

Aggregate-Surplus Maximizing Monetary Policy

David Baqaee

UCLA

Ariel Burstein

UCLA

Joao Guerreiro

UCLA

October 2026

Introduction

How should heterogeneity affect monetary policy? Depends on **central bank's objective**.

How should heterogeneity affect monetary policy? Depends on **central bank's objective**.

Literature typically uses **Social Welfare Functions**:

- SWF requires taking a stance on distributional conflict across households, before we are able to discuss monetary policy. [Who should get what?]
- Different SWFs imply different positions on what constitutes good monetary policy.
- But how to discipline? And whose social preferences to use? Reasonable people may disagree.
- E.g., standard utilitarian: unconstrained optimum is to equalize consumption for all.
- But, this is a controversial objective with which to endow the central bank.

How should heterogeneity affect monetary policy? Depends on **central bank's objective**.

Literature typically uses **Social Welfare Functions**:

- SWF requires taking a stance on distributional conflict across households, before we are able to discuss monetary policy. [Who should get what?]
- Different SWFs imply different positions on what constitutes good monetary policy.
- But how to discipline? And whose social preferences to use? Reasonable people may disagree.
- E.g., standard utilitarian: unconstrained optimum is to equalize consumption for all.
- But, this is a controversial objective with which to endow the central bank.

Question: How can we analyze optimal MP without taking a strong stance on distributional issues?

Define notion of aggregate surplus, without taking a stance on how it ought to be split.

Surplus measured by resources left if winners from policy reform were to compensate losers.

- Related to **cost-benefit** where surplus is “money” left over once winners comp. losers.
- Traditional cost-benefit analysis measures gains in dollars at constant prices.
- But, MP is a general-equilibrium problem $\Rightarrow$ use agg. surplus from Baqaee–Burstein (2025).

Define notion of aggregate surplus, without taking a stance on how it ought to be split.

Surplus measured by resources left if winners from policy reform were to compensate losers.

- Related to **cost-benefit** where surplus is “money” left over once winners comp. losers.
- Traditional cost-benefit analysis measures gains in dollars at constant prices.
- But, MP is a general-equilibrium problem $\Rightarrow$ use agg. surplus from Baqaee–Burstein (2025).

Characterize **surplus-maximizing monetary policy** in general HANK economy:

- Sticky wages/prices, incomplete markets, heterogeneous labor wedges and incidence of cycles.
- Reform releases resources while delivering pre-reform welfare, by alleviating inefficiencies.
- So, surplus resources can be used to generate Pareto improvement. [Depends on surplus-division rule]

Define notion of aggregate surplus, without taking a stance on how it ought to be split.

Surplus measured by resources left if winners from policy reform were to compensate losers.

- Related to **cost-benefit** where surplus is “money” left over once winners comp. losers.
- Traditional cost-benefit analysis measures gains in dollars at constant prices.
- But, MP is a general-equilibrium problem $\Rightarrow$ use agg. surplus from Baqaee–Burstein (2025).

Characterize **surplus-maximizing monetary policy** in general HANK economy:

- Sticky wages/prices, incomplete markets, heterogeneous labor wedges and incidence of cycles.
- Reform releases resources while delivering pre-reform welfare, by alleviating inefficiencies.
- So, surplus resources can be used to generate Pareto improvement. [Depends on surplus-division rule]

Why? Surplus maximization is an important benchmark because it focuses on efficiency. It's useful to know what policy maximizes efficiency, even if you think CB should have redistributive objectives.

Main result: Heterogeneity matters if it introduces **additional distortions** that **policy can affect**.

Main result: Heterogeneity matters if it introduces **additional distortions** that **policy can affect**.

We use **linear-quadratic methods**. Show that maximizing surplus equivalent to minimizing

Quadratic obj. = RANK loss + incomplete market loss + het. labor wedges loss.

Optimal Policy = Time-dependent component + State-dependent component (shock response).

Homogeneous labor wedges: Time-dependent = 0 & State-dependent = RANK.

- Idiosyncratic risk is costly (efficiency losses), but MP cannot provide insurance/affect distortions.

Heterogeneous labor wedges:

- Stronger incentive to stabilize output gap from new labor wedges (state-dependent component)
- Small ability to provide insurance (time-dependent component)
- Quantitatively: dual-mandate with higher weight on output gap very close to optimal.

1. Positive HANK, transmission and aggregate demand: Kaplan–Moll–Violante (2018); Auclert (2019); Werning (2015); Auclert–Rognlie–Straub (2024); Auclert et al. (2021); Bilbiie (2008, 2025)

Here: adopt the key HANK ingredients from this literature

2. SWF-Normative HANK: Bhandari et al. (2021); Bilbiie (2008, 2025); Bilbiie–Ragot (2021); Nuño–Thomas (2020); Smirnov (2022); Acharya–Challe–Dogra (2023); Dávila–Schaab (2023); La’O–Morrison (2024); Le Grand et al. (2022)

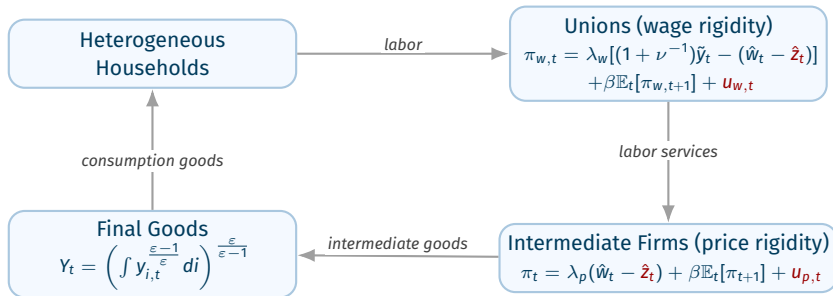
Here: change CB objective. Existing work uses SWF (e.g., utilitarian); we maximize aggregate surplus. No cardinal utility, no Negishi weights, different results.

3. LQ Methods in Optimal Policy Analysis: Barro (1979); Clarida–Gali–Gertler (1999); Woodford (2003); Benigno–Woodford (2005); Gali (2008); Bilbiie (2008, 2025); Bilbiie–Ragot (2021); McKay–Wolf (2022)

Here: derive microfounded LQ problem for heterogeneous agents via maximum surplus
General approach in companion paper: Baqaee–Burstein–Guerreiro, 2026

Model environment

General HANK environment with four building blocks:



Policy: Monetary policy sets time/state-contingent path of interest rate $\{i_t\}$.

Aggregate shocks: Technology, cost-push, wage-push, and demand/financial shocks.

Incomplete markets model with idiosyncratic labor productivity risk $z_{h,t}$.

Preferences $\succeq$ over stochastic sequences $\mathbf{x}_h \equiv \{c_{h,t}, l_{h,t}\}$: consumption and leisure.

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

Budget constraint ($t \geq 1$)

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

- Nominal bond $a_{h,t}$ and securities $\tilde{a}_{k,h,t}$ spanning aggregate risk.
- **Demand shock:** financial wedge δ_t , rebated back in $\tilde{t}_{h,t}$, distorts consumption-savings.

[Like discount factor shock, Smets & Wouters (2007)]

The budget constraint at $t = 0$ also includes a one-off transfer T_h .

Household also faces general borrowing and portfolio constraints.

- Literature documents **heterogeneous consequences of business cycles** for employment and income.
[Guvenen et al. (2017); Patterson (2023); Coglianesi et al. (2025); Bilbiie et al. (2025)]
- Standard to capture this using incidence function.
[Werning (2015); Auclert–Rognlie (2018); Alves–Kaplan–Moll–Violante (2020)]

Labor incidence

Hours set by union with incidence:

$$n_{h,t} = \Gamma(N_t; \zeta_{h,t}) N_t$$

At steady state, assume $n_{h,t}$ is individually optimal.

Relative employment elasticity:

$$\gamma_h \equiv \frac{d \log \Gamma}{d \log N_t}$$

- Literature documents **heterogeneous consequences of business cycles** for employment and income.
[Guvenen et al. (2017); Patterson (2023); Coglianesi et al. (2025); Bilbiie et al. (2025)]
- Standard to capture this using incidence function.
[Werning (2015); Auclert–Rognlie (2018); Alves–Kaplan–Moll–Violante (2020)]

Labor incidence

Hours set by union with incidence:

$$n_{h,t} = \Gamma(N_t; \zeta_{h,t})N_t$$

At steady state, assume $n_{h,t}$ is individually optimal.

