NBER WORKING PAPER SERIES

OPTIMAL MONETARY POLICY WITH REDISTRIBUTION

Jennifer La'O
Wendy A. Morrison

Working Paper 32921
http://www.nber.org/papers/w32921

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2024

We thank Joao Guerreiro, Paolo Martellini, Victor Rios-Rull, and Elisa Rubbo for their excellent discussions and constructive comments on our paper. We furthermore thank Fernando Alvarez, Adrien Auclert, Marina Azzimonti, Marco Bassetto, Anmol Bhandari, V.V. Chari, Eduardo Dávila, Mike Golosov, Greg Kaplan, Pat Kehoe, Rishabh Kirpalani, Juanpa Nicolini, Andreas Schaab, Balint Szoke, Harald Uhlig, Ivan Werning, Nicolas Werquin, and Mike Woodford for insightful comments, feedback, and suggestions. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System or the National Bureau of Economic Research.

Optimal Monetary Policy with Redistribution
Jennifer La'O and Wendy A. Morrison
NBER Working Paper No. 32921
September 2024
JEL No. D61, D63, E32, E52, E63, H21

ABSTRACT

We study optimal monetary policy in a general equilibrium economy with heterogeneous agents and nominal rigidities. Households differ in type-specific, state-contingent labor productivity and initial firm ownership, yet markets are complete. The fiscal authority has access to a linear tax schedule with non-state-contingent tax rates and uniform, lump-sum taxes (or transfers). We derive sufficient conditions under which implementing flexible-price allocations is optimal. We then show that when there are fluctuations in relative labor productivity across households, it is optimal for monetary policy to abandon the flexible-price benchmark and target a state-contingent markup. The optimal markup covaries positively with a sufficient statistic for labor income inequality. In a calibrated version of the model, countercyclical earnings inequality implies countercyclical optimal markups.

Jennifer La'O
Columbia University
Department of Economics
1014C International Affairs Building
420 West 118th Street
New York, NY 10027
and NBER
jenlao@columbia.edu

Wendy A. Morrison
Federal Reserve Bank of Minneapolis
90 Hennepin Ave
Minneapolis, MN 55401
wam2133@columbia.edu

1 Introduction

Empirical studies find large, systematic, and forecastable differences in labor earnings profiles across households. One prominent feature of the data is the unequal exposure of household earnings to the business cycle: the labor earnings of low-income households exhibit a greater covariance with aggregate fluctuations than those of mid- and high-income households.[1] This heterogeneity in the covariance of individual earnings with aggregate fluctuations contributes to countercyclical earnings inequality. Notably, these differences are to a large extent systematic: according to Guvenen, Ozkan, and Song (2014), "fortunes during recessions are predictable by observable characteristics before the recession."

Monetary policy is not an ideal tool for achieving distributional objectives—questions of redistribution typically fall within the purview of the fiscal authority. Yet countercyclical earnings inequality calls this division of labor into question: if fiscal instruments cannot respond to short-run movements in the labor earnings distribution, monetary instruments might be the next best alternative. That said, even if monetary policymakers were to respond to business cycle variation in the labor earnings distribution, it is not obvious from a theoretical perspective whether monetary policy should do so and in what manner.

What is the optimal conduct of monetary policy in light of distributional concerns? In this paper we seek to answer this question. We study a dynamic, general equilibrium economy with systematic heterogeneity across households, nominal rigidities, and complete markets. By assuming complete markets, we focus on *ex ante* heterogeneity, rather than ex post, and therefore on the question of redistribution rather than a lack of insurance.[2] We follow the Ramsey approach: given a restricted set of tax instruments, we solve for optimal monetary and fiscal policy jointly. We find that optimal monetary policy targets a state-contingent markup; the optimal markup covaries positively with a sufficient statistic for labor income inequality.

Framework and Methodology. Our framework is 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. Types map to heterogeneous labor productivities and initial (time-0) firm ownership.

Type-specific labor productivities are state-contingent. We allow these contingencies to be fully general—they can therefore nest any exogenous labor income process. We assume that markets are complete: in every period, households can trade a complete set of Arrow securities

[1] Parker and Vissing-Jorgensen (2009); Guvenen, Ozkan, and Song (2014); Guvenen, Schulhofer-Wohl, Song, and Yogo (2017); Alves, Kaplan, Moll, and Violante (2020).

[2] By focusing on ex ante heterogeneity and complete markets, our framework stands in contrast to heterogeneous-agent New Keynesian models (HANK) as in, e.g., Kaplan, Moll, and Violante (2018). These models typically feature idiosyncratic labor income risk and incomplete asset markets. We discuss the relationship to the HANK literature below.

in addition to a nominal bond and firm equity.

A continuum of intermediate-good firms employ workers, produce differentiated goods, and face aggregate productivity shocks. These firms are monopolistically-competitive and set prices subject to a nominal rigidity. We model the nominal rigidity as an informational friction as in Mankiw and Reis (2002); Woodford (2003). For tractability we adopt a particular specification consistent with Correia, Nicolini, and Teles (2008): we assume that a fixed fraction of randomly-selected firms set their nominal prices before perfectly observing realized demand. In our baseline model we assume that firm profits are fully taxed—this is not a crucial assumption, and we relax it in an extension of the model.

The desirability and efficacy 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 state-contingent and risk-free debt in order to finance uniform, lump-sum transfers (or taxes).

We follow the Ramsey approach and allow for linear taxes on consumption, labor income, and firm revenue (sales). 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 in the Ramsey literature, we allow for state-contingent, lump-sum taxes or transfers as in Werning (2007). That is, while the fiscal authority cannot alter the slope of the tax schedule in response to shocks, it can freely adjust the intercept. Crucially, however, we restrict lump-sum taxes and transfers to be uniform across households.[3]

We adopt a utilitarian welfare function with arbitrary Pareto weights. We solve for optimal fiscal and monetary policy jointly under commitment using the primal approach (Lucas and Stokey, 1983; Chari and Kehoe, 1999). In particular we adapt the primal approach used in Werning (2007) for a flexible-price economy with heterogeneous agents, and the primal representation employed by Correia, Nicolini, and Teles (2008) and Angeletos and La'O (2020) for representative agent economies with nominal rigidities, to our setting that features both heterogeneous agents and nominal rigidities.

Main Results. Our first main result provides sufficient conditions under which it is optimal for monetary policy to implement flexible-price allocations. Specifically, we show that when preferences are separable and homothetic and shocks to the labor skill distribution are proportional—that is, when *relative* productivities across types are fixed—the Ramsey optimum can be implemented under flexible prices with the available set of fiscal instruments. The best that monetary policy can do in this case, is to replicate flexible-price allocations. It can do so by targeting price stability.

[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.

To understand this first result, consider the case in which society desires a more equal distribution of resources across households than under laissez-faire (zero taxes and flexible prices). A positive, linear tax on labor income (or consumption, or sales) is beneficial for redistribution in the following sense: although all households face the same positive marginal tax rate, high-skilled, high labor income households face a higher average tax rate than low-skilled, low labor income households. Total tax revenue finances lump-sum transfers—distributed equally across households. It follows that a higher labor income tax rate, coupled with uniform transfers, lowers inequality; see Werning (2007); Correia (2010); Hall and Rabushka (1995).

The planner optimally trades off the redistributional benefit of taxation with its distortionary cost. Under separable and homothetic preferences and proportional shocks to the labor skill distribution, both the marginal benefit and the marginal cost of distortionary taxation are constant across states. It follows that the optimal tax rate, at which marginal benefit equals marginal cost, is also constant and hence can be implemented with the available set of fiscal tools. Monetary policy, under these conditions, should play no redistributive role.

Our second result is the following: we show that under these same conditions, it is optimal for monetary policy to implement flexible price allocations *even if* tax rates are set suboptimally. We thereby generalize the first result: if preferences are separable and homothetic and shocks to the labor skill distribution are proportional, it is optimal for monetary policy to replicate flexible price allocations, *irrespective* of fiscal policy. This result highlights a key limitation of monetary policy: state-contingent monetary policy is useful to the Ramsey planner only if the marginal benefit (or cost) of distortionary taxation varies across states.

Our third and primary result concerns the more general and realistic case in which shocks to the labor skill distribution are disproportional. When relative productivities of households vary over the business cycle, the available set of fiscal instruments is insufficient to implement the Ramsey optimum. The Ramsey optimum calls for monetary policy to deviate from replicating flexible-price allocations and to target a state-contingent markup that co-varies positively with a sufficient statistic for labor income inequality. In particular, we show that the optimal markup is high when labor income inequality is high and the optimal markup is low when labor income inequality is low.[4]

To understand this, consider again the case in which society desires a more equal distribution of resources across households than under laissez-faire. A fixed linear tax rate, uniform lump-sum transfers, and state-contingent monetary policy can jointly attain a more desirable allocation by lowering overall earnings inequality at the expense of some efficiency. In particu-

[4]In an extension, we relax the non-state-contingency of tax rates and allow taxes to be set one period in advance. We obtain an explicit, closed-form solution for the derivative of the optimal monetary wedge with respect to the sufficient statistic for labor income inequality. We show that this derivative is strictly positive and strictly less than one. We furthermore provide comparative statics for this derivative with respect to the elasticity of substitution across goods and the strength of the nominal rigidity.

lar, by raising the markup when labor income inequality is high and lowering the markup when labor income inequality is low, state-contingent monetary policy compresses the *lifetime* labor earnings distribution in the desired direction. Monetary policy thereby complements the fiscal instruments, and together these tools implement the Ramsey optimum.

In our baseline model we assume that profits are fully taxed, rendering heterogeneity in initial firm ownership irrelevant. In an extension of the model, we assume only partial taxation of profits. With less than full profit taxation, heterogeneity in initial shares introduces an additional distributional channel of monetary policy. Initial equity is a lifetime claim to firm profits. By shifting firm markups, monetary policy affects equilibrium profits and, by implication, household financial wealth.

We find that our primary qualitative result on the optimal conduct of monetary policy is robust. In particular, we show that the extent to which profits are taxed and the cross-sectional covariance between financial and lifetime labor earnings changes the slope of the optimal response of monetary policy to labor income inequality—yet this slope remains positive. A constraint on profit taxation thereby does not alter the general lesson that the optimal markup covaries positively with a sufficient statistic for labor income inequality.

The contribution of this paper is primarily theoretical. However, in a simple, quantitative illustration of the model, we calibrate the labor income distribution to match estimates of "worker betas"—the percent change in household labor income growth associated with a percent change in GDP growth—from Guvenen, Schulhofer-Wohl, Song, and Yogo (2017). In this way, business cycle movements in the earnings distribution in our calibrated model directly reflect the unequal incidence of GDP fluctuations documented in the data.

We find that the optimal markup is countercyclical. In our baseline calibration, the elasticity of the optimal markup with respect to real GDP ranges from –.25 to –.39. The countercyclicality of the optimal markup stems from two features: countercyclical earnings inequality as documented in the data, and our main theoretical result that the optimal markup covaries positively with earnings inequality. The behavior of the optimal markup in our calibrated model is thereby consistent with work that documents countercyclical price markups (Bils, 1987; Rotemberg and Woodford, 1999; Bils, Klenow, and Malin, 2018) and, more generally, a countercyclical labor wedge (Hall, 1997; Chari, Kehoe, and McGrattan, 2007).

Related literature. Our paper falls squarely within the Ramsey literature on optimal fiscal and monetary policy. The standard Ramsey problem consists of choosing an optimal tax structure in a representative agent economy when only distorting taxes are available. We follow in the tradition of solving the Ramsey problem using the primal approach (Atkinson and Stiglitz, 1980; Lucas and Stokey, 1983; Chari, Christiano, and Kehoe, 1991, 1994; Chari and Kehoe, 1999). The primal approach recasts the problem of choosing optimal policy as a problem of choosing allocations subject to a set of feasibility and implementabilty constraints (Chari and Kehoe, 1999). In relation to the standard problem, our model has three distinguishing characteristics: (i) heterogeneous agents with complete markets and lump-sum transfers; (ii) nominal rigidities; and (iii) state-contingent monetary policy but non-state-contingent fiscal policy. We elaborate on these three features below.

Earlier work on optimal policy with heterogeneous agents and complete markets include Judd (1985), Chari and Kehoe (1999), Niepelt (2004), and Bassetto (2014). The paper that we are closest to in this vein is Werning (2007). Werning (2007) studies optimal fiscal policy in a model with heterogeneous agents, complete markets, a linear tax schedule, and uniform lump-sum transfers (or taxes). He focuses primarily on the question of optimal taxation in the face of government spending and aggregate technology shocks when a key motive for distortionary taxation is redistribution.

We build directly on the Werning (2007) model. Relative to Werning (2007), we study optimal monetary policy and focus specifically on how monetary policy should respond to short-run movements in the distribution of labor productivity.[5] In order to do so, we incorporate monopolistically-competitive firms, nominal rigidities, and state-contingent monetary policy. As a natural consequence our model features equilibrium firm profits; we therefore consider the effects of heterogeneous firm ownership. Despite these key differences, we rely heavily on the results, insights, and intuition provided in Werning (2007).

Second, Correia, Nicolini, and Teles (2008), Correia, Farhi, Nicolini, and Teles (2013) and Angeletos and La'O (2020) use the primal approach to characterize optimal policy in economies with nominal rigidities. We follow the same approach and build on these contributions. Relative to this work, our paper makes a methodological contribution—to the best of our knowledge, ours is the first to show how the primal approach can be used to characterize optimal monetary policy when the Ramsey optimum *does not* coincide with a flexible-price allocation.[6]

Third, we show how state-contingent monetary policy can be useful in the absence of state-contingent fiscal policy. In this sense our results are reminiscent of insights found in Chari, Christiano, and Kehoe (1991) and Chari and Kehoe (1999). These papers show, among other things, that when the only asset available to the government is non-state-contingent nominal debt, inflation can be used to make real returns state-contingent. State-contingent monetary policy therefore enables the government to use nominal debt as a shock absorber.

Aside from the aforementioned Ramsey literature, the recent heterogeneous agent New Keynesian (HANK) literature incorporates heterogeneity into the New Keynesian framework

[5] In Section 5 of Werning (2007) the author considers shocks to the labor productivity distribution and studies the optimal response of state-contingent tax rates. This section of Werning (2007) is the closest antecedent to our paper.

[6] Dávila and Schaab (2023) use a modified version of the traditional primal approach to characterize optimal monetary policy in a HANK model. In their primal form, allocations and prices are explicit control variables for the planner.

via incomplete markets; see e.g. Kaplan, Moll, and Violante (2018), McKay, Nakamura, and Steinsson (2016), and Auclert, Rognlie, and Straub (2020), among many others. In these models, of the Bewley-Imrohoroglu-Huggett-Aiyagari variety, uninsurable idiosyncratic income risk and precautionary savings give rise to an endogenous wealth distribution featuring heterogeneous marginal propensities to consume. This changes not only the primary transmission channel for monetary policy, but it moreover implies that monetary policy can play a novel role of providing insurance or facilitating greater self-insurance. A number of papers have studied optimal monetary policy in this framework, including: Bhandari, Evans, Golosov, and Sargent (2021); Acharya, Challe, and Dogra (2023); Nuño and Thomas (2022); Dávila and Schaab (2023); McKay and Wolf (2022); Le Grand, Martin-Baillon, and Ragot (2024).

In contrast to the canonical HANK model, we assume markets are complete. We thus abstract entirely from the insurance motive for monetary policy and focus solely on the redistributive motive. As noted above, empirical evidence suggest that systematic, forecastable, between-group variation accounts for a large share of the total variation in earnings growth over the business cycle (Guvenen, Ozkan, and Song, 2014; Guvenen, Schulhofer-Wohl, Song, and Yogo, 2017).[7] While no consensus exists on the exact share of variation in lifetime earnings accounted for by systematic heterogeneity—as well as by the "insurable" component of labor income shocks—a number of structural estimations place it above 50 percent and some as high as 90 percent.[8] This evidence motivates the focus of our paper.

Layout. This paper is organized as follows. In Section 2 we describe the economic environment and in Section 3 we characterize equilibrium allocations. In Section 4 we solve a relaxed planning problem and provide sufficient conditions under which implementing flexible-price allocations is optimal. In Section 5 we solve the Ramsey problem and characterize optimal monetary policy. In Section 6 we analyze an extension of the model in which profits are not fully taxed. In Section 7 we explore a calibrated version of the baseline model. Section 8 concludes. All proofs, except for those explicitly provided in the text, are found in the Appendix.

[7] Both of these studies use a large, panel data set on individual earnings from the US Social Security Administration in which the same individuals can be tracked over time. Guvenen, Ozkan, and Song (2014) find that "between-group differences are large and systematic," and that the factor structures they estimate "are consistent with countercyclical earnings inequality without needing within-group (idiosyncratic) shocks that have countercyclical variances."

[8] See, e.g., Keane and Wolpin (1997), Storesletten, Telmer, and Yaron (2004), Huggett, Ventura, and Yaron (2011), Guvenen and Smith (2014), and Heathcote, Storesletten, and Violante (2014). Keane and Wolpin (1997) find that between-type variation accounts for 90 percent of the total variance in lifetime utility. Storesletten, Telmer, and Yaron (2004) conclude that roughly half of the variance of lifetime earnings is attributable to initial heterogeneity, while Huggett, Ventura, and Yaron (2011) and Heathcote, Storesletten, and Violante (2014) instead estimate that fraction at 62 and 63 percent, respectively. Heathcote, Storesletten, and Violante (2014) furthermore find that around half of all shocks to wages are insurable and that the most important source of lifetime consumption inequality is initial heterogeneity.

2 The Environment

We study a general equilibrium model with heterogeneous agents and a form of nominal rigidity. 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 s_t. We let $\mu(s^t|s^{t-1})$ denote the probability of history s^t conditional on s^{t-1}, and with slight abuse of notation we let $\mu(s^t)$ denote the unconditional probability of history 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:

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

The parameters γ and η denote the inverse elasticity of intertemporal substitution 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 $\sum_{i \in I} \pi^i = 1$. Households are born a type and remain that type throughout their (infinite) lifetime. 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: $\ell^i(s^t) = \theta^i(s_t)h^i(s^t)$. The household maximizes its lifetime expected utility given by:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)). \tag{2}$$

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

$$\begin{aligned} &(1+\tau_c)P(s^t)c^i(s^t) + b^i(s^t) + \sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)z^i(s^{t+1}|s^t) + V(s^t)[\sigma^i(s^t) - \sigma^i(s^{t-1})] \\ &\leq (1-\tau_\ell)W(s^t)\ell^i(s^t) + (1-\tau_\Pi)(1+\sigma^i(s^{t-1}))\Pi(s^t) + z^i(s^t|s^{t-1}) + (1+i(s^{t-1}))b^i(s^{t-1}) + P(s^t)T(s^t). \end{aligned} \tag{3}$$

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 income tax rates, τ_c and τ_ℓ, respectively.

The household can borrow and save via three separate instruments. The first is a one-period, non-state-contingent nominal bond, $b^i(s^t)$, which the household can buy or sell at time t, history s^t, and which pays $(1+i(s^t))b^i(s^t)$ units of money one period later. The second is a complete set of state-contingent Arrow securities, indexed by $s^{t+1}|s^t$. 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 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 nominal bond is a redundant asset but it allows us to represent the one-period interest rate, $i(s^t)$. We assume that initial wealth from bond holdings is zero: $b^i_0 = 0$ for all $i \in I$.

The third instrument is equity: the household can buy and sell shares of a fully diversified portfolio of firms. Equity ownership is a claim to aggregate firm profits, denoted in nominal terms by $\Pi(s^t)$ and taxed at a constant rate of $\tau_\Pi \in [0,1]$. If the household enters time t, history s^t, with $1 + \sigma^i(s^{t-1})$ shares, it receives dividend $(1-\tau_\Pi)\Pi(s^t)$ per share and it can trade shares at ex-dividend price $V(s^t)$. We assume that the type-i household is endowed with $1 + \sigma^i_0$ shares at time 0, with $\sum_{i\in I} \pi^i \sigma^i_0 = 0$.

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

Household's Problem. *Given initial bond holdings and initial equity shares, the type-i household chooses a complete contingent plan for consumption, efficiency units of labor, bond holdings, equity holdings, and Arrow security holdings: $\{c^i(s^t), \ell^i(s^t), b^i(s^t), \sigma^i(s^t), (z^i(s^{t+1}|s^t))_{s^{t+1}}\}_{t\geq 0, s^t\in S^t}$, in order to maximize its lifetime expected utility* (2) *subject to its per-period budget constraint* (3) *for all $s^t \in S^t$ and no-Ponzi conditions.*

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

$$y^j(s^t) = A(s_t) n^j(s^t), \tag{4}$$

where $A(s_t)$ is an exogenous, 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 $f^j(s^t) = (1-\tau_r)p^j_t(\cdot)y^j(s^t) - W(s^t)n^j(s^t)$ where τ_r is a constant tax on firm revenue. We postpone for a moment our discussion of the nominal rigidity and the firms' optimization problem—that is, how the price $p^j_t(\cdot)$ is set.

Final good production. A representative firm produces the final good with a constant elasticity of substitution (CES) production 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 elasticity of substitution $\rho > 1$. 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 and $P(s^t)$ is the price of the final good.

The final good producer is perfectly competitive and takes prices as given. Profit maximization implies the standard CES demand function for intermediate good j:

$$y^j(s^t) = \left(\frac{p_t^j(\cdot)}{P(s^t)}\right)^{-\rho} Y(s^t), \qquad \forall s^t \in S^t. \tag{5}$$

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. We let:

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

denote nominal tax revenue collected at time t, history 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 in efficiency units, and aggregate profits, respectively. The government can issue both state-contingent and non-state-contingent debt. The government's period-t nominal budget constraint is given by:

$$(1 + i(s^{t-1}))B(s^{t-1}) + Z(s^t|s^{t-1}) + P(s^t)T(s^t) = B(s^t) + \sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)Z(s^{t+1}|s^t) + \mathcal{T}(s^t), \tag{6}$$

where $B(s^t)$ is aggregate bond issuance and $Z(s^{t+1}|s^t)$ denotes aggregate Arrow security issuance for each $s^{t+1}|s^t$.[9]

Finally, for monetary policy we assume that the monetary authority directly controls nominal aggregate demand according to the following ad hoc, cash-in-advance constraint: $M(s^t) = P(s^t)C(s^t)$. We therefore avoid well-known issues of indeterminacy. Monetary policy is state-contingent: the monetary authority can freely choose $M(s^t) > 0$ in every history.

Market Clearing. Market clearing in the goods and labor markets imply: $C(s^t) = Y(s^t)$ and $L(s^t) = \int_{j\in J} n^j(s^t)\mathrm{d}j$. Market clearing in the financial markets imply: $B(s^t) = \sum_{i\in I} \pi^i b^i(s^t)$, $Z(s^{t+1}|s^t) = \sum_{i\in I} \pi^i z^i(s^{t+1}|s^t)$ for all $s^{t+1}|s^t$, and $\sum_{i\in I} \pi^i \sigma^i(s^t) = 0$.

[9] Throughout, we abstract from the zero lower bound on the nominal interest rate.

2.1 The Nominal Rigidity

At each date t, Nature draws the state $s_t \in S$ according to probability distribution μ. The aggregate state determines period t total factor productivity and the relative skills for each type $i \in I$. Formally, we define functions $A : S \to \mathbb{R}_+$ and $\theta^i : S \to \mathbb{R}_+$, for all $i \in I$, as exogenous mappings from the state space to aggregate productivity and type-specific labor productivities.

Intermediate good firms are price-setters. We equate the nominal rigidity in our model with an informational friction, following in the tradition of Mankiw and Reis (2002) and Woodford (2003). For tractability we assume a particular specification employed by Correia, Nicolini, and Teles (2008): all firms set prices in every period, but only a subset of firms are attentive to the realized, current state.[10]

Formally, we assume that in every period a mass $\kappa \in [0, 1)$ of randomly-selected firms are inattentive, or "sticky." All other firms, of mass $1 - \kappa$, are attentive, or "flexible." We let $\mathcal{J}^s \subset \mathcal{J}$ denote the set of "sticky-price" firms and $\mathcal{J}^f \subset \mathcal{J}$ denote the set of "flexible-price" firms, with $\mathcal{J}^f = (\mathcal{J}^s)'$.

Sticky-price firms at time t are inattentive to the current state, s_t, and hence set their price based solely on their knowledge of the history of past states, s^{t-1}. We denote the price they set by $p^s_t(s^{t-1})$. The subscript t indicates that this is the nominal price set *at time* t by the sticky-price firm, even though the price function itself is measurable in s^{t-1}.

Flexible-price firms at time t are attentive to the current state, s_t, as well as the history of past states, s^{t-1}. It follows that these firms can set their price based on knowledge of the entire history, s^t. We denote the price they set by $p^f_t(s^t)$. The subscript t similarly indicates that this is the nominal price set at time t by the flexible-price firm. However, unlike the sticky-price function, the flexible-price function is measurable in s^t.

