Optimal Monetary Policy with Redistribution*

Jennifer La'O† Wendy Morrison‡

June 13, 2023

Abstract

We study optimal monetary policy in a dynamic, general equilibrium economy with heterogeneous agents. All heterogeneity is ex-ante: workers differ in type-specific, state-contingent labor productivity, yet markets are complete. The fiscal authority has access to a uniform, state-contingent lump-sum tax (or transfer), but linear taxes are restricted to be non-state contingent. We derive sufficient conditions under which implementing flexible-price allocations is optimal. We show that such allocations are not optimal when the relative labor income distribution varies with the business cycle; in such cases, optimal monetary policy implements a state-contingent mark-up that co-moves positively with a sufficient statistic for labor income inequality.

Keywords: monetary policy, inequality, redistribution, household heterogeneity, fiscal policy, nominal rigidity, informational frictions.

JEL codes: E52, D63, H23

*We thank Joao Guerreiro and Elisa Rubbo for excellent and constructive discussions of our paper. We furthermore thank Fernando Alvarez, Marina Azzimonti, Andreas Schaab, Balint Szoke, Harald Uhlig, Ivan Werning, and Mike Woodford for insightful comments, feedback, and suggestions.

†Columbia University, NBER, and CEPR. jenlao@columbia.edu.

‡Columbia University. wam2133@columbia.edu.

1 Introduction

Monetary policy has traditionally been regarded as a tool best suited for macroeconomic stabilization. In recent years, however, there has been growing public opinion that central bankers take heed of rising inequality and acknowledge their potential role in reducing it. While this shift in perspective may or may not be gaining traction among practitioners, it is not obvious from a theoretical standpoint whether monetary policy should be used for such purposes and if so, in what manner.

In this paper we study the optimal conduct of monetary and fiscal policy in a dynamic, general equilibrium model in which households are *ex-ante* heterogeneous and markets are complete. We focus on ex ante heterogeneity, rather than ex post heterogeneity, and therefore on the issue of redistribution rather than lack of insurance.[1] In this context, we ask two questions. First, given a restricted set of fiscal instruments, under what conditions should monetary policy play an active role in redistribution? And second, if such circumstances exist, in what manner should monetary policy be conducted in order to reduce inequality and improve welfare?

Framework and Methodology. We study a general equilibrium, heterogeneous agent economy with nominal rigidities. We model household heterogeneity following Werning (2007). Households are assigned a "type" at birth and remain that type throughout their lifetime. Type-specific labor productivities are stochastic and contingent on the aggregate state; we allow these contingencies to be fully general and can therefore nest any exogenous labor income process. We assume markets are complete: in every period, households can trade a complete set of Arrow securities. It follows that there are no missing insurance markets.

A continuum of intermediate-good firms employ workers, produce differentiated goods, and are subject to aggregate productivity shocks. These firms are monopolistically-competitive and set prices subject to nominal rigidities. We model the nominal rigidity as an informational friction as in Woodford (2003a); Mankiw and Reis (2002); Mackowiak and Wiederholt (2009); Angeletos and La'O (2020); in particular we assume that firms set their nominal prices before perfectly observing realized demand. We initially assume that equity shares of the intermediate-good firms are evenly distributed among all households, but we relax this assumption in our extended model.

The desirability of monetary policy in any context depends on the available set of fiscal instruments. We consider a consolidated government that controls both fiscal and monetary policy. The government raises tax revenue and issues nominal bonds in order to finance exogenous

[1]By focusing on ex ante heterogeneity rather than ex post, our framework stands in contrast to heterogeneous-agent New Keynesian models (HANK), e.g. Kaplan, Moll and Violante (2018); Auclert (2019), that typically feature idiosyncratic labor income risk and incomplete asset markets. We discuss the relationship to the HANK literature below.

shocks to government spending (Lucas and Stokey, 1983) and uniform, lump-sum transfers.

We follow the Ramsey approach and allow for linear taxes on consumption, labor income, firm revenue (sales), and profits. We assume that all tax rates are non-state-contingent, in line with the New Keynesian literature. One can think of this lack of fiscal state-contingency as a political constraint: the fiscal authority cannot change tax rates at business cycle frequency. Furthermore, and in contrast to the typical restriction imposed in the Ramsey literature, we allow for state-contingent, lump-sum taxes or transfers (Werning, 2007). That is, while the fiscal authority cannot change the *slope* of the tax schedule in response to shocks, it can freely move the intercept (subject to the government budget constraint).[2] Crucially, however, we restrict the lump-sum taxes or transfers to be uniform across household types.[3]

Finally, we adopt a utilitarian welfare function with arbitrary Pareto weights. We solve for optimal fiscal and monetary policy jointly using the primal approach (Lucas and Stokey, 1983; Chari, Christiano and Kehoe, 1991, 1994; Chari and Kehoe, 1999). In particular we adapt the primal approach used in Werning (2007) for a flexible-price economy with heterogeneous agents, and that employed in Correia, Nicolini and Teles (2008) for a representative agent economy with nominal rigidities, to our setting that features both heterogeneous households and nominal rigidities.

Results. We first derive sufficient conditions under which implementation of flexible-price allocations is optimal. Specifically, we show that when shocks to the labor skill distribution affect all households proportionally—that is, when there are no movements in workers' *relative* skills—the optimal level of redistribution is achieved through the tax system. In this case, non-state-contingent distortionary taxes and lump-sum transfers are sufficient, and monetary policy should play no redistributive role.

A distortionary tax on consumption or on labor income implies that high-skilled, wealthy households pay more taxes (in levels) than low-skilled, poor households. In combination with a uniform lump-sum transfer, a higher tax rate lowers wealth inequality (Werning, 2007; Correia, 2010). The planner in our environment optimally trades off the redistributional benefit of distortionary taxation with its efficiency cost. When shocks to the labor skill distribution affect all workers proportionally (and preferences are homothetic), both the marginal benefit and the marginal cost to this tax are invariant to the aggregate state. It follows that the optimal wedge is invariant to the the business cycle and, as a result, the restricted set of fiscal instruments is sufficient to implement the planner's optimum. The best that monetary policy can do, in this case, is replicate flexible-price allocations.

[2] For example, while it is difficult for the U.S. Congress to change the tax code, fiscal policymakers were able to issue stimulus checks in response to the Covid-19 pandemic.

[3] One can motivate this restriction with an informational constraint on the government: the fiscal authority cannot tell apart high-type households from low-types.

We show that this is not the case when shocks alter the workers' relative skill distribution. When the labor income of certain households are disproportionally affected by business cycle fluctuations than others, the available set of fiscal instruments is insufficient. It is then optimal for monetary policy to deviate from implementing flexible-price allocations and play an active role in redistribution. In particular, we find that optimal monetary policy targets a state-contingent markup that co-varies positively with a sufficient statistic for labor income inequality.

To understand this result, consider again the case in which labor skill shocks are proportional. A constant tax rate is sufficient to implement the planner's optimum because both the marginal benefit of distortionary taxation (greater redistribution) and the marginal cost (efficiency) are invariant to the aggregate state. When instead labor skill shocks are disproportional, the marginal redistributional benefit of distortionary taxation increases with labor income inequality, while the marginal cost remains the same; it follows that the optimal tax rate in such states should increase.

However, tax rates are assumed to be non-state-contingent. This restriction on fiscal state-contingency is what opens the door for monetary policy to step in and play a redistributive role. In particular, we show that it is optimal for monetary policy to target a higher mark-up when labor market inequality is high and, conversely, a lower mark-up when labor market inequality is low. In doing so, monetary policy imperfectly replicates the missing tax instrument with an "inflation tax" in high inequality states and an "inflation subsidy" in low inequality states.

We show that our results are robust to heterogeneous equity shares. When monetary policy increases the "inflation tax" by targeting a higher mark-up, firm profits increase. Depending on how profit shares co-vary with lifetime income, this can either curb or exacerbate overall income inequality. We show that the presence of heterogeneous equity shares changes both the slope and the intercept of the response of monetary policy to labor income inequality, depending in part on this covariance as well as the degree of firm market power, but it does not alter the general lesson that the optimal markup should covary positively with a sufficient statistic for labor income inequality.

Related literature. The most widely used framework for analyzing monetary policy is the New Keynesian (NK) framework (Woodford, 2003b; Galí, 2008). While much research exists on optimal monetary policy in the New Keynesian model, see e.g. Benigno and Woodford (2003); Schmitt-Grohe and Uribe (2004); Correia, Nicolini and Teles (2008), most studies assume a single representative agent, and therefore cannot speak to distributional concerns.

The more recent heterogeneous agent New Keynesian (HANK) literature explicitly incorporates heterogeneity into the NK model by introducing uninsurable idiosyncratic income risk (Kaplan, Moll and Violante, 2018; Auclert, Rognlie and Straub, 2018). In these models of the

Bewley-Huggett-Aiyagari variety, households use precautionary savings to self-insure against income shocks and smooth their consumption. HANK models can therefore generate an endogenous wealth distribution with heterogeneous marginal propensities to consume, affecting both the amplification and transmission of monetary shocks. Furthermore, monetary policy can play a novel role of providing insurance by transferring resources from savers to borrowers (Acharya, Challe and Dogra, 2020; Dávila and Schaab, 2022; McKay and Wolf, 2022; Bhandari, Evans, Golosov and Sargent, 2021).

In contrast to the HANK model, in our framework markets are complete: households are able to fully insure themselves against idiosyncratic labor income shocks. We thus focus solely on ex-ante heterogeneity rather than ex-post. In doing so, we abstract entirely from the insurance motive for monetary policy and focus solely on the *redistributive* motive.

Empirical evidence suggests that systematic differences in household income are quantitatively important. In particular, Guvenen and Smith (2014) and Schulhofer-Wohl (2011) find that households are able to smooth their consumption to a large degree and that systematic differences between households account for a large share of differences in household income growth. Furthermore, by allowing for fully general labor income processes, we are able to nest those that feature a high degree of heterogeneity in the covariance of individual labor income with aggregate fluctuations. The unequal exposure of individual earnings to business cycles not only appears to be a prominent feature of the data, see e.g. Parker and Vissing-Jorgensen (2009); Guvenen, Schulhofer-Wohl, Song and Yogo (2017); Alves, Kaplan, Moll and Violante (2020); Patterson (Forthcoming), but it is also an important driver of our results.

We show that monetary policy can exploit the redistributional benefits of an "inflation tax" in a manner that is similar to a distortionary tax rate coupled with a lump-sum transfer. The theoretical insight that a flat tax with a lump-sum transfer can reduce inequality has theoretical underpinnings in the macro-public finance literature. In particular, we build on previous insights found in Werning (2007) and Correia (2010).

Our paper is most closely related to the Ramsey literature on optimal taxation, in particular those that apply the primal approach (Lucas and Stokey, 1983; Chari, Christiano and Kehoe, 1991, 1994; Chari and Kehoe, 1999). A subset uses the primal approach to characterize optimal monetary policy in economies with nominal rigidities (Correia, Nicolini and Teles, 2008; Correia, Farhi, Nicolini and Teles, 2013; Angeletos and La'O, 2020). As a methodological contribution, to the best of our knowledge we are the first to show how the primal approach can be used to characterize optimal monetary policy even when the Ramsey optimum cannot be implemented under flexible prices.

Layout. This paper is organized as follows. In Section 2 we describe the model. In Section 3 we characterize the set of allocations that can be implemented as competitive equilibria in this economy. In Section 4 we set up and solve the Ramsey problem. In Section 5 we present implementation of the Ramsey optimum via fiscal and monetary policy. In Section 6 we analyze an extension of our model in which households hold heterogeneous equity shares of all firms. Section 7 concludes. All proofs, except for those explicitly provided in the text, are found in the appendix.

2 The Model

We study a general equilibrium economy with heterogeneous agents and nominal rigidities.

2.1 The Environment

Time is discrete, indexed by $t = 0, 1, \ldots, \infty$. We denote the aggregate state at time t by $s_t \in S$ where S is a finite set. We let $s^t = \{s_0, ..., s_t\} \in S^t$ denote a history of states up to and including time t. We let $\mu(s^t|s^{t-1})$ denote the probability of history s^t conditional on s^{t-1}. Finally, with slight abuse of notation, we denote the unconditional probability of history s^t by $\mu(s^t)$.

Households. There is a measure one continuum of households. Households have identical preferences; in each period, a household receives flow utility $U(c, h)$ from consumption c and work effort h. We assume throughout that preferences are additively-separable and iso-elastic:

$$U(c, h) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{h^{1+\eta}}{1+\eta}, \qquad \text{with} \qquad \gamma, \eta > 0.$$

The parameters γ and η denote the coefficient of relative risk aversion and the inverse Frisch elasticity of labor supply, respectively.

Households are divided into a finite number of types $i \in I$ of relative size π^i, with $\sum_{i \in I} \pi^i = 1$. The worker of a type-i household has "skill" level $\theta^i(s_t)$ in time t, state s_t. If the worker puts in $h^i(s^t)$ units of effort, then its labor in efficiency units are given by: $\ell^i(s^t) = \theta^i(s_t) h^i(s^t)$. Thus, the household maximizes lifetime expected utility given by:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)). \tag{1}$$

The household's budget constraint at time t, history s^t is written in nominal terms by:

$$\begin{aligned} (1+\tau_c)P(s^t)c^i(s^t) + b^i(s^t) - (1+i(s^{t-1}))b^i(s^{t-1}) + \sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t) z^i(s^{t+1}|s^t) \leq \\ (1-\tau_\ell)W(s^t)\ell^i(s^t) + (1-\tau_\Pi)\Pi(s^t) + P(s^t)T(s^t) + z^i(s^t|s^{t-1}) \end{aligned} \tag{2}$$

where $P(s^t)$ is the nominal price of the final good at time t and $W(s^t)$ is the nominal wage per efficiency unit. The household faces constant consumption and labor tax rates, τ_c and τ_ℓ, respectively.

For our baseline analysis we assume that households own equal shares of the intermediate good firms. Equity ownership is a claim to intermediate good firm profits, denoted in nominal terms by $\Pi(s^t)$ and taxed at a constant rate of $\tau_\Pi \in [0,1]$. We relax this assumption and consider heterogeneous equity shares in Section 6.

The household may choose to borrow or save via two separate instruments. The first is a one-period, non-state-contingent bond, $b^i(s^t)$ which the household may buy or sell at time t, history s^t,and which pay $(1+i(s^t))b^i(s^t)$ units of money one period later. The second is a complete set of state-contingent Arrow securities, indexed by $s^{t+1} \in S^{t+1}$. We let $Q(s^{t+1}|s^t)$ denote the price at time t, history s^t, of an Arrow security that pays 1 unit of money in period $t+1$ if state s^{t+1} is realized and 0 otherwise. We denote the corresponding quantities purchased of this Arrow security by $z^i(s^{t+1}|s^t)$. Note that the non-state-contingent bond is a redundant asset but allows us to represent the one-period interest rate, $i(s^t)$.

Finally, $T(s^t)$ is a real, lump-sum transfer and is unrestricted; it can be either positive (a transfer) or negative (a tax) and can depend on the realized history of aggregate states s^t. We state the household's problem as follows.

Household's Problem. *Given initial bond holdings* $b^i(s^0) = 0$ *and Arrow securities* $z^i(s^0) = 0$*, the type-*i *household chooses a complete contingent plan for consumption, efficiency units of labor, bond holdings, and Arrow security holdings:* $\{c^i(s^t), \ell^i(s^t), b^i(s^t), (z^i(s^{t+1}|s^t))_{s^{t+1}}\}_{t\geq 0, s^t \in S^t}$*, in order to maximize its lifetime expected utility* (2) *subject to its budget constraint* (2) *and no-Ponzi conditions.*

Intermediate good production. There is a measure one continuum of intermediate-good firms, indexed by $j \in \mathcal{J} = [0,1]$, with identical technologies. The production function of intermediate-good firm $j \in \mathcal{J}$ is given by the constant returns to scale production function

$$y^j(s^t) = A(s_t) n^j(s^t), \tag{3}$$

where $A(s_t)$ is an aggregate productivity shock and $n^j(s^t)$ is firm j's input of efficiency units of labor.

Intermediate-good firms are monopolistically-competitive: they produce differentiated goods and set nominal prices. The nominal profits of firm j in history s^t are given by $f^j(s^t) = (1-\tau_r)p_t^j(\cdot)y^j(s^t) - W(s^t)n^j(s^t)$ where τ_r is a constant marginal tax on firm revenue. We postpone for the moment our discussion of the nominal rigidity—that is, how the price $p_t^j(\cdot)$ is set.

Final good production. A representative final good firm produces the final good with the following constant elasticity of substitution (CES) technology over intermediate-good varieties:

$$Y(s^t) = \left[\int_{j\in\mathcal{J}} y^j(s^t)^{\frac{\rho-1}{\rho}} \mathrm{d}j\right]^{\frac{\rho}{\rho-1}},$$

with constant elasticity of substitution parameter $\rho > 1$. The final good producer is perfectly competitive and takes prices as given. Its nominal profits are given by $P(s^t)Y(s^t) - \int_{j\in\mathcal{J}} p_t^j(\cdot)y^j(s^t)\mathrm{d}j$ where $p_t^j(\cdot)$ is the price of variety j at time t and $P(s^t)$ is the nominal price of the final good.

Given nominal prices, profit maximization of the representative final good producer implies the standard downward-sloping CES demand function for intermediate good j given by:

$$y^j(s^t) = \left(\frac{p_t^j(\cdot)}{P(s^t)}\right)^{-\rho} Y(s^t), \qquad \forall s^t \in S^t. \tag{4}$$

At its optimum, the representative final good producer makes zero profits.

The government. The government consists of a consolidated monetary and fiscal authority with commitment. Let $\mathcal{T}(s^t)$ denote the nominal tax revenue collected at time t, history s^t, given by:

$$\mathcal{T}(s^t) \equiv \tau_c P(s^t)C(s^t) + \tau_\ell W(s^t)L(s^t) + \tau_r P(s^t)Y(s^t) + \tau_\Pi \Pi(s^t),$$

where

$$C(s^t) \equiv \sum_{i\in I} \pi^i c^i(s^t), \qquad L(s^t) \equiv \sum_{i\in I} \pi^i \ell^i(s^t), \qquad \text{and} \qquad \Pi(s^t) \equiv \int_{j\in\mathcal{J}} f^j(s^t)\mathrm{d}j$$

denote aggregate consumption, aggregate labor supply in efficiency units, and aggregate profits of the intermediate-good firms, respectively.

The government's period-t budget constraint, written in nominal terms, is given by:

$$(1+i(s^{t-1}))B(s^{t-1})+Z(s^t)+P(s^t)T(s^t)+P(s^t)G(s_t) \leq B(s^t)+\sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)Z(s^{t+1})+\mathcal{T}(s^t), \tag{5}$$

where $G(s_t)$ is a government spending shock, and

$$B(s^t) \equiv \sum_{i\in I} \pi^i b^i(s^t), \qquad \text{and} \qquad Z(s^t) \equiv \sum_{i\in I} \pi^i z^i(s^t|s^{t-1}),$$

denote aggregate bond holdings and aggregate Arrow security holdings, respectively.

We assume that the monetary authority directly controls nominal aggregate demand according to the following "ad-hoc" cash-in-advance constraint:

$$M(s^t) = P(s^t)C(s^t).$$

Finally, we abstract from the zero lower bound on the nominal interest rate.

Market Clearing. Market clearing in the goods and labor markets are given by:

$$C(s^t) + G(s_t) = Y(s^t), \qquad \text{and} \qquad L(s^t) = \int_{j\in J} n^j(s^t)\mathrm{d}j,$$

respectively. That is, aggregate consumption and government purchases are equated with aggregate output, and aggregate labor supply (in efficiency units) is equated with labor demand.

2.2 Shocks and the Nominal Rigidity

At each date t, Nature draws the aggregate state $s_t \in S$ according to the probability distribution μ. The aggregate state determines period t total factor productivity, government spending, and relative skills for each type $i \in I$. Formally, we define functions $A : S \to \mathbb{R}_+, G : S \to \mathbb{R}_+$ and $\theta^i : S \to \mathbb{R}_+$ for all $i \in I$, mapping the state s_t at time t to aggregate productivity, government spending, and the relative skill distribution.

The nominal rigidity. Intermediate good firms are price-setters. We model the nominal rigidity as an informational friction as in Woodford (2003a), and Mankiw and Reis (2002). For tractability we follow a particular specification assumed in Correia, Nicolini and Teles (2008); that is, we assume that all firms can set their nominal prices in every period, but in each period a fraction of firms are inattentive to the current state.

Formally, we assume that in every period a mass κ of firms are inattentive, or "sticky," and a mass $1-\kappa$ firms are attentive, or "flexible." We let $\mathcal{J}^s \subset \mathcal{J}$ denote the set of firms that are sticky and $\mathcal{J}^f \subset \mathcal{J}$ denote the set of firms that are flexible, with $\mathcal{J}^f = (\mathcal{J}^s)'$.

Sticky-price firms are inattentive to the current state s_t at time t. They choose their price based only on their knowledge of the history of previous states, s^{t-1}. We denote the price they set by $p^s_t(s^{t-1})$. The subscript t on the price indicates that this is the nominal price set *at time* t by the sticky-price firm, however, the price itself is a function of the history of states only up to time $t-1$. That is, we impose that this price is measurable only up to history s^{t-1}.

The flexible-price firms, on the other hand, are attentive to the current state s_t as well as the entire history of previous states, s^{t-1}. Hence, these firms can set their price as functions of s^t. We denote the price they set by $p^f_t(s^t)$. The subscript t on the price similarly indicates that this is the nominal price set *at time* t by the flexible-price firm. However, unlike the sticky price firms, the flexible-price firms are attentive to the current state s_t, and hence their price is measurable in the current history s^t.

Implicit Timing Assumption. Implicit in the above measurability constraints is the following within-period timing assumption. Nature draws the aggregate state $s_t \in S$ at the beginning of the period and randomly selects which firms are sticky, $j \in \mathcal{J}^s$, and which firms are flexible,

$j \in \mathcal{J}^f$. Intermediate good firms make their nominal pricing decisions given their available information set: s^{t-1}if sticky, s^t if flexible.

Once nominal prices are set, the aggregate state becomes common knowledge. Given prices, households and final good firms make their respective decisions. All allocations adjust so that supply equals demand and markets clear.[4]

2.3 Remarks on the model

This concludes our description of the model. That said, we have made several modeling choices that depart from the standard New Keynesian model, the typical Ramsey framework, as well as more recent HANK models. We discuss these choices below.

Heterogeneity with market completeness. Household types remain fixed, however household labor income can vary over time and over states in a general and flexible manner characterized by the arbitrary function $\theta^i : S \to \mathbb{R}_+$. This formulation nests all labor income processes, including those with a high degree of heterogeneity in the covariance of household labor income with aggregate shocks. However, in the proceeding analysis we show that the complete markets assumption implies that households fully insure themselves against idiosyncratic consumption risk: equilibrium household consumption varies only with aggregate consumption. In this sense there are no missing insurance markets; household heterogeneity in consumption is entirely "ex-ante" rather than "ex post."

Lump-sum transfers. In the standard, representative-agent Ramsey framework, lump-sum taxes or transfers—or any combination of taxes that may replicate them—are a priori ruled out. Were it not the case, the first best would be achievable. When instead households are heterogeneous, Werning (2007) shows that one can incorporate a lump-sum tax or transfer into a Ramsey taxation-style model without sacrificing the earlier lessons from the optimal taxation literature. In such a framework, it is the uniformity of the lump-sum transfer *across* types that ensures that the first best is unattainable. We follow Werning (2007) in this vein and assume the existence of a lump-sum transfer that is uniform across household types. One can think of the uniformity restriction as an informational constraint on the government: the fiscal authority cannot distinguish household types.

Fiscal instrument state-contingency (or lack thereof). The nature of optimal monetary policy often depends on the set of available fiscal instruments. We assume linear taxation as

[4]We make the simplifying assumption that all intermediate-good firms learn the aggregate state at the end of each period. This assumption is compatible with the notion that all firms can observe end-of-period equilibrium outcomes and from these endogenous objects infer the realized state at time t.

in Ramsey; accordingly we allow for consumption, labor income, sales, and profit taxes. We therefore do not artificially restrict the *type* of linear taxes in our model.

However, we constrain these tax instruments to be fixed, i.e. non-state-contingent. This lack of state-contingency is what opens the door for a potential role for monetary policy. State-contingency of monetary policy but non-state-contingency taxes is the typical assumption made in New Keynesian models; it is motivated by the idea that the monetary authority is better suited for responding to business cycle shocks than the fiscal authority.

At the same time we allow the uniform lump-sum transfer to be state-contingent. We find this particular choice of fiscal state-contingency to be reasonable: while legislation of tax rates is often a prolonged and difficult political process, the same is not necessarily true for non-targeted fiscal transfers. Indeed, as a policy response to the past two recessions, Congress sent out non-targeted "stimulus checks." Therefore, fiscal state-contingency of this sort appears feasible.

Nominal Rigidity. Finally, we equate the nominal rigidity in our model with an informational friction. We do so for two reasons. The first is tractability vis-a-vis time- and state-dependent pricing models (e.g. Calvo or menu cost). By assuming only a measurability constraint on firm pricing, the firm's problem becomes static. Every firm is free to adjust its price in every period; it follows that no firm needs to take into account future periods and future states when setting its current period price. The second reason is that by assuming the exact same nominal rigidity present in Correia, Nicolini and Teles (2008), we can directly tie our results to the relevant literature.

3 Equilibrium Definition and Characterization

In this section we define a competitive equilibrium in our economy and characterize the set of equilibrium allocations.

3.1 Equilibrium Definition

We denote an allocation in this economy by:

$$x \equiv \{(c^i(s^t), \ell^i(s^t))_{i \in I}, (y^j(s^t), n^j(s^t))_{j \in \mathcal{J}}, C(s^t), G(s_t), Y(s^t), L(s^t)\}_{t \geq 0, s^t \in S^t}$$

Formally, we say that an allocation x is feasible if it satisfies the economy's technology and resource constraints.

Definition 1. *An allocation x is* feasible *if, for all $s^t \in S^t$:*

$$y^j(s^t) = A(s_t) n^j(s^t), \qquad \forall j \in \mathcal{J}; \tag{6}$$

$$Y(s^t) = \left[\int_{j\in\mathcal{J}} y^j(s^t)^{\frac{\rho-1}{\rho}}\mathrm{d}j\right]^{\frac{\rho}{\rho-1}}; \qquad L(s^t) = \int_{j\in\mathcal{J}} n^j(s^t)\mathrm{d}j; \tag{7}$$

$$C(s^t) = \sum_{i\in I}\pi^i c^i(s^t); \quad L(s^t) = \sum_{i\in I}\pi^i \ell^i(s^t); \quad \textit{and} \qquad C(s^t) + G(s_t) = Y(s^t). \tag{8}$$

Let $\mathcal{X}$ denote the set of all feasible allocations. We are interested in the set of feasible allocations $x \in \mathcal{X}$ that can be supported as part of a competitive equilibrium in our economy. Prior to defining our equilibrium concept(s), we introduce some simplifying notation. We denote a policy by:

$$\Omega \equiv \{\tau_c, \tau_\ell, \tau_r, \tau_\Pi, T(s^t), i(s^t), M(s^t)\}_{t\geq 0, s^t\in S^t},$$

a price system by:

$$\varrho \equiv \{p_t^f(s^t), p_t^s(s^{t-1}), P(s^t), W(s^t), (Q(s^{t+1}|s^t))_{s^{t+1}\in S^{t+1}}\}_{t\geq 0, s^t\in S^t},$$

and a set of financial asset positions by:

$$\zeta \equiv \{(b^i(s^t))_{i\in I}, B(s^t), (z^i(s^{t+1}|s^t), Z(s^{t+1}))_{s^{t+1}\in S^{t+1}}\}_{t\geq 0, s^t\in S^t}.$$

We define an equilibrium in this economy as follows.