Relative employment elasticity:

$$\gamma_h \equiv \frac{d \log \Gamma}{d \log N_t}$$

Income incidence

Labor income plus profits with incidence:

$$y_{h,t} = z_{h,t} \Gamma^Y(N_t; \zeta_{h,t}) Y_t$$

Income can respond differently to employment.

Income incidence elasticity:

$$\gamma_h^Y \equiv \frac{d \log \Gamma^Y}{d \log N_t}$$

- If elasticity is high (> 0), employment/income increases relatively more in expansion.
- Nests **frictionless labor markets**.

SWF-Optimal Monetary Policy

Difficulty: policy is sensitive to welfare weights and how ordinal pref. are mapped to cardinal utils.

Example: Utilitarian social welfare function $\mathcal{W} = \int \bar{\chi}_h U(\mathbf{x}_h) dh$ [$\bar{\chi}_h$: welfare weights.]

[► Details](#)

- Let's make it simple: complete markets, het. initial assets $a_{h,-1}$, perfect foresight.
- Steady-state is Pareto efficient.
- Equal weights: **inflationary bias** to compress wealth distribution $\Rightarrow$ inefficient.
 - Here $\pi_0 = 2$ p.p., but sensitive to how we cardinalize utils (more concavity, π_0 can grow arbitrarily large).
- One solution: find **Negishi weights** so policymaker happy with steady-state distribution.

Difficulty: policy is sensitive to welfare weights and how ordinal pref. are mapped to cardinal utils.

Example: Utilitarian social welfare function $\mathcal{W} = \int \bar{\chi}_h U(\mathbf{x}_h) dh$ [$\bar{\chi}_h$: welfare weights.]

[Details](#)

- Let's make it simple: complete markets, het. initial assets $a_{h,-1}$, perfect foresight.
- Steady-state is Pareto efficient.
- Equal weights: **inflationary bias** to compress wealth distribution $\Rightarrow$ inefficient.
 - Here $\pi_0 = 2$ p.p., but sensitive to how we cardinalize utils (more concavity, π_0 can grow arbitrarily large).
- One solution: find **Negishi weights** so policymaker happy with steady-state distribution.
- But, even then, **redistributive concerns creep into stabilization policy**. Why?
 - Policymaker wants to preserve wealth distribution. So, stabilizes real rates (because they redistribute).
 - E.g, in response to demand shock, tolerates inefficient recession $y_0 = 1.73\%$ & deflation $\pi_0 = -0.15$ p.p.
 - Again, results are sensitive to cardinalization of preferences. More concavity, deeper recession

[Here](#)

The Efficiency Mandate

After reforming policy, how much can we reduce endowments while keeping every household whole?

1. **Status-quo:** policy $\mathcal{P}^0$ generates $\mathbf{x}^0$. Normalize endowments of time $H = 1$.

After reforming policy, how much can we reduce endowments while keeping every household whole?

1. **Status-quo:** policy $\mathcal{P}^0$ generates $\mathbf{x}^0$. Normalize endowments of time $H = 1$.
2. **Reform:** choose policy $\mathcal{P}$, allowing one-time transfers satisfying $\int T_h dh = 0$. [Time-0, cannot insure]

After reforming policy, how much can we reduce endowments while keeping every household whole?

1. **Status-quo:** policy $\mathcal{P}^0$ generates $\mathbf{x}^0$. Normalize endowments of time $H = 1$.
2. **Reform:** choose policy $\mathcal{P}$, allowing one-time transfers satisfying $\int T_h dh = 0$. [Time-0, cannot insure]
3. **Resources saved:** biggest number we can divide endowments by, preserving everyone's welfare.

$$A(\mathcal{P}) = \max \left\{ H : \text{given } \mathcal{P} \text{ and time endowment } 1/H, \text{ there exist transfers } \int T_h dh = 0 \right. \\ \left. \text{supporting equilibrium allocations } \mathbf{x}, \text{ such that } \mathbf{x}_h \succeq \mathbf{x}_h^0 \text{ for all } h \right\}$$

Example: $A = 1.01$ means all can be equally well off with 1% less time.

➤ So, have surplus resources to engineer Pareto improvement.

Definition 1 (The Efficiency Mandate)

The optimal policy solves

$$\max_{\mathcal{P}} A(\mathcal{P})$$

- Depends only on ordinal preference relations and recoverable from **revealed preferences**.
- So, it is invariant to monotone transformations of utility.
- This approach takes a stance on how to conduct monetary policy,
- but, no stance on how to distribute the gains across households. [No redistributive concerns]
- If $A(\mathcal{P}) > 1$, there are Pareto improvements relative to status quo, i.e., there can be consensus.

► Key Properties of Efficiency Criterion

► No Redistributive Concerns

Illustration in RANK Economy

RANK: No heterogeneity, complete markets, so no transfers $T_h = 0$

- Homothetic preferences: consumption of market goods and leisure scale with H

$$\mathbf{x}(H^{-1}, \mathcal{P}) = H^{-1} \cdot \mathbf{x}(1, \mathcal{P})$$

So, aggregate surplus is:

$$A(\mathcal{P}) = \max \left\{ H : \frac{\mathbf{x}(1, \mathcal{P})}{H} \succeq \mathbf{x}^0 \right\}$$

RANK: No heterogeneity, complete markets, so no transfers $T_h = 0$

- Homothetic preferences: consumption of market goods and leisure scale with H

$$\mathbf{x}(H^{-1}, \mathcal{P}) = H^{-1} \cdot \mathbf{x}(1, \mathcal{P})$$

So, aggregate surplus is:

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

RANK: No heterogeneity, complete markets, so no transfers $T_h = 0$

- Homothetic preferences: consumption of market goods and leisure scale with H

$$\mathbf{x}(H^{-1}, \mathcal{P}) = H^{-1} \cdot \mathbf{x}(1, \mathcal{P})$$

So, aggregate surplus is:

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

RANK: No heterogeneity, complete markets, so no transfers $T_h = 0$

- Homothetic preferences: consumption of market goods and leisure scale with H

$$\mathbf{x}(H^{-1}, \mathcal{P}) = H^{-1} \cdot \mathbf{x}(1, \mathcal{P})$$

So, aggregate surplus is:

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

RANK: No heterogeneity, complete markets, so no transfers $T_h = 0$

➤ Homothetic preferences: consumption of market goods and leisure scale with H

$$\mathbf{x}(H^{-1}, \mathcal{P}) = H^{-1} \cdot \mathbf{x}(1, \mathcal{P})$$

So, aggregate surplus is:

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

- With RANK, equivalent to Lucas' welfare measure of consumption-equivalent gain
- RANK optimal policy is equal to standard Ramsey

[Lucas, 1987]

Example: Second-Order Approximation in RANK Economy

In RANK, standard to use linear-quadratic approach.

[Clarida, Gali, and Gertler (1999)]

Example: Second-Order Approximation in RANK Economy

In RANK, standard to use linear-quadratic approach.

[Clarida, Gali, and Gertler (1999)]

$$\log A \approx \underbrace{\log \mathcal{D}}_{\text{First-best surplus}} - \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left\{ (1+\nu^{-1})\tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right\}$$

where $\omega_C \equiv C/(C + wL)$

- $\mathcal{D}$ = distance to first-best (surplus gain from moving to Pareto frontier).
- If policy can stabilize output gap and price and wage inflation, surplus gain equals first best.
- Maximizing RHS subject to linearized equilibrium conditions, delivers approx. optimal policy.

Example: Second-Order Approximation in RANK Economy

In RANK, standard to use linear-quadratic approach.

[Clarida, Gali, and Gertler (1999)]

$$\log A \approx \underbrace{\log \mathcal{D}}_{\text{First-best surplus}} - \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left\{ (1+\nu^{-1})\tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right\}$$

where $\omega_C \equiv C/(C + wL)$

- $\mathcal{D}$ = distance to first-best (surplus gain from moving to Pareto frontier).
- If policy can stabilize output gap and price and wage inflation, surplus gain equals first best.
- Maximizing RHS subject to linearized equilibrium conditions, delivers approx. optimal policy.

Next: Let's do this in HANK.

