

Monetary Policy without Redistributive Concerns*

David Baqaee
UCLA

Ariel Burstein
UCLA

Joao Guerreiro
UCLA
FRB Minneapolis

May 22, 2026

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 deadweight-loss triangles. This objective nests the standard NK 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 idiosyncratic-income risk, incomplete-market losses are locally orthogonal to monetary policy, so optimal policy coincides with the representative-agent NK (RANK) benchmark. In richer environments with cyclical income risk, 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.

*We thank Enrique Martínez García for an excellent discussion. We also thank Fernando Alvarez, Marco Bassetto, Laura Castillo Martinez, Greg Kaplan, Pablo Ottonello, Christina Patterson, Elisa Rubbo, Pedro Teles, as well as seminar participants at the Bank of Portugal, the University of Chicago Money and Banking Workshop, and the 2026 Texas Monetary Conference for useful comments. 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, and Auclert, 2019). 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: how much does household heterogeneity change optimal monetary policy when the central bank is concerned only with economic efficiency? In heterogeneous-agent environments, analyses of optimal policy based on maximizing social-welfare functions generally combine two distinct motives: improving aggregate efficiency and redistributing resources across households. For example, under a “behind-the-veil-of-ignorance” utilitarian criterion, the policymaker values policies partly because they compress cross-sectional differences in marginal utility. When consumption and leisure are separable, this creates pressure to equalize consumption across households. Monetary policy may therefore be used to redistribute across ex-ante heterogeneous households, even when the underlying inequality is not itself an inefficiency that monetary policy can correct. Indeed, under such a criterion, the policymaker is typically willing to sacrifice efficiency in order to pursue redistributive objectives. Optimal policy therefore depends not only on efficient stabilization considerations, but also on subjective 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 of this idea developed by Baqaee and Burstein (2025a). We 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?

The objective of the policymaker is to maximize *aggregate productivity*, a money-metric-style measure of surplus. Formally, to measure the gains associated with a policy reform,

we ask: if winners were allowed to compensate losers through one-time lump-sum transfers, by how much could aggregate factor endowments be reduced while still keeping every household at least indifferent to the status quo? The reduction in endowments measures the surplus generated by the reform. This criterion isolates efficiency from redistribution by construction: it rises only if a Pareto improvement is feasible. A policy that raises aggregate productivity allows everyone to be made indifferent to the status quo with some factor endowments left over. The criterion does not specify how that surplus should then be divided among agents — hence it does not take a stance on redistribution. Throughout, we restrict compensating transfers to be paid at time 0, so their only role is to implement the compensation required to make everyone whole. In particular, the transfers themselves cannot provide households with any insurance or be used to complete missing financial markets.¹

Unlike a standard social welfare function, this measure is invariant to monotone transformations of individual utility functions and does not require cardinal interpersonal utility comparisons. Although our approach is related to social-welfare-based evaluations with transfers, it generically yields distinct policy prescriptions from that approach in second-best environments, i.e., in situations where the policymaker cannot close all wedges and transfers cannot implement the technological frontier (Baqae 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. Heterogeneity alters the efficient stabilization problem only when aggregate fluctuations have heterogeneous labor-income incidence. Otherwise, incomplete-market losses are present but locally orthogonal to monetary policy, so the efficient-mandate optimal policy coincides with the RANK benchmark.

We establish this result in a general HANK environment with nominal price and wage rigidities, incomplete financial markets, and cyclical labor-income risk. Aggregate demand depends on the distribution of household balance sheets, income expectations, and marginal propensities to consume, which we represent using sequence-space Jacobian methods of Auclert et al. (2021, 2024). On the labor side, a flexible incidence rule governs how aggregate labor fluctuations are distributed across households. When labor-income incidence is homogeneous, the framework nests canonical HANK models; when incidence varies across households, the model generates cyclical income risk consistent with the evidence in Guvenen et al. (2017).

¹Appendix A.5.3 considers *state-contingent transfers*. With a sufficiently rich transfer system, the policymaker 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.

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 mandate asks the central bank to choose policy to maximize aggregate productivity, rather than a social welfare function. This provides a way to isolate efficiency from redistribution in an environment where monetary policy affects both aggregate allocations and cross-sectional outcomes. This measure is also related in spirit to the classic Kaldor-Hicks compensation test (Kaldor, 1939; Hicks, 1939) and its generalization by Schulz et al. (2023).

Deadweight-loss 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 the change in aggregate productivity due to a policy change can be written as the distance of the initial equilibrium from the technological frontier minus a sum of deadweight loss triangles, one for each source of inefficiency in the economy given the new policy rule. In our environment these inefficiencies come from price-markup wedges (due to sticky prices), wage-markup wedges (due to sticky wages), person-specific labor wedges induced by heterogeneous labor-income incidence, and incomplete-markets wedges (due to incomplete markets).

This representation nests the standard New Keynesian quadratic loss function and augments it with additional terms capturing labor-income incidence heterogeneity, incomplete markets, 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.

By invoking results from a companion paper, Baqaee et al. (2026), we show that minimizing this quadratic loss function yields a first-order approximation to optimal policy. Hence, the optimal policy can be characterized with standard linear-quadratic methods as in RANK economies (Woodford, 2003). 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 individual labor wedges and market incompleteness.

Optimal policy under the Efficiency Mandate. In sticky-price and sticky-wage HANK models with acyclical idiosyncratic-income risk, incomplete-market losses are locally orthogonal to monetary policy, so optimal policy coincides with RANK.²

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_h and the incomplete-market wedges. When the covariance is positive, as it is in the data, the optimal policy is to temporarily expand output. When the covariance is negative, the optimal policy is to contract output. However, this second effect is quantitatively short-lived and relatively small.

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 also derive closed-form solutions for optimal time-invariant 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.

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 cyclicity 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 pol-

²The condition under which optimal policy coincides with RANK, acyclical idiosyncratic-income risk, is interestingly related to the positive RANK representation derived in a seminal contribution by Werning (2015). Albeit related, these results are different, since Werning (2015) is focused on deriving conditions under which aggregate quantities behave identically in HANK and RANK, whereas we instead consider policy such that it is possible to engineer individual welfare improvements for all agents. As a technical matter, we also focus on different limits of market completeness.

icy 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. 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 is grounded in the standard New Keynesian tradition for studying optimal policy. In that approach, one studies policy in a linear-quadratic environment: equilibrium conditions are linearized and the objective function is approximated to a second order. This delivers transparent targets and tractable policy prescriptions and has become the workhorse methodology for characterizing the inflation-activity trade-off.

This approach has been applied in many different settings: 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 RANK models with many sectors and input-output linkages, 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 also adopt a linear-quadratic approach 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. This literature shows that the policymaker 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 the conventional inflation-output stabilization trade-offs in RANK (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. The optimal policy problem we study is deliberately different in scope from this body of work because we focus exclusively on efficiency and shut down redistributive concerns. This results in policy prescriptions that are much closer to traditional RANK economies.

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. The linear-quadratic framework we use to approximate optimal policy is detailed in a companion paper Baqaee et al. (2026).

Paper Outline. In Section 2, we introduce our general HANK environment. In Section 3, we define 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 consider the sticky-price and wage only limits. In Section 6 we quantify the optimal policy in a calibrated sticky-wage HANK model. 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. The model allows for both nominal price and wage rigidities, general financial constraints limiting households’ ability to smooth consumption, and endogenous cyclical income risk.

Environment. Time is discrete, infinite, and indexed by $t = 0, 1, 2, \dots$. There is a continuum of households indexed by $h \in [0, 1]$, a representative final goods producer, intermediate-goods producers 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 from productivity state z today to 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 so 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 stochastic sequences of market-goods and leisure consumption bundle for household h . Households have ordinal preferences \succeq over stochastic processes of \mathbf{x}_h . We assume that these preferences are represented by the utility function:

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

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.³ 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 steady-state Frisch elasticity of labor supply is controlled by the parameter χ .

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

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

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:

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

where Z_t denotes an aggregate technology 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. The demand for labor services from union u is given by $n_{u,t}^d \equiv \int n_{i,u,t} di$. Intermediate-goods firms set sticky prices, and unions set sticky wages.

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

$$\int \int n_{i,u,t} didu = 1 - L_t, \quad (4)$$

³In 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, and it is possible to extend the analysis beyond log utility.

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

$$Y_t = C_t, \quad (5)$$

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 to maximize profits $P_t Y_t - \int p_{i,t} y_{i,t} di$, 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

$$y_{i,t} = \left(\frac{p_{i,t}}{P_t} \right)^{-\varepsilon} Y_t, \text{ and } P_t = \left(\int p_{i,t}^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}.$$

Intermediate Goods Firms. Intermediate goods firms are monopolistically competitive in the product market and competitive in input markets. The cost minimization problem implies that

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

where $w_{u,t}$ is the wage of labor variety u and W_t denotes the wage index of labor services.

Intermediate goods prices are sticky à la Calvo (1983). 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. In the appendix, we show that the optimal price is given by

$$p_t^* = \frac{\varepsilon}{\varepsilon - 1} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} \tilde{Q}_{t,t+k} \theta_p^k P_{t+k}^{\varepsilon+1} Y_{t+k} \frac{W_{t+k}}{Z_{t+k} P_{t+k}}}{\mathbb{E}_t \sum_{k=0}^{\infty} \tilde{Q}_{t,t+k} (1 + \tau_{p,t+k}) \theta_p^k P_{t+k}^{\varepsilon} Y_{t+k}}, \quad (6)$$

where $\tilde{Q}_{t,t+k}$ 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 model cost-push shocks and to undo the monopoly distortion in steady state.

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

$$P_t c_{h,t} + \frac{a_{h,t}}{(1+i_t)(1+\delta_t)} + \sum_k q_{k,t} \tilde{a}_{k,h,t} = y_{h,t} + a_{h,t-1} + \sum_k R_{k,t} \tilde{a}_{k,h,t-1} - \tilde{t}_{h,t}, \quad (7)$$

where $y_{h,t}$ is the income of household h , which depends on the household idiosyncratic characteristics $\zeta_{h,t} \equiv (z_{h,t}, a_{h,t-1}, \{\tilde{a}_{k,h,t-1}\}_k)$. Following the NK tradition, we allow 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 time-0 budget constraint also includes a person specific transfer $T_{h,0}$, which we assume to be zero in the status-quo equilibrium. These transfers satisfy $\int T_{h,0} dh = 0$, so they are budget-neutral in aggregate. We introduce these transfers to allow the policy-maker to implement the compensations from winners to losers from a monetary reform, which is a key step in our analysis of optimal policy under the Efficiency Mandate.⁴

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

$$\Omega(a_{h,t}, \{\tilde{a}_{k,h,t}\}_k) \geq 0, \quad (8)$$

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 $\Omega(0, \{0\}_k) > 0$.⁵

Income. Household h has labor income $y_{h,t}$. We allow for nominal rigidities and heterogeneous incidence of aggregate labor fluctuations. Due to the presence of nominal wage rigidities, the household does not optimally choose its labor supply and takes labor

⁴In Appendix A.5.3, 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 .

⁵This assumption ensures that the household can hold some financial debt, which rules out the zero liquidity limit.

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

$$z_{h,t} \int w_{u,t} n_{h,u,t}^s du, \quad (9)$$

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

The literature has emphasized the importance of heterogeneous labor-income incidence of aggregate fluctuations, which are a key driver of the distributional effects of monetary policy and business cycles more broadly (Guvenen et al., 2017; Patterson, 2023; Coglianesi et al., 2025; Del Canto et al., 2025). To capture this feature, we assume that the labor hours of household h used by union u are determined by a flexible function:

$$n_{h,u,t}^s = \Gamma(1 - L_t; \zeta_{h,t}) \int n_{i,u,t} di, \quad (10)$$

where $\int n_{i,u,t} di$ is the total of hours union u supplies to firms i and $\zeta_{h,t}$ indexes household h 's state variables. 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).⁶ 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.

The function Γ satisfies $\int z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) dh = 1$, 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. This ensures that we can nest flexible, efficient labor markets as a special case.

In addition to labor income, the household also receives income from profits. We use a profit-distribution rule, as in Alves et al. (2020), and for simplicity, assume that it is the same incidence function as labor income. So, total household income is given by

$$y_{h,t} = z_{h,t} \int w_{u,t} n_{h,u,t}^s du + z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) \Xi_t. \quad (11)$$

Note that, under these assumptions and the fact that $\Xi_t = \int p_{i,t} y_{i,t} - \int w_{u,t} n_{i,u,t} du di =$

⁶Following the HANK literature, we model the cyclical risk of income with a reduced form incidence rule. Instead, Rubbo (2026) develops a microfounded model of the heterogeneous incidence of aggregate demand shocks on workers' labor incomes.

$P_t Y_t - \int w_{u,t} n_{i,u,t} di$, then

$$y_{h,t} = z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) \left[\int w_{u,t} \int n_{i,u,t} di du + \Xi_t \right] = z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) P_t Y_t.$$

Thus, γ_h controls the elasticity of household h 's total income to fluctuations in aggregate output Y_t .

Unions. Following the sticky-wage NK literature, we assume that each labor variety is supplied by a monopolistically competitive union that demands labor hours from households (e.g., Erceg et al., 2000, Schmitt-Grohé and Uribe, 2005, and Auclert et al., 2024). 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.⁷ As in the case of the intermediate goods producer, we assume that there is a subsidy $\tau_{w,t}$ that undoes monopoly distortions and allows us to capture cost-push shocks. The optimal reset wage is given by

$$w_t^* = \frac{\psi}{\psi - 1} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta \theta_w)^k \chi u'(L_{t+k}) \left(\frac{1}{W_{t+k}} \right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta \theta_w)^k u'(C_{t+k}) \frac{(1 + \tau_{w,t+k})}{P_{t+k}} \left(\frac{1}{W_{t+k}} \right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}. \quad (12)$$

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 technology shocks 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 , which stands in for aggregate demand shocks.

⁷Alternatively, 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.

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

The log-linearized Phillips curves for price and wage inflation are standard and given, in log deviations, by

$$\pi_t = \lambda_p (\hat{w}_t - \hat{z}_t) + \beta \mathbb{E}_t[\pi_{t+1}] + u_{p,t} \quad (13)$$

$$\pi_{w,t} = \lambda_w \left(\left\{ 1 + \nu^{-1} \right\} (\hat{y}_t - \hat{y}_t^*) - (\hat{w}_t - \hat{z}_t) \right) + \beta \mathbb{E}_t[\pi_{w,t+1}] + u_{w,t}, \quad (14)$$

where \hat{y}_t^* is the efficient level of output, \hat{z}_t is the aggregate technology 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

$$\hat{w}_t = \hat{w}_{t-1} + \pi_{w,t} - \pi_t. \quad (15)$$

We can represent the aggregate demand ($C_t = \int c_{i,t} di$) 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 $d\mathbf{C}_{t:\infty} = [dC_t, dC_{t+1}, \dots]$ be the vector of changes in consumption over time. Then,

$$\mathbb{E}_t[d\mathbf{C}_{t:\infty}] = \mathbb{E}_{t-1}[d\mathbf{C}_{t:\infty}] + \sum_{\Theta} \mathbf{M}^{\Theta} (\mathbb{E}_t[d\mathbf{\Theta}_{t:\infty}] - \mathbb{E}_{t-1}[d\mathbf{\Theta}_{t:\infty}]), \quad (16)$$

given $\mathbb{E}_{-1}[d\mathbf{C}_{t:\infty}] = \mathbb{E}_{-1}[d\mathbf{\Theta}_{t:\infty}] = 0$. Here, \mathbf{M}^{Θ} is the sequence-space Jacobian associated with input of the household block Θ . In equilibrium, $dC_t = dY_t$. Collectively, these equations pin down the way equilibrium variables respond to exogenous shocks.

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. Following Baqaee and Burstein (2025a), the change in *aggregate productivity* associated to a change in policy 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 given compensating transfers. This construction maps arbitrary shocks to the economy into a productivity-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 shifter H that shifts the total endowment of time available for market production and home production/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 $\mathbf{x}^0 \equiv \{c_{h,t}^0, l_{h,t}^0\}$, which is the equilibrium consumption and leisure processes under the status quo monetary policy, \mathcal{P}^0 (i.e., the equilibrium in the data). We then consider how a 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 equilibria under the new policy \mathcal{P} with some transfers.

Definition 1 (Consumption Possibility Set). 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 one-off time-0 lump-sum transfers $\{T_{h,0}\}$ that satisfy $\int T_{h,0} dh = 0$.⁸

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

Definition 2 (Aggregate Productivity). Fix a status-quo allocation $\mathbf{x}^0 \equiv \{c_{h,t}^0, l_{h,t}^0\}$. 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,

$$A(\mathcal{P}) \equiv \max \left\{ H > 0 : \text{there exists } \mathbf{x} \in \mathcal{X}(1/H; \mathcal{P}) \text{ with } \mathbf{x}_h \succeq \mathbf{x}_h^0, \text{ for all } h \right\}. \quad (17)$$

In words, $A(\mathcal{P})$ measures the amount of time left-over if we implement policy \mathcal{P} and every household is made indifferent to the status quo allocation using transfers. So, for

⁸The transfers are not contingent on future idiosyncratic state and cannot be used to provide insurance. If the policy-maker has access to state-contingent transfers, then the optimal policy is the one in RANK (see Appendix A.5.3).

example, if $A = 1.1$, then under the new policy, the economy can achieve the same level of welfare for every household as under the status quo allocation, while only using roughly 90% of the total time available. The remaining free time, the surplus, can then be devoted to production or leisure in some unspecified way.

By construction, this measure of efficiency does not rely on cardinal utility functions and interpersonal comparisons of utility. If $A(\mathcal{P}) > 1$, then the policy \mathcal{P} is more efficient than the status quo policy because it can achieve the same level of welfare for every household 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 every household.

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

$$\max_{\mathcal{P}} A(\mathcal{P}). \tag{18}$$

Remark (Variations on the definition of aggregate productivity). Following Debreu (1951) and Baqaee and Burstein (2025a), we define the productivity gain from policy \mathcal{P} via a uniform proportional contraction of all factor endowments. Under this definition, $A(\mathcal{P})$ has a total-factor-augmenting interpretation. Alternative definitions are also possible. For instance, one could scale only the productivity of market production, rather than that of the entire time endowment. Similarly, instead of assuming that lump-sum compensating transfers are paid at time 0, one could consider other lump-sum payout schemes. These modifications do not affect our characterization of the approximately optimal policy. We discuss these issues in more detail in Appendix A.5

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 there are no idiosyncratic productivity shocks. Financial markets are complete and the household is not subject to ad hoc borrowing constraints. In this case, the set of feasible allocations under a policy \mathcal{P} is a singleton, since there is only one equilibrium allocation — there are no transfers with a single agent.