Implicit Timing Assumption. Implicit in these 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 information sets: s^{t-1}if sticky, s^t if flexible. Once nominal prices are set, the aggregate state becomes common knowledge. Given intermediate good prices, the representative final good firm purchases inputs and produces the final good, and households make their consumption, savings, and effort choices. All allocations adjust so that supply equals demand and markets clear.[11]

[10]Furthermore, by assuming the exact same nominal rigidity present in Correia, Nicolini, and Teles (2008), our equilibrium analysis becomes directly comparable.

[11]We make the simplifying assumption that all firms learn the state by the end of each period. This assumption is compatible with the notion that all firms, including sticky-price firms, can observe end-of-period equilibrium outcomes and from these endogenous objects infer the realized state at time t.

Firm problems. Given the above description of the nominal rigidity, we now state the problems of the two types of firms. We start with the flexible-price firms.

Flexible-Price Firm's Problem. *At time t, history s^t, a flexible-price firm $j \in \mathcal{J}^f$ solves:*

$$p_t^f(s^t) \in \arg\max_{p_t^j} \left\{ (1-\tau_r) p_t^j y^j(s^t) - \frac{W(s^t)}{A(s_t)} y^j(s^t) \right\}$$

subject to (5).

The flexible-price firm sets its nominal price so as to maximize firm profits, and it does so state-by-state. We state the problem of the sticky-price firms in a similar fashion.

Sticky-Price Firm's Problem. *At time t, given history s^{t-1}, a sticky-price firm $j \in \mathcal{J}^s$ solves:*

$$p_t^s(s^{t-1}) \in \arg\max_{p_t^j} \sum_{s^t|s^{t-1}} Q(s^t|s^{t-1}) \left\{ (1-\tau_r) p_t^j y^j(s^t) - \frac{W(s^t)}{A(s_t)} y^j(s^t) \right\}$$

subject to (5).

The sticky-price firm sets its nominal price so as to maximize its expectation, conditional on s^{t-1}, of the investors' valuation of firm profits.

The firm is owned by its investors. For this reason it weighs profits across states not only by the true conditional probabilities, $\mu(s^t|s^{t-1})$, but also by the stochastic discount factor of the marginal investor. Given that markets are complete, it will not matter which household's stochastic discount factor we use—in equilibrium, marginal rates of substitution between consumption across states will be equated across all households.

That said, complete markets gives us an even simpler way of stating the firm's problem: by no arbitrage we can equivalently value firm profits using the Arrow prices $Q(s^t|s^{t-1})$. While this may appear obvious, later on in our analysis we verify that the equilibrium Arrow prices indeed reflect the true conditional probabilites and the appropriate pricing kernel.

Finally, note that all firms—sticky and flexible—solve a static problem. This is because every firm is free to adjust its price in every period; it follows that no firm need take into account future periods or states when setting its current price.

2.2 Equilibrium Definition

We denote an allocation in this economy by:

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

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 $y^j(s^t) = A(s_t)n^j(s^t)$ for all $j \in \mathcal{J}$;*

$$\sum_{i \in I} \pi^i c^i(s^t) = C(s^t) = Y(s^t) = \left[\int_{j \in \mathcal{J}} y^j(s^t)^{\frac{\rho-1}{\rho}} \mathrm{d}j\right]^{\frac{\rho}{\rho-1}}; \qquad \textit{and} \tag{7}$$

$$\sum_{i \in I} \pi^i \ell^i(s^t) = L(s^t) = \int_{j \in \mathcal{J}} n^j(s^t) \mathrm{d}j \tag{8}$$

for all $s^t \in S^t$.

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

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

a price system by:

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

and a set of financial market positions by:

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

We define an equilibrium in this economy as follows.

Definition 2. *A sticky-price equilibrium is an allocation x, a price system $\mathcal{R}$, a policy $\mathcal{P}$, and a set of financial market positions $\mathcal{A}$ such that: (i) in every t, s^t: $p_t^s(s^{t-1})$ solves the sticky-price firm's problem, for all $j \in \mathcal{J}^s$; the price $p_t^f(s^t)$ solves the flexible-price firm's problem, for all $j \in \mathcal{J}^f$; (ii) the aggregate price level is given by:*

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

(iii) for all t, s^t: prices and allocations satisfy (5) for all $j \in \mathcal{J}$; (iv) given the price system and the policy, $\{c^i(s^t), \ell^i(s^t), b^i(s^t), \sigma^i(s^t), (z^i(s^{t+1}|s^t))_{s^{t+1}}\}_{t \geq 0, s^t \in S^t}$ solves the household's problem of type i, for every $i \in I$; (v) for all t, s^t: the government budget constraint is satisfied and $M(s^t) = P(s^t)C(s^t)$; 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 $\mathcal{R}$, a policy $\mathcal{P}$, and a set of financial market positions $\mathcal{A}$ such that: (i) in every t, s^t: $p_t^f(s^t)$ solves the flexible-price firm's problem, for all $j \in \mathcal{J}$; (ii) the aggregate price level given by: $P(s^t) = p_t^f(s^t)$, and parts (iii)-(vi) of Definition 2 hold.*

The flexible-price environment will serve as a natural benchmark for isolating the role of fiscal policy in this environment.

2.3 Remarks on the model

We close this section with a few general remarks on modeling choices.

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

Lump-sum taxes and transfers. In the standard, single-agent Ramsey framework, only distorting taxes are available. Lump-sum taxes—or any combination of taxes that may replicate them—are a priori ruled out; otherwise, the first best would be attainable. When instead households are heterogeneous, one can incorporate a lump-sum tax or transfer without sacrificing the earlier lessons from the Ramsey literature on optimal taxation (Werning, 2007). We follow Werning (2007) and assume the existence of a uniform lump-sum tax or transfer. It is the uniformity of the lump-sum tax or transfer *across* types that ensures that the first best is unattainable. One can think of the uniformity restriction as an informational constraint on the government: the fiscal authority cannot distinguish household types.

The lack of fiscal state-contingency. The nature of optimal monetary policy depends on the set of available fiscal instruments. We assume distortionary, linear taxation as in Ramsey. In particular, we allow for taxes on consumption, labor income, firm sales, and firm profits; we therefore do not artificially restrict *what* can be taxed in our model.

However, we constrain these tax rates to be fixed, i.e. non-state-contingent. The lack of fiscal state-contingency is what opens the door to a potential distributive role for monetary policy. State-contingency of monetary policy but non-state-contingency of tax rates is a typical assumption made in New Keynesian models; it is motivated by the notion that monetary policy is better suited to respond at the business cycle frequency than fiscal policy.

At the same time, we allow the lump-sum tax or transfer to be state-contingent. It turns out that the state-contingency of the lump-sum transfer is without loss of generality—markets are complete, the government can issue state-contingent debt, and the infinitely-lived households have rational expectations. It follows that all agents are Ricardian.

3 Equilibrium Characterization

In this section we characterize the set of equilibrium allocations and state the Ramsey problem.

3.1 Household optimality

Consider the individual household's problem. 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 are equated across households. The Negishi (1960) characterization of competitive equilibria follows.

Lemma 1. *(Negishi, 1960; Werning, 2007). For any equilibrium there exist "Negishi" or "market" weights $\varphi \equiv (\varphi^i)_{i\in I}$ with $\varphi^i \geq 0$ such that in every history $s^t \in S^t$, the individual assignments of consumption and labor solve the following static sub-problem:*

$$U^m(C(s^t), L(s^t), s_t, \varphi) \equiv \max_{(c^i(s^t), \ell^i(s^t))_{i\in I}} \sum_{i\in I} \varphi^i \pi^i U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)) \tag{10}$$

subject to

$$C(s^t) = \sum_{i\in I} \pi^i c^i(s^t), \qquad \textit{and} \qquad L(s^t) = \sum_{i\in I} \pi^i \ell^i(s^t). \tag{11}$$

Proof. See Appendix A.2. □

In any equilibrium there is 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 representative household with utility function $U^m(C, L; \varphi)$; we follow the notation in Werning (2007) and let the superscript m stand for "market." Relative prices satisfy the representative household's intratemporal and intertemporal conditions:[12]

$$-\frac{U^m_L(s^t)}{U^m_C(s^t)} = \left[\frac{1-\tau_\ell}{1+\tau_c}\right] \frac{W(s^t)}{P(s^t)}, \tag{12}$$

$$\frac{U^m_C(s^t)}{P(s^t)} = \beta(1+i(s^t)) \sum_{s^{t+1}|s^t} \mu(s^{t+1}|s^t) \frac{U^m_C(s^{t+1})}{P(s^{t+1})}, \tag{13}$$

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

for all $s^t \in S^t$, where we let $U_C^m(s^t) \equiv \partial U^m(\cdot)/\partial C(s^t)$ and $U_L^m(s^t) \equiv \partial U^m(\cdot)/\partial L(s^t)$ denote the representative household's marginal utilities with respect to aggregate consumption and labor. Condition (12) indicates that the marginal rate of substitution between aggregate consumption and aggregate labor is equal to the after-tax real wage; condition (13) is the Euler equation corresponding to the one-period nominal bond. Furthermore, the set of Arrow prices and the ex-dividend share price satisfy, respectively:

$$Q(s^{t+1}|s^t) = \frac{\beta U_C^m(s^{t+1})}{U_C^m(s^t)} \frac{P(s^t)}{P(s^{t+1})} \mu(s^{t+1}|s^t), \quad \forall s^{t+1}|s^t; \tag{14}$$

$$V(s^t) = \sum_{s^{t+1}|s^t} Q(s^{t+1}|s^t)[(1-\tau_\Pi)\Pi(s^{t+1}) + V(s^{t+1})]. \tag{15}$$

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.[13] Therefore equations (12)-(15) hold with U^i in place of U^m. This verifies our earlier claim that the Arrow prices appropriately reflect the investors' conditional probabilities and stochastic discount factor.

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 $(C(s^t), L(s^t))$, the Negishi weights φ, and the distribution $(\theta^i(s_t))_{i\in I}$ alone. With the separable and iso-elastic preferences assumed in (2), 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

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

Individual consumption and labor are thereby proportional to their aggregates.

Household i's shares of aggregate consumption and aggregate labor are given by $\omega_C^i(\varphi)$ and $\omega_L^i(\varphi, s_t)$, respectively. Its consumption share is fixed and depends only on the Negishi weights, φ, and the coefficient of relative risk aversion. Markets are complete—as a result, households insure all idiosyncratic risk and face only aggregate risk in consumption. In contrast, its share of labor is a function of the Negishi weights, φ, the Frisch elasticity of labor supply, 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 labor productivity—but it is allocated efficiently given market weights.

At every household's optimum, its lifetime budget constraint holds with equality. Using equations (12)-(15) to substitute out after-tax prices, we obtain the following conditions cor-

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

responding to the time-0 lifetime budget constraint for each household type $i \in I$:[14]

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U^m_C(s^t)\omega^i_C(\varphi)C(s^t) + U^m_L(s^t)\omega^i_L(\varphi, s_t)L(s^t)\right] \tag{18}$$
$$= U^m_C(s_0)\bar{T} + \sigma^i_0 \frac{1-\tau_\Pi}{1+\tau_c} \sum_t \sum_{s^t} \beta^t \mu(s^t) U^m_C(s^t) \frac{\Pi(s^t)}{P(s^t)},$$

where

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

The conditions in (18) indicate that for any household, its lifetime expenditure on consumption is equal to its lifetime wealth. These conditions are similar to the implementability conditions in Werning (2007), themselves reminiscent of the standard constraint in the Ramsey literature (Lucas and Stokey, 1983; Chari and Kehoe, 1999). However, in contrast to the standard implementability condition in a single-agent Ramsey economy with distortionary taxation, there are a few key differences.

The first is that in an economy with a single household type, there is a single implementability condition corresponding to the household's budget constraint (the government's budget constraint holds by Walras's law). In our economy with multiple household types, there is a *set* of implementability conditions: one for each type $i \in I$.[15]

The second key difference with the standard Ramsey framework is the existence of lump-sum taxes and transfers. As in Werning (2007), the combination of linear, distortionary taxes and uniform 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 transfers are non-targeted. To see this, note that the $\bar{T}$ on the right hand side of equation (18) represents, in part, the time-0 value of lifetime transfers; this value is the same across all types $i \in I$. It follows that the conditions in (18) are joint restrictions on equilibrium allocations.[16] Furthermore, is clear from these conditions that the assumed state-contingency of the lump-sum transfer is without loss of generality: what matters for the household's lifetime budget constraint is the value of $\bar{T}$. Ricardian equivalence holds.

The final key difference differentiates our conditions in (18) from the corresponding conditions in Werning (2007) and any model with perfect competition. In our economy,

[14] See Appendix A.3 for the full derivation of these conditions.

[15] This is true in Werning (2007) as well as in Niepelt (2004) and Bassetto (2014).

[16] In the standard single-household Ramsey framework with distortionary taxation, 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 lump sum taxes exist. It follows that we need no such restriction on the initial period consumption tax; in fact, we subsume τ_c into our definition of $\bar{T}$.

monopolistically-competitive intermediate-good firms can make equilibrium profits. Real profits, $\Pi(s^t)/P(s^t)$, enter the household's budget constraints as dividend payouts. We subsume the "common" component of dividend payouts into $\bar{T}$.

However, there is an "uncommon" component due to heterogeneity in initial endowments of equity. The final term on the right-hand side of equation (18) represents the household's heterogeneous exposure, σ_0^i, to the time-0 value of lifetime after-tax real profits. Complete markets imply that it is irrelevant whether the household holds on to its initial shares from time-0 onward, or trades them away—it is only the initial claim on firm profits that matters for its lifetime budget constraint.

The aforementioned term disappears when there is either no heterogeneity in initial equity ($\sigma_0^i = 0$ for all $i \in I$) or when profits are fully taxed ($\tau_\Pi = 1$). In order to isolate the role of labor income heterogeneity in our model, we begin with a baseline that taxes all profits.

Assumption 1. *Profits (dividends) are fully taxed:* $\tau_\Pi = 1$.

Full profit taxation renders heterogeneity in initial firm ownership irrelevant. Herein, we impose Assumption 1. Under this assumption, the final term in (18) disappears and the implementability conditions reduce to:

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi, s_t)L(s^t) \right] = U_C^m(s_0)\bar{T}, \quad \forall i \in I, \tag{20}$$

as in Werning (2007). We later discard Assumption 1 in Section 6 of the paper and study the role of untaxed dividends and heterogeneous equity shares.

3.2 Firm optimality

We now turn to the firms. The unique solution to the flexible-price firm's problem is given by:

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

Firm optimality equates marginal cost with after-tax marginal revenue. The firm's optimal price, therefore, is a constant markup over its nominal marginal cost $W(s^t)/A(s_t)$. The markup is a function of the CES parameter ρ and the marginal tax (or subsidy) on revenue.

The unique solution to the sticky-price firm's problem is similarly 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 we let $q(s^t|s^{t-1})$ denote the risk-adjusted conditional probabilities, conditional on history s^{t-1}.[17] These probabilities satisfy $\sum_{s^t|s^{t-1}} q(s^t|s^{t-1}) = 1$, by construction. Equation (22) states

[17]We provide the definition of $q(s^t|s^{t-1})$ in Appendix A.4 along with the derivation of (22).

that the sticky-price 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 conditions (21) and (22), one can infer the following relationship:

$$p_t^s(s^{t-1}) = \sum_{s^t|s^{t-1}} q(s^t|s^{t-1}) p_t^f(s^t).$$

That is, the optimal price of the sticky-price firm is equal to its risk-adjusted expectation of the optimal price of the flexible-price firm, conditional on information set s^{t-1}, as in Correia, Nicolini, and Teles (2008).

3.3 Flexible-Price Equilibrium Allocations

We next characterize the set of allocations that can be implemented as competitive equilibria under flexible prices. In any such equilibrium, all firms set their price according to (21).

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) for all $s^t \in S^t$,*

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi A(s_t); \tag{23}$$

and (iii) condition (20) holds for every $i \in I$.

Proof. See Appendix A.5. □

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. All firms share the same technology and face the same nominal wages; as a result they set the same prices. It follows from the demand functions (5) that, in any flexible-price equilibrium, all firms produce identical levels of output.

Part (ii) of Proposition 1 states that in any flexible-price equilibrium, condition (23) must hold in every history. This condition follows from taking the optimal price of the flexible-price firms, (21), noting that in equilibrium all firms set the same nominal price: $p_t^f(s^t) = P(s^t)$, and combining this with the representative household's intratemporal condition in (12).

Therefore, 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 labor wedge, denoted by χ. This wedge is given by:

$$\chi \equiv \left(\frac{\rho-1}{\rho}\right)\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c}. \tag{24}$$

The wedge is 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 time and state-invariant 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) states that in any flexible-price equilibrium, condition (20) must hold for every $i \in I$. 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 (24), this wedge. However, note that the fiscal authority's power to influence allocations using this instrument is limited: χ is a scalar, but condition (23) must hold in every history, $s^t \in S^t$. The fiscal authority can furthermore use lump-sum transfers (or taxes) to control the level of $\bar{T}$. However, the power of fiscal policy to influence allocations using this instrument is also limited: condition (20) must hold for every household type $i \in I$, as we have assumed transfers are non-targeted. Therefore, the non-state-contingency of tax rates and the uniformity of lump sum transfers (or taxes) together imply that the set of allocations that can be implemented as flexible-price equilibria is constrained relative to the feasible set.

3.4 Sticky-Price Equilibrium Allocations

We now turn to the set of allocations that can be implemented as competitive equilibria under sticky prices. In any sticky-price equilibrium, all sticky-price firms set their prices according to (22) and all flexible-price firms set their prices according to (21). It follows from the demand functions (5) 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. By the same logic, we denote the output, labor, and profits of the flexible-price firms by $y^f(s^t)$, $n^f(s^t)$, and $\pi^f(s^t)$, respectively. 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 $\bar{T} \in \mathbb{R}$, and a strictly positive scalar $\chi \in \mathbb{R}_+$, such that the following three sets of conditions are jointly satisfied: (i) for all $s^t \in S^t$, $y^j(s^t) = y^f(s^t)$ for all $j \in \mathcal{J}^f$, and $y^j(s^t) = y^s(s^t)$ for all $j \in \mathcal{J}^s$; (ii) for all $s^t \in S^t$,*

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{\chi A(s_t)} = 0; \tag{25}$$

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

$$\sum_{s^t|s^{t-1}} U_C^m(s^t) y^s(s^t) \left\{ \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{\chi A(s_t)} \right\} \mu(s^t|s^{t-1}) = 0; \tag{26}$$

and (iii) condition (20) holds for every $i \in I$.

Proof. See Appendix A.6. □

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) 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 production across the two sets of firms.

Part (ii) states that in any sticky-price equilibrium, condition (25) must hold in every history. This condition follows from combining the optimality condition of the flexible-price firms with the fictitious representative household's intratemporal condition (12). The resulting condition simply states that the marginal cost of producing an extra unit of output of the flexible-price firm is equated with its marginal revenue. Note that this condition is similar to condition (23) in Proposition 1, and in fact is identical when $y^f(s^t) = Y(s^t)$.

Condition (26) similarly follows from combining the optimality condition for the sticky-price firms with the fictitious representative household's intratemporal optimality condition. This condition states that the marginal cost of producing an extra unit of output of the sticky-price firm is equated with its marginal revenue "on average." It is essentially the same as condition (25) corresponding to flexible-price firm optimality, the only difference being that in (26), the marginal cost and marginal revenue of the sticky-price firm are equated in risk-weighted expectation, conditional on information set s^{t-1}.

Finally, part (iii) of Proposition 2 is identical to part (iii) of Proposition 1; these conditions ensure that all budget constraints are satisfied in equilibrium.

The power of monetary policy. To understand the power of monetary policy vis-à-vis fiscal policy in this economy, it is instructive to rewrite our equilibrium conditions in the following manner. First, 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 $\epsilon(s^t)$ is defined by:

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

Formally, $\epsilon(s^t)$ is defined as the firm's optimal forecast error of $W(s^t)/A(s_t)$, conditional on information set s^{t-1}. Therefore, $\epsilon(s^t)$ acts as a stochastic wedge between the firm's price and its ex-post optimal price, i.e. the markup over nominal marginal cost. Because the sticky-price firm has incomplete information, it cannot perfectly forecast its nominal marginal cost, and as a result, a state-contingent wedge emerges that can be interpreted as the firm's "pricing mistake."

Next, aggregating over the sticky- and flexible-price firms' prices according to (9) and combining the aggregate price level with the representative household's intratemporal optimality condition (12), we obtain the following equilibrium condition in the sticky-price economy:

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

This condition looks similar to condition (23) in the flexible-price economy. As in the flexible-price economy, it indicates that the marginal rate of substitution between consumption and labor is equated with the marginal rate of transformation, $A(s_t)$, modulo a labor wedge. In this case, though, the labor wedge is the product of two components. The first is the scalar χ defined in (24) that corresponds to the markup and taxes. The second is a new, state-contingent component that contains the state-contingent "pricing errors," $\epsilon(s^t)$, made by the fraction κ of inattentive firms.

Therefore, the nominal rigidity gives rise to a state-contingent component of the labor wedge. While χ is a lever of fiscal policy, the state-contingent component represents an *additional* lever of the government, one driven by monetary policy. By shifting $\epsilon(s^t)$, the monetary authority can move around allocations in a manner that the fiscal authority cannot.

The power of monetary policy, however, is limited in two ways, corresponding to parts (i) and (ii) of Proposition 2. First, $\epsilon(s^t)$ introduces a wedge between the prices of the sticky- and flexible-price firms: $p_t^s(s^{t-1}) = \epsilon(s^t)p_t^f(s^t)$. This in turn drives a wedge between the sticky-price and flexible-price firms' output, implying misallocation—a loss in production efficiency. Second, by construction, the forecast errors $\epsilon(s^t)$ must "average out" to 1. This is the meaning of the implementability condition in (26): monetary policy cannot surprise firms "on average." This constraint on equilibrium allocations follows directly from the optimal price-setting behavior of sticky-price firms; it is therefore a natural consequence of rational expectations.

These limits on the power of monetary policy notwithstanding, the nominal rigidity enlarges the set of implementable allocations.

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 $x \in \mathcal{X}^f$; x satisfies the conditions in Proposition 1. This allocation satisfies all conditions in Proposition 2 with $\frac{y^s(s^t)}{Y(s^t)} = \frac{y^f(s^t)}{Y(s^t)} = 1$ for all $s^t \in S^t$. Therefore, $x \in \mathcal{X}^s$. □

Any allocation that can be implemented under flexible-prices can also be implemented under sticky prices. It can be implemented with a monetary policy that targets price stability.[18]

3.5 Welfare function and Ramsey problem definition

The goal of this paper is to solve the Ramsey problem in this economy. We assume a utilitarian social welfare function given by:

$$\mathcal{U} \equiv \sum_{i \in I} \lambda^i \pi^i \sum_t \sum_{s^t} \beta^t \mu(s^t) U(c^i(s^t), \ell^i(s^t)/\theta^i(s_t)) \tag{29}$$

where $\lambda \equiv (\lambda^i)_{i \in I}$ denotes an arbitrary set of Pareto weights, with $\lambda^i > 0$ for all $i \in I$. We state the Ramsey problem as follows.

Definition 4. *A Ramsey optimum x^* is an allocation x that maximizes (29) subject to $x \in \mathcal{X}^s$.*

4 The Relaxed Ramsey Problem

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 implementability constraints that must be satisfied. We thus proceed in this section by first solving an *easier* problem, that of a "relaxed" Ramsey planner.

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.

[18] Equivalently, it sets $\epsilon(s^t) = 1$ for all $s^t \in S^t$.

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 $\varphi \equiv (\varphi^i)$ and a a scalar $\bar{T} \in \mathbb{R}$ such that condition (20) holds for all $i \in I$. A relaxed Ramsey optimum x^{R*} is an allocation x that maximizes social welfare (29) subject to $x \in \mathcal{X}^R$.*

Relative to the set of sticky-price allocations, the relaxed set is constructed by dropping all implementability conditions stated in parts (i) and (ii) of Proposition 2, but maintaining those stated in part (iii). Given this definition, the following observation is self-evident.

Observation 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 and intermediate good-specific linear 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.

The relaxed Ramsey problem in our economy is equivalent to the Ramsey problem in Werning (2007). For our purposes, solving this problem is useful in the following sense. We will first characterize a relaxed Ramsey optimum x^{R*}. We will then derive 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 a relaxed Ramsey optimum and an *unrelaxed* Ramsey optimum!