Definition 2. *A sticky-price equilibrium is an allocation x, a price system ϱ, a policy Ω, and asset holdings ζ such that: (i) at time t, history s^t, the price $p_t^s(s^{t-1})$ is optimal for all sticky-price firms $j \in \mathcal{J}^s$, the price $p_t^f(s^t)$ is optimal for all flexible-price firms $j \in \mathcal{J}^f$, and the aggregate price level given by:*

$$P(s^t) = \left[\kappa p_t^s(s^{t-1})^{1-\rho} + (1-\kappa)p_t^f(s^t)^{1-\rho}\right]^{\frac{1}{1-\rho}}; \tag{9}$$

(ii) prices and allocations satisfy the CES demand function (4) for all $j \in \mathcal{J}$ at time t; (iii) given the price system and the policy, the allocation and financial asset holdings of type i solve the household problem of type i, for every $i \in I$; (iv) the government budget constraint is satisfied; (v) aggregate nominal demand satisfies $P(s^t)C(s^t) = M(s^t)$; and (vi) markets clear.

In addition to sticky-price equilibria, we will also consider a hypothetical benchmark economy in which we abstract from nominal rigidities. To construct this benchmark we relax the measurability constraints on firms so that all firms have complete information about current fundamentals s_t when making their respective decisions. Formally we call this the "flexible-price" environment and define a competitive equilibrium in this environment accordingly.

Definition 3. *A flexible-price equilibrium is an allocation x, a price system ϱ, a policy Ω, and asset holdings ζ such that: (i) at time t, history s^t, the price $p_t^f(s^t)$ is optimal for all firms $j \in \mathcal{J}^f = \mathcal{J}$, and the aggregate price level given by:*

$$P(s^t) = p_t^f(s^t), \qquad \forall s^t \in S^t; \tag{10}$$

and parts (ii)-(vi) of Definition 2 hold.

The flexible-price environment will serve as a natural benchmark for separating the roles of fiscal and monetary policy in our model.

3.2 Household and Firm optimality

Households. Consider first the individual household's problem; for this we follow the analysis found in Werning (2007).[5] Markets are complete and taxes are linear; this implies that all households face the same after-tax prices. As a result, marginal rates of substitution across all goods and states are equated across households. The next result then follows.

Lemma 1. (Werning, 2007.) *For any equilibrium there exist "market" weights $\varphi \equiv (\varphi^i)_{i\in I}$ with $\varphi^i \geq 0$ so that the individual assignments of consumption and labor solve the following static sub-problem*

$$U^m(C(s^t), L(s^t); \varphi) \equiv \max_{(c^i(s^t), \ell^i(s^t))_{i\in I}} \sum_{i\in I} \varphi^i \pi^i U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)) \tag{11}$$

subject to

$$C(s^t) = \sum_{i\in I} \pi^i c^i(s^t), \qquad \text{and} \qquad L(s^t) = \sum_{i\in I} \pi^i \ell^i(s^t) \tag{12}$$

where the superscript m stands for "market."

Proof. See Appendix A.2. □

That is, any equilibrium delivers an efficient assignment of individual consumption and labor $(c^i(s^t), \ell^i(s^t))_{i\in I}$ given aggregates $(C(s^t), L(s^t))$ and market weights φ. The economy thus behaves *as if* there exists a fictitious representative household with utility function $U^m(C, L; \varphi)$. Relative prices satisfy the representative household's intratemporal condition:

$$-\frac{U^m_L(s^t)}{U^m_C(s^t)} = \left(\frac{1-\tau_\ell}{1+\tau_c}\right)\frac{W(s^t)}{P(s^t)}, \qquad \forall s^t \in S^t, \tag{13}$$

and intertemporal conditions:

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U^m_C(s^{t+1})}{U^m_C(s^t)}\frac{P(s^t)}{P(s^{t+1})}, \qquad \forall s^{t+1} \in S^{t+1}, \tag{14}$$

$$\frac{U^m_C(s^t)}{P(s^t)} = \beta(1+i(s^t)) \sum_{s^{t+1}|s^t} \mu(s^{t+1}|s^t)\frac{U^m_C(s^{t+1})}{P(s^{t+1})}, \qquad \forall s^t \in S^t. \tag{15}$$

where we let $U^m_C(s^t) \equiv \partial U^m(\cdot)/\partial C(s^t)$ and $U^m_L(s^t) \equiv \partial U^m(\cdot)/\partial L(s^t)$ denote the representative household's marginal utilities with respect to aggregate consumption and aggregate labor. Condition (13) states that the representative household's marginal rate of substitution between

[5] See Appendix A.1 for the full derivation of the households' optimality conditions.

consumption and labor is equal to the after-tax real wage. Condition (15) is the bond Euler equation and conditions (14) are the Euler equations for each specific Arrow security.

From the envelope condition of the static sub-problem, $U_C^m(s^t) = \varphi^i U_c^i(s^t)$ and $U_L^m(s^t) = \varphi^i U_\ell^i(s^t)$, where we let $U_c^i(s^t) \equiv \partial U(\cdot)/\partial c^i(s^t)$ and $U_\ell^i(s^t) \equiv \partial U(\cdot)/\partial \ell^i(s^t)$ denote household i's marginal utilities with respect to individual consumption and labor.[6] Therefore equations (13)-(15) hold with U^i in place of U^m, and individual household's marginal rates of substitution are equated to after-tax prices.

With general preferences, the unique solution to the static sub-problem in Lemma 1 implies that individual household consumption and labor can be written as functions of aggregates $(C(s^t), L(s^t))$, market weights φ, and the distribution $(\theta^i(s_t))_{i\in I}$ alone; see Werning (2007). With the additively-separable and iso-elastic preferences assumed in (1), the solution can be written in closed form:

$$c^i(s^t) = \omega_C^i(\varphi)C(s^t) \qquad \text{and} \qquad \ell^i(s^t) = \omega_L^i(\varphi, s_t)L(s^t), \tag{16}$$

with

$$\omega_C^i(\varphi) \equiv \frac{(\varphi^i)^{1/\gamma}}{\sum_{k\in I} \pi^k (\varphi^k)^{1/\gamma}} \qquad \text{and} \qquad \omega_L^i(\varphi, s_t) \equiv \frac{(\varphi^i)^{-1/\eta}\theta^i(s_t)^{\frac{1+\eta}{\eta}}}{\sum_{k\in I} \pi^k (\varphi^k)^{-1/\eta}\theta^k(s_t)^{\frac{1+\eta}{\eta}}}. \tag{17}$$

Therefore, with these preferences, individual consumption and labor are proportional to their aggregates.

The household's shares of aggregate consumption and aggregate labor are given by $\omega_C^i(\varphi)$ and $\omega_L^i(\varphi, s_t)$, respectively. The consumption share is fixed and depends only on the market weights, φ, and the risk aversion parameter, γ. Markets are complete—as a result, individual households insure away all idiosyncratic risk in consumption and face only aggregate risk. In contrast, the share of labor is a function of the market weights, φ, the Frisch elasticity of labor supply, η, as well as the entire distribution of worker productivities $(\theta^i(s_t))_{i\in I}$. The household's share of labor supply is thereby state-contingent: it depends on the household's relative skill in state s_t.[7]

In equilibrium, each household's budget constraint (2) must hold with equality. Using equations (13)-(15) to substitute out after-tax prices, we obtain the following implementability conditions:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi, s_t)L(s^t)\right] \leq U_C^m(s_0)\bar{T}, \qquad \forall i \in I, \tag{18}$$

where

$$\bar{T} \equiv \frac{(1+\tau_c)^{-1}}{U_c^m(s_0)} \sum_t \sum_{s^t} \beta^t \mu(s^t) U_C^m(s^t) \left[T(s^t) + (1-\tau_\Pi)\frac{\Pi(s^t)}{P(s^t)}\right]. \tag{19}$$

[6] Note that $\frac{\partial U(\cdot)}{\partial \ell^i(s^t)} = \frac{1}{\theta^i(s_t)}\frac{\partial U(\cdot)}{\partial h^i(s^t)}$.

[7] Although our model features ex-post differences in labor supply across households, these differences reflect an efficient allocation of labor supply for a given level of aggregate labor. For a discussion of this point, see Werning (2007).

The above implementability conditions are expressed entirely in terms of the aggregate allocation $(C(s^t), L(s^t))$ and the market weights φ. See Appendix A.3 for their derivation.

Condition (18) corresponds to household i's lifetime budget constraint and is similar to the standard implementability condition found in the Ramsey taxation literature; see Chari et al. (1994); Chari and Kehoe (1999). However, in contrast to representative agent economies, rather than equilibrium imposing just one implementability condition of the form in (18), in our economy there exists a *set* of conditions: one for each type $i \in I$.

As noted previously, one stark difference between our framework and the representative-agent Ramsey framework is the existence of lump-sum taxes and transfers, as in Werning (2007). When coupled with labor income taxes, these lump-sum transfers give the planner some ability to redistribute. This power, however, is limited: the planner cannot achieve *any* desired distribution of resources across households because lump-sum transfers are non-targeted. To see this, note that the right hand side of equation (18) represents the present discounted value of lifetime transfers and after-tax profits, denoted by $\bar{T}$, and this value is the same across all types $i \in I$. It follows that the conditions in (18) are joint restrictions on the planner's problem.

Furthermore, in the representative agent Ramsey framework, not only does one typically rule out lump-taxes, but also any combination of taxes that may replicate them. When consumption and labor income taxes are available, this applies to the initial period consumption tax—one can set the initial period consumption tax arbitrarily high and achieve the undistorted optimum. Typically to rule this out, one must treat the initial consumption tax as separate from all other period consumption taxes and impose a binding upper bound; see Chari and Kehoe (1999). Here we have no such issue because we assume the existence of lump sum taxes. It follows that we need no such restriction on the initial period consumption tax; in fact, we simply subsume it into our definition of $\bar{T}$.

Finally, in our framework, due to the monopolistic competition assumption, intermediate-good firms earn equilibrium profits. Equilibrium profits would presumably complicate our analysis as they enter endogenously into household budget constraints as dividend payouts. However, from condition (19) it is evident that profits are isomorphic to lump-sum transfers. This equivalence relies on the assumption of homogeneous equity shares across households; we relax this assumption in Section 6.

Firms. We now turn to the firms' problems and begin by considering that of the flexible-price firms, $j \in \mathcal{J}^f$. We state these firms' problem as follows.

Flexible Price Firm's Problem. *Firm $j \in \mathcal{J}^f$ chooses a nominal price $p_t^j(s^t)$ to maximize firm profits:*

$$p_t^j(s^t) \in \arg\max_{p'} \left\{(1-\tau_r)p' y^j(s^t) - W(s^t)n^j(s^t)\right\}$$

subject to its production function (3) and demand function:

$$y^j(s^t) = \left(\frac{p'}{P(s^t)}\right)^{-\rho} Y(s^t), \qquad \forall s^t \in S^t, \tag{20}$$

The flexible-price firms are attentive to the current state and can therefore choose a price measurable in the history s^t. The solution to this problem is given by:

$$p_t^f(s^t) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}, \qquad \forall s^t \in S^t. \tag{21}$$

That is, the firm optimally equates its marginal cost with its after-tax marginal revenue. This implies that the firm's optimal nominal price equal to a constant mark-up over its nominal marginal cost $W(s^t)/A(s_t)$. The mark-up is a function of the CES parameter ρ and the marginal tax (or subsidy) on revenue.

Consider next the problem of the sticky-price firms, $j \in \mathcal{J}^s$. Sticky-price firms are inattentive to the current state s_t and hence make their nominal pricing decisions based only on their knowledge of the history of previous states, s^{t-1}. Recall that all firms are owned by the households; and the fictitious household's stochastic discount factor in state s^t is given by $U_C^m(s^t)/P(s^t)$. From our previous derivation of household optimality, the Arrow security price $Q(s^t|s^{t-1})$ satisfies (14) and hence can be interpreted as the firm's pricing kernel. We may therefore write the sticky price firm's problem as follows.

Sticky Price Firm's Problem. *Firm $j \in \mathcal{J}^s$ chooses a nominal price $p_t^j(s^{t-1})$ such that it maximizes the expected value of firm profits (weighted appropriately by the market's stochastic discount factor):*

$$p_t^j(s^{t-1}) \in \arg\max_{p'} \sum_{s^t|s^{t-1}} Q(s^t|s^{t-1}) \left\{(1-\tau_r)p' y^j(s^t) - W(s^t) n^j(s^t)\right\}$$

subject to its production function (3) and demand function (20).

Let $y^s(s^t)$ denote the equilibrium output of a sticky-price firm.[8] The solution to the sticky-price firm's problem is given by:

$$p_t^s(s^{t-1}) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1}) \tag{22}$$

where

$$q(s^t|s^{t-1}) \equiv \frac{Q(s^t|s^{t-1}) y^s(s^t)}{\sum_{s^t|s^{t-1}} Q(s^t|s^{t-1}) y^s(s^t)} \tag{23}$$

denote the risk-adjusted conditional probabilities of sticky-price firm j, conditional on history s^{t-1}. Note that these probabilities satisfy $\sum_{s^t} q(s^t|s^{t-1}) = 1$, by construction. Therefore,

[8] We will verify shortly that all sticky-price firms produce the same level of output in equilibrium.

the firm's optimal price is equal to a markup over its risk-weighted expectation of its nominal marginal cost, $W(s^t)/A(s_t)$, conditional on information set s^{t-1}. Comparing this to the optimal price of the flexible-price firm, (21), one can rewrite (22) in the following manner: $p_t^s(s^{t-1}) = \sum_{s^t} q(s^t|s^{t-1})p_t^f(s^t)$. That is, the optimal price of the sticky-price firm is equal to its risk-weighted expectation, conditional on information set s^{t-1}, of the optimal price of the flexible-price firm (Correia, Nicolini and Teles, 2008).

3.3 Equilibrium Allocations

We now characterize the set of allocations that can be implemented as a competitive equilibrium under flexible prices as well as under sticky prices.

Flexible-price equilibria. Consider first the equilibrium under flexible prices. In any such equilibrium, all firms set their price according to (21). This implies that the aggregate price level is given by:

$$P(s^t) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}, \qquad \forall s^t \in S^t. \tag{24}$$

Combining this with the fictitious representative household's optimality conditions, we obtain the following result.

Proposition 1. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a flexible-price equilibrium if and only if there exist market weights $\varphi \equiv (\varphi^i)$, a scalar $\bar{T} \in \mathbb{R}$, and a strictly positive scalar $\chi \in \mathbb{R}_+$, such that the following three sets of conditions are jointly satisfied: (i) for all $s^t \in S^t$, $y^j(s^t) = Y(s^t)$ for all $j \in \mathcal{J}$; (ii)*

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi A(s_t); \tag{25}$$

for all $s^t \in S^t$; and (iii) condition (18) holds for every $i \in I$.

Proof. See Appendix A.4. □

Proposition 1 characterizes the entire set of allocations that can be supported as a flexible-price equilibrium; for shorthand we call such allocations "flexible-price allocations." In addition to resource and technology constraints, any flexible-price allocation satisfies three sets of constraints described in parts (i)-(iii) of the proposition.

Part (i) of Proposition 1 indicates that in any flexible-price equilibrium, there is no output dispersion across firms. Firms share the same technology and face the same nominal wages; as a result they choose the same prices as in (21). It follows from the demand functions (4) that in any flexible-price equilibrium, all firms produce equal levels of output.

Next, part (ii) of Proposition 1 states that in any flexible-price equilibrium, condition (25) must hold in every history. Condition (25) follows from combining the equilibrium price level

(24) with the fictitious representative household's intratemporal optimality condition (13). It follows that, in any flexible-price equilibrium, the marginal rate of substitution between aggregate consumption and aggregate labor is equated with the marginal rate of transformation, $A(s_t)$, modulo a constant wedge, denoted by χ. This wedge is given by:

$$\chi \equiv \left(\frac{\rho-1}{\rho}\right)\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c}. \tag{26}$$

It is thereby the product of multiple terms: the consumption, sales, and labor income taxes levied by the government, and the markup that arises due to monopolistic-competition among intermediate-good producers. It is important to note that χ is a constant—this follows from the assumption that the tax rates, as well as the elasticity of substitution parameter, ρ, are not contingent on the aggregate state.

Finally part (iii) of Proposition 1 states that in any flexible-price equilibrium, condition (18) must hold for each $i \in I$. These implementability conditions ensure that every households' lifetime budget constraint is satisfied. The government's budget constraint holds by Walras's Law.

The power of fiscal policy. The flexible-price economy allows us to isolate the role of fiscal policy in our environment. In particular, the power of the fiscal authority is parameterized by the scalars χ and $\bar{T}$. Consider χ: the fiscal policy can control, via the linear taxes in (26), this wedge. However, note that the fiscal authority's power to influence allocations using this instrument is limited: χ is a scalar, but condition (25) must hold in every history, $s^t \in S^t$. Therefore, because we have assumed non-state-contingent tax rates, the set of feasible allocations that can be implemented as a flexible price equilibrium is constrained.

Next, consider the scalar $\bar{T}$. The fiscal policy can use lump-sum transfers (or taxes) to control the level of the households' budget constraints. However, again the fiscal authority's power to influence allocations using this instrument is limited: condition (18) must hold for every household type $i \in I$, as we have assumed lump-sum transfers are non-targeted. Therefore conditions (18) constrain the set of feasible allocations that can be implemented as a flexible price equilibrium.

Sticky-Price Equilibria. We turn now to the set of allocations that can be supported as part of an equilibrium under sticky prices. In any sticky-price equilibrium, all sticky-price firms set their prices according to (27). It follows from the demand functions (4) that all sticky-price firms produce the same level of output, hire the same amount of labor, and earn the same level of profits; we henceforth denote these objects by $y^s(s^t)$, $n^s(s^t)$, and $\pi^s(s^t)$, respectively. Similarly, in any sticky-price equilibrium, all flexible-price firms set their prices according to (21); by the same logic, we denote their output, labor, and profits by $y^f(s^t)$, $n^f(s^t)$, and $\pi^f(s^t)$, respectively.

Next, note that from we can rewrite the optimal price of the sticky-price firm in (22) as follows:

$$p_t^s(s^{t-1}) = \epsilon(s^t) \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}, \tag{27}$$

where

$$\epsilon(s^t) \equiv \frac{\sum_{s^t|s^{t-1}} q(s^t|s^{t-1}) W(s^t)/A(s_t)}{W(s^t)/A(s_t)} \tag{28}$$

denotes a stochastic wedge between the optimal prices of the sticky- and flexible-price firms. Because the sticky-price firm has incomplete information, it cannot perfectly forecast its ex-post optimal price, i.e. a markup over its nominal marginal cost. The wedge $\epsilon(s^t)$ can therefore be interpreted as the sticky-price firm's "pricing mistake." Formally, $\epsilon(s^t)$ is defined in (28) as the firm's optimal "forecast error" of its nominal marginal cost, $W(s^t)/A(s_t)$, given its incomplete information set s^{t-1}. This brings us to the following characterization.

Proposition 2. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a sticky-price equilibrium if and only if there exist market weights $\varphi \equiv (\varphi^i)$, a scalar $\bar{T} \in \mathbb{R}$, a strictly positive scalar $\chi \in \mathbb{R}_+$, and a positively-valued function $\epsilon : S^t \to \mathbb{R}_+$, such that the following three sets of conditions are jointly satisfied: (i) for all $s^t \in S^t$, $y^j(s^t) = y^f(s^t)$ for all $j \in \mathcal{J}^f$, and $y^j(s^t) = y^s(s^t)$ for all $j \in \mathcal{J}^s$, and*

$$y^s(s^t) = \epsilon(s^t)^{-\rho} y^f(s^t), \tag{29}$$

(ii) for all $s^t \in S^t$:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi \left[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t); \tag{30}$$

and for all $s^{t-1} \in S^{t-1}$:

$$\sum_{s^t|s^{t-1}} U_C^m(s^t) Y(s^t) \left[\frac{\epsilon(s^t)-1}{\kappa\epsilon(s^t) + (1-\kappa)\epsilon(s^t)^{\rho}}\right] \mu(s^t|s^{t-1}) = 0; \tag{31}$$

and (iii) condition (18) holds for every $i \in I$.

Proof. See Appendix A.5. ☐

Proposition 2 characterizes the entire set of allocations that can be supported as a sticky-price equilibrium; for shorthand we call such allocations "sticky-price allocations." Similar to Proposition 1, Proposition 2 states that, aside from satisfying resource and technology constraints, any sticky-price allocation satisfies three additional sets of constraints.

Part (i) of Proposition 2 indicates that in any sticky-price equilibrium, there is no output dispersion within the set of sticky-price firms and similarly no output dispersion within the set of flexible-price firms. However, there can be differences in output across the two sets of firms as indicated by equation (29). The CES demand function in equation (4) implies that relative quantities across the two types of firms must satisfy:

$$\frac{y^s(s^t)}{y^f(s^t)} = \left(\frac{p_t^s(s^{t-1})}{p_t^f(s^t)}\right)^{-\rho}.$$

where $p_t^s(s^{t-1}) = \epsilon(s^t)p_t^f(s^t)$. Combining this with our characterization of the optimal price of the sticky-price firm in equation (27), we obtain equilibrium condition (29) where $\epsilon(s^t)$ is the forecast error of the sticky-price firms in history s^t.

Part (ii) of Proposition 2 states that in any sticky-price equilibrium, condition (30) must hold in every history. Similar to condition (25) in Proposition 1, condition (30) follows from aggregating over individual firm prices according to (9), then combining the aggregate price level with the fictitious representative household's intratemporal optimality condition (13). This equilibrium condition indicates that the marginal rate of substitution between aggregate consumption and aggregate labor is equated with the marginal rate of transformation, $A(s_t)$, modulo a wedge. In this case the wedge is a product of two components. The first is the constant scalar denoted by χ that corresponds to the mark-up and taxes (26). The second is a new, state-contingent component given by $\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}$. This component comes from aggregation over the optimal prices of the sticky- and flexible-price firms, and therefore contains the state-contingent pricing "mistakes" made by the fraction κ of inattentive firms.

Condition (31) is a re-arrangement of our definition of the forecast error $\epsilon(s^t)$ in 28. This condition simply states that "on average," conditional on history s^{t-1},the forecast error must be equal to 1; this follows from the optimal price-setting behavior of the sticky-price firm.

Finally, part (iii) of Proposition 2 is identical to part (iii) of Proposition 1; these conditions ensure that the budget constraint is satisfied for every household type.

The power of monetary policy. Vis-a-vis the flexible-price economy, the variable $\epsilon(s^t)$ in Proposition 2 represents an *additional* control variable of the planner in the sticky-price economy, one that encapsulates the power of monetary policy over real allocations. In particular, the state-contingency of this variable allows the monetary authority to move around the equilibrium intratemporal condition (30) in a manner that the fiscal authority cannot. However, note that this power is limited by conditions (29) and (31): the variable $\epsilon(s^t)$ is indeed the forecast error of the sticky-price firms and, as such, introduces a stochastic wedge between the sticky-price and flexible-price firms' output. That is, by changing $\epsilon(s^t)$, monetary policy is now able to alter the implicit tax wedge in response to changes in the state, but doing so introduces an efficiency cost due to price dispersion.

Lemma 2. *Let $\mathcal{X}^f$ denote the set of all flexible-price allocations and let $\mathcal{X}^s$ denote the set of all sticky-price allocations.*

$$\mathcal{X}^f \subset \mathcal{X}^s.$$

Proof. Take any allocation x that can be implemented under flexible prices—that is, x satisfies the conditions stated in Proposition 1. This allocation satisfies all conditions stated in Proposition 2 with $\epsilon(s^t) = 1$ for all $s^t \in S^t$. As a result, x can be implemented under sticky prices. □

Therefore, any allocation that can be implemented under flexible-prices can also be implemented under sticky prices.

4 The Ramsey Optimum

In this section we define and characterize the Ramsey optimum in his economy. We consider a utilitarian planner with social welfare function given by:

$$\mathcal{U} \equiv \sum_{i \in I} \lambda^i \pi^i \sum_t \sum_{s^t} \beta^t \mu(s^t) U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)) \tag{32}$$

where $\lambda \equiv (\lambda^i)$ denote an arbitrary set of Pareto weights, with $\lambda^i > 0$ for all $i \in I$.

Definition 4. *A Ramsey optimum x^* is an allocation x that maximizes social welfare (32) subject to $x \in \mathcal{X}^s$.*

The goal of our analysis is to characterize the social welfare-maximizing allocation among the set of sticky-price allocations. However, the set of sticky-price allocations, $\mathcal{X}^s$, is fairly complicated: there are a number of constraints that must be satisfied in order for an allocation to be supported as an equilibrium. We thus proceed by first solving an *easier* problem, that of a "relaxed" Ramsey planner.

4.1 The Relaxed Ramsey Problem

The "relaxed" Ramsey planning problem is one in which we maximize over a larger, relaxed set of allocations relative to the set of sticky-price allocations; see Correia, Nicolini and Teles (2008) and Angeletos and La'O (2020) for similar analyses. We define the relaxed set of allocations and an optimum within this set as follows.

Definition 5. *The relaxed set of allocations $\mathcal{X}^R$ is the set of all feasible allocations $x \in \mathcal{X}$ for which there exists a set of market weights $\varphi \equiv (\varphi^i)$ such that condition (18) holds for all $i \in I$. A relaxed Ramsey optimum x^{R*} is an allocation x that maximizes social welfare (32) subject to $x \in \mathcal{X}^R$.*

Relative to the set of sticky-price allocations characterized in Proposition 2, the relaxed set is constructed by dropping all equilibrium conditions stated in parts (i) and (ii) of the proposition, but maintaining those stated in part (iii). The following corollary is the direct result of Proposition 2, Lemma 2, and Definition 5.

Corollary 1. $\mathcal{X}^f \subset \mathcal{X}^s \subset \mathcal{X}^R \subset \mathcal{X}$.

The relaxed set is a strict superset of $\mathcal{X}^s$, the set of sticky-price allocations, and by implication, $\mathcal{X}^f$, the set of flexible-price allocations. One can think of the relaxed Ramsey planner as a planner that has access to a complete set of state-contingent, firm- and/or good-specific tax instruments, and can thus freely choose the equilibrium price of *any* good in *any* state, but does not have access to type-specific lump-sum transfers, therefore must respect the lifetime budget constraints of the households.[9]

Why study the relaxed Ramsey planning problem? This problem is useful for our analysis in the following sense. We will first characterize the relaxed Ramsey optimum x^{R*}. We will then derive necessary and sufficient conditions under which $x^{R*} \in \mathcal{X}^f$, and by implication, $x^{R*} \in \mathcal{X}^s$. Finally, because the relaxed set is a strict superset of the set of sticky-price allocations, it follows that under these conditions, x^{R*} is both the relaxed Ramsey optimum and the *unrelaxed* Ramsey optimum!