Since the set of feasible allocations, $\mathcal{X}(1/H;\mathcal{P})$, is a singleton, let \mathbf{x}^0 denote the status-quo equilibrium allocation and let $\mathbf{x}(1/H;\mathcal{P})$ denote the equilibrium allocation given

productivity shifter $1/H$ and alternative policy \mathcal{P} . By using the definition, and the fact that utility is monotone in $1/H$, we know that $A(\mathcal{P})$ must satisfy the indifference condition

$$U(\mathbf{x}(1/A(\mathcal{P});\mathcal{P})) = U(\mathbf{x}^0).$$

Since allocations scale linearly with $A(\mathcal{P})$, we can write:

$$U\left(\frac{\mathbf{x}(1, \mathcal{P})}{A(\mathcal{P})}\right) = U(\mathbf{x}^0).$$

So, $A(\mathcal{P})$ is exactly the Lucas (2003) consumption-equivalent variation in this example. Using the functional form of the utility function, we further write:

$$\log A(\mathcal{P}) = U(\mathbf{x}(1, \mathcal{P})) - U(\mathbf{x}^0).$$

Since \mathbf{x}^0 is fixed, the policy that maximizes aggregate productivity is also the one that maximizes utility.

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 Arrow-Debreu economy with wedges that decentralizes the same allocations. 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).

We use four types of wedges to decentralize the equilibrium allocation in Arrow-Debreu: (1) a price-markup wedge to capture nominal price rigidity, (2) a wage-markup wedge to capture nominal wage rigidity, (3) a labor-wedge to capture the incidence function Γ , and (4) household-specific consumption wedges to capture incomplete financial markets.

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. The price-markup wedges, denoted by $\mu_{i,t}^p$, are given by the gap between prices and marginal costs:

$$\log \mu_{i,t}^p \equiv \log p_{i,t} - \log \frac{W_t}{Z_t}. \tag{19}$$

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 $\mu_{u,t}^w$, and define them to be

$$\log \mu_{u,t}^w \equiv \log w_{u,t} - \log \tilde{W}_t, \quad (20)$$

where $\tilde{W}_t \equiv W_t^{1-\psi} \cdot \left(\int w_{u,t}^{-\psi} du \right)^{-1}$ is the average wage per effective labor unit.

Third, the incidence function Γ can also create an inefficient allocation of labor across households captured via household-specific labor wedges. Household h 's labor wedge, denoted by $\mu_{h,t}^L$, satisfies

$$\log \mu_{h,t}^L \equiv \log \left(\frac{\chi u'(l_{h,t})}{u'(c_{h,t})} \right) - \log \frac{\tilde{W}_t}{P_t}. \quad (21)$$

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

$$\log \mu_{h,t}^{IM} \equiv \log \left(\frac{u'(c_{h,t})}{u'(C_t)} \right) - \log \left(\frac{u'(c_{h,0})}{u'(C_0)} \right). \quad (22)$$

In Appendix A.2.1, we show that the equilibrium allocation in our economy can be decentralized as an Arrow-Debreu equilibrium with the wedges defined above. In particular, we show that there exists a set of transfers such that the equilibrium allocation in our economy is supported by an AD equilibrium with these wedges. This result is key for our analysis of optimal monetary policy, since it allows us to leverage the general results in Baqaee et al. (2026) to characterize the approximately optimal policy.

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 first-best Pareto efficient frontier, and there will be some residual inefficiency.

We define the distance to the Pareto efficient frontier as in Debreu (1951).

Definition 3 (Distance to the Frontier). Let $\mathcal{X}^*(H)$ denote the set of allocations that 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.,

$$\mathcal{D} \equiv \max \left\{ H > 0 : \text{there exists } \mathbf{x} \in \mathcal{X}^*(1/H) \text{ with } \mathbf{x}_h \succeq \mathbf{x}_h^0, \text{ for all } h \right\}. \quad (23)$$

Since $\mathcal{X}^*(1/H)$ is the set of technologically feasible allocations, it trivially follows that $\mathcal{X}(1/H; \mathcal{P}) \subset \mathcal{X}^*(1/H)$. 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 next proposition establishes a key property of the frontier allocations, which are allocations that solve (23). For any status-quo allocation \mathbf{x}^0 , the frontier allocations satisfy

$$Y_t^* = C_t^* = \frac{Z_t}{1 + \chi}, \quad L_t^* = \frac{\chi}{1 + \chi}. \quad (24)$$

That is, although there are many Pareto efficient allocations, depending on how resources are distributed across agents, aggregate quantities in every efficient allocation must satisfy the equations above.

Proposition 2 (Distance to Pareto Frontier). *The distance to the Pareto efficient frontier is*

$$\log \mathcal{D} = U(\{C_t^*, L_t^*\}) - \log \int \exp\{U(x_h^0)\} dh. \quad (25)$$

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 still remains open relative to the frontier under a given policy $\log A(\mathcal{P}) = \log \mathcal{D} + \log A(\mathcal{P})/\mathcal{D}$. 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 $\hat{x}_t = d \log(X_t)$ denotes 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 goods-price 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, $\omega_C \equiv C/(wL + C)$ denotes the value of consumption relative to the time endowment, 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.⁹

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

⁹Note 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).

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 Deadweight-Loss 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 approximate optimal policy around an efficient allocation. Effectively, this is done by choosing policy to maximize a second-order approximation of the objective function around 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 3. 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 collection of general-equilibrium deadweight loss triangles.

Proposition 3 (Second-Order Approximation of Aggregate Productivity). *To a second-order approximation, aggregate productivity can be written as*

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \frac{(1 - \beta)\omega_C}{2} \mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left\{ (1 + \nu^{-1}) (\hat{y}_t - \hat{y}_t^*)^2 + \frac{\varepsilon}{\lambda_p} (\pi_t)^2 + \frac{\psi}{\lambda_w} (\pi_{w,t})^2 \right\} \right\} \quad (26)$$

$$\begin{aligned}
& + \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_{t=0}^{\infty} \beta^t (\hat{y}_t - \hat{y}_t^*)^2 + \nu^{-1} (1 - \beta) \text{Var}_{\chi_h}[\gamma_h] \left[\sum_t \beta^t (\hat{y}_t - \hat{y}_t^*) \right]^2 \\
& + \omega_C^{-1} (1 - \beta)^{-1} E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] - \sum_{t=0}^{\infty} \beta^t \text{Cov}_{\chi_h} \left(d \log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) (\hat{y}_t - \hat{y}_t^*) \Bigg\},
\end{aligned}$$

where $\text{Var}_{\chi_h}[\cdot]$ and $\text{Cov}_{\chi_h}[\cdot]$ use cross-sectional consumption-shares χ_h as weights, and $\text{Var}_{\beta, \pi}[\cdot]$ uses time-series discounted value $(1 - \beta)\beta^t \pi(s)$ as weights, and $d \log \bar{c}_h = \mathbb{E}_0[(1 - \beta) \sum_t \beta^t d \log c_{h,t}] = E_{\beta, \pi}[d \log c_{h,t}]$ is expected discounted average deviation in h 's consumption.

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 on the last line coincides with the costs of incomplete markets in Baqaee and Burstein (2025b), while the final term captures the additional costs of incomplete markets in the presence of a heterogeneous distribution of labor wedges across households.

In the RANK model, there are complete markets and no heterogeneity across households. So, $\text{Var}_{\chi_h}(\gamma_h) = 0$ and $E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] = 0$. 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 idiosyncratic-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.

Below we provide a derivation of Proposition 3. Readers that are not interested in the derivation can skip ahead to the next section.

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 (i.e., wedge times the distortion in the quantity divided by two).

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

$$\log A(\mathcal{P}) \approx \log \mathcal{D} + \triangleright^{IM} + \triangleright^p + \triangleright^w + \triangleright^L, \quad (27)$$

where

i. *The losses from incomplete markets are given by*

$$\triangleright^{IM} \equiv (1 - \beta) \mathbb{E}_0 \sum_t \beta^t \int_0^1 \chi_h \cdot \frac{d \log \mu_{h,t}^{IM} \cdot \partial \log x_{h,t}}{2} dh, \quad (28)$$

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-leisure composite 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*

$$\triangleright^p \equiv (1 - \beta) \omega_C \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 \frac{d \log \mu_{i,t}^p \cdot \partial \log y_{i,t}}{2} di, \quad (29)$$

where $\omega_C \equiv C / (wL + C)$ is the market consumption share of the aggregate time endowment.

iii. *The losses from wage markups are given by*

$$\triangleright^w \equiv (1 - \beta) \omega_C \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 \frac{d \log \mu_{u,t}^w \cdot \partial \log n_{u,t}^d}{2} du. \quad (30)$$

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

$$\triangleright^L \equiv (1 - \beta)(1 - \omega_C) \cdot \mathbb{E}_0 \sum_t \beta^t \int_0^1 \chi_h \cdot \frac{d \log \mu_{h,t}^L \cdot \partial \log l_{h,t}}{2} dh. \quad (31)$$

where all partial derivatives, $\partial \log x_{h,t}$, $\partial \log y_{i,t}$, $\partial \log n_{u,t}^d$, and $\partial \log l_{h,t}$, denote the changes in the corresponding quantities, around the efficient allocation, caused only by the wedges in the Arrow-Debreu general equilibrium with compensating transfers. In contrast, we use $d \log \mu_{h,t}^{IM}$, $d \log \mu_{i,t}$, $d \log \mu_{u,t}^w$, and $d \log \mu_{h,t}^L$ to denote changes in wedges in the decentralized equilibrium due to all shocks.

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 deadweight-loss 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 deadweight-loss triangles. In this way, the problem of the efficiency-mandate central bank is to choose policy to minimize the losses associated with the deadweight-loss triangles.

Our next set of results characterize how the different wedges respond to changes in aggregate variables, and ultimately monetary policy. In the main text, we provide some key equations that guide the intuition for the results which are derived in full detail in Appendix A.2.

In equilibrium, the change in $y_{i,t}$ and $n_{u,t}^d$ caused by the wedges satisfy

$$\begin{aligned} \partial \log y_{i,t} &= -\varepsilon \left(d \log \mu_{i,t}^p - \int d \log \mu_{i',t}^p di' \right) + \partial \log Y_t, \\ \partial \log n_{u,t}^d &= -\psi \left(d \log \mu_{u,t}^w - \int d \log \mu_{u',t}^w du' \right) + \partial \log Y_t, \end{aligned}$$

where $\partial \log Y_t$ is the response of aggregate output to wedges only. We can write this, in

more familiar terms, as

$$\partial \log Y_t = d \log Y_t - d \log Y_t^* = (\hat{y}_t - \hat{y}_t^*),$$

which is the gap between output and its efficient level. Using these expressions, we can rewrite the deadweight-loss triangles for price and wage markups using the fact that

$$\begin{aligned} \int_0^1 d \log \mu_{i,t}^p \cdot \partial \log y_{i,t} di &= -\varepsilon \text{Var}_i(d \log \mu_{i,t}^p) + E_i[d \log \mu_{i,t}^p] \partial \log Y_t, \\ \int_0^1 d \log \mu_{u,t}^w \cdot \partial \log n_{u,t}^d du &= -\psi \text{Var}_u(d \log \mu_{u,t}^w) + E_u[d \log \mu_{u,t}^w] \partial \log Y_t, \end{aligned}$$

where $\text{Var}_i(d \log \mu_{i,t}^p)$ and $\text{Var}_u(d \log \mu_{u,t}^w)$ denote the cross-sectional variance of price and wage markups across firms and unions, respectively, $E_i[d \log \mu_{i,t}^p]$ and $E_u[d \log \mu_{u,t}^w]$ denote the cross-sectional average of price and wage markups across firms and unions, respectively. Using these expressions, we obtain the following result.

Lemma 2 (Characterizing the Deadweight-Loss Triangles for Price and Wage Markups).

The Harberger triangles for price and wage markups can be written as

$$\triangleright^p = \frac{(1-\beta)\omega_C}{2} \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\varepsilon}{\lambda_p} (\pi_t)^2 + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right]$$

and

$$\triangleright^w = \frac{(1-\beta)\omega_C}{2} \cdot \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\psi}{\lambda_w} (\pi_{w,t})^2 + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right]$$

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 \triangleright^p and \triangleright^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 $\gamma_h \equiv \frac{\partial \log \Gamma(1-L, \zeta_{h,0})}{\partial \log(1-L)}$ is the incidence elasticity of aggregate labor fluctuations. Using this expression, we can rewrite the deadweight-loss triangle for person-specific labor wedges as

$$\triangleright^L \equiv -\frac{(1 - \beta)\omega_C}{2} \cdot \mathbb{E}_0 \sum_t \beta^t \left\{ E_{\chi_h} \left[d \log \mu_{h,t}^L \right] \partial \log Y_t + Cov_{\chi_h} \left(d \log \mu_{h,t}^L, \gamma_h \right) \partial \log Y_t \right\}$$

where $E_{\chi_h} \left[d \log \mu_{h,t}^L \right]$ and $Cov_{\chi_h} \left(d \log \mu_{h,t}^L, \gamma_h \right)$ 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 Deadweight-Loss Triangles for Person-Specific Labor Wedges). *The deadweight-loss triangle for person-specific labor wedges can be written as*

$$\begin{aligned} \triangleright^L = & -\frac{\omega_C(1 - \beta)}{2} \left[\mathbb{E}_0 \sum_t \beta^t \left[\left[(1 + \nu^{-1}) \partial \log Y_t + E_i [d \log \mu_{i,t}] + E_u [d \log \mu_{u,t}] \right] \partial \log Y_t \right. \right. \\ & \left. \left. + \nu^{-1} Var_{\chi_h} (\gamma_h) (\partial \log Y_t)^2 \right] + \nu^{-1} Var_{\chi_h} [\gamma_h] \cdot (1 - \beta) \mathbb{E}_0 \left(\sum_t \beta^t \partial \log Y_t \right)^2 \right]. \quad (32) \end{aligned}$$

Note that the terms in the first row capture the effect of the average labor wedge, $E_{\chi_h} \left[d \log \mu_{h,t}^L \right]$, on aggregate productivity, while the terms in the second row capture how the incidence of fluctuations induces heterogeneity in the labor wedges across households, $Cov_{\chi_h} \left(d \log \mu_{h,t}^L, \gamma_h \right)$, which also lowers aggregate productivity. If $Var_{\chi_h} (\gamma_h) = 0$, 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 $Var_{\chi_h} (\gamma_h) > 0$, then the heterogeneity in labor wedges across households also creates losses in aggregate productivity.

Denote the lifetime discounted expectation operator, which weighs dates and states by $(1 - \beta)\beta^t \pi(s)$, using $E_{\beta, \pi}[\cdot]$ and let $Cov_{\beta, \pi}[\cdot]$ and $Var_{\beta, \pi}[\cdot]$ denote the associated covariance and variance operators. Then we can rewrite the deadweight-loss triangle for

incomplete markets, solving out for $\partial \log x_{h,t}$, as

$$\triangleright^{IM} = -\frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta,\pi} \left(d \log \left(\mu_{h,t}^{IM} \right) \right) - (1 - \omega_C) \text{Cov}_{\beta,\pi} \left(d \log \mu_{h,t}^{IM}, d \log \mu_{h,t}^L \right) \right].$$

Hence, the losses scale with the cross-sectional expectation of the discounted life-time variance and covariance of incomplete markets and labor wedges. We now use the expressions for wedges in the decentralized equilibrium to rewrite this expression in terms of the deviations of quantities from their efficient allocations.

Lemma 4 (Characterizing the Deadweight-Loss Triangles for Incomplete-Market Wedges).

The deadweight-loss triangles for incomplete markets can be written as

$$\triangleright^{IM} = -\frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \frac{(1 - \beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \text{Cov}_{\chi_h} \left(d \log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \partial \log Y_t.$$

where $d \log c_{h,t}$ represents the log-deviation of household h 's consumption from that in the efficient allocation induced by the wedges, and $d \log \bar{c}_h = (1 - \beta) \mathbb{E}_0 \sum_t \beta^t d \log c_{h,t}$ is the average lifetime consumption of household h .

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 average life-time volatility of each households' share of aggregate consumption. 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 incidence elasticity of aggregate labor fluctuations, γ_h , weighted by the change in aggregate output relative to the efficient allocation, $\partial \log Y_t = (\hat{y}_t - \hat{y}_t^*)$. The intuition for the second component is as follows. With incomplete markets, households over- and under-consume consumption goods, but also leisure, relative to the efficient allocation. By introducing an output gap, monetary policy can affect the distribution of labor and leisure in the cross section provided γ_h is heterogeneous across households. The covariance term therefore captures whether aggregate stabilization pushes labor effort toward households that are currently consuming relatively too much, and away from households that are currently consuming relatively too little.

Combining these lemmas results in Proposition 3.

5 Optimal Monetary Policy

We now use the results formally established in Baqaee et al. (2026) to approximate optimal policy in this setting. To do this, we maximize the second-order approximation of aggregate productivity, which is given by the expression in Proposition 3. Baqaee et al. (2026) show that this produces 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 (26), subject to the equilibrium conditions (13)–(16). Then, \mathcal{P}^* is a first-order approximation of the optimal policy in the non-linear problem in (18).*

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

Proposition 4 (Full accommodation 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 accommodate demand shocks, i.e., to set the nominal interest rate to ensure 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 (16) 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 (13)–(15). 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 with Homogeneous Incidence

In this subsection, we consider optimal policy in the cases in which the incidence of labor wedges is homogeneous across households, i.e., $\gamma_h = 0$ for all h . This includes both the sticky-price only HANK model, in which households choose labor supply optimally, and the canonical sticky-wage HANK model, in which unions demand equal labor supply from all households. As it turns out, in both cases, the optimal policy coincides with the optimal policy in the corresponding RANK model. This result follows from the fact that, since household incidence is homogeneous, monetary policy cannot provide insurance by moving aggregates and so it is optimal to disregard the incomplete market wedges.

Definition 4 (Sticky-Price Models). The *Sticky-Price HANK* model is the HANK model where the slope of the wage Phillips curve $\lambda_w \rightarrow \infty$, there are no wage cost-push shocks $u_{w,t} = 0$, and there is no heterogeneity in incidence $\gamma_h = 0$ for all h .

In the Sticky-Price HANK model, the second-order approximation of aggregate productivity simplifies since $Var_{\chi_h}(\gamma_h) = 0$ and $Cov_{\chi_h}\left(d \log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h\right) = 0$. Furthermore, since $\lambda_w \rightarrow \infty$, then the losses associated with wage inflation disappear. So, 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:

$$\frac{1}{2} E_{\chi_h} \left[Var_{\beta, \pi} \left(d \log \frac{c_{h,t}}{Y_t} \right) \right].$$

This leads to 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 5 (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*

$$-\frac{1}{2} E_{\chi_h} \left[Var_{\beta, \pi} \left(d \log \frac{c_{h,t}}{Y_t} \right) \right], \quad (33)$$

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 deadweight loss

triangles. So, the optimal policy in a Sticky-Price HANK model coincides with the optimal policy in the Sticky-Price RANK model. Note, however, that the incomplete market wedges still affect aggregate productivity. The result shows that, despite the presence of these wedges, when the incidence of demand fluctuations is homogeneous, there is nothing monetary policy can do to provide insurance against idiosyncratic income risk, and so the optimal policy coincides with that in the RANK model.