Approximate Optimal Policy Minimizes Deadweight-Loss Triangles

To make progress, we adopt a **linear-quadratic approach** to solve

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

To make progress, we adopt a **linear-quadratic approach** to solve

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

Theory developed in companion paper: “An Approximate General Theory of Second Best”

- General results for approx. policy in distorted het-agents economies [in spirit Woodford (2003)]
- **Key result:** approximate optimal policy minimizes sum of DWL triangles [To 1st-order in wedges]
- Paper characterizes DWL triangles in terms of expenditure shares, micro price elasticities, and elasticity of wedges to policy [Only observable statistics e.g., no Pareto weights, utils]

To make progress, we adopt a **linear-quadratic approach** to solve

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

Theory developed in companion paper: “An Approximate General Theory of Second Best”

- General results for approx. policy in distorted het-agents economies [in spirit Woodford (2003)]
- **Key result:** approximate optimal policy minimizes sum of DWL triangles [To 1st-order in wedges]
- Paper characterizes DWL triangles in terms of expenditure shares, micro price elasticities, and elasticity of wedges to policy [Only observable statistics e.g., no Pareto weights, utils]
- Extension to distortionary compensating transfers.
- Analytical applications to labor taxes, market power, fiscal unions,...

How does the approximation work?

- For any allocation, implement it in an Arrow-Debreu economy with wedges μ_g
 - Wedges account for market failures (markups, incomplete markets, externalities,...)

From a Global Criterion to a Tractable Policy Problem

How does the approximation work?

- For any allocation, implement it in an Arrow-Debreu economy with wedges μ_g
 - Wedges account for market failures (markups, incomplete markets, externalities,...)
- **Key result:** To second-order in wedges, aggregate surplus is given by:

$$\log A \approx \underbrace{\log \mathcal{D}}_{\text{First-best surplus}} - \sum_g \underbrace{\frac{p_g y_g}{\sum_{g'} p_{g'} y_{g'}}}_{\text{sales of market } g} \underbrace{\frac{d \log y_g^{\text{comp}} \cdot d \log \mu_g}{2}}_{\text{DWL triangle}} + O(\|d \log \mu_g\|^3),$$

$d \log y_g^{\text{comp}}$ is response of quantity to wedges (formula in paper)

From a Global Criterion to a Tractable Policy Problem

How does the approximation work?

- For any allocation, implement it in an Arrow-Debreu economy with wedges μ_g
 - Wedges account for market failures (markups, incomplete markets, externalities,...)
- **Key result:** To second-order in wedges, aggregate surplus is given by:

$$\log A \approx \underbrace{\log \mathcal{D}}_{\text{First-best surplus}} - \sum_g \underbrace{\frac{p_g y_g}{\sum_{g'} p_{g'} y_{g'}}}_{\text{sales of market } g} \underbrace{\frac{d \log y_g^{\text{comp}} \cdot d \log \mu_g}{2}}_{\text{DWL triangle}} + O(\|d \log \mu_g\|^3),$$

$d \log y_g^{\text{comp}}$ is response of quantity to wedges (formula in paper)

- Figure out how each wedge responds to policy $d \log \mu_g / d\mathcal{P}$
- Minimizing sum of DWL triangles by choosing $d\mathcal{P}$ yields first-order approx. of $\mathcal{P}^*$
 - Note that if $d \log y_g^{\text{comp}}$ and $d \log \mu_g$ do not respond to policy, then policy ignores the term.
 - Only take into account DWL triangles you can interact with.

Generalize 2nd order approx. to HANK environment: Three step process

Step 1: Construct Arrow-Debreu economy with wedges

Step 2: Identify the wedges

Step 3: Construct DWL as before

Generalize 2nd order approx. to HANK environment: Three step process

Step 1: Construct Arrow-Debreu economy with wedges

- Wedges are engineered to construct AD economy that replicates original economy.
- Household-specific consumption and labor wedges decentralize the original allocation.
- Firms price $p_{i,t} = \mu_{i,t}^p W_t / Z_t$ and pay wage $w_{u,t} = \mu_{u,t}^w P_t C_t / L_t$.

Step 2: Identify the wedges

Step 3: Construct DWL as before

Generalize 2nd order approx. to HANK environment: Three step process

Step 1: Construct Arrow-Debreu economy with wedges

Step 2: Identify the wedges

➤ Four sources of inefficiency imply four wedges:

- **Price markup:** $\log \mu_{i,t}^P \equiv \log p_{i,t} - \log(W_t/Z_t)$

[Calvo prices]

- **Wage markup:** $\log \mu_{u,t}^W \equiv \log(w_{u,t}/P_t) - \log(C_t/L_t)$

[Calvo wages]

- **Household-specific labor wedge:** $\log \mu_{h,t}^L \equiv \log\left(\frac{c_{h,t}}{z_{h,t}l_{h,t}}\right) - \log(C_t/L_t)$

- **Incomplete market wedge:** $\log \mu_{h,t}^{IM} \equiv \log(C_t/C_0) - \log(c_{h,t}/c_{h,0})$

Step 3: Construct DWL as before

Aggregate surplus equals surplus from first-best minus remaining efficiency losses.

$$\begin{aligned}
 \log A(\mathcal{P}) \approx \log D - & \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\
 & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^Y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\
 & - \underbrace{\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}.
 \end{aligned}$$

Special Cases

Special Case 1: Complete Markets and Homogeneous Labor Wedges

Full insurance and frictionless labor market ($\gamma_h = 0$) leave only the familiar RANK losses.

$$\begin{aligned} \log A(\mathcal{P}) \approx \log \mathcal{D} & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\ & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\ & - \underbrace{\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}. \end{aligned}$$

➤ **Heterogeneity matters** only if it introduces **additional distortions**. In this case, it doesn't.

Special Case 2: Complete Markets and Heterogeneous Labor Wedges

Full insurance removes incomplete-market losses, but heterogeneous labor wedges remain.

$$\begin{aligned}
 \log A(\mathcal{P}) \approx \log \mathcal{D} & - \underbrace{\frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\
 & - \underbrace{\frac{(1-\beta)\omega_C}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\
 & - \underbrace{\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}.
 \end{aligned}$$

➤ **Heterogeneous labor wedges** imply output-gap fluctuations are more costly.

Frictionless labor market without nominal rigidities.

$$\begin{aligned}
 \log A(\mathcal{P}) \approx \log \mathcal{D} & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\
 & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\
 & \underbrace{- \frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}.
 \end{aligned}$$

➤ **Uninsured risk** implies inefficient consumption smoothing.

Frictionless labor market, allowing nominal rigidities and incomplete markets.

$$\begin{aligned}
 \log A(\mathcal{P}) \approx \log D - & \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\
 & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\
 & - \underbrace{\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}.
 \end{aligned}$$

➤ **Canonical HANK** = RANK losses + inefficient consumption smoothing.

Special Case 5: Full Model

Incomplete markets and heterogeneous labor wedges also generate a signed interaction.

$$\begin{aligned}
 \log A(\mathcal{P}) \approx \log D - & \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK losses}} \\
 & - \underbrace{\frac{(1-\beta)\omega_c}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 + (1-\beta) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]}_{\text{Heterogeneous labor wedges}} \\
 & - \underbrace{\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t}_{\text{Incomplete markets}}.
 \end{aligned}$$

- **Additional term** is interaction of all three ingredients. If $\text{Cov}_{\chi_h} > 0$, can improve consumption smoothing by boosting the economy initially.

[Quantitatively negligible]

Optimal Policy with Homogeneous Labor Wedges

Proposition

*Suppose homogeneous incidence on employment $\gamma_h = 0$, so that labor wedges are homogenous. Then optimal policy under the Efficiency Mandate is **identical to the corresponding RANK policy**.*

Heterogeneity does not change the efficient stabilization tradeoff.

Intuition: Monetary policy cannot offer insurance because IM losses are orthogonal to policy.

Examples:

- Sticky-price HANK – workers supply labor optimally given wage [Kaplan-Moll-Violante, 2018]
- Canonical sticky-wage model – equal incidence rule [Auclert-Rognlie-Straub, 2021]

Proposition

*Suppose homogeneous incidence on employment $\gamma_h = 0$, so that labor wedges are homogenous. Then optimal policy under the Efficiency Mandate is **identical to the corresponding RANK policy**.*

Heterogeneity does not change the efficient stabilization tradeoff.