Let $\pi^i \nu^i$ denote the Lagrange multiplier on the implementability condition (18) of type $i \in I$; let $\nu \equiv (\nu^i)_{i \in I}$ denote the set of multipliers. Following Werning (2007), we incorporate these constraints into the planner's maximand and define the pseudo-welfare function $\mathcal{W}(\cdot)$:

$$\begin{aligned}\mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) \equiv \sum_{i \in I} \pi^i \{ & \lambda^i U(\omega_C^i(\varphi) C(s^t), \omega_L^i(\varphi, s_t) L(s^t)/\theta^i(s_t)) \\ & + \nu^i \left[U_C^m(s^t) \omega_C^i(\varphi) C(s^t) + U_L^m(s^t) \omega_L^i(\varphi, s_t) L(s^t) \right] \}\end{aligned} \tag{33}$$

We then write the relaxed Ramsey planning problem as follows.

Relaxed Ramsey Planner's Problem. *The Relaxed Ramsey planner chooses an allocation x, market weights $\varphi \equiv (\varphi^i)$, and $\bar{T} \in \mathbb{R}$, in order to maximize*

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) - U_C^m(s_0) \bar{T} \sum_{i \in I} \pi^i \nu^i \tag{34}$$

subject to feasibility: $x \in \mathcal{X}$.

The pseudo-utility function is stated in terms of aggregates alone, making the relaxed Ramsey planning problem quite tractable. One can think of the pseudo-welfare function as a social welfare function that incorporates not only the distributive motives of society, as captured by the Pareto weights, but also the constraints imposed by the households' heterogeneous budget sets.

[9] With a complete set of state-contingent, firm- and/or good-specific tax instruments, parts (i) and (ii) of Proposition 2 are no longer necessary conditions.

Relaxed Ramsey optimum. The following proposition characterizes a relaxed Ramsey optimum given an arbitrary set of Pareto weights. For shorthand, we let $\mathcal{W}_C(s^t) \equiv \partial\mathcal{W}(\cdot)/\partial C(s^t)$ and $\mathcal{W}_L(s^t) \equiv \partial\mathcal{W}(\cdot)/\partial L(s^t)$ denote the marginal pseudo-utility of aggregate consumption and of aggregate labor, respectively.

Proposition 3. *An allocation is a relaxed Ramsey optimum x^{R*} if (i) for all $s^t \in S^t$, $y^j(s^t) = Y(s^t)$ for all $j \in \mathcal{J}$; and (ii)*

$$-\frac{\mathcal{W}_L(s^t)}{\mathcal{W}_C(s^t)} = A(s_t), \qquad \forall s^t \in S^t. \tag{35}$$

Proof. See Appendix A.6. □

Consider first part (ii) of Proposition 3. It is optimal, from the relaxed Ramsey planner's perspective, to set the social marginal rate of substitution between consumption and labor equal to the marginal rate of transformation, $A(s_t)$. In this formulation, the social marginal rate of substitution between consumption and labor is given by the ratio of the marginal pseudo-utility of labor to the marginal pseudo-utility of consumption. The social marginal rate of substitution therefore takes into account the Pareto weights, i.e. the planner's appetite for redistribution, as well as the constraints imposed by household budget sets.

Consider now part (i). It is furthermore optimal, from the relaxed Ramsey planner's perspective, that there be zero output dispersion across intermediate good firms. The relaxed Ramsey optimum thereby preserves production efficiency in the sense of Diamond and Mirrlees (1971). Although the planner chooses to tax certain margins in order to raise money to support lump-sum transfers (or taxes), it does so at the intratemporal margin.

Preservation of production efficiency indicates that a relaxed Ramsey optimum *could be* a flexible-price allocation—in any flexible-price equilibrium, there is zero cross-sectional dispersion in output—but it does not yet tell us *when* such an allocation is implementable under flexible prices. The following result provides an answer.

Theorem 1. *If there exist positive scalars $(\vartheta^1, \vartheta^2, \ldots \vartheta^I) \in \mathbb{R}^I_+$ and a positively-valued function $\Theta : S \to \mathbb{R}_+$ such that the skill distribution satisfies:*

$$\theta^i(s_t) = \vartheta^i \Theta(s_t), \qquad \forall s_t \in S, \tag{36}$$

then: (i) the relaxed Ramsey optimum is implementable as a flexible-price equilibrium, $x^{R*} \in \mathcal{X}^f$; (ii) the relaxed Ramsey optimum is implementable as a sticky-price equilibrium, $x^{R*} \in \mathcal{X}^s$; and (iii) the relaxed Ramsey optimum x^{R*} is an unrelaxed Ramsey optimum, x^*.

Proof. Suppose there exists positive scalars $(\vartheta^1, \vartheta^2, \ldots \vartheta^I) \in \mathbb{R}^I_+$ and a function $\Theta : S \to \mathbb{R}_+$ such that (36) is satisfied. The individual household shares defined in (17) reduce to:

$$\omega^i_C(\varphi) \equiv \frac{(\varphi^i)^{1/\gamma}}{\sum_{j\in I} \pi^j (\varphi^j)^{1/\gamma}} \qquad \text{and} \qquad \omega^i_L(\varphi) \equiv \frac{(\varphi^i)^{-1/\eta}(\vartheta^i)^{\frac{1+\eta}{\eta}}}{\sum_{k\in I} \pi^k (\varphi^k)^{-1/\eta}(\vartheta^k)^{\frac{1+\eta}{\eta}}},$$

and are therefore non-state-contingent. The relaxed Ramsey optimality condition in (35) can be written as follows:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left[\frac{\sum_{i\in I}\pi^i\omega_L^i(\varphi)\left(\lambda^i/\varphi^i+\nu^i(1+\eta)\right)}{\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left(\lambda^i/\varphi^i+\nu^i(1-\gamma)\right)}\right]=A(s_t) \tag{37}$$

Comparing this to the flexible-price intratemporal condition (25), it is clear that (37) can be replicated under flexible prices with an appropriate choice of scalar χ. This proves part (i) of the theorem; part (ii) follows directly from Lemma 2. Finally, part (iii) follows from the fact that x^{R*} is the welfare-maximizing allocation in $\mathcal{X}^R$, and $\mathcal{X}^s \subset \mathcal{X}^R$. Therefore, if $x^{R*} \in \mathcal{X}^s$, then x^{R*} must be the welfare-maximizing allocation in $\mathcal{X}^s$. □

Theorem 1 provides a sufficient condition under which a relaxed Ramsey optimum can be implemented under flexible prices. We henceforth refer to a skill distribution that satisfies this property as one that "exhibits only proportional aggregate shocks."

To understand the intuition behind Theorem 1, it is helpful to first think about the problem of the relaxed Ramsey planner, a planner constrained only by the feasibility of allocations and the household budget implementability conditions (18). This planner faces a trade-off between the benefit of redistribution and its cost. The cost of redistribution is efficiency: if the planner would like to achieve a more equal distribution of resources across households than under laissez-faire, the planner must distort the fictitious household's intratemporal margin between aggregate consumption and aggregate labor, i.e. the after-tax real wage, in order to raise tax revenue. The relaxed Ramsey planner's optimum is thus the point at which, in every state, the marginal benefit of redistribution is equal to the marginal cost; this state-by-state trade-off is captured in the planner's intratemporal optimality condition (35).

Now consider whether this optimum can be achieved under flexible prices. Suppose first that there are no shocks in the economy: TFP, government spending, and the labor skill of each household type is fixed; one can think of this as the economy's "non-stochastic steady state." In the absence of shocks, the marginal benefit of redistribution and its marginal cost are both constant over time. It follows that the relaxed Ramsey optimum can be implemented as an equilibrium under flexible prices with some constant level of distortion χ. A higher tax rate (equivalently, a lower χ) means that greater tax revenue can be collected from high-skilled, wealthy households than from low-skilled, poor households. Because such tax revenue is redistributed via a uniform, lump sum transfer, a greater tax rate implies greater redistribution. The optimal, constant tax rate thus balances the relaxed Ramsey planner's distributional concerns against efficiency in the "non-stochastic steady state" of the economy.

Now consider the case in which there are shocks: TFP shocks, government spending shocks, and shocks to the labor skill distribution. Suppose further that we restrict the latter to feature only proportional aggregate shocks $\Theta(s_t)$ as described in Lemma 1. When such is the case, the ratio of labor productivity between any two household types remains constant over time and over states:

$$\frac{\theta^i(s_t)}{\theta^j(s_t)} = \frac{\vartheta^i}{\vartheta^j}, \qquad \forall s_t \in S.$$

As a result, because there are no shocks to the *relative* skill distribution, the marginal benefit from redistribution *does not vary* over the business cycle. Because the marginal cost also does not vary over the business cycle (technology and preferences are homothetic), the optimum at which the marginal benefit of redistribution equals the marginal cost is invariant to the aggregate state. It follows that the optimal level of redistribution can be achieved under flexible prices with a constant level of distortion χ; this is the result described in Lemma 1.

Finally, when the relaxed Ramsey optimum can be achieved under flexible prices—that is, when the tax system is sufficient to achieve the optimal level of redistribution—then the best that monetary policy can do is to replicate flexible price allocations. We show in Section 5 that it can do so by targeting a constant price level.[10]

Note that the key property that drives this result is the preservation of Diamond and Mirrlees (1971) production efficiency at the relaxed Ramsey optimum. In this sense Theorem 1 is similar to the insight of Correia, Nicolini and Teles (2008). Although the planner in our environment trades-off redistribution with a wedge that distorts the fictitious representative household's intratemporal margin, under no circumstances does the relaxed planner find it optimal to misallocate resources *across* firms. Thus, with only proportional aggregate shocks to the labor skill distribution, there is no reason for monetary policy to introduce such distortions.

Remark. The homotheticity assumption on preferences plays a role in generating the above results. In the proof of Theorem 1, we use the fact that the equilibrium allocation of consumption and labor across households take the form given in (16), which itself relies on the iso-elastic preference specification.

Furthermore, Theorem 1 provides a sufficient condition under which a relaxed Ramsey optimum can be implemented under flexible prices but it is not necessary. In addition there exists a knife-edge, degenerate case, in which the Pareto weights are such that $\lambda^i = \varphi^i$ for all $i \in I$ at the laissez-faire equilibrium, in which case $\nu^i = 0$ for all $i \in I$. In this degenerate case, the laissez-faire equilibrium under flexible prices is always a relaxed Ramsey optimum, and furthermore first-best efficient, for any stochastic process of the skill distribution.

4.2 The (Unrelaxed) Ramsey Problem

The relaxed Ramsey optimum can be implemented as a sticky-price equilibrium only under very special circumstances: proportional aggregate shocks to the labor skill distribution. Away from

[10] Equivalently, by setting $\epsilon(s^t) = 1$ for all $s^t \in S^t$.

this case, though, it is not yet obvious what the optimal sticky-price allocation is, and hence what monetary policy should do. In order to answer this question, we now solve our original, slightly more difficult problem, that of the "unrelaxed" Ramsey planner, as in Definition 4.

Again letting $\pi^i \nu^i$ denote the Lagrange multipliers on the budget implementability conditions (18), we show that we can write the Ramsey planning problem in the following manner.

Ramsey Planner's Problem. *The Ramsey planner chooses an aggregate allocation,* $(C(s^t), Y(s^t), L(s^t))_{s^t \in S^t}$*, market weights* $\varphi \equiv (\varphi^i)$*, constants* $\bar{T} \in \mathbb{R}$ *and* $\chi \in \mathbb{R}_+$*, and* $\epsilon : S^t \to \mathbb{R}_+$*, in order to maximize the pseudo-utility function in* (34) *subject to* (30)*,* (31)*, and the aggregate resource constraint:*

$$C(s^t) + G(s_t) \leq Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t), \qquad \forall s^t \in S^t \tag{38}$$

where $\Delta : \mathbb{R}_+ \to \mathbb{R}_+$ *is a function defined by:*

$$\Delta(\epsilon) \equiv \left\{ \frac{\left[\kappa\epsilon^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\left[\kappa\epsilon^{-\rho} + (1-\kappa)\right]^{1/\rho}} \right\}^{\rho} > 0. \tag{39}$$

Given an aggregate allocation and market weights φ*, the individual allocations of household consumption and labor satisfy* (16) *and* (17)*, and the allocation of production across sticky- and flexible-price firms are given by* (29).

Following the same steps as before, we incorporate the budget implementability conditions into the planner's maximand via the pseudo-utility function. However, relative to the relaxed Ramsey planning problem, the unrelaxed Ramsey planner must satisfy *all* implementability conditions in Proposition 2. This includes the equilibrium intratemporal condition of the fictitious representative household, (30), as well as (31).

The planning problem stated above simplifies the Ramsey problem by restating it in terms of aggregates alone. Recall that the pseudo-utility function in (34) is a function of aggregate consumption and aggregate labor and hence already incorporates the heterogeneity across households.

The Ramsey planner must furthermore respect the heterogeneity that might occur across sticky- and flexible-price firms within each period. We show that the effects of such heterogeneity can be captured solely by its impact on TFP: in particular, the multiplicative term $\Delta(\epsilon)$ in the aggregate resource constraint (38) represents TFP loss due to misallocation of inputs across sticky- and flexible-price firms. This term is otherwise known as the efficiency wedge (Chari, Kehoe and McGrattan, 2007).[11] To see how Δ captures misallocation more clearly, we provide the following characterization.

[11] See Appendix A.7 for the derivation of the aggregate resource constraint (38).

Lemma 3. *The function* $\Delta : \mathbb{R}_+ \to \mathbb{R}_+$ *is strictly concave and satisfies* $\max_{\epsilon>0} \Delta(\epsilon) = 1$. *Furthermore, it attains its unique maximum at* $\epsilon = 1$.

Proof. See Appendix A.8. □

When monetary policy implements flexible-price allocations—that is, when it sets $\epsilon = 1$ in all states—then $\Delta(\epsilon)$ attains its unique maximum of 1. In this case, there is no misallocation across firms and therefore no loss in production efficiency. On the other hand, when monetary policy deviates from implementing flexible-price allocations—that is, when it sets $\epsilon \neq 1$ in some or all states—then in those states, $\Delta(\epsilon)$ is strictly below 1. In this case, the "active" use of monetary policy leads to forecast errors of the sticky-price firms. Dispersion of prices across sticky- and flexible-price firms implies misallocation of inputs and results in TFP loss.

The term $\Delta(\epsilon)$ therefore represents the consequence of using monetary policy in an active manner. The benefit of using monetary policy, however, as mentioned previously, is that it can serve as an imperfect substitute for the missing state-contingency of taxes.

Finally, for tractability, we make the following assumption:

Assumption 1. *Let* $G(s_t) = 0$ *for all* $s_t \in S$.

That is, we shut down government spending. We herein impose Assumption 1 in order to simplify the statements and proofs of the following results.

Optimal monetary wedge. Appendix A.9 provides a characterization of the Ramsey optimum.[12] While we do not provide this characterization here for expositional purposes, the essential necessary condition of the planner's optimum appears as follows:

$$-\frac{\mathcal{W}_L(s^t)}{\mathcal{W}_C(s^t)} = \left[\text{Ramsey wedge}(s^t)\right] \frac{Y(s^t)}{L(s^t)}. \tag{40}$$

Condition (40) is the Ramsey planner's intratemporal optimality condition; it is the counterpart to condition (35) of the relaxed Ramsey planner. The Ramsey planner sets the social marginal rate of substitution between consumption and labor, $-\mathcal{W}_L(s^t)/\mathcal{W}_C(s^t)$ equal to the marginal rate of transformation, $Y(s^t)/L(s^t)$, modulo a state-contingent wedge. This wedge is a function of the state-contingent Lagrange multiplier on equilibrium condition (30).

Relative to the relaxed planner, the Ramsey planner is subject to condition (30): this condition must be satisfied in order for an allocation to be implementable as a competitive equilibrium under sticky prices. When condition (30) is non-binding, the multiplier on it is zero, and therefore the social MRS is equalized with the MRT as in the relaxed Ramsey optimum. When instead condition (30) is binding, the multiplier on it is non-zero, in which case the social MRS departs from the MRT at the Ramsey optimum.

[12] See Proposition 8 in Appendix A.9.

Furthermore, in contrast to the relaxed Ramsey optimum, in the Ramsey optimum the marginal rate of transformation between labor and consumption is no more $A(s_t)$, but instead $Y(s^t)/L(s^t) = A(s_t)\Delta(\epsilon(s^t))$. As long as the Ramsey planner finds it optimal to deviate from flexible-price allocations, and can do so utilizing monetary policy, the efficiency wedge $\Delta(\epsilon(s^t))$ is strictly less than one and therefore affects the marginal rate of transformation.

We use our characterization of the Ramsey optimum to characterize optimal monetary policy. Following the Ramsey literature, we begin by presenting the implicit labor wedge that implements the planner's optimum. That is, for a Ramsey optimum x^*, we define an implicit "monetary wedge," $1 - \tau_M^*(s^t)$, by:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi^*(1 - \tau_M^*(s^t))\frac{Y(s^t)}{L(s^t)}, \tag{41}$$

where χ^* denotes the implicit fiscal wedge at this allocation. The following theorem provides a characterization of $\tau_M^*(s^t)$, the optimal "monetary tax."

Theorem 2. *Let $\mathcal{I} : S \to \mathbb{R}_+$ be a positively-valued function defined as:*

$$\mathcal{I}(s_t) \equiv \frac{\sum_{i\in I} \tilde{\pi}^i(\varphi^i)^{-1/\eta}(\theta^i(s_t))^{\frac{1+\eta}{\eta}}}{\sum_{i\in I} \pi^i(\varphi^i)^{-1/\eta}(\theta^i(s_t))^{\frac{1+\eta}{\eta}}} > 0, \qquad \textit{where} \qquad \tilde{\pi}^i \equiv \pi^i\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right] \tag{42}$$

There exists a threshold $\bar{\mathcal{I}}(s^{t-1}) > 0$, measurable in history s^{t-1}, such that optimal implicit monetary tax $\tau_M^*(s^t)$ satisfies:

$$\begin{array}{lll} \tau_M^*(s^t) > 0 & \textit{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1}), \\ \tau_M^*(s^t) = 0 & \textit{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}), \\ \tau_M^*(s^t) < 0 & \textit{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1}). \end{array}$$

Proof. See Appendix A.10. □

Recall that λ^i are the Pareto weights, φ^i are the market weights, and ν^i are the multipliers on the implementability conditions in (18). Notably these are all non-state-contingent constants. As a result, $\mathcal{I}$ is a function of the current state s_t alone, not the history, and in particular depends only on the labor skill distribution, $(\theta^i(s_t))_{i\in I}$.

The function $\mathcal{I}(s_t)$ can be interpreted as a sufficient statistic for the level of labor income inequality in this economy. As we show in an example below, the term $\lambda^i/\varphi^i + \nu^i(1+\eta)$ is increasing in the household's type: households with greater labor productivity have a greater weight, $\lambda^i/\varphi^i + \nu^i(1+\eta)$ at the Ramsey optimum than households with lower labor market productivity.[13] As the labor productivity $\theta^i(s_t)$ of the high-type households increases relative to that of the low types, the numerator of $\mathcal{I}(s_t)$ grows relative to its denominator. As a result, $\mathcal{I}(s_t)$

[13] While it is true that high-type households have high market weights, φ^i, at the Ramsey optimum, their multipliers ν^i, are also high and dominate the overall direction of this term.

is high in states in which high types are relatively more productive than low types, and $\mathcal{I}(s_t)$ is low in states where the converse is true.

Theorem 2 states that the optimal monetary tax varies with the state and depends on the level of labor income inequality, as proxied by $\mathcal{I}(s_t)$. When labor income inequality is strictly greater than a threshold $\bar{\mathcal{I}}(s^{t-1})$, the implicit monetary tax is positive. On the other hand, when labor income inequality is strictly less than this threshold, the implicit monetary tax is negative (i.e. a subsidy).

To understand the intuition for this result, recall the intuition behind Theorem 1. When the labor skill distribution exhibits only proportional aggregate shocks, the tax system is sufficient to achieve the optimal level of redistribution. In this case, fiscal policy optimally trades off the benefit of redistribution with its cost—that is, the efficiency cost from distorting aggregate consumption and aggregate labor—and this trade-off does not vary over time. Note that Theorem 2 nests this as a special case: when the labor skill distribution satisfies (36), the function $\mathcal{I}(s_t)$ reduces to a constant equal to $\bar{\mathcal{I}}$ in all states, and the optimal monetary tax is zero.

Starting from $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$, now consider a small shock to the *relative* skill distribution that raises $\mathcal{I}(s_t)$ above $\bar{\mathcal{I}}(s^{t-1})$. The marginal benefit to redistribution increases in response to this shock while the marginal cost remains the same. It follows that if state-contingent tax rates were available, the optimal tax rate would increase. A higher tax rate implies that high-skilled, rich workers pay more taxes than low-skilled, poor workers, but everyone receives the same lump-sum transfer; hence an increase in the optimal tax rate provides greater redistribution.

When state-contingent taxes are unavailable, however, it becomes optimal for monetary policy to step in and fill this role. Thus, when $\mathcal{I}(s_t)$ rises above $\bar{\mathcal{I}}(s^{t-1})$, it is optimal for monetary policy to abandon implementing flexible-price allocations and instead *mimic* an increase in the tax rate. Conversely, when $\mathcal{I}(s_t)$ falls below $\bar{\mathcal{I}}(s^{t-1})$, it is optimal for monetary policy to mimic a fall in the tax rate, that is, to act as a subsidy: $\tau_M^*(s^t) < 0$.

The Loss in Production Efficiency. There is a clear distinction between using monetary and fiscal policy. Both monetary and fiscal policy can be used to drive a wedge between the MRS and the MRT of aggregate consumption and aggregate labor. However, unlike state-contingent fiscal policy, state-contingent monetary policy leads to an additional type of distortion: a wedge between the prices of sticky-price firms and those of flexible-price firms. Equilibrium price dispersion results in misallocation and, ultimately, a loss in production efficiency.

For this reason, monetary policy should be considered an imperfect substitute for missing tax instruments. Were state-contingent tax instruments readily available, one could use these instruments to implement a relaxed Ramsey optimum x^{R*} without any corresponding loss in production efficiency. However, when such fiscal state-contingency is ruled out, as we have assumed a priori, the next best tool is monetary policy. In this case, the best possible allocation is a Ramsey optimum x^* which necessarily features misallocation across sticky- and flexible-price firms whenever $\mathcal{I}(s_t) \neq \bar{\mathcal{I}}(s^{t-1})$.

Note that this additional efficiency cost of using monetary policy does not negate the intuition provide above. To see this, start again from $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$. Here, the fiscal policy is set so that the marginal benefit of redistribution is equal to the marginal cost of distorting the intratemporal margin, and monetary policy implements the flexible price allocation ($\tau_M^*(s^t) = 0$).

Now consider a small deviation of $\mathcal{I}(s_t)$ above $\bar{\mathcal{I}}(s^{t-1})$. The marginal benefit to redistribution increases, the marginal cost of distorting the intratemporal margin stays the same, which implies that monetary policy should abandon the flexible-price benchmark. But the marginal cost in production efficiency due to such abandonment is, to a first-order, zero. This is because at the flexible-price allocation, production efficiency is maximized:

$$\Delta(1) = \max_{\epsilon > 0} \Delta(\epsilon) = 1.$$

Therefore, any loss in production efficiency due to misallocation of intermediate goods around this benchmark must be of second-order. It follows that for small deviations of $\mathcal{I}(s_t)$ above $\bar{\mathcal{I}}(s^{t-1})$, the optimal monetary tax must be strictly positive: $\tau_M^*(s^t) > 0$.

The intuition provided above holds for small shocks around $\bar{\mathcal{I}}(s^{t-1})$, but what about for large shocks? In fact, Theorem 2 has makes no provision that $\mathcal{I}(s_t)$ be close to $\bar{\mathcal{I}}(s^{t-1})$.

The intuition does not change even when $\mathcal{I}(s_t)$ is far from $\bar{\mathcal{I}}(s^{t-1})$. As monetary policy moves further and further away from implementing flexible-price allocations, it is true that losses in production efficiency become first-order. However, these losses can never be strong enough to reverse the sign of monetary policy—such an occurrence would lead to a contradiction. This is because the only reason monetary policy abandons the flexible-price benchmark in the first place is planner's distributional motive. Even if the loss in production efficiency may dampen the extent to which monetary policy mimics a missing tax instrument as $\mathcal{I}(s_t)$ moves further and further away from $\bar{\mathcal{I}}$, it can never force monetary policy to reverse sign: were that the case, monetary policy could always do better by reverting to implementation of flexible-price allocations, therefore contradicting the Ramsey optimality of the allocation.

4.3 Numerical Illustration

We illustrate the mechanism underlying Theorem 2 with a simple numerical example. Suppose there are only two household types, $i \in \{H, L\}$ of equal sizes ($\pi^H = \pi^L = 1/2$). We consider a labor skill distribution that features non-proportional shocks: in particular, we let the ratio of θ^H/θ^L fluctuate across $N = 10$ possible states. For simplicity we assume states are i.i.d. and uniformly distributed so that $\mu(s'|s) = 1/N$ for all $s, s' \in S$. Finally, we set $\eta = .5$, $\gamma = 2$, $\beta = .98$, $\kappa = .5$ and $\rho = 2$.

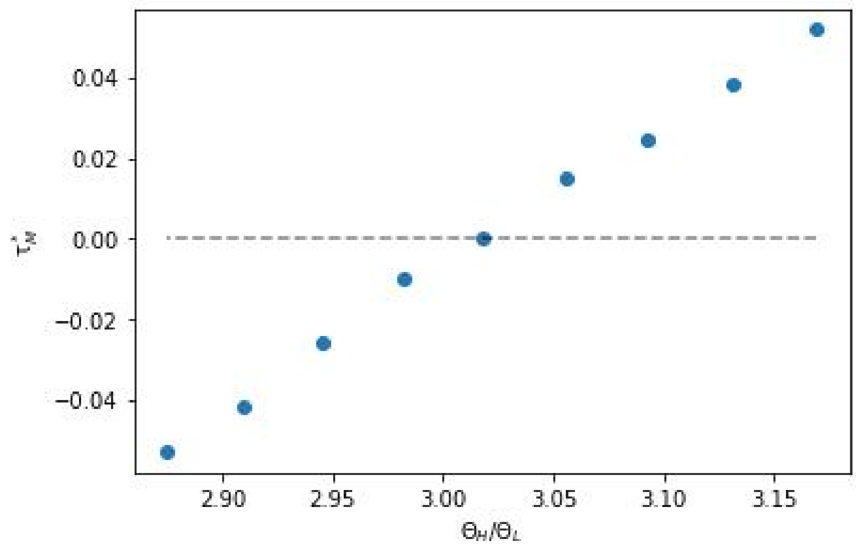

Figure 1. The optimal monetary tax $\tau_M^*(s^t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$

We numerically solve for optimal fiscal and monetary policy given equal (egalitarian) Pareto weights ($\lambda^H = \lambda^L = 1$). Figure 1 plots the implied optimal monetary tax for different values of $\theta^H(s_t)/\theta^L(s_t)$. As this ratio increases, the optimal monetary tax increases.