Similarly, we define the canonical Sticky-Wage HANK model to be the sticky-wage only model with $\gamma_h = 0$ for all h .

Definition 5 (Canonical Sticky-Wage HANK Model). The *canonical HANK* model is the HANK model where the slope of the price Phillips curve $\lambda_p \rightarrow \infty$, there are no cost-push shocks $u_{p,t} = 0$, and there is no heterogeneity in incidence $\gamma_h = 0$ for all h .

As in the previous case, the losses relative to the frontier, $\log \mathcal{D} - \log A(\mathcal{P})$, are equal to the losses in the sticky-wage RANK model, up to the same losses from incomplete markets. We obtain the same irrelevance result.

Proposition 6 (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*

$$-\frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \frac{c_{h,t}}{Y_t} \right) \right], \quad (34)$$

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.

5.2 Optimal Policy with Heterogeneous Incidence and Sticky Wages

We now consider optimal policy in the model with sticky wages and heterogeneous incidence. The key implication is that, when incidence is heterogeneous, the losses from incomplete markets are no longer orthogonal to aggregate conditions. Hence, changes in monetary policy can affect the incomplete market wedges, and so the optimal policy in the sticky-wage HANK model does not coincide with the optimal policy in the sticky-wage RANK model.

Definition 6 (Sticky-Wage HANK Model). The *Sticky-Wage HANK* model is the HANK model where the slope of the price Phillips curve $\lambda_p \rightarrow \infty$ and there are no cost-push shocks $u_{p,t} = 0$.

In the 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. The next proposition derives the optimal targeting criterion that characterizes optimal monetary policy in this environment.

Proposition 7 (Targeting Criterion). *Consider the Sticky-Wage HANK model with $Var_{\chi_h}(\gamma_h) > 0$. The optimal policy satisfies the following targeting criterion*

$$\pi_{w,t} = -\hat{\psi}^{-1}(\tilde{y}_t - \tilde{y}_{t-1}) + \zeta_{IM} \cdot \frac{Cov_{\chi_h}(d \log c_{h,t}, \gamma_h) - Cov_{\chi_h}(d \log c_{h,t-1}, \gamma_h)}{Var_{\chi_h}(\gamma_h)} \quad (35)$$

with the convention that $\tilde{y}_{-1} = 0$ and $Cov_{\chi_h}(d \log c_{h,-1}, \gamma_h) = 0$, and where $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$ is the output gap, and the parameters $\hat{\psi} \equiv \frac{\psi}{1 + \frac{v-1}{1+v-1} Var_{\chi_h}(\gamma_h)}$ and $\zeta_{IM} \equiv \frac{\hat{\psi}^{-1} - \psi^{-1}}{2v^{-1}}$.

The Targeting Criterion in Proposition 7 shows that the optimal wage-inflation target in a Sticky-Wage HANK model is a combination of two terms. The first term captures the optimal response to the output gap. The weight on the output gap term, $\hat{\psi}$, is affected by the dispersion of labor wedges across households $Var_{\chi_h}(\gamma_h)$. The second term captures the optimal response to the interaction between incomplete markets and the distribution 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. To see the mechanism, rewrite the numerator in the covariance term as $Cov_{\chi_h}(d \log(c_{h,t}/c_{h,t-1}), \gamma_h)$. With incomplete markets, households do not smooth consumption and leisure perfectly across dates and states, so some households are consuming relatively too much while others are consuming relatively too little. When this covariance is positive, consumption rises relatively more for households whose hours are more exposed to aggregate labor fluctuations. An

expansion from $t - 1$ to t then raises aggregate hours and, through the heterogeneous incidence elasticities γ_h , shifts labor effort toward precisely the households that have too much leisure, attenuating the incomplete-market wedge. If the covariance were negative, the same logic would call for a contraction.

The targeting criterion in Proposition 7 established that the optimal policy deviates from that of the RANK model to affect the losses arising from incomplete markets. The next result shows that, in the long-run, this ability to affect the losses from incomplete markets vanishes, and so in the long-run, the optimal targeting criterion converges to a simple implementable targeting criterion that only depends on the output gap and wage inflation. It follows that, in the long-run, the optimal policy in the HANK model converges to a simple modified version of the optimal policy in the RANK model.

Corollary 1 (Long-Run Targeting Criterion). *Consider the Sticky-Wage HANK model. Suppose that the optimal equilibrium path is bounded and converges back to the stationary allocation. In the long run, $t \rightarrow \infty$, the optimal targeting rule converges to*

$$\tilde{y}_t - \tilde{y}_{t-1} = -\hat{\psi}\pi_{w,t}. \quad (36)$$

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 7 (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:

$$\mathcal{L} = \frac{1}{2}\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\omega_y(\hat{y}_t - \hat{y}_t^*)^2 + \omega_p\pi_t^2 + \omega_w\pi_{w,t}^2 \right) \right] \quad (37)$$

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 $S_t^{DM} \equiv \{\hat{y}_t^{DM}, \pi_t^{DM}, \pi_{w,t}^{DM}, \hat{w}_t^{DM}\}$.

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 8 (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^{DM} denote the allocations implemented by the Dual-Mandate CB with weights $\omega_y = [1 + \nu^{-1}(1 + \text{Var}_{\chi_h}(\gamma_h))]$, $\omega_p = \frac{\varepsilon}{\lambda_p}$, and $\omega_w = \frac{\psi}{\lambda_w}$. Then, there exists a deterministic sequence D_t , such that*

$$S_t = S_t^{DM} + D_t, \quad \forall t \geq 0 \quad (38)$$

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

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 or in the case of the canonical HANK model with homogeneous incidence, $D_t = 0$ for all t .

5.4 Optimal Simple Policy Rules

The previous section characterizes the optimal non-parametric optimal policy, which solve for the best achievable allocations among all possible policies. It is possible to fully

characterize those optimal policies using numerical methods, which we do in Section 6. However, as is well-known, this optimal policy may involve complex time- and state-dependent policies that depend on the entire history of shocks. In this section, we analyze the optimal policy within a restricted class of simple policy rules, as in Woodford (1999). This analysis has the additional benefit that we can provide closed-form expressions for the optimal policy rules.

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: the demand shock δ_t has persistence parameter ρ_δ , the natural output shock z_t has persistence parameter ρ_z , the price cost-push shock $u_{p,t}$ has persistence parameter ρ_{u_p} , and the wage cost-push shock $u_{w,t}$ has persistence parameter ρ_{u_w} . We assume that the initial conditions for each shock are given by their zero unconditional mean.

We now characterize the optimal simple policy rules in the Sticky-Price and Sticky-Wage HANK models, separately.¹⁰

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

Proposition 9. *Consider the Sticky-Price models with simple policy rules of the form:*

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{y}_t^* + \phi_{y,u}u_{p,t} \quad (39)$$

$$\hat{\pi}_t = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{y}_t^* + \phi_{\pi,u}u_{p,t}. \quad (40)$$

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

- i. Zero average inflation and output gap: $\phi_y = \phi_y^{\text{RANK}} \equiv 0$ and $\phi_\pi = \phi_\pi^{\text{RANK}} \equiv 0$.
- ii. Full accommodation of demand shocks: $\phi_{y,\delta} = \phi_{y,\delta}^{\text{RANK}} \equiv 0$ and $\phi_{\pi,\delta} = \phi_{\pi,\delta}^{\text{RANK}} \equiv 0$.
- iii. Optimal response to natural output shocks: $\phi_{y,z} = \phi_{y,z}^{\text{RANK}} \equiv 1$ and $\phi_{\pi,z} = \phi_{\pi,z}^{\text{RANK}} \equiv 0$.
- iv. Leaning against cost-push shocks:

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

¹⁰Focusing on one source of rigidity at a time allows us to derive closed form expressions for the optimal policy.

where $\kappa_p \equiv \lambda_p \{1 + v^{-1}\}$.

The optimal policy rules in the Sticky-Wage HANK model are a little more subtle, since in general optimal policy will not coincide with that in its RANK counterpart. However, we can still characterize what these optimal policies imply. 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. Since the simple policy rules are time-invariant, the interaction between incomplete markets and the policy (which decay asymptotically for the non-parametric optimal policy) no longer appear.

Proposition 10 (Simple Rules in a Sticky-Wage HANK Model). *Consider the Sticky-Wage model with policy rules of the form:*

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{y}_t^* + \phi_{y,u}u_{w,t} \quad (41)$$

$$\hat{\pi}_{w,t} = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{y}_t^* + \phi_{\pi,u}u_{w,t}. \quad (42)$$

Then, the optimal policy rules in HANK satisfy:

- i. Zero average inflation and output gap: $\phi_y = \phi_y^{\text{RANK}} \equiv 0$ and $\phi_\pi = \phi_\pi^{\text{RANK}} \equiv 0$.
- ii. Full accommodation of demand shocks: $\phi_{y,\delta} = \phi_{y,\delta}^{\text{RANK}} \equiv 0$ and $\phi_{\pi,\delta} = \phi_{\pi,\delta}^{\text{RANK}} \equiv 0$.
- iii. Optimal response to natural output shocks: $\phi_{y,z} = \phi_{y,z}^{\text{RANK}} \equiv 1$ and $\phi_{\pi,z} = \phi_{\pi,z}^{\text{RANK}} \equiv 0$.
- iv. Weaker leaning against cost-push shocks:

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

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

where $\kappa_w \equiv \lambda_w \{1 + v^{-1}\} > 0$ and $\tilde{\psi} \equiv \frac{\psi}{1 + \frac{\text{Var}\chi_H[\gamma_H]}{(1+v^{-1})} \left[v^{-1} + \frac{(1-\beta)}{(1-\beta\rho_u)^2(1+\beta)} \right]} > 0$.

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 central bank focuses more heavily on output stabilization than in the RANK counterpart.

6 Numerical Illustration in a Sticky-Wage HANK Model

In this section, we quantify the optimal policy in a calibrated HANK model. The goal is to understand the quantitative importance of the differences between optimal policy in the HANK and RANK model. 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 calibrate the model to match key features of the US economy, and we compute the optimal policy using the second-order perturbation method described in Theorem 1. We also compute the optimal simple policy rules as in Proposition 10, and we compare the optimal policies in the HANK model to those in the RANK model. We find that, while the optimal policy in the HANK model does not coincide with that in the RANK model, the differences are quantitatively small.

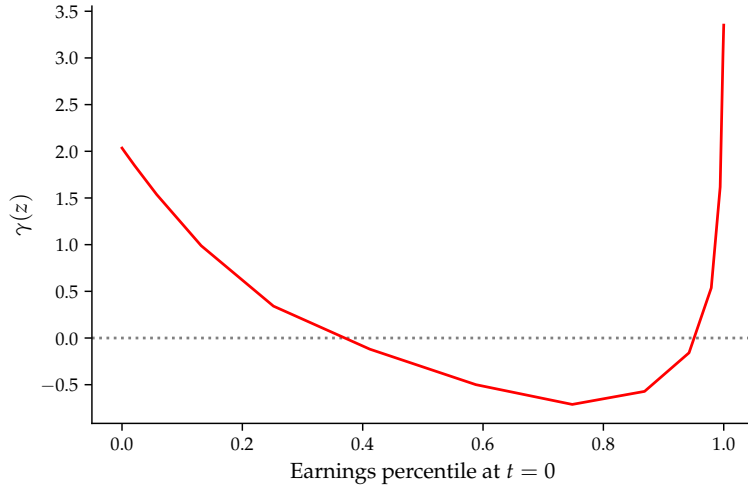
Calibration. We calibrate the model to a period length of one quarter. The elasticity of intertemporal substitution is one, and we set χ so that in steady state the Frisch elasticity of labor supply is $\nu = \frac{L}{1-L} = 0.75$, which are standard parameter values. We set technology shifter $Z = 1.75$ so as to normalize steady-state output to one. We assume that the aggregate technology shock process follows a log AR(1) process with persistence $\rho_z = 0.95$ and standard deviation $\sigma_z = 0.75$. We assume that 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 $\lambda_p \rightarrow \infty$, 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 .

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 employment. The function is calibrated using worker-level earnings betas from Guvenen et al. (2017); see Appendix A.4.1 for details. Higher values of $\gamma(z)$ indicate that the corresponding productivity group bears a disproportionately large share of aggregate employment fluctuations.

To model the cyclical risk of income, we follow Auclert and Rognlie (2018) and assume that the incidence elasticity is only a function of the household’s current productivity 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 A.4.1 for details. We plot the implied $\gamma(z)$ in Figure 1. The calibrated incidence function is U-shaped — the income of very low and very high productivity agents are more procyclical.

Targeting Criterion. The optimal policy obeys the targeting criterion in Proposition 7. Table 2 lists the parameters of the optimum targeting criterion implied by our calibration. In our parametrization, the (expenditure-weighted) variance of the incidence elasticity is given by $Var_{\chi_h}(\gamma_h) = 0.78$. This value implies that $\hat{\psi} = 2.76$, which represents an increase in the weight attributed to output stabilization relative to inflation stabilization of 45%.

The second component of the targeting criterion, whose magnitude is captured by ζ_{IM} , captures the optimal response to the interaction between incomplete markets and the dis-

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

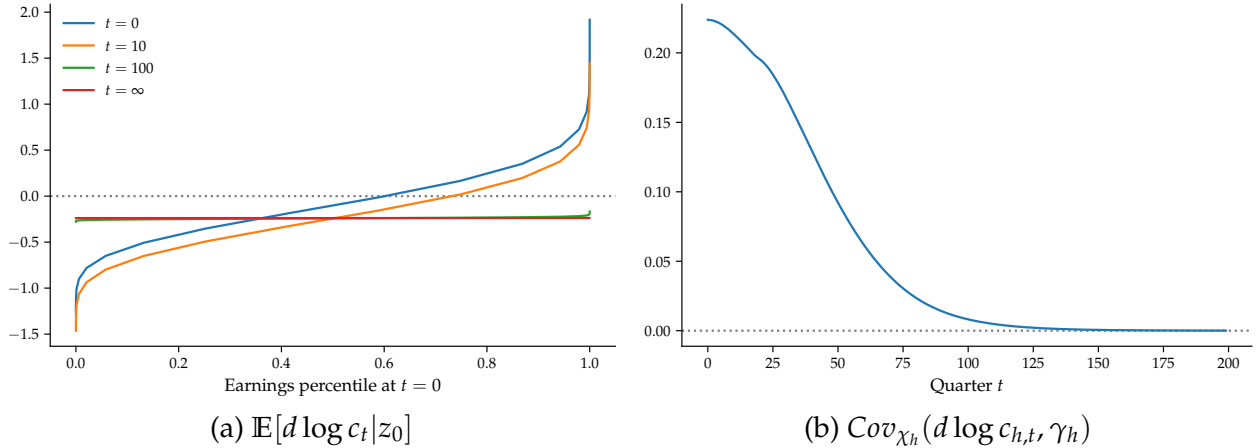
Notes: This table reports the key coefficients entering the targeting criterion in Proposition 7. The upper panel lists the primitive parameters: the elasticity of substitution across labor varieties (ψ), the Frisch elasticity (ν), and the variance of the incidence elasticity $Var_{\chi_h}(\gamma_h)$. 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.

tribution of labor wedges across households. This covariance term depends on the covariance of expected consumption at time t as a function of initial productivity $\mathbb{E}[d \log c_{h,t} | z_0]$ and the cyclical income $\gamma(z_0)$. The first panel of Figure 2 plots expected consumption at time t as a function of the initial productivity level, $\mathbb{E}[d \log c_{h,t} | z_0]$. Recall that $d \log c_{h,t}$ represents the log-change in consumption at time t relative to the $\sigma = 0$ allocation. Early on, for t close to zero, expected consumption is strongly increasing in initial productivity. However, since the model is ergodic, in the long-run, consumption is constant as a function of initial productivity. This helps explain why the covariance term in the targeting principle asymptotically decays to zero in Corollary 1.

The second panel of Figure 2 plots the covariance term, $Cov_{\chi_h}(d \log c_{h,t}, \gamma_h)$, over time t . Recall that, for the targeting criterion, the relevant object is the difference between the covariance at times t and $t - 1$: $Cov_{\chi_h}(d \log c_{h,t}, \gamma_h) - Cov_{\chi_h}(d \log c_{h,t-1}, \gamma_h)$. The covariance is highest at $t = 0$ and falls over time eventually reaching zero in the long-run. The fact that the covariance is positive reflects the fact that the cyclical income rises particularly fast at the top of the income distribution, where households consume more and have a higher expenditure share 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 8, we decompose the response of optimal policy into a decaying deterministic component, D_t , that captures the deviation of optimal policy from the Dual-Mandate policy, and a stochastic component

Figure 2: Incomplete Market Terms in Targeting Criterion



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 7. Both panels are computed under the baseline calibration reported in Table 1.

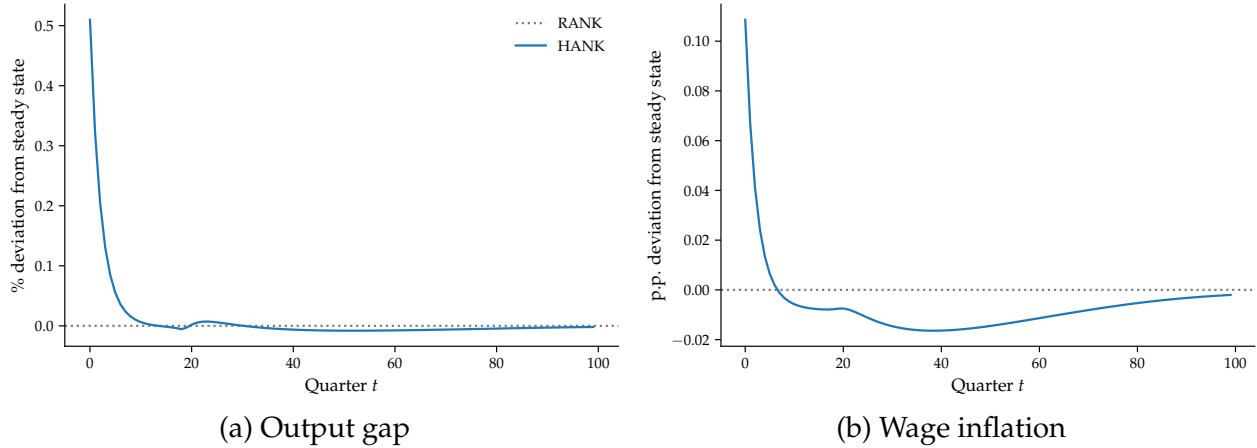
that captures the response of optimal policy to shocks.