Intuition: Monetary policy cannot offer insurance because IM losses are orthogonal to policy.

Examples:

- Sticky-price HANK – workers supply labor optimally given wage [Kaplan-Moll-Violante, 2018]
- Canonical sticky-wage model – equal incidence rule [Auclert-Rognlie-Straub, 2021]

Additional result: The approximate optimal policy is the same even if we impose $T_h = 0$ for all h , provided shocks are mean zero.

Optimal Policy with Heterogeneous Labor Wedges

Heterogeneous labor wedges arise due to incidence $\gamma_h \neq 0$

Some households are more exposed to aggregate fluctuations than others:

- Output gaps create dispersion in person-specific labor wedges

$$\text{Var}_{x_h}(\gamma_h) > 0$$

- By providing insurance, monetary policy can affect IM wedges

$$\text{Cov}_{x_h}(\log(c_{h,t}/\bar{c}_h), \gamma_h) \neq 0$$

- If covariance is positive, people who overconsume market and leisure goods have labor supply more sensitive to output gap.

Shutdown sticky prices for now.

Proposition

Suppose prices are flexible, optimal policy implies the following inflation output-gap trade-off:

$$\pi_{w,t} + \hat{\psi}^{-1}(\tilde{y}_t - \tilde{y}_{t-1}) = \zeta_{IM} \cdot \frac{\text{Cov}_{\chi_h} \left(\log \left(\frac{c_{h,t}}{c_{h,t-1}} \right), \gamma_h \right)}{\text{Var}_{\chi_h}(\gamma_h)}$$

where $\hat{\psi} < \psi$, $\zeta_{IM} > 0$, and $\tilde{y}_t$ is the output gap.

Two differences relative to RANK:

- $\hat{\psi}^{-1} = \psi^{-1} \left(1 + \frac{\nu^{-1}}{1+\nu^{-1}} \text{Var}_{\chi_h}(\gamma_h) \right) > \psi^{-1}$ more weight on stabilizing output gap than RANK
- Covariance reflects whether exposed people have higher expected consumption growth

Proposition

The short-run IM correction vanishes asymptotically:

$$\lim_{t \rightarrow \infty} \frac{\text{Cov}_{\chi_h} \left(\log \left(\frac{c_{h,t}}{c_{h,t-1}} \right), \gamma_h \right)}{\text{Var}_{\chi_h}(\gamma_h)} = \lim_{t \rightarrow \infty} \frac{\text{Cov}_{\chi_h}(\log c_{h,t}, \gamma_h) - \text{Cov}_{\chi_h}(\log c_{h,t-1}, \gamma_h)}{\text{Var}_{\chi_h}(\gamma_h)} = 0.$$

Hence optimal policy converges to:

$$\pi_{w,t} + \hat{\psi}^{-1}(\tilde{y}_t - \tilde{y}_{t-1}) = 0.$$

This is a dual mandate, with output weight

$$1 + \nu^{-1} [1 + \text{Var}_{\chi_h}(\gamma_h)]$$

- Asymptotically, cannot provide insurance
- Long-run policy is RANK-like, but with a larger output weight
- Heterogeneous labor wedges raise the cost of output fluctuations

General HANK Model with Sticky Prices and Wages

Previous insight carry over readily to sticky price and wage HANK.

Proposition

Optimal allocation is the sum of a time-dependent and a state-contingent component

$$X_t^* \equiv (\tilde{y}_t, \pi_t, \pi_{w,t}, \hat{w}_t) = D_t + S_t, \quad \lim_{t \rightarrow \infty} D_t = 0,$$

where D_t is deterministic and S_t solves the dual-mandate problem (shock response)

$$\min \frac{1}{2} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left\{ \omega_y \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right\}, \quad \omega_y = 1 + \nu^{-1} [1 + \text{Var}_{\chi_h}(\gamma_h)].$$

Optimal policy is the **sum of time-dependent and state-contingent** components

- **Time-dependent** = deterministic path that alleviates incomplete market wedges
- **State-contingent** = respond to shocks as if Dual-Mandate, but higher weight on output gap

So, optimal to stabilize output gap and inflation as in standard RANK

- Fully offset demand shocks.
- Fully accomodate productivity shocks if only have one source of nominal rigidity.

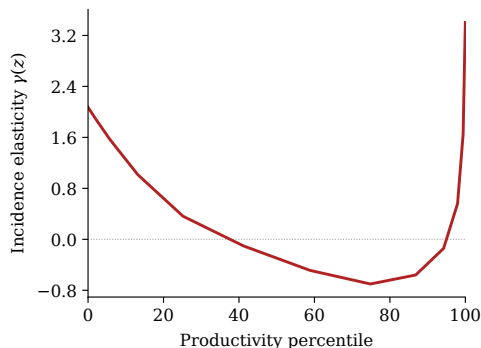
Additional result: Suppose central bank only implements Dual Mandate, then the compensatory transfers are approximately zero $T_h \approx 0$, provided shocks have mean zero.

Quantitative Illustration

Baseline: sticky-wage HANK model (quarterly, U.S.; $\gamma^Y = \gamma$)

Parameter	Description	Value
β	Discount factor	0.9816
$4r$	Annual real interest rate	2%
ν	Steady-state Frisch elasticity	0.75
$\underline{a}$	Borrowing limit (quarterly MPC = 0.25)	-5.66
λ_w	Wage Phillips curve slope	0.0364
ψ	Elasticity across labor varieties	4

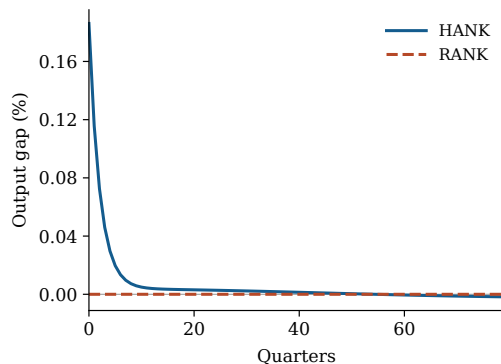
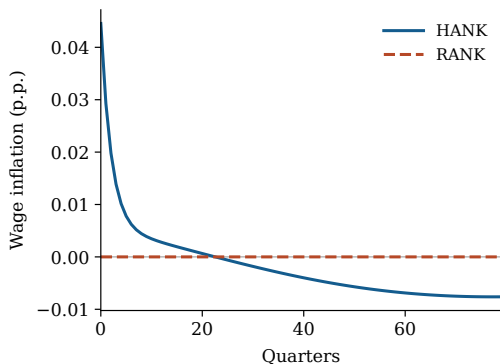
Idiosyncratic productivity: $\rho_z = 0.9767$, $\sigma_z = 0.7084$. Technology $Z = 1.75$ normalizes steady-state output to one.



Calibration: Map worker earnings betas into labor-income incidence.

- Low- and high-productivity households have more cyclical labor income.
- $\text{Var}_{x_h}(\gamma_h) = 0.6443$ implies $\hat{\psi} = 2.92 < 4 = \psi$.
- **36.82% more weight on output-gap stabilization** than in RANK.

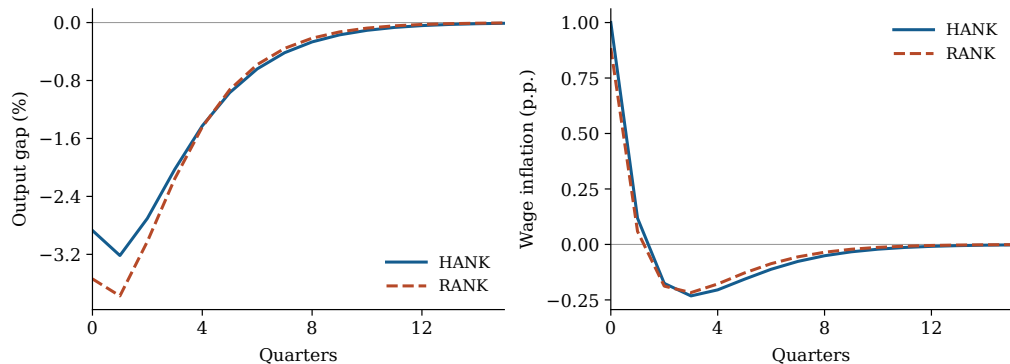
Deterministic Component of Optimal Policy



A small initial expansion: output rises 0.186% and quarterly wage inflation rises 0.045 p.p. The expansion fades, with a small undershooting before returning to steady state.