In this simple example, the weight $\lambda^i/\varphi^i + \nu^i(1+\eta)$ of the high-productivity types is 3.486, while the weight of the low-productivity type is -5.09. As a result, our sufficient statistic for inequality, $\mathcal{I}(s_t)$ is increasing in $\theta^H(s_t)/\theta^L(s_t)$. Figure 2 plots the relationship between $\mathcal{I}(s_t)$ and $\theta^H(s_t)/\theta^L(s_t)$.

5 Implementation

We now turn to implementation and show how the optimum characterized in Theorem 2 can be implemented with the available fiscal and monetary policy instruments.

Fiscal policy. The optimal fiscal wedge is χ^*. Clearly there is no unique implementation of this wedge, and any implementation of χ^* results in the same behavior for optimal monetary policy. For the sake of exposition, in this section we set the firm sales tax such that it directly neutralizes the monopolistic markup and the labor income and consumption taxes such that they implement the appropriate fiscal wedge. Specifically:

$$(1-\tau_r)\left(\frac{\rho-1}{\rho}\right) = 1, \qquad \text{and} \qquad \frac{1-\tau_\ell}{1+\tau_c} = \chi^*. \tag{43}$$

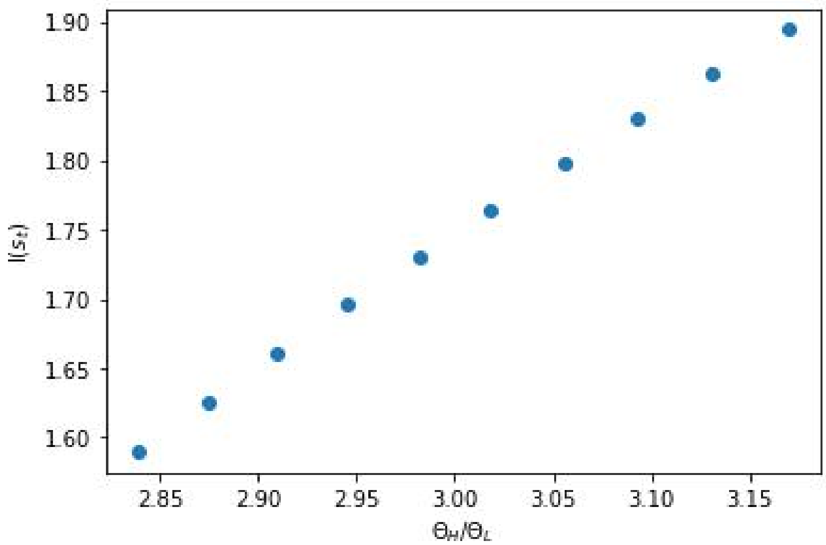

Figure 2. The inequality statistic $\mathcal{I}(s_t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$

Monetary Policy. We define the aggregate markup $\mathcal{M}(s^t)$ in the economy as the nominal price level divided by the nominal marginal cost; in logs:

$$\log \mathcal{M}(s^t) \equiv \log P(s^t) - \log(W(s^t)/A(s^t)). \tag{44}$$

We can now express optimal monetary policy in terms of this target.

Proposition 4. *It is optimal for monetary policy to target an aggregate mark-up that satisfies:*

$$\begin{array}{lll} \log \mathcal{M}(s^t) > 0 & \textit{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1}), \\ \log \mathcal{M}(s^t) = 0 & \textit{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}), \\ \log \mathcal{M}(s^t) < 0 & \textit{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1}). \end{array}$$

Proof. See Appendix A.11. □

Proposition 4 is essentially a restatement of Theorem 2 in terms of nominal targets instead of wedges. The nominal target is the markup of the aggregate price level over the marginal cost: when households pay a higher price for the final good than the cost to produce it, it is as if they pay an "inflation tax." Were we to shut down TFP shocks and set $A(s^t) = 1$ in all states, then the markup would simply be the price level over the nominal wage, i.e. the inverse of the real wage.

Recall that when the labor income distribution exhibits only proportional aggregate shocks, it is optimal for monetary policy to implement flexible-price allocations. This possibility is nested in Proposition 4 as the case in which $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$ in all states, and monetary policy targets a constant mark-up: $\log \mathcal{M}(s^t) = 0$. Here, the level of zero for the log markup under flexible-prices is arbitrary: it is only zero because we have set the sales tax in (43) such that it cancels out the monopolistic markup. Had we not made that choice, the markup under flexible prices would be a non-zero constant.[14]

Away from this special case, when instead shocks affect the relative labor skill distribution, the tax system is insufficient to implement the optimal level of redistribution. It is then optimal for monetary policy to deviate from implementing flexible price allocations and target a state-contingent markup. Proposition 4 tells us that the optimal state-contingent markup covaries positively with $\mathcal{I}(s_t)$, our measure of inequality. When inequality is greater than $\bar{\mathcal{I}}$, the optimal mark-up is positive, meaning that the aggregate price level should rise above the nominal marginal cost of production.

The markup, or inflation tax, works much like a standard, fiscal tax rate. Recall that a higher tax rate in this environment is more redistributive: although all households face the same tax *rate*, high-skilled households pay more taxes (in levels) than low-skilled households. The same is true with the inflation tax. Although all households face the same markup of prices over marginal costs, high-skilled households buy more goods and pay more of the "inflation tax" than low-skilled households.

Note, however, that the manner by which the proceeds of the inflation tax are collected and distributed back to households differs from how tax revenue is collected and distributed. With standard fiscal instruments, tax revenue is collected by the government and redistributed to households via uniform, lump-sum transfers. In contrast, when the price level rises above marginal cost, firms make positive profits; these profits, in turn, are distributed equally across households as households have uniform equity shares. It follows that a higher inflation tax is more redistributive.

Therefore, proceeds from the inflation tax, firm profits, are isomorphic to lump-sum transfers in our model.[15] As we have mentioned previously, this equivalence relies on the homogeneous equity shares assumption—an assumption that one might argue is unrealistic. We agree, and have only assumed this as a benchmark for our analysis. In the following section, Section 6, we relax this assumption and study an extension of our model in which households own unequal shares of the firms.

We next provide a characterization of optimal monetary policy in terms of the nominal interest rate.

Proposition 5. *Directly following an arbitrary history, s^{t-1}, take three states s_ℓ, s_m, s_h, such that $\mathcal{I}(s_\ell) < \mathcal{I}(s_m) = \bar{\mathcal{I}}(s^{t-1}) < \mathcal{I}(s_h)$, and $A(s_\ell) = A(s_m) = A(s_h)$. If shocks are i.i.d., then*

$$1 + i(s_\ell) > 1 + i(s_m) > 1 + i(s_h)$$

[14] More generally, under flexible prices: $\mathcal{M}(s^t) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}$ for all $s^t \in S^t$.

[15] This can be seen directly in the budget implementability conditions (18), in particular our definition of $\bar{T}$ in (19).

Proof. See Appendix A.12. □

Therefore the nominal interest rate should fall in states in which $\mathcal{I}(s_t)$ is high.

6 Heterogeneous Equity Shares

In this section we relax the assumption that households own equal shares of the intermediate good firms. We let $1+\sigma^i$ denote the share of equity held by a household of type $i \in I$, with $\sum_{i\in I}\pi^i\sigma^i = 0$, and assume that equity shares are fixed. The household's budget constraint in nominal terms is then given by:

$$(1+\tau_c)P(s^t)c^i(s^t) + b^i(s^t) - (1+i(s^{t-1}))b^i(s^{t-1}) + \sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)z^i(s^{t+1}|s^t) \leq \\ (1-\tau_\ell)W(s^t)\ell^i(s^t) + (1-\tau_\Pi)(1+\sigma^i)\Pi(s^t) + P(s^t)T(s^t) + z^i(s^t|s^{t-1}) \tag{45}$$

For now, we impose no restrictions on the cross-sectional covariance between labor skill type and equity, but we will discuss in detail the implications of this covariance for optimal policy.

6.1 Equilibrium Characterization

We begin by characterizing the set of equilibrium allocations in this economy. Income from profits is exogenous from the point of view of households, implying that the household's intratemporal and intertemporal optimality conditions in (13)-(15) continue to hold (both at the individual and the representative household level). The flexible-price and sticky-price firms' optimal pricing equations, (21) and (22), respectively, remain similarly unchanged.

What must be adjusted, however, are the implementability conditions corresponding to the households' budget constraints. The lifetime budget constraint of type i is now equivalent to the following condition:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi,s_t)L(s^t) - \sigma^i U_C^m(s^t)\frac{1-\tau_\Pi}{1+\tau_c}\frac{\Pi(s^t)}{P(s^t)} \right] \leq U_C^m(s_0)\bar{T}, \tag{46}$$

where $\bar{T}$ is as in (19) and $\Pi(s^t)/P(s^t)$ are real profits in history s^t.[16]

The only difference between this condition and the corresponding implementability condition in our baseline model is the third term on the left-hand side of the equation which includes after-tax real profits. The emergence of this term is a direct result of the heterogeneity in equity shares; note that it disappears when $\sigma^i = 0$. Using this condition, we provide the following

[16] We derive this in the usual way by substituting in equilibrium prices into the household's budget constraint. See Appendix B.1 for the specific derivation.

proposition which characterizes the entire set of allocations that can be supported as a sticky-price equilibrium.

Proposition 6. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a sticky-price equilibrium if and only if there exist market weights $\varphi \equiv (\varphi^i)$, a scalar $\bar{T} \in \mathbb{R}$, a strictly positive scalar $\chi \in \mathbb{R}_+$, a weakly positive scalar δ, and a positively-valued function $\epsilon : S^t \to \mathbb{R}_+$, such that parts (i) and (ii) of Proposition 2 are satisfied and the following holds for all $i \in I$:*

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t) \omega_C^i(\varphi) C(s^t) + U_L^m(s^t) \omega_L^i(\varphi, s_t) L(s^t) + \sigma^i \delta U_L^m(s^t) L(s^t) \Phi(\epsilon(s^t)) \right] \leq U_C^m(s_0) \bar{T}, \tag{47}$$

where the function $\Phi : \mathbb{R}_+ \to \mathbb{R}$ is defined by:

$$\Phi(\epsilon) \equiv \frac{(1-\kappa)\frac{1}{\rho-1}\epsilon^\rho + \kappa\left(\frac{\rho}{\rho-1}\epsilon - 1\right)}{\kappa + (1-\kappa)\epsilon^\rho}. \tag{48}$$

Proof. See Appendix B.2. □

Proposition 6 is the heterogeneous equity shares economy analog of Proposition 2 in our baseline model. Again, the only difference between the two propositions is the replacement of the baseline economy's implementability conditions in (18) with the heterogeneous equity share implementability conditions in (47).

In order to derive the latter, we obtain an expression for real profits in terms of the allocation and the monetary wedge function, $\epsilon(s^t)$, alone. In particular, in the appendix we show how real profits can be written as a function $\Phi(\epsilon(s^t))$ where $\Phi(\cdot)$ is convex and reaches a minimum at some $\epsilon(s^t) < 1$; see Appendix B.3. Furthermore, the third term on the left-hand side of (47) includes a weakly-positive scalar denoted by δ. This scalar is given by $\delta \equiv \frac{1-\tau_\Pi}{1-\tau_\ell}$; it is the product of the profit tax and the labor income tax.

6.2 The Ramsey Optimum and Optimal Monetary Wedge

We turn now to the Ramsey problem in the heterogeneous equity share economy. We first present the following trivial result.

Proposition 7. *If profits are taxed fully, $\tau_\Pi = 1$, then the optimal monetary tax $\tau_M^*(s^t)$ in the heterogeneous equity shares economy is the same as in the baseline homogeneous shares economy.*

Proof. If profits are taxed fully, $\delta = 0$. In this case (47) collapses to (18). □

When profits are taxed fully, it is irrelevant whether equity shares are homogeneous or heterogeneous; in either case the implementability conditions for the household budget constraints are the same. As a result, Theorem 2 holds. Therefore, in order to make the analysis in this extension more interesting, we make the following ad hoc assumption on δ.

Assumption 2. *There exists a strictly-positive, lower bound* $\bar{\delta} > 0$ *such that* $\delta \geq \bar{\delta}$.

For the remainder of our analysis, we assume the following: $\delta > 0$.

The Ramsey planner has the same utilitarian social welfare function (32) as before. We define the Ramsey problem in the following way. Provided that $\delta > 0$, we define a new pseudo-welfare function $\mathcal{W}^{\sigma}(\cdot)$ as follows:

$$\mathcal{W}^{\sigma}(C(s^t), L(s^t), \epsilon(s^t), s_t; \varphi, \nu, \lambda, \sigma) \equiv \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) + \delta U_L^m(s^t) L(s^t) \Phi(\epsilon(s^t)) \sum_{i \in I} \pi^i \nu^i \sigma^i$$

where $\mathcal{W}$ is as in (33). With the new pseudo-welfare function so defined, we can recast the Ramsey planning problem as follows.

Ramsey Planner's Problem. *The Ramsey planner chooses an aggregate allocation,* $(C(s^t), Y(s^t), L(s^t))_{s^t \in S^t}$, *market weights* $\varphi \equiv (\varphi^i)$, *constants* $\bar{T} \in \mathbb{R}$ *and* $\chi \in \mathbb{R}_+$, *and* $\epsilon : S^t \to \mathbb{R}_+$, *in order to maximize the pseudo-utility function:*

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}^{\sigma}(C(s^t), L(s^t), \epsilon(s^t), s_t; \varphi, \nu, \lambda, \sigma) - U_C^m(s_0) \bar{T} \sum_{i \in I} \pi^i \nu^i \tag{49}$$

subject to (30), (31), *and the aggregate resource constraint* (38).

Appendix B.4 provides a characterization of the Ramsey optimum.[17] Again, we do not provide this characterization here for expositional purposes; we instead move directly to characterizing the implicit monetary tax wedge at the optimum. We do so in the following theorem.

Theorem 3. *Let* $\sum_{i \in I} \pi^i \nu^i \sigma^i \geq 0$. *There exists a threshold* $\bar{\mathcal{I}}^{\sigma}(s^{t-1}) > 0$, *measurable in history* s^{t-1}, *such that optimal implicit monetary tax* $\tau_M^*(s^t)$ *satisfies:*

$$\begin{array}{lll} \tau_M^*(s^t) > 0 & \text{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}^{\sigma}(s^{t-1}), \\ \tau_M^*(s^t) = 0 & \text{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}^{\sigma}(s^{t-1}), \\ \tau_M^*(s^t) < 0 & \text{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}^{\sigma}(s^{t-1}). \end{array}$$

where $\bar{\mathcal{I}}^{\sigma}(s^{t-1}) \equiv \bar{\mathcal{I}}(s^{t-1}) - \delta \frac{1}{\rho - 1} [(1+\eta) - (\gamma + \eta)\rho] \mathcal{C}$ *and* $\mathcal{C} = \sum_{i \in I} \pi^i \nu^i \sigma^i$.

Proof. See Appendix B.5. □

6.3 Numerical Illustration

To be added.

[17] See Proposition 9 in Appendix B.4.

7 Conclusion

In this paper we study optimal monetary policy in a dynamic, general equilibrium economy with heterogeneous agents, complete markets, and a motive for redistribution. We find that when preferences are iso-elastic and there are no shocks to the relative skill distribution, all redistribution is done via the tax system. In this case it is optimal for monetary policy to implement flexible-price allocations. It does so by targeting a constant mark-up in response to TFP, government spending, and aggregate skill shocks.

On the other hand, when there are shocks to the relative skill distribution, the available tax instruments are insufficient. In this case it is optimal for monetary policy to deviate from implementing flexible-price allocations by instead targeting a state-contingent markup. We find that the optimal markup co-varies positively with a sufficient statistic for labor income inequality.

There are many interesting channels that we have abstracted from in this paper. First, although we allow for differences in labor productivity, there is perfect reallocation of efficiency units of labor across firms in our economy. Furthermore, labor productivity is modeled as a type-specific, stochastic process that is contingent on the aggregate state, but this process is exogenous—we do not allow for endogeneity of labor productivity to the monetary policy. Finally, several papers document the distributional effects of monetary policy shocks; see e.g. Doepke and Schneider (2006); Coibion, Gorodnichenko, Kueng and Silvia (2017); Auclert (2019). We abstract from the heterogeneous effects of monetary shocks, and focus solely on how an inflation tax that affects all households uniformly can be useful for redistribution. We leave exploration of these channels open for future work.

References

Acharya, S., Challe, E., Dogra, K., 2020. Optimal monetary policy according to hank .

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

Angeletos, G.M., La'O, J., 2020. Optimal monetary policy with informational frictions. Journal of Political Economy 128, 1027–1064.

Auclert, A., 2019. Monetary policy and the redistribution channel. American Economic Review 109, 2333–67.

Auclert, A., Rognlie, M., Straub, L., 2018. The intertemporal keynesian cross. NBER Working Paper 25020 .

Benigno, P., Woodford, M., 2003. Optimal monetary and fiscal policy: A linear-quadratic approach. NBER Macroeconomics Annual 18, 271–333.

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

Chari, V., Christiano, L., Kehoe, P., 1994. Optimal fiscal policy in a business cycle model. Journal of Political Economy 102, 617–52.

Chari, V., Christiano, L.J., Kehoe, P.J., 1991. Optimal fiscal and monetary policy: Some recent results. Journal of Money, Credit and Banking 23, 519–539.

Chari, V.V., Kehoe, P.J., 1999. Optimal fiscal and monetary policy, in: Handbook of Macroeconomics. Elsevier. volume 1, pp. 1671–1745.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2007. Business cycle accounting. Econometrica 75, 781–836.

Coibion, O., Gorodnichenko, Y., Kueng, L., Silvia, J., 2017. Innocent bystanders? monetary policy and inequality. Journal of Monetary Economics 88, 70–89.

Correia, I., 2010. Consumption taxes and redistribution. American Economic Review 100, 1673–94.

Correia, I., Farhi, E., Nicolini, J.P., Teles, P., 2013. Unconventional fiscal policy at the zero bound. American Economic Review 103, 1172–1211.

Correia, I., Nicolini, J.P., Teles, P., 2008. Optimal fiscal and monetary policy: Equivalence results. Journal of Political Economy 116, 141–170.

Dávila, E., Schaab, A., 2022. Optimal monetary policy with heterogeneous agents: A timeless ramsey approach. Available at SSRN 4102028.

Diamond, P., Mirrlees, J., 1971. Optimal taxation and public production: I–production efficiency. American Economic Review 61, 8–27.

Doepke, M., Schneider, M., 2006. Inflation and the redistribution of nominal wealth. Journal of Political Economy 114, 1069–1097.

Galí, J., 2008. Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework. Princeton University Press.

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

Guvenen, F., Smith, A.A., 2014. Inferring labor income risk and partial insurance from economic choices. Econometrica 82, 2085–2129.

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

Lucas, R.J., Stokey, N.L., 1983. Optimal fiscal and monetary policy in an economy without capital. Journal of Monetary Economics 12, 55–93.

Mackowiak, B., Wiederholt, M., 2009. Optimal sticky prices under rational inattention. American Economic Review 99, 769–803.

Mankiw, N.G., Reis, R., 2002. Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. Quarterly Journal of Economics 117, 1295–1328.

McKay, A., Wolf, C., 2022. Optimal policy rules in hank. Technical Report. Working Paper, FRB Minneapolis.

Parker, J.A., Vissing-Jorgensen, A., 2009. Who bears aggregate fluctuations and how? American Economic Review 99, 399–405.

Patterson, C., Forthcoming. The matching multiplier and the amplification of recessions. American Economic Review .

Schmitt-Grohe, S., Uribe, M., 2004. Optimal fiscal and monetary policy under sticky prices. Journal of Economic Theory 114, 198–230.

Schulhofer-Wohl, S., 2011. Heterogeneity and tests of risk sharing. Journal of Political Economy 119, 925–958.

Werning, I., 2007. Optimal fiscal policy with redistribution. The Quarterly Journal of Economics 122, 925–967.

Woodford, M., 2003a. Imperfect common knowledge and the effects of monetary policy. Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps .

Woodford, M., 2003b. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.

A Appendix: Proofs

A.1 Household optimality

In this section of the appendix, we derive the optimality conditions for household i. We let $\beta^t\mu(s^t)\Lambda^i(s^t)$ denote the Lagrange multiplier on household i's budget set at time t, history s^t.

The first-order conditions for household i with respect to consumption and labor are given by, respectively:

$$\mu(s^t)U_c^i(s^t) - \mu(s^t)\Lambda^i(s^t)(1+\tau_c)P(s^t) = 0, \tag{50}$$

$$\mu(s^t)\frac{1}{\theta^i(s_t)}U_\ell^i(s^t) + \mu(s^t)\Lambda^i(s^t)(1-\tau_\ell)W(s^t) = 0, \tag{51}$$

where $U_c^i(s^t) \equiv \partial U(\cdot)/\partial c^i(s^t)$ and $U_\ell^i(s^t) \equiv \partial U(\cdot)/\partial h^i(s^t)$ denote the marginal utilities of the household of type i with respect to individual consumption and work effort. The first-order condition with respect to nominal bonds $b^i(s^t)$ is given by:

$$-\beta^t\mu(s^t)\Lambda^i(s^t) + \beta^{t+1}\sum_{s^{t+1}|s^t}\mu(s^{t+1})\Lambda^i(s^{t+1})(1+i(s^t)) = 0. \tag{52}$$

The first-order condition with respect to Arrow security $z^i(s^{t+1})$ is given by:

$$-\beta^t\mu(s^t)\Lambda^i(s^t)Q(s^{t+1}|s^t) + \beta^{t+1}\mu(s^{t+1})\Lambda^i(s^{t+1}) = 0. \tag{53}$$

The household's transversality conditions for nominal bonds and Arrow securities are given by:

$$\lim_{t\to\infty}\beta^t\mu(s^t)\Lambda^i(s^t)b^i(s^t) = 0 \qquad \text{and} \qquad \lim_{t\to\infty}\beta^t\mu(s^t)\Lambda^i(s^t)Q(s^{t+1}|s^t)z^i(s^{t+1}) = 0$$

Combining (50) and (51), we obtain the household's intratemporal condition:

$$-\frac{1}{\theta^i(s_t)}\frac{U_\ell^i(s^t)}{U_c^i(s^t)} = \frac{(1-\tau_\ell)W(s^t)}{(1+\tau_c)P(s^t)} \tag{54}$$

Using the fact that $U_c^i(s^t) = \Lambda(s^t)(1+\tau_c)P(s^t)$, we may rewrite the Euler equation for bonds as

$$\frac{U_c^i(s^t)}{P(s^t)} = \beta(1+i(s^t))\sum_{s^{t+1}|s^t}\mu(s^{t+1}|s^t)\frac{U_c^i(s^{t+1})}{P(s^{t+1})}, \tag{55}$$

Finally, the Arrow security price must satisfy

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U_c^i(s^{t+1})}{P(s^{t+1})}\frac{P(s^t)}{U_c^i(s^t)} \tag{56}$$

where $\mu(s^{t+1}|s^t) \equiv \mu(s^{t+1})/\mu(s^t)$ is the probability of s^{t+1} conditional on s^t.

A.2 Proof of Lemma 1

Condition (54) of the household's problem implies that in any equilibrium, the following condition must hold:

$$-\frac{1}{\theta^i(s_t)}\frac{U^i_\ell(s^t)}{U^i_c(s^t)} = \frac{(1-\tau_\ell)W(s^t)}{(1+\tau_c)P(s^t)} = -\frac{1}{\theta^k(s_t)}\frac{U^k_\ell(s^t)}{U^k_c(s^t)}$$

for all types $i,k \in I$. Consider now the static subproblem described in Lemma 1. Let $\rho_C(s^t)$ and $\rho_L(s^t)$ be the Lagrange multipliers on the constraints in (12). The first-order conditions of this subproblem are given by

$$\varphi^i U^i_c(s^t) - \rho_C(s^t) = 0, \qquad \forall i \in I \tag{57}$$

$$\varphi^i \frac{1}{\theta^i(s_t)} U^i_\ell(s^t) + \rho_L(s^t) = 0, \qquad \forall i \in I \tag{58}$$

Therefore

$$-\frac{1}{\theta^i(s_t)}\frac{U^i_\ell(s^t)}{U^i_c(s^t)} = \frac{\rho_L(s^t)}{\rho_C(s^t)} = -\frac{1}{\theta^k(s_t)}\frac{U^k_\ell(s^t)}{U^k_c(s^t)}$$

for all types $i,k \in I$. It follows that the solution to the sub-problem coincides with the equilibrium allocation. Finally, the envelope conditions for this static sub-problem are given by:

$$U^m_C(s^t) = \varphi^i U^i_c(\mathcal{C}^i(C(s^t), L(s^t); \varphi)),$$
$$U^m_L(s^t) = \varphi^i \frac{1}{\theta^i(s_t)} U^i_\ell(\mathcal{L}^i(C(s^t), L(s^t); \varphi)),$$

for all $i \in I$. Next, with the separable and isoelatic preferences assumed in (1), the FOCs in (57)-(58) can be written as

$$\varphi^i c^i(s^t)^{-\gamma} - \rho_C(s^t) = 0$$
$$\varphi^i \frac{1}{\theta^i(s_t)} \left[\frac{\ell^i(s^t)}{\theta^i(s_t)}\right]^\eta + \rho_L(s^t) = 0$$

Combining these conditions with the resource constraints in (12), we obtain the linear expressions in (16) for individual consumption and labor with shares given by (17).

A.3 Derivation of Budget Implementability Conditions

We derive condition (18). We take the household's budget constraint in (2) for type $i \in I$, multiply both sides by $\Lambda^i(s^t)$, and use the household's FOCs in (50) and (51) to substitute out consumption and labor prices. Doing so, we obtain:

$$U^i_c(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)}U^i_\ell(s^t)\ell^i(s^t) = \Lambda^i(s^t)z^i(s^t|s^{t-1}) - \Lambda^i(s^t)\sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)z^i(s^{t+1}|s^t) - \Lambda^i(s^t)b^i(s^t)$$
$$+ \Lambda^i(s^t)(1+i(s^{t-1}))b^i(s^{t-1}) + \Lambda^i(s^t)P(s^t)\bar{T}(s^t)$$

where we let

$$\bar{T}(s^t) \equiv T(s^t) + \frac{1}{P(s^t)}(1-\tau_\Pi)\Pi(s^t). \tag{59}$$

Multiplying both sides by $\beta^t\mu(s^t)$, summing over t and s^t, and using the household's intertemporal optimality conditions (55)-(54) to cancel terms, we obtain:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t)\left[U_c^i(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)}U_\ell^i(s^t)\ell^i(s^t)\right] \leq U_c^i(s_0)\bar{T},$$

where

$$\bar{T} \equiv \frac{1}{\Lambda^i(s_0)(1+\tau_c)P(s_0)}\sum_t\sum_{s^t}\beta^t\mu(s^t)\Lambda^i(s^t)P(s^t)\bar{T}(s^t).$$

Therefore

$$\bar{T} \equiv \frac{1}{U_c^i(s_0)(1+\tau_c)}\sum_t\sum_{s^t}\beta^t\mu(s^t)U_c^i(s^t)\bar{T}(s^t)$$

for all $i \in I$. Finally, using the solution and the envelope conditions for the static sub-problem described in Lemma 1, as well as the fact that individual allocations satisfy (16), we can rewrite the above conditions as:

$$\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi, s_t)L(s^t)\right] \leq U_C^m(s_0)\bar{T}$$

where

$$\bar{T} \equiv \frac{(1+\tau_c)^{-1}}{U_c^m(s_0)}\sum_t\sum_{s^t}\beta^t\mu(s^t)U_C^m(s^t)\left[T(s^t) + (1-\tau_\Pi)\frac{\Pi(s^t)}{P(s^t)}\right]$$

for all $i \in I$, as was to be shown.