Deterministic Component. Figure 3 plots the deterministic component, D_t . The figure shows that the optimal policy implements a small expansion initially. Output peaks at 0.5% and wage inflation at 0.1% above steady-state on impact and decays quickly. The effects are mostly gone after one year.

The intuition for this initial boom is as follows. In the targeting criterion (35), the relevant object is the change in the covariance across two consecutive periods. The right panel of 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 efficiency can be improved by a policy that causes high-productivity households to work more (since they are overconsuming leisure compared to first-best) and to subsidize leisure for households with relatively low productivity. In our cali-

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 8). 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 $t \rightarrow \infty$. Baseline calibration; see Table 1 for parameter values.

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

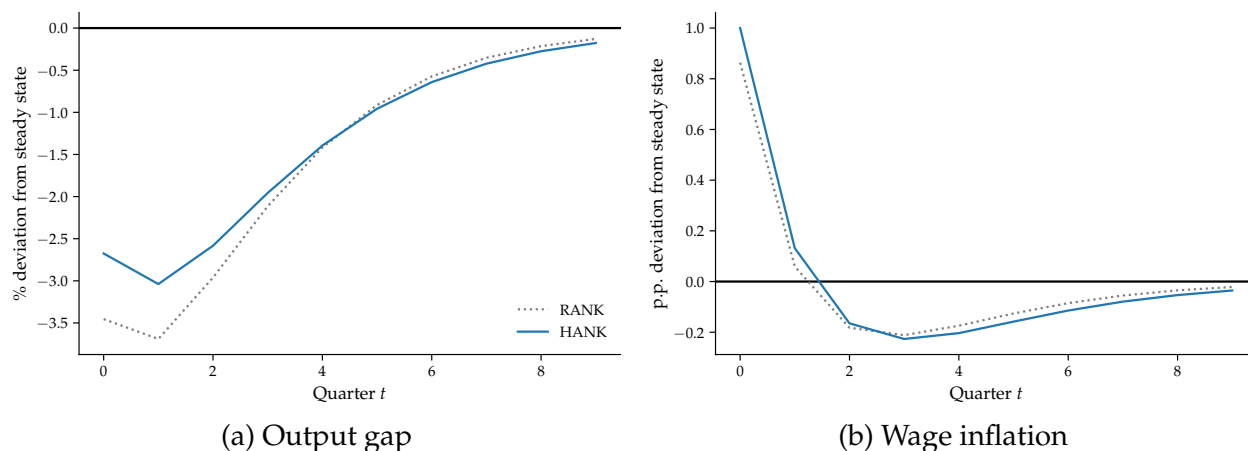
Response to aggregate shocks. We now turn to the response of optimal policy to aggregate shocks. We have already established that the optimal policy in the general HANK model effectively offsets the demand shock δ_t , so that output and inflation are unaffected by this shock. In this model with only 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 11% 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 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. These rules take the form in Proposition 10

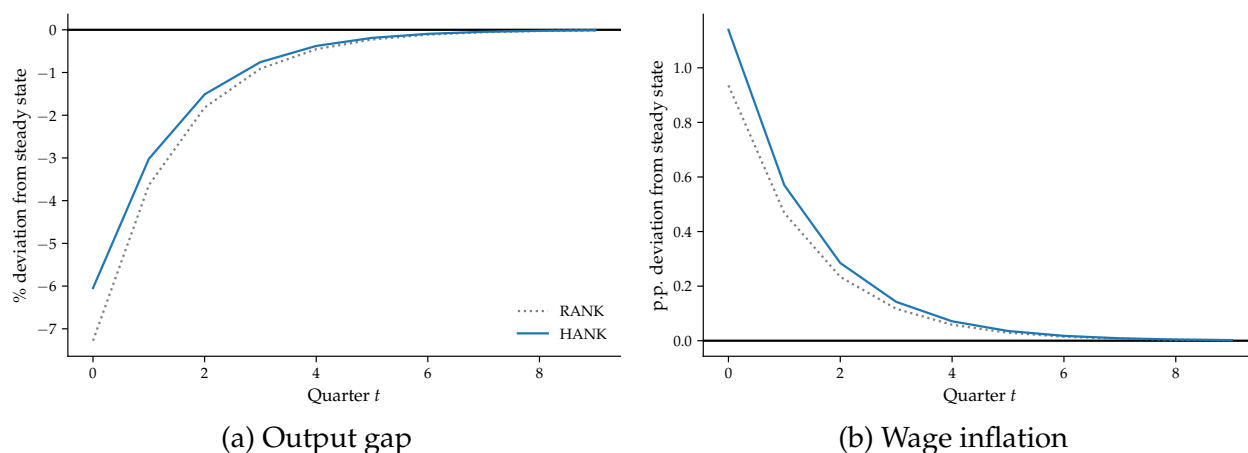
The HANK simple rule responds less aggressively to the output gap (in absolute value) and more aggressively to wage inflation than the RANK rule, consistent with the

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 ($\frac{\phi_{u,y}}{\phi_{u,y}^{RANK}} - 1 = -0.17$) and a 22% larger wage-inflation response than the RANK simple rule ($\frac{\phi_{u,\pi}}{\phi_{u,\pi}^{RANK}} - 1 = 0.22$). Note that, because inflation and the output gap are the product of these coefficients and the cost push shock, the percentage differences in the coefficients directly translate into percentage differences in the impact and cumulative responses of inflation and the output gap.

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.

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 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.

7 Conclusion

This paper studies optimal monetary policy in HANK economies under an Efficiency Mandate that ranks policies by their implications for the measure of aggregate productivity in Baqaee and Burstein (2025a). We measure the increase in aggregate productivity due to a policy reform by the amount of factor endowments that can be taken away after the reform while compensating every household to keep them indifferent to the status quo. 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 Harberger triangles representation. This quadratic objective nests the standard New Keynesian loss function and augments it with additional deadweight loss 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 in the presence of heterogeneity in the incidence of aggregate demand fluctuations. Our analysis suggests that, the key difference is that the optimal policy in HANK puts more weight on output-gap stabilization than in RANK, and this difference is quantitatively meaningful.

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. (2019). Monetary policy and the redistribution channel. *American Economic Review* 109(6), 2333–2367.
- 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.
- Ball, L., N. G. Mankiw, and R. Reis (2005). Monetary policy for inattentive economies. *Journal of monetary economics* 52(4), 703–725.
- Baqaae, D. and A. Burstein (2025a, August). Aggregate productivity with heterogeneous agents. Working Paper 34176, National Bureau of Economic Research.
- Baqaae, D. and A. Burstein (2026). Aggregate productivity with discrete choice. Technical report, National Bureau of Economic Research.
- Baqaae, D., A. Burstein, and J. Guerreiro (2026). The approximate general theory of second best. Working paper.
- Baqaae, D. R. and A. Burstein (2025b, September). Efficiency costs of incomplete markets. Working Paper 34233, National Bureau of Economic Research.
- Baqaae, 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 mone-*

- tary 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.
- Coglianesi, J., M. Olsson, and C. Patterson (2025). Monetary policy and the labor market: A quasi-experiment in sweden. *American Economic Review* 115(10), 3451–3486.
- 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.
- Del Canto, F., J. Grigsby, E. Qian, and C. Walsh (2025). Are inflationary shocks regressive? a feasible set approach. *The Quarterly Journal of Economics* 140(4), 2685–2747.
- 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.
- King, R. G. and S. T. Rebelo (1999). Chapter 14 resuscitating real business cycles. Volume 1

- of *Handbook of Macroeconomics*, pp. 927–1007. Elsevier.
- 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.
- Lucas, R. E. (2003). Macroeconomic priorities. *American economic review* 93(1), 1–14.
- 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.
- Patterson, C. (2023). The matching multiplier and the amplification of recessions. *American Economic Review* 113(4), 982–1012.
- Rubbo, E. (2023). Networks, phillips curves, and monetary policy. *Econometrica* 91(4), 1417–1455.
- Rubbo, E. (2026). A theory of worker betas: Aggregate demand incidence across the labor income distribution. Technical report.
- 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.
- Schulz, K., A. Tsyvinski, and N. Werquin (2023). Generalized compensation principle. *Theoretical Economics* 18(4), 1665–1710.
- 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.

Appendix A Proofs and Additional Results

A.1 Appendix to Section 2

A.1.1 Intermediate Goods Firms.

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

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

where

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

where W_t 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

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

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 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:

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

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

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

A.1.2 Aggregate Demand

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

$$y_{h,t} = \tilde{W}_t \cdot (1 - L_t) \cdot z_{h,t} \cdot \Gamma(1 - L_t; \zeta_{h,t}), \quad (43)$$

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

The aggregate demand of the economy is given by

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

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

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

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.

A.1.3 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.¹¹ As in the case of the intermediate goods producer, we assume that there is a subsidy $\tau_{w,t}$ that undoes monopoly distortions and allows us to capture cost-push shocks.

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

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

The wage subsidy undoes the monopoly distortion in the labor market and is financed by taxes on households, so that households receive 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

$$w_t^* = \frac{\psi}{\psi-1} \frac{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k \chi u'(L_{t+k}) \left(\frac{1}{W_{t+k}}\right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}{\mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k u'(C_{t+k}) \frac{(1+\tau_{w,t+k})}{P_{t+k}} \left(\frac{1}{W_{t+k}}\right)^{-\psi} \frac{\int y_{i,t+k} di}{Z_{t+k}}}. \quad (44)$$

A.1.4 Linearized Equilibrium Conditions

We approximate the equilibrium conditions around an efficient steady state, where all variables are constant over time. To achieve this steady state, we index exogenous variables by their deviations from steady state: $\log Z_t^\sigma = (1 - \sigma_Z) \log Z + \sigma_Z \log Z_t$, $\log(1 + \tau_{p,t}^\sigma) = (1 - \sigma_{\tau,p}) \log(1 + \tau_p) + \sigma_{\tau,p} \log(1 + \tau_{p,t})$, and $\log(1 + \tau_{w,t}^\sigma) = (1 - \sigma_{\tau,w}) \log(1 + \tau_w) + \sigma_{\tau,w} \log(1 + \tau_{w,t})$, $\log(1 + \delta_t^\sigma) = \sigma_\delta \cdot \log(1 + \delta_t)$, $\log z_{h,t}^\sigma = (1 - \sigma_z) \log z_{h,0} + \sigma_z \log z_{h,t}$, $a_{h,0}^\sigma = \sigma_a a_{h,0}$, and $\log(1 + i_t^\sigma) = (1 - \sigma_i) \log(1 + i) + \sigma_i \log(1 + i_t)$. We assume that $(1 + \tau_p) \frac{\varepsilon-1}{\varepsilon} = (1 + \tau_w) \frac{\psi-1}{\psi} = 1$, so that the steady state is efficient. We assume that i is such that the steady state has zero inflation, i.e., $i = \beta^{-1} - 1$. Note that our approx-

¹¹Alternatively, 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.

imation preserves the cross-sectional distribution of labor income, which is crucial for capturing the efficiency implications of cyclical heterogeneity.

Let $\sigma = \{\sigma_Z, \sigma_{\tau,p}, \sigma_{\tau,w}, \sigma_\delta, \sigma_z, \sigma_a\}$. When $\sigma = 1$, we have the full non-linear model. We approximate the economy around $\sigma = 0$ and evaluate at $\sigma = 1$. At $\sigma = 0$, the equilibrium is efficient and all wedges are zero.

We denote by dx_t a variable deviation from its steady-state value and by $\hat{x}_t = d \log x_t$ the log deviation between $\sigma = 1$ and $\sigma = 0$.

The Price New Keynesian Phillips Curve (13). The firm's optimal reset price can be written as

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

Let $MC_{t+k} \equiv \frac{W_{t+k}}{Z_{t+k} P_{t+k}}$ be the real marginal cost. Log-linearizing around the efficiency steady-state allocation, we obtain

$$\begin{aligned} \hat{p}_t^* - \hat{p}_{t-1} &= (1 - \beta\theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [(1 - \varepsilon)(\hat{p}_{t+k} - \hat{p}_{t-1}) + \hat{y}_{t+k} + \hat{m}c_{t+k}] \\ &\quad - (1 - \beta\theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\hat{\tau}_{p,t+k} + \varepsilon(\hat{p}_{t+k} - \hat{p}_{t-1}) + \hat{y}_{t+k}] \end{aligned}$$

where $\hat{m}c_{t+k} = d \log MC_{t+k} = \hat{w}_t - \hat{z}_t$, $\hat{w}_t = d \log(W_t/P_t)$ and $\hat{z}_t \equiv d \log Z_t$, $\hat{p}_t^* = d \log p_t^*$, $\hat{p}_t = d \log P_t$, $\hat{\tau}_{p,t} = d \log(1 + \tau_{p,t})$, and $\hat{y}_t = d \log Y_t$. Rearranging terms and defining $\pi_t \equiv \hat{p}_t - \hat{p}_{t-1}$, we obtain

$$\hat{p}_t^* - \hat{p}_{t-1} = (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\hat{m}c_{t+k}] + \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\pi_{t+k}] - (1 - \beta\theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\hat{\tau}_{p,t+k}].$$

Now, we use the fact that in the Calvo model, the price level is given by

$$P_t^{1-\varepsilon} = \theta_p P_{t-1}^{1-\varepsilon} + (1 - \theta_p)(p_t^*)^{1-\varepsilon} \Rightarrow \pi_t = (1 - \theta_p)(\hat{p}_t^* - \hat{p}_{t-1}).$$

Combining the last two equations, we obtain

$$\begin{aligned} \pi_t &= (1 - \beta\theta_p)(1 - \theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\hat{m}c_{t+k}] + (1 - \theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\pi_{t+k}] \\ &\quad - (1 - \beta\theta_p)(1 - \theta_p) \sum_{k=0}^{\infty} (\beta\theta_p)^k \mathbb{E}_t [\hat{\tau}_{p,t+k}], \end{aligned}$$

which we can write recursively as

$$\pi_t = \frac{(1 - \beta\theta_p)(1 - \theta_p)}{\theta_p} \hat{m}c_t + \beta \mathbb{E}_t[\pi_{t+1}] - \frac{(1 - \beta\theta_p)(1 - \theta_p)}{\theta_p} \hat{\tau}_{p,t}.$$

Defining $\lambda_p \equiv \frac{(1 - \beta\theta_p)(1 - \theta_p)}{\theta_p}$ and $u_{p,t} \equiv -\lambda_p \hat{\tau}_{p,t}$, we obtain the Price New Keynesian Phillips Curve:

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

The Wage New Keynesian Phillips Curve (14). The union's optimal reset wage can be written as

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

where $N_{t+k} = \int y_{i,t+k} / Z_{t+k} di$ is aggregate labor supply. Let $MRS_{t+k}^w \equiv \frac{\chi u'(L_{t+k})}{u'(C_{t+k}) W_{t+k} / P_{t+k}}$ denote the ratio of the marginal rate of substitution to the real wage. Log-linearizing around the efficiency steady-state allocation, we obtain

$$\begin{aligned} \hat{w}_t^* - \hat{w}_{t-1}^{nom} &= (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t \left[-\hat{l}_{t+k} + \psi(\hat{w}_{t+k}^{nom} - \hat{w}_{t-1}^{nom}) + \hat{n}_{t+k} \right] \\ &\quad - (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t \left[-\hat{y}_{t+k} + \hat{\tau}_{w,t+k} - (\hat{p}_{t+k} - \hat{p}_{t-1}) + (\hat{w}_{t-1}^{nom} - \hat{p}_{t-1}) \right. \\ &\quad \left. + \psi(\hat{w}_{t+k}^{nom} - \hat{w}_{t-1}^{nom}) + \hat{n}_{t+k} \right] \end{aligned}$$

where $\hat{w}_t^* = d \log w_t^*$, $\hat{w}_t^{nom} = d \log W_t$, $\hat{l}_t = d \log L_t$, $\hat{n}_t = d \log N_t$, and $\hat{\tau}_{w,t} = d \log(1 + \tau_{w,t})$. Note that the $\psi(\hat{w}_{t+k}^{nom} - \hat{w}_{t-1}^{nom})$ and \hat{n}_{t+k} terms cancel between numerator and denominator. After cancelling these terms, using $\hat{w}_{t-1}^{nom} - \hat{p}_{t-1} = \hat{w}_{t-1}$, and collecting terms using $\hat{m}rs_{t+k}^w \equiv \hat{y}_{t+k} - \hat{l}_{t+k} - \hat{w}_{t+k}$, we can write the difference as

$$\hat{w}_t^* - \hat{w}_{t-1}^{nom} = (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t \left[\hat{m}rs_{t+k}^w + (\hat{w}_{t+k}^{nom} - \hat{w}_{t-1}^{nom}) \right] - (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t \left[\hat{\tau}_{w,t+k} \right]$$

where $m\hat{r}s_{t+k}^w \equiv d \log MRS_{t+k}^w = \hat{y}_{t+k} - \hat{l}_{t+k} - \hat{w}_{t+k}$. Defining $\pi_{w,t} \equiv \hat{w}_t^{nom} - \hat{w}_{t-1}^{nom}$ and noting that $(1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k (\hat{w}_{t+k}^{nom} - \hat{w}_{t-1}^{nom}) = \sum_{k=0}^{\infty} (\beta\theta_w)^k \pi_{w,t+k}$, we obtain

$$\hat{w}_t^* - \hat{w}_{t-1}^{nom} = (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [m\hat{r}s_{t+k}^w] + \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [\pi_{w,t+k}] - (1 - \beta\theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [\hat{\tau}_{w,t+k}].$$

Now, we use the fact that in the Calvo model, the wage index is given by

$$W_t^{1-\psi} = \theta_w W_{t-1}^{1-\psi} + (1 - \theta_w)(w_t^*)^{1-\psi} \Rightarrow \pi_{w,t} = (1 - \theta_w)(\hat{w}_t^* - \hat{w}_{t-1}^{nom}).$$

Combining the last two equations, we obtain

$$\begin{aligned} \pi_{w,t} &= (1 - \beta\theta_w)(1 - \theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [m\hat{r}s_{t+k}^w] + (1 - \theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [\pi_{w,t+k}] \\ &\quad - (1 - \beta\theta_w)(1 - \theta_w) \sum_{k=0}^{\infty} (\beta\theta_w)^k \mathbb{E}_t [\hat{\tau}_{w,t+k}], \end{aligned}$$

which we can write recursively as

$$\pi_{w,t} = \frac{(1 - \beta\theta_w)(1 - \theta_w)}{\theta_w} m\hat{r}s_t^w + \beta \mathbb{E}_t [\pi_{w,t+1}] - \frac{(1 - \beta\theta_w)(1 - \theta_w)}{\theta_w} \hat{\tau}_{w,t}.$$

Defining $\lambda_w \equiv \frac{(1 - \beta\theta_w)(1 - \theta_w)}{\theta_w}$ and $u_{w,t} \equiv -\lambda_w \hat{\tau}_{w,t}$, we obtain

$$\pi_{w,t} = \lambda_w m\hat{r}s_t^w + \beta \mathbb{E}_t [\pi_{w,t+1}] + u_{w,t}.$$

To express $m\hat{r}s_t^w$ in terms of the output gap and the real wage gap, note that with log utility, $\hat{l}_t = -\frac{1-L}{L}(\hat{y}_t - \hat{z}_t) = -\nu^{-1}(\hat{y}_t - \hat{z}_t)$, where $\nu \equiv L/(1-L)$ is the Frisch elasticity of labor supply. This follows from $Y_t = Z_t(1-L)$ and $C_t = Y_t$. Furthermore, in the efficient allocation, leisure is constant at $L^* = \chi/(1+\chi)$, so that $\hat{y}_t^* = \hat{z}_t$. Therefore,

$$\begin{aligned} m\hat{r}s_t^w &= \hat{y}_t + \nu^{-1}(\hat{y}_t - \hat{z}_t) - \hat{w}_t = (1 + \nu^{-1})\hat{y}_t - \nu^{-1}\hat{y}_t^* - \hat{w}_t \\ &= \left\{1 + \nu^{-1}\right\} (\hat{y}_t - \hat{y}_t^*) - (\hat{w}_t - \hat{z}_t). \end{aligned}$$

Substituting, we obtain the Wage New Keynesian Phillips Curve:

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

The Law of Motion for the Real Wage (15). The real wage is defined as $\hat{w}_t \equiv d \log(W_t/P_t) = \hat{w}_t^{nom} - \hat{p}_t$. Therefore,

$$\hat{w}_t - \hat{w}_{t-1} = (\hat{w}_t^{nom} - \hat{w}_{t-1}^{nom}) - (\hat{p}_t - \hat{p}_{t-1}) = \pi_{w,t} - \pi_t,$$

which gives the law of motion for the real wage:

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

Aggregate Demand (16). The aggregate demand of the economy is given by

$$C_t = \int c_{h,t} dh. \quad (45)$$

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

$$C_t = \mathcal{C} \left(\{ \{ \mathbb{E}_s[\Theta_{s+h}] \}_{h=0}^\infty \}_{s=-\infty}^t \right), \quad (46)$$

where $\Theta_t \equiv \{L_t, Y_t, i_t, \delta_t, P_t/P_{t-1}, \{q_k\}, \{z_{h,t}\}\}$, for $t \geq 1$, and

$\Theta_0 \equiv \{L_0, Y_0, i_0, \delta_0, P_0/P_{-1}, \{q_k\}, \{T_{h,0}, a_{h,0}, \{z_{h,0}\}\}\}$ 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 $s < 0$. Linearizing around the steady state, we represent by $d\Theta_t$ the deviation of each variable in Θ_t from its steady state value, induced by the deviation of the exogenous variables from their steady state values.