Response to Cost-Push Shocks

AR(1) shock with persistence 0.5, normalized to raise quarterly HANK wage inflation by 1 p.p.



HANK tolerates more wage inflation to limit the contraction.

- **Present value of the output-gap contraction is 8.66% smaller** than under RANK.

Efficiency Gains from Optimal Monetary Policy

Status quo: $i_t = 1.5\pi_t$, with productivity, demand, and wage cost-push shocks.

Loss component	$\log \mathcal{D}$	$\log A(\mathcal{P}^*)$	Share closed (%)
Output gap	0.0476	0.0458	96.09
Wage dispersion	0.0498	0.0466	93.70
Heterogeneous labor wedges	0.0178	0.0170	95.97
Business-cycle subtotal	0.1152	0.1094	95.04
Incomplete markets	17.4374	0.0001	0.00
Total	17.5525	0.1095	0.62

Incomplete market loss is hundreds of times bigger, but policy has limited ability to address them.

The dual mandate delivers 99.97% of the gains from the full optimum.

- With mean-zero aggregate shocks, it requires no first-order compensatory transfers.

Conclusions

Question: What is optimal MP in HANK for a CB with an efficiency mandate?

Question: What is optimal MP in HANK for a CB with an efficiency mandate?

Main result: Heterogeneity matters if it introduces **additional distortions** that **policy can affect**.

- Homogeneous labor wedges: IM losses are large, but **optimal policy = RANK**.
- Heterogeneous labor wedges: output-gap stabilization weight $\hat{\psi}^{-1}$ is larger than RANK, limited ability to affect IM losses.

Quantitatively: Dual mandate with modified weights achieves almost all of the full policy gains.

Question: What is optimal MP in HANK for a CB with an efficiency mandate?

Main result: Heterogeneity matters if it introduces **additional distortions** that **policy can affect**.

- Homogeneous labor wedges: IM losses are large, but **optimal policy = RANK**.
- Heterogeneous labor wedges: output-gap stabilization weight $\hat{\psi}^{-1}$ is larger than RANK, limited ability to affect IM losses.

Quantitatively: Dual mandate with modified weights achieves almost all of the full policy gains.

Broader message: General theory of 2nd best for policy design in a cost-benefit framework

[No cardinal utils, no SWF, no Pareto weights]

- Tools are broadly applicable, beyond monetary policy. E.g., fiscal unions in companion paper.
- Policies that focus on improving efficiency.

Thank you!

Appendix

Step 1: Identify the wedges

Step 2: Construct Arrow-Debreu economy with wedges

Step 3: Construct Harberger Triangles

Step 4: Figure out how wedges are affected by policy

Step 1: Identify the wedges

➤ Four sources of inefficiency imply four wedges:

- **Price markup:** $\log \mu_{i,t}^P \equiv \log p_{i,t} - \log(W_t/Z_t)$ *[Calvo prices]*
- **Wage markup:** $\log \mu_{u,t}^W \equiv \log(w_{u,t}/P_t) - \log(C_t/L_t)$ *[Calvo wages]*
- **Household-specific labor wedge:** $\log \mu_{h,t}^L \equiv \log\left(\frac{c_{h,t}}{z_{h,t}l_{h,t}}\right) - \log(C_t/L_t)$
- **Incomplete market wedge:** $\log \mu_{h,t}^{IM} \equiv \log(C_t/C_0) - \log(c_{h,t}/c_{h,0})$

Step 2: Construct Arrow-Debreu economy with wedges

Step 3: Construct Harberger Triangles

Step 4: Figure out how wedges are affected by policy

Step 1: Identify the wedges

Step 2: Construct Arrow-Debreu economy with wedges

- Can use these wedges to construct AD economy that replicates original economy
- Household-specific consumption and leisure wedges decentralize the original allocation.
- Firms price $p_{i,t} = \mu_{i,t}^p W_t / Z_t$ and pay wage $w_{u,t} = \mu_{u,t}^w P_t C_t / L_t$.

Step 3: Construct Harberger Triangles

Step 4: Figure out how wedges are affected by policy

Step 1: Identify the wedges

Step 2: Construct Arrow-Debreu economy with wedges

Step 3: Construct Harberger Triangles

➤ The second-order approximation around an efficient reference allocation is:

[Baqae-Burstein-Guerreiro, 2026]

$$\log A(\mathcal{P}) \approx \log \mathcal{D} + \sum_g \underbrace{\frac{p_g y_g}{\sum_{g'} p_{g'} y_{g'}}}_{\text{Sales Share}} \overbrace{\frac{d \log y_g^{\text{comp}} \cdot d \log \mu_g}{2}}^{\text{— Harberger Triangle}}$$

- Aggregate surplus approximated around the efficient reference allocation
 - As in Woodford (2003) for RANK, can also incorporate steady-state/permanent distortions
- Baqaee-Burstein (2026) show that second-order approximation of $\mathcal{D}$ is good in stationary Aiyagari model with idiosyncratic risk and borrowing constraints

Step 4: Figure out how wedges are affected by policy

Across households: expenditure-share weights.

$$\mathbb{E}_{\chi_h}[x_h] = \int_0^1 \chi_h x_h dh, \quad \int_0^1 \chi_h dh = 1.$$

Var_{χ_h} and Cov_{χ_h} use these same weights.

Across dates and states: discounted probability weights.

$$\mathbb{E}_{\beta, \pi}[x_{h,t}] = (1 - \beta) \mathbb{E} \sum_{t=0}^{\infty} \beta^t x_{h,t}, \quad \log \bar{c}_h = \mathbb{E}_{\beta, \pi}[\log c_{h,t}],$$

$$\text{Var}_{\beta, \pi}(x_{h,t}) = \mathbb{E}_{\beta, \pi} \left[(x_{h,t} - \mathbb{E}_{\beta, \pi}[x_{h,t}])^2 \right].$$

Here π in the subscript denotes state probabilities.

Policy gains are reductions in remaining losses. With $g \in \{\text{RANK}, \text{HetLab}, \text{IM}\}$ and $\log A(\mathcal{P}^0) = 0$,

$$\log \mathcal{D} \approx \sum_g \mathcal{H}_g^0, \quad \log A(\mathcal{P}) \approx \sum_g (\mathcal{H}_g^0 - \mathcal{H}_g).$$

Superscript 0 on $\mathcal{H}_g^0$ denotes the status quo. The first-best surplus $\mathcal{D}$ is fixed across policies.

A1. Model: full household problem

Household h maximizes:

$$(1 - \beta) \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1 + \nu} \log c_{h,t} + \frac{\nu}{1 + \nu} \log l_{h,t} \right]$$

Budget constraint:

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

Portfolio constraint: $\Omega(a_{h,t}, \{\tilde{a}_{k,h,t}\}_k) \geq 0$

Idiosyncratic risk: $z_{h,t}$ follows a Markov process with stationary distribution $\pi(z)$

Labor incidence: $n_{h,u,t}^s = \Gamma(1 - L_t; \zeta_{h,t}) \int n_{i,u,t} di, \quad \gamma_h \equiv \frac{\partial \log \Gamma}{\partial \log(1 - L_t)}$

Total income: $y_{h,t} = z_{h,t} \Gamma^Y(1 - L_t; \zeta_{h,t}) Y_t, \quad \gamma_h^Y \equiv \frac{\partial \log \Gamma^Y}{\partial \log(1 - L_t)}$

Transfers: time-0 transfers T_h are budget-neutral, $\int T_h dh = 0$, and are used to implement compensation under the Efficiency Mandate

A2. Aggregate Surplus: Formal Definition

Given a fixed status-quo allocation $\mathbf{x}^0$, aggregate surplus under policy $\mathcal{P}$ is

$$A(\mathcal{P}) = \max \left\{ H > 0 : \exists \mathbf{x} \in \mathcal{X}(1/H; \mathcal{P}), \quad U_h(\mathbf{x}_h) \geq U_h(\mathbf{x}_h^0) \quad \forall h \right\}.$$

$\mathcal{X}(1/H; \mathcal{P})$ allows time-0 lump-sum compensation and scales every household's time endowment by $1/H$.