Let $\pi^i \nu^i$ denote the Lagrange multiplier on the implementability condition (20) of type $i \in I$; let $\nu \equiv (\nu^i)_{i \in I}$ denote the set of multipliers. We incorporate these constraints into the planner's maximand and define the pseudo-welfare function $\mathcal{W}(\cdot)$ as follows:

$$\begin{aligned} \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) \equiv \sum_{i \in I} \pi^i \{ & \lambda^i U(\omega_C^i(\varphi) C(s^t), \omega_L^i(\varphi, s_t) L(s^t)/\theta^i(s_t)) \\ & + \nu^i \left[U_C^m(s^t) \omega_C^i(\varphi) C(s^t) + U_L^m(s^t) \omega_L^i(\varphi, s_t) L(s^t) \right] \} \end{aligned} \tag{30}$$

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 $\bar{T} \in \mathbb{R}$, so as to maximize*

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) - U_C^m(s_0) \bar{T} \sum_{i \in I} \pi^i \nu^i \tag{31}$$

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

The pseudo-welfare function is stated in terms of aggregates alone, making the relaxed Ramsey planning problem surprisingly tractable. One can think of the pseudo-welfare function as a social welfare function that not only reflects the distributional motives of society but also incorporates the constraints imposed by the absence of type-specific transfers.

4.1 Relaxed Ramsey Optimum

The following proposition characterizes a relaxed Ramsey optimum given an arbitrary set of Pareto weights.[19] For shorthand, we let $\mathcal{W}_C(s^t) \equiv \partial\mathcal{W}(\cdot)/\partial C(s^t)$ and $\mathcal{W}_L(s^t) \equiv \partial\mathcal{W}(\cdot)/\partial L(s^t)$.

Proposition 3. *A feasible 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)*

$$-\frac{\mathcal{W}_L(s^t)}{\mathcal{W}_C(s^t)} = A(s_t), \qquad \forall s^t \in S^t. \tag{32}$$

Proof. See Appendix A.7. □

Part (i) of Proposition 3 indicates that a relaxed Ramsey optimum features zero dispersion in intermediate good output. Part (ii) provides the planner's first-order conditions: it sets the social marginal rate of substitution between labor and consumption, $-\mathcal{W}_L(s^t)/\mathcal{W}_C(s^t)$, equal to the marginal rate of transformation, $A(s_t)$, state-by-state. Therefore, although the relaxed Ramsey planner has the ability to tax (or subsidize) different margins in order to finance lump-sum transfers, it is optimal to distort only the intratemporal margin. A relaxed Ramsey optimum thereby preserves production efficiency in the sense of Diamond and Mirrlees (1971).

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:*

$$\theta^i(s_t) = \vartheta^i \Theta(s_t), \qquad \forall s_t \in S, \tag{33}$$

then: (i) the relaxed Ramsey optimum is implementable as a flexible-price allocation, $x^{R*} \in \mathcal{X}^f$; (ii) the relaxed Ramsey optimum is implementable as a sticky-price allocation, $x^{R*} \in \mathcal{X}^s$; 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 (33) is satisfied. Then the individual household shares defined in (17) reduce to:

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

Notably, $\omega^i_L(\varphi)$ is non-state-contingent. The relaxed Ramsey optimality condition in (32) can be written as follows:

$$-\frac{U^m_L(s^t)}{U^m_C(s^t)} \left[\frac{\sum_{i\in I} \pi^i \omega^i_L(\varphi) \left(\lambda^i/\varphi^i + \nu^i(1+\eta)\right)}{\sum_{i\in I} \pi^i \omega^i_C(\varphi) \left(\lambda^i/\varphi^i + \nu^i(1-\gamma)\right)}\right] = A(s_t) \tag{34}$$

[19] See also Werning (2007).

Comparing this to the flexible-price implementability condition (23), it is clear that (34) 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 $x^{R*} \in \mathcal{X}^s \subset \mathcal{X}^R$. □

Theorem 1 provides sufficient conditions under which a relaxed Ramsey optimum can be implemented under flexible prices. These conditions are: separable and homothetic preferences and proportional shocks to the labor productivity distribution.[20]

To understand the intuition behind Theorem 1, consider the problem of the relaxed Ramsey planner constrained only by the feasibility of allocations and the budget set implementability conditions. Suppose society desires a more equal distribution of resources across households than under laissez-faire (zero taxes and flexible prices). It is optimal, as discussed above, to distort only the intratemporal margin—the wedge between the marginal rate of substitution, $-U_L^m(s^t)/U_C^m(s^t)$, and the marginal rate of transformation, $A(s_t)$. The relaxed planner thereby faces a trade-off between the benefit of intratemporal taxation and the cost.

The cost of intratemporal taxation is inefficiency: distorting the intratemporal margin moves $C(s^t)$ and $L(s^t)$ away from their undistorted levels. The benefit of intratemporal taxation is redistribution: the tax is used to finance lump-sum transfers, bringing about a more equal distribution of resources across households. To understand this last point, note that given a strictly positive linear tax, all households face the same marginal tax rate, but high-skilled types face a higher *average* tax rate than low-skilled types. This is because more tax revenue is collected from high-skilled types than from low-skilled types, yet total tax revenue finances uniform, lump-sum transfers. It follows that a greater tax rate coincides with greater redistribution.

The relaxed Ramsey planner's optimum is the point at which, in every state, the marginal benefit of intratemporal taxation is equated with the marginal cost; this state-by-state trade-off is captured by the relaxed planner's first-order conditions in (32).

Now consider whether this optimum can be achieved under flexible prices. When preferences are separable and homothetic, the marginal cost of intratemporal taxation is constant across states (Lucas and Stokey, 1983).[21] One can understand this as an application of the classic uniform commodity taxation result: under homothetic and separable preferences, it is optimal to tax goods at a uniform rate (Chari and Kehoe, 1999; Atkinson and Stiglitz, 1980).

Furthermore, with proportional shocks to the labor skill distribution, as in (33), the ratio of

[20]Note that in the proof of Theorem 1, we use the fact that the allocation of consumption and labor across individual households take the form given in (16), which itself relies on the separable and iso-elastic preference specification in 1.

[21]In a standard single-agent Ramsey framework without lump-sum taxes, linear taxation finances exogenous government spending, as in Lucas and Stokey (1983). It is a well known result that with separable and iso-elastic preferences, the Ramsey optimum features "perfect tax smoothing."

labor productivity across any two types, $i, j \in I$, is constant across states:

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

Homothetic preferences and no movement whatsoever in the *relative* skill distribution imply that there are no states in which intratemporal taxation is more beneficial than others. More specifically, although greater tax revenue is collected from high-skilled types than from low-skilled types, the *relative* tax revenue collected across types is invariant. It follows that the marginal redistributive benefit of taxation is constant across states.

If both the marginal benefit and the marginal cost of intratemporal taxation are constant across states, then the optimum at which marginal benefit equals marginal cost is also constant. It follows that the relaxed Ramsey optimum can be achieved with a uniform distortion χ, implemented with fiscal tools, and a monetary policy that replicates flexible prices.

Finally, note that a 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 key insight in Correia, Nicolini, and Teles (2008). Although the relaxed planner in our environment trades off distributional motives with inefficiency, it distorts only the intratemporal margin—under no circumstances does it introduce misallocation *across* firms. It follows that a relaxed Ramsey optimum is implementable under sticky prices if and only if it is implementable under flexible prices.

Necessity. Theorem 1 provides sufficient conditions under which a relaxed Ramsey optimum can be implemented under flexible prices. These conditions, however, are not necessary. To see why, note that there exists a degenerate case in which the Pareto weights are exactly equal to the Negishi weights under laissez-faire (zero taxes and flexible prices). Formally, if $\lambda^i = \varphi^i$ for all $i \in I$, then $\nu^i = 0$ for all $i \in I$. In other words, given any stochastic process of the skill distribution, there exists a knife edge case of Pareto weights such that the relaxed Ramsey optimum can be implemented under flexible prices with $\chi = 1$. Furthermore, this allocation is on the Pareto frontier: all implementability conditions are slack at this optimum.

4.2 Proportional shocks but suboptimal fiscal policy

We can strengthen the result on monetary policy provided in Theorem 1. In our previous analysis we have assumed that the planner controls both monetary *and* fiscal policy. We now consider the special case in which labor productivity shocks are proportional, but the Ramsey planner can no longer control fiscal policy and can only control monetary policy. Even if fiscal policy is set suboptimally, i.e. $\chi \neq \chi^*$, we show that optimal monetary policy remains unchanged.

Theorem 2. *Let 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* (33)*, and taxes are set such that:*

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

It is optimal for monetary policy to replicate the flexible-price allocation.

Proof. See Appendix A.9. □

If preferences are separable and homothetic and the labor skill distribution exhibits proportional shocks, then it is optimal for monetary policy to implement flexible price allocations, *regardless* of fiscal policy. This generalizes the result provided in Theorem 1 to all fiscal policies (within the affine tax structure), including sub-optimal policies.

If the tax rate were set suboptimally, one might presume that monetary policy should try to substitute for the missing tax (wedge). Theorem 2 states that any attempt to do so, or any deviation from flexible prices for that matter, would be counterproductive.

To understand why, note that the missing tax wedge is constant across all states and histories. But recall that the only power monetary policy has over allocations is through the pricing errors, $\epsilon(s^t)$, and these errors must, by construction, "average out" to 1 over all states immediately following a given history. Therefore, if the monetary authority were to raise the labor wedge in one state, it cannot do so unless it lowers it in another state. If the missing tax wedge is constant across all states, then there is no reason why one state should be more distorted than any other, and shifting distortions across states only worsens the allocation. As a result, it remains optimal for monetary policy to do absolutely nothing at all and target price stability.

5 The Ramsey Problem and Optimal Monetary Policy

We return to our original problem of interest, that of the "unrelaxed" Ramsey planner (Definition 4). Given the pseudo-welfare function $\mathcal{W}(\cdot)$ defined in (30), we can write the Ramsey planning problem in the following way.

Ramsey Planner's Problem. *The Ramsey planner chooses an allocation* $x \equiv \{y^s(s^t), y^f(s^t), C(s^t), Y(s^t), L(s^t)\}_{t\geq 0, s^t \in S^t}$*, market weights* $\varphi \equiv (\varphi^i)$*, constants* $\bar{T} \in \mathbb{R}$ *and* $\chi \in \mathbb{R}_+$*, in order to maximize* (31)*, subject to:*

$$Y(s^t) = \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}, \qquad L(s^t) = \kappa \frac{y^s(s^t)}{A(s_t)} + (1-\kappa)\frac{y^f(s^t)}{A(s_t)}, \tag{35}$$

$C(s^t) \leq Y(s^t)$, (25)*, and* (26)*.*

Unlike the relaxed Ramsey planner, the unrelaxed Ramsey planner is subject to *all* implementability conditions, including conditions (25) and (26) of Proposition 2. In Appendix Section A.8 we provide a complete characterization of the Ramsey optimum.[22] While we do not provide the full characterization here for exposition and conciseness, the essential necessary condition of the planner's optimum appears as follows:

$$-\frac{\mathcal{W}_L(s^t)}{\mathcal{W}_C(s^t)}\left[\text{Ramsey wedge}(s^t)\right] = \frac{Y(s^t)}{L(s^t)}. \tag{36}$$

Condition (36) is the Ramsey planner's intratemporal optimality condition; it is the counterpart to condition (32) of the relaxed Ramsey optimum. The Ramsey planner sets the social marginal rate of substitution between labor and consumption, $-\mathcal{W}_L(s^t)/\mathcal{W}_C(s^t)$, equal to the marginal rate of transformation, $Y(s^t)/L(s^t)$, modulo a state-contingent wedge. Relative to the relaxed planner, the Ramsey planner is subject to implementability conditions (25) and (26); the wedge in condition (36) is a function of their state-contingent Lagrange multipliers. When these conditions are slack, their corresponding multipliers are equal to zero and condition (36) reduces to (32). When these conditions are binding, the ratio $-\mathcal{W}_L(s^t)/\mathcal{W}_C(s^t)$ departs from the marginal rate of transformation at the Ramsey optimum.

Furthermore, in contrast to the relaxed Ramsey optimum, note that the marginal rate of transformation between labor and consumption at the Ramsey optimum is no longer $A(s_t)$, but instead is equal to $Y(s^t)/L(s^t)$. As long as the planner finds it optimal to deviate from flexible-price allocations, a wedge arises between sticky- and flexible-price firm output, resulting in a loss in production efficiency.

We are interested in what the Ramsey optimum implies for optimal monetary policy. To that extent, we follow the primal approach and characterize the implicit labor wedge that supports the Ramsey optimum in equilibrium. Given a Ramsey optimum x^*, we define the "monetary wedge," $1-\tau_M^*(s^t)$, implicitly as follows:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi^*(1-\tau_M^*(s^t))\frac{Y(s^t)}{L(s^t)}, \tag{37}$$

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

Theorem 3. *Let $\mathcal{I}: S \to \mathbb{R}_+$ be a positively-valued function defined by:*

$$\mathcal{I}(s_t) \equiv \frac{\sum_{i\in I}\tilde{\pi}^i(\varphi^i)^{-1/\eta}(\theta^i(s_t))^{\frac{1+\eta}{\eta}}}{\sum_{i\in I}\pi^i(\varphi^i)^{-1/\eta}(\theta^i(s_t))^{\frac{1+\eta}{\eta}}} > 0, \qquad \textit{where} \qquad \tilde{\pi}^i \equiv \pi^i\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right]. \tag{38}$$

[22] See Proposition 6 in Appendix A.8 and its proof.

There exists a threshold $\bar{\mathcal{I}}(s^{t-1}) > 0$, measurable in history s^{t-1}, such that the optimal 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. □

The function $\mathcal{I}(s_t)$ can be interpreted as a sufficient statistic for the level of labor income inequality in state s_t. Recall that λ^i are the Pareto weights, φ^i are the Negishi weights, and ν^i are the planner's multipliers on the implementability conditions in (20). Notably all are scalars—they do not depend on the realized state or history. As a result, $\mathcal{I}(s_t)$ is history-independent and in particular depends only on the current realization of the labor skill distribution, $(\theta^i(s_t))_{i\in I}$.

Furthermore, as we show in an example below, the term $\lambda^i/\varphi^i + \nu^i(1+\eta)$ is increasing in the household's human wealth: households with high lifetime labor earnings have larger weights, $\lambda^i/\varphi^i + \nu^i(1+\eta)$, at the Ramsey optimum than households with low lifetime labor earnings.[23] As the labor productivities $\theta^i(s_t)$ of the high-type households increase relative to those of lower-types, the numerator of $\mathcal{I}(s_t)$ grows relative to its denominator. As a result, $\mathcal{I}(s_t)$ is high in states in which high human wealth households are relatively more productive and low in states when they are relatively less productive. Furthermore, the extent to which $\mathcal{I}(s_t)$ responds to relative movements in the labor skill distribution depends on the Frisch elasticity of labor supply, $1/\eta$.

Theorem 3 states that the optimal monetary tax varies with the state and depends on the level of labor income inequality, as proxied for by $\mathcal{I}(s_t)$. There exists a threshold $\bar{\mathcal{I}}(s^{t-1})$ such that when labor income inequality is strictly greater than this threshold, the implied optimal monetary tax is positive. On the other hand, when labor income inequality is below this threshold, the implied optimal monetary tax is negative (i.e. a subsidy). When $\mathcal{I}(s_t)$ is exactly equal to the threshold, the optimal monetary tax is zero.

Recall from Theorem 1 that when preferences are separable and homothetic and shocks to the labor skill distribution are proportional, the tax system is sufficient to achieve the optimal level of redistribution. Theorem 3 nests this result as a special case: when the labor skill distribution satisfies (33), the function $\mathcal{I}(s_t)$ reduces to a constant equal to the threshold in all states and histories; in this case the optimal monetary tax is always zero.

Numerical Illustration. We illustrate Theorem 3 with a simple numerical example with two household types—a high-type and a low-type—indexed by $i \in \{H, L\}$, of equal sizes ($\pi^H = \pi^L = 1/2$). We consider a labor skill distribution in which the high-type is always more productive

[23]While high-type households have high market weights, φ^i, at the Ramsey optimum, the multipliers ν^i on their budget implementability conditions are also high and dominate the overall direction of this term.

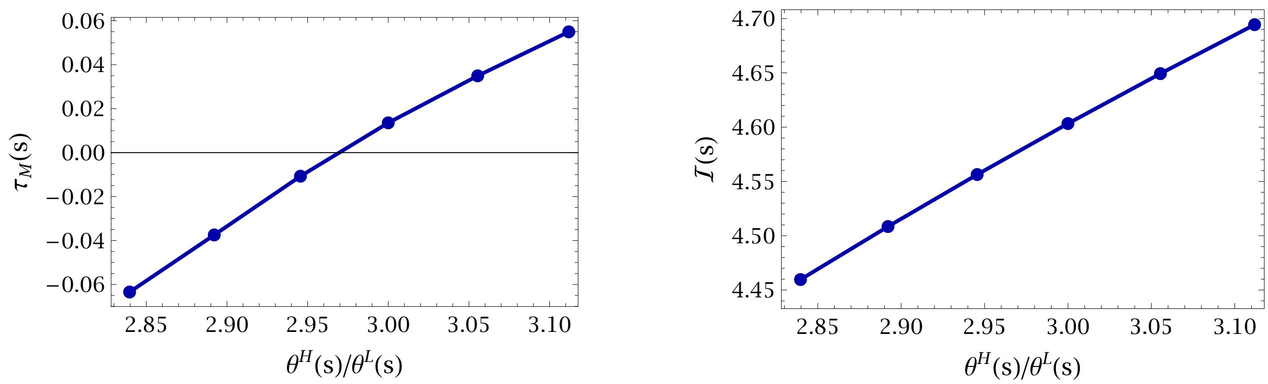

Figure 1. Optimal $\tau_M^*(s^t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$ (left panel). $\mathcal{I}(s_t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$ (right panel).

than the low-type, but we let the ratio $\theta^H(s_t)/\theta^L(s_t)$ fluctuate across 6 possible states. We assume states are uniformly distributed and i.i.d.: $\mu(s'|s) = 1/6$ for all $s, s' \in S$. Finally, we set $\beta = .98$, $\eta = 1$, $\gamma = 2$, $\kappa = .25$, and $\rho = 2$.

We numerically solve for the Ramsey optimum with equal Pareto weights: $\lambda^H = \lambda^L = 1$. The left panel of Figure 1 plots the optimal monetary tax for different values of $\theta^H(s_t)/\theta^L(s_t)$. As this ratio increases, i.e. as the high-type becomes more productive relative to the low-type, $\tau_M^*(s^t)$ increases. To check that this is in line with the predictions of our theory, in the right panel of Figure 1 we plot $\mathcal{I}(s_t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$.[24]

Intuition. Suppose society desires a more equal distribution of resources across households than under laissez-faire (zero taxes and flexible prices). If the Ramsey planner had access to type-specific lump-sum transfers, it would use those to redistribute; in fact, it would simply implement the most desirable location on the Pareto frontier. In the absence of such transfers, the Ramsey planner cannot achieve that allocation.

Recall that the *relaxed* Ramsey planner finds it optimal to distort, state-by-state, the intratemporal margin in order to attain a more desirable distribution of wealth. If the Ramsey planner had access to state-contingent tax rates, it would use those to implement the relaxed Ramsey optimum. In particular, the state-contingent tax that implements (32) satisfies:

$$\chi(s_t) \equiv 1 - \tau(s_t) \propto \frac{1}{\mathcal{I}(s_t)}. \tag{39}$$

Therefore, the optimal tax rate $\tau(s_t)$ is strictly increasing in $\mathcal{I}(s_t)$. But again: in the absence of such state-contingent taxes, the Ramsey planner cannot achieve the relaxed optimum.

The Ramsey planner thereby uses the available instruments—a non-state-contingent fiscal wedge, and state-contingent monetary policy—to imperfectly mimic the missing state-

[24] In this example, the weight $\lambda^i/\varphi^i + \nu^i(1+\eta)$ of the high-type is greater than that of the low-type.

contingent tax. Roughly speaking, while the constant fiscal wedge is chosen to achieve the desired *average* level of the optimal distortion, monetary policy varies across states in a way that mimics the desired state-contingency.

In fact, in order for monetary policy to be an effective policy tool, the benefit of intratemporal distortion must vary across states. As we have emphasized with Theorem 2, monetary policy can only raise the labor wedge in one state if it lowers it in another. In the case of proportional shocks to the labor skill distribution, there is no such state-contingency in the benefit of intratemporal distortion. Relative productivities across households are constant; it follows, under appropriate preference conditions, that the optimal tax rate is also constant (Theorem 1).

But in the case in which relative labor productivities vary across states, the benefit of intratemporal distortion also varies—rendering state-contingent monetary policy useful. The optimal monetary tax rises in states in which households with high lifetime labor earnings are relatively more productive (than other households) and falls in states in which high wealth households are relatively less productive. Distorting the economy in this state-contingent manner—contracting when labor income inequality is high and expanding when labor income inequality is low—compresses the *lifetime* labor earnings distribution in the desired direction.

Monetary policy thus assumes the desired state-contingent pattern of the missing tax in (39): the optimal monetary tax is positive when $\mathcal{I}(s_t)$ is high and negative when $\mathcal{I}(s_t)$ is low. Monetary policy, however, is an imperfect substitute for this tax as abandoning the flexible-price allocation results in a loss of production efficiency. Nevertheless, starting from the flexible-price benchmark, any loss in production efficiency from abandonment is, to a first-order, zero. This is because at the flexible-price allocation, production efficiency is maximized. It follows that if there is any incentive to move monetary policy away from replicating flexible-prices—in this case, when relative productivities across households fluctuate—the planner finds it optimal to do so.

5.1 Implementation: Optimal Monetary Policy

We now turn to implementation of the Ramsey optimum. We begin with fiscal policy. Clearly there is no unique implementation of the optimal fiscal wedge χ^*, and any implementation of χ^* results in the same behavior for optimal monetary policy. For the sake of exposition, we set the sales subsidy such that it directly offsets the monopolistic markup and let the labor income and consumption tax rates jointly implement the optimal fiscal wedge:

$$1-\tau_r = \frac{\rho}{\rho-1}, \qquad \text{and} \qquad \frac{1-\tau_\ell}{1+\tau_c} = \chi^*. \tag{40}$$

We now turn to monetary policy. We define the aggregate markup $\mathcal{M}(s^t)$ in the economy as the price level over the nominal marginal cost; in logs:

$$\log \mathcal{M}(s^t) \equiv \log P(s^t) - \log(W(s^t)/A(s^t)). \tag{41}$$

Note that if we shut down aggregate productivity shocks, i.e. $A(s^t) = 1$ for all s^t,then the aggregate markup is equal to the inverse of the real wage, $W(s^t)/P(s^t)$. We express optimal monetary policy in terms of the aggregate markup as follows.

Proposition 4. *With tax rates set according to (40), the optimal markup satisfies:*

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

Proof. See Appendix A.12. □

Proposition 4 is simply a restatement of Theorem 3 but in terms of the optimal state-contingent markup instead of the implicit monetary tax. In fact, the two are essentially equivalent: if the final good price rises relative to marginal cost, it is as if households are paying a labor income tax. Conversely, if the price falls relative to marginal cost, it is as if they are receiving a subsidy.

As with the monetary tax, we find that when $\mathcal{I}(s_t)$ rises above the critical threshold, the optimal markup is positive; conversely when $\mathcal{I}(s_t)$ falls below the threshold, the optimal markup is negative. When $\mathcal{I}(s_t)$ is exactly equal to the threshold, the optimal markup is zero. Again, by distorting the economy in this state-contingent manner—contracting when labor income inequality is high and expanding when labor income inequality is low—monetary policy compresses the lifetime labor earnings distribution in the desired direction.

The special case in which shocks to the labor distribution are proportional is nested in Proposition 4. When the labor skill distribution satisfies (33), the function $\mathcal{I}(s_t)$ reduces to a constant equal to the threshold in all states and histories. In this case, it is optimal for monetary policy to implement flexible-price allocations—it can do so by targeting a constant markup, $\log \mathcal{M}(s^t) = 0$, which can itself be achieved by targeting zero inflation (price stability).[25] In Appendix C we expand on our discussion of implementation and provide results on the behavior of aggregate price levels and nominal interest rates consistent with the Ramsey optimum.

5.2 Optimal Monetary Policy with Partially State-Contingent Taxes

Thus far in our analysis we have made the stark assumption that monetary policy is state-contingent while fiscal policy is not—a standard assumption found throughout the New Keynesian literature. We now partially relax this restriction on fiscal tools and allow tax rates to be set one period in advance; specifically, we let τ_c, τ_ℓ, and τ_r at time t be contingent on s^{t-1}. While this does not go all the way to full fiscal state-contingency, it provides the fiscal authority with the

[25] Here, the level of zero for the log markup under flexible prices is arbitrary: it is only equal to zero because we have set the sales subsidy to exactly cancel out $\frac{\rho}{\rho-1}$. Had we not made that choice, the markup under flexible prices would be equal to a non-zero constant, specifically: $\mathcal{M} = (1-\tau_r)^{-1}\frac{\rho}{\rho-1}$.

flexibility of responding to shocks with a one-period lag. To the extent that shocks are persistent, fiscal policy can thus absorb some of the state-contingent pressure placed on monetary policy.