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.

Following Angeletos et al. (2025) for properties of this functional, we can write the linearized aggregate demand as:

$$\mathbb{E}_t[d\mathbf{C}_{t:\infty}] = \mathbb{E}_{t-1}[d\mathbf{C}_{t:\infty}] + \sum_{\Theta} \mathbf{M}^\Theta (\mathbb{E}_t[d\Theta_{t:\infty}] - \mathbb{E}_{t-1}[d\Theta_{t:\infty}]).$$

A.1.5 Proof of Proposition 1

See Baqaee et al. (2026).

A.1.6 Proof of Proposition 2

In the technological frontier, the marginal rates of intertemporal substitution are equalized across households and the marginal rates of substitution between consumption and leisure divided by idiosyncratic productivity are equalized across households. It follows that there exists $\omega_h \geq 0$ such that

$$c_{h,t} = \omega_h C_t, \quad \text{and} \quad l_{h,t} = \frac{\omega_h L_t}{z_{h,t}}.$$

and $\int \omega_h dh = 1$.

Furthermore, these optimality conditions, together with the production function imply that the frontier allocations satisfy the following conditions:

$$\frac{\chi C_t}{L_t} = Z_t, \quad \text{and} \quad Y_t = C_t = Z_t(1 - L_t).$$

Solving this system of equations implies (24).

Finally, as in Example 1, since preferences are homothetic, the frontier allocations can also be derived in scaled variables, which implies that

$$c_{h,t} = \omega_h \frac{C_t^*}{H}, \quad \text{and} \quad l_{h,t} = \frac{\omega_h L_t^*}{z_{h,t} H}.$$

Then, the utility of agent h is given by

$$U(\mathbf{x}_h^*/H) = \log \omega_h - \frac{\chi}{1 + \chi} \log \bar{z}_h - \log H + U(\mathbf{X}^*),$$

where $\log \bar{z}_h \equiv (1 - \beta) \mathbb{E}_0 [\sum_{t=0}^{\infty} \beta^t \log z_{h,t}]$.

At $H = \mathcal{D}$, it must be that

$$U(\mathbf{x}_h^*/\mathcal{D}) = \log \omega_h - \frac{\chi}{1 + \chi} \log \bar{z}_h - \log \mathcal{D} + U(\mathbf{X}^*) = U(\mathbf{x}_h^0)$$

for all h .

Solving these equations first for $\tilde{\omega}_h$ implies

$$\begin{aligned}\omega_h &= \frac{\bar{z}_h^{\frac{\chi}{1+\chi}} \exp\{U(\mathbf{x}_h^0)\}}{\exp\{U(\mathbf{X}^*)\}} \cdot \mathcal{D} \Rightarrow 1 = \int \omega_h dh = \frac{\int \bar{z}_h^{\frac{\chi}{1+\chi}} \exp\{U(\mathbf{x}_h^0)\} dh}{\exp\{U(\mathbf{X}^*)\}} \cdot \mathcal{D} \\ \Leftrightarrow \log \mathcal{D} &= U(\mathbf{X}^*) - \log \int \bar{z}_h^{\frac{\chi}{1+\chi}} \exp\{U(\mathbf{x}_h^0)\} dh.\end{aligned}$$

A.1.7 Further Details on Examples 1 and 2

In the RANK model, $\mathcal{X}(1/H, \mathcal{P})$ is a singleton for all H , since there is a unique equilibrium allocation for each policy, i.e., no transfers are possible since all agents are identical.

Then, note that, because preferences are homothetic, it is easy to show that the equilibrium can be derived in scaled variables, i.e., for each variable x , we can write $x = H\tilde{x}$, where \tilde{x} is the equilibrium value of the variable in the RANK model under policy \mathcal{P} and $H = 1$.

It follows that the utility of the representative agent is $U(\mathbf{x}(1, \mathcal{P})/H)$. Then, aggregate productivity is given by:

$$A(\mathcal{P}) \equiv \max\{H > 0 : U(\mathbf{x}(1, \mathcal{P})/H) \geq U(\mathbf{x}^0)\}.$$

Since U is strictly increase in all arguments we can write

$$\begin{aligned}U(\mathbf{x}(1, \mathcal{P})/A(\mathcal{P})) = U(\mathbf{x}^0) &\Leftrightarrow U(\mathbf{x}(1, \mathcal{P})) - \log A(\mathcal{P}) = U(\mathbf{x}^0) \\ \Leftrightarrow \log A(\mathcal{P}) &= U(\mathbf{x}(1, \mathcal{P})) - U(\mathbf{x}^0).\end{aligned}$$

Note that, using the results of Proposition 2, we can write

$$\log A(\mathcal{P}) = U(\mathbf{x}(1, \mathcal{P})) - U(\mathbf{x}^0) = \log \mathcal{D} + [U(\mathbf{x}(1, \mathcal{P})) - U(\mathbf{X}^*)].$$

Then, to second order, $\log A(\mathcal{P})$ is equal to the difference between $\log \mathcal{D}$ and the second-order approximation of the loss $\mathcal{L} \equiv U(\mathbf{X}^*) - U(\mathbf{x}(1, \mathcal{P}))$. This loss \mathcal{L} is the standard object approximated in Woodford (2003) for the case with only one source of nominal rigidity, or in Erceg et al. (2000) for the case with both price and wage rigidities. So, following the standard approximation procedure, we obtain that

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

A.2 Appendix to Section 4

A.2.1 The Arrow-Debreu Economy with Wedges

We now show formally that the baseline equilibrium allocations can be implemented as the allocation of an Arrow-Debreu economy with wedges. Fix any equilibrium of the baseline economy and denote its allocation and prices by

$$\left\{ c_{h,t}, l_{h,t}, C_t, Y_t, y_{i,t}, n_{i,u,t}, n_{u,t}^d, P_t, p_{i,t}, W_t, w_{u,t}, \tilde{W}_t \right\}_{h,i,u,t}.$$

The Arrow-Debreu economy is static in the sense that all state-contingent commodities dated at all histories are traded at date 0. Let Q_t denote the Arrow-Debreu price of one unit of the final good in period t , normalized so that

$$Q_t = \beta^t \frac{u'(C_t)}{P_t}.$$

Household h chooses $\{c_{h,t}, l_{h,t}\}_{t=0}^{\infty}$ to maximize (1) subject to

$$\mathbb{E}_0 \sum_{t=0}^{\infty} Q_t \mu_{h,t}^{IM} \left(P_t c_{h,t} + \mu_{h,t}^L \tilde{W}_t l_{h,t} \right) \leq \mathcal{W}_h + T_{h,0}, \quad (47)$$

where \mathcal{W}_h is the value of household h 's endowments and profit claims in the Arrow-Debreu economy, and $T_{h,0}$ is a time-0 transfer. The transfers are chosen to satisfy (47) at equality for the baseline allocation; aggregating the household budget constraints and the firms' zero-profit conditions gives $\int T_{h,0} dh = 0$.

The wedges are chosen from the baseline equilibrium prices and allocations and are exactly those defined in (19)–(22). The price wedge makes each intermediate producer's price satisfy

$$p_{i,t} = \mu_{i,t}^p \frac{W_t}{Z_t}.$$

The wage wedge makes each union wage satisfy

$$w_{u,t} = \mu_{u,t}^w \tilde{W}_t,$$

where $\tilde{W}_t \equiv W_t^{1-\psi} \left(\int w_{u,t}^{-\psi} du \right)^{-1}$. The household labor wedge makes the household's marginal rate of substitution satisfy

$$\frac{\chi u'(l_{h,t})}{u'(c_{h,t})} = \mu_{h,t}^L \frac{\tilde{W}_t}{P_t},$$

and the incomplete-markets wedge satisfies

$$\mu_{h,t}^{IM} = \frac{u'(c_{h,t})/u'(C_t)}{u'(c_{h,0})/u'(C_0)}.$$

We verify that the baseline allocation satisfies the equilibrium conditions of this Arrow-Debreu economy. Let λ_h be the multiplier on household h 's Arrow-Debreu budget constraint and set

$$\lambda_h = \frac{1 - \beta u'(c_{h,0})}{1 + \chi u'(C_0)}.$$

The first-order condition for consumption in the Arrow-Debreu economy is

$$\frac{1 - \beta}{1 + \chi} \beta^t u'(c_{h,t}) = \lambda_h Q_t \mu_{h,t}^{IM} P_t.$$

Using the definition of Q_t and (22), the right-hand side is

$$\frac{1 - \beta}{1 + \chi} \frac{u'(c_{h,0})}{u'(C_0)} \beta^t u'(C_t) \frac{u'(c_{h,t})/u'(C_t)}{u'(c_{h,0})/u'(C_0)} = \frac{1 - \beta}{1 + \chi} \beta^t u'(c_{h,t}),$$

so the consumption first-order condition is satisfied. The first-order condition for leisure is

$$\frac{1 - \beta}{1 + \chi} \beta^t \chi u'(l_{h,t}) = \lambda_h Q_t \mu_{h,t}^{IM} \mu_{h,t}^L \tilde{W}_t.$$

By the consumption first-order condition, this is equivalent to

$$\frac{\chi u'(l_{h,t})}{u'(c_{h,t})} = \mu_{h,t}^L \frac{\tilde{W}_t}{P_t},$$

which holds exactly by (21). Thus each household optimally chooses the baseline paths $\{c_{h,t}, l_{h,t}\}_{t=0}^{\infty}$.

On the production side, firms in the Arrow-Debreu economy take wedges as taxes or markups that separate marginal rates of transformation from relative prices. Given (19), the baseline price $p_{i,t}$ is exactly the price implied by the Arrow-Debreu price wedge. The final-good producer is competitive and has the CES technology from the baseline economy. Its cost-minimization conditions imply

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

which are the same demand system and price index that hold in the baseline equilib-

rium. Hence the baseline allocation of intermediate goods is optimal for the final-good producer.

Similarly, (20) implies that the baseline wage $w_{u,t}$ is exactly the wage implied by the Arrow-Debreu wage wedge. Intermediate firms minimize costs over labor varieties using the same CES labor aggregator as in the baseline economy, so their labor demand satisfies

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

Since the wages $w_{u,t}$ and the wage index W_t are the baseline ones, these conditions implement exactly the baseline allocation of labor across unions and firms. The induced demand for union u is therefore

$$n_{u,t}^d = \int n_{i,u,t} di.$$

It remains to check the household allocation of labor. In the baseline economy, the household-level incidence of union labor demand is given by

$$n_{h,u,t}^s = \Gamma(1 - L_t; \zeta_{h,t}) \int n_{i,u,t} di,$$

with $\int z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) dh = 1$. Therefore

$$\int \int z_{h,t} n_{h,u,t}^s dh du = \int n_{u,t}^d du = 1 - L_t,$$

and the household leisure allocation in the Arrow-Debreu economy coincides with the baseline allocation. The wedge $\mu_{h,t}^L$ is exactly the wedge that makes this incidence-implied leisure allocation optimal for household h at the common price \tilde{W}_t/P_t .

Finally, goods and labor markets clear because they clear in the baseline equilibrium:

$$C_t = \int c_{h,t} dh = Y_t, \quad \int \int n_{i,u,t} di du = 1 - L_t.$$

Thus, for any baseline equilibrium allocation, the wedges defined in (19)–(22), together with Arrow-Debreu prices Q_t and budget-balancing transfers $\{T_{h,0}\}$, support exactly the same allocation as an equilibrium of the Arrow-Debreu economy with wedges. The four wedges therefore provide a complete decentralization of the sources of inefficiency in the baseline economy: price dispersion, wage dispersion, heterogeneous incidence of aggregate labor fluctuations, and incomplete financial markets.

A.2.2 Proof of Lemma 1

See Baqaee et al. (2026).

A.2.3 Proof of Lemma 2

Note that, in equilibrium, the introduction of wedges causes a change in the the final good's demand for intermediate good i as well as firm i 's demand for labor. The change in the final good's demand for intermediate good i is given by

$$\partial \log y_{i,t} = -\varepsilon \left(d \log \mu_{i,t}^p - \int d \log \mu_{i',t}^p di' \right) + \partial \log Y_t,$$

where $\partial \log Y_t$ captures the change in aggregate demand associated with the introduction of wedges, i.e., the output gap $\hat{y}_t - \hat{y}_t^*$.

The change in the demand for labor input u by firm i is given by

$$\partial \log n_{i,u,t} = -\psi \left(d \log \mu_{u,t}^w - \int d \log \mu_{u',t}^w du' \right) + \partial \log y_{i,t}.$$

Note that, to first order around the efficient allocation, $\int \partial \log y_{i,t} di = \partial \log Y_t$. So, we can write the change in the demand for labor input u , $n_{u,t}^d = \int n_{i,u,t} di$ as

$$\partial \log n_{u,t}^d = -\psi \left(d \log \mu_{u,t}^w - \int d \log \mu_{u',t}^w du' \right) + \partial \log Y_t.$$

We first solve the integral in the definition of the price-wedge contribution. Substituting the expression for $\partial \log y_{i,t}$ gives

$$\begin{aligned} \triangleright^p &\equiv (1 - \beta) \omega_C \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 \frac{d \log \mu_{i,t}^p \cdot \partial \log y_{i,t}}{2} di \\ &= (1 - \beta) \omega_C \mathbb{E}_0 \frac{1}{2} \sum_{t=0}^{\infty} \beta^t \int d \log \mu_{i,t}^p \left[-\varepsilon \left(d \log \mu_{i,t}^p - \int d \log \mu_{i',t}^p di' \right) + \partial \log Y_t \right] di \\ &= (1 - \beta) \omega_C \mathbb{E}_0 \frac{1}{2} \sum_{t=0}^{\infty} \beta^t \left[-\varepsilon \text{Var}_i(d \log \mu_{i,t}^p) + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right]. \end{aligned}$$

The variance of price markups equals the variance of prices, since W_t/Z_t is common across firms in period t . Hence, the standard Calvo price-setting recursion implies

$$\text{Var}_i(d \log \mu_{i,t}^p) = \text{Var}_i(d \log p_{i,t}) = \theta_p \text{Var}_i(d \log p_{i,t-1}) + \frac{\theta_p}{1 - \theta_p} \pi_t^2.$$

Plugging this recursion into the last equation and solving backwards, using $Var_i(d \log p_{i,-1}) = 0$, gives

$$\triangleright^p = \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\varepsilon \frac{\theta_p}{(1-\theta_p)(1-\beta\theta_p)} \pi_t^2 + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right].$$

Using $\lambda_p \equiv (1-\beta\theta_p)(1-\theta_p)/\theta_p$, this becomes

$$\triangleright^p = \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\varepsilon}{\lambda_p} \pi_t^2 + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right].$$

We next solve the integral in the definition of the wage-wedge contribution. Substituting the expression for $\partial \log n_{u,t}^d$ gives

$$\begin{aligned} \triangleright^w &\equiv (1-\beta)\omega_C \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 \frac{d \log \mu_{u,t}^w \cdot \partial \log n_{u,t}^d}{2} du \\ &= (1-\beta)\omega_C \mathbb{E}_0 \frac{1}{2} \sum_{t=0}^{\infty} \beta^t \int d \log \mu_{u,t}^w \left[-\psi \left(d \log \mu_{u,t}^w - \int d \log \mu_{u',t}^w du' \right) + \partial \log Y_t \right] du \\ &= (1-\beta)\omega_C \mathbb{E}_0 \frac{1}{2} \sum_{t=0}^{\infty} \beta^t \left[-\psi Var_u(d \log \mu_{u,t}^w) + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right]. \end{aligned}$$

The variance of wage markups equals the variance of wages, since \tilde{W}_t is common across unions in period t . Hence, the standard Calvo wage-setting recursion implies

$$Var_u(d \log \mu_{u,t}^w) = Var_u(d \log w_{u,t}) = \theta_w Var_u(d \log w_{u,t-1}) + \frac{\theta_w}{1-\theta_w} \pi_{w,t}^2.$$

Plugging this recursion into the last equation and solving backwards, using $Var_u(d \log w_{u,-1}) = 0$, gives

$$\triangleright^w = \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\psi \frac{\theta_w}{(1-\theta_w)(1-\beta\theta_w)} \pi_{w,t}^2 + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right].$$

Using $\lambda_w \equiv (1-\beta\theta_w)(1-\theta_w)/\theta_w$, this becomes

$$\triangleright^w = \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\psi}{\lambda_w} \pi_{w,t}^2 + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right],$$

which is the desired wage-wedge expression.