A.4 Proof of Proposition 1

Necessity. Condition (25) follows from combining the aggregate price in (24) with the household's intratemporal optimality condition (13), and letting χ denote the labor wedge as follows:

$$\chi \equiv \left(\frac{\rho-1}{\rho}\right)\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c}. \tag{60}$$

Next, condition (21) implies that all firms set the same nominal price. The demand functions in (4) then imply that all firms produce the same level of output, proving necessity of $y^j(s^t) = Y(s^t)$ for all $j \in \mathcal{J}$. Finally, the derivation of the set of necessary conditions in (18) is provided in Appendix A.3.

Sufficiency. Take any feasible allocation $x \in \mathcal{X}$, vector $\varphi \equiv (\varphi^i)$, and constants $\bar{T} \in \mathbb{R}$ and $\chi \in \mathbb{R}_+$ that satisfy conditions (i)-(iii) of Proposition 1. We show that there exists a price system ϱ, a policy Ω, and asset holdings ζ, that support x as a flexible-price equilibrium; we construct these objects as follows.

First, we set intermediate-good prices according to:

$$p_t^j(s^t) = p_t^f(s^t) = P(s^t), \qquad \forall j \in \mathcal{J}$$

where we normalize the aggregate price level to one: $P(s^t) = 1$ for all s^t. These prices, combined with $y^j(s^t) = Y(s^t)$, ensure that the CES demand function (4) is satified for all $j \in \mathcal{J}$.

Second, we set the tax rates $(\tau_\ell, \tau_c, \tau_r)$ such that they jointly satisfy:

$$\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c} = \left(\frac{\rho-1}{\rho}\right)^{-1} \chi. \tag{61}$$

For any strictly positive χ and $\rho > 1$, such tax rates exist. Combining this with condition (25), we obtain the following:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{1+\tau_c}{1-\tau_\ell}\right) = \left(\frac{\rho-1}{\rho}\right)(1-\tau_r)A(s_t). \tag{62}$$

Given tax rates $(\tau_\ell, \tau_c, \tau_r)$, we set the real wage $W(s^t)$ as follows:

$$W(s^t) = -\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{1+\tau_c}{1-\tau_\ell}\right), \tag{63}$$

and therefore satisfy the household's intratemporal condition in (13). Substituting the above expression for the real wage into (62) and re-arranging gives us:

$$1 = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}. \tag{64}$$

Therefore the flexible-price firm's optimality condition (21) is satisfied.

Next, we set Arrow prices and the nominal interest rate as follows:

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)} \qquad \text{and} \qquad 1 = \beta(1+i(s^t))\sum_{s^{t+1}|s^t}\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)},$$

and therefore satisfy the household's intertemporal conditions in (14) and (14).

What remains to be shown is that we may construct bond holdings such that the household's budget constraints are satisfied at this allocation in every history. To do so, we first choose any sequence $\bar{T}(s^t)$ that satifies the following condition:

$$\bar{T} = \frac{1}{U_c^m(s_0)(1+\tau_c)}\sum_t\sum_{s^t}\beta^t\mu(s^t)U_c^m(s^t)\bar{T}(s^t).$$

Next we take the household's budget constraint in (2) for type $i \in I$. Multiplying both sides by $\beta^t\mu(s^t)\Lambda^i(s^t)$ and summing over all periods and states following period r, history s^r, we get:

$$\sum_{t=r+1}^{\infty}\sum_{s^t}\beta^t\mu(s^t)\Lambda^i(s^t)\left[(1+\tau_c)c^i(s^t) + b^i(s^t) + \sum_{s^{t+1}|s^t}Q(s^{t+1}|s^t)z^i(s^{t+1}|s^t)\right]$$
$$= \sum_{t=r+1}^{\infty}\sum_{s^t}\beta^t\mu(s^t)\Lambda^i(s^t)\left[(1-\tau_\ell)W(s^t)\ell^i(s^t) + \bar{T}(s^t) + (1+i(s^{t-1}))b^i(s^{t-1}) + z^i(s^t|s^{t-1})\right]$$

where we let $T(s^t) + (1-\tau_\Pi)\Pi(s^t) = \bar{T}(s^t)$. Substituting in the household's FOCs for bonds (52) and Arrow securities (53) we get:

$$\sum_{t=r+1}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \Lambda^i(s^t) \left[(1+\tau_c) c^i(s^t)\right] = \sum_{t=r+1}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \Lambda^i(s^t) \left[(1-\tau_\ell) W(s^t) \ell^i(s^t) + \bar{T}(s^t)\right] + \sum_{s^{r+1}|s^r} \beta^{r+1} \mu(s^{r+1}) \Lambda^i(s^{r+1})(1+i(s^r)) b^i(s^r)$$

Rearranging gives us:

$$\beta^r \mu(s^r) \Lambda^i(s^r) b^i(s^r) = \sum_{t=r+1}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \Lambda^i(s^t) \left[(1+\tau_c) c^i(s^t) - (1-\tau_\ell) W(s^t) \ell^i(s^t) - \bar{T}(s^t)\right]$$

Next, using the household's FOCs for consumption and labor (50) and (51), we obtain:

$$\frac{\beta^r \mu(s^r) U_c^i(s^r)}{1+\tau_c} b^i(s^r) = \sum_{t=r+1}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \left[U_c^i(s^t) c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t) \ell^i(s^t) - \frac{U_c^i(s^t)}{(1+\tau_c)} \bar{T}(s^t)\right]$$

which we may rewrite as follows

$$\frac{U_c^i(s^r)}{1+\tau_c} b^i(s^r) = \sum_{t=r+1}^{\infty} \sum_{s^t} \beta^{t-r} \mu(s^t|s^r) \left[U_c^i(s^t) c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t) \ell^i(s^t) - \frac{U_c^i(s^t)}{(1+\tau_c)} \bar{T}(s^t)\right]$$

Therefore real bond holdings of household i are given by

$$b^i(s^r) = \left(\frac{U_c^i(s^r)}{1+\tau_c}\right)^{-1} \left\{\sum_{t=r+1}^{\infty} \sum_{s^t} \beta^{t-r} \mu(s^t|s^r) \left[U_c^i(s^t) c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t) \ell^i(s^t) - \frac{U_c^i(s^t)}{(1+\tau_c)} \bar{T}(s^t)\right]\right\}$$

for any period r, history s^r.

A.5 Proof of Proposition 2

Necessity. The sticky-price firm's optimality condition is given by

$$\sum_{s^t} Q(s^t|s^{t-1}) y^s(s^t) \left[\frac{W(s^t)}{A(s_t)} - (1-\tau_r)\left(\frac{\rho-1}{\rho}\right) p_t^s(s^{t-1})\right] = 0.$$

Rearranging the above equation allows us to write the optimal price as:

$$p_t^s(s^{t-1}) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1}) \tag{65}$$

with $q(s^t|s^{t-1})$ defined in (23). We define a random variable $\epsilon(s^t)$ in (28) as the forecast error of the sticky price firm. We can then rewrite the optimal price of the sticky-price firm as we have in (27).

Next, condition (21) indicates that all flexible-price firms set the same nominal price, and similarly condition (27) indicates that all sticky-price firms set the same nominal price. Combining these with the demand function in (4) implies that all flexible-price firms produce the same level of output and all sticky-price firms produce the same level of output, denoted by $y^f(s^t)$ and $y^s(s^t)$, respectively. The demand function (4) further implies that the ratio of $y^s(s^t)$ to $y^f(s^t)$ satisfies:

$$\frac{y^s(s^t)}{y^f(s^t)} = \left(\frac{p_t^s(s^{t-1})}{p_t^f(s^t)}\right)^{-\rho}.$$

Combining this condition with our expressions for the optimal prices in (21) and (22) provides necessity of condition (29).

Next, aggregating over the optimal prices in (21) and (27) according to (9) we obtain the following expression for the aggregate price level:

$$P(s^t) = \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}, \qquad \forall s^t \in S^t. \tag{66}$$

Necessary condition (30) follows from combining this expression for the aggregate price level with the household's intratemporal optimality condition (13) with χ defined in (60).

Moreover, it is necessary that the forecast error $\epsilon(s^t)$ satisfies (28). Rewriting the latter, we have that

$$\epsilon(s^t)\frac{W(s^t)}{A(s_t)} = \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1}), \qquad \forall s^t|s^{t-1}. \tag{67}$$

We define a positively-valued function $h : S^{t-1} \to \mathbb{R}_+$ as the (risk-adjusted) expected nominal marginal cost given information set s^{t-1}:

$$h(s^{t-1}) \equiv \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1}). \tag{68}$$

Then condition (67) implies:

$$\epsilon(s^t)\frac{W(s^t)}{A(s_t)} = h(s^{t-1}), \qquad \forall s^t|s^{t-1} \tag{69}$$

Solving this for $W(s^t)/A(s^t)$ and substituting back into the definition for $h(s^{t-1})$ in (68), gives us:

$$h(s^{t-1}) = \sum_{s^t|s^{t-1}} \frac{h(s^{t-1})}{\epsilon(s^t)} q(s^t|s^{t-1})$$

which reduces to:

$$1 = \sum_{s^t|s^{t-1}} \epsilon(s^t)^{-1} q(s^t|s^{t-1}). \tag{70}$$

This condition is a necessary condition—it must hold in any sticky-price equilibrium.

Next, we write the risk-adjusted probabilities $q(s^t|s^{t-1})$ as functions of the allocation and forecast errors alone. To do so, we substitute our characterization of the equilibrium Arrow prices $Q(s^t|s^{t-1})$ from (14) into our definition for $q(s^t|s^{t-1})$ in (23); doing so yields:

$$q(s^t|s^{t-1}) = \frac{\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)P(s^t)^{-1}}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)P(s^t)^{-1}}. \tag{71}$$

Combining our characterization of the aggregate price level in (66) with (69), we obtain the following expression for the price level:

$$P(s^t) = \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{h(s^{t-1})}{\epsilon(s^t)},$$

and substituting this into (71) gives us:

$$q(s^t|s^{t-1}) \equiv \frac{\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}.$$

Next, we substitute the above expression for $q(s^t|s^{t-1})$ into (70) and obtain the following equilibrium necessary condition:

$$1 = \frac{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}.$$

Rearranging, we can write this as follows:

$$\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}\left(\epsilon(s^t) - 1\right) = 0 \tag{72}$$

Next, we use the fact that aggregate output satisfies:

$$Y(s^t) = \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} \tag{73}$$

Having already proven that $y^f(s^t) = y^s(s^t)\epsilon(s^t)^{\rho}$, we combine this with (73), and obtain the following equilibrium relationship between $y^s(s^t)$ and $Y(s^t)$:

$$y^s(s^t) = Y(s^t)\epsilon(s^t)^{-\rho}\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{\frac{\rho}{1-\rho}}$$

Substituting this into (72), yields the following equation:

$$\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)Y(s^t)\epsilon(s^t)^{-\rho}\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-1}\left(\epsilon(s^t) - 1\right) = 0,$$

which reduces to:

$$\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)Y(s^t)\left\{\frac{\epsilon(s^t) - 1}{\kappa\epsilon(s^t) + (1-\kappa)\epsilon(s^t)^{\rho}}\right\} = 0.$$

Condition (31) is therefore a necessary equilibrium condition—it must hold in any sticky-price equilibrium—and follows directly from our definition of $\epsilon(s^t)$ in (28). Finally, the derivation of the set of necessary conditions in (18) is provided in Appendix A.3.

Sufficiency. Take any feasible allocation $x \in \mathcal{X}$, vector $\varphi \equiv (\varphi^i)$, constants $\bar{T} \in \mathbb{R}$ and $\chi \in \mathbb{R}_+$, and a positively-valued function $\epsilon : S^t \to \mathbb{R}_+$ that satisfy conditions (i)-(iii) of Proposition 2. We show that there exists a price system ϱ, a policy Ω, and asset holdings ζ, that support x as a sticky-price equilibrium; we construct these as follows.

First, we set the tax rates $(\tau_\ell, \tau_c, \tau_r)$ such that they jointly satisfy (61). For any strictly positive χ and $\rho > 1$, such tax rates exist.

Next we set prices such that: $p_t^j(s^t) = p_t^f(s^t)$ for all $j \in \mathcal{J}^f$ and $p_t^j(s^t) = p_t^s(s^{t-1})$ for all $j \in \mathcal{J}^s$. That is, all flexible-price firms set the same price and all sticky-price firms set the same price. We define an arbitrary, positively-valued function $h : S^{t-1} \to \mathbb{R}_+$. Given tax rates, we set the sticky price and the nominal wage as follows:

$$p_t^s(s^{t-1}) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} h(s^{t-1}) \qquad \text{and} \qquad W(s^t) = \frac{h(s^{t-1})}{\epsilon(s^t)} A(s_t). \tag{74}$$

Furthermore, we set the flexible price such that the following condition holds:

$$p_t^f(s^t) = \epsilon(s^t)^{-1} p_t^s(s^{t-1}). \tag{75}$$

These prices, in conjunction with condition (29), ensure that the equilibrium demand function 4 is satified for all firms $j \in \mathcal{J}$.

Furthermore, note that (74) and (75) imply:

$$p_t^s(s^{t-1}) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \epsilon(s^t)\frac{W(s^t)}{A(s_t)} \qquad \text{and} \qquad p_t^f(s^t) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)}. \tag{76}$$

The latter indicates that the flexible-price firm's optimality condition in (21) is satisfied at these prices.

Next, aggregating over the prices constructed in (76) according to (9), we obtain the following expression for the aggregate price level:

$$P(s^t) = \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \frac{W(s^t)}{A(s_t)} \tag{77}$$

Combining this with the tax rates $(\tau_\ell, \tau_c, \tau_r)$ set according to (61) and rearranging we get:

$$\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t) = \left(\frac{1-\tau_\ell}{1+\tau_c}\right)\frac{W(s^t)}{P(s^t)}$$

Combining this with equation (30), we obtain the following condition:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \left(\frac{1-\tau_\ell}{1+\tau_c}\right)\frac{W(s^t)}{P(s^t)}. \tag{78}$$

Therefore, the household's intratemporal condition in (13) is satisfied at these prices.

Given the price level in (77), we set the state-contingent debt prices and the nominal interest rate as follows:

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)}\frac{P(s^t)}{P(s^{t+1})} \qquad \text{and} \qquad 1 = \beta(1+i(s^t))\sum_{s^{t+1}|s^t}\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)}. \tag{79}$$

We therefore satisfy the household's intertemporal conditions in (14) and (14). We set the money supply such that $M(s^t) = P(s^t)C(s^t)$ and therefore satisfy the cash-in-advance constraint.

We next prove that the price of the sticky-price firm constructed in (74) is optimal from the sticky-price firm's perspective. To do so we use equilibrium condition (31); note that this condition is equivalent to the following one:

$$\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}(\epsilon(s^t)-1) = 0.$$

Rearranging, we obtain:

$$1 = \frac{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}. \tag{80}$$

Next, using the Arrow prices constructed in (77), we have that the risk-adjusted probabilities defined in (23) can be written as follows:

$$q(s^t|s^{t-1}) = \frac{\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)P(s^t)^{-1}}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)P(s^t)^{-1}}. \tag{81}$$

Furthermore, the aggregate price level constructed in (77) can be written as:

$$P(s^t) = \left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{h(s^{t-1})}{\epsilon(s^t)}$$

where we have substituted in the nominal wage constructed in (74). Substituting this expression for the price level into (81) gives us the following expression for $q(s^t|s^{t-1})$:

$$q(s^t|s^{t-1}) \equiv \frac{\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}{\sum_{s^t|s^{t-1}}\mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}\epsilon(s^t)}. \tag{82}$$

Using (82), we can re-write condition (80) as follows:

$$1 = \sum_{s^t|s^{t-1}}\epsilon(s^t)^{-1}q(s^t|s^{t-1})$$

Substituting in the nominal wage constructed in (74), this becomes:

$$1 = \sum_{s^t|s^{t-1}}\frac{W(s^t)}{A(s_t)}h(s^{t-1})^{-1}q(s^t|s^{t-1}).$$

and multiplying both sides by $h(s^{t-1})$ gives us:

$$h(s^{t-1}) = \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1}).$$

That is, nominal wages have been constructed such that $h(s^{t-1})$ is equal to the (risk-adjusted) expected nominal marginal cost given information set s^{t-1}. Substituting this into our expression for the sticky-price in (74), we obtain the following:

$$p_t^s(s^{t-1}) = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1} \sum_{s^t|s^{t-1}} \frac{W(s^t)}{A(s_t)} q(s^t|s^{t-1})$$

Therefore the sticky-price firm's optimality condition (22) is satisfied.

Finally, what remains to be shown is that we may construct bond holdings such that the household's budget sets are satisfied at this allocation at every history. For this we follow the exact same steps above used to obtain equilibrium bond holdings in the sufficiency portion of the proof of Proposition 1. Following these steps, real bond holdings of household i are given by

$$\frac{b^i(s^r)}{P(s^t)} = \left(\frac{U_c^i(s^r)}{1+\tau_c}\right)^{-1} \left\{ \sum_{t=r+1}^{\infty} \sum_{s^t} \beta^{t-r} \mu(s^t|s^r) \left[U_c^i(s^t) c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t) \ell^i(s^t) - \frac{U_c^i(s^t)}{(1+\tau_c)} \bar{T}(s^t) \right] \right\}$$

for any period r, history s^r.

A.6 Proof of Proposition 3

The Relaxed Ramsey planner's problem is to choose an allocation $x \in \mathcal{X}$, market weights $\varphi \equiv (\varphi^i)$, and $\bar{T} \in \mathbb{R}$, that maximize the pseudo-utility function in (34) subject to technology and resource constraints (6)-(8). First, note that at any history s^t, the planner can solve a static sub-problem: maximize final good output $Y(s^t)$ given productivity $A(s_t)$ and aggregate labor supply, $L(s^t)$. Specifically:

$$Y(s^t) = \max_{(n^j(s^t))_{j\in\mathcal{J}}} \left[\int_{j\in\mathcal{J}} (A(s_t) n^j(s^t))^{\frac{\rho-1}{\rho}} \mathrm{d}j \right]^{\frac{\rho}{\rho-1}} \qquad \text{subject to} \qquad L(s^t) = \int_{j\in\mathcal{J}} n^j(s^t) \mathrm{d}j.$$

The first-order conditions for this problem yield: $n^j(s^t) = n^{j'}(s^t) = L(s^t)$ for all $j, j' \in \mathcal{J}$, which implies that at the planner's optimum $y^j(s^t) = Y(s^t) = A(s_t)L(s^t)$ for all $j \in \mathcal{J}$.

Using this result, we can rewrite the relaxed planner's problem in terms of aggregates alone:

$$\max_{\{C(s^t),Y(s^t),L(s^t)\},\varphi,\bar{T}} \sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}(C(s^t), L(s^t); \varphi, \nu, \lambda) - U_C^m(s_0) \sum_{i\in I} \pi^i \nu^i \bar{T}$$

subject to

$$C(s^t) + G(s^t) = Y(s^t) = A(s_t)L(s^t), \qquad \forall s^t \in S^t. \tag{83}$$

We let $\beta^t \mu(s^t)\hat{\varsigma}(s^t)$ denote the Lagrange multiplier on the time t, history s^t resource constraints in (83). The first-order conditions of this problem give us:

$$\beta^t \mu(s^t)\mathcal{W}_C(s^t) - \beta^t \mu(s^t)\hat{\varsigma}(s^t) = 0,$$
$$\beta^t \mu(s^t)\mathcal{W}_L(s^t) + \beta^t \mu(s^t)\hat{\varsigma}(s^t)A(s_t) = 0.$$

Combining, we obtain the relaxed planner's optimality condition in (35).

A.7 Derivation of (38)

Under sticky prices, $y^s(s^t) = \epsilon(s^t)^{-\rho} y^f(s^t)$. From the firm production functions:

$$n^s(s^t) = \frac{y^s(s^t)}{A(s_t)}, \qquad \text{and} \qquad n^f(s^t) = \frac{y^f(s^t)}{A(s_t)}$$

Aggregate output and aggregate labor thereby satisfy:

$$Y(s^t) = \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} = \left[\kappa \epsilon(s^t)^{-(\rho-1)} y^f(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}$$

and

$$L(s^t) = \kappa n^s(s^t) + (1-\kappa) n^f(s^t) = \kappa \frac{y^s(s^t)}{A(s_t)} + (1-\kappa)\frac{y^f(s^t)}{A(s_t)} = \kappa \frac{\epsilon(s^t)^{-\rho} y^f(s^t)}{A(s_t)} + (1-\kappa)\frac{y^f(s^t)}{A(s_t)},$$

respectively. Therefore

$$Y(s^t) = y^f(s^t)\left[\kappa \epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}$$

and

$$L(s^t) = \frac{y^f(s^t)}{A(s_t)}\left[\kappa \epsilon(s^t)^{-\rho} + (1-\kappa)\right]$$

Taking the ratio of aggregate output to aggregate labor, we get:

$$\frac{Y(s^t)}{L(s^t)} = \frac{y^f(s^t)\left[\kappa \epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}}{\frac{y^f(s^t)}{A(s_t)}\left[\kappa \epsilon(s^t)^{-\rho} + (1-\kappa)\right]} = A(s_t)\frac{\left[\kappa \epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}}{\left[\kappa \epsilon(s^t)^{-\rho} + (1-\kappa)\right]}$$

Therefore, aggregate output satisfies $Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t)$ with Δ defined in (39).

A.8 Proof of Lemma 3

Note that $\Delta(\epsilon)$ is a continuous function of ϵ. The first derivative of $\Delta(\epsilon)$ is given by:

$$\frac{d\Delta(\epsilon)}{d\epsilon} = \rho \Delta(\epsilon)^{1-\frac{1}{\rho}} \frac{d}{d\epsilon}\left\{\frac{\left[\kappa \epsilon^{-(\rho-1)} + (1-\kappa)\right]^{\frac{1}{\rho-1}}}{\left[\kappa \epsilon^{-\rho} + (1-\kappa)\right]^{1/\rho}}\right\}$$

where the last term satisfies:

$$\frac{d}{d\epsilon}\left\{\frac{\left[\kappa\epsilon^{-(\rho-1)}+(1-\kappa)\right]^{\frac{1}{\rho-1}}}{\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{1/\rho}}\right\}=\kappa\Delta(\epsilon)^{\frac{1}{\rho}}\epsilon^{-\rho-1}\left\{\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{-1}-\left[\kappa\epsilon^{-\rho+1}+(1-\kappa)\right]^{-1}\epsilon\right\}.$$

Therefore:

$$\frac{d\Delta(\epsilon)}{d\epsilon}=\kappa\rho\Delta(\epsilon)\epsilon^{-\rho-1}\left\{\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{-1}-\left[\kappa\epsilon^{-\rho+1}+(1-\kappa)\right]^{-1}\epsilon\right\} \tag{84}$$

To obtain a maxima or minima, we set the first derivative equal to zero as follows:

$$\Delta(\epsilon)\epsilon^{-\rho-1}\left\{\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{-1}-\left[\kappa\epsilon^{-\rho+1}+(1-\kappa)\right]^{-1}\epsilon\right\}=0.$$

Noting that both $\Delta(\epsilon)$ and $\epsilon^{-\rho-1}$ are strictly positive, this implies:

$$\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{-1}-\left[\kappa\epsilon^{-\rho+1}+(1-\kappa)\right]^{-1}\epsilon=0.$$

Solving this for ϵ, we obtain a unique solution of $\epsilon = 1$. Furthermore, note that from (84), $d\Delta(\epsilon)/d\epsilon > 0$ if and only if $\epsilon < 1$. Finally, we evaluate the second derivative of $\Delta(\epsilon)$ at $\epsilon = 1$, and find that it is unambiguously negative:

$$\Delta''(1)=-\rho\kappa(1-\kappa)<0$$

We conclude that the function $\Delta(\epsilon)$ attains a global maximum at $\epsilon = 1$. The function $\Delta(\epsilon)$ is strictly increasing in ϵ when $\epsilon < 1$ and is strictly decreasing in ϵ when $\epsilon > 1$. Finally, the maximal value of this function is given by:

$$\max_{\epsilon>0}\Delta(\epsilon)=\Delta(1)\equiv\left\{\frac{\left[\kappa+(1-\kappa)\right]^{\frac{1}{\rho-1}}}{\left[\kappa+(1-\kappa)\right]^{1/\rho}}\right\}^{\rho}=1$$

as was to be shown.

A.9 Ramsey Optimum

We solve the Ramsey problem given an arbitrary set of Pareto weights. For this problem we let $\beta^t\mu(s^t)\xi(s^t)$ and $\beta^t\mu(s^{t-1})\upsilon(s^{t-1})$ denote the Lagrange multipliers on the implementability conditions (30) and (31), respectively. We let $\beta^t\mu(s^t)\varsigma(s^t)$ denote the the Lagrange multipliers on the aggregate resource constraint in (38). We obtain the following characterization.

Proposition 8. *Given Assumption 1, an allocation x^* is a Ramsey optimum if, for all $s^t \in S^t$,*

$$-\frac{\mathcal{W}_L(s^t)+\xi(s^t)U^m_{LL}(s^t)+\upsilon(s^{t-1})\left[U^m_{LL}(s^t)L(s^t)+U^m_L(s^t)\right]F(\epsilon(s^t))}{\mathcal{W}_C(s^t)+\xi(s^t)U^m_{CC}(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}=\frac{Y(s^t)}{L(s^t)},\qquad \forall s^t\in S^t \tag{85}$$

and
$$\frac{\xi(s^t)}{L(s^t)} = \frac{\varsigma(s^t)A(s_t)\Delta'(\epsilon(s^t)) + \upsilon(s^{t-1})U_L^m(s^t)F'(\epsilon(s^t))}{U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)} \tag{86}$$

where $\varsigma(s^t) > 0$,
$$F(\epsilon(s^t)) \equiv \frac{\epsilon(s^t) - 1}{\kappa + (1-\kappa)\epsilon(s^t)^\rho}, \tag{87}$$

and
$$F'(\epsilon(s^t)) = \frac{\kappa + (1-\kappa)(1-\rho)\epsilon(s^t)^\rho + (1-\kappa)\rho\epsilon(s^t)^{\rho-1}}{(\kappa + (1-\kappa)\epsilon(s^t)^\rho)^2}. \tag{88}$$

Proof. In the Ramsey problem, we have the following constraints:

$$\sum_{s^t|s^{t-1}} U_C^m(s^t)Y(s^t)\frac{\epsilon(s^t) - 1}{\epsilon(s^t)^\rho\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]}\mu(s^t|s^{t-1}) = 0; \tag{89}$$

and
$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t), \qquad \forall s^t \in S^t;$$

Thus
$$U_C^m(s^t) = -\frac{U_L^m(s^t)}{\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)} \qquad \forall s^t \in S^t.$$

Substituting this into (89), we obtain:

$$\sum_{s^t|s^{t-1}} \frac{U_L^m(s^t)Y(s^t)}{\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}\frac{\epsilon(s^t) - 1}{\epsilon(s^t)^\rho\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]}\mu(s^t|s^{t-1}) = 0$$

where $Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t)$. Thus:

$$\sum_{s^t|s^{t-1}} \frac{U_L^m(s^t)\Delta(\epsilon(s^t))L(s^t)}{\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}\left[\frac{\epsilon(s^t) - 1}{\epsilon(s^t)^\rho\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]}\right]\mu(s^t|s^{t-1}) = 0 \tag{90}$$

where
$$\Delta(\epsilon) \equiv \frac{\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{\rho}{1-\rho}}}{\left[\kappa\epsilon(s^t)^{-\rho} + (1-\kappa)\right]}$$

Equation (90) thus reduces to:

$$\sum_{s^t|s^{t-1}} U_L^m(s^t)L(s^t)F(\epsilon(s^t))\mu(s^t|s^{t-1}) = 0 \tag{91}$$

where F is a function defined as in (87).

