANK policy responds less aggressively to the output gap and more aggressively to wage inflation than RANK policy, remain evident in the simple rules solution.

Figure 5: Optimal Policy – Simple-Rules Response to Cost-Push Shock



Notes: This figure plots the impulse responses of the output gap (left panel) and wage inflation (right panel) to a cost-push wage shock under optimal simple rules in the HANK (solid) and RANK (dashed) models. The simple rules express the policy response as a linear function of the output gap and wage inflation with shock-specific coefficients $\phi_{y,u}$ and $\phi_{\pi,u}$. The cost-push shock follows an AR(1) process with persistence $\rho_u = 0.5$. Baseline calibration; see Table 1 for parameter values.

7 Conclusion

This paper studies optimal monetary policy in HANK under an Efficiency Mandate that ranks policies by aggregate productivity—the largest feasible contraction of endowments consistent with compensating every household. This criterion is invariant to monotone transformations of utility and separates efficiency from redistribution by construction. A second-order approximation of aggregate productivity yields a tractable quadratic objective that nests the standard New Keynesian loss function and augments it with Harberger

triangles for person-specific labor wedges, incomplete markets, and their interaction with cyclical income risk.

In canonical sticky-price and sticky-wage HANK models, incomplete-market losses are locally orthogonal to monetary policy, so the efficiency-maximizing policy coincides with the RANK benchmark. When the incidence of aggregate labor-demand fluctuations is heterogeneous across households, the orthogonality breaks down: optimal policy deviates from RANK by placing more weight on output-gap stabilization. Asymptotically, the optimal policy converges to a modified dual-mandate rule isomorphic to RANK but with a higher output-gap coefficient that reflects the dispersion of incidence elasticities.

These results show that departures from RANK prescriptions in HANK can arise from pure efficiency considerations—without invoking redistributive objectives—once cyclical heterogeneity alters the mapping from aggregate stabilization to wedge dispersion.

References

Acharya, S., E. Challe, and K. Dogra (2023). Optimal monetary policy according to hank. *American Economic Review 113*(7), 1741–1782.

Allais, M. (1979). La théorie générale des surplus.

Alves, F., G. Kaplan, B. Moll, and G. L. Violante (2020). A further look at the propagation of monetary policy shocks in hank. *Journal of Money, Credit and Banking 52*(S2), 521–559.

Angeletos, G.-M., J. Guerreiro, and D. R. Zhang (2025). From rank to hank, without fire. Technical report, National Bureau of Economic Research.

Aoki, K. (2001). Optimal monetary policy responses to relative-price changes. *Journal of monetary economics 48*(1), 55–80.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica 89*(5), 2375–2408.

Auclert, A. and M. Rognlie (2018). Inequality and aggregate demand. Technical report, National Bureau of Economic Research.

Auclert, A., M. Rognlie, and L. Straub (2024). The intertemporal keynesian cross. *Journal of Political Economy 132*(12), 4068–4121.

Auclert, A., M. Rognlie, and L. Straub (2025). Fiscal and monetary policy with heterogeneous agents. *Annual Review of Economics 17*.

Ball, L., N. G. Mankiw, and R. Reis (2005). Monetary policy for inattentive economies. *Journal of monetary economics 52*(4), 703–725.

Baqaee, D. and A. Burstein (2025a, August). Aggregate productivity with heterogeneous agents. Working Paper 34176, National Bureau of Economic Research.

Baqaee, D. and A. Burstein (2026). Aggregate productivity with discrete choice. Technical report, National Bureau of Economic Research.

Baqaee, D., A. Burstein, and J. Guerreiro (2026). The approximate general theory of second best. Working paper.

Baqaee, D. R. and A. Burstein (2025b, September). Efficiency costs of incomplete markets. Working Paper 34233, National Bureau of Economic Research.

Baqaee, D. R. and E. Farhi (2024). Networks, barriers, and trade. *Econometrica 92*(2), 505–541.

Benigno, P. and M. Woodford (2003). Optimal monetary and fiscal policy: A linear-quadratic approach. *NBER macroeconomics annual 18*, 271–333.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021). Inequality, business cycles, and monetary-fiscal policy. *Econometrica 89*(6), 2559–2599.