In this case we obtain a sharper characterization of the behavior of $\tau_M^*(s^t)$ around zero.

Theorem 4. *Let tax rates be set one period in advance. There exists a threshold $\bar{\mathcal{I}}(s^{t-1}) > 0$ such that $\tau_M^*(s^t) = 0$ if and only if $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$, $\tau_M^*(s^t) > 0$ if and only if $\mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1})$, and $\tau_M^*(s^t) < 0$ if and only if $\mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1})$. To a first-order Taylor approximation around $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$,*

$$\tau_M^*(s^t) \approx \delta_0[\mathcal{I}(s_t)/\bar{\mathcal{I}}(s^{t-1}) - 1] \tag{42}$$

where

$$\delta_0 = \frac{1}{1 + \rho(\eta + \gamma)\frac{1-\kappa}{\kappa}} \in (0, 1). \tag{43}$$

Proof. See Appendix D.3. □

When we allow tax rates to be set one-period in advance, our main result on the optimal conduct of monetary policy remains intact. However, with this greater level of fiscal flexibility, we obtain a sharper characterization of the optimal monetary tax near the benchmark of zero. In particular, we show that to a first-order Taylor approximation around $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$, the optimal monetary tax is strictly increasing in $\mathcal{I}(s_t)/\bar{\mathcal{I}}(s^{t-1})$, with a slope of $\delta_0 \in (0, 1)$.

The slope δ_0 characterizes the extent to which the optimal monetary tax responds to an increase in $\mathcal{I}(s_t)$: a larger value for δ_0 indicates a more aggressive response, whereas a lower value indicates a less aggressive response. An explicit, closed-form expression for this derivative is given in (43). In particular, δ_0 is strictly positive, strictly less than 1, and a function of the primitives ρ, γ, η, and κ.

First, note that δ_0 is decreasing in ρ, the elasticity of substitution across goods. Deviations of monetary policy away from the flexible-price allocation results in intermediate good price dispersion. In response, the final good firm substitutes away from high-priced intermediates towards low-priced intermediates. The greater the substitutability across goods, the greater the misallocation and corresponding loss in production efficiency. It follows that when ρ is high, monetary policy responds less aggressively to movements in $\mathcal{I}(s_t)$.

Second, δ_0 is increasing in $\kappa/(1-\kappa)$, the mass of sticky-price firms relative to the mass of flexible-price firms. Consider the limit in which $\kappa \to 1$. In this case δ_0 approaches one. When nearly all firms in the economy are sticky, movements in monetary policy away from flexible-rice allocations result in near zero losses in production efficiency; monetary policy therefore approximates a labor income tax. In this limit, monetary policy perfectly mimics the optimal state-contingent tax rate, which responds one-for-one with changes in $\mathcal{I}(s_t)$. In the opposite limit in which $\kappa \to 0$, δ_0 approaches zero. When nearly all firms in the economy are flexible, monetary policy has no power.

6 Optimal Policy with Constrained Profit Taxation

In this section we extend the baseline model by discarding Assumption 1. We study how untaxed profits and heterogeneous initial equity shares affect optimal monetary policy.

We return to our general characterization of the households' lifetime budget constraints in (18). We do not impose any restrictions on the cross-sectional covariance between labor skill type and initial equity across households, but we will discuss the implications of this covariance for optimal policy. In particular we will focus on the case in which initial equity shares covary positively with lifetime labor earnings.

6.1 Equilibrium Characterization

We begin by characterizing the set of equilibrium allocations.

Proposition 5. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a sticky-price equilibrium if and only if there exist $\varphi \equiv (\varphi^i)$, $\bar{T} \in \mathbb{R}$, $\chi \in \mathbb{R}_+$, and a weakly-positive scalar $\hat{\vartheta} \in \mathbb{R}_{\geq 0}$, such that parts (i)-(ii) of Proposition 2 are satisfied, and*

$$\begin{aligned}&\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi, s_t)L(s^t)\right] \\ &= U_C^m(s_0)\bar{T} + \sigma_0^i \hat{\vartheta} \sum_t \sum_{s^t} \beta^t \mu(s^t) \left[\chi \frac{\rho}{\rho - 1} U_C^m(s^t)C(s^t) + U_L^m(s^t)L(s^t)\right]\end{aligned} \tag{44}$$

holds for all $i \in I$.

Proof. See Appendix E.1. □

Proposition 5 is the analog of Proposition 2 in the baseline economy. Parts (i) and (ii) of Proposition 2 remain intact; the only distinction is that we have replaced the implementability conditions in (20) with those in (44). The latter correspond to the lifetime budget constraints in (18) but with real profits, $\Pi(s^t)/P(s^t)$, expressed in terms of the allocation.

The final term on the right-hand side of (44) represents type-i's heterogeneous exposure, σ_0^i, to the lifetime value of after-tax real profits. Note that this term includes a weakly-positive scalar given by $\hat{\vartheta} \equiv (1 - \tau_\Pi)/(1 - \tau_\ell)$. This scalar parameterizes an additional lever by which fiscal policy can influence real allocations. If profits are fully taxed ($\tau_\Pi = 1$), as assumed in our baseline model, then $\hat{\vartheta} = 0$ and (44) reduces to the implementability conditions in (20). Furthermore, to the extent that initial equity shares covary positively with lifetime labor earnings, and the Ramsey planner wishes to redistribute from high wealth households to low wealth households, it is optimal to fully tax profits. Therefore, in order to make our analysis in this section interesting, we make the following ad hoc assumption on τ_Π and τ_ℓ.

Assumption 2. *Let $\vartheta \geq 0$ be a weakly positive scalar. The tax rates τ_Π and τ_ℓ are such that $\hat{\vartheta} = \vartheta$.*

Assumption 2 is an ad hoc constraint on fiscal policy. If $\vartheta > 0$, then the fiscal authority cannot fully tax profits nor drive τ_ℓ to negative infinity. For the remainder of our analysis, we impose Assumption 2 and index economies by ϑ. We let $\mathcal{X}^s(\vartheta)$ denote the set of sticky-price allocations in economy ϑ: those that satisfy Proposition 5 with $\hat{\vartheta} = \vartheta$. Our baseline economy is nested in this more general formulation with $\vartheta = 0$.

6.2 The Ramsey Problem and Optimal Monetary Policy

We return to the Ramsey problem. A Ramsey optimum x^* is an allocation x that maximizes social welfare (29) subject to $x \in \mathcal{X}^s(\vartheta)$. We solve this problem in Appendix E.2 and characterize the Ramsey optimum. Here, we present our main result on the optimal implicit monetary tax.

Theorem 5. *There exists a threshold $\bar{\mathcal{I}}_\vartheta(s^{t-1}) > 0$, such that the optimal implicit monetary tax $\tau_M^*(s^t)$ satisfies:*

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

Proof. See Appendix E.3. □

When profit taxation is constrained, the behavior of optimal monetary policy resembles that in the baseline economy with full profit taxation. In particular, the optimal monetary tax is state-contingent and depends on $\mathcal{I}(s_t)$, our sufficient statistic of labor income inequality. There exists a threshold $\bar{\mathcal{I}}_\vartheta(s^{t-1})$ such that when $\mathcal{I}(s_t)$ is strictly greater than the threshold the optimal monetary tax is positive, when $\mathcal{I}(s_t)$ is strictly below the threshold the optimal monetary tax is negative, and when $\mathcal{I}(s_t)$ is equal to the threshold the optimal monetary tax is zero.

Therefore, our main qualitative result on the optimal conduct of monetary policy is robust to the constraint on profit taxation. Initially this result may seem surprising. When monetary policy abandons flexible price allocations and increases the markup, firm profits increase; when it lowers the markup, firm profits decrease. However, as emphasized previously, monetary policy can only raise the markup in one state if it lowers it in another. On average, such movements in profits may not fully cancel each other out. Therefore, when profits are only partially taxed, the abandonment of flexible-price allocations by monetary policy leads to some amount of redistribution of financial wealth.

Yet, it remains the case that when relative productivities vary over the business cycle, the benefit of intratemporal distortion varies across states in the same pattern as when profits are fully taxed. By distorting the economy in a state-contingent manner—by raising the markup when labor income inequality is high and lowering the markup when labor income inequality is low—monetary policy still compresses the lifetime labor earnings distribution in the desired direction, regardless of who owns the firms. Fiscal policy (in the form of χ) can then be used to balance out all "average" costs and benefits of intratemporal distortion, including the redistributional effect of monetary policy on financial wealth.[26]

6.3 Constrained Profit Taxation and Partially State-Contingent Taxes

While it is difficult to characterize the behavior of $\tau_M^*(s^t)$ in general, we can again provide a sharper characterization of the optimal monetary tax in the particular version of our economy in which tax rates are set one period in advance.

Theorem 6. *Let tax rates be set one period in advance. There exists a threshold $\bar{\mathcal{I}}_\vartheta(s^{t-1}) > 0$, such that $\tau_M^*(s^t) = 0$ if and only if $\mathcal{I}(s_t) = \bar{\mathcal{I}}_\vartheta(s^{t-1})$, $\tau_M^*(s^t) > 0$ if and only if $\mathcal{I}(s_t) > \bar{\mathcal{I}}_\vartheta(s^{t-1})$, and $\tau_M^*(s^t) < 0$ if and only if $\mathcal{I}(s_t) < \bar{\mathcal{I}}_\vartheta(s^{t-1})$. To a first-order Taylor approximation around $\mathcal{I}(s_t) = \bar{\mathcal{I}}_\vartheta(s^{t-1})$,*

$$\tau_M^*(s^t) \approx \delta_0 \frac{1}{\mathcal{H}_\vartheta(s^{t-1}) + \frac{\rho}{\rho-1}(\gamma-1)\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i} [\mathcal{I}(s_t) - \bar{\mathcal{I}}_\vartheta(s^{t-1})] \tag{45}$$

where $\delta_0 \in (0,1)$ is given in (43) and $\mathcal{H}_\vartheta(s^{t-1})$ is a positively-valued function measurable in s^{t-1} defined by:

$$\mathcal{H}_\vartheta(s^{t-1}) \equiv \frac{1}{\chi^*(s^{t-1})} \frac{\sum_{i\in I} \tilde{\pi}^i (\varphi^i)^{1/\gamma}}{\sum_{i\in I} \pi^i (\varphi^i)^{1/\gamma}} > 0. \tag{46}$$

Proof. See Appendix E.5 and for the definition of $\mathcal{H}_\vartheta(s^{t-1})$. □

Theorem 6 provides a first-order approximation of $\tau_M^*(s^t)$ near the benchmark of $\tau_M^*(s^t) = 0$. When profit taxation is constrained, the behavior of optimal monetary policy resembles that in our baseline economy with full profit taxation. In particular, the slope of $\tau_M^*(s^t)$ with respect to $\mathcal{I}(s_t)$ is strictly positive.

We compare this behavior to that in the baseline economy, $\vartheta = 0$, with tax rates set one period in advance. In this economy, to a first order around $\tau_M^*(s^t) = 0$,

$$\tau_M^*(s^t) \approx \delta_0 \frac{1}{\mathcal{H}_0(s^{t-1})} [\mathcal{I}(s_t) - \bar{\mathcal{I}}_0(s^{t-1})]. \tag{47}$$

where $\mathcal{H}_0(s^{t-1}) > 0$ takes the same form as in (46), but evaluated at the Ramsey optimum for economy-0. The above equation directly corresponds to equation (42) in Theorem 4, noting that $\bar{\mathcal{I}}_0(s^{t-1}) = \mathcal{H}_0(s^{t-1})$.

We will make the following heuristic argument comparing the slope in (45) to that in (47). The terms $\mathcal{H}_\vartheta(s^{t-1})$ and $\mathcal{H}_0(s^{t-1})$, although they take the same functional form, are difficult to compute and compare. We therefore focus on the other term determining the difference in

[26]To see this new role for χ, note that it appears in the last term of equation (44), the term that corresponds to after-tax real profits.

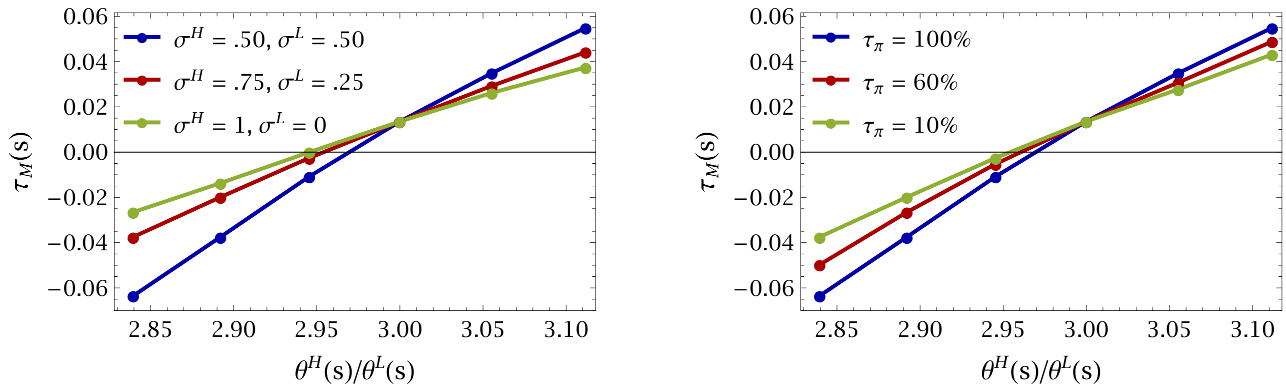

Figure 2. Optimal $\tau_M^*(s^t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$ for different initial distributions of equity (left panel) and for different levels of the profit tax (right panel).

slopes: $\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i$. This key term is the product of two components. The first is ϑ which parameterizes the extent to which profits are left untaxed. The second is $\sum_{i\in I} \pi^i \nu^i \sigma_0^i$, the cross-sectional covariance between initial equity shares, σ_0^i, and the planner's multipliers on the implementability conditions in (44), ν^i.

Recall that households with high lifetime labor earnings have high values of ν^i, and households with low lifetime labor earnings have low values of ν^i. A positive value for $\sum_{i\in I} \pi^i \nu^i \sigma_0^i$ thereby indicates a positive cross-sectional covariance between initial equity and lifetime labor earnings: high human wealth households own greater shares of the firm.

When $\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i$ is strictly positive—when profits are not fully taxed and when initial equity and lifetime labor earnings are positively correlated—this term contributes to a "lower" slope in economy-ϑ than in economy-0. This is consistent with the intuition that when profits are not fully taxed and high human wealth households own greater shares of the firm, an increase in the markup in high-inequality states reduces overall labor income inequality but increases profits in those states. As a result, optimal monetary policy responds less aggressively to movements in $\mathcal{I}(s_t)$. While this is only a heuristic argument, we show that it holds in the following numerical example.

Numerical Illustration. We return to the simple numerical example of two household types, a high-type and a low-type, of equal sizes described in Section 5. We again let the the ratio $\theta^H(s_t)/\theta^L(s_t)$ fluctuate across 6 possible states and we fix all parameter values as in our previous example.[27] In this exercise we vary the distribution of initial equity and the profit tax itself.

In the left panel of Figure 2 we plot the optimal monetary tax $\tau_M^*(s^t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$ for three different economies: one with equal initial firm ownership, one in which the low-type owns 25% and the high-type owns 75%, and one in which the low-type owns zero

[27]We set $\tau_\ell = 0$ so that $\vartheta = 1 - \tau_\Pi$.

shares and the high-type owns 100% of the firm. We keep the profit tax constant, set at $\tau_\Pi = 10\%$. Our baseline ($\vartheta = 0$) is nested by the economy with equal firm ownership. We find that in all three economies, the optimal monetary tax is increasing in the ratio $\theta^H(s_t)/\theta^L(s_t)$. As the distribution of initial shares becomes more unequal, the slope of the optimal monetary tax with respect to the ratio of productivities falls but remains positive.

In the right panel of Figure 2 we plot the optimal monetary tax $\tau_M^*(s^t)$ as a function of $\theta^H(s_t)/\theta^L(s_t)$ for three different economies: one with $\tau_\Pi = 100\%$ as in our baseline, another with $\tau_\Pi = 60\%$, and a third with $\tau_\Pi = 10\%$. We keep constant the initial distribution of equity: the low-type owns 20% of the firm and the high type owns 80%. We find that in all three economies the optimal monetary tax is increasing in the ratio $\theta^H(s_t)/\theta^L(s_t)$. As the profit tax falls, the slope of the optimal monetary tax schedule falls but remains positive.

7 Quantitative Illustration

In this final section we use a simple, calibrated version of the model to compute the model-implied elasticity of the optimal markup with respect to aggregate output.

We use estimates of "worker betas"—the percent change in the growth rate of labor income associated with a percent change in GDP growth—from Guvenen, Schulhofer-Wohl, Song, and Yogo (2017) to construct the functions $\theta^i : S \to \mathbb{R}_+$. We assume 5 household types: the first four types capture the bottom 90 percent of the labor income distribution, while the last type captures the top decile. We partition the type space in this way to capture the non-monoticity of worker betas over the income distribution observed in the data. Worker-betas are monotonically falling in earnings levels throughout most of the income distribution then rise again at the very top of the distribution. Our type partition is able to capture this U-shape of household earnings exposure to GDP growth.[28]

Specifically, we use Treasury Department estimates of the labor income distribution in 2019 to construct the long-run labor productivity distribution.[29] We interpret a period in our model as a year and use annual data on real GDP from the Bureau of Economic Analysis from 1948 to 2023. We detrend the data and, using annual growth rates, we calculate unconditional probabilities for 4 distinct economic states: a severe recession (growth lower than 4.5 percent below trend), a mild recession (growth between 4.5 and 1.9 percent below trend), normal times (between 1.9 percent below trend and 1.7 percent above trend), and high growth (greater than 1.7 percent above trend). Equating the average labor income growth rate with the average growth rate of GDP, we use the worker betas to translate the percent change in GDP growth in each state into percent changes in labor income growth for each household. We then translate these

[28] See Figure 1 in Guvenen, Schulhofer-Wohl, Song, and Yogo (2017).

[29] We use the U.S. Department of the Treasury Table on the Distribution of Income by Source for 2019.

Table 1. Model-implied optimal markups and elasticities

			$\kappa = .25$		$\kappa = .06$	
State	$d\log Y$	μ	$d\log \mathcal{M}$	Elasticity	$d\log \mathcal{M}$	Elasticity
severe recession	–5.23	.09	1.32	–.25	.22	–.04
mild recession	–3.30	.09	1.30	–.39	.20	–.06
normal times	0	.49	0	—	0	—
high growth	3.21	.33	–.8	–.25	–.15	–.05

Notes. This table reports model-implied optimal markups and elasticities when price change frequencies are calibrated to match Nakamura and Steinsson (2008) (κ = .25) and Bils and Klenow (2004) (κ = .06), respectively.

changes in labor income into changes in labor productivity.

We set $\beta = .98$. We use an elasticity of intertemporal substitution of .5 ($\gamma = 2$), following Hall (2009). We set the Frisch elasticity of labor supply to 2 ($\eta = .5$), in line with "macro" elasticities (Hall, 2009; Rogerson and Wallenius, 2009). The elasticity of substitution across goods, ρ, is set to 6, a value used commonly throughout the New Keynesian literature (McKay, Nakamura, and Steinsson, 2016). For aggregate productivity, we use the annual series on total factor productivity (TFP) growth from Fernald (2014) to calculate average TFP growth in the years corresponding to each state.

We calibrate the share of sticky-price firms, κ, by converting estimates of the monthly frequency of price changes from Nakamura and Steinsson (2008) and Bils and Klenow (2004) into annual probabilities of a price remaining unchanged. Nakamura and Steinsson (2008) report that roughly 11 percent of prices change per month; the corresponding number from Bils and Klenow (2004) is 21 percent. Assuming that price changes are i.i.d. across months, these estimates imply κ =.25 and $\kappa = .06$, respectively.[30]

With these parameter values, we solve numerically for the Ramsey optimum with equal Pareto weights: $\lambda_i = 1$ for all $i \in I$. In terms of fiscal policy, we assume profits are fully taxed and follow the implementation in (40).

Results. The first two columns of Table 1 report percent deviations of real GDP from trend and unconditional probabilities for each state. Model-implied optimal markups expressed in percent deviations from normal times, and their elasticities with respect to real GDP, are reported for the two specifications of κ.

We find that in either specification, the optimal markup is counter-cyclical. In our preferred

[30] For example, if 11 percent of prices change within a month, then 89 percent of prices do not change. Assuming that price changes are i.i.d., the probability of a price remaining unchanged within a given year is $(.89)^{12} \approx .247$.

specification of $\kappa = .25$, the optimal markup grows by 1.32 percent in severe recessions and falls by .8 percent in periods of high growth. These numbers imply a range for the elasticity of the optimal markup with respect to real GDP of –.25 to –.39. In the specification with more flexible prices ($\kappa = .06$), the optimal markup grows by .22 percent in severe recessions and falls by .15 percent in high growth periods; implied elasticities range from –.04 to –.06.

The countercyclicality of the optimal markup is the natural consequence of two features: countercyclical earnings inequality (in the data) and optimal monetary policy as prescribed by the model. As noted above, estimates of worker betas feature a striking pattern: earnings exposure to GDP growth is monotonically falling in income throughout the majority of the distribution.[31] As output falls in a recession, the labor income of low-skilled workers declines disproportionately, resulting in an increase in earnings inequality. Countercyclical earnings inequality, coupled with the positive covariance between earnings inequality and the optimal markup, together imply countercyclical optimal markups.

The behavior of the optimal markup in our model is thereby consistent with work that documents countercyclical price markups and, more generally, a countercyclical labor wedge. It is firmly established that the labor wedge, defined as the ratio of the marginal product of labor to the marginal rate of substitution between consumption and leisure, is countercyclical (Hall, 1997; Chari, Kehoe, and McGrattan, 2007). While the labor wedge can arise from both product and labor market distortions, a number of studies find evidence of countercyclical price markups: Bils (1987); Rotemberg and Woodford (1999); Bils, Klenow, and Malin (2018).[32] Given the countercyclicality of earnings inequality in the data, optimal monetary policy in our model is broadly consistent with these findings.

In terms of magnitudes, Bils, Klenow, and Malin (2018) show that, depending on the wage measure used, the price markup elasticity with respect to real GDP can range from –.32 to –2.17. When using the frequency of price changes estimated by Nakamura and Steinsson (2008), the elasticities for the optimal markup implied by our model are consistent with the lower end (in absolute value) of this range.

8 Conclusion

In this paper we study optimal monetary policy in a dynamic, general equilibrium economy with heterogeneous agents and nominal rigidities. Markets are complete; fiscal and monetary policy can be used for redistributional purposes. We find that when household labor productivities fluctuate disproportionately over the business cycle, it is optimal for monetary policy

[31]Although worker betas are large in the top income decile, these workers comprise only 10% of the population and are therefore too small to overturn the counter-cyclicality of income inequality.

[32]See also Bils and Kahn (2000) and Kryvtsov and Midrigan (2013).

to deviate from implementing flexible-price allocations and target a state-contingent markup. The optimal markup co-varies positively with a sufficient statistic for labor income inequality.

In a quantitative illustration of the model, we calibrate the labor income distribution in order to reflect the unequal incidence of GDP fluctuations documented in the data. We show that countercyclical earnings inequality implies countercyclical optimal markups. In our baseline calibration, the elasticity of the optimal markup with respect to real GDP ranges from –.25 to –.39. The behavior of the optimal markup is thereby consistent with work that documents countercyclical price markups and, more broadly, a countercyclical labor wedge.

References

ACHARYA, S., E. CHALLE, AND K. DOGRA (2023): "Optimal Monetary Policy According to HANK," *American Economic Review*, 113, 1741–1782.

ALVES, F., G. KAPLAN, B. MOLL, AND G. L. VIOLANTE (2020): "A Further Look at the Propagation of Monetary Policy Shocks in HANK," *Journal of Money, Credit and Banking*, 52, 521–559.

ANGELETOS, G.-M. AND J. LA'O (2020): "Optimal Monetary Policy with Informational Frictions," *Journal of Political Economy*, 128, 1027–1064.

ATKINSON, A. B. AND J. E. STIGLITZ (1980): *Lectures on Public Economics*, McGraw-Hill, New York.

AUCLERT, A., M. ROGNLIE, AND L. STRAUB (2020): "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model," *NBER Working Paper 26647*.

BASSETTO, M. (2014): "Optimal fiscal policy with heterogeneous agents," *Quantitative Economics*, 5, 675–704.

BHANDARI, A., D. EVANS, M. GOLOSOV, AND T. J. SARGENT (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, 89, 2559–2599.

BILS, M. (1987): "The Cyclical Behavior of Marginal Cost and Price," *The American Economic Review*, 77, 838–855.