A.2.4 Proof of Lemma 3

Start from the Harberger triangle for the household-specific labor wedge, written in the normalized form used in the lemma,

$$\triangleright^L \equiv (1 - \beta)(1 - \omega_C)\mathbb{E}_0 \sum_t \beta^t \int_0^1 \chi_h \frac{d \log \mu_{h,t}^L \cdot \partial \log l_{h,t}}{2} dh.$$

The incidence rule implies that, to a first order around the efficient allocation,

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

Substituting this expression and collecting cross-sectional terms gives

$$\triangleright^L = -\frac{(1 - \beta)\omega_C}{2}\mathbb{E}_0 \sum_t \beta^t \left\{ E_{\chi_h} [d \log \mu_{h,t}^L] \partial \log Y_t + Cov_{\chi_h} (d \log \mu_{h,t}^L, \gamma_h) \partial \log Y_t \right\}. \quad (48)$$

Thus, the proof reduces to characterizing the average labor wedge and the component of the labor wedge that covaries with the incidence elasticity γ_h .

We first compute the average labor wedge. Using log utility, the definition of the labor wedge implies

$$d \log \mu_{h,t}^L = -(d \log l_{h,t} - d \log c_{h,t}) - (d \log \tilde{W}_t - d \log P_t).$$

The price and wage indices imply, to a first order,

$$d \log P_t = E_i [d \log \mu_{i,t}^p] + d \log W_t - d \log Z_t, \quad d \log W_t = E_u [d \log \mu_{u,t}^w] + d \log \tilde{W}_t.$$

Therefore,

$$\begin{aligned} d \log \tilde{W}_t - d \log P_t &= d \log W_t - E_u [d \log \mu_{u,t}^w] - \left(E_i [d \log \mu_{i,t}^p] + d \log W_t - d \log Z_t \right) \\ &= d \log Z_t - E_i [d \log \mu_{i,t}^p] - E_u [d \log \mu_{u,t}^w]. \end{aligned}$$

Taking the consumption-share weighted average across households gives

$$\begin{aligned} E_{\chi_h} [d \log \mu_{h,t}^L] &= -(E_{\chi_h} [d \log l_{h,t}] - E_{\chi_h} [d \log c_{h,t}]) - (d \log \tilde{W}_t - d \log P_t) \\ &= -(d \log L_t - d \log C_t) - d \log Z_t + E_i [d \log \mu_{i,t}^p] + E_u [d \log \mu_{u,t}^w]. \end{aligned}$$

Around the efficient allocation, $d \log C_t = d \log Z_t + \partial \log Y_t$ and $d \log L_t = -\nu^{-1}\partial \log Y_t$.

Hence,

$$E_{\chi_h}[d \log \mu_{h,t}^L] = (1 + \nu^{-1})\partial \log Y_t + E_i[d \log \mu_{i,t}^p] + E_u[d \log \mu_{u,t}^w]. \quad (49)$$

This is the average labor-wedge component. It combines the usual aggregate labor wedge, $(1 + \nu^{-1})\partial \log Y_t$, with the average price and wage markup wedges.

We next compute the dispersion component. Since the aggregate price term $d \log \tilde{W}_t - d \log P_t$ is common across households, it drops out of the covariance with γ_h . Therefore,

$$Cov_{\chi_h}(d \log \mu_{h,t}^L, \gamma_h) = Cov_{\chi_h}(-(d \log l_{h,t} - d \log c_{h,t}), \gamma_h).$$

Recall that we have assumed that $\Gamma(\cdot)$ is such that labor supply is optimal in the absence of aggregate shocks. This implies that, to a first order around the efficient allocation, the labor wedge does not respond to changes in idiosyncratic characteristics. It follows that

$$d \log \mu_{h,t}^L = - \left[(1 + \gamma_h)\partial \log L_t - \mathbb{E}_0 \sum_{s=0}^{\infty} \sum_{X_s \in \hat{\Theta}_s} \frac{\partial \log c_{h,t}}{\partial X_s} d X_s \right],$$

where $\hat{\Theta}_t \equiv \{Y_t, i_t, \delta_t, P_t/P_{t-1}, \{q_{k,t}\}\}$.

Now, using the fact that all aggregates except for L_t affect households symmetrically (i.e., generate $\partial \log c_{h,t} \partial \log X_s$ which are constant across h), we can write

$$\begin{aligned} Cov_{\chi_h}(d \log \mu_{h,t}^L, \gamma_h) &= Cov_{\chi_h} \left(-((1 + \gamma_h)d \log L_t - \mathbb{E}_0 \sum_{s=0}^{\infty} \overbrace{\frac{\partial \log c_{h,t}}{\partial \log L_s}} d \log L_s), \gamma_h \right) \\ &= -Var_{\chi_h}(\gamma_h) \partial \log L_t + Cov_{\chi_h} \left(\mathbb{E}_0 \sum_{s=0}^{\infty} \overbrace{\frac{\partial \log c_{h,t}}{\partial \log L_s}} d \log L_s, \gamma_h \right) \\ &= -Var_{\chi_h}(\gamma_h) \cdot d \log L_t - Var_{\chi_h}(\gamma_h) \mathbb{E}_0 \sum_{s=0}^{\infty} (1 - \beta)\beta^s d \log L_s \\ &= \nu^{-1} Var_{\chi_h}(\gamma_h) \cdot \partial \log Y_t + \nu^{-1} Var_{\chi_h}(\gamma_h) \mathbb{E}_0 \sum_{s=0}^{\infty} (1 - \beta)\beta^s \partial \log Y_s. \end{aligned}$$

So, substituting above we get

$$\begin{aligned} \triangleright^L &= -\frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\sum_t\beta^t\left\{\left[(1+\nu^{-1})\partial\log Y_t+E_i[d\log\mu_{i,t}^p]+E_u[d\log\mu_{u,t}^w]\right]\partial\log Y_t\right. \\ &\quad \left.+ \nu^{-1}\text{Var}_{\chi_h}(\gamma_h)\cdot(\partial\log Y_t)^2+\nu^{-1}\text{Var}_{\chi_h}(\gamma_h)\mathbb{E}_0\sum_{s=0}^{\infty}(1-\beta)\beta^s\partial\log Y_s\partial\log Y_t\right\}. \\ \Leftrightarrow \triangleright^L &= -\frac{(1-\beta)\omega_C}{2}\mathbb{E}_0\sum_t\beta^t\left\{\left[(1+\nu^{-1})\partial\log Y_t+E_i[d\log\mu_{i,t}^p]+E_u[d\log\mu_{u,t}^w]\right]\partial\log Y_t\right. \\ &\quad \left.+ \nu^{-1}\text{Var}_{\chi_h}(\gamma_h)\cdot(\partial\log Y_t)^2+\nu^{-1}(1-\beta)\text{Var}_{\chi_h}(\gamma_h)\mathbb{E}_0\left(\sum_{s=0}^{\infty}\beta^s\partial\log Y_s\right)^2\right\}. \end{aligned}$$

A.2.5 Proof of Lemma 4

The Harberger triangle for incomplete markets is given by

$$\triangleright^{IM}\equiv(1-\beta)\mathbb{E}_0\sum_t\beta^tE_{\chi_h}\left[\frac{d\log\mu_{h,t}^{IM}\cdot\partial\log x_{h,t}}{2}\right],$$

Note that, using household optimality, we find that

$$\beta^t\left(\frac{x_{h,t}}{x_{h,0}}\right)^{-1}=Q_t\frac{\mu_{h,t}^{IM}p_{h,t}^x}{\mu_{h,0}^{IM}p_{h,0}^x},$$

where $p_{h,t}^x\equiv P_t^{\frac{1}{1+\chi}}(\mu_{h,t}^L\tilde{W}_tz_{h,t})^{\frac{\chi}{1+\chi}}$ is the price of the composite good $x_{h,t}$ in the Arrow-Debreu economy. Therefore,

$$\frac{\mu_{h,t}^{IM}p_{h,t}^xx_{h,t}}{\mu_{h,0}^{IM}p_{h,0}^xx_{h,0}}=K_t,$$

for some variable K_t that is common across households. Taking logs and approximating

around the efficient allocation gives us

$$\begin{aligned}\partial \log x_{h,t} - \partial \log x_{h,0} &= -(d \log \mu_{h,t}^{IM} - d \log \mu_{h,0}^{IM}) - (\partial \log p_{h,t}^x - \partial \log p_{h,0}^x) + \partial \log K_t \\ \partial \log x_{h,t} - \partial \log x_{h,0} &= -(d \log \mu_{h,t}^{IM} - d \log \mu_{h,0}^{IM}) - (1 - \omega_C)(d \log \mu_{h,t}^L - d \log \mu_{h,0}^L) + \partial \log K_t \\ (\partial \log x_{h,t} - \partial \log x_{h,0}) - (\partial \log x_{0,t} - \partial \log x_{0,0}) &= -(d \log \mu_{h,t}^{IM} - d \log \mu_{h,0}^{IM} - (d \log \mu_{0,t}^{IM} - d \log \mu_{0,0}^{IM})) \\ &\quad - (1 - \omega_C)(d \log \mu_{h,t}^L - d \log \mu_{h,0}^L - (d \log \mu_{0,t}^L - d \log \mu_{0,0}^L)).\end{aligned}$$

By Corollary 1 in Baqaee et al. (2026) we know that

$$\mathbb{E}_0 \sum_{t=0}^{\infty} (1 - \beta) \beta^t \partial \log x_{h,t} = \mathbb{E}_0 \sum_{t=0}^{\infty} (1 - \beta) \beta^t \partial \log x_{0,t} = 0.$$

So, we can write

$$\begin{aligned}\partial \log x_{h,t} - \partial \log x_{0,t} &= -[(d \log \mu_{h,t}^{IM} - d \log \bar{\mu}_h^{IM}) - (d \log \mu_{0,t}^{IM} - d \log \bar{\mu}_0^{IM})] \\ &\quad - (1 - \omega_C)[(d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L) - (d \log \mu_{0,t}^L - d \log \bar{\mu}_0^L)],\end{aligned}$$

where $d \log \bar{\mu}_h = (1 - \beta) \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t d \log \mu_{h,t}$ is the average lifetime wedge for household h .

Now, note that since $\log c_{h,t} = \log(p_{h,t}^x x_{h,t}) - \log P_t$ and $\log l_{h,t} = \log(p_{h,t}^l x_{h,t}) - \log \mu_{h,t}^L - \log \tilde{W}_t$, then

$$\omega_C \partial \log c_{h,t} + (1 - \omega_C) \partial \log l_{h,t} = \partial \log x_{h,t},$$

and, since by market clearing

$$E_{\chi_h} [\omega_C \partial \log c_{h,t} + (1 - \omega_C) \partial \log l_{h,t}] = 0 \Leftrightarrow E_{\chi_h} [\partial \log x_{h,t}] = 0.$$

Using these findings in the expression above, we can rewrite

$$\begin{aligned}\partial \log x_{h,t} &= -[(d \log \mu_{h,t}^{IM} - d \log \bar{\mu}_h^{IM}) - E_{\chi_{h'}} [(d \log \mu_{h',t}^{IM} - d \log \bar{\mu}_{h'}^{IM})]] \\ &\quad - (1 - \omega_C)[(d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L) - E_{\chi_{h'}} [(d \log \mu_{h',t}^L - d \log \bar{\mu}_{h'}^L)]],\end{aligned}$$

and note that $E_{\chi_h} [d \log \mu_{h,t}^{IM}] = -E_{\chi_h} [\partial \log c_{h,t} - \partial \log C_t - (\partial \log c_{h,0} - \partial \log C_0)] = 0$ and similarly $E_{\chi_h} [d \log \bar{\mu}_{h,t}^{IM}] = 0$.

Replacing $\partial \log x_{h,t}$ in the definition of the Harberger triangle for incomplete markets

gives

$$\begin{aligned}
\triangleright^{IM} &= -\frac{1-\beta}{2}\mathbb{E}_0 \sum_t \beta^t E_{\chi_h} \left[d \log \mu_{h,t}^{IM} [(d \log \mu_{h,t}^{IM} - d \log \bar{\mu}_h^{IM})] \right] \\
&\quad - \frac{1-\beta}{2}(1-\omega_C)\mathbb{E}_0 \sum_t \beta^t E_{\chi_h} \left[d \log \mu_{h,t}^{IM} [(d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L) - E_{\chi_{h'}}[(d \log \mu_{h',t}^L - d \log \bar{\mu}_{h'}^L)]] \right] \\
&\Leftrightarrow \triangleright^{IM} = -\frac{1-\beta}{2}E_{\chi_h} \left[\mathbb{E}_0 \sum_t \beta^t d \log \mu_{h,t}^{IM} [(d \log \mu_{h,t}^{IM} - d \log \bar{\mu}_h^{IM})] \right] \\
&\quad - \frac{1-\beta}{2}(1-\omega_C)\mathbb{E}_0 \sum_t \beta^t Cov_{\chi_h} \left[d \log \mu_{h,t}^{IM}, (d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L) \right] \\
&\Leftrightarrow \triangleright^{IM} = -\frac{1}{2}E_{\chi_h} \left[Var_{\beta,\pi} \left(d \log \mu_{h,t}^{IM} \right) \right] \\
&\quad - \frac{1-\beta}{2}(1-\omega_C)\mathbb{E}_0 \sum_t \beta^t Cov_{\chi_h} \left[d \log \mu_{h,t}^{IM}, (d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L) \right].
\end{aligned}$$

Now, using the fact that

$$\begin{aligned}
d \log \mu_{h,t}^{IM} &= - \left[d \log \left(\frac{c_{h,t}}{C_t} \right) - d \log \left(\frac{c_{h,0}}{C_0} \right) \right] \\
d \log \mu_{h,t}^L - d \log \bar{\mu}_h^L &= \gamma_h \frac{\omega_C}{1-\omega_C} \left(\partial \log Y_t - (1-\beta)\mathbb{E}_0 \sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right),
\end{aligned}$$

and $C_t = Y_t$, we can write

$$\begin{aligned}
\triangleright^{IM} &= -\frac{1}{2}E_{\chi_h} \left[Var_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] \\
&\quad + \frac{1-\beta}{2}\omega_C \mathbb{E}_0 \sum_t \beta^t Cov_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{c_{h,0}} \right), \gamma_h \right] \left(\partial \log Y_t - (1-\beta)\mathbb{E}_0 \sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right) \\
&\Leftrightarrow \triangleright^{IM} = -\frac{1}{2}E_{\chi_h} \left[Var_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \frac{1-\beta}{2}\omega_C \left[\mathbb{E}_0 \sum_t \beta^t Cov_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{c_{h,0}} \right), \gamma_h \right] \partial \log Y_t \right. \\
&\quad \left. - Cov_{\chi_h} \left[d \log \left(\frac{\bar{c}_h}{c_{h,0}} \right), \gamma_h \right] \mathbb{E}_0 \sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right]
\end{aligned}$$

where $d \log (\bar{c}_h) \equiv (1-\beta)\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t d \log c_{h,t}$ is the average lifetime consumption of house-

hold h . So,

$$\triangleright^{IM} = -\frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \frac{1-\beta}{2}\omega_C \mathbb{E}_0 \sum_t \beta^t \text{Cov}_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right), \gamma_h \right] \partial \log Y_t$$

A.2.6 Proof of Proposition 3

We have found that

$$\begin{aligned} \triangleright^p &= \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\varepsilon}{\lambda_p} \pi_t^2 + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right] \\ \triangleright^w &= \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\psi}{\lambda_w} \pi_{w,t}^2 + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right], \\ \triangleright^L &= -\frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_t \beta^t \left\{ \left[(1+\nu^{-1}) \partial \log Y_t + E_i[d \log \mu_{i,t}^p] + E_u[d \log \mu_{u,t}^w] \right] \partial \log Y_t \right. \\ &\quad \left. + \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \cdot (\partial \log Y_t)^2 + \nu^{-1}(1-\beta) \text{Var}_{\chi_h}(\gamma_h) \mathbb{E}_0 \left(\sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right)^2 \right\}, \\ \triangleright^{IM} &= -\frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \frac{1-\beta}{2}\omega_C \mathbb{E}_0 \sum_t \beta^t \text{Cov}_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right), \gamma_h \right] \partial \log Y_t. \end{aligned}$$

Putting these together, we get

$$\begin{aligned} \triangleright^p + \triangleright^w + \triangleright^L + \triangleright^{IM} &= \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\varepsilon}{\lambda_p} \pi_t^2 + E_i[d \log \mu_{i,t}^p] \partial \log Y_t \right] \\ &\quad + \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[-\frac{\psi}{\lambda_w} \pi_{w,t}^2 + E_u[d \log \mu_{u,t}^w] \partial \log Y_t \right] \\ &\quad - \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_t \beta^t \left\{ \left[(1+\nu^{-1}) \partial \log Y_t + E_i[d \log \mu_{i,t}^p] + E_u[d \log \mu_{u,t}^w] \right] \partial \log Y_t \right. \\ &\quad \left. + \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \cdot (\partial \log Y_t)^2 + \nu^{-1}(1-\beta) \text{Var}_{\chi_h}(\gamma_h) \mathbb{E}_0 \left(\sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right)^2 \right\} \\ &\quad - \frac{1}{2}E_{\chi_h} \left[\text{Var}_{\beta,\pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] \\ &\quad + \frac{1-\beta}{2}\omega_C \mathbb{E}_0 \sum_t \beta^t \text{Cov}_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right), \gamma_h \right] \partial \log Y_t \end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \log A(\mathcal{P}) = & \log \mathcal{D} - \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1+\nu^{-1})(\partial \log Y_t)^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right] \\
& - \frac{(1-\beta)\omega_C}{2} \mathbb{E}_0 \sum_t \beta^t \left\{ \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \cdot (\partial \log Y_t)^2 \right. \\
& \left. + \nu^{-1}(1-\beta) \text{Var}_{\chi_h}(\gamma_h) \mathbb{E}_0 \left(\sum_{s=0}^{\infty} \beta^s \partial \log Y_s \right)^2 \right\} \\
& - \frac{1}{2} E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \frac{1-\beta}{2} \omega_C \mathbb{E}_0 \sum_t \beta^t \text{Cov}_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right), \gamma_h \right] \partial \log Y_t
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \log A(\mathcal{P}) = & \log \mathcal{D} - \frac{1}{2} \left\{ \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \omega_C (1-\beta) \beta^t \left[(1+\nu^{-1})(\partial \log Y_t)^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right] \right] \right. \\
& + \mathbb{E}_0 \left[\sum_t \omega_C (1-\beta) \beta^t \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \cdot (\partial \log Y_t)^2 \right] \\
& + \omega_C \nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \mathbb{E}_0 \left(\sum_{s=0}^{\infty} (1-\beta) \beta^s \partial \log Y_s \right)^2 \\
& \left. + E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \left(\frac{c_{h,t}}{Y_t} \right) \right) \right] + \mathbb{E}_0 \sum_t \omega_C (1-\beta) \beta^t \text{Cov}_{\chi_h} \left[d \log \left(\frac{c_{h,t}}{\bar{c}_h} \right), \gamma_h \right] \partial \log Y_t \right\}
\end{aligned}$$

Finally, replacing the change in output induced by the wedges

$$\partial \log Y_t = d \log(Z_t L_t) - d \log \left(\frac{Z_t}{1+\chi} \right) = d \log Y_t - d \log Y_t^* = \hat{y}_t - \hat{y}_t^*,$$

we obtain the desired result.