- Measures the resources that policy can save while keeping everyone indifferent to the status quo.
- Invariant to increasing transformations of individual utility.
- Takes no stance on how to distribute the gains from policy reform.

Surplus from first-best: allow all feasible allocations, $\mathcal{X}^*$.

$$\mathcal{D} = \max \left\{ H > 0 : \exists \mathbf{x} \in \mathcal{X}^*(1/H), \quad U_h(\mathbf{x}_h) \geq U_h(\mathbf{x}_h^0) \quad \forall h \right\}.$$

A3. Aggregate Surplus as Harberger Triangles

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

$$\triangleright^{IM} \equiv \frac{1-\beta}{2} \mathbb{E} \sum_t \beta^t \int \chi_h d \log \mu_{h,t}^{IM} \partial \log x_{h,t} dh,$$

$$\triangleright^p \equiv \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_t \beta^t \int d \log \mu_{i,t}^p \partial \log y_{i,t} di,$$

$$\triangleright^w \equiv \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_t \beta^t \int d \log \mu_{u,t}^w \partial \log n_{u,t}^d du,$$

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

$\log x_{h,t} \equiv [\log c_{h,t} + \nu \log l_{h,t}]/(1+\nu)$ is the consumption-leisure composite.

$d \log \mu$ denotes a change in a wedge; $\partial \log x$ denotes a quantity change caused by wedges.

A4. Incomplete-Market Losses with Homogeneous Labor Wedges

Suppose $\gamma_h = \gamma_h^Y = 0$ for every household. Then

$$\mathcal{H}_{IM} = \frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right].$$

Key result: around the efficient reference allocation, this loss is orthogonal to aggregate monetary conditions.

- Idiosyncratic risk still generates efficiency losses.
- Monetary policy cannot reduce those losses at the order relevant for optimal stabilization.
- The remaining policy tradeoff is the corresponding RANK tradeoff.

Thus the efficiency mandate implements the same first-order policy as RANK.

A5. Efficiency vs. SWF-based HANK analyses

Why do SWF-based analyses find larger departures from RANK?

SWF approach:

- Welfare = $\sum_h \omega_h U_h$, weights ω_h are *normative*
- Redistribution directly enters welfare
- Monetary policy can look attractive because it changes who bears fluctuations
- Result: optimal policy can depart materially from RANK

Efficiency Mandate:

- Aggregate surplus asks how much surplus a policy creates
- Redistribution is separated from the objective by time-0 lump-sum compensation
- The paper's departures from RANK come from efficiency effects, not redistributive motives

The gap between SWF and EM is precisely the redistributive component of optimal MP

A6. Sequence-space methods

We use sequence-space Jacobian (SSJ) methods

[Auclert et al., 2021, 2024]

Key ingredients:

1. Solve for household decision rules in steady state
2. Compute fake-news matrices (heterogeneous impulse responses to news shocks)
3. Compose Jacobians across blocks to get general equilibrium responses

Application to Efficiency Mandate:

- Represent aggregate demand as a functional of expected fundamentals
- Compose block Jacobians to characterize equilibrium responses
- Use those linear responses inside the quadratic approximation of aggregate surplus

Advantage: Handles rich asset-market heterogeneity without curse of dimensionality

Three quantitative objects determine the policy departure from RANK

Object	Value	Role
$\text{Var}_{\chi_h}(\gamma_h)$	0.6443	Raises the output-stabilization weight
$\hat{\psi}$	2.92	Modified RANK tradeoff coefficient
ζ_{IM}	0.0345	Short-run incomplete-markets correction

Consumption dynamics:

- Initial productivity matters less for future expected log consumption over time.
- The consumption-incidence covariance eventually declines toward zero.
- Optimal policy initially expands, followed by a small undershooting.

The dual mandate delivers 99.97% of the full optimal-policy gains.

Optimal Simple Rules

A8. Optimal Simple Rules

Simple rules: restrict policy to time-invariant linear functions of current shocks

[Woodford, 1999]

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{z}_t + \phi_{y,u}u_{w,t}, \quad \pi_{w,t} = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{z}_t + \phi_{\pi,u}u_{w,t}$$

Result 1: Sticky-price HANK = RANK when $\gamma_h = \gamma_h^Y = 0$

Result 2: Sticky-wage incomplete markets drop out

- In the sticky-wage HANK model, the IM covariance term does *not* affect the ex-ante optimal simple rule
- Simple rules are “as if RANK” but with a **modified elasticity** $\tilde{\psi}$

If $\text{Var}_{x_h}(\gamma_h) > 0$ and $\text{Cov}_{x_h}(\gamma_h, \gamma_h^Y) \geq 0$, then $\tilde{\psi} < \psi$: more weight on output stabilization.

A9. Optimal Simple Rules: Cost-Push Shocks

Sticky-wage simple rules with heterogeneous labor wedges:

$$\hat{y}_t = \phi_y + \phi_{y,\delta}\delta_t + \phi_{y,z}\hat{z}_t + \phi_{y,u}u_{w,t}, \quad \pi_{w,t} = \phi_\pi + \phi_{\pi,\delta}\delta_t + \phi_{\pi,z}\hat{z}_t + \phi_{\pi,u}u_{w,t}$$

- Demand shock: zero inflation and output gap = RANK
- Productivity shock: zero inflation and output gap = RANK
- Cost-push shocks: more inflation and less output volatility if $\text{Cov}_{\chi_h}(\gamma_h, \gamma_h^Y) \geq 0$

Cost-push coefficients:

$$\phi_{y,u} = -\frac{\tilde{\psi}}{\lambda_w(1+\nu^{-1})\tilde{\psi} + (1-\beta\rho_u)^2}, \quad \phi_{\pi,u} = \frac{1-\beta\rho_u}{\lambda_w(1+\nu^{-1})\tilde{\psi} + (1-\beta\rho_u)^2}$$

$$\tilde{\psi} = \frac{\psi}{1 + \frac{1}{1+\nu^{-1}} \left[\nu^{-1} \text{Var}_{\chi_h}(\gamma_h) + \frac{(1-\beta)(1-\beta\rho_u^2) \text{Cov}_{\chi_h}(\gamma_h, \gamma_h^Y)}{(1-\beta\rho_u)^2(1+\beta)} \right]}.$$

- If $\text{Cov}_{\chi_h}(\gamma_h, \gamma_h^Y) \geq 0$, then $\tilde{\psi} < \psi$: more weight on output

Cost-push coefficients:

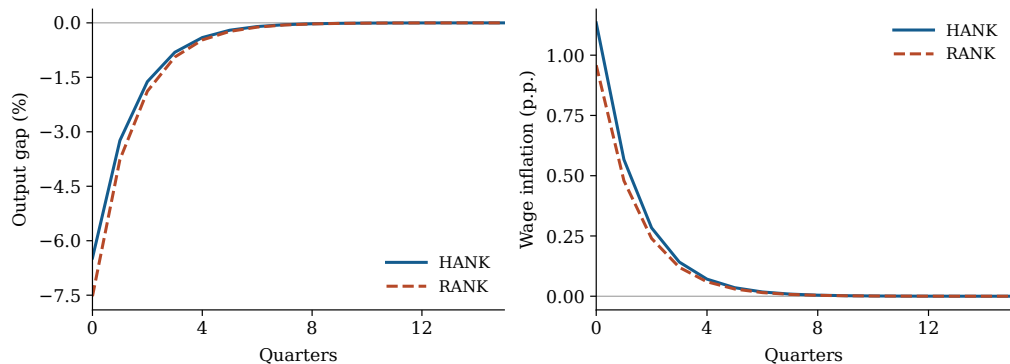
$$\phi_{y,u} = -\frac{\tilde{\psi}}{\lambda_w(1+\nu^{-1})\tilde{\psi} + (1-\beta\rho_u)^2}, \quad \phi_{\pi,u} = \frac{1-\beta\rho_u}{\lambda_w(1+\nu^{-1})\tilde{\psi} + (1-\beta\rho_u)^2}.$$

Calibrated responses relative to RANK ($\rho_u = 0.5$):

	RANK	HANK
Absolute output-gap coefficient	1	0.8602
Wage-inflation coefficient	1	1.1833

- HANK contracts output 13.98% less and raises inflation 18.33% more.
- These differences apply to both impact and cumulative responses.