We then write the planner's Lagrangian as follows:

$$\begin{aligned}\mathcal{L} = & \sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}(C(s^t), L(s^t); \varphi, \nu, \lambda) - U_C^m(s_0)\bar{T} \sum_{i \in I} \pi^i \nu^i \\ & + \sum_t \sum_{s^t} \beta^t \mu(s^t) \xi(s^t) \left\{ U_C^m(s^t) \chi \left[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa) \right]^{-\frac{1}{1-\rho}} A(s_t) + U_L^m(s^t) \right\} \\ & + \sum_t \sum_{s^t} \beta^t \mu(s^t) \varsigma(s^t) \left\{ A(s_t) \Delta(\epsilon(s^t)) L(s^t) - C(s^t) - G(s_t) \right\} \\ & + \sum_t \sum_{s^t} \beta^{t+1} \mu(s^t) \upsilon(s^t) \sum_{s^t | s^{t-1}} U_L^m(s^{t+1}) L(s^{t+1}) F(\epsilon(s^{t+1})) \mu(s^{t+1} | s^t)\end{aligned}$$

with complementary slackness conditions:

$$A(s_t)\Delta(\epsilon(s^t))L(s^t) - C(s^t) - G(s_t) \geq 0, \qquad \varsigma(s^t) \geq 0, \tag{92}$$

and

$$\varsigma(s^t) \left\{ A(s_t)\Delta(\epsilon(s^t))L(s^t) - C(s^t) - G(s_t) \right\} = 0, \qquad \forall s^t \in S^t. \tag{93}$$

The FOC with respect to $C(s^t)$ is given by:

$$0 = \mathcal{W}_C(s^t) + \xi(s^t) U_{CC}^m(s^t) \chi \left[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa) \right]^{-\frac{1}{1-\rho}} A(s_t) - \varsigma(s^t), \tag{94}$$

the FOC with respect to $L(s^t)$ is given by:

$$0 = \mathcal{W}_L(s^t) + \xi(s^t) U_{LL}^m(s^t) + \varsigma(s^t) \frac{Y(s^t)}{L(s^t)} + \upsilon(s^{t-1}) \left[U_{LL}^m(s^t) L(s^t) + U_L^m(s^t) \right] F(\epsilon(s^t)), \tag{95}$$

and the FOC with respect to $\epsilon(s^t)$ is given by:

$$\begin{aligned}0 = & - \xi(s^t) U_C^m(s^t) \chi \left[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa) \right]^{-\frac{1}{1-\rho} - 1} \kappa \epsilon(s^t)^{-\rho} A(s_t) + \varsigma(s^t) A(s_t) \Delta'(\epsilon(s^t)) L(s^t) \\ & + \upsilon(s^{t-1}) U_L^m(s^t) L(s^t) F'(\epsilon(s^t)).\end{aligned}$$

where $F'$is given in (88).

Combining (94) and (94), we get:

$$\begin{aligned}0 = & \mathcal{W}_L(s^t) + \xi(s^t) U_{LL}^m(s^t) + \left[\mathcal{W}_C(s^t) + \xi(s^t) U_{CC}^m(s^t) \chi \left[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa) \right]^{-\frac{1}{1-\rho}} A(s_t) \right] \frac{Y(s^t)}{L(s^t)} \\ & + \upsilon(s^{t-1}) \left[U_{LL}^m(s^t) L(s^t) + U_L^m(s^t) \right] F(\epsilon(s^t)),\end{aligned}$$

Therefore

$$- \frac{\mathcal{W}_L(s^t) + \xi(s^t) U_{LL}^m(s^t) + \upsilon(s^{t-1}) \left[U_{LL}^m(s^t) L(s^t) + U_L^m(s^t) \right] F(\epsilon(s^t))}{\mathcal{W}_C(s^t) + \xi(s^t) U_{CC}^m(s^t) \chi [\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{-\frac{1}{1-\rho}} A(s_t)} = \frac{Y(s^t)}{L(s^t)} \tag{96}$$

as we have stated in 85.

The FOC with respect to $\epsilon(s^t)$ gives us

$$\xi(s^t)U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)=\varsigma(s^t)A(s_t)\Delta'(\epsilon(s^t))L(s^t) + \upsilon(s^{t-1})U_L^m(s^t)L(s^t)F'(\epsilon(s^t)) \tag{97}$$

We can write this as in 86. Finally, the FOC with respect to χ is given by

$$\sum_t\sum_{s^t}\beta^t\mu(s^t)\xi(s^t)\left\{U_C^m(s^t)\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)\right\}=0.$$

□

A.10 Proof of Theorem 2

At the Ramsey optimum:

$$-\frac{\mathcal{W}_L(s^t)+\xi(s^t)U_{LL}^m(s^t)+\upsilon(s^{t-1})\left[U_{LL}^m(s^t)L(s^t)+U_L^m(s^t)\right]F(\epsilon(s^t))}{\mathcal{W}_C(s^t)+\xi(s^t)U_{CC}^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}=\frac{Y(s^t)}{L(s^t)}$$

Iso-elastic preferences imply

$$\frac{U_{CC}^m(s^t)C(s^t)}{U_C^m(s^t)}=-\gamma \qquad \text{and} \qquad \frac{U_{LL}^m(s^t)L(s^t)}{U_L^m(s^t)}=\eta$$

Therefore

$$-\frac{\mathcal{W}_L(s^t)+\eta\xi(s^t)U_L^m(s^t)L(s^t)^{-1}+(1+\eta)\upsilon(s^{t-1})U_L^m(s^t)F(\epsilon(s^t))}{\mathcal{W}_C(s^t)-\gamma\xi(s^t)U_C^m(s^t)C(s^t)^{-1}\chi^*\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}=\frac{Y(s^t)}{L(s^t)} \tag{98}$$

With separable and iso-elastic utility, the derivatives of the pseudo-utility in (34) with respect to $C(s^t)$ and $L(s^t)$ are given by, respectively:

$$\mathcal{W}_C(s^t)=U_C^m(s^t)\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right] \tag{99}$$

$$\mathcal{W}_L(s^t)=U_L^m(s^t)\sum_{i\in I}\pi^i\omega_L^i(\varphi,s_t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]. \tag{100}$$

Substituting these expressions for $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ into (98), we get:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{\sum_{i\in I}\pi^i\omega_L^i(s^t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]+\eta\xi(s^t)L(s^t)^{-1}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}{\sum_{i\in I}\pi^i\omega_C^i\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]-\gamma\xi(s^t)C(s^t)^{-1}\chi^*\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}\right)=\frac{Y(s^t)}{L(s^t)}$$

Therefore the optimal monetary wedge satisfies:

$$1-\tau_M^*(s^t)=\frac{(\chi^*)^{-1}\sum_{i\in I}\pi^i\omega_C^i\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]-\gamma\xi(s^t)C(s^t)^{-1}\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}{\sum_{i\in I}\pi^i\omega_L^i(s^t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]+\eta\xi(s^t)L(s^t)^{-1}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

Using the fact that

$$C(s^t) = Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t)$$

we can write the optimal monetary wedge as follows

$$1 - \tau_M^*(s^t) = \frac{(\chi^*)^{-1}\sum_{i\in I}\pi^i\omega_C^i\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1-\gamma)\right] - \gamma\frac{\xi(s^t)}{L(s^t)}\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\sum_{i\in I}\pi^i\omega_L^i(s^t)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right] + \eta\frac{\xi(s^t)}{L(s^t)} + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

Next we define a function $\mathcal{I}(s_t)$ and a constant $\bar{\mathcal{I}}_C$ as follows:

$$\mathcal{I}(s_t) \equiv \sum_{i\in I}\pi^i\omega_L^i(s^t)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right], \qquad \text{and} \qquad \bar{\mathcal{I}}_C \equiv (\chi^*)^{-1}\sum_{i\in I}\pi^i\omega_C^i\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1-\gamma)\right].$$

The optimal monetary wedge can then be written as follows:

$$1 - \tau_M^*(s^t) = \frac{\bar{\mathcal{I}}_C - \gamma\frac{\xi(s^t)}{L(s^t)}\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t) + \eta\frac{\xi(s^t)}{L(s^t)} + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

where $\xi(s^t)/L(s^t)$ satisfies (86).

Next we decompose $\xi(s^t)/L(s^t)$ into two terms:

$$\frac{\xi(s^t)}{L(s^t)} = \hat{\varsigma}(s^t) + \hat{\xi}(s^t)$$

where

$$\hat{\varsigma}(s^t) \equiv \varsigma(s^t)\frac{A(s_t)\Delta'(\epsilon(s^t))}{U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)} \tag{101}$$

and

$$\hat{\xi}(s^t) \equiv \frac{\upsilon(s^{t-1})U_L^m(s^t)F'(\epsilon(s^t))}{U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)} \tag{102}$$

Note that in any sticky price equilibrium:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t);$$

Therefore, we can write $\hat{\xi}(s^t)$ as follows:

$$\hat{\xi}(s^t) = -\kappa^{-1}\upsilon(s^{t-1})\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]\epsilon(s^t)^{\rho}F'(\epsilon(s^t))$$

Threshold. We first consider the conditions under which $\tau_M^*(s^t) = 0$, equivalently, $\epsilon(s^t) = 1$. In this case:

$$\mathcal{I}(s_t) + \eta\frac{\xi(s^t)}{L(s^t)} + (1+\eta)\upsilon(s^{t-1})F(1) = \bar{\mathcal{I}}_C - \gamma\frac{\xi(s^t)}{L(s^t)}$$

where $F(1) = 0$. Therefore

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C - \gamma \frac{\xi(s^t)}{L(s^t)} - \eta \frac{\xi(s^t)}{L(s^t)}$$

Thus

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C - (\eta + \gamma) \frac{\xi(s^t)}{L(s^t)} \tag{103}$$

Next, the $\xi(s^t)/L(s^t)$ reduces to:

$$\frac{\xi(s^t)}{L(s^t)} = -\kappa^{-1} \upsilon(s^{t-1}) F'(1)$$

Plugging this into (103) we get:

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C + \kappa^{-1}(\eta + \gamma)\upsilon(s^{t-1}) F'(1)$$

where

$$F'(1) = \kappa + (1-\kappa)(1-\rho) + (1-\kappa)\rho = 1$$

Therefore there exists a threshold that depends only on past states given by:

$$\bar{\mathcal{I}}(s^{t-1}) \equiv \bar{\mathcal{I}}_C + \kappa^{-1}(\eta + \gamma)\upsilon(s^{t-1}).$$

When $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$, the optimal $\tau_M^*(s^t) = 0$.

First fictitious tax wedge. We first define a fictitious tax wedge as follows:

$$1 - \hat{\tau}_0(s^t) \equiv \frac{\bar{\mathcal{I}}_C + \gamma\kappa^{-1}\upsilon(s^{t-1})}{\mathcal{I}(s_t) - \eta\kappa^{-1}\upsilon(s^{t-1})}$$

This wedge is unambiguously falling in $\mathcal{I}(s_t)$, as all other terms are constants. Furthermore, note that when $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}) \equiv \bar{\mathcal{I}}_C + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1})$, this wedge is equal to one. As a result, the fictitious tax $\hat{\tau}_0(s^t)$ trivially satisfies:

$$\begin{array}{lll} \hat{\tau}_0(s^t) > 0 & \text{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1}), \\ \hat{\tau}_0(s^t) = 0 & \text{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}), \\ \hat{\tau}_0(s^t) < 0 & \text{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1}). \end{array}$$

Second fictitious tax wedge. We next define a second fictitious tax as the one that jointly satisfies the following equations:

$$1 - \hat{\tau}_1(s^t) = \frac{\bar{\mathcal{I}} - \gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

where $\hat{\xi}(s^t)$ is defined in (102) and satisfies:

$$\hat{\xi}(s^t) = -\kappa^{-1}\upsilon(s^{t-1})\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]\epsilon(s^t)^{\rho}F'(\epsilon(s^t)).$$

Therefore

$$\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}=-\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))$$

Substituting this in, we have:

$$1-\hat{\tau}_1(s^t)=\frac{\bar{\mathcal{I}}+\gamma\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))}{\mathcal{I}(s_t)-\eta\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

The goal is to compare this to $1-\hat{\tau}_0(s^t)$. First note that when $\epsilon(s^t)=1$, $\hat{\tau}_1(s^t)=\hat{\tau}_0(s^t)$, therefore $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$.

Suppose $\upsilon(s^{t-1})$ is negative. We consider the case in which $\epsilon(s^t)>1$. When $\epsilon(s^t)>1$:

$$F'(\epsilon(s^t))\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]<1$$

which implies

$$\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))>\upsilon(s^{t-1})$$

since $\upsilon(s^{t-1})$ is negative, which further implies

$$\bar{\mathcal{I}}+\gamma\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))>\bar{\mathcal{I}}+\gamma\kappa^{-1}\upsilon(s^{t-1})$$

Therefore the numerator of $1-\hat{\tau}_1(s^t)$ is strictly greater than the numerator of $1-\hat{\tau}_0(s^t)$.

Furthermore when $\epsilon(s^t)>1$:

$$-\eta\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))<-\eta\kappa^{-1}\upsilon(s^{t-1})$$

and

$$(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))<0$$

since $F(\epsilon(s^t))>1$. Together, these imply:

$$\mathcal{I}(s_t)-\eta\kappa^{-1}\upsilon(s^{t-1})\left[\kappa+(1-\kappa)\epsilon(s^t)^{\rho}\right]F'(\epsilon(s^t))+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))<\mathcal{I}(s_t)-\eta\kappa^{-1}\upsilon(s^{t-1})$$

Therefore the denominator of $1-\hat{\tau}_1(s^t)$ is strictly less than the denominator of $1-\hat{\tau}_0(s^t)$.

Therefore, when $\upsilon(s^{t-1})$ is negative and $\epsilon(s^t)>1$, the following is true: $1-\hat{\tau}_1(s^t)>1-\hat{\tau}_0(s^t)$. This implies:

$$0<\hat{\tau}_1(s^t)<\hat{\tau}_0(s^t)$$

Similarly when $\upsilon(s^{t-1})$ is negative and $\epsilon(s^t)<1$, the following is true: $1-\hat{\tau}_1(s^t)<1-\hat{\tau}_0(s^t)$. This implies:

$$\hat{\tau}_0(s^t)<\hat{\tau}_1(s^t)<0$$

We next consider the case in which $\upsilon(s^{t-1})$ is positive. Again consider the case in which $\epsilon(s^t) > 1$. When $\epsilon(s^t) > 1$:

$$F'(\epsilon(s^t))\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right] < 1$$

which implies

$$\upsilon(s^{t-1})\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right] F'(\epsilon(s^t)) < \upsilon(s^{t-1})$$

which further implies

$$\bar{\mathcal{I}}_C + \gamma\kappa^{-1}\upsilon(s^{t-1})\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right] F'(\epsilon(s^t)) < \bar{\mathcal{I}}_C + \gamma\kappa^{-1}\upsilon(s^{t-1})$$

Therefore the numerator of $1 - \hat{\tau}_1(s^t)$ is strictly less than the numerator of $1 - \hat{\tau}_0(s^t)$.

Furthermore when $\epsilon(s^t) > 1$:

$$-\eta\kappa^{-1}\upsilon(s^{t-1})\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right] F'(\epsilon(s^t)) > -\eta\kappa^{-1}\upsilon(s^{t-1})$$

and

$$(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) > 0$$

since $F(\epsilon(s^t)) > 1$. Together, these imply:

$$\mathcal{I}(s_t) - \eta\kappa^{-1}\upsilon(s^{t-1})\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right] F'(\epsilon(s^t)) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) > \mathcal{I}(s_t) - \eta\kappa^{-1}\upsilon(s^{t-1})$$

Therefore the denominator of $1 - \hat{\tau}_1(s^t)$ is strictly greater than the denominator of $1 - \hat{\tau}_0(s^t)$.

Therefore, when $\upsilon(s^{t-1})$ is positive and $\epsilon(s^t) > 1$, the following is true: $1 - \hat{\tau}_1(s^t) < 1 - \hat{\tau}_0(s^t)$. This implies $\hat{\tau}_0(s^t) < \hat{\tau}_1(s^t)$. But note that $\epsilon(s^t) > 1$ corresponds with $\hat{\tau}_0(s^t) > 0$. Therefore

$$0 < \hat{\tau}_0(s^t) < \hat{\tau}_1(s^t)$$

Similarly when $\upsilon(s^{t-1})$ is positive and $\epsilon(s^t) < 1$, the following is true: $1 - \hat{\tau}_1(s^t) > 1 - \hat{\tau}_0(s^t)$. This implies $\hat{\tau}_0(s^t) > \hat{\tau}_1(s^t)$. But note that $\epsilon(s^t) < 1$ corresponds with $\hat{\tau}_0(s^t) < 0$. Therefore

$$\hat{\tau}_1(s^t) < \hat{\tau}_0(s^t) < 0$$

As a result, the fictitious tax rate $\hat{\tau}_1(s^t)$ satisfies:

$$\begin{array}{lll} \hat{\tau}_1(s^t) > 0 & \text{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1}), \\ \hat{\tau}_1(s^t) = 0 & \text{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}), \\ \hat{\tau}_1(s^t) < 0 & \text{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1}). \end{array}$$

Define a third fictitious tax wedge. We define a third fictitious tax as the one that jointly satisfies the following equations:

$$1 - \hat{\tau}_2(s^t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

where $\hat{\xi}(s^t)$ is defined in (102). The goal is to compare this to the fictitious tax wedge $1-\hat{\tau}_1(s^t)$. First note that when $\epsilon(s^t)=1$, $\hat{\tau}_2(s^t)=\hat{\tau}_1(s^t)$, therefore $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$.

Furthermore note that the numerators of $1-\hat{\tau}_2(s^t)$ and $1-\hat{\tau}_1(s^t)$ are identical. Therefore, one needs only to consider the denominator.

First, note that

$$\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}<1$$

if and only if $1<\epsilon(s^t)$.

Suppose first $\upsilon(s^{t-1})$ is negative. We consider the case in which $\epsilon(s^t)>1$. When $\epsilon(s^t)>1$:

$$\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}<1$$

and $\upsilon(s^{t-1})<0$ implies $\hat{\xi}(s^t)>0$. This implies

$$\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))>\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))$$

Therefore the denominator of $1-\hat{\tau}_2(s^t)$ is strictly greater than the denominator of $1-\hat{\tau}_1(s^t)$.

Therefore, when $\upsilon(s^{t-1})$ is negative and $\epsilon(s^t)>1$,the following is true: $1-\hat{\tau}_1(s^t)>1-\hat{\tau}_2(s^t)$. This implies:

$$0<\hat{\tau}_1(s^t)<\hat{\tau}_2(s^t)$$

Similarly when $\upsilon(s^{t-1})$ is negative and $\epsilon(s^t)<1$,the following is true: $1-\hat{\tau}_1(s^t)<1-\hat{\tau}_2(s^t)$. This implies:

$$\hat{\tau}_2(s^t)<\hat{\tau}_1(s^t)<0$$

We next consider the case in which $\upsilon(s^{t-1})$ is positive. Again consider the case in which $\epsilon(s^t)>1$. In this case, $\upsilon(s^{t-1})<0$ implies $\hat{\xi}(s^t)<0$. This implies

$$\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))<\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))$$

Therefore the denominator of $1-\hat{\tau}_2(s^t)$ is strictly less than the denominator of $1-\hat{\tau}_1(s^t)$.

Therefore, when $\upsilon(s^{t-1})$ is positive and $\epsilon(s^t)>1$, the following is true: $1-\hat{\tau}_1(s^t)<1-\hat{\tau}_2(s^t)$. This implies:

$$0<\hat{\tau}_2(s^t)<\hat{\tau}_1(s^t).$$

Similarly when $\upsilon(s^{t-1})$ is positive and $\epsilon(s^t)<1$, the following is true: $1-\hat{\tau}_1(s^t)>1-\hat{\tau}_2(s^t)$. This implies:

$$\hat{\tau}_1(s^t)<\hat{\tau}_2(s^t)<0.$$

As a result, the fictitious tax rate $\hat{\tau}_2(s^t)$ satisfies:

$$\begin{aligned}
\hat{\tau}_2(s^t)>0 &\qquad \text{if and only if} \quad \mathcal{I}(s_t)>\bar{\mathcal{I}}(s^{t-1}),\\
\hat{\tau}_2(s^t)=0 &\qquad \text{if and only if} \quad \mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1}),\\
\hat{\tau}_2(s^t)<0 &\qquad \text{if and only if} \quad \mathcal{I}(s_t)<\bar{\mathcal{I}}(s^{t-1}).
\end{aligned}$$

The optimal monetary wedge. Finally we consider the optimal monetary wedge:

$$1-\tau_M^*(s^t)=\frac{\bar{\mathcal{I}}_C-\gamma\frac{\xi(s^t)}{L(s^t)}\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t)+\eta\frac{\xi(s^t)}{L(s^t)}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

where

$$\frac{\xi(s^t)}{L(s^t)}=\hat{\varsigma}(s^t)+\hat{\xi}(s^t)$$

where $\hat{\varsigma}(s^t)$ is defined in (101) and $\hat{\xi}(s^t)$ is defined in (102). Therefore the optimal monetary wedge satisfies

$$1-\tau_M^*(s^t)=\frac{\bar{\mathcal{I}}_C-\gamma\left[\hat{\varsigma}(s^t)+\hat{\xi}(s^t)\right]\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t)+\eta\left[\hat{\varsigma}(s^t)+\hat{\xi}(s^t)\right]+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}$$

The goal is to compare this to the fictitious tax wedge $1-\hat{\tau}_2(s^t)$. First note that when $\epsilon(s^t)=1$, $\tau_M^*(s^t)=\hat{\tau}_2(s^t)$, therefore $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$.

We consider the case in which $\epsilon(s^t)>1$. When $\epsilon(s^t)>1$:

$$\Delta'(\epsilon(s^t))<0$$

which implies $\hat{\varsigma}(s^t)<0$. Therefore

$$\bar{\mathcal{I}}_C-\gamma\left[\hat{\varsigma}(s^t)+\hat{\xi}(s^t)\right]\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}>\bar{\mathcal{I}}_C-\gamma\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}$$

Therefore the numerator of $1-\tau_M^*(s^t)$ is strictly greater than the numerator of $1-\hat{\tau}_2(s^t)$. Furthermore:

$$\mathcal{I}(s_t)+\eta\left[\hat{\varsigma}(s^t)+\hat{\xi}(s^t)\right]+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))<\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))$$

Therefore the denominator of $1-\tau_M^*(s^t)$ is strictly less than the denominator of $1-\hat{\tau}_2(s^t)$.

Therefore, when $\epsilon(s^t)>1$, the following is true: $1-\tau_M^*(s^t)>1-\hat{\tau}_2(s^t)$. This implies

$$0<\tau^*(s^t)<\hat{\tau}_2(s^t)$$

Similarly, when $\epsilon(s^t)<1$, the following is true: $1-\tau_M^*(s^t)<1-\hat{\tau}_2(s^t)$. This implies

$$\hat{\tau}_2(s^t)<\tau^*(s^t)<0$$

To conclude, the optimal monetary tax rate $\tau_M^*(s^t)$ satisfies:

$$\begin{array}{lll}\tau_M^*(s^t)>0 & \text{if and only if} & \mathcal{I}(s_t)>\bar{\mathcal{I}}(s^{t-1}),\\ \tau_M^*(s^t)=0 & \text{if and only if} & \mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1}),\\ \tau_M^*(s^t)<0 & \text{if and only if} & \mathcal{I}(s_t)<\bar{\mathcal{I}}(s^{t-1}).\end{array}$$

A.11 Proof of Proposition 4

In any sticky price equilibrium, the aggregate price level satisfies (66). Using the fiscal implementation that sets $(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)=1$, we have that the optimal markup satisfies:

$$\log\mathcal{M}(s^t)=\log\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}=\frac{1}{1-\rho}\log\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]$$

with $\rho>1$. Therefore

$$\begin{array}{lll}\log\mathcal{M}(s^t)>0 & \text{if and only if} & \epsilon(s^t)>1,\\ \log\mathcal{M}(s^t)=0 & \text{if and only if} & \epsilon(s^t)=1,\\ \log\mathcal{M}(s^t)<0 & \text{if and only if} & \epsilon(s^t)<1.\end{array}$$

From our proof of Theorem 2, we show that (i) $\epsilon(s^t)=1$ if and only if $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$; (ii) $\epsilon(s^t)>1$ if and only if $\mathcal{I}(s_t)>\bar{\mathcal{I}}(s^{t-1})$; and (iii) $\epsilon(s^t)<1$ if and only if $\mathcal{I}(s_t)<\bar{\mathcal{I}}(s^{t-1})$. The result stated in Proposition 4 then follows.