BILS, M. AND J. A. KAHN (2000): "What Inventory Behavior Tells Us about Business Cycles," *The American Economic Review*, 90, 458–481.

BILS, M. AND P. J. KLENOW (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.

BILS, M., P. J. KLENOW, AND B. A. MALIN (2018): "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, 108, 1118–46.

CHARI, V., L. CHRISTIANO, AND P. KEHOE (1994): "Optimal Fiscal Policy in a Business Cycle Model," *Journal of Political Economy*, 102, 617–52.

CHARI, V., L. J. CHRISTIANO, AND P. J. KEHOE (1991): "Optimal fiscal and monetary policy: Some recent results," *Journal of Money, Credit and Banking*, 23, 519–539.

CHARI, V. V. AND P. J. KEHOE (1999): "Optimal Fiscal and Monetary Policy," in *Handbook of Macroeconomics*, Elsevier, vol. 1, 1671–1745.

CHARI, V. V., P. J. KEHOE, AND E. R. MCGRATTAN (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.

CORREIA, I. (2010): "Consumption Taxes and Redistribution," *American Economic Review*, 100, 1673–94.

CORREIA, I., E. FARHI, J. P. NICOLINI, AND P. TELES (2013): "Unconventional Fiscal Policy at the Zero Bound," *American Economic Review*, 103, 1172–1211.

CORREIA, I., J. P. NICOLINI, AND P. TELES (2008): "Optimal Fiscal and Monetary Policy: Equivalence Results," *Journal of Political Economy*, 116, 141–170.

DÁVILA, E. AND A. SCHAAB (2023): "Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy," *NBER Working Paper 30961*.

DIAMOND, P. AND J. MIRRLEES (1971): "Optimal Taxation and Public Production: I–Production Efficiency," *American Economic Review*, 61, 8–27.

FERNALD, J. G. (2014): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Working Paper Series 2012-19, Federal Reserve Bank of San Francisco.

GUVENEN, F., S. OZKAN, AND J. SONG (2014): "The nature of countercyclical income risk," *Journal of Political Economy*, 122, 621–660.

GUVENEN, F., S. SCHULHOFER-WOHL, J. SONG, AND M. YOGO (2017): "Worker Betas: Five Facts about Systematic Earnings Risk," *American Economic Review: Papers and Proceedings*, 107, 398–403.

GUVENEN, F. AND A. A. SMITH (2014): "Inferring labor income risk and partial insurance from economic choices," *Econometrica*, 82, 2085–2129.

HALL, R. (2009): “Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor,” *Journal of Political Economy*, 117, 281–323.

HALL, R. AND A. RABUSHKA (1995): *The Flat Tax*, Hoover Institution Press publication, Hoover Institution Press, Stanford University.

HALL, R. E. (1997): “Macroeconomic Fluctuations and the Allocation of Time,” *Journal of Labor Economics*, 15, pp. S223–S250.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): “Consumption and Labor Supply with Partial Insurance: An Analytical Framework,” *American Economic Review*, 104, 2075–2126.

HUGGETT, M., G. VENTURA, AND A. YARON (2011): “Sources of Lifetime Inequality,” *The American Economic Review*, 101, 2923–2954.

JUDD, K. (1985): “Redistributive taxation in a simple perfect foresight model,” *Journal of Public Economics*, 28, 59–83.

KAPLAN, G., B. MOLL, AND G. L. VIOLANTE (2018): “Monetary Policy According to HANK,” *American Economic Review*, 108, 697–743.

KEANE, M. P. AND K. I. WOLPIN (1997): “The Career Decisions of Young Men,” *Journal of Political Economy*, 105, 473–522.

KRYVTSOV, O. AND V. MIDRIGAN (2013): “Inventories, Markups, and Real Rigidities in Menu Cost Models,” *The Review of Economic Studies*, 80, 249–276.

LE GRAND, F., A. MARTIN-BAILLON, AND X. RAGOT (2024): “Should monetary policy care about redistribution? Optimal monetary and fiscal policy with heterogeneous agents,” *Working Paper*.

LUCAS, R. J. AND N. L. STOKEY (1983): “Optimal Fiscal and Monetary policy in an Economy without Capital,” *Journal of Monetary Economics*, 12, 55–93.

MANKIW, N. G. AND R. REIS (2002): “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 117, 1295–1328.

MCKAY, A., E. NAKAMURA, AND J. STEINSSON (2016): “The Power of Forward Guidance Revisited,” *American Economic Review*, 106, 3133–3158.

MCKAY, A. AND C. WOLF (2022): “Optimal Policy Rules in HANK,” *Working Paper*.

NAKAMURA, E. AND J. STEINSSON (2008): “Five Facts about Prices: A Reevaluation of Menu Cost Models,” *The Quarterly Journal of Economics*, 123, 1415–1464.

NEGISHI, T. (1960): “Welfare Economics and Existence of an Equilibrium for a Competitive Economy,” *Metroeconomica*, 12, 92–97.

NIEPELT, D. (2004): “Tax smoothing versus tax shifting,” *Review of Economic Dynamics*, 7, 27–51.

NUÑO, G. AND C. THOMAS (2022): “Optimal Redistributive Inflation,” *Annals of Economics and Statistics*, 146, 3–64.

PARKER, J. A. AND A. VISSING-JORGENSEN (2009): “Who Bears Aggregate Fluctuations and How?” *American Economic Review*, 99, 399–405.

ROGERSON, R. AND J. WALLENIUS (2009): “Micro and macro elasticities in a life cycle model with taxes,” *Journal of Economic Theory*, 144, 2277–2292, dynamic General Equilibrium.

ROTEMBERG, J. J. AND M. WOODFORD (1999): “The cyclical behavior of prices and costs,” in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, Elsevier, vol. 1 of *Handbook of Macroeconomics*, chap. 16, 1051–1135.

STORESLETTEN, K., C. I. TELMER, AND A. YARON (2004): “Consumption and risk sharing over the life cycle,” *Journal of Monetary Economics*, 51, 609–633.

U.S. DEPARTMENT OF THE TREASURY (2019): “Distribution of Income by Source (2019 Income Levels),” Https://home.treasury.gov/system/files/131/Distribution-of-Income-by-Source-2019.pdf.

WERNING, I. (2007): “Optimal Fiscal Policy with Redistribution,” *The Quarterly Journal of Economics*, 122, 925–967.

WOODFORD, M. (2003): “Imperfect Common Knowledge and the Effects of Monetary Policy,” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.*

A Proofs for the Baseline Economy

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:

$$\mu(s^t)U_c^i(s^t) - \mu(s^t)\Lambda^i(s^t)(1+\tau_c)P(s^t) = 0, \tag{48}$$

$$\mu(s^t)\frac{1}{\theta^i(s_t)}U_\ell^i(s^t) + \mu(s^t)\Lambda^i(s^t)(1-\tau_\ell)W(s^t) = 0, \tag{49}$$

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:

$$-\beta^t\mu(s^t)\Lambda^i(s^t) + \beta^{t+1}\sum_{s^{t+1}|s^t}\mu(s^{t+1})\Lambda^i(s^{t+1})(1+i(s^t)) = 0. \tag{50}$$

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

$$-\beta^t\mu(s^t)\Lambda^i(s^t)Q(s^{t+1}|s^t) + \beta^{t+1}\mu(s^{t+1})\Lambda^i(s^{t+1}) = 0. \tag{51}$$

The first-order condition with respect to equity shares $\sigma^i(s^t)$ is given by:

$$-\beta^t\mu(s^t)\Lambda^i(s^t)V(s^t) + \beta^{t+1}\sum_{s^{t+1}|s^t}\mu(s^{t+1})\Lambda^i(s^{t+1})[(1-\tau_\Pi)\Pi(s^{t+1}) + V(s^{t+1})] = 0 \tag{52}$$

The household's transversality conditions are given by:

$$\lim_{t\to\infty}\sum_{s^t}\mu(s^t)\Lambda^i(s^t)b^i(s^t) = 0,$$

$$\lim_{t\to\infty}\sum_{s^t}\mu(s^t)\Lambda^i(s^t)V(s^t)\sigma^i(s^t) = 0,$$

$$\lim_{t\to\infty}\sum_{s^t}\mu(s^t)\Lambda^i(s^t)Q(s^{t+1}|s^t)z^i(s^{t+1}) = 0.$$

Combining (48) and (49), we obtain the household's intratemporal condition:

$$-\frac{1}{\theta^i(s_t)}\frac{U_\ell^i(s^t)}{U_c^i(s^t)} = \frac{(1-\tau_\ell)W(s^t)}{(1+\tau_c)P(s^t)} \tag{53}$$

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

$$\frac{U_c^i(s^t)}{P(s^t)} = \beta(1+i(s^t))\sum_{s^{t+1}|s^t}\mu(s^{t+1}|s^t)\frac{U_c^i(s^{t+1})}{P(s^{t+1})}, \tag{54}$$

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.

Furthermore, each Arrow security price satisfies:

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

and the equity share price satisfies:

$$V(s^t) = \beta \sum_{s^{t+1}|s^t} \mu(s^{t+1}|s^t)\frac{U_c^i(s^{t+1})}{P(s^{t+1})}\frac{P(s^t)}{U_c^i(s^t)}[(1-\tau_\Pi)\Pi(s^{t+1}) + V(s^{t+1})]. \tag{56}$$

A.2 Proof of Lemma 1

Markets are complete. The single, lifetime budget constraint of the household of type i can be represented as:

$$\sum_t \sum_{s^t} \hat{q}(s^t)\left[(1+\tau_c)c^i(s^t) - (1-\tau_\ell)\frac{W(s^t)}{P(s^t)}\ell^i(s^t)\right] = \sum_t \sum_{s^t} \hat{q}(s^t)\left[T(s^t) + (1+\sigma_0^i)(1-\tau_\Pi)\frac{\Pi(s^t)}{P(s^t)}\right]$$

where $\hat{q}(s^t)$ represents the Arrow-Debreu price of one unit of consumption in period t, history s^t, normalized so that $\hat{q}(s_0) = 1$, $W(s^t)/P(s^t)$ is the real wage, and $\Pi(s^t)/P(s^t)$ are real profits. Let $1/\varphi^i$ denote the Lagrange multiplier on this budget constraint. The household's first-order conditions with respect to $c^i(s^t)$ and $\ell^i(s^t)$ can be written as follows:

$$\varphi^i U_c^i(s^t) - (1+\tau_c)\frac{\hat{q}(s^t)}{\beta^t\mu(s^t)} = 0,$$
$$\varphi^i \frac{1}{\theta^i(s_t)}U_\ell^i(s^t) + (1-\tau_\ell)\frac{\hat{q}(s^t)}{\beta^t\mu(s^t)}\frac{W(s^t)}{P(s^t)} = 0,$$

These conditions hold for all t, s^t and for all types $i \in I$. These conditions imply that in equilibrium, for any period t, history s^t:

$$\varphi^i U_c^i(s^t) = \varphi^j U_c^j(s^t), \qquad \forall i,j \in I; \tag{57}$$
$$\varphi^i \frac{1}{\theta^i(s_t)}U_\ell^i(s^t) = \varphi^j \frac{1}{\theta^j(s_t)}U_\ell^j(s^t), \qquad \forall i,j \in I; \tag{58}$$
$$-\frac{1}{\theta^i(s_t)}\frac{U_\ell^i(s^t)}{U_c^i(s^t)} = -\frac{1}{\theta^j(s_t)}\frac{U_\ell^j(s^t)}{U_c^j(s^t)}, \qquad \forall i,j \in I. \tag{59}$$

These conditions and the resource constraints in (11) pin down the equilibrium allocation.

Consider now the static subproblem described in Lemma 1. Take any t, s^t and let $\rho_C(s^t)$ and $\rho_L(s^t)$ be the Lagrange multipliers on the constraints in (11). The first-order conditions to this problem are given by

$$\varphi^i U_c^i(s^t) - \rho_C(s^t) = 0, \qquad \forall i \in I$$
$$\varphi^i \frac{1}{\theta^i(s_t)}U_\ell^i(s^t) + \rho_L(s^t) = 0, \qquad \forall i \in I$$

These conditions imply (57)-(59). Again, these conditions, along with the resource constraints (11), pin down the allocation. It follows that the solution to the sub-problem coincides with the equilibrium allocation. The envelope conditions for the static sub-problem are given by:

$$U_C^m(s^t) = \varphi^i U_c^i(s^t) \qquad \text{and} \qquad U_L^m(s^t) = \varphi^i \frac{1}{\theta^i(s_t)} U_\ell^i(s^t), \qquad \forall i \in I.$$

Next, with the separable and iso-elastic preferences assumed in (2), conditions (57) and (58) can be written as:

$$\begin{aligned} \varphi^i c^i(s^t)^{-\gamma} =& \varphi^j c^j(s^t)^{-\gamma}, \qquad \forall i, j \in I; \\ \varphi^i \frac{1}{\theta^i(s_t)} \left[\frac{\ell^i(s^t)}{\theta^i(s_t)}\right]^\eta =& \varphi^j \frac{1}{\theta^j(s_t)} \left[\frac{\ell^j(s^t)}{\theta^j(s_t)}\right]^\eta, \qquad \forall i, j \in I. \end{aligned}$$

Combining these conditions with the resource constraints in (11), 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 (20). We take the household's budget constraint in (3) for type $i \in I$, multiply both sides by $\Lambda^i(s^t)$, and use the household's FOCs in (48) and (49) to substitute out consumption and labor prices. Doing so, we obtain:

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

where we let

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

Multiplying both sides by $\beta^t \mu(s^t)$, summing over t and s^t, and using the household's intertemporal optimality conditions (54)-(56) to cancel terms, we obtain:

$$\begin{aligned} \sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_c^i(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t)\ell^i(s^t)\right] =& \sum_t \sum_{s^t} \beta^t \mu(s^t) \Lambda^i(s^t) P(s^t) \bar{T}(s^t) \\ &+ \sigma_0^i \Lambda^i(s_0)[(1 - \tau_\Pi)\Pi((s_0) + V(s_0)] \end{aligned}$$

We can rewrite this as:

$$\begin{aligned} \sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_c^i(s^t)c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t)\ell^i(s^t)\right] =& U_c^i(s_0)\bar{T}_i \\ &+ \sigma_0^i \Lambda^i(s_0)[(1 - \tau_\Pi)\Pi((s_0) + V(s_0)] \end{aligned} \tag{60}$$

where

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

Next, we use the household optimality condition 52 to write the equity share price as follows:

$$V(s^t) = \beta \sum_{s^{t+1}|s^t} \mu(s^{t+1}|s^t) \frac{\Lambda^i(s^{t+1})}{\Lambda^i(s^t)} [(1-\tau_\Pi)\Pi(s^{t+1}) + V(s^{t+1})].$$

Iterating this forward, we have that the share price satisfies:

$$V(s^t) = \sum_{\tau=t+1}^{\infty} \sum_{s^\tau|s^t} \beta^{\tau-t} \mu(s^\tau|s^t) \frac{\Lambda^i(s^\tau)}{\Lambda^i(s^t)} (1-\tau_\Pi)\Pi(s^\tau)$$

Therefore at time 0, the price per share is given by:

$$V(s_0) = \sum_{t=1}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \frac{\Lambda^i(s^t)}{\Lambda^i(s_0)} (1-\tau_\Pi)\Pi(s^t)$$

Substituting this into 60 and using the fact that $U_c^i(s^t) = \Lambda^i(s^t)(1+\tau_c)P(s^t)$, we obtain:

$$\begin{aligned} &\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_c^i(s^t) c^i(s^t) + \frac{1}{\theta^i(s_t)} U_\ell^i(s^t) \ell^i(s^t) \right] \\ &= U_c^i(s_0)\bar{T}_i + \sigma_0^i (1-\tau_\Pi) \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \frac{U_c^i(s^t)}{(1+\tau_c)} \frac{\Pi(s^t)}{P(s^t)} \end{aligned}$$

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:

$$\begin{aligned} &\sum_t \sum_{s^t} \beta^t \mu(s^t) \left[U_C^m(s^t) \omega_C^i(\varphi) C(s^t) + U_L^m(s^t) \omega_L^i(\varphi, s_t) L(s^t) \right] \\ &= U_C^m(s_0)\bar{T} + \sigma_0^i (1-\tau_\Pi) \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \mu(s^t) \frac{U_C^m(s^t)}{(1+\tau_c)} \frac{\Pi(s^t)}{P(s^t)} \end{aligned}$$

where

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

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

A.4 Derivation of Sticky-Price Firm Optimality

The sticky-price firm solves the following problem:

$$\max_{p'} \sum_{s^t|s^{t-1}} Q(s^t|s^{t-1}) \left\{ (1-\tau_r) p' \left(\frac{p'}{P(s^t)} \right)^{-\rho} Y(s^t) - \frac{W(s^t)}{A(s_t)} \left(\frac{p'}{P(s^t)} \right)^{-\rho} Y(s^t) \right\}.$$

The first-order condition with respect to p' is given by

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

Rearranging gives us:

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

Substituting in the equilibrium Arrow prices from (14) yields:

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1}) \frac{U_C^m(s^t)}{P(s^t)} Y(s^t) P(s^t)^{\rho} \left\{ p_t^s(s^{t-1}) - \left[(1-\tau_r) \left(\frac{\rho-1}{\rho} \right) \right]^{-1} \frac{W(s^t)}{A(s_t)} \right\} = 0.$$

Solving this for $p_t^s(s^{t-1})$ gives us (22) with $q(s^t|s^{t-1})$ defined as follows:

$$q(s^t|s^{t-1}) \equiv \frac{\mu(s^t|s^{t-1}) U_C^m(s^t) C(s^t) P(s^t)^{\rho-1}}{\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1}) U_C^m(s^t) C(s^t) P(s^t)^{\rho-1}}. \tag{61}$$

A.5 Proof of Proposition 1

Necessity. In any flexible-price equilibrium, all firms set the same nominal price. The demand functions in (5) imply that all firms produce the same level of output, proving necessity of $y^j(s^t) = Y(s^t)$ for all $j \in \mathcal{J}$.

Aggregation over the optimal price (21) implies that the aggregate price level is given by:

$$P(s^t) = \left[(1-\tau_r) \left(\frac{\rho-1}{\rho} \right) \right]^{-1} \frac{W(s^t)}{A(s_t)}, \qquad \forall s^t \in S^t. \tag{62}$$

Condition (23) follows from combining (62) with the household's intratemporal optimality condition (12), and letting χ denote the labor wedge as follows:

$$\chi \equiv \left(\frac{\rho-1}{\rho} \right) \frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c}. \tag{63}$$

Finally, the derivation of the set of necessary conditions (20) is provided in Appendix A.3.

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

First, for all $s^t \in S^t$, we normalize the aggregate price level to one and set intermediate-good prices according to:

$$p_t^j(s^t) = p_t^f(s^t) = P(s^t) = 1, \qquad \forall j \in \mathcal{J}.$$

These prices, combined with condition (i) of Proposition 1, ensure that the CES demand function (5) is satisfied for all goods, $j \in \mathcal{J}$.

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

$$\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c} = \left(\frac{\rho-1}{\rho}\right)^{-1}\chi. \tag{64}$$

For any strictly positive χ and $\rho > 1$, such tax rates exist. Combining this with condition (23), we obtain the following:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{1+\tau_c}{1-\tau_\ell}\right) = \left(\frac{\rho-1}{\rho}\right)(1-\tau_r)A(s_t). \tag{65}$$

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

$$W(s^t) = -\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{1+\tau_c}{1-\tau_\ell}\right), \tag{66}$$

and therefore satisfy the household's intratemporal condition in (12). Substituting the above expression for the real wage into (65) and re-arranging gives us:

$$p_t^f(s^t) = 1 = \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{A(s_t)}. \tag{67}$$

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

Next, for all $s^t \in S^t$, we set the Arrow prices, the nominal interest rate, and the ex-dividend share price as follows:

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)}, \quad \forall s^{t+1}|s^t;$$

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

$$V(s^t) = \sum_{\tau=t+1}^{\infty}\sum_{s^\tau|s^t}\beta^{\tau-t}\mu(s^\tau|s^t)\frac{U_C^m(s^\tau)}{U_C^m(s^t)}(1-\tau_\Pi)\Pi(s^\tau).$$

We therefore satisfy equilibrium conditions (13)-(15).

What remains to be shown is that we can construct financial asset holdings such that the household's budget constraints are satisfied at this allocation in every history. Given $\bar{T}$, we first construct a sequence $\{\bar{T}(s^t)\}_{t\geq 0, s^t \in S^t}$ that satisfies the following condition:

$$\bar{T} = \frac{1}{U_c^m(s_0)(1+\tau_c)}\sum_t\sum_{s^t}\beta^t\mu(s^t)U_c^m(s^t)\bar{T}(s^t). \tag{68}$$

Given such a sequence $\{\bar{T}(s^t)\}$, we set transfers such that $T(s^t) = \bar{T}(s^t) - (1-\tau_\Pi)\Pi(s^t)$ for all $s^t \in S^t$. Next, we take the household's budget constraint in (3) for type $i \in I$ for all periods and states following and including period r, history s^r; we multiply these budget constraints by $\beta^{t-r}\mu(s^t|s^r)\Lambda^i(s^t)$ and sum over all periods and states following and including period r, history s^r. Doing so, we get:

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

Using the household's FOCs for financial assets, (50)-(52), the above equation reduces to:

$$\sum_{t=r}^{\infty}\sum_{s^t|s^r}\beta^{t-r}\mu(s^t|s^r)\Lambda^i(s^t)\left[(1+\tau_c)c^i(s^t)\right]=\sum_{t=r}^{\infty}\sum_{s^t|s^r}\beta^{t-r}\mu(s^t|s^r)\Lambda^i(s^t)\left[(1-\tau_\ell)W(s^t)\ell^i(s^t)+\bar{T}(s^t)\right]$$
$$+\Lambda^i(s^r)\left[(1+i(s^{r-1}))b^i(s^{r-1})+z^i(s^r|s^{r-1})\right]$$
$$+\Lambda^i(s^r)\sigma^i(s^{r-1})\left[(1-\tau_\Pi)\Pi(s^r)+V(s^r)\right] \tag{69}$$

Next, we define $a^i(s^r)$ as:

$$a^i(s^r)\equiv(1+i(s^{r-1}))b^i(s^{r-1})+z^i(s^r|s^{r-1})+\sigma^i(s^{r-1})\left[(1-\tau_\Pi)\Pi(s^r)+V(s^r)\right] \tag{70}$$

Therefore $a^i(s^r)$ represents the total financial assets (cash-on-hand) that household i carries into period r, history s^r. Rearranging (69) gives us:

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

Next, using the household's FOCs for consumption and labor, (48) and (49), we obtain the following expression for the total financial assets that household i carries into period r, history s^r:

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

This can equivalently be written as follows:

$$a^i(s^r)=\left(\frac{U_C^m(s^r)}{1+\tau_c}\right)^{-1}\sum_{t=r+1}^{\infty}\sum_{s^t}\beta^{t-r}\mu(s^t|s^r)\left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t)+U_L^m(s^t)\omega_L^i(\varphi,s_t)L(s^t)\right]$$
$$-\sum_{t=r+1}^{\infty}\sum_{s^t}\beta^{t-r}\mu(s^t|s^r)\frac{U_C^m(s^t)}{U_C^m(s^r)}\bar{T}(s^t).$$

Finally, the government's budget constraint holds by Walras's law.

A.6 Proof of Proposition 2

Necessity. Condition (21) indicates that all flexible-price firms set the same nominal price; similarly condition (22) indicates that all sticky-price firms set the same nominal price. Combining this observation with the demand functions,

$$\frac{y^f(s^t)}{Y(s^t)} = \left[\frac{p_t^f(s^t)}{P(s^t)}\right]^{-\rho} \quad \text{and} \quad \frac{y^s(s^t)}{Y(s^t)} = \left[\frac{p_t^s(s^{t-1})}{P(s^t)}\right]^{-\rho}. \tag{71}$$

we infer 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 flexible price firm sets its price according to (21). Rearranging and dividing through by $P(s^t)$ gives us:

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

The flexible-price firm optimality condition can be written as follows:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} - \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{P(s^t)A(s_t)} = 0.$$

Combining the above condition with the household's intratemporal optimality condition (12) yields equilibrium necessary condition (25) with χ defined in (63).

As shown in Section A.4 of the Appendix, the sticky price firm sets its price according to

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)Y(s^t)\left(\frac{p_t^s(s^{t-1})}{P(s^t)}\right)^{-\rho}\left\{\frac{p_t^s(s^{t-1})}{P(s^t)} - \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{P(s^t)A(s_t)}\right\} = 0.$$

Using the CES demand function (71), the previous condition can be written as:

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)Y(s^t)\frac{y^s(s^t)}{Y(s^t)}\left\{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} - \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{P(s^t)A(s_t)}\right\} = 0$$

Combining the above condition with the household's intratemporal optimality condition (12) yields the equilibrium necessary condition (26) with χ defined in (63). Finally, the derivation of the set of necessary conditions (20) is provided in Appendix A.3.