A11. Optimal Simple Rules: Cost-Push Shocks



Simple rules preserve the main HANK–RANK comparison: more wage inflation and a smaller output contraction under HANK. Same cost-push innovation as in the commitment exercise. Output gaps in percent, wage inflation in quarterly percentage points.

Optimal monetary policy in HANK: [Bilbiie, 2008; Acharya-Dogra, 2020; Le Grand-Ragot, 2022; Challe, 2020; Debortoli-Galí, 2017]

Efficiency criteria and GE welfare: [Baqae-Burstein, 2025a, 2025b; Harberger, 1964; Dixit-Norman, 1986]

HANK models: [Kaplan-Moll-Violante, 2018; Auclert, 2019; McKay-Nakamura-Steinsson, 2016; Bilbiie, 2019]

Sequence-space methods: [Auclert-Bardóczy-Rognlie-Straub, 2021; Auclert-Bardóczy-Rognlie-Straub, 2024]

Labor income risk and cyclical risk: [Floden-Lindé, 2001; Storesletten-Telmer-Yaron, 2004; Guvenen-Ozkan-Song, 2014]

Separation of efficiency and redistribution: [Diamond-Mirrlees, 1971; Atkinson-Stiglitz, 1976; Werning, 2007]

With the wedge definitions above, average price and wage markups satisfy

$$\begin{aligned}\mathbb{E}_i[d \log \mu_{i,t}^p] &= \hat{z}_t - \hat{w}_t, \\ \mathbb{E}_u[d \log \mu_{u,t}^w] &= \hat{w}_t - \hat{z}_t - (1 + \nu^{-1})\tilde{y}_t.\end{aligned}$$

Here $\hat{w}_t = d \log(W_t/P_t)$ and $\hat{l}_t = d \log L_t$. Since $\hat{y}_t^* = \hat{z}_t$, aggregate leisure satisfies $\hat{l}_t = -\nu^{-1}\tilde{y}_t$. Thus

$$\mathbb{E}_i[d \log \mu_{i,t}^p] + \mathbb{E}_u[d \log \mu_{u,t}^w] = -(1 + \nu^{-1})\tilde{y}_t.$$

Together with price and wage dispersion, this gives

$$\triangleright^p + \triangleright^w = -\frac{(1 - \beta)\omega c}{2} \mathbb{E} \sum_t \beta^t \left[(1 + \nu^{-1})\tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right] = -\mathcal{H}_{\text{RANK}}.$$

Combining the price, wage, and labor triangles gives

$$\triangleright^p + \triangleright^w + \triangleright^L = -\mathcal{H}_{\text{RANK}} - \mathcal{H}_{\text{HetLab}}.$$

$$\mathcal{H}_{\text{RANK}} = \frac{(1-\beta)\omega_C}{2} \mathbb{E} \sum_t \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\varepsilon}{\lambda_p} \pi_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right],$$

$$\begin{aligned} \mathcal{H}_{\text{HetLab}} = \frac{(1-\beta)\omega_C}{2} \mathbb{E} \left[\nu^{-1} \text{Var}_{x_h}(\gamma_h) \sum_t \beta^t \tilde{y}_t^2 \right. \\ \left. + (1-\beta) \text{Cov}_{x_h}(\gamma_h, \gamma_h^Y) \left(\sum_t \beta^t \tilde{y}_t \right)^2 \right]. \end{aligned}$$

Labor incidence γ_h determines the physical assignment of labor. Total-income incidence γ_h^Y also affects lifetime wealth.

If $\gamma_h^Y = \gamma_h$, both heterogeneous-labor terms are nonnegative.

Solving for the consumption–leisure response to incomplete-market wedges gives

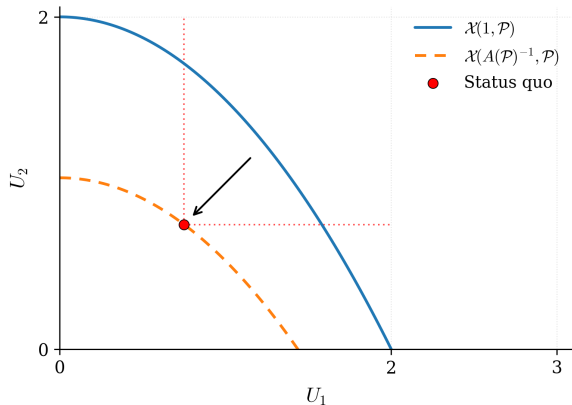
$$\begin{aligned} \triangleright^{IM} = & -\frac{1}{2} \mathbb{E}_{\chi_h} \left[\text{Var}_{\beta, \pi} \left(\log \frac{c_{h,t}}{Y_t} \right) \right] \\ & + \frac{(1-\beta)\omega_c}{2} \mathbb{E} \sum_t \beta^t \text{Cov}_{\chi_h} \left(\log \frac{c_{h,t}}{\bar{c}_h}, \gamma_h \right) \tilde{y}_t. \end{aligned}$$

The first term measures variation in each household's consumption share across dates and states.

The signed interaction captures how an output gap changes labor effort among households with different consumption profiles. It loads on labor incidence γ_h .

Define $\mathcal{H}_{IM} \equiv -\triangleright^{IM}$. Adding all four triangles yields

$$\log A(\mathcal{P}) \approx \log \mathcal{D} - \mathcal{H}_{\text{RANK}} - \mathcal{H}_{\text{HetLab}} - \mathcal{H}_{IM}.$$



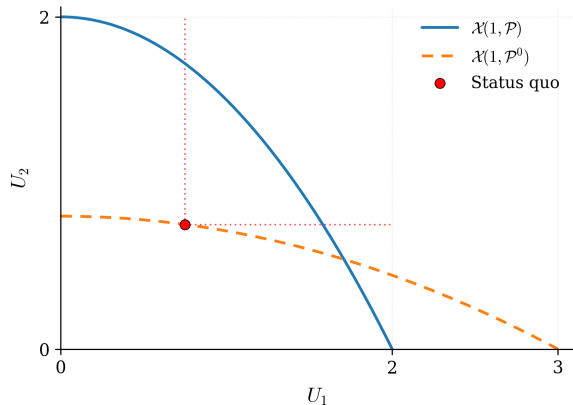
$\mathcal{X}(1; \mathcal{P})$ generates a new frontier – given compensatory transfers

Agg Productivity = how much can contract factor productivity (and so frontier), while keeping everyone at least as well off as under status quo

Status quo is part of the theory and is disciplined by data

Policy problem: look for a policy that generates consensus – feasible Pareto improvements

Theory does not take a stance on what point on new frontier to choose



Scitovsky paradox: Kaldor-Hicks criterion can generate reversals

This does not happen here for two reasons:

- Status quo is part of the theory (the data) and we do not change it – changing status quo is like changing preferences
- We compare status quo to new frontier, not a point. There are points on that frontier that Pareto dominate the status quo – if ultimately those were chosen, then no reversal...

1. **Surplus/Cost-benefit focus:** separates “is there surplus?” from “who gets surplus?” [~~redistributive concerns~~]
2. **Potential-consensus property:** can generate Pareto improvements, process/status-quo matters
 - New allocations reached by consensus, not by coercion [Lockean Pareto principle]
 - Not subject to Scitovsky paradox [▶ More](#)
3. **Ordinal invariance:** no interpersonal utility comparisons, not dependent on cardinal utility
 - No explicit aggregation of utils (SWF), no Pareto weights, no representative-agent fiction required
4. **TFP Equivalent:** measured in factor productivity (interpretable units), preserves Hulten
 - Measures misallocation as resource waste, not as a change in spending
5. **GE consistency:** accounts for GE feedbacks from policy and transfers
6. **Global criterion:** also applies to large changes, not just marginal ones
7. **Empirical implementation:** characterized using only observables (supply and demand curves)
8. **Generalizations:** (1) can incorporate altruism/people’s equity concerns in preferences, (2) can allow for costly redistribution
9. **Nesting and fixes:** nests familiar measures when “well defined”, but also avoids pathologies
 - (1) No Boadway/Kaldor-Hicks paradox, (2) avoids real GDP pathologies with heterogeneous preferences, (3) robust to including general market distortions unlike real GDP/KH
10. **Tractable:** Baqaee-Burstein (2025) develop tools to analyze this criterion in general distorted economies with heterogeneous agents

Policy problem ensures that **Pareto improvements** are possible relative to status quo

But, takes **no stance** over which of the Pareto improvements to choose.