A.12 Proof of Proposition 5

The Euler equation of the fictional household in (15) is given by:

$$\frac{U_C^m(s^t)}{P(s^t)}=\beta(1+i(s^t))\sum_{s^{t+1}|s}\mu(s^{t+1}|s)\frac{U_C^m(s^{t+1})}{P(s^{t+1})}.$$

With i.i.d. shocks, the expected marginal utility of consumption, $\sum_{s^{t+1}|s^t}\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{P(s^{t+1})}$ is equalized across states s and s'. Therefore:

$$\frac{1+i(s)}{1+i(s')}=\frac{U_C^m(s)/P(s)}{U_C^m(s')/P(s')}\tag{104}$$

From our proof of proposition ??, the aggregate price level that implements this allocation satisfies::

$$P(s^t)=\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}\frac{W(s^t)}{A(s_t)}$$

where we have used the fact that $(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)=1$. The nominal wage satisfies:

$$W(s^t)=\frac{h(s^{t-1})}{\epsilon(s^t)}A(s_t)$$

Therefore the aggregate price level is given by:

$$P(s)=\left[\kappa\epsilon(s)^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}\frac{h(s^{t-1})}{\epsilon(s)}$$

Substituting this into (104)

$$\frac{1+i(s)}{1+i(s')}=\frac{C(s)^{-\gamma}}{C(s')^{-\gamma}}\frac{\left[\kappa\epsilon(s')^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}\epsilon(s')^{-1}}{\left[\kappa\epsilon(s)^{1-\rho}+(1-\kappa)\right]^{\frac{1}{1-\rho}}\epsilon(s)^{-1}}$$

where we have used the fact that $U_C^m(s) = C(s)^{-\gamma}$. Note that

$$\left[\kappa\epsilon(s)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \epsilon(s)^{-1} = \left[\kappa\epsilon(s)^{1-\rho}\epsilon(s)^{-(1-\rho)} + (1-\kappa)\epsilon(s)^{-(1-\rho)}\right]^{\frac{1}{1-\rho}}$$
$$= \left[\kappa + (1-\kappa)\epsilon(s)^{-(1-\rho)}\right]^{\frac{1}{1-\rho}}$$

therefore

$$\frac{1+i(s)}{1+i(s')} = \frac{C(s)^{-\gamma}}{C(s')^{-\gamma}} \frac{\left[\kappa + (1-\kappa)\epsilon(s')^{-(1-\rho)}\right]^{\frac{1}{1-\rho}}}{\left[\kappa + (1-\kappa)\epsilon(s)^{-(1-\rho)}\right]^{\frac{1}{1-\rho}}}$$

$$\left[\kappa + (1-\kappa)\epsilon(s')^{-(1-\rho)}\right]^{\frac{1}{1-\rho}} < \left[\kappa + (1-\kappa)\epsilon(s)^{-(1-\rho)}\right]^{\frac{1}{1-\rho}}$$

$$\kappa + (1-\kappa)\epsilon(s')^{-(1-\rho)} > \kappa + (1-\kappa)\epsilon(s)^{-(1-\rho)}$$

$$\epsilon(s')^{\rho-1} > \epsilon(s)^{\rho-1}$$

Next, note that $C(s)$ and $L(s)$ jointly solve:and

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi^*(1-\tau_M^*(s^t))\frac{C(s^t)}{L(s^t)}$$

$$C(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t)$$

Thus

$$\frac{L(s^t)^\eta}{C(s')^{-\gamma}} = \chi^*(1-\tau_M^*(s^t))\frac{C(s^t)}{L(s^t)}$$

We next solve forsubproblem:

$$L(s^t)^{\eta+\gamma} = \chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t)^{1-\gamma}\Delta(\epsilon(s^t))^{-\gamma}$$

$$L(s^t) = \left[\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t)^{1-\gamma}\Delta(\epsilon(s^t))^{-\gamma}\right]^{\frac{1}{\eta+\gamma}}$$

$$L(s^t)^\eta = \left[\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t)^{1-\gamma}\Delta(\epsilon(s^t))^{-\gamma}\right]^{\frac{\eta}{\eta+\gamma}}$$

$$\frac{L(s^t)^\eta}{C(s^t)^{-\gamma}} = \chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t),$$

$$C(s)^{-\gamma} = \left[\kappa\epsilon(s)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} A(s)^{-1}L(s)^\eta$$

$$\frac{1+i(s)}{1+i(s')} = \frac{C(s)^{-\gamma}}{C(s')^{-\gamma}} \frac{\left[\kappa\epsilon(s')^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \frac{1}{\epsilon(s')}}{\left[\kappa\epsilon(s)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \frac{1}{\epsilon(s)}}$$

$$\frac{1+i(s)}{1+i(s')} = \frac{\left[\kappa\epsilon(s)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} A(s)^{-1}L(s)^\eta}{\left[\kappa\epsilon(s')^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} A(s')^{-1}L(s')^\eta} \frac{\left[\kappa\epsilon(s')^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \frac{1}{\epsilon(s')}}{\left[\kappa\epsilon(s)^{1-\rho} + (1-\kappa)\right]^{\frac{1}{1-\rho}} \frac{1}{\epsilon(s)}}$$

$$\frac{1+i(s)}{1+i(s')} = \frac{A(s)^{-1}L(s)^{\eta}}{A(s')^{-1}L(s')^{\eta}} \frac{\frac{1}{\epsilon(s')}}{\frac{1}{\epsilon(s)}}$$

$$\frac{1+i(s)}{1+i(s')} = \frac{A(s)^{-1}L(s)^{\eta}\epsilon(s)}{A(s')^{-1}L(s')^{\eta}\epsilon(s')}$$

$$\frac{1+i(s)}{1+i(s')} = \frac{L(s)^{\eta}\epsilon(s)}{L(s')^{\eta}\epsilon(s')}$$

$$\frac{1+i(s)}{1+i(s')} = \frac{\left[\left[\kappa\epsilon(s)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}\Delta(\epsilon(s))^{-\gamma}\right]^{\frac{\eta}{\eta+\gamma}}\epsilon(s)}{\left[\left[\kappa\epsilon(s')^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}\Delta(\epsilon(s'))^{-\gamma}\right]^{\frac{\eta}{\eta+\gamma}}\epsilon(s')}$$

Theorem 2 implies that with $\bar{\mathcal{I}}(s^{t-1}) < \mathcal{I}(s) < \mathcal{I}(s')$, we have that:

$$\tau_M^*(s') > \tau_M^*(s) > 0$$

and that

$$\epsilon(s') > \epsilon(s) > 1$$

$$1+i(s) < 1+i(s')$$

B Appendix: Proofs for Section 6

B.1 Derivation of Implementability Conditions (46)

We derive condition (46). We take the household's budget constraint in (45) for type $i \in I$, multiply both sides by $\Lambda^i(s^t)$, and use the household's FOCs in (50) and (51) to substitute out consumption and labor prices. Doing so, we obtain:

$$U_c^i(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)}U_\ell^i(s^t)\ell^i(s^t) - U_c^i(s^t)\frac{(1-\tau_\Pi)}{(1+\tau_c)}\sigma^i\frac{\Pi(s^t)}{P(s^t)} = \Lambda^i(s^t)z^i(s^t|s^{t-1}) - \Lambda^i(s^t)\sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)z^i(s^{t+1}|s^t) - \\ + \Lambda^i(s^t)(1+i(s^{t-1}))b^i(s^{t-1}) + \Lambda^i(s^t)P(s^t)\bar{T}(s^t)$$

where we let $\bar{T}(s^t) = T(s^t) + (1-\tau_\Pi)\Pi(s^t)/P(s^t)$ as before. Multiplying both sides by $\beta^t\mu(s^t)$, summing over t and s^t, and using the household's intertemporal optimality conditions (55)-(54) to cancel terms, we obtain:

$$\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[U_c^i(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)}U_\ell^i(s^t)\ell^i(s^t) - U_c^i(s^t)\frac{(1-\tau_\Pi)}{(1+\tau_c)}\sigma^i\frac{\Pi(s^t)}{P(s^t)}\right] \le U_c^i(s_0)\bar{T},$$

where

$$\bar{T} \equiv \frac{1}{U_c^i(s_0)(1+\tau_c)}\sum_t\sum_{s^t}\beta^t\mu(s^t)U_c^i(s^t)\bar{T}(s^t)$$

for all $i \in I$. Finally, using the solution and envelope conditions for the static sub-problem described in Lemma 1, as well as the fact that individual allocations satisfy (16), we can rewrite the above conditions as:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi, s_t)L(s^t) - U_C^m(s^t)\frac{1-\tau_\Pi}{1+\tau_c}\sigma^i \frac{\Pi(s^t)}{P(s^t)} \right] \leq U_C^m(s_0)\bar{T}$$

where $\bar{T}$ is as in (19), for all $i \in I$, as was to be shown.

B.2 Proof of Proposition 6

Necessity. Necessity of parts (i) and (ii) of the proposition follow from the same steps as those used to prove Proposition 2.

In order to prove part (iii) we must first obtain an expression for real profits in terms of the allocation and the monetary wedge function, $\epsilon(s^t)$, alone. We write aggregate profits, $\Pi(s^t)$ in the following way:

$$\Pi(s^t) = (1-\kappa)\Pi^f(s^t) + \kappa \Pi^s(s^t)$$

where $\Pi^f(s^t)$ denotes profits of the flexible-price firms and $\Pi^s(s^t)$ denotes profits of the sticky price firms in history s^t. Flexible-price firms profits are given by:

$$\Pi^f(s^t) = \left[(1-\tau_r)p_t^f(s^t) - \frac{W(s^t)}{A(s^t)} \right] y^f(s^t) = \frac{1}{\rho-1}\frac{W(s^t)}{A(s^t)}y^f(s^t)$$

where we have replaced $p_t^f(s^t)$ using the flexible-price firm's optimality condition (21). Doing the same for sticky price firms using (22) gives us:

$$\Pi^s(s^t) = \left[(1-\tau_R)p_t^s(s^{t-1}) - \frac{W(s^t)}{A(s^t)} \right] y^s(s^t) = \left(\frac{\rho}{\rho-1}\epsilon(s^t) - 1 \right) \frac{W(s^t)}{A(s^t)} y^s(s^t)$$

This implies aggregate profits can be written as:

$$\Pi(s^t) = (1-\kappa)\frac{1}{\rho-1}\frac{W(s^t)}{A(s^t)}y^f(s^t) + \kappa \left(\frac{\rho}{\rho-1}\epsilon(s^t) - 1 \right) \frac{W(s^t)}{A(s^t)} y^s(s^t)$$

Next, we use the necessary equilibrium condition $y^f(s^t) = \epsilon(s^t)^\rho y^s(s^t)$. Substituting this into the above expression and rewriting in terms of real profits yields:

$$\frac{\Pi(s^t)}{P(s^t)} = \left[(1-\kappa)\frac{\epsilon(s^t)^\rho}{\rho-1} + \kappa \left(\frac{\rho}{\rho-1}\epsilon(s^t) - 1 \right) \right] \frac{W(s^t)}{P(s^t)} \frac{y^s(s^t)}{A(s^t)}$$

Next, we replace the real wage $W(s^t)/P(s^t)$ in the above expression using the representative household's intratemporal condition, equation (13). This gives us the following expression for real profits:

$$\frac{\Pi(s^t)}{P(s^t)} = -\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1+\tau_c}{1-\tau_\ell} \left[(1-\kappa)\frac{\epsilon(s^t)^\rho}{\rho-1} + \kappa \left(\frac{\rho\epsilon(s^t)}{\rho-1} - 1 \right) \right] \frac{y^s(s^t)}{A(s^t)} \tag{105}$$

Finally, we have that aggregate output is given by

$$Y(s^t) = \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}$$

Again using the fact that $y^f(s^t) = \epsilon(s^t)^\rho y^s(s^t)$, we can write this as:

$$Y(s^t) = \left[\kappa + (1-\kappa)\epsilon(s^t)^{\rho-1}\right]^{\frac{\rho}{\rho-1}} y^s(s^t)$$

As a result, $y^s(s^t)$ satisfies:

$$y^s(s^t) = \frac{Y(s^t)}{\left[\kappa + (1-\kappa)\epsilon(s^t)^{\rho-1}\right]^{\frac{\rho}{\rho-1}}} = \frac{\Delta(\epsilon(s^t))A(s_t)L(s^t)}{\left[\kappa + (1-\kappa)\epsilon(s^t)^{\rho-1}\right]^{\frac{\rho}{\rho-1}}}$$

where $\Delta(\epsilon)$ is defined in (39). This implies:

$$y^s(s^t) = A(s_t)L(s^t)\frac{\left[\kappa\epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}}{\left[\kappa\epsilon(s^t)^{-\rho} + (1-\kappa)\right]\left[\kappa + (1-\kappa)\epsilon(s^t)^{\rho-1}\right]^{\frac{\rho}{\rho-1}}}$$

Thus

$$y^s(s^t) = A(s_t)L(s^t)\frac{1}{\kappa\epsilon(s^t)^{-\rho} + (1-\kappa)}\left(\frac{\left[\kappa\epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}}{\epsilon(s^t)^\rho\left[\kappa\epsilon(s^t)^{-(\rho-1)} + (1-\kappa)\right]^{\frac{\rho}{\rho-1}}}\right)$$

Thus

$$y^s(s^t) = A(s_t)L(s^t)\frac{1}{\kappa + (1-\kappa)\epsilon(s^t)^\rho}$$

Substituting this expression for $y^s(s^t)$ into (105), we obtain:

$$\frac{\Pi(s^t)}{P(s^t)} = -\frac{U_L^m(s^t)L(s^t)}{U_C^m(s^t)}\frac{1+\tau_c}{1-\tau_\ell}\left[(1-\kappa)\frac{1}{\rho-1}\epsilon(s^t)^\rho + \kappa\left(\frac{\rho}{\rho-1}\epsilon(s^t) - 1\right)\right]\frac{1}{\kappa + (1-\kappa)\epsilon(s^t)^\rho}$$

Therefore, real profits, $\Pi(s^t)/P(s^t)$, satisfy:

$$\frac{\Pi(s^t)}{P(s^t)} = -\frac{1+\tau_c}{1-\tau_\ell}\frac{U_L^m(s^t)L(s^t)}{U_C^m(s^t)}\Phi(\epsilon(s^t)) \tag{106}$$

where the function Φ is defined in (48). Finally, the implementability conditions in (47) follow from combining condition (46) with our expression for real profits in (106).

Sufficiency. Need to add.

B.3 Properties of the Φ function

The function Φ, defined in (48), can be rewritten as follows:

$$\Phi(\epsilon) = \frac{\alpha\frac{1}{\rho-1}\epsilon^\rho + \frac{\rho}{\rho-1}\epsilon - 1}{1 + \alpha\epsilon^\rho}$$

where $\alpha \equiv \frac{1-\kappa}{\kappa} > 0$. Then

$$\Phi'(\epsilon) = \frac{(1+\alpha\epsilon^{\rho})\left(\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1} + \frac{\rho}{\rho-1}\right) - \left(\alpha\frac{\epsilon^{\rho}}{\rho-1} + \frac{\rho\epsilon}{\rho-1} - 1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})^2}$$

Furthermore, the second derivative is given by:

$$\Phi''(\epsilon) =$$

Evaluating these at $\epsilon = 1$ we get:

$$\Phi(1) = \frac{1}{\rho-1} > 0, \qquad \Phi'(1) = \kappa\frac{\rho}{\rho-1} > 0, \quad \text{and} \qquad \Phi''(1) = -2\frac{\rho^2\kappa(1-\kappa)}{\rho-1} < 0.$$

B.4 Ramsey Optimum

We solve the Ramsey problem given an arbitrary set of Pareto weights. We again let $\beta^t\mu(s^t)\xi(s^t)$ and $\beta^t\mu(s^{t-1})\upsilon(s^{t-1})$ denote the Lagrange multipliers on the implementability conditions (30) and (31), respectively. We let $\beta^t\mu(s^t)\varsigma(s^t)$ denote the the Lagrange multipliers on the aggregate resource constraint in (38). We obtain the following characterization.

Proposition 9. *Given Assumption 1, an allocation x^* is a Ramsey optimum if, for all $s^t \in S^t$,*

$$-\frac{\mathcal{W}^{\sigma}_L(s^t) + \xi(s^t)U^m_{LL}(s^t) + \upsilon(s^{t-1})\left[U^m_{LL}(s^t)L(s^t) + U^m_L(s^t)\right]F(\epsilon(s^t))}{\mathcal{W}^{\sigma}_C(s^t) + \xi(s^t)U^m_{CC}(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)} = \frac{Y(s^t)}{L(s^t)}, \qquad \forall s^t \in S^t \tag{107}$$

and

$$\frac{\xi(s^t)}{L(s^t)} = \frac{\mathcal{W}^{\sigma}_{\epsilon}(s^t)/L(s^t) + \varsigma(s^t)A(s_t)\Delta'(\epsilon(s^t)) + \upsilon(s^{t-1})U^m_L(s^t)F'(\epsilon(s^t))}{U^m_C(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)} \tag{108}$$

where $\varsigma(s^t) > 0$, and $F(\epsilon(s^t))$ and $F'(\epsilon(s^t))$ given by (87) and (88).

Proof. Following the same steps as in the proof of Proposition 8, we can show that the equilibrium constraint in (31) is equivalent to (91) with the function F as defined in (87).

We then write the planner's Lagrangian as follows:

$$\begin{aligned}
\mathcal{L}^{\sigma} = &\sum_t\sum_{s^t}\beta^t\mu(s^t)\mathcal{W}^{\sigma}(C(s^t), L(s^t), \epsilon(s^t), s_t; \varphi, \nu, \lambda, \sigma) - U^m_C(s_0)\bar{T}\sum_{i\in I}\pi^i\nu^i \\
&+ \sum_t\sum_{s^t}\beta^t\mu(s^t)\xi(s^t)\left\{U^m_C(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t) + U^m_L(s^t)\right\} \\
&+ \sum_t\sum_{s^t}\beta^t\mu(s^t)\varsigma(s^t)\left\{A(s_t)\Delta(\epsilon(s^t))L(s^t) - C(s^t) - G(s_t)\right\} \\
&+ \sum_t\sum_{s^t}\beta^{t+1}\mu(s^t)\upsilon(s^t)\sum_{s^t|s^{t-1}}U^m_L(s^{t+1})L(s^{t+1})F(\epsilon(s^{t+1}))\mu(s^{t+1}|s^t)
\end{aligned}$$

with complementary slackness conditions (92) and (92).

The FOCs wrt $C(s^t)$ and $L(s^t)$ are given by:

$$0 = \mathcal{W}^{\sigma}_{C}(s^t) + \xi(s^t)U^m_{CC}(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t) - \varsigma(s^t),$$
$$0 = \mathcal{W}^{\sigma}_{L}(s^t) + \xi(s^t)U^m_{LL}(s^t) + \varsigma(s^t)\frac{Y(s^t)}{L(s^t)} + \upsilon(s^{t-1})\left[U^m_{LL}(s^t)L(s^t) + U^m_{L}(s^t)\right] F(\epsilon(s^t)),$$

Combining () and (), as in our previous analysis, we obtain:

$$\begin{aligned}0 =&\mathcal{W}^{\sigma}_{L}(s^t) + \xi(s^t)U^m_{LL}(s^t) + \upsilon(s^{t-1})\left[U^m_{LL}(s^t)L(s^t) + U^m_{L}(s^t)\right] F(\epsilon(s^t))\\ &+\left[\mathcal{W}^{\sigma}_{C}(s^t) + \xi(s^t)U^m_{CC}(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t)\right]\frac{Y(s^t)}{L(s^t)},\end{aligned}$$

Rearranging, we obtain the expression in 107.

Next, the FOC with respect to $\epsilon(s^t)$ is given by:

$$\begin{aligned}0 =&\mathcal{W}^{\sigma}_{\epsilon}(s^t) - \xi(s^t)U^m_{C}(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t) + \varsigma(s^t)A(s_t)\Delta'(\epsilon(s^t))L(s^t)\\ &+ \upsilon(s^{t-1})U^m_{L}(s^t)L(s^t)F'(\epsilon(s^t)).\end{aligned}$$

where $F'(\epsilon(s^t))$ is given in (88). Rearranging, we obtain the expression in 108. □

B.5 Proof of Theorem 3.

Consider first the Ramsey planner's intratemporal optimality condition in 107. Iso-elastic preferences imply that we may rewrite 107 as follows:

$$-\frac{\mathcal{W}^{\sigma}_{L}(s^t) + \eta\xi(s^t)U^m_{L}(s^t)L(s^t)^{-1} + (1+\eta)\upsilon(s^{t-1})U^m_{L}(s^t)F(\epsilon(s^t))}{\mathcal{W}^{\sigma}_{C}(s^t) - \gamma\xi(s^t)U^m_{C}(s^t)C(s^t)^{-1}\chi\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} A(s_t)} = \frac{Y(s^t)}{L(s^t)}$$

Next, $\mathcal{W}^{\sigma}_{C}(s^t)$ remains the same as before:

$$\mathcal{W}^{\sigma}_{C}(s^t) = \mathcal{W}_{C}(s^t) = U^m_{C}(s^t)\sum_{i\in I}\pi^i\omega^i_C(\varphi)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1-\gamma)\right]$$

While $\mathcal{W}^{\sigma}_{L}(s^t)$ is now given by:

$$\mathcal{W}^{\sigma}_{L}(s^t) = \mathcal{W}_{L}(s^t) + \delta\left[U^m_{LL}(s^t)L(s^t) + U^m_{L}(s^t)\right]\Phi(\epsilon(s^t))\sum_{i\in I}\pi^i\nu^i\sigma^i$$

Therefore

$$\mathcal{W}^{\sigma}_{L}(s^t) = U^m_{L}(s^t)\sum_{i\in I}\pi^i\omega^i_L(\varphi, s^t)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right] + (1+\eta)\delta U^m_{L}(s^t)\Phi(\epsilon(s^t))\sum_{i\in I}\pi^i\nu^i\sigma^i$$

Relative to our previous analysis, there is an extra term in $\mathcal{W}_L^\sigma(s^t)$. This extra term is due to the fact that aggregate labor affects profits and hence the budget constraints of houseeholds. Substituting these expressions for $\mathcal{W}_C^\sigma(s^t)$ and $\mathcal{W}_L^\sigma(s^t)$ into (), we get:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{\sum_{i\in I}\pi^i\omega_L^i(\varphi,s^t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]+(1+\eta)\delta\Phi(\epsilon(s^t))\sum_{i\in I}\pi^i\nu^i\sigma^i+\eta\xi(s^t)L(s^t)^{-1}+(1+\eta)\upsilon(s^{t-1})F(\epsilon(}{\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]-\gamma\xi(s^t)C(s^t)^{-1}\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}\right.$$

Therefore the optimal monetary wedge in this environment satisfies:

$$1-\tau_M^*(s^t)=\frac{(\chi^*)^{-1}\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]-\gamma\xi(s^t)C(s^t)^{-1}\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}{\sum_{i\in I}\pi^i\omega_L^i(\varphi,s^t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]+(1+\eta)\delta\Phi(\epsilon(s^t))\sum_{i\in I}\pi^i\nu^i\sigma^i+\eta\xi(s^t)L(s^t)^{-1}+(1+\eta)\upsilon(s^{t-1})F(\epsilon}$$

Defining $\mathcal{I}(s^t)$ and $\bar{\mathcal{I}}_C$ as in the previous section, and letting

$$\mathcal{V}\equiv\delta\sum_{i\in I}\pi^i\nu^i\sigma^i$$

this can be written as:

$$1-\tau_M^*(s^t)=\frac{\bar{\mathcal{I}}_C-\gamma\xi(s^t)C(s^t)^{-1}\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t)}{\mathcal{I}(s_t)+(1+\eta)\mathcal{V}\Phi(\epsilon(s^t))+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))+\eta\xi(s^t)L(s^t)^{-1}}$$

Next using the fact that

$$C(s^t)=A(s_t)\Delta(\epsilon(s^t))L(s^t)$$

we can write the optimal monetary wedge as follows

$$1-\tau_M^*(s^t)=\frac{\bar{\mathcal{I}}_C-\gamma\frac{\xi(s^t)}{L(s^t)}\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t)+\eta\frac{\xi(s^t)}{L(s^t)}+(1+\eta)\mathcal{V}\Phi(\epsilon(s^t))+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))}\tag{109}$$

where $\xi(s^t)/L(s^t)$ satisfies (108).

Next we decompose $\xi(s^t)/L(s^t)$ into three terms:

$$\frac{\xi(s^t)}{L(s^t)}=\hat{\varsigma}(s^t)+\hat{\xi}(s^t)+\hat{\beta}(s^t)$$

where $\hat{\varsigma}(s^t)$ and $\hat{\xi}(s^t)$ are defined in (101) and (102), respectively, and

$$\hat{\beta}(s^t)\equiv\frac{\mathcal{W}_\epsilon^\sigma(s^t)/L(s^t)}{U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}-1}\kappa\epsilon(s^t)^{-\rho}A(s_t)}\tag{110}$$

Note that in any sticky price equilibrium:

$$-U_L^m(s^t)=U_C^m(s^t)\chi\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}A(s_t);$$

Therefore, we can write $\hat{\beta}(s^t)$ as follows:

$$\hat{\beta}(s^t) = -\kappa^{-1}\epsilon(s^t)^\rho \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right] \frac{\mathcal{W}^\sigma_\epsilon(s^t)}{U^m_L(s^t)L(s^t)}$$

Next, the derivative of $\mathcal{W}^\sigma$ with respect to $\epsilon(s^t)$ is given by:

$$\mathcal{W}^\sigma_\epsilon(s^t) = U^m_L(s^t)L(s^t)\Phi'(\epsilon(s^t))\mathcal{V}.$$

Substituting this into $\hat{\beta}(s^t)$ we obtain:

$$\hat{\beta}(s^t) = -\kappa^{-1}\epsilon(s^t)^\rho \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right] \Phi'(\epsilon(s^t))\mathcal{V}$$

Threshold. We first consider the conditions under which $\tau^*_M(s^t) = 0$, equivalently, $\epsilon(s^t) = 1$. In this case:

$$\mathcal{I}(s_t) + \eta\frac{\xi(s^t)}{L(s^t)} + (1+\eta)\mathcal{V}\Phi(1) + (1+\eta)\upsilon(s^{t-1})F(1) = \bar{\mathcal{I}}_C - \gamma\frac{\xi(s^t)}{L(s^t)}$$

where $F(1) = 0$. Therefore

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C - (1+\eta)\mathcal{V}\Phi(1) - (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)} \tag{111}$$

Next, the $\xi(s^t)/L(s^t)$ reduces to:

$$\frac{\xi(s^t)}{L(s^t)} = \hat{\xi}^\sigma(s^t) = -\kappa^{-1}\left[\mathcal{V}\Phi'(1) + \upsilon(s^{t-1})F'(1)\right]$$

Plugging this into (111) we get:

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C - (1+\eta)\mathcal{V}\Phi(1) + (\eta+\gamma)\kappa^{-1}\left[\mathcal{V}\Phi'(1) + \upsilon(s^{t-1})F'(1)\right]$$

where

$$F'(1) = \kappa + (1-\kappa)(1-\rho) + (1-\kappa)\rho = 1$$

and

$$\Phi(1) = \frac{1}{\rho-1} > 0 \qquad \text{and} \qquad \Phi'(1) = \kappa\frac{\rho}{\rho-1} > 0$$

Plugging all of these in:

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C - (1+\eta)\frac{1}{\rho-1}\mathcal{V} + (\eta+\gamma)\frac{\rho}{\rho-1}\mathcal{V} + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1})$$

Therefore

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}_C + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1}) + \frac{1}{\rho-1}\left[(\eta+\gamma)\rho - (1+\eta)\right]\mathcal{V}$$

Therefore there exists a threshold that depends only on the past history given by:

$$\bar{\mathcal{I}}^{\sigma}(s^{t-1}) \equiv \bar{\mathcal{I}}_C + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1}) + \frac{1}{\rho-1}\left[\rho(\eta+\gamma)-(1+\eta)\right]\mathcal{V}$$

When $\mathcal{I}(s_t) = \bar{\mathcal{I}}^{\sigma}(s^{t-1})$, the optimal $\tau_M^*(s^t) = 0$.