Sufficiency. Take any feasible allocation $x \in \mathcal{X}$, vector $\varphi \equiv (\varphi^i)$, and scalars $\bar{T} \in \mathbb{R}$ and $\chi \in \mathbb{R}_+$ that satisfy conditions (i)-(iii) of Proposition 2. We show that there exists a price system $\mathcal{R}$, a policy $\mathcal{P}$, and a set of financial market positions $\mathcal{A}$, that support x as a sticky-price equilibrium; we construct these as follows.

First, we construct nominal prices as follows. Let $\mathcal{B}_t(s^{t-1}) > 0$ denote the common belief of the aggregate price level at time t based on history s^{t-1}; aside from being strictly positive,

$\mathcal{B}_t(s^{t-1}) > 0$ is a free parameter in our model. We set $p_t^s(s^{t-1}) = \mathcal{B}_t(s^{t-1})$. Next, we can decompose the sticky- and flexible-price firm output each into two components:

$$y^s(s^t) = \phi^s(s^{t-1})\Phi(s^t) \qquad \text{and} \qquad y^f(s^t) = \phi^f(s^t)\Phi(s^t). \tag{72}$$

where we set $\phi^s(s^{t-1}) \equiv \mathcal{B}_t(s^{t-1})^{-\rho}$. Therefore,

$$p_t^s(s^{t-1}) = \phi^s(s^{t-1})^{-1/\rho}.$$

The output decomposition in (72) implies $\Phi(s^t) = y^s(s^t)/\mathcal{B}_t(s^{t-1})^{-\rho}$ and $\phi^f(s^t) = y^f(s^t)/\Phi(s^t)$. Finally, we set the price of the flexible-price firm as follows:

$$p_t^f(s^t) = \phi^f(s^t)^{-1/\rho}.$$

Note that these prices, along with the feasibility constraint (7), imply that the aggregate price level is given by:

$$P(s^t) = \left[\frac{Y(s^t)}{\Phi(s^t)}\right]^{-1/\rho}. \tag{73}$$

These prices furthermore ensure that the CES demand curves in (71) are satisfied. We set the money supply such that $M(s^t) = P(s^t)Y(s^t)$.

Next, we set the tax rates $(\tau_\ell, \tau_c, \tau_r)$ such that they jointly satisfy (64). For any strictly positive χ and $\rho > 1$, such tax rates exist. Combining this with conditions (25) and (26), we obtain the following two conditions:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1+\tau_c}{1-\tau_\ell}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)} = 0, \tag{74}$$

and

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left\{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1+\tau_c}{1-\tau_\ell}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)}\right\} = 0. \tag{75}$$

Given tax rates $(\tau_\ell, \tau_c, \tau_r)$, we set the nominal wage as follows:

$$W(s^t) = -\frac{U_L^m(s^t)}{U_C^m(s^t)}\left(\frac{1+\tau_c}{1-\tau_\ell}\right)P(s^t), \tag{76}$$

and therefore satisfy the household's intratemporal condition in (12). Substituting the above expression for the real wage into (74) and (75) rearranging gives us:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} - \frac{W(s^t)}{P(s^t)}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)} = 0.$$

and

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left\{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} - \frac{W(s^t)}{P(s^t)}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)}\right\} = 0.$$

Combining these with the CES demand functions in (71) we get the following two conditions:

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

and

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)Y(s^t)\left(\frac{p_t^s(s^{t-1})}{P(s^t)}\right)^{-\rho}\left\{\frac{p_t^s(s^{t-1})}{P(s^t)} - \frac{W(s^t)}{P(s^t)}\left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)}\right\} = 0.$$

Finally, with some rearrangement, these imply:

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

and

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})\frac{U_C^m(s^t)}{P(s^t)}Y(s^t)P(s^t)^\rho\left\{p_t^s(s^{t-1}) - \left[(1-\tau_r)\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{A(s_t)}\right\} = 0.$$

Therefore both the flexible-price and the sticky-price firm's optimality conditions, (21) and (22), are satisfied.

Next, for all $s^t \in S^t$, we set the Arrow prices, the nominal interest rate, and the ex-dividend share price as follows:

$$Q(s^{t+1}|s^t) = \beta\mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{U_C^m(s^t)}\frac{P(s^t)}{P(s^{t+1})}, \quad \forall s^{t+1}|s^t;$$

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

$$V(s^t) = \sum_{\tau=t+1}^{\infty}\sum_{s^\tau|s^t}\beta^{\tau-t}\mu(s^\tau|s^t)\frac{U_C^m(s^\tau)}{U_C^m(s^t)}\frac{P(s^t)}{P(s^\tau)}(1-\tau_\Pi)\Pi(s^\tau).$$

We therefore satisfy equilibrium conditions (13)-(15).

What remains to be shown is that we can construct financial asset holdings such that the household's budget constraints are satisfied at this allocation in every history. For this we follow the same steps as in the sufficiency portion of the proof of Proposition 1. Given $\bar{T}$, we first construct a sequence $\{\bar{T}(s^t)\}_{t\geq 0, s^t\in S^t}$ that satisfies (68). Given such a sequence $\{\bar{T}(s^t)\}$, we set transfers such that $T(s^t) = \bar{T}(s^t) - (1-\tau_\Pi)\Pi(s^t)/P(s^t)$ for all $s^t \in S^t$. Next, we take the household's budget constraint in (3) for type $i \in I$ for all periods and states following and including period r, history s^r; we multiply these budget constraints by $\beta^{t-r}\mu(s^t|s^r)\Lambda^i(s^t)$ and sum over all periods and states following and including period r, history s^r. Doing so, we get:

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

Using the household's FOCs for financial assets, (50)-(52), the above equation reduces to:

$$\begin{aligned}&\sum_{t=r}^{\infty}\sum_{s^t|s^r}\beta^{t-r}\mu(s^t|s^r)\Lambda^i(s^t)\left[(1+\tau_c)P(s^t)c^i(s^t)-(1-\tau_\ell)W(s^t)\ell^i(s^t)-P(s^t)\bar{T}(s^t)\right]\\&=\Lambda^i(s^r)\left[(1+i(s^{r-1}))b^i(s^{r-1})+z^i(s^r|s^{r-1})\right]\\&\quad+\Lambda^i(s^r)\sigma^i(s^{r-1})\left[(1-\tau_\Pi)\Pi(s^r)+V(s^r)\right]\end{aligned} \tag{77}$$

We define $a^i(s^r)$ as the total *nominal* financial assets (cash-on-hand) that household i carries into period r, history s^r, given by (70). Rearranging (77) gives us:

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

Next, using conditions (48) and (49), we obtain the following expression for the total *real* financial assets that household i carries into period r, history s^r:

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

This can equivalently be written as follows:

$$\begin{aligned}\frac{a^i(s^r)}{P(s^t)}=&\left(\frac{U_C^m(s^r)}{1+\tau_c}\right)^{-1}\sum_{t=r+1}^{\infty}\sum_{s^t}\beta^{t-r}\mu(s^t|s^r)\left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t)+U_L^m(s^t)\omega_L^i(\varphi,s_t)L(s^t)\right]\\&-\sum_{t=r+1}^{\infty}\sum_{s^t}\beta^{t-r}\mu(s^t|s^r)\frac{U_C^m(s^t)}{U_C^m(s^r)}\bar{T}(s^t).\end{aligned}$$

Finally, the government's budget constraint holds by Walras's law.

A.7 Proof of Proposition 3

The Relaxed Ramsey planner's problem is to choose an allocation $x\in\mathcal{X}$, a vector $\varphi\equiv(\varphi^i)$, and scalar $\bar{T}\in\mathbb{R}$, in order to maximize the pseudo-welfare function in (31) subject to technology and resource constraints (7)-(8). First, note that in 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 sub-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, we can rewrite the relaxed planner's problem in terms of aggregates alone:

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

subject to

$$C(s^t) = A(s_t)L(s^t), \qquad \forall s^t \in S^t. \tag{78}$$

We let $\beta^t\mu(s^t)\hat{\varsigma}(s^t)$ denote the Lagrange multiplier on the time t, history s^t resource constraint (78). The first-order conditions of this problem are given by:

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

Combining, we obtain the relaxed planner's optimality condition in (32).

A.8 The Ramsey Optimum

In this section of the appendix, we solve the Ramsey problem stated in Section 5. Recall that in our statement of the Ramsey problem, we allow for the inequality constraint: $C(s^t) \le Y(s^t)$; that is, the planner has free disposal of the final good. We let $\beta^t\mu(s^t)(1-\kappa)\xi(s^t)$ and $\beta^t\mu(s^{t-1})\kappa\upsilon(s^{t-1})$ denote the Lagrange multipliers on the implementability conditions (25) and (26), respectively. We obtain the following Ramsey optimality condition.

Proposition 6. *A Ramsey optimum x^* satisfies, for all $s^t \in S^t$,*

$$-\frac{\mathcal{W}_L(s^t) + (U_L^m(s^t) + U_{LL}^m(s^t)L(s^t))\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{A(s_t)L(s^t)} + (1-\kappa)\xi(s^t)\frac{y^f(s^t)}{A(s_t)L(s^t)}\right]}{\mathcal{W}_C(s^t) + \chi(U_C^m(s^t) + U_{CC}^m(s^t)C(s^t))\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1-1/\rho}\right]} = \frac{Y(s^t)}{L(s^t)}. \tag{79}$$

The proof of Proposition 6 is found below. Note first that we can rewrite condition (79) as it is stated in the main text in equation (36), with

$$[\text{Ramsey wedge}(s^t)] \equiv \frac{1 + \left(\frac{U_L^m(s^t)+U_{LL}^m(s^t)L(s^t)}{\mathcal{W}_L(s^t)}\right)\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{A(s_t)L(s^t)} + (1-\kappa)\xi(s^t)\frac{y^f(s^t)}{A(s_t)L(s^t)}\right]}{1 + \chi\left(\frac{U_C^m(s^t)+U_{CC}^m(s^t)C(s^t)}{\mathcal{W}_C(s^t)}\right)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1-1/\rho}\right]}.$$

Proof. We write the planner's Lagrangian as follows:

$$\begin{aligned}\mathcal{L} =&\sum_t \sum_{s^t} \beta^t \mu(s^t) \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda)\\
&+\sum_t \sum_{s^t} \beta^t \mu(s^t) \varsigma^Y(s^t) \left\{ \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} - Y(s^t) \right\}\\
&+\sum_t \sum_{s^t} \beta^t \mu(s^t) \varsigma^L(s^t) \left\{ \kappa \frac{y^s(s^t)}{A(s_t)} + (1-\kappa) \frac{y^f(s^t)}{A(s_t)} - L(s^t) \right\}\\
&+\sum_t \sum_{s^t} \beta^t \mu(s^t) \varsigma^C(s^t) \left\{ Y(s^t) - C(s^t) \right\}\\
&+\sum_t \sum_{s^t} \beta^t \mu(s^{t-1}) \kappa \upsilon(s^{t-1}) \sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1}) y^s(s^t) \left\{ \chi U_C^m(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} + U_L^m(s^t) \frac{1}{A(s_t)} \right\}\\
&+\sum_t \sum_{s^t} \beta^t \mu(s^t)(1-\kappa)\xi(s^t) y^f(s^t) \left\{ \chi U_C^m(s^t) \left(\frac{y^f(s^t)}{Y(s^t)} \right)^{-1/\rho} + U_L^m(s^t) \frac{1}{A(s_t)} \right\}\end{aligned}$$

with Karush-Kuhn-Tucker conditions:

$$Y(s^t) - C(s^t) \geq 0, \quad \varsigma^C(s^t) \geq 0, \quad \text{and} \quad \varsigma^C(s^t)[Y(s^t) - C(s^t)] = 0, \quad \forall s^t \in S^t.$$

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

$$\begin{aligned}0 =&\kappa \varsigma^Y(s^t) \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}-1} y^s(s^t)^{\frac{\rho-1}{\rho}-1} + \kappa \varsigma^L(s^t) \frac{1}{A(s_t)}\\
&+\kappa \upsilon(s^{t-1}) \left\{ \chi U_C^m(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} + U_L^m(s^t) \frac{1}{A(s_t)} - \frac{1}{\rho} \chi U_C^m(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} \right\}\end{aligned} \tag{80}$$

The FOC with respect to $y^f(s^t)$

$$\begin{aligned}0 =&(1-\kappa) \varsigma^Y(s^t) \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}-1} y^f(s^t)^{\frac{\rho-1}{\rho}-1} + (1-\kappa) \varsigma^L(s^t) \frac{1}{A(s_t)}\\
&+(1-\kappa)\xi(s^t) \left\{ \chi U_C^m(s^t) \left(\frac{y^f(s^t)}{Y(s^t)} \right)^{-1/\rho} + U_L^m(s^t) \frac{1}{A(s_t)} - \frac{1}{\rho} \chi U_C^m(s^t) \left[\frac{y^f(s^t)}{Y(s^t)} \right]^{-1/\rho} \right\}\end{aligned} \tag{81}$$

Note that (80) can equivalently be written as:

$$\begin{aligned}0 =&\kappa \varsigma^Y(s^t) \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}-1} y^s(s^t)^{\frac{\rho-1}{\rho}} + \kappa \varsigma^L(s^t) \frac{y^s(s^t)}{A(s_t)}\\
&+\kappa \upsilon(s^{t-1}) y^s(s^t) \left\{ \chi U_C^m(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} \frac{\rho-1}{\rho} + U_L^m(s^t) \frac{1}{A(s_t)} \right\}\end{aligned} \tag{82}$$

or

$$0 = \varsigma^Y(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} + \varsigma^L(s^t) \frac{1}{A(s_t)} + \upsilon(s^{t-1}) \left\{ \chi U_C^m(s^t) \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{-1/\rho} \frac{\rho-1}{\rho} + U_L^m(s^t) \frac{1}{A(s_t)} \right\} \tag{83}$$

Similarly, note that (81) can equivalently be written as:

$$0 =(1-\kappa)\varsigma^Y(s^t)\left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}-1} y^f(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\varsigma^L(s^t)\frac{y^f(s^t)}{A(s_t)} \\ + (1-\kappa)\xi(s^t)y^f(s^t)\left\{\chi U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{\rho-1}{\rho} + U_L^m(s^t)\frac{1}{A(s_t)}\right\} \tag{84}$$

or

$$0 = \varsigma^Y(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \varsigma^L(s^t)\frac{1}{A(s_t)} + \xi(s^t)\left\{\chi U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{\rho-1}{\rho} + U_L^m(s^t)\frac{1}{A(s_t)}\right\} \tag{85}$$

Adding (82) to (83) gives us:

$$0 =\varsigma^Y(s^t)Y(s^t) + \varsigma^L(s^t)L(s^t) \\ + \chi\frac{\rho-1}{\rho}U_C^m(s^t)Y(s^t)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho} + (1-\kappa)\xi(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{1-1/\rho}\right] \\ + U_L^m(s^t)\frac{1}{A(s_t)}\left[\kappa\upsilon(s^{t-1})y^s(s^t) + (1-\kappa)\xi(s^t)y^f(s^t)\right] \tag{86}$$

We can rewrite the above condition as follows:

$$-\frac{\varsigma^L(s^t) + U_L^m(s^t)\frac{1}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})y^s(s^t) + (1-\kappa)\xi(s^t)y^f(s^t)\right]}{\varsigma^Y(s^t) + \chi\left(1-\frac{1}{\rho}\right)U_C^m(s^t)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho} + (1-\kappa)\xi(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{1-1/\rho}\right]} = \frac{Y(s^t)}{L(s^t)} \tag{87}$$

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

$$0 = \mathcal{W}_C(s^t) - \varsigma^C(s^t) + \kappa\upsilon(s^{t-1})\chi y^s(s^t)U_{CC}^m(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} + (1-\kappa)\xi(s^t)\chi y^f(s^t)U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}, \tag{88}$$

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

$$0 = -\varsigma^Y(s^t) + \varsigma^C(s^t) + \frac{1}{\rho}\kappa\upsilon(s^{t-1})\chi U_C^m(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{y^s(s^t)}{Y(s^t)} + \frac{1}{\rho}(1-\kappa)\xi(s^t)\chi U_C^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}\frac{y^f(s^t)}{Y(s^t)} \tag{89}$$

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

$$0 = \mathcal{W}_L(s^t) - \varsigma^L(s^t) + \kappa\upsilon(s^{t-1})y^s(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)} + (1-\kappa)\xi(s^t)y^f(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)}, \tag{90}$$

Combining (88) and (89) we get:

$$\varsigma^Y(s^t) =\mathcal{W}_C(s^t) + \kappa\upsilon(s^{t-1})\chi y^s(s^t)U_{CC}^m(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} + (1-\kappa)\xi(s^t)\chi y^f(s^t)U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho} \\ + \frac{1}{\rho}\kappa\upsilon(s^{t-1})\chi U_C^m(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{y^s(s^t)}{Y(s^t)} + \frac{1}{\rho}(1-\kappa)\xi(s^t)\chi U_C^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}\frac{y^f(s^t)}{Y(s^t)}$$

This reduces to:

$$\varsigma^Y(s^t) = \mathcal{W}_C(s^t)+\chi\left\{U^m_{CC}(s^t)Y(s^t)+\frac{1}{\rho}U^m_C(s^t)\right\}\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho}+(1-\kappa)\xi(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{1-1/\rho}\right] \tag{91}$$

and from (90) we have:

$$\varsigma^L(s^t) = \mathcal{W}_L(s^t)+U^m_{LL}(s^t)\frac{1}{A(s_t)}\left[\kappa\upsilon(s^{t-1})y^s(s^t)+(1-\kappa)\xi(s^t)y^f(s^t)\right] \tag{92}$$

Substituting these into (87) and noting that $Y(s^t)=C(s^t)$, for all $s^t\in S^t$, we obtain:

$$-\frac{\mathcal{W}_L(s^t)+\left\{U^m_{LL}(s^t)\frac{1}{A(s_t)}+U^m_L(s^t)\frac{1}{A(s_t)L(s^t)}\right\}\left[\kappa\upsilon(s^{t-1})y^s(s^t)+(1-\kappa)\xi(s^t)y^f(s^t)\right]}{\mathcal{W}_C(s^t)+\chi\left\{U^m_{CC}(s^t)C(s^t)+U^m_C(s^t)\right\}\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}+(1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]}=\frac{Y(s^t)}{L(s^t)},$$

and $\varsigma^C(s^t)>0$ for all $s^t\in S^t$. The above equation coincides with (79). □

A.9 Proof of Theorem 2

Iso-elastic preferences satisfy:

$$\frac{U^m_{CC}(s^t)C(s^t)}{U^m_C(s^t)}=-\gamma \qquad \text{and} \qquad \frac{U^m_{LL}(s^t)L(s^t)}{U^m_L(s^t)}=\eta$$

This implies that (79) can be written as follows:

$$-\frac{\mathcal{W}_L(s^t)+(1+\eta)U^m_L(s^t)\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{Y(s^t)}+(1-\kappa)\xi(s^t)\frac{y^f(s^t)}{Y(s^t)}\right]}{\mathcal{W}_C(s^t)+\chi(1-\gamma)U^m_C(s^t)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}+(1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]}=\frac{Y(s^t)}{L(s^t)} \tag{93}$$

This holds for any arbitrary χ. We combine this with the implementability condition (25) and obtain:

$$\frac{\frac{\mathcal{W}_L(s^t)}{U^m_L(s^t)}+(1+\eta)\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{Y(s^t)}+(1-\kappa)\xi(s^t)\frac{y^f(s^t)}{Y(s^t)}\right]}{\chi^{-1}\frac{\mathcal{W}_C(s^t)}{U^m_C(s^t)}+(1-\gamma)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}+(1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]}=\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{1/\rho}\frac{Y(s^t)}{A(s_t)L(s^t)} \tag{94}$$

With proportional shocks to the labor skill distribution, we have that:

$$\mathcal{W}_C(s^t)=U^m_C(s^t)\Omega_C \qquad \text{and} \qquad \mathcal{W}_L(s^t)=U^m_L(s^t)\Omega_L.$$

where

$$\Omega_C\equiv\sum_{i\in I}\pi^i\omega^i_C(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right] \qquad \text{and} \qquad \Omega_L\equiv\sum_{i\in I}\pi^i\omega^i_L(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]$$

are constants. Substituting this into (94) gives us the following optimality condition:

$$\frac{\Omega_L + (1+\eta)\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{Y(s^t)} + (1-\kappa)\xi(s^t)\frac{y^f(s^t)}{Y(s^t)}\right]}{\chi^{-1}\Omega_C + (1-\gamma)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]} = \left(\frac{y^f(s^t)}{Y(s^t)}\right)^{1/\rho}\frac{Y(s^t)}{A(s_t)L(s^t)} \tag{95}$$

Next we combine FOCs (83) and (85) in order to obtain:

$$\frac{\varsigma^Y(s^t) + \upsilon(s^{t-1})\left\{\chi U_C^m(s^t)\frac{\rho-1}{\rho} + U_L^m(s^t)\frac{1}{A(s_t)}\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}}{\varsigma^Y(s^t) + \xi(s^t)\left\{\chi U_C^m(s^t)\frac{\rho-1}{\rho} + U_L^m(s^t)\frac{1}{A(s_t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}} = \frac{\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}}{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}} \tag{96}$$

Furthermore, condition (91) can be written as:

$$\varsigma^Y(s^t) = U_C^m(s^t)\Omega_C(\varphi) + \chi U_C^m(s^t)\left\{\frac{1}{\rho} - \gamma\right\}\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]$$

Substituting this into (96) we and using the implementability condition in (25) we obtain:

$$\frac{\chi^{-1}\Omega_C(\varphi) + \left\{\frac{1}{\rho} - \gamma\right\}\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right] + \upsilon(s^{t-1})\left[\frac{\rho-1}{\rho} - \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1/\rho}\right]}{\chi^{-1}\Omega_C(\varphi) + \left\{\frac{1}{\rho} - \gamma\right\}\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right] - \frac{1}{\rho}\xi(s^t)} \tag{97}$$

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

We thus have a system of four equations in four unknowns. The four equations are (95), (97), along with the two resource constraints in (35). The four unknowns are:

$$\left\{\frac{Y(s^t)}{A(s_t)L(s^t)}, \frac{y^s(s^t)}{Y(s^t)}, \frac{y^f(s^t)}{Y(s^t)}, \xi(s^t)\right\}$$

We relabel these variables as follows:

$$\left\{\tilde{Y}(s^t), \tilde{y}^s(s^t), \tilde{y}^f(s^t), \xi(s^t)\right\} \equiv \left\{\frac{Y(s^t)}{A(s_t)L(s^t)}, \frac{y^s(s^t)}{Y(s^t)}, \frac{y^f(s^t)}{Y(s^t)}, \xi(s^t)\right\} \tag{98}$$

We rewrite the four equations with the relabeled variables below:

$$\frac{\Omega_L(\varphi) + (1+\eta)\tilde{Y}(s^t)\left[\kappa\upsilon(s^{t-1})\tilde{y}^s(s^t) + (1-\kappa)\xi(s^t)\tilde{y}^f(s^t)\right]}{(\chi^*)^{-1}\Omega_C(\varphi) + (1-\gamma)\left[\kappa\upsilon(s^{t-1})\tilde{y}^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\tilde{y}^f(s^t)^{\frac{\rho-1}{\rho}}\right]} = \tilde{y}^f(s^t)^{1/\rho}\tilde{Y}(s^t)$$

$$\frac{\mathcal{H} + \left(\frac{1}{\rho} - \gamma\right)\left[\kappa\upsilon(s^{t-1})\tilde{y}^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\tilde{y}^f(s^t)^{\frac{\rho-1}{\rho}}\right] + \upsilon(s^{t-1})\left[\frac{\rho-1}{\rho} - \left[\frac{\tilde{y}^s(s^t)}{\tilde{y}^f(s^t)}\right]^{\frac{1}{\rho}}\right]}{\mathcal{H} + \left(\frac{1}{\rho} - \gamma\right)\left[\kappa\upsilon(s^{t-1})\tilde{y}^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\tilde{y}^f(s^t)^{\frac{\rho-1}{\rho}}\right] - \frac{1}{\rho}\xi(s^t)} = \left[\frac{\tilde{y}^s(s^t)}{\tilde{y}^f(s^t)}\right]^{\frac{1}{\rho}}$$

$$1 = \kappa \tilde{y}^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\tilde{y}^f(s^t)^{\frac{\rho-1}{\rho}},$$
$$1 = \kappa \tilde{y}^s(s^t)\tilde{Y}(s^t) + (1-\kappa)\tilde{y}^f(s^t)\tilde{Y}(s^t).$$

where $\mathcal{H} \equiv (\chi^*)^{-1}\Omega_C(\varphi)$. Note that these equations are identical across all states s, s' conditional on s^{t-1}. Therefore, the quadruplet in (98) satisfies:

$$\left\{ \tilde{Y}(s), \tilde{y}^s(s), \tilde{y}^f(s), \xi(s) \middle| s^{t-1} \right\} = \left\{ \tilde{Y}(s'), \tilde{y}^s(s'), \tilde{y}^f(s'), \xi(s') \middle| s^{t-1} \right\}, \qquad \forall s, s' \in S|s^{t-1} \tag{99}$$

In other words, conditional on history s^{t-1}, there is no variation in these endogenous variables across realizations of states.