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

Bilbiie, F. O. (2025). Monetary policy and heterogeneity: An analytical framework. *Review of Economic Studies 92*(4), 2398–2436.

Bilbiie, F. O. and X. Ragot (2021). Optimal monetary policy and liquidity with heterogeneous households. *Review of Economic Dynamics 41*, 71–95.

Blanchard, O. and J. Galí (2010). Labor markets and monetary policy: A new keynesian model with unemployment. *American economic journal: macroeconomics 2*(2), 1–30.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of monetary Economics 12*(3), 383–398.

Clarida, R., J. Gali, and M. Gertler (1999). The science of monetary policy: a new keynesian perspective. *Journal of economic literature 37*(4), 1661–1707.

Cúrdia, V. and M. Woodford (2016). Credit frictions and optimal monetary policy. *Journal of Monetary Economics 84*, 30–65.

Dávila, E. and A. Schaab (2023, February). Optimal monetary policy with heterogeneous agents: Discretion, commitment, and timeless policy. NBER Working Paper 30961, National Bureau of Economic Research.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 273–292.

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

Gabaix, X. (2020). A behavioral new keynesian model. *American Economic Review 110*(8), 2271–2327.

Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications*. Princeton University Press.

Gali, J. and T. Monacelli (2005). Monetary policy and exchange rate volatility in a small open economy. *The Review of Economic Studies 72*(3), 707–734.

Gali, J. and T. Monacelli (2008). Optimal monetary and fiscal policy in a currency union. *Journal of international economics 76*(1), 116–132.

Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo (2017). Worker betas: Five facts about systematic earnings risk. *American Economic Review 107*(5), 398–403.

Harberger, A. C. (1964). The measurement of waste. *The American Economic Review 54*(3), 58–76.

Harberger, A. C. (1971). Three basic postulates for applied welfare economics: an interpretive essay. *Journal of Economic literature 9*(3), 785–797.

Hicks, J. R. (1939). The foundations of welfare economics. *The economic journal 49*(196), 696–712.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The economic journal 49*(195), 549–552.

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

La'O, J. and W. A. Morrison (2024). Optimal monetary policy with redistribution. Technical report, National Bureau of Economic Research.

La'O, J. and A. Tahbaz-Salehi (2022). Optimal monetary policy in production networks. *Econometrica 90*(3), 1295–1336.

Le Grand, F., A. Martin-Baillon, and X. Ragot (2022). Should monetary policy care about redistribution? optimal monetary and fiscal policy with heterogeneous agents. Technical report, Technical report, Paris School of Economics. 1, 5.1, D. 2.

Luenberger, D. G. (1996). Welfare from a benefit viewpoint. *Economic Theory 7*(3), 445–462.

McKay, A. and C. Wolf (2022a). Optimal policy rules in hank. *Manuscript, January*, 121–144.

McKay, A. and C. K. Wolf (2022b). Optimal policy rules in hank. Working paper, version July 17, 2022.

Monacelli, T. (2008). Optimal monetary policy with collateralized household debt and borrowing constraints. In *Asset Prices and Monetary Policy*, pp. 103–146. University of Chicago Press.

Nuño, G. and C. Thomas (2020). Optimal monetary policy with heterogeneous agents. CESifo Working Paper 8670, CESifo.

Rubbo, E. (2023). Networks, phillips curves, and monetary policy. *Econometrica 91*(4),

1417–1455.

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

Schmitt-Grohé, S. and M. Uribe (2005). Optimal fiscal and monetary policy in a medium-scale macroeconomic model. *NBER Macroeconomics Annual 20*, 383–425.

Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American economic review 97*(3), 586–606.

Smirnov, D. (2022). Optimal monetary policy in hank. Technical report.

Werning, I. (2015). Incomplete markets and aggregate demand. Technical report, National Bureau of Economic Research.

Woodford, M. (1999, July). Optimal monetary policy inertia. (7261).

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.

Woodford, M. and Y. Xie (2022). Fiscal and monetary stabilization policy at the zero lower bound: Consequences of limited foresight. *Journal of Monetary Economics 125*, 18–35.