Next, recall that

$$\bar{\mathcal{I}}(s^{t-1}) = \bar{\mathcal{I}}_C + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1})$$

Therefore:

$$\bar{\mathcal{I}}^{\sigma}(s^{t-1}) > \bar{\mathcal{I}}(s^{t-1}) \qquad \text{iff} \qquad \left[\rho(\eta+\gamma)-(1+\eta)\right]\mathcal{V} > 0$$

If $\mathcal{V}$ is positive, then:

$$\bar{\mathcal{I}}^{\sigma}(s^{t-1}) > \bar{\mathcal{I}}(s^{t-1}) \qquad \text{iff} \qquad \rho > \frac{\eta+1}{\eta+\gamma}$$

First fictitious tax wedge. We first define a fictitious tax wedge as follows:

$$1-\hat{\tau}_0^{\sigma}(s^t) \equiv \frac{\bar{\mathcal{I}}_C + \gamma\kappa^{-1}\upsilon(s^{t-1}) + \gamma\kappa^{-1}\Phi'(1)\mathcal{V}}{\mathcal{I}(s_t) - \eta\kappa^{-1}\upsilon(s^{t-1}) + \left[(1+\eta)\Phi(1) - \eta\kappa^{-1}\Phi'(1)\right]\mathcal{V}}$$

This can be rewritten as

$$1-\hat{\tau}_0^{\sigma}(s^t) = \frac{\bar{\mathcal{I}}_C + \gamma\kappa^{-1}\upsilon(s^{t-1}) + \frac{1}{\rho-1}\rho\gamma\mathcal{V}}{\mathcal{I}(s_t) - \eta\kappa^{-1}\upsilon(s^{t-1}) + \frac{1}{\rho-1}\left[1+\eta-\eta\rho\right]\mathcal{V}}$$

This wedge is unambiguously falling in $\mathcal{I}(s_t)$, as all other terms are constants. Furthermore, note that when

$$\mathcal{I}(s_t) = \bar{\mathcal{I}}^{\sigma}(s^{t-1}) \equiv \bar{\mathcal{I}}_C + \kappa^{-1}(\eta+\gamma)\upsilon(s^{t-1}) + \frac{1}{\rho-1}\left[\rho(\eta+\gamma)-(1+\eta)\right]\mathcal{V},$$

this wedge is equal to one. As a result, the fictitious tax $\hat{\tau}_0^{\sigma}(s^t)$ trivially satisfies:

$$\begin{array}{lll} \hat{\tau}_0^{\sigma}(s^t) > 0 & \text{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}^{\sigma}(s^{t-1}), \\ \hat{\tau}_0^{\sigma}(s^t) = 0 & \text{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}^{\sigma}(s^{t-1}), \\ \hat{\tau}_0^{\sigma}(s^t) < 0 & \text{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}^{\sigma}(s^{t-1}). \end{array}$$

Second fictitious tax wedge. We define a second fictitious tax as the one that jointly satisfies:

$$1-\hat{\tau}_1^{\sigma}(s^t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + \gamma\kappa^{-1}\Phi'(1)\mathcal{V}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + \left[(1+\eta)\Phi(1) - \eta\kappa^{-1}\Phi'(1)\right]\mathcal{V}}$$

where $\hat{\xi}(s^t)$ is defined in (102), and recall that it satisfies

$$\hat{\xi}(s^t) = -\kappa^{-1}\upsilon(s^{t-1})\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]\epsilon(s^t)^{\rho}F'(\epsilon(s^t))$$

The goal is to compare this to the fictitious tax wedge $1 - \hat{\tau}_0^\sigma(s^t)$.

First note that when $\epsilon(s^t) = 1$,

$$\hat{\xi}(s^t) = -\kappa^{-1}\upsilon(s^{t-1}).$$

In this case, $\hat{\tau}_1^\sigma(s^t) = \hat{\tau}_0^\sigma(s^t)$, therefore $\mathcal{I}(s_t) = \bar{\mathcal{I}}^\sigma(s^{t-1})$. Following the same steps as in the proof of Proposition ??, one can show that:

$$\begin{array}{lll} \hat{\tau}_1^\sigma(s^t) > 0 & \text{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}^\sigma(s^{t-1}), \\ \hat{\tau}_1^\sigma(s^t) = 0 & \text{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}^\sigma(s^{t-1}), \\ \hat{\tau}_1^\sigma(s^t) < 0 & \text{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}^\sigma(s^{t-1}). \end{array}$$

Third fictitious tax wedge. We define a third fictitious tax as the one that jointly satisfies:

$$1-\hat{\tau}_2^\sigma(s_t) = \frac{\bar{\mathcal{I}}_C - \gamma\left[\hat{\xi}(s^t) + \hat{\beta}(s^t)\right]\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + \eta\hat{\beta}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + (1+\eta)\mathcal{V}\Phi(\epsilon(s^t))}$$

where $\hat{\xi}(s^t)$ is defined in (102) and $\hat{\beta}(s^t)$ is defined in ?? and recall that it satisfies:

$$\hat{\beta}(s^t) = -\kappa^{-1}\epsilon(s^t)^\rho\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]\Phi'(\epsilon(s^t))\mathcal{V}$$

The goal is to compare this to the fictitious tax wedge $1 - \hat{\tau}_1^\sigma(s^t)$.

First note that when $\epsilon(s^t) = 1$,

$$\hat{\xi}(s^t) = -\kappa^{-1}\upsilon(s^{t-1}) \qquad \text{and} \qquad \hat{\beta}(s^t) = -\kappa^{-1}\Phi'(1)\mathcal{V}$$

In this case, $\hat{\tau}_2^\sigma(s_t) = \hat{\tau}_1^\sigma(s^t) = \hat{\tau}_0^\sigma(s^t)$, therefore $\mathcal{I}(s_t) = \bar{\mathcal{I}}^\sigma(s^{t-1})$.

Next, we have that

$$\hat{\beta}(s^t)\frac{\left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} = -\kappa^{-1}\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\Phi'(\epsilon(s^t))\mathcal{V}$$

Substituting this into () we get:

$$1-\hat{\tau}_2^\sigma(s_t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + \gamma\kappa^{-1}\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\Phi'(\epsilon(s^t))\mathcal{V}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + (1+\eta)\Phi(\epsilon(s^t))\mathcal{V} - \eta\kappa^{-1}\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\Phi'(\epsilon(s^t))\mathcal{V}}$$

we can rewrite this as

$$1-\hat{\tau}_2^\sigma(s_t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + \gamma\kappa^{-1}\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\Phi'(\epsilon(s^t))\mathcal{V}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + \left[(1+\eta)\Phi(\epsilon(s^t)) - \eta\kappa^{-1}\left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\Phi'(\epsilon(s^t))\right]\mathcal{V}}$$

Let

$$g(\epsilon(s^t)) \equiv \left[\kappa + (1-\kappa)\epsilon(s^t)^\rho\right]\frac{\Phi'(\epsilon(s^t))}{\Phi'(1)}$$

$$1-\hat{\tau}_2^{\sigma}(s_t)=\frac{\bar{\mathcal{I}}_C-\gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))}+\gamma\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\mathcal{V}}{\mathcal{I}(s_t)+\eta\hat{\xi}(s^t)+(1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t))+\left[(1+\eta)\Phi(\epsilon(s^t))-\eta\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\right]\mathcal{V}}$$

Note that

$$\Phi'(\epsilon)=\frac{(1+\alpha\epsilon^{\rho})\left(\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1}+\frac{\rho}{\rho-1}\right)-\left(\alpha\frac{\epsilon^{\rho}}{\rho-1}+\frac{\rho\epsilon}{\rho-1}-1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})^2}$$

where $\alpha\equiv\frac{1-\kappa}{\kappa}>0$. Then

$$g(\epsilon(s^t))\equiv\kappa\left[1+\alpha\epsilon^{\rho}\right]\frac{\Phi'(\epsilon(s^t))}{\Phi'(1)}$$

$$g(\epsilon(s^t))\equiv\kappa\left[1+\alpha\epsilon^{\rho}\right]\left[\frac{(1+\alpha\epsilon^{\rho})\left(\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1}+\frac{\rho}{\rho-1}\right)-\left(\alpha\frac{\epsilon^{\rho}}{\rho-1}+\frac{\rho\epsilon}{\rho-1}-1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})^2}\right]\frac{1}{\Phi'(1)}$$

$$g(\epsilon(s^t))\equiv\kappa\left[\frac{(1+\alpha\epsilon^{\rho})^2\left(\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1}+\frac{\rho}{\rho-1}\right)-(1+\alpha\epsilon^{\rho})\left(\alpha\frac{\epsilon^{\rho}}{\rho-1}+\frac{\rho\epsilon}{\rho-1}-1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})^2}\right]\frac{1}{\Phi'(1)}$$

$$g(\epsilon(s^t))\equiv\kappa\left[\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1}+\frac{\rho}{\rho-1}-\frac{\left(\alpha\frac{\epsilon^{\rho}}{\rho-1}+\frac{\rho\epsilon}{\rho-1}-1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})}\right]\frac{1}{\Phi'(1)}$$

where

$$\Phi'(1)=\kappa\frac{\rho}{\rho-1}>0$$

$$g(\epsilon(s^t))\equiv\left[\alpha\frac{\rho\epsilon^{\rho-1}}{\rho-1}+\frac{\rho}{\rho-1}-\frac{\left(\alpha\frac{\epsilon^{\rho}}{\rho-1}+\frac{\rho\epsilon}{\rho-1}-1\right)\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})}\right]\frac{\rho-1}{\rho}$$

$$g(\epsilon(s^t))\equiv\left[\alpha\rho\epsilon^{\rho-1}+\rho-\frac{(\alpha\epsilon^{\rho}+\rho\epsilon-\rho+1)\,\alpha\rho\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})}\right]\frac{1}{\rho}$$

$$g(\epsilon(s^t))\equiv\left[\alpha\epsilon^{\rho-1}+1-\frac{(\alpha\epsilon^{\rho}+\rho\epsilon-\rho+1)\,\alpha\epsilon^{\rho-1}}{(1+\alpha\epsilon^{\rho})}\right]$$

$$g(\epsilon(s^t))\equiv\left[\alpha\epsilon^{\rho-1}(1+\alpha\epsilon^{\rho})+(1+\alpha\epsilon^{\rho})-(\alpha\epsilon^{\rho}+\rho\epsilon-\rho+1)\,\alpha\epsilon^{\rho-1}\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t))\equiv\left[\alpha\epsilon^{\rho-1}+\alpha\epsilon^{\rho}\alpha\epsilon^{\rho-1}+1+\alpha\epsilon^{\rho}-\left(\alpha\epsilon^{\rho}\alpha\epsilon^{\rho-1}+\rho\epsilon\alpha\epsilon^{\rho-1}-\rho\alpha\epsilon^{\rho-1}+\alpha\epsilon^{\rho-1}\right)\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t))\equiv\left[\alpha\epsilon^{\rho-1}+\alpha\epsilon^{\rho}\alpha\epsilon^{\rho-1}+1+\alpha\epsilon^{\rho}-\alpha\epsilon^{\rho}\alpha\epsilon^{\rho-1}-\rho\epsilon\alpha\epsilon^{\rho-1}+\rho\alpha\epsilon^{\rho-1}-\alpha\epsilon^{\rho-1}\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t))\equiv\left[1+\alpha\epsilon^{\rho}-\rho\epsilon\alpha\epsilon^{\rho-1}+\rho\alpha\epsilon^{\rho-1}\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t))\equiv\left[1+\alpha\epsilon^{\rho}-\rho\alpha\epsilon^{\rho}+\rho\alpha\epsilon^{\rho-1}\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t))\equiv\left[1+(1-\rho)\alpha\epsilon^{\rho}+\rho\alpha\epsilon^{\rho-1}\right]\frac{1}{(1+\alpha\epsilon^{\rho})}$$

$$g(\epsilon(s^t)) < 1$$

if and only if

$$\left[1 + (1-\rho)\alpha\epsilon^\rho + \rho\alpha\epsilon^{\rho-1}\right] \frac{1}{(1+\alpha\epsilon^\rho)} < 1$$

$$1 + (1-\rho)\alpha\epsilon^\rho + \rho\alpha\epsilon^{\rho-1} < 1 + \alpha\epsilon^\rho$$

$$1 - \rho\alpha\epsilon^\rho + \rho\alpha\epsilon^{\rho-1} < 1$$

$$-\rho\alpha\epsilon^\rho + \rho\alpha\epsilon^{\rho-1} < 0$$

$$\rho\alpha\epsilon^{\rho-1} < \rho\alpha\epsilon^\rho$$

$$\epsilon^{\rho-1} < \epsilon^\rho$$

$$\epsilon^{-1} < 1$$

$$1 < \epsilon$$

we want to compare:

$$1 - \hat{\tau}_2^\sigma(s_t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + \gamma\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\mathcal{V}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + \left[(1+\eta)\Phi(\epsilon(s^t)) - \eta\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\right]\mathcal{V}}$$

to:

$$1 - \hat{\tau}_1^\sigma(s^t) = \frac{\bar{\mathcal{I}}_C - \gamma\hat{\xi}(s^t)\frac{[\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)]^{-\frac{1}{1-\rho}}}{\Delta(\epsilon(s^t))} + \gamma\kappa^{-1}\Phi'(1)\mathcal{V}}{\mathcal{I}(s_t) + \eta\hat{\xi}(s^t) + (1+\eta)\upsilon(s^{t-1})F(\epsilon(s^t)) + \left[(1+\eta)\Phi(1) - \eta\kappa^{-1}\Phi'(1)\right]\mathcal{V}}$$

Suppose $\mathcal{V}$ is positive. We consider the case in which $\epsilon(s^t) > 1$. When $\epsilon(s^t) > 1$:

$$g(\epsilon(s^t)) < 1$$

which implies

$$\gamma\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\mathcal{V} < \gamma\kappa^{-1}\Phi'(1)\mathcal{V}$$

Therefore the numerator of $1 - \hat{\tau}_2^\sigma(s_t)$ is strictly less than the numerator of $1 - \hat{\tau}_1^\sigma(s^t)$.

Furthermore when $\epsilon(s^t) > 1$:

$$g(\epsilon(s^t)) < 1 \qquad \text{and} \qquad \Phi(\epsilon(s^t)) > \Phi(1)$$

Together, these imply:

$$\left[(1+\eta)\Phi(\epsilon(s^t)) - \eta\kappa^{-1}g(\epsilon(s^t))\Phi'(1)\right]\mathcal{V} > \left[(1+\eta)\Phi(1) - \eta\kappa^{-1}\Phi'(1)\right]\mathcal{V}$$

Therefore the denominator of $1 - \hat{\tau}_2^\sigma(s_t)$ is strictly greater than the numerator of $1 - \hat{\tau}_1^\sigma(s^t)$.

Therefore, when $\mathcal{V}$ is positive and $\epsilon(s^t) > 1$, the following is true: $1 - \hat{\tau}_2^\sigma(s_t) < 1 - \hat{\tau}_1^\sigma(s^t)$. This implies:

$$\hat{\tau}_1^\sigma(s^t) < \hat{\tau}_2^\sigma(s_t)$$

but it could still be the case that $\hat{\tau}_1^\sigma(s^t) < 0$?

B.6 Proof of Proposition 10.

Furthermore, the presence of heterogeneous profits also changes the rate at which the optimal implicit tax responds to increasing inequality. When $\mathcal{C}$ is negative, the optimal markup is (locally) less responsive to increases in inequality than when shares are homogeneous. Intuitively, if targeting higher markup leads to both a more equitable labor income distribution and a less equitable profit income distribution, the monetary authority will require an incrementally higher level of labor market inequality to justify increasing profits. These results are summarize in the following Proposition.

Proposition 10. *Define the derivative of the optimal markup with respect to inequality evaluated at the flexible price threshold,* $\Theta_{ho} \equiv \frac{d\mathcal{M}^*}{d\mathcal{I}}(\bar{\mathcal{I}}(s^{t-1}))$ *when profits are homogeneous. Define an analogous term when profits are heterogeneous,* $\Theta_{he} \equiv \frac{d\mathcal{M}^*}{d\mathcal{I}}(\bar{\mathcal{I}}^{\sigma}(s^{t-1}))$*. If* $\lambda^i = 1 \; \forall \; i \in I$*, and* s^t *affects the relative skill distribution.*

$$\begin{array}{lll} \Theta_{he} < \Theta_{ho} & \textit{if and only if} & \mathcal{C} < 0, \\ \Theta_{he} = \Theta_{ho} & \textit{if and only if} & \mathcal{C} = 0, \\ \Theta_{he} > \Theta_{ho} & \textit{if and only if} & \mathcal{C} > 0. \end{array}$$

Homogeneous Profits Case. Equation X states that the optimal monetary tax should be set according to the following.

$$1 - \tau^* = \frac{\bar{\mathcal{I}}(s^{t-1}) - \gamma \frac{\xi(s^t)}{L(s^t)} A(s^t) L(s^t) C(s^t)^{-1} [\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{-\frac{1}{1-\rho}}}{\mathcal{I}(s^t) + \eta \xi(s^t) L(s^t)^{-1}}$$

$$1 - \tau^* = \frac{\bar{\mathcal{I}}(s^{t-1}) - \gamma \frac{\xi(s^t)}{L(s^t)} \Delta(\epsilon(s^t))^{-1} [\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{-\frac{1}{1-\rho}}}{\mathcal{I}(s^t) + \eta \frac{\xi(s^t)}{L(s^t)}}$$

Where:

$$\begin{aligned} & \Delta(\epsilon(s^t))^{-1} [\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{-\frac{1}{1-\rho}} \\ = & \frac{[\kappa \epsilon(s^t)^{-\rho} + (1-\kappa)]}{[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{\frac{\rho}{\rho-1}}} [\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{\frac{1}{\rho-1}} \\ & \phi(\epsilon(s^t)) = \frac{[\kappa \epsilon(s^t)^{-\rho} + (1-\kappa)]}{[\kappa \epsilon(s^t)^{1-\rho} + (1-\kappa)]^{\rho}} \end{aligned}$$

So:

$$1 - \tau^* = \Delta(\epsilon(s^t))^{-1} [\kappa \epsilon(s^t)^{1-\rho} + 1 - \kappa]^{\frac{1}{\rho-1}} = \phi(\epsilon(s^t)) = \frac{\bar{\mathcal{I}}(s^{t-1}) - \gamma \frac{\xi(s^t)}{L(s^t)} \phi(\epsilon(s^t))}{\mathcal{I}(s^t) + \eta \frac{\xi(s^t)}{L(s^t)}}$$

Recall that:

$$\frac{\xi}{L}(\epsilon(s^t)) = \frac{\zeta(s^t)\Delta'(\epsilon(s^t))\epsilon(s^t)^{\rho}[\kappa\epsilon(s^t)^{1-\rho}+1-\kappa]^{1+\frac{1}{1-\rho}}}{\chi^*\kappa U_C^m(s^t)}$$

$$\frac{\xi}{L}(\epsilon(s^t)) = \frac{\zeta(s^t)\Delta'(\epsilon(s^t))\Xi(\epsilon(s^t))}{\chi^*\kappa U_C^m(s^t)}$$

Here I've defined, $\Xi(\epsilon(s^t)) = \epsilon(s^t)^{\rho}[\kappa\epsilon(s^t)^{1-\rho}+1-\kappa]^{1+\frac{1}{1-\rho}}$.
Note that $\Xi(1) = 1$ and:

$$\Xi'(\epsilon(s^t)) = \rho\epsilon(s^t)^{\rho-1}[\kappa\epsilon(s^t)^{1-\rho}+1-\kappa]^{1+\frac{1}{1-\rho}} + \epsilon(s^t)^{\rho}\frac{2-\rho}{1-\rho}\kappa(1-\rho)\epsilon(s^t)^{-\rho} * [\kappa\epsilon(s^t)^{1-\rho}+1-\kappa]^{\frac{1}{1-\rho}}$$

$$\Xi'(1) = \rho + 2\kappa - \rho\kappa$$

Define the following implicit function:

$$\begin{aligned} F(\epsilon(s^t),\mathcal{I}(s^t)) &= \phi(\epsilon(s^t))[\mathcal{I}(s^t) + \eta\frac{\xi(s^t)}{L(s^t)}] - \bar{\mathcal{I}}(s^{t-1}) + \gamma\frac{\xi(s^t)}{L(s^t)}\phi(\epsilon(s^t)) \\ &= \phi(\epsilon(s^t))[\mathcal{I}(s^t) + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}] - \bar{\mathcal{I}}(s^{t-1}) \end{aligned}$$

By the IFT, $\frac{d\epsilon(s^t)}{d\mathcal{I}(s^t)} = -\frac{F_{\mathcal{I}}}{F_{\epsilon}}$

$$F_{\mathcal{I}} = \phi(s^t)$$

$$F_{\epsilon} = \phi'(\epsilon(s^t))[\mathcal{I}(s^t) + (\gamma+\eta)\frac{\xi(s^t)}{L(s^t)}] + \phi(\epsilon(s^t))[(\gamma+\eta)\frac{\xi(s^t)}{L(s^t)}'(\epsilon)]$$

Note that $\xi(s^t)/L(s^t) = 0$ at $\epsilon = 1$. Evaluate $\frac{d\epsilon(s^t)}{d\mathcal{I}(s^t)}$ at the point, $(\epsilon(s^t) = 1, \mathcal{I}(s^t) = \bar{\mathcal{I}}(s^{t-1}))$:

$$-1 * (\phi'(1)\bar{\mathcal{I}} + (\gamma+\eta)\frac{\xi}{L}'(1))^{-1} > 0$$

Where:

$$\frac{\xi}{L}'(1) = \frac{\zeta(s^t)\Xi(1)\Delta''(1)}{\chi^*\kappa U_C^m(s^t)} = \frac{\zeta(s^t)\Delta''(1)}{\chi^*\kappa U_C^m(s^t)} < 0$$

In the heterogeneous agents case,

$$\phi(\epsilon(s^t)) = \frac{\bar{\mathcal{I}}(s^{t-1}) - \gamma\frac{\xi(s^t)}{L(s^t)}\phi(\epsilon(s^t))}{\mathcal{I}(s^t) + (1+\eta)\delta\Phi(\epsilon(s^t))\mathcal{C} + \eta\frac{\xi(s^t)}{L(s^t)}}$$

and we have:

$$\frac{\xi}{L}(\epsilon(s^t)) = \frac{\epsilon(s^t)^\rho[\kappa\epsilon(s^t)^{1-\rho}+1-\kappa]^{1+\frac{1}{1-\rho}}}{\chi^*\kappa U_C^m(s^t)}\left(\zeta(s^t)\Delta'(\epsilon(s^t)) + \delta U_L^m(s^t)\Phi'(\epsilon(s^t))\mathcal{C}A(s^t)^{-1}\right)$$

$$= \frac{\Xi(s^t)\zeta(s^t)\Delta'(\epsilon(s^t))}{\chi^*\kappa U_C^m(s^t)} + \frac{\Xi(s^t)}{\kappa}\delta\Phi'(\epsilon(s^t))\mathcal{C}\frac{U_L^m(s^t)}{U_C^m(s^t)A(s^t)\chi^*}$$

$$= \frac{\Xi(s^t)\zeta(s^t)\Delta'(\epsilon(s^t))}{\chi^*\kappa U_C^m(s^t)} - \frac{\mathcal{C}\delta}{\epsilon}\left(\frac{\Xi(s^t)}{\kappa}\Phi'(\epsilon(s^t))[\kappa\epsilon^{1-\rho}+1-\kappa]^{\frac{1}{\rho-1}}\right)$$

Note that now, $\frac{\xi}{L}(1)$ is:

$$-\frac{\delta\Phi'(1)}{\kappa}\mathcal{C} > 0 \qquad \text{when} \qquad \mathcal{C} < 0$$

Now define the implicit function:

$$F^\sigma(\epsilon(s^t), \mathcal{I}(s^t)) = \phi(\epsilon(s^t))[\mathcal{I}(s^t) + (1+\eta)\delta\Phi(\epsilon(s^t))\mathcal{C} + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}] - \bar{\mathcal{I}}(s^{t-1})$$

Again, $F^\sigma_{\mathcal{I}} = \phi(\epsilon(s^t)) = 1$ when $\epsilon = 1$.

$$F^\sigma_\epsilon = \phi'(\epsilon)[\mathcal{I}(s^t) + (1+\eta)\delta\Phi(\epsilon(s^t))\mathcal{C} + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}] +$$
$$\phi(\epsilon(s^t)[(1+\eta)\delta\Phi'(\epsilon(s^t))\mathcal{C} + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}'(\epsilon)]$$

Evaluate at $(\epsilon(s^t) = 1, \mathcal{I}(s^t) = \bar{\mathcal{I}}^\sigma)$

$$F^\sigma_\epsilon = \phi'(1)[\bar{\mathcal{I}}^\sigma(s^{t-1}) + (1+\eta)\delta\Phi(1)\mathcal{C} - (\eta+\gamma)\delta\Phi'(1)\kappa^{-1}\mathcal{C}] +$$
$$[(1+\eta)\delta\Phi'(1)\mathcal{C} + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}'(1)]$$

$$= \phi'(1)\bar{\mathcal{I}}(s^{t-1}) + (1+\eta)\delta\frac{\rho\kappa}{\rho-1}\mathcal{C} + (\eta+\gamma)\frac{\xi(s^t)}{L(s^t)}'(1)$$

$$\frac{\xi}{L}'(1) = \frac{\Delta''(1)\Xi(1)\zeta(s^t)}{\chi^*\kappa U_C^m(s^t)} - \frac{\mathcal{C}\delta}{\omega}\left(\Phi''(1)\Xi(1) + \Phi'(1)\Xi'(1) - \Phi'(1)\Xi(1)\right)$$

$$\frac{\xi}{L}'(1) = \frac{\Delta''(1)\zeta(s^t)}{\chi^* \kappa U_C^m(s^t)} - \frac{\mathcal{C}\delta}{\kappa}\bigg(\Phi''(1) + \Phi'(1)(\Xi'(1) - 1)\bigg)$$

Homogeneous Case.

$$\frac{d\epsilon}{d\mathcal{I}}_{ho} = -1 * \bigg(\phi'(1)\bar{\mathcal{I}}(s^{t-1}) + (\gamma + \eta)\frac{\xi}{L}'(1)\bigg)^{-1} > 0$$

Where:

$$\frac{\xi}{L}'(1) = \frac{\zeta(s^t)\Xi(1)\Delta''(1)}{\chi^* \kappa U_C^m(s^t)} = \frac{\zeta(s^t)\Delta''(1)}{\chi^* \kappa U(s^t)} < 0$$

Heterogeneous Case.

$$\frac{d\epsilon}{d\mathcal{I}}_{he} = -1 * \bigg(\phi'(1)\bar{\mathcal{I}}(s^{t-1}) + (1+\eta)\delta\frac{\rho\kappa}{\rho - 1}\mathcal{C} + (\eta + \gamma)\frac{\xi(s^t)}{L(s^t)}'(1)\bigg)^{-1} > 0$$

Where.

$$\frac{\xi}{L}'(1) = \frac{\Delta''(1)\zeta(s^t)}{\chi^* \kappa U_C^m(s^t)} - \frac{\mathcal{C}\delta}{\kappa}\bigg(\Phi''(1) + \Phi'(1)(\Xi'(1) - 1)\bigg)$$

$$\frac{\xi}{L}'(1) = \frac{\Delta''(1)\zeta(s^t)}{\chi^* \kappa U_C^m(s^t)} - \frac{\mathcal{C}\delta}{\kappa}\bigg((\rho^2(\kappa - 1) - \rho\kappa(1 - 2\kappa))(\rho - 1)^{-1}\bigg)$$

Note that:

$$\frac{d\mathcal{I}}{d\epsilon}_{he} = \frac{d\mathcal{I}}{d\epsilon}_{ho} - \mathcal{C}\delta\bigg(\frac{((1+\eta) + (\eta + \gamma)(1 - 2\kappa))\rho\kappa + \rho^2(1 - \kappa)}{\rho - 1}\bigg)$$

Which means that the derivative is *smaller* if $\mathcal{C}$ is negative and *larger* if $\mathcal{C}$ is positive.