Finally we use the implementability condition (26). By combining it with (25) it can be written as:

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1}) U_C^m(s^t) y^s(s^t) \left\{ \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} - \left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho} \right\} = 0$$

or,

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1}) U_C^m(s^t) y^s(s^t) \left\{ \tilde{y}^s(s^t)^{-1/\rho} - \tilde{y}^f(s^t)^{-1/\rho} \right\} = 0$$

This is consistent with the property stated in (99) if and only if:

$$\tilde{y}^s(s^t) = \tilde{y}^f(s^t) = 1, \qquad \forall s^t|s^{t-1}.$$

It is therefore optimal for monetary policy to implement the flexible-price allocation given any arbitrary χ.

A.10 Proof of Theorem 3

At the Ramsey optimum, we have:

$$-\frac{\mathcal{W}_L(s^t) + (1+\eta)U_L^m(s^t)\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{Y(s^t)} + (1-\kappa)\xi(s^t)\frac{y^f(s^t)}{Y(s^t)}\right]}{\mathcal{W}_C(s^t) + \chi(1-\gamma)U_C^m(s^t)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]} = \frac{Y(s^t)}{L(s^t)}. \tag{100}$$

With separable and iso-elastic utility, $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ satisfy:

$$\mathcal{W}_C(s^t) = U_C^m(s^t)\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1-\gamma)\right] \tag{101}$$

$$\mathcal{W}_L(s^t) = U_L^m(s^t)\sum_{i\in I}\pi^i\omega_L^i(\varphi, s_t)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right]. \tag{102}$$

Substituting these expressions for $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ into (100), and defining a function $\mathcal{I}(s_t)$ and a scalar $\mathcal{H}$ as follows:

$$\mathcal{I}(s_t) \equiv \sum_{i\in I}\pi^i\omega_L^i(\varphi, s_t)\left[\frac{\lambda^i}{\varphi^i} + \nu^i(1+\eta)\right] \qquad \text{and} \qquad \mathcal{H} \equiv (\chi^*)^{-1}\Omega_C, \tag{103}$$

we infer that the optimal monetary wedge satisfies:

$$1-\tau_M^*(s^t) = \frac{\mathcal{H} + (1-\gamma)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}} + (1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\right]}{\mathcal{I}(s_t) + (1+\eta)\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{Y(s^t)} + (1-\kappa)\xi(s^t)\frac{y^f(s^t)}{Y(s^t)}\right]}. \tag{104}$$

Threshold. We first consider the conditions under which $\tau_M^*(s^t) = 0$. In this state: $y^s(s^t) = y^f(s^t) = Y(s^t) = A(s_t)L(s^t)$. Condition (104) reduces to:

$$1 = \frac{\mathcal{H} + (1-\gamma)\left[\kappa\upsilon(s^{t-1}) + (1-\kappa)\xi(s^t)\right]}{\mathcal{I}(s_t) + (1+\eta)\left[\kappa\upsilon(s^{t-1}) + (1-\kappa)\xi(s^t)\right]}$$

Furthermore, by optimality conditions (83) and (85), $y^s(s^t) = y^f(s^t) = Y(s^t)$ if and only if $\xi(s^t) = \upsilon(s^{t-1})$; we state and prove this formally in Lemma (3). Therefore:

$$1 = \frac{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1})}{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1})}$$

Solving this for $\mathcal{I}(s_t)$ we obtain the following threshold:

$$\bar{\mathcal{I}}(s^{t-1}) = \mathcal{H} - (\eta+\gamma)\upsilon(s^{t-1})$$

Therefore if $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$, the optimal monetary tax is equal to zero: $\tau_M^*(s^t) = 0$.

A fictitious tax wedge. We next define a fictitious tax wedge:

$$1-\hat{\tau}(s^t) \equiv \frac{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1})}{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1})} \tag{105}$$

The wedge $1-\hat{\tau}(s^t)$ is continuous and strictly decreasing in $\mathcal{I}(s_t)$, as all other terms are constants (conditional on s^{t-1}). Furthermore, note that $\hat{\tau}(s^t) = 0$ if and only if $\mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1})$. As a result, the fictitious tax $\hat{\tau}(s^t)$ trivially satisfies:

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

The optimal monetary wedge. The next step of our proof involves characterizing $y^f(s^t)$, $y^s(s^t)$, and the multipliers $\xi(s^t)$ and $\upsilon(s^{t-1})$ at the optimal allocation.[33]

[33] Note that if $\upsilon(s^{t-1}) < 0$, then $\xi(s^t)/\upsilon(s^{t-1}) < 1$ implies:

$$\upsilon(s^{t-1}) < \xi(s^t) < 0.$$

Lemma 3. *(i) At the optimal allocation:*

$$\begin{array}{lll} y^f(s^t) > y^s(s^t) & \textit{if and only if} & \tau_M^*(s^t) > 0; \\ y^f(s^t) = y^s(s^t) & \textit{if and only if} & \tau_M^*(s^t) = 0; \\ y^f(s^t) < y^s(s^t) & \textit{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

(ii) At the optimal allocation:

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

Proof. See Section A.11 of this appendix. □

We will use this lemma in what follows. We consider the optimal monetary wedge (104), which can be written as:

$$1 - \tau_M^*(s^t) = \frac{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1}) \left[\kappa \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}} + (1-\kappa) \frac{\xi(s^t)}{\upsilon(s^{t-1})} \left[\frac{y^f(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}} \right]}{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1}) \left[\kappa \frac{y^s(s^t)}{A(s_t)L(s^t)} + (1-\kappa) \frac{\xi(s^t)}{\upsilon(s^{t-1})} \frac{y^f(s^t)}{A(s_t)L(s^t)} \right]}$$

From our resource constraints, note that aggregate labor and aggregate output satisfy:

$$1 = \kappa \frac{y^s(s^t)}{A(s_t)L(s^t)} + (1-\kappa) \frac{y^f(s^t)}{A(s_t)L(s^t)}, \quad \text{and} \tag{106}$$

$$1 = \kappa \left[\frac{y^s(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}} + (1-\kappa) \left[\frac{y^f(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}}. \tag{107}$$

Substituting these into our previous expression for the optimal monetary wedge and rearranging, we obtain the following expression:

$$1 - \tau_M^*(s^t) = \frac{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1}) + (1-\kappa)(1-\gamma) \left[\frac{y^f(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}} (\xi(s^t) - \upsilon(s^{t-1}))}{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1}) + (1-\kappa)(1+\eta) \frac{y^f(s^t)}{Y(s^t)} \frac{Y(s^t)}{A(s_t)L(s^t)} (\xi(s^t) - \upsilon(s^{t-1}))} \tag{108}$$

We want to compare this to the fictitious tax wedge defined in (105). In order to do so, we take the inverse wedges from (108) and (105):

$$\frac{1}{1-\tau_M^*(s^t)} = \frac{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1}) + (1-\kappa)(1+\eta) \frac{y^f(s^t)}{Y(s^t)} \frac{Y(s^t)}{A(s_t)L(s^t)} (\xi(s^t) - \upsilon(s^{t-1}))}{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1}) + (1-\kappa)(1-\gamma) \left[\frac{y^f(s^t)}{Y(s^t)} \right]^{\frac{\rho-1}{\rho}} (\xi(s^t) - \upsilon(s^{t-1}))},$$

$$\frac{1}{1-\hat{\tau}(s^t)} = \frac{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1})}{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1})}.$$

Combining these two, we get:

$$\frac{1}{1-\tau_M^*(s^t)}\left\{1+\frac{(1-\kappa)(1-\gamma)}{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}(\xi(s^t)-\upsilon(s^{t-1}))\right\}$$
$$=\frac{1}{1-\hat{\tau}(s^t)}+\frac{(1-\kappa)(1+\eta)}{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}\frac{y^f(s^t)}{Y(s^t)}\frac{Y(s^t)}{A(s_t)L(s^t)}(\xi(s^t)-\upsilon(s^{t-1}))$$

which furthermore implies:

$$\begin{aligned}\frac{1}{1-\tau_M^*(s^t)}=&\frac{1}{1-\hat{\tau}(s^t)}+\frac{(1-\kappa)(1+\eta)}{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}\left[\frac{y^f(s^t)}{Y(s^t)}\frac{Y(s^t)}{A(s_t)L(s^t)}\right](\xi(s^t)-\upsilon(s^{t-1}))\\&-\frac{(1-\kappa)(1-\gamma)}{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{\frac{\rho-1}{\rho}}\frac{1}{1-\tau_M^*(s^t)}(\xi(s^t)-\upsilon(s^{t-1}))\end{aligned}\tag{109}$$

Next, we combine the monetary wedge defined in (37) with the implementability condition (25). Doing so yields the following equation:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}=(1-\tau_M^*(s^t))\frac{Y(s^t)}{A(s_t)L(s^t)}\tag{110}$$

Substituting this into our expression in (109), we obtain the following condition:

$$\frac{1}{1-\tau_M^*(s^t)}=\frac{1}{1-\hat{\tau}(s^t)}+\frac{(1-\kappa)(\eta+\gamma)}{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}\frac{y^f(s^t)}{Y(s^t)}\frac{Y(s^t)}{A(s_t)L(s^t)}(\xi(s^t)-\upsilon(s^{t-1}))\tag{111}$$

It follows that

$$\begin{array}{lll}\tau_M^*(s^t)<\hat{\tau}(s^t) & \text{if and only if} & \xi(s^t)<\upsilon(s^{t-1}),\\ \tau_M^*(s^t)=\hat{\tau}(s^t) & \text{if and only if} & \xi(s^t)=\upsilon(s^{t-1}),\\ \tau_M^*(s^t)>\hat{\tau}(s^t) & \text{if and only if} & \xi(s^t)>\upsilon(s^{t-1}).\end{array}$$

Consider again the case in which:

$$y^s(s^t)=y^f(s^t)=Y(s^t)=A(s_t)L(s^t).$$

From Lemma (3) we have that in this state, $\tau_M^*(s^t)=0$ and $\xi(s^t)=\upsilon(s^{t-1})$. Therefore:

$$1-\tau_M^*(s^t)=\frac{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})}{\mathcal{I}(s_t)+(1+\eta)\upsilon(s^{t-1})}=1-\hat{\tau}(s^t)=1.$$

Therefore $\tau_M^*(s^t)=0$ if and only if $\hat{\tau}(s^t)=0$. This implies $\tau_M^*(s^t)=0$ if and only if $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$.

Consider second the case in which $y^f(s^t)>y^s(s^t)$. From Lemma (3) we have that in this state, $\tau_M^*(s^t)>0$ and $\xi(s^t)<\upsilon(s^{t-1})$. From the expression above, the latter implies $\tau_M^*(s^t)<\hat{\tau}(s^t)$. Therefore:

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

Finally, $\hat{\tau}(s^t)>0$ implies $\mathcal{I}(s_t)>\bar{\mathcal{I}}(s^{t-1})$.

Consider third the case in which $y^f(s^t) < y^s(s^t)$. From Lemma (3) we have that in this state, $\tau_M^*(s^t) < 0$ and $\xi(s^t) > \upsilon(s^{t-1})$. From the expression above, the latter implies $\tau_M^*(s^t) > \hat{\tau}(s^t)$. Therefore:

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

Finally, $\hat{\tau}(s^t) < 0$ implies $\mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1})$.

We now prove the converse statements by contradiction. Let $\mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1})$ and suppose that $\tau_M^*(s^t) < 0$. From Lemma (3), this implies $\xi(s^t) > \upsilon(s^{t-1})$, which further implies $\tau_M^*(s^t) > \hat{\tau}(s^t)$. It follows that $\hat{\tau}(s^t) < \tau_M^*(s^t) < 0$. But $\hat{\tau}(s^t) < 0$ is a contradiction of $\mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1})$. Therefore $\tau_M^*(s^t) > 0$.

Similarly let $\mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1})$ and suppose that $\tau_M^*(s^t) > 0$. From Lemma (3), this implies $\xi(s^t) < \upsilon(s^{t-1})$, which further implies $\tau_M^*(s^t) < \hat{\tau}(s^t)$. It follows that $0 < \tau_M^*(s^t) < \hat{\tau}(s^t)$. But $\hat{\tau}(s^t) > 0$ is a contradiction of $\mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1})$. Therefore $\tau_M^*(s^t) < 0$.

We thus conclude that 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 Lemma 3

Part (i). Substituting in our expression for $L(s^t)$ from (106) into (110) we obtain the following expression:

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

which we may rewrite as:

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

Next we combine this with (107) and obtain:

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

It follows that:

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

This implies:

$$\text{sign}(\tau_M^*(s^t)) = \text{sign}\left\{1 - \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right\}.$$

Therefore:

$$\begin{array}{lll} y^f(s^t) > y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) > 0, \\ y^f(s^t) = y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) = 0, \\ y^f(s^t) < y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

Part (ii). Combining the planner's optimality conditions (83) and (85), we obtain the following condition which must hold at the planner's optimum:

$$\frac{\varsigma^Y(s^t) + \chi U_C^m(s^t)\upsilon(s^{t-1})\left\{\frac{\rho-1}{\rho} + \frac{U_L^m(s^t)}{\chi U_C^m(s^t)}\frac{1}{A(s_t)}\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}}{\varsigma^Y(s^t) + \chi U_C^m(s^t)\xi(s^t)\left\{\frac{\rho-1}{\rho} + \frac{U_L^m(s^t)}{\chi U_C^m(s^t)}\frac{1}{A(s_t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}} = \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}$$

Next, using implementability condition (25), we rewrite the above equation as follows:

$$\frac{\varsigma^Y(s^t) + \chi U_C^m(s^t)\upsilon(s^{t-1})\left\{\frac{\rho-1}{\rho} - \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}}{\varsigma^Y(s^t) + \chi U_C^m(s^t)\xi(s^t)\left\{\frac{\rho-1}{\rho} - 1\right\}} = \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}$$

which reduces to:

$$\frac{\frac{\varsigma^Y(s^t)}{\chi U_C^m(s^t)} + \upsilon(s^{t-1})\left\{\frac{\rho-1}{\rho} - \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right\}}{\frac{\varsigma^Y(s^t)}{\chi U_C^m(s^t)} + \xi(s^t)\left\{\frac{\rho-1}{\rho} - 1\right\}} = \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}$$

Rearranging, we obtain the following equilibrium condition:

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho} - \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right) - \xi(s^t)\left(\frac{\rho-1}{\rho} - 1\right)\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho} = \frac{\varsigma^Y(s^t)}{\chi U_C^m(s^t)}\left(\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho} - 1\right) \tag{112}$$

Consider the term $\varsigma^Y(s^t)/\chi U_C^m(s^t)$ on the right-hand side of condition (112). The scalar χ is strictly positive by definition. Next, because individual utility is strictly increasing in consumption, it is clear that $U_C^m(s^t)$ is strictly positive for all $s^t \in S^t$. Finally, $\varsigma^Y(s^t)$ is strictly positive for all $s^t \in S^t$; we verify this statement later on in the proof.

We first consider the case in which $y^s(s^t) = y^f(s^t)$. In this case, condition (112) reduces to:

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho} - 1\right) - \xi(s^t)\left(\frac{\rho-1}{\rho} - 1\right) = 0,$$

which implies: $\xi(s^t) = \upsilon(s^{t-1})$.

We now prove the converse. Consider the case in which $\xi(s^t) = \upsilon(s^{t-1})$. In this case, condition (112) reduces to:

$$\upsilon(s^{t-1})\frac{\rho-1}{\rho}\left(1 - \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right) = \frac{\varsigma^Y(s^t)}{\chi U_C^m(s^t)}\left(\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho} - 1\right),$$

which implies: $y^s(s^t) = y^f(s^t)$.

Consider next the case in which $y^s(s^t) > y^f(s^t)$. Then condition (112) implies

$$\xi(s^t)\left(\frac{\rho-1}{\rho} - 1\right)\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho} < \upsilon(s^{t-1})\left(\frac{\rho-1}{\rho} - \left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right),$$

which furthermore implies

$$\xi(s^t)\left(\frac{\rho-1}{\rho}-1\right) < \upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{-1/\rho}-1\right). \tag{113}$$

Furthermore note that

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{-1/\rho}-1\right) < \upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}-1\right). \tag{114}$$

Combining (113) and (114), we infer:

$$\xi(s^t)\left(\frac{\rho-1}{\rho}-1\right) < \upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}-1\right)$$

which finally implies: $\xi(s^t) > \upsilon(s^{t-1})$. We can prove the converse statement by contradiction.

Consider next the case in which $y^s(s^t) < y^f(s^t)$. Then condition (112) implies

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}-\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho}\right) < \xi(s^t)\left(\frac{\rho-1}{\rho}-1\right)\left[\frac{y^s(s^t)}{y^f(s^t)}\right]^{1/\rho},$$

which furthermore implies

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}\left[\frac{y^f(s^t)}{y^s(s^t)}\right]^{1/\rho}-1\right) < \xi(s^t)\left(\frac{\rho-1}{\rho}-1\right). \tag{115}$$

Furthermore note that

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}-1\right) < \upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}\left[\frac{y^f(s^t)}{y^s(s^t)}\right]^{1/\rho}-1\right). \tag{116}$$

Combining (115) and (116), we infer:

$$\upsilon(s^{t-1})\left(\frac{\rho-1}{\rho}-1\right) < \xi(s^t)\left(\frac{\rho-1}{\rho}-1\right)$$

which finally implies: $\xi(s^t) < \upsilon(s^{t-1})$. We can prove the converse statement by contradiction.

We have thus shown that:

$$\begin{array}{lll} \xi(s^t) < \upsilon(s^{t-1}) & \text{if and only if} & y^f(s^t) > y^s(s^t), \\ \xi(s^t) = \upsilon(s^{t-1}) & \text{if and only if} & y^f(s^t) = y^s(s^t), \\ \xi(s^t) > \upsilon(s^{t-1}) & \text{if and only if} & y^f(s^t) < y^s(s^t). \end{array}$$

Combining this with the result stated in part (i) of Lemma 3, we obtain the result stated in part (ii).

What remains to be shown is that the multiplier $\varsigma^Y(s^t)$ is strictly positive for all $s^t \in S^t$. We combine the planner's optimality conditions (88) and (89) and obtain the following expression for $\varsigma^Y(s^t)$:

$$\varsigma^Y(s^t) = \varsigma^C(s^t) + \frac{1}{\rho}\frac{U^m_C(s^t)}{U^m_{CC}(s^t)C(s^t)}(\varsigma^C(s^t) - \mathcal{W}_C(s^t))$$

Next, iso-elastic utility implies:

$$\varsigma^Y(s^t) = \left[1 - \frac{1}{\gamma\rho}\right]\varsigma^C(s^t) + \frac{1}{\gamma\rho}\mathcal{W}_C(s^t)$$

with $\rho > 1$ and $\gamma > 1$. It is clear from the definition of $\mathcal{W}(\cdot)$ that its first derivative with respect to aggregate consumption, denoted by $\mathcal{W}_C(s^t)$, is strictly positive. Furthermore, recall that in our proof of Proposition 6 [in Appendix A.8], we show that the Karush–Kuhn–Tucker multiplier $\varsigma^C(s^t)$ is strictly positive for all $s^t \in S^t$. Therefore $\varsigma^Y(s^t)$ is strictly positive for all $s^t \in S^t$.

A.12 Proof of Proposition 4

We combine (37) with the intratemporal condition in (12) and obtain the following condition:

$$\frac{W(s^t)}{P(s^t)} = (1-\tau_r)\left(\frac{\rho-1}{\rho}\right)(1-\tau_M^*(s^t))\frac{Y(s^t)}{L(s^t)}$$

To simplify, we consider the fiscal implementation that sets $(1-\tau_r)\left(\frac{\rho-1}{\rho}\right) = 1$. Using this fiscal implementation, we combine the above expression with the implementability condition in (110) and infer that the aggregate price level satisfies:

$$P(s^t) = \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\frac{W(s^t)}{A(s_t)}.$$

Therefore the optimal markup satisfies:

$$\log\mathcal{M}(s^t) = \frac{1}{\rho}(\log y^f(s^t) - \log Y(s^t))$$

with $\rho > 1$. It follows that:

$$\begin{aligned} \log\mathcal{M}(s^t) > 0 \quad &\text{if and only if} \quad y^f(s^t) > y^s(s^t), \\ \log\mathcal{M}(s^t) = 0 \quad &\text{if and only if} \quad y^f(s^t) = y^s(s^t), \\ \log\mathcal{M}(s^t) < 0 \quad &\text{if and only if} \quad y^f(s^t) < y^s(s^t). \end{aligned}$$

Combining this result with Lemma 3, we obtain:

$$\begin{aligned} \log\mathcal{M}(s^t) > 0 \quad &\text{if and only if} \quad \tau_M^*(s^t) > 0, \\ \log\mathcal{M}(s^t) = 0 \quad &\text{if and only if} \quad \tau_M^*(s^t) = 0, \\ \log\mathcal{M}(s^t) < 0 \quad &\text{if and only if} \quad \tau_M^*(s^t) < 0. \end{aligned}$$

The result stated in Proposition 4 follows by combining the above with Theorem 3.

B Equivalent Equilibrium Representation

In this section of the appendix we provide an equivalent characterization of the equilibrium in our economy using the forecast errors $\epsilon(s^t)$ defined in (28). Such a representation gives rise to a few auxiliary results, Lemmas (4), (5), and (6), that we find useful in later proofs.

First, note that equation (27) implies $p_t^s(s^{t-1}) = \epsilon(s^t)p_t^f(s^t)$. It follows from the CES demand equation (5) that relative quantities across the two types of firms satisfy:

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

Therefore:

$$y^s(s^t) = \epsilon(s^t)^{-\rho}y^f(s^t) \tag{117}$$

Note that the flexible-price allocation coincides with $\epsilon(s^t) = 1$ for all $s^t \in S^t$.

Lemma 4. *For any $s^t \in S^t$, equilibrium aggregate output satisfies:*

$$Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t) \tag{118}$$

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

$$\Delta(\epsilon) \equiv \left\{\frac{\left[\kappa\epsilon^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}}}{\left[\kappa\epsilon^{-\rho} + (1-\kappa)\right]^{1/\rho}}\right\}^{\rho} > 0. \tag{119}$$

The function Δ is continuous, differentiable, strictly concave, and satisfies $\max_{\epsilon>0}\Delta(\epsilon) = 1$. Furthermore, it attains its unique maximum at $\epsilon = 1$.

Proof. See Appendix B.1. □

Lemma 4 provides a succinct characterization of the efficiency wedge in this economy. When monetary policy implements flexible-price allocations—that is, when it sets $\epsilon(s^t) = 1$ 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 $\epsilon(s^t)$ deviates from one 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 leads to misallocation of inputs. This manifests as an efficiency wedge, or TFP loss. The term $\Delta(\epsilon)$ represents this efficiency wedge.

The following lemma provides a similar result for the equilibrium labor wedge in this economy.

Lemma 5. *For any $s^t \in S^t$, aggregate output and labor joint satisfy:*

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)} = \chi\Gamma(\epsilon(s^t))A(s_t) \tag{120}$$

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

$$\Gamma(\epsilon) \equiv \left[\kappa\epsilon^{1-\rho} + (1-\kappa)\right]^{\frac{1}{\rho-1}} > 0. \tag{121}$$

The function Γ is continuous, differentiable, and satisfies the following two properties (i) $\Gamma(1) = 1$ and (ii) $\Gamma'(\epsilon) < 0$ for all $\epsilon > 0$. It follows that:

$$\begin{array}{lll} \Gamma(\epsilon) < 1 & \textit{if and only if} & \epsilon > 1, \\ \Gamma(\epsilon) = 1 & \textit{if and only if} & \epsilon = 1, \\ \Gamma(\epsilon) > 1 & \textit{if and only if} & \epsilon \in (0,1). \end{array}$$

Proof. See Appendix B.2. □

Finally, we can relate implicit monetary tax $\tau_M(s^t)$ defined in 37 to the forecast error. In this representation we can think of the monetary tax as a function of ϵ, that is, $\tau_M : \mathbb{R}_+ \to (-\infty, 1)$.