A.3 Appendix to Section 5

A.3.1 Proof of Theorem 1

See Baqaee et al. (2026).

The approximate optimal policy can be computed from choosing $\{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$ to maximize the second order approximation of aggregate productivity, subject to the constraints

$$\begin{aligned}\pi_t &= \lambda_p (\hat{w}_t - \hat{z}_t) + \beta \mathbb{E}_t[\pi_{t+1}] + u_{p,t} \\ \pi_{w,t} &= \lambda_w \left(\left\{ 1 + \nu^{-1} \right\} (\hat{y}_t - \hat{y}_t^*) - (\hat{w}_t - \hat{z}_t) \right) + \beta \mathbb{E}_t[\pi_{w,t+1}] + u_{w,t} \\ \hat{w}_t &= \hat{w}_{t-1} + \pi_{w,t} - \pi_t.\end{aligned}$$

A.3.2 Proof of Proposition 4

Note that δ_t does not show up in the reduced form program for optimal monetary policy. It follows that the quantities $\{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$ that maximize the second order approximation of aggregate productivity are the same as those in the case without demand shocks. In particular, the optimal policy does not respond to demand shocks. So, the central bank chooses i_t to fully offset the demand shock, i.e., $d \log(1 + i_t) = -d \log(1 + \delta_t)$.

A.3.3 Proof of Proposition 5

Let $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$. In the sticky-price model, $\gamma_h = 0$ for every h , wages are flexible, and there are no wage-push shocks. Hence the second-order objective in Proposition 3 reduces to

$$\begin{aligned}\log A(\mathcal{P}) &= \log \mathcal{D} - \frac{(1 - \beta)\omega_C}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1 + \nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 \right] \\ &\quad - \frac{1}{2} E_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(d \log \frac{c_{h,t}}{Y_t} \right) \right],\end{aligned}\tag{50}$$

up to terms that are independent of policy. Thus the only point to prove is that the last term has zero derivative with respect to aggregate conditions.

Consider first the household block with idiosyncratic risk and initial asset heterogeneity shut down, $\sigma_z = \sigma_a = 0$, but with arbitrary aggregate shocks and arbitrary monetary policy. Since $\gamma_h = 0$, aggregate fluctuations do not change household h 's share of aggregate labor and profit income to a first order. In particular, household income can be

written as

$$y_{h,t} = z_{h,0} P_t Y_t,$$

where $z_{h,0}$ is time invariant and independent of the aggregate state to a first order. The financial constraint is slack at the efficient allocation because $\Omega(0, \{0\}_k) > 0$. Therefore, for a sufficiently small perturbation around that allocation, the household's Euler equations with respect to assets that span aggregate risk hold with equality.

Let Q_t denote the date-0 Arrow price of one unit of nominal wealth in the aggregate history at date t . The first-order condition for any aggregate Arrow claim is

$$\frac{(1 - \beta)\beta^t}{1 + \chi} \frac{1}{P_t c_{h,t}} = \Lambda_h Q_t, \quad (51)$$

where Λ_h is the household-specific multiplier on lifetime wealth. This first order condition implies that

$$c_{h,t} = \alpha_h Y_t,$$

for some $\alpha_h \geq 0$ with $\int \alpha_h dh = 0$. It follows that aggregate shocks and monetary policy can affect the common factor C_t , but they do not affect the time profile of household consumption relative to aggregate consumption. Equivalently, for any aggregate perturbation, the incomplete market wedges are zero around the complete markets allocation.

A.3.4 Proof of Proposition 6

Follows from the same arguments as in the proof of Proposition 5, but with the wage Phillips curve replaced by the one in the canonical HANK model.

A.3.5 Proof of Proposition 7

In the sticky-wage model, prices are flexible and there are no price cost-push shocks. Hence, the price Phillips curve implies $\hat{w}_t = \hat{z}_t$, and the wage Phillips curve reduces to

$$\pi_{w,t} = \lambda_w (1 + \nu^{-1}) \tilde{y}_t + \beta \mathbb{E}_t[\pi_{w,t+1}] + u_{w,t}, \quad \tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*. \quad (52)$$

Terms in the objective that are independent of policy can be dropped. After multiplying by the positive constant $(1 - \beta)\omega_C$, the relevant maximization problem is

$$\max_{\{\tilde{y}_t, \pi_{w,t}\}} -\frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{\psi}{\lambda_w} \pi_{w,t}^2 + \left(1 + \nu^{-1} + \nu^{-1}V_\gamma\right) \tilde{y}_t^2 - C_t \tilde{y}_t \right] - \frac{1}{2} \nu^{-1} (1 - \beta) V_\gamma \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \tilde{y}_t \right)^2,$$

where

$$V_\gamma \equiv \text{Var}_{\chi_h}(\gamma_h), \quad C_t \equiv \text{Cov}_{\chi_h} \left(d \log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right).$$

Since \bar{c}_h is time invariant for a given household,

$$C_t - C_{t-1} = \text{Cov}_{\chi_h}(d \log c_{h,t}, \gamma_h) - \text{Cov}_{\chi_h}(d \log c_{h,t-1}, \gamma_h).$$

Let ξ_t be the multiplier on the wage Phillips curve. The Lagrangian first-order conditions are

$$\left(1 + \nu^{-1} + \nu^{-1}V_\gamma\right) \tilde{y}_t + \nu^{-1}(1 - \beta)V_\gamma \mathbb{E}_0 \sum_{s=0}^{\infty} \beta^s \tilde{y}_s = \frac{1}{2}C_t - \lambda_w(1 + \nu^{-1})\xi_t, \quad (53)$$

$$\frac{\psi}{\lambda_w} \pi_{w,t} - \xi_t + \xi_{t-1} = 0, \quad (54)$$

with $\xi_{-1} = 0$. The present-value term in (53) is common across dates, so it cancels when subtracting the condition at $t - 1$ from the condition at t . Using (54), for $t \geq 1$ we obtain

$$\psi(1 + \nu^{-1})\pi_{w,t} = \frac{1}{2}(C_t - C_{t-1}) - \left(1 + \nu^{-1} + \nu^{-1}V_\gamma\right) (\tilde{y}_t - \tilde{y}_{t-1}). \quad (55)$$

Rearranging,

$$\pi_{w,t} = -\frac{1 + \nu^{-1} + \nu^{-1}V_\gamma}{\psi(1 + \nu^{-1})} (\tilde{y}_t - \tilde{y}_{t-1}) + \frac{1}{2\psi(1 + \nu^{-1})} (C_t - C_{t-1}). \quad (56)$$

Now define

$$\hat{\psi} \equiv \frac{\psi}{1 + \frac{\nu^{-1}}{1 + \nu^{-1}} V_\gamma}.$$

Then

$$\hat{\psi}^{-1} = \frac{1 + \nu^{-1} + \nu^{-1}V_\gamma}{\psi(1 + \nu^{-1})}.$$

Moreover,

$$\frac{\hat{\psi}^{-1} - \psi^{-1}}{2\nu^{-1}} = \frac{V_\gamma}{2\psi(1 + \nu^{-1})} \equiv \zeta_{IM}.$$

Therefore

$$\frac{1}{2\psi(1 + \nu^{-1})}(C_t - C_{t-1}) = \zeta_{IM} \cdot \frac{Cov_{\chi_h}(d \log c_{h,t}, \gamma_h) - Cov_{\chi_h}(d \log c_{h,t-1}, \gamma_h)}{V_\gamma},$$

which gives (35). For $t = 0$, use $\xi_{-1} = 0$ and the normalization $\tilde{y}_{-1} = 0$ and $C_{-1} = 0$. The same algebra gives

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

A.3.6 Proof of Corollary 1

The targeting rule in Proposition 7 can be written as

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

Along a bounded equilibrium path converging back to the stationary allocation, the cross-sectional distribution induced by a date-0 perturbation becomes independent of the initial household state. Hence $Cov_{\chi_h}(d \log c_{h,t}, \gamma_h)$ converges to a constant, equal to zero under the normalization of log deviations around the efficient steady state. It follows that

$$Cov_{\chi_h}(d \log c_{h,t}, \gamma_h) - Cov_{\chi_h}(d \log c_{h,t-1}, \gamma_h) \rightarrow 0.$$

Taking limits in the targeting rule yields

$$\pi_{w,t} = -\hat{\psi}^{-1}(\tilde{y}_t - \tilde{y}_{t-1}),$$

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

A.3.7 Proof of Proposition 8

Let $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$, $C_t \equiv Cov_{\chi_h}(d \log \frac{c_{h,t}}{c_h}, \gamma_h)$, and $V_\gamma \equiv Var_{\chi_h}(\gamma_h)$. Use the Dual-Mandate weights in Proposition 8,

$$\omega_y = 1 + \nu^{-1}(1 + V_\gamma), \quad \omega_p = \frac{\varepsilon}{\lambda_p}, \quad \omega_w = \frac{\psi}{\lambda_w}.$$

Dropping terms independent of policy and multiplying by positive constants, the policy-relevant part of the Efficiency Mandate objective in the sticky-price and sticky-wage model can be written as

$$-\frac{1}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\omega_y \tilde{y}_t^2 + \omega_p \pi_t^2 + \omega_w \pi_{w,t}^2 - C_t \tilde{y}_t \right] - \frac{1}{2} \nu^{-1} (1 - \beta) V_\gamma \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \tilde{y}_t \right)^2. \quad (57)$$

The Dual-Mandate problem in Proposition 8 has the same linearized private-sector constraints,

$$\begin{aligned} \pi_t &= \lambda_p (\hat{w}_t - \hat{z}_t) + \beta \mathbb{E}_t [\pi_{t+1}] + u_{p,t}, \\ \pi_{w,t} &= \lambda_w \left((1 + \nu^{-1}) \tilde{y}_t - (\hat{w}_t - \hat{z}_t) \right) + \beta \mathbb{E}_t [\pi_{w,t+1}] + u_{w,t}, \\ \hat{w}_t &= \hat{w}_{t-1} + \pi_{w,t} - \pi_t, \end{aligned}$$

and the same quadratic weights ω_y , ω_p , and ω_w on \tilde{y}_t , π_t , and $\pi_{w,t}$, but it omits the deterministic linear term $C_t \tilde{y}_t$ and the intertemporal incidence term.

Let $\xi_{w,t}$ denote the multiplier on the wage Phillips curve. The first-order condition of the Efficiency Mandate problem with respect to \tilde{y}_t is

$$\omega_y \tilde{y}_t + \nu^{-1} (1 - \beta) V_\gamma \mathbb{E}_0 \sum_{s=0}^{\infty} \beta^s \tilde{y}_s - \frac{1}{2} C_t = -\lambda_w (1 + \nu^{-1}) \xi_{w,t}. \quad (58)$$

The present-value term is constant across dates. Therefore, after differencing the date- t and date- $(t-1)$ conditions, it drops out. For $t \geq 1$,

$$\lambda_w (1 + \nu^{-1}) (\xi_{w,t} - \xi_{w,t-1}) = -\omega_y (\tilde{y}_t - \tilde{y}_{t-1}) + \frac{1}{2} (C_t - C_{t-1}). \quad (59)$$

In the Dual-Mandate problem, the same differenced condition holds with $C_t - C_{t-1} = 0$. All other first-order conditions and all private-sector constraints have the same time-invariant coefficients in the two problems. Thus the only non-common object in the system is an additive deterministic forcing term proportional to $C_t - C_{t-1}$, together with a possible date-0 difference in initial multipliers induced by the present-value term.

We now use the fact that the incomplete-market covariance is a deterministic transition term. Since \bar{c}_h is the discounted lifetime average of household h 's consumption response, the discounted average of C_t is zero. Moreover, along the unique bounded solution, the cross-sectional distribution induced by a date-0 perturbation converges back to the stationary distribution, so C_t converges to a finite limit and therefore $C_t - C_{t-1} \rightarrow 0$.

Because the problem is linear-quadratic with linear constraints, the full system of constraints and first-order conditions can be written as a linear rational-expectations system in a vector containing allocations and commitment multipliers. Under the maintained uniqueness and boundedness condition in the proposition, this system has a stable representation

$$X_{t+1} = AX_t + B\varepsilon_{t+1} + b_t, \quad (60)$$

for the Efficiency Mandate problem, where ε_{t+1} collects the one-step-ahead innovations in the exogenous shocks. The deterministic vector b_t is proportional to $C_t - C_{t-1}$ and hence satisfies $b_t \rightarrow 0$. The corresponding Dual-Mandate problem has the same stable matrices A and B , but no deterministic forcing term:

$$X_{t+1}^{DM} = AX_t^{DM} + B\varepsilon_{t+1}. \quad (61)$$

Any difference in the date-0 multipliers is absorbed in the initial condition.

Define $\Delta_t \equiv X_t - X_t^{DM}$. Subtracting (61) from (60) gives

$$\Delta_{t+1} = A\Delta_t + b_t.$$

The stochastic innovations enter the two systems through the same matrix B , so they cancel exactly; Δ_t is deterministic. Since the solution is stable, the spectral radius of A is strictly less than one. Let $\|\cdot\|$ denote any fixed vector norm and its induced matrix norm. Hence there are constants $M < \infty$ and $\alpha \in (0, 1)$ such that $\|A^k\| \leq M\alpha^k$ for all $k \geq 0$. Iterating,

$$\Delta_t = A^t\Delta_0 + \sum_{j=0}^{t-1} A^{t-1-j}b_j.$$

The first term converges to zero. For the second, fix any $\epsilon > 0$ and choose T such that $\|b_j\| \leq \epsilon$ for all $j \geq T$. Then

$$\begin{aligned} \left\| \sum_{j=0}^{t-1} A^{t-1-j}b_j \right\| &\leq \sum_{j=0}^{t-1} \left\| A^{t-1-j}b_j \right\| \\ &\leq \sum_{j=0}^{t-1} \left\| A^{t-1-j} \right\| \|b_j\| \\ &\leq \sum_{j=0}^{T-1} \left\| A^{t-1-j} \right\| \|b_j\| + \epsilon \sum_{k=0}^{\infty} \|A^k\|, \end{aligned}$$

where the first step uses the triangle inequality and the second step uses compatibility of

the matrix norm with the vector norm. The finite sum converges to zero as $t \rightarrow \infty$, and the second term is bounded by $\epsilon M / (1 - \alpha)$. Since ϵ is arbitrary, the whole convolution converges to zero. Thus $\Delta_t \rightarrow 0$.

The allocation vector $S_t = \{\hat{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t\}$ is a linear projection of the state vector X_t and contemporaneous shocks. Since the shock component is common across the two problems and Δ_t is deterministic with $\Delta_t \rightarrow 0$, there exists a deterministic sequence D_t such that

$$S_t = S_t^{DM} + D_t, \quad D_t \rightarrow 0.$$

This proves the proposition.

A.3.8 Proof of Proposition 9

In the sticky-price model, wages are flexible, $u_{w,t} = 0$, and $\gamma_h = 0$ for all h . By Proposition 5, the incomplete-market term is locally orthogonal to monetary policy and the reduced-form policy problem coincides with the sticky-price RANK problem. Thus it is enough to solve the RANK simple-rule problem.

Let $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$ denote the output gap and let

$$\kappa_p \equiv \lambda_p(1 + \nu^{-1}).$$

In the sticky-price model the price Phillips curve reduces to

$$\pi_t = \kappa_p \tilde{y}_t + \beta \mathbb{E}_t[\pi_{t+1}] + u_{p,t}.$$

The relevant loss is

$$\frac{1}{2} \mathbb{E}_{-1} \sum_{t=0}^{\infty} \beta^t \left[(1 + \nu^{-1}) \tilde{y}_t^2 + \frac{\epsilon}{\lambda_p} \pi_t^2 \right].$$

Under the proposed rules,

$$\tilde{y}_t = \phi_y + \phi_{y,\delta} \delta_t + (\phi_{y,z} - 1) \hat{y}_t^* + \phi_{y,u} u_{p,t}, \quad \pi_t = \phi_\pi + \phi_{\pi,\delta} \delta_t + \phi_{\pi,z} \hat{y}_t^* + \phi_{\pi,u} u_{p,t}.$$

The shocks are mean zero and mutually orthogonal, so the problem separates across the constant, demand-shock, natural-output, and cost-push components. Substituting the rules into the Phillips curve and using $\mathbb{E}_t[\vartheta_{t+1}] = \rho_\vartheta \vartheta_t$ for $\vartheta \in \{\delta, \hat{y}^*, u_p\}$ gives the coeffi-

cient restrictions

$$\begin{aligned}
(1 - \beta)\phi_\pi &= \kappa_p\phi_y, \\
(1 - \beta\rho_\delta)\phi_{\pi,\delta} &= \kappa_p\phi_{y,\delta}, \\
(1 - \beta\rho_z)\phi_{\pi,z} &= \kappa_p(\phi_{y,z} - 1), \\
(1 - \beta\rho_u)\phi_{\pi,u} &= \kappa_p\phi_{y,u} + 1.
\end{aligned}$$

For the first three components, the feasible choice that sets both inflation and the output gap to zero is available and uniquely minimizes the quadratic loss. Hence

$$\phi_y = \phi_\pi = 0, \quad \phi_{y,\delta} = \phi_{\pi,\delta} = 0, \quad \phi_{y,z} = 1, \quad \phi_{\pi,z} = 0.$$

For the cost-push component, write $a_u \equiv 1 - \beta\rho_u$. The Phillips-curve restriction gives

$$\phi_{y,u} = -\frac{1 - a_u\phi_{\pi,u}}{\kappa_p}.$$

The cost-push part of the loss is proportional to

$$\frac{\varepsilon}{\lambda_p}\phi_{\pi,u}^2 + (1 + \nu^{-1})\left(\frac{1 - a_u\phi_{\pi,u}}{\kappa_p}\right)^2.$$

The first-order condition is

$$\kappa_p\varepsilon\phi_{\pi,u} = a_u(1 - a_u\phi_{\pi,u}),$$

so

$$\phi_{\pi,u} = \frac{a_u}{\kappa_p\varepsilon + a_u^2}, \quad \phi_{y,u} = -\frac{\varepsilon}{\kappa_p\varepsilon + a_u^2}.$$

Substituting back $a_u = 1 - \beta\rho_u$ gives the expressions in the proposition.