With **individualistic preferences**, this implies policy problem has no redistributive concerns.

But, if people were altruistic, policy problem could have redistributive concerns. [Baqae-Burstein, 2025]

Key: respect people's preferences, but do not impose any additional social preferences.

Complete markets, perfect foresight, flexible prices, sticky wages, $\nu = 0.75$.

$$\mathcal{W} = \int \bar{\chi}_h \frac{1-\beta}{1+\nu} \mathbb{E} \sum_{t=0}^{\infty} \beta^t (\log c_{h,t} + \nu \log l_{h,t}) dh, \quad \int \bar{\chi}_h dh = 1.$$

$$\chi_h = 1 + \frac{a_{h,-1}}{P_0 V}, \quad V \equiv \sum_{t=0}^{\infty} q_t w_t.$$

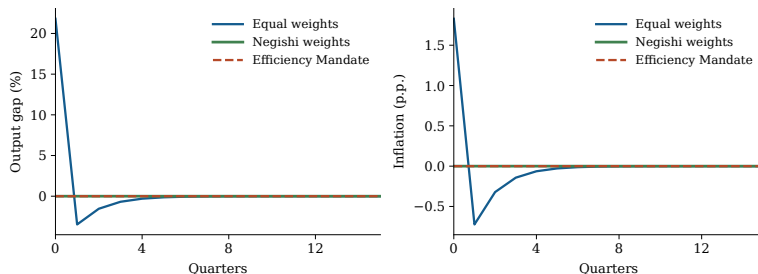
Initial inflation and lower real rates compress wealth shares by raising $P_0 V$.

$$\mathcal{W} - \mathcal{W}^{\text{FB}} \approx - \underbrace{\frac{(1-\beta)\omega_c}{2} \sum_{t=0}^{\infty} \beta^t \left[(1+\nu^{-1}) \tilde{y}_t^2 + \frac{\psi}{\lambda_w} \pi_{w,t}^2 \right]}_{\text{RANK stabilization loss}} - \underbrace{\frac{1}{2} \text{Var}_{\chi_h^0}(\log \chi_h - \log \bar{\chi}_h)}_{\text{Redistributional loss}}.$$

Equal weights target equal shares. Negishi weights target inherited shares, $\bar{\chi}_h = \chi_h^0$.

Here $\omega_c = 1/(1+\nu) = 4/7$. The variance uses fixed reference expenditure weights χ_h^0 . Welfare weights remain fixed across policy and shock counterfactuals.

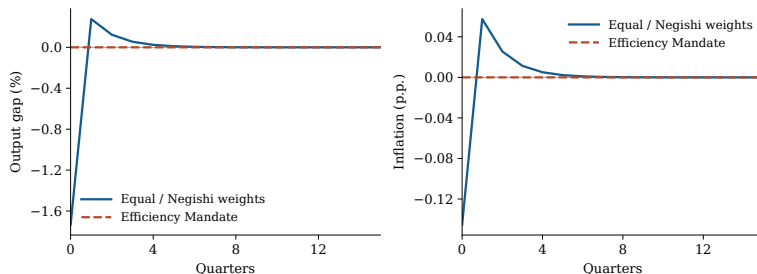
Equal welfare weights induce inflation to redistribute from savers to borrowers.



- Equal weights: initial output gap 21.80% and inflation 1.83 pp.
- Negishi weights and the efficiency mandate keep output and inflation at zero.

Quarters. Output gap in percent, inflation in quarterly pp. $\beta = 0.99$, $\nu = 0.75$, $\psi = 6$, $\lambda_w = 0.05$. Inherited claims use the calibrated asset distribution in GDP units. Unconstrained quadratic policy solutions.

Lower real rates stabilize demand but also compress the wealth distribution.



- Equal and Negishi weights give the same impact responses: output gap -1.73% , inflation -0.145 pp.
- The efficiency mandate fully stabilizes output and inflation.

A 1 pp contractionary demand innovation with persistence 0.9. Responses are relative to each criterion's own no-shock optimum. Quarters, output in percent, inflation in quarterly pp. Quadratic policy solutions.

Keep preferences unchanged, recenter utility at the no-shock status quo, and vary curvature.

$$U_h \equiv \frac{1-\beta}{1+\nu} \mathbb{E} \sum_{t=0}^{\infty} \beta^t (\log c_{h,t} + \nu \log l_{h,t}), \quad \tilde{U}_h \equiv 1 + \frac{U_h - U_h^0}{1-\beta}.$$

$$\mathcal{W}_\rho = \int \bar{\chi}_h f_\rho(\tilde{U}_h) dh, \quad f_\rho(x) = \begin{cases} x^\rho / \rho, & \rho \neq 0, \\ \log x, & \rho = 0, \end{cases} \quad \rho \leq 1.$$

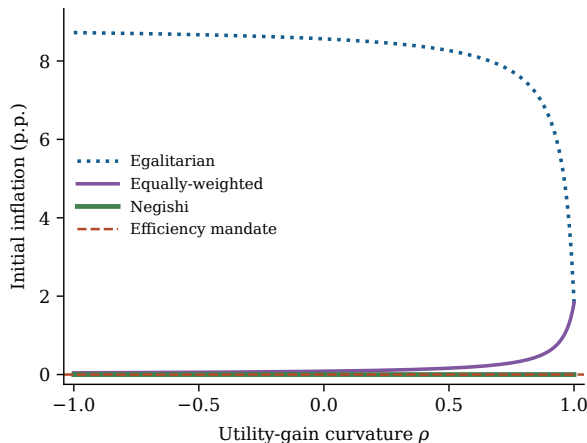
$$(1-\beta) (\mathcal{W}_\rho - \mathcal{W}_\rho^{\text{FB}}) \approx -\mathcal{H}_{\text{RANK}} - \frac{\kappa_\rho}{2} \text{Var}_{\chi_h^0} (\log \chi_h - \log \chi_h^{\text{FB}}), \quad \kappa_\rho = 1 + \frac{1-\rho}{1-\beta}.$$

- **Equal utilitarian weights:** $\bar{\chi}_h = 1$, so $\chi_h^{\text{FB}} \propto (\chi_h^0)^{1-1/\kappa_\rho}$.
- **Egalitarian target:** $\chi_h^{\text{FB}} = 1$. **Negishi target:** $\chi_h^{\text{FB}} = \chi_h^0$.

Targets refer to the quadratic approximation. $\rho = 1$ recovers the baseline. $\mathcal{H}_{\text{RANK}}$ is the baseline stabilization loss. The transformed objective is defined for $\tilde{U}_h > 0$.

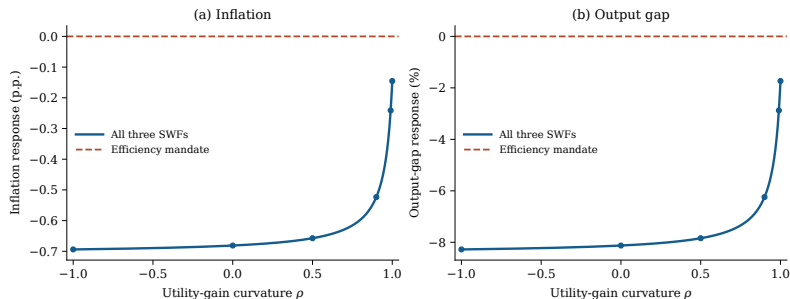
Moving from $\rho = 1$ to $\rho = 0$:

- **Equal utilitarian weights:** initial inflation falls from 1.83 to 0.085 pp.
- **Fixed egalitarian target:** inflation rises from 1.83 to 8.57 pp.
- **Negishi weights:** inflation stays at zero.



Quarterly inflation. $\kappa_\rho = 1 + 100(1 - \rho)$. Curves solve the unconstrained quadratic problem. Large responses are formal implications of this approximation, not exact nonlinear optima.

The three SWFs have the same incremental response at each ρ .



- As ρ falls from 1 to 0, the initial output contraction grows from 1.73% to 8.13%.
- The efficiency mandate fully stabilizes output and inflation for every ρ .

A 1 pp demand innovation with persistence 0.9, relative to each criterion's own no-shock optimum. Output in percent, inflation in quarterly pp. Unconstrained quadratic solutions.