A.3.9 Proof of Proposition 10

Let $\tilde{y}_t \equiv \hat{y}_t - \hat{y}_t^*$ and

$$V_\gamma \equiv \text{Var}_{\chi_h}(\gamma_h), \quad \kappa_w \equiv \lambda_w(1 + \nu^{-1}).$$

In the sticky-wage model the wage Phillips curve is

$$\pi_{w,t} = \kappa_w\tilde{y}_t + \beta\mathbb{E}_t[\pi_{w,t+1}] + u_{w,t}.$$

The policy-dependent part of the second-order loss is proportional to

$$\mathbb{E}_{-1} \sum_{t=0}^{\infty} \beta^t \left[\frac{\psi}{\lambda_w} \pi_{w,t}^2 + \left(1 + \nu^{-1} + \nu^{-1} V_\gamma\right) \tilde{y}_t^2 \right] + (1 - \beta) V_\gamma \mathbb{E}_{-1} \left(\sum_{t=0}^{\infty} \beta^t \tilde{y}_t \right)^2.$$

The linear incomplete-market term does not affect the optimal time-invariant simple rules. Indeed,

$$(1 - \beta) \sum_{t=0}^{\infty} \beta^t \text{Cov}_{\chi_h} \left(d \log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) = 0$$

by the definition of \bar{c}_h , and the aggregate shocks entering the simple rules have zero unconditional means.

Under the proposed rules,

$$\tilde{y}_t = \phi_y + \phi_{y,\delta} \delta_t + (\phi_{y,z} - 1) \hat{y}_t^* + \phi_{y,u} u_{w,t}, \quad \pi_{w,t} = \phi_\pi + \phi_{\pi,\delta} \delta_t + \phi_{\pi,z} \hat{y}_t^* + \phi_{\pi,u} u_{w,t}.$$

Substituting these rules into the wage Phillips curve gives

$$\begin{aligned} (1 - \beta) \phi_\pi &= \kappa_w \phi_y, \\ (1 - \beta \rho_\delta) \phi_{\pi,\delta} &= \kappa_w \phi_{y,\delta}, \\ (1 - \beta \rho_z) \phi_{\pi,z} &= \kappa_w (\phi_{y,z} - 1), \\ (1 - \beta \rho_u) \phi_{\pi,u} &= \kappa_w \phi_{y,u} + 1. \end{aligned}$$

The constant, demand-shock, and natural-output components are minimized by setting both wage inflation and the output gap to zero. Therefore

$$\phi_y = \phi_\pi = 0, \quad \phi_{y,\delta} = \phi_{\pi,\delta} = 0, \quad \phi_{y,z} = 1, \quad \phi_{\pi,z} = 0.$$

It remains to solve for the cost-push coefficients. Let $a_u \equiv 1 - \beta \rho_u$. The wage Phillips curve implies

$$\phi_{y,u} = -\frac{1 - a_u \phi_{\pi,u}}{\kappa_w}.$$

For the AR(1) cost-push experiment used in the simple-rule problem,

$$\frac{\mathbb{E}_{-1} \left(\sum_{t=0}^{\infty} \beta^t u_{w,t} \right)^2}{\mathbb{E}_{-1} \sum_{t=0}^{\infty} \beta^t u_{w,t}^2} = \frac{1 - \beta}{(1 - \beta \rho_u)^2 (1 + \beta)},$$

which follows from the fact that since:

$$\sum_{t=0}^{\infty} \beta^t u_{w,t} = \sum_{t=0}^{\infty} \beta^t \sum_{s=0}^t \rho^{t-s} \varepsilon_s = \sum_{s=0}^{\infty} \varepsilon_s \sum_{t=s}^{\infty} \beta^t \rho^{t-s} = \frac{1}{1-\beta\rho} \sum_{s=0}^{\infty} \beta^s \varepsilon_s,$$

then

$$\begin{aligned} \mathbb{E} \left(\sum_{t=0}^{\infty} \beta^t u_{w,t} \right)^2 &= \frac{\sigma_u^2}{(1-\beta\rho)^2} \sum_{s=0}^{\infty} \beta^{2s} = \frac{\sigma_u^2}{(1-\beta\rho)^2(1-\beta^2)} \\ \mathbb{E} \sum_{t=0}^{\infty} \beta^t u_{w,t}^2 &= \sigma_u^2 \sum_{t=0}^{\infty} \beta^t \sum_{j=0}^t \rho^{2j} = \sigma_u^2 \sum_{t=0}^{\infty} \beta^t \frac{1-\rho^{2(t+1)}}{1-\rho^2} = \frac{\sigma_u^2}{(1-\beta)(1-\beta\rho^2)}. \end{aligned}$$

Hence the cost-push component of the loss is proportional to

$$\frac{\psi}{\lambda_w} \phi_{\pi,\mu}^2 + \left[1 + \nu^{-1} + V_\gamma \left(\nu^{-1} + \frac{(1-\beta)}{(1-\beta\rho_u)^2(1+\beta)} \right) \right] \phi_{y,\mu}^2.$$

Equivalently, using $\kappa_w = \lambda_w(1 + \nu^{-1})$, this is

$$\frac{\psi}{\lambda_w} \phi_{\pi,\mu}^2 + (1 + \nu^{-1}) \kappa_\gamma \phi_{y,\mu}^2, \quad \kappa_\gamma \equiv 1 + \frac{V_\gamma}{1 + \nu^{-1}} \left[\nu^{-1} + \frac{(1-\beta)}{(1-\beta\rho_u)^2(1+\beta)} \right].$$

Thus the sticky-wage HANK simple-rule problem is isomorphic to the RANK problem with the wage-inflation coefficient ψ replaced by

$$\tilde{\psi} \equiv \frac{\psi}{\kappa_\gamma}.$$

The first-order condition for $\phi_{\pi,\mu}$ is

$$\kappa_w \tilde{\psi} \phi_{\pi,\mu} = a_u (1 - a_u \phi_{\pi,\mu}),$$

so

$$\phi_{\pi,\mu} = \frac{a_u}{\kappa_w \tilde{\psi} + a_u^2}, \quad \phi_{y,\mu} = -\frac{\tilde{\psi}}{\kappa_w \tilde{\psi} + a_u^2}.$$

Since $V_\gamma > 0$ implies $\kappa_\gamma > 1$, we have $\tilde{\psi} < \psi$. Therefore $\phi_{y,\mu}$ is less negative than the corresponding RANK coefficient and $\phi_{\pi,\mu}$ is larger than the corresponding RANK coefficient. Substituting $a_u = 1 - \beta\rho_u$ gives the proposition.

A.4 Appendix to Quantitative Section

A.4.1 Calibrating the Incidence Function

In the economy with sticky-wages only that we use for the quantitative analysis, profits are zero and so

$$\frac{y_{h,t}}{P_t} = z_{h,t} \Gamma(1 - L_t; \zeta_{h,t}) Y_t,$$

so, to first order in $\sigma = 0$, we find that

$$\begin{aligned} \log \frac{y_{h,t}}{P_t} &= \log z_{h,t} + \gamma_h d \log(1 - L_t) + d \log Y_t \\ &= \log z_{h,t} + (1 + \gamma_h) d \log(Y_t) - \gamma_h d \log(Z_t), \end{aligned}$$

where the second equality follows from the fact that $d \log Y_t = d \log Z_t + d \log(1 - L_t)$. So, income growth is given by

$$\Delta \log \frac{y_{h,t}}{P_t} = (1 + \gamma_h) \Delta d \log(Y_t) - \gamma_h \Delta d \log(Z_t).$$

To map the elasticity to the evidence in Guvenen et al. (2017), we assume that γ_h is only a function of the household's current productivity, i.e., with a slight abuse of notation, $\gamma_h = \gamma_g$ where g denotes the productivity group. This procedure involves choosing γ_g to match the cross-sectional distribution of the worker beta regressions

$$\Delta \log \frac{y_{h,t}}{P_t} = b_{0,g} + b_{1,g} \Delta d \log(Y_t) + \varepsilon_{h,t},$$

where g is the productivity group to which household h belongs.

$$\begin{aligned} \text{plim } \hat{b}_{1,g} &= 1 + \gamma_g - \gamma_g \frac{\text{Cov}(\Delta \log Z_t, \Delta \log Y_t)}{\text{Var}(\Delta \log Z_t)} \frac{\text{Var}(\Delta \log Z_t)}{\text{Var}(\Delta \log Y_t)} \\ &= 1 + \gamma_g - \gamma_g \frac{\text{Cov}(\Delta \log Z_t, \Delta \log Y_t)}{\text{SD}(\Delta \log Z_t) \text{SD}(\Delta \log Y_t)} \frac{\text{SD}(\Delta \log Z_t)}{\text{SD}(\Delta \log Y_t)} \end{aligned}$$

We use the evidence in King and Rebelo (1999) to calibrate the covariance and variance terms:

$$\frac{\text{SD}(\Delta \log Z_t)}{\text{SD}(\Delta \log Y_t)} = 0.54, \quad \frac{\text{Cov}(\Delta \log Z_t, \Delta \log Y_t)}{\text{SD}(\Delta \log Z_t) \text{SD}(\Delta \log Y_t)} = 0.78.$$

It follows that we can map

$$b_g = 1 + (1 - 0.42) \gamma_g \Leftrightarrow \gamma_g = \frac{b_g - 1}{1 - 0.42}.$$

Our parameterization uses the GDP betas from Guvenen et al. (2017) for males aged 36 to 45. We map these coefficients to b_g in our model by interpolating the betas across income groups. We normalize the coefficients such that $\sum \pi_g z_g b_g = 1$. Finally, we then apply the above mapping to obtain γ_g . The final result can be found in Figure 1.

A.5 Variations on the Definition of Aggregate Productivity

In this section, we discuss two modifications to the definition of aggregate productivity. What happens if we: (1) define productivity using a factor-biased technology shock, (2) alter the timing of the lump-sum compensations. We show that our characterization of the approximately optimal policy is invariant to these changes.

A.5.1 Biased Productivity Variation

Our baseline definition measures aggregate productivity by asking how much the entire time endowment can be contracted while keeping every household indifferent to the status quo allocation. One could instead measure the same efficiency gain in biased productivity units. For example, the productivity shifter could apply only to market production time, leaving the productivity of home-production/leisure unchanged. More generally, let $\tilde{A}(\mathcal{P})$ denote the largest contraction in the productivity of a subset of factor endowments, rather than a uniform contraction of all factor endowments, such that the compensated allocation remains feasible and every household is kept indifferent to the status quo allocation \mathbf{x}^0 .

This alternative changes the units in which aggregate productivity is reported, but not the local policy problem. The formal argument is given in Appendix B of Baqaee et al. (2026). Around the efficient allocation, the Hessian of the biased productivity measure with respect to wedges is proportional to the Hessian of the baseline TFP-equivalent measure. If the biased shifter applies to a factor, or set of factors, with income share $\lambda_{F^*} > 0$, then, to a second-order approximation,

$$\log \tilde{A}(\mathcal{P}) \approx \frac{1}{\lambda_{F^*}} \log A(\mathcal{P}).$$

Thus the biased measure rescales the deadweight-loss-triangle objective by a positive con-

stant.

It follows that all first-order policy prescriptions in the paper are unchanged. The quadratic approximation in Proposition 3 would be multiplied by a positive factor, and the distance-to-frontier term would be expressed in different productivity units. But the policy-relevant first-order conditions are identical. Hence the orthogonality results for the canonical HANK cases, the targeting criterion with cyclical income risk, and the long-run equivalence to a modified dual-mandate policy do not depend on whether productivity variation is measured in uniform time-endowment units or in biased factor-productivity units.

A.5.2 Timing of Transfers

In the baseline definition of aggregate productivity, one-off compensating transfers are paid out at date 0. This is a normalization for the local policy problem rather than a substantive restriction on the date at which compensation is delivered.¹² This subsection explains why.

Consider an alternative definition in which household h receives a non-state-contingent transfer schedule $\{T_{h,t}\}_{t \geq 0}$ rather than a single date-0 transfer. The transfer schedule may vary across households and across calendar dates, but it is fixed independently of the household's future idiosyncratic shocks and satisfies budget balance at each date,

$$\int T_{h,t} dh = 0.$$

Let Q_t denote the date-0 price of one unit of the final good delivered at date t , evaluated at the allocation around which we approximate. The present value of household h 's transfer schedule is

$$\mathcal{T}_h \equiv \sum_{t=0}^{\infty} Q_t T_{h,t}.$$

If portfolio constraints are slack, the household's intertemporal problem depends on the schedule $\{T_{h,t}\}$ only through this present value. Hence the schedule is locally equivalent to a date-0 transfer \mathcal{T}_h .

The required slackness holds at our point of approximation. We approximate around the efficient allocation with zero household asset positions. By assumption, $\Omega(0, \{0\}_k) > 0$, so the portfolio constraints are slack at that allocation. Since $\Omega(\cdot)$ is continuously differentiable, there is an open neighborhood of zero asset positions in which the constraints

¹²The substantive restriction is that compensating transfers are not contingent on households' future idiosyncratic histories. We explore this in more detail in Section A.5.3.

remain slack. Therefore, for any sufficiently small compensating transfer schedule, households can use the available financial assets to shift resources across dates without changing which portfolio constraints bind. Delaying or accelerating a non-state-contingent payment only changes the household's financing position inside this neighborhood.

This also shows that the timing of such transfers does not create incomplete-market wedges to first order. Holding fixed aggregate prices and quantities at the efficient allocation, a non-state-contingent transfer changes household h 's lifetime wealth and therefore its multiplier on the intertemporal budget constraint. Since the constraint set is locally interior, the first-order conditions imply

$$d \log c_{h,t} - d \log C_t = \eta_h \quad \text{for all } t,$$

for some household-specific constant η_h that depends on the present value \mathcal{T}_h , but not on the timing of the schedule. With log utility, the incomplete-market wedge in (22) satisfies

$$\begin{aligned} d \log \mu_{h,t}^{IM} &= - [d \log c_{h,t} - d \log C_t] + [d \log c_{h,0} - d \log C_0] \\ &= -\eta_h + \eta_h = 0. \end{aligned}$$

Thus a non-state-contingent compensating transfer can redistribute the present value of resources across households, but it does not by itself provide insurance across idiosyncratic states or alter intertemporal wedges at the margin. In particular, replacing date-0 transfers with any locally small non-state-contingent payout schedule leaves unchanged the first-order incomplete-market wedges that enter the second-order approximation in Proposition 3. The quadratic policy problem and all the first-order policy prescriptions are therefore the same.

State-contingent transfers are different. If transfers can depend on the household's realized idiosyncratic history, the policymaker is no longer only choosing the timing of compensation. It is expanding the set of insurance instruments available to households. With sufficiently rich person-time-state-contingent transfers, the policymaker can equate the ratio of marginal utilities across dates and states, thereby setting the eliminating incomplete-market wedges. This is the case studied in Appendix A.5.3. Hence our date-0 formulation should be read as a convenient representative of the class of non-state-contingent compensation schemes, not as an assumption that the calendar timing of compensation is economically important for the local analysis.

A.5.3 Optimal Monetary Policy with State-Contingent Transfers

In this appendix, we study the problem of the efficiency-maximizing central bank under the assumption that it can implement arbitrary transfers in each period. This setting, while unrealistic, is instructive and allows us to characterize the efficiency-maximizing policy in a simple way.

Without any restriction on transfers, the efficiency-maximizing allocations must be such that the incomplete market wedges are always equal to zero. In other words, the Euler equation is always satisfied with equality for all households and states, since person-time-state contingent transfers are used to substitute incomplete markets. So, in this case

$$c_{h,t} = \alpha_h C_t$$

for some $\alpha_h \geq 0$ such that $\int \alpha_h dh = 1$.

Suppose further that labor-income incidence function is such that there is no dispersion of labor wedges across households, i.e., $\log \mu_{h,t}^L = \log \mu_t^L$ for all h , and so

$$\frac{\chi u'(l_{h,t})}{u'(c_{h,t})} = \frac{\chi u'(L_t)}{u'(C_t)}, \quad \text{for all } h.$$

This would be the case, for instance, in a model without nominal wage rigidity in which households would choose labor supply optimally. In this case, it follows that

$$z_{h,t} l_{h,t} = \alpha_h L_t.$$

Using the fact that, in the optimum, $\tilde{U}_h(\mathbf{x}_h)$ is equalized across h , and that $\int \alpha_h dh = 1$, it follows that $\alpha_h = \frac{\exp\{U(\mathbf{x}_h^0)\}}{\int \exp\{U(\mathbf{x}_{h'}^0)\} dh'}$. The utility of the compensated agent is

$$\min_h \tilde{U}_h(\mathbf{x}_h) = \frac{\exp\{U(\{C_t, L_t\})\}}{\int \exp\{U(\mathbf{x}_{h'}^0)\} dh'}. \quad (62)$$

It follows that the equilibrium is as if the aggregates were determined by a representative agent with the original preferences $U(\{C_t, L_t\})$, i.e., the equilibrium aggregate quantities coincide with those in the RANK model under policy \mathcal{P} .

Proposition 11. *Consider the case with state-contingent transfers. Suppose furthermore that labor-income incidence is such that there is no dispersion of labor wedges across households. Then,*

the aggregate productivity of policy \mathcal{P} is given by

$$A(\mathcal{P}) = \frac{\exp\{U(\{C_t^{\text{RANK}}(\mathcal{P}), L_t^{\text{RANK}}(\mathcal{P})\})\}}{\int \exp\{U(x_{h'}^0)\} dh'}, \quad (63)$$

where $C_t^{\text{RANK}}(\mathcal{P})$ and $L_t^{\text{RANK}}(\mathcal{P})$ are the equilibrium aggregate consumption and leisure under policy \mathcal{P} in the RANK model. Hence, the efficiency-maximizing policy coincides with the optimal policy prescriptions in RANK, i.e.,

$$\mathcal{P}^* = \arg \max_{\mathcal{P}} A(\mathcal{P}) = \arg \max_{\mathcal{P}} U(\{C_t^{\text{RANK}}(\mathcal{P}), L_t^{\text{RANK}}(\mathcal{P})\}). \quad (64)$$

With access to state-contingent transfers, the efficient allocations are such that the incomplete market wedges are always equal to zero. If, on top of that, labor-income incidence is such that there is a common labor wedge across households, then the efficient allocations are such that aggregate allocations coincide with those in the RANK model. Furthermore, aggregate productivity is a monotone transformation of the representative agent's utility in the RANK model and the efficiency-mandate optimal policy coincides with the optimal policy prescriptions for the RANK model.

Intuitively, in the absence of incomplete market wedges and with a common labor wedge across households, the only sources of inefficiencies in this economy coincide with those in the RANK model, i.e., the price and wage markups. Our efficiency-mandate central bank chooses the policy that best addresses these sources of inefficiency, ignoring the underlying heterogeneity across households. So, the optimal policy exactly coincides with the optimal policy in the RANK model.