Lemma 6. *The monetary tax satisfies:*

$$\tau_M(\epsilon) = 1 - \frac{\Gamma(\epsilon)}{\Delta(\epsilon)} = \frac{\kappa\epsilon^{-\rho}(\epsilon - 1)}{\kappa\epsilon^{-(\rho-1)} + (1-\kappa)}.$$

The function τ_M is continuous, differentiable, and satisfies: sign$(\tau_M(\epsilon))$ = sign$(\epsilon - 1)$ for all $\epsilon > 0$. It follows that:

$$\begin{array}{lll} \tau_M(\epsilon) > 0 & \textit{if and only if} & \epsilon > 1, \\ \tau_M(\epsilon) = 0 & \textit{if and only if} & \epsilon = 1, \\ \tau_M(\epsilon) < 0 & \textit{if and only if} & \epsilon \in (0,1). \end{array}$$

Proof. See Appendix B.3. □

B.1 Proof of Lemma 4

We combine (117) with (106) and (107) and obtain the following expressions for aggregate output and labor:

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

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

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

It follows that the aggregate production function can be expressed as (118) with

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

Next, note that $\Delta(\epsilon)$ is continuous and differentiable. The first derivative of $\Delta(\epsilon)$ with respect to ϵ 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:

$$\frac{d\Delta(\epsilon)}{d\epsilon}=\kappa\rho\Delta(\epsilon)\epsilon^{-\rho-1}\left\{\left[\kappa\epsilon^{-\rho}+(1-\kappa)\right]^{-1}-\left[\kappa\epsilon^{-\rho+1}+(1-\kappa)\right]^{-1}\epsilon\right\} \tag{122}$$

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 (122), $d\Delta(\epsilon)/d\epsilon > 0$ 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.

B.2 Proof of Lemma 5

The aggregate price level satisfies:

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

Substituting in the firms' optimal prices, we obtain:

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

Combining the above equation with the household's intratemporal condition (12), we get:

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

It follows that in equilibrium, aggregate output and aggregate labor jointly satisfy (120) with

$$\Gamma(\epsilon) \equiv \left[\kappa\epsilon(s^t)^{1-\rho} + (1-\kappa)\right]^{-\frac{1}{1-\rho}} > 0.$$

Next, note that $\Gamma(\epsilon)$ is continuous and differentiable. Furthermore, $\Gamma(1) = 1$. Finally, the first derivative of $\Gamma(\epsilon)$ is given by:

$$\frac{d\Gamma(\epsilon)}{d\epsilon} \equiv -\kappa\left[\kappa\epsilon^{-(\rho-1)} + (1-\kappa)\right]^{\frac{1}{\rho-1}-1}\epsilon^{-\rho}$$

Therefore $d\Gamma(\epsilon)/d\epsilon < 0$ for all $\epsilon > 0$.

B.3 Proof of Lemma 6

Combining the definition of the implicit monetary tax in 37 with (120) yields:

$$(1-\tau_M(s^t))\frac{Y(s^t)}{L(s^t)} = \Gamma(\epsilon(s^t))A(s_t)$$

Combining the above expression with (118) we infer that in equilibrium the monetary tax satisfies:

$$1-\tau_M(s^t) = \frac{\Gamma(\epsilon(s^t))}{\Delta(\epsilon(s^t))}$$

Substituting the functions Γ and Δ from () and (), we find that the implicit monetary tax satisfies:

$$1-\tau_M(\epsilon) = \frac{\kappa\epsilon^{-\rho} + (1-\kappa)}{\kappa\epsilon^{-(\rho-1)} + (1-\kappa)}. \tag{123}$$

Solving this for $\tau_M(\epsilon)$, we get:

$$\tau_M(\epsilon) = \frac{\kappa\epsilon^{-\rho}(\epsilon-1)}{\kappa\epsilon^{-(\rho-1)} + (1-\kappa)}$$

Note that $\tau_M(\epsilon)$ is continuous and differentiable on the domain $\epsilon > 0$. With this expression we can prove the last part of Lemma 6. Note that the denominator is strictly positive for all $\epsilon > 0$. Furthermore $\kappa\epsilon^{-\rho} > 0$ for all $\epsilon > 0$. Therefore sign$(\tau_M(\epsilon)) =$ sign$(\epsilon - 1)$ for all $\epsilon > 0$.

C Implementation

In this section of the appendix, we expand on our discussion of implementation. In particular, we consider the behavior of aggregate price levels and nominal interest rates consistent with the Ramsey optimum. As in the sufficiency portion of our proof of Proposition 2 (in Appendix A.6), we let $\mathcal{B}_t(s^{t-1}) > 0$ denote the common belief of the aggregate price level at time t based on history s^{t-1}. Aside from being strictly positive, $\mathcal{B}_t(s^{t-1}) > 0$ is a free parameter in our model. For a given $\mathcal{B}_t(s^{t-1})$, when $P(s^t) = \mathcal{B}_t(s^{t-1})$, the economy replicates the flexible price outcome. Let $\hat{\imath}(s^t)$ denote the nominal interest rate consistent with the flexible-price outcome; one can think of $\hat{\imath}(s^t)$ as the "natural" rate of interest.

Proposition 7. *Given a common belief* $\mathcal{B}_t(s^{t-1}) > 0$*, the aggregate price level,* $P(s^t)$*, at the Ramsey optimum and the nominal interest rate,* $i(s^t)$*, consistent with that price level satisfies:*

$$\begin{array}{llll} P(s^t) < \mathcal{B}_t(s^{t-1}) & \textit{and} \quad i(s^t) > \hat{\imath}(s^t) & \textit{if and only if} & \mathcal{I}(s_t) > \bar{\mathcal{I}}(s^{t-1}), \\ P(s^t) = \mathcal{B}_t(s^{t-1}) & \textit{and} \quad i(s^t) = \hat{\imath}(s^t) & \textit{if and only if} & \mathcal{I}(s_t) = \bar{\mathcal{I}}(s^{t-1}), \\ P(s^t) > \mathcal{B}_t(s^{t-1}) & \textit{and} \quad i(s^t) < \hat{\imath}(s^t) & \textit{if and only if} & \mathcal{I}(s_t) < \bar{\mathcal{I}}(s^{t-1}). \end{array}$$

Proof. See Appendix C.1. □

The behavior of the aggregate price level and nominal interest rate described in Proposition 7 is consistent with the optimal markup described in Proposition 4. To understand this, recall that prices are "sticky" in our model, while nominal wages are fully flexible. Specifically, given a common belief $\mathcal{B}_t(s^{t-1})$ for the aggregate price level, the prices of the sticky-price firms are "stuck" at $p^s_t(s^{t-1}) = \mathcal{B}_t(s^{t-1})$, while the prices of the flexible-price firms and the nominal wage respond to the realized state.

Therefore, in order to generate an increase in the aggregate markup, the nominal wage must (unexpectedly) fall. In this case, the prices of the flexible-price firms fall, in line with the realized nominal wage, while those of the sticky price firms remain at $p^s_t(s^{t-1}) = \mathcal{B}_t(s^{t-1})$. As a result, the aggregate price level falls but not to the same extent as the nominal wage, and the aggregate markup increases. Conversely, in order to generate a fall in the aggregate markup, the nominal wage must (unexpectedly) rise. In this case, the aggregate price level also rises, but less so than wages, so that the aggregate markup indeed falls.

Furthermore, there exists a nominal interest rate that satisfies the bond Euler equation in (13) and is consistent with the movements of the aggregate price level and the optimal markup. An increase in the aggregate markup in our model is consistent with a tightening of the nominal interest rate relative to the natural rate; conversely a fall in the markup is consistent with a loosening of the nominal interest rate.

Robustness. Finally, it is important to note that the price and interest rate movement consistent with the Ramsey optimum depend on the relative stickiness of prices versus wages. More specifically, our characterization of the aggregate price level (and nominal interest rate) in Proposition 7 relies on our assumption of sticky prices but flexible wages. If instead wages were sticky and prices were flexible, an increase in the aggregate markup would require an unexpected *increase* in the aggregate price level rather than a fall.

For this reason, in terms of monetary implementation we wish to put less emphasis on Proposition 7 and more emphasis on Proposition 4—what is robust to the relative stickiness of prices versus wages is the movement in the optimal markup. Furthermore, although monetary policy typically centers on changes in nominal interest rates, the emphasis of the primal approach on characterizing allocations proves highly useful in our analysis. In particular, the economics behind (and robustness of) optimal monetary policy in our environment can best be understood by studying the state-contingent wedge at the Ramsey optimum (Section 5).

C.1 Proof of Proposition 7

Recall from the sufficiency argument of Proposition 2 (Appendix A.6) that for any sticky-price allocation, we can construct nominal prices as follows. Given $\mathcal{B}_t(s^{t-1}) > 0$, we set:

$$p_t^s(s^{t-1}) = \mathcal{B}_t(s^{t-1}).$$

This implies, in terms of our output decomposition in 72, that $\phi^s(s^{t-1})^{-1/\rho} = \mathcal{B}_t(s^{t-1})$. Therefore:

$$\phi^s(s^{t-1}) = \mathcal{B}_t(s^{t-1})^{-\rho} \qquad \text{and} \qquad y^s(s^t) = \mathcal{B}_t(s^{t-1})^{-\rho}\Phi(s^t).$$

which implies that $\Phi(s^t) = y^s(s^t)/\mathcal{B}_t(s^{t-1})^{-\rho}$. The aggregate price level thereby satisfies:

$$P(s^t) = \left[\frac{Y(s^t)}{\Phi(s^t)}\right]^{-1/\rho} = \left[\frac{Y(s^t)}{y^s(s^t)}\mathcal{B}_t(s^{t-1})^{-\rho}\right]^{-1/\rho} = \left[\frac{Y(s^t)}{y^s(s^t)}\right]^{-1/\rho}\mathcal{B}_t(s^{t-1}).$$

Therefore the deviation of the price level from the expected price satisfies:

$$\log P(s^t) - \log \mathcal{B}_t(s^{t-1}) = -\frac{1}{\rho}(\log Y(s^t) - \log y^s(s^t)) \tag{124}$$

with $\rho > 1$. Therefore

$$\begin{array}{lll} P(s^t) < \mathcal{B}_t(s^{t-1}) & \text{if and only if} & y^f(s^t) > y^s(s^t), \\ P(s^t) = \mathcal{B}_t(s^{t-1}) & \text{if and only if} & y^f(s^t) = y^s(s^t), \\ P(s^t) > \mathcal{B}_t(s^{t-1}) & \text{if and only if} & y^f(s^t) < y^s(s^t). \end{array}$$

Combining this with our characterization of the optimal monetary tax in Lemma 3, we find:

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

Combining this with our result in Theorem 3, the result about the aggregate price level stated in Proposition 4 then follows.

Next we turn to the nominal interest rate. The nominal interest rate satisfies the Euler equation in 13:

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

Let $\hat{C}(s^t) = \hat{Y}(s^t)$ denote the flexible-price level of output. The natural, flexible-price, interest rate satisfies:

$$\frac{\hat{C}(s^t)^{-\gamma}}{\mathcal{B}_t(s^{t-1})} = \beta(1 + \hat{\imath}(s^t)) \sum_{s^{t+1}|s^t} \mu(s^{t+1}|s^t)\frac{U_C^m(s^{t+1})}{P(s^{t+1})}.$$

Therefore

$$\frac{1+i(s^t)}{1+\hat{\imath}(s^t)} = \frac{C(s^t)^{-\gamma}/P(s^t)}{\hat{C}(s^t)^{-\gamma}/\mathcal{B}_t(s^{t-1})}$$

In logs:

$$\log\left[\frac{1+i(s^t)}{1+\hat{\imath}(s^t)}\right] = -\gamma[\log Y(s^t) - \log \hat{Y}(s^t)] - [\log P(s^t) - \log \mathcal{B}_t(s^{t-1})]$$

Next we substitute the price level from 124 into the above expression. Doing so, we obtain the following expression:

$$\log(1+i(s^t)) - \log(1+\hat{\imath}(s^t)) = \frac{1}{\rho}(\log Y(s^t) - \log y^s(s^t)) - \gamma(\log Y(s^t) - \log \hat{Y}(s^t)).$$

First, from our characterization of the optimal monetary tax in Lemma 3, we have that:

$$\begin{array}{lll} \log Y(s^t) > \log y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) > 0, \\ \log Y(s^t) = \log y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) = 0, \\ \log Y(s^t) < \log y^s(s^t) & \text{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

The next step in our proof requires the following lemma:

Lemma 7. *Let $\hat{Y}(s^t)$ denote the level of output in history s^t under flexible prices. Then*

$$\begin{array}{lll} \log Y(s^t) < \log \hat{Y}(s^t) & \textit{if and only if} & \tau_M^*(s^t) > 0, \\ \log Y(s^t) = \log \hat{Y}(s^t) & \textit{if and only if} & \tau_M^*(s^t) = 0, \\ \log Y(s^t) > \log \hat{Y}(s^t) & \textit{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

Proof. See Section C.2 of this appendix. □

Therefore, using Lemmas 3 and 7, it follows that:

$$\begin{array}{lll} i(s^t) > \hat{\imath}(s^t) & \text{if and only if} & \tau_M^*(s^t) > 0, \\ i(s^t) = \hat{\imath}(s^t) & \text{if and only if} & \tau_M^*(s^t) = 0, \\ i(s^t) < \hat{\imath}(s^t) & \text{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

The result stated in Proposition 7 follows by combining the above with Theorem 3.

C.2 Proof of Lemma 7

First, we solve for the natural level of output. Let $\hat{Y}(s^t)$ and $\hat{L}(s^t)$ denote the flexible-price level of output and employment, respectively. These jointly satisfy:

$$\frac{\hat{L}(s^t)^\eta}{\hat{Y}(s^t)^{-\gamma}} = \chi A(s_t) \qquad \text{and} \qquad \frac{\hat{Y}(s^t)}{\hat{L}(s^t)} = A(s_t)$$

This is two equations in two unknowns. Solving for $\hat{Y}(s^t)$, we obtain the following expression for the flex-price level of output:

$$\hat{Y}(s^t)^{\eta+\gamma} = \chi A(s_t)^{1+\eta}.$$

Next, we solve for the realized level of output. Using the equivalent equilibrium representation articulated in Appendix B, realized output $Y(s^t)$ and employment $L(s^t)$ jointly satisfy:

$$\frac{L(s^t)^{\eta}}{Y(s^t)^{-\gamma}} = \chi\Gamma(\epsilon(s^t))A(s_t) \qquad \text{and} \qquad Y(s^t) = A(s_t)\Delta(\epsilon(s^t))L(s^t).$$

This is a system of two equations in two unknowns. Solving for $Y(s^t)$, we obtain the following expression for realized output:

$$Y(s^t)^{\eta+\gamma} = \chi A(s_t)^{1+\eta}\Gamma(\epsilon(s^t))\Delta(\epsilon(s^t))^{\eta}.$$

Combining this with the flexible-price level of output we get:

$$\frac{Y(s^t)^{\eta+\gamma}}{\hat{Y}(s^t)^{\eta+\gamma}} = \Gamma(\epsilon(s^t))\Delta(\epsilon(s^t))^{\eta}.$$

In logs:

$$\log Y(s^t) - \log \hat{Y}(s^t) = \frac{1}{\eta+\gamma}\log\Gamma(\epsilon(s^t)) + \frac{\eta}{\eta+\gamma}\log\Delta(\epsilon(s^t)).$$

First, recall from Lemmas (4) and (5) that $\Gamma(1) = 1$ and $\Delta(1) = 1$. It follows that if $\epsilon(s^t) = 1$, then $Y(s^t) = \hat{Y}(s^t)$.

Second, note that, to a first order around $\epsilon(s^t) = 1$,

$$\log\Delta(\epsilon(s^t)) \approx 0.$$

To see this, note that:

$$\log\Delta(\epsilon(s^t)) \approx \log\Delta(\epsilon(s^t))\big|_{\epsilon=1} + \left.\frac{d\log\Delta(\epsilon)}{d\epsilon}\right|_{\epsilon=1}\epsilon(s^t) = 0$$

since:

$$\Delta(1) = 1, \qquad \frac{d\log\Delta(\epsilon)}{d\epsilon} = \frac{1}{\Delta(\epsilon)}\frac{d\Delta(\epsilon)}{d\epsilon}, \qquad \text{and} \qquad \left.\frac{d\Delta(\epsilon)}{d\epsilon}\right|_{\epsilon=1} = 0.$$

This implies that for small shocks around $\epsilon(s^t) = 1$,

$$\log Y(s^t) - \log\hat{Y}(s^t) \approx \frac{1}{\eta+\gamma}\log\Gamma(\epsilon(s^t)) \tag{125}$$

From Lemma (5), we have that the function $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the following:

$$\begin{array}{lcr} \log\Gamma(\epsilon) < 0 & \text{if and only if} & \epsilon > 1, \\ \log\Gamma(\epsilon) = 0 & \text{if and only if} & \epsilon = 1, \\ \log\Gamma(\epsilon) > 0 & \text{if and only if} & \epsilon \in (0,1). \end{array}$$

Finally, using this in equation (125) and the result of Lemma (6), it follows that:

$$\begin{array}{lcr} \log Y(s^t) < \log Y^n(s^t) & \text{if and only if} & \tau_M^*(s^t) > 0, \\ \log Y(s^t) = \log Y^n(s^t) & \text{if and only if} & \tau_M^*(s^t) = 0, \\ \log Y(s^t) > \log Y^n(s^t) & \text{if and only if} & \tau_M^*(s^t) < 0. \end{array}$$

D One-Period-Ahead Tax Rates

In this section of the appendix, we characterize the economy with one-period-ahead tax rates. We first state some auxiliary results, followed by their proofs. We begin with our characterization of the set of sticky price allocations, $\mathcal{X}^s$.

Proposition 8. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a sticky-price equilibrium with one-period-ahead taxes if and only if there exist market weights $\varphi \equiv (\varphi^i)$ and a scalar $\bar{T} \in \mathbb{R}$, such that the following three sets of conditions are satisfied: (i) $y^j(s^t) = y^f(s^t)$ for all $j \in \mathcal{J}^f$, and $y^j(s^t) = y^s(s^t)$ for all $j \in \mathcal{J}^s$, for all $s^t \in S^t$; (ii) for all $s^{t-1} \in S^{t-1}$,*

$$\left[\frac{y^f(s|s^{t-1})}{Y(s|s^{t-1})}\right]^{-1/\rho} \frac{A(s)U_C^m(s|s^{t-1})}{-U_L^m(s|s^{t-1})} = \left[\frac{y^f(s'|s^{t-1})}{Y(s'|s^{t-1})}\right]^{-1/\rho} \frac{A(s')U_C^m(s'|s^{t-1})}{-U_L^m(s'|s^{t-1})}, \qquad \forall s, s'|s^{t-1}; \quad (126)$$

and (iii) condition (20) holds for every $i \in I$.

Proof. See Appendix D.1. □

Proposition 8 characterizes the set $\mathcal{X}^s$ when tax rates can be set one period in advance. Note that the conditions stated in part (ii) of the proposition are equivalent to the following statement: for all $s^{t-1} \in S^{t-1}$, there exists a positive scalar $\chi(s^{t-1}) \in \mathbb{R}_+$ such that:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} \frac{A(s_t)U_C^m(s^t)}{-U_L^m(s^t)} = \frac{1}{\chi(s^{t-1})}, \qquad \forall s^t|s^{t-1}. \quad (127)$$

This allows us to state the Ramsey planner's problem as follows.

Ramsey Planner's Problem. *The Ramsey planner chooses an allocation,*

$$x \equiv \{y^s(s^t), y^f(s^t), C(s^t), Y(s^t), L(s^t)\}_{t\geq 0, s^t \in S^t},$$

market weights $\varphi \equiv (\varphi^i)$, and scalar $\bar{T} \in \mathbb{R}$, in order to maximize (31), subject to

$$C(s^t) = Y(s^t) = \left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa) y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}, \qquad L(s^t) = \kappa\frac{y^s(s^t)}{A(s_t)} + (1-\kappa)\frac{y^f(s^t)}{A(s_t)}, \quad (128)$$

and (127).

We let $\beta^t \mu(s^t)(1-\kappa)\xi(s^t)$ denote the Lagrange multiplier on the implementability condition (127). The Ramsey optimum can be characterized as follows.

Proposition 9. *A Ramsey optimum x^* satisfies*

$$-\frac{\mathcal{W}_L(s^t) + \xi(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)}}{\mathcal{W}_C(s^t) + \xi(s^t)\chi(s^{t-1})U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}} = \frac{Y(s^t)}{L(s^t)}, \qquad \forall s^t \in S^t. \quad (129)$$

Proof. See Appendix D.2. □

This result is the counterpart to Proposition 6 for the economy with one-period ahead tax rates. Finally, we use this result to characterize the optimal monetary wedge in this economy; this characterization is presented in Theorem 4 in the main text and its proof is in Appendix D.3.

D.1 Proof of Proposition 8

Necessity. The necessity argument follows similar steps as the proof of Proposition 2 [see Appendix A.6]. In particular, we combine the flexible-price firm's optimality condition (21) with the CES demand function (71) and the household's intratemporal optimality condition (12) and obtain the following equilibrium condition:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \left(\frac{\rho-1}{\rho}\right)^{-1}\left[\frac{1+\tau_c(s^{t-1})}{(1-\tau_r(s^{t-1}))(1-\tau_\ell(s^{t-1}))}\right]\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{A(s_t)} = 0.$$

We let

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

denote the wedge due to the markup and taxes, and rewrite the previous condition as follows:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{1}{\chi(s^{t-1})}\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{A(s_t)} = 0.$$

This is a necessary condition for an allocation to be supported in equilibrium. Note that the above is equivalent to the conditions stated in (126).

We can similarly combine the sticky-price firm optimality condition with the CES demand function (71) and the household's intratemporal optimality condition (12) and obtain the following equilibrium condition:

$$\sum_{s^t|s^{t-1}} U_C^m(s^t)y^s(s^t)\left\{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{\chi(s^{t-1})A(s_t)}\right\}\mu(s^t|s^{t-1}) = 0; \qquad (130)$$

for all $s^{t-1} \in S^{t-1}$. Therefore (130) is a necessary condition for an allocation to be supported in equilibrium. The remainder of the proof of necessity follows the same steps as in the proof of Proposition 2.

Sufficiency. Take any feasible allocation $x \in \mathcal{X}$, vector $\varphi \equiv (\varphi^i)$, and scalar $\bar{T} \in \mathbb{R}$ that satisfy conditions (i)-(iii) of Proposition 8. We show that there exists a price system $\mathcal{R}$, a policy $\mathcal{P}$, and a set of financial market positions $\mathcal{A}$, that support x as a sticky-price equilibrium; we construct these as follows.

First, we construct nominal prices as in the sufficiency portion of the proof of Proposition 2 [Appendix A.6]. We decompose output as in (72). Given $\mathcal{B}_t(s^{t-1}) > 0$, we set $\phi^s(s^{t-1}) \equiv \mathcal{B}_t(s^{t-1})^{-\rho}$ and prices as follows:

$$p_t^s(s^{t-1}) = \phi^s(s^{t-1})^{-1/\rho} \qquad \text{and} \qquad p_t^f(s^t) = \phi^f(s^t)^{-1/\rho}$$

This implies $\Phi(s^t) = y^s(s^t)/\mathcal{B}_t(s^{t-1})^{-\rho}$ and $\phi^f(s^t) = y^f(s^t)/\Phi(s^t)$. These prices, along with the feasibility constraint (7), imply that the aggregate price level is given by (73). These prices furthermore ensure that the CES demand curves in (71) are satisfied. We set the money supply such that $M(s^t) = P(s^t)Y(s^t)$.

Next, note that the conditions stated in (126) are equivalent to the following statement: for all $s^{t-1} \in S^{t-1}$, there exists a positive scalar $\chi(s^{t-1}) \in \mathbb{R}_+$ such that:

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \frac{1}{\chi(s^{t-1})}\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{A(s_t)} = 0. \tag{131}$$

These conditions imply that such a constant exists but is not unique. In fact, for any $s^{t-1} \in S^{t-1}$, we can choose $\chi(s^{t-1})$ freely, provided it remain strictly positive. In particular, we set $\chi(s^{t-1})$ as follows:

$$\chi(s^{t-1}) = -\frac{\sum_{s^t|s^{t-1}} y^s(s^t)U_L^m(s^t)\frac{1}{A(s_t)}\mu(s^t|s^{t-1})}{\sum_{s^t|s^{t-1}} y^s(s^t)U_C^m(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\mu(s^t|s^{t-1})} > 0. \tag{132}$$

Next, we set tax rates $\{\tau_\ell(s^{t-1}), \tau_c(s^{t-1}), \tau_r(s^{t-1})\}$ such that they jointly satisfy:

$$\frac{(1-\tau_\ell(s^{t-1}))(1-\tau_r(s^{t-1}))}{1+\tau_c(s^{t-1})} = \left(\frac{\rho-1}{\rho}\right)^{-1}\chi(s^{t-1}). \tag{133}$$

For any strictly positive $\chi(s^{t-1})$ and $\rho > 1$, such tax rates exist.

Combining (133) with condition (131), we obtain:

$$0 = \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} + \left(\frac{\rho-1}{\rho}\right)^{-1}\left[\frac{1+\tau_c(s^{t-1})}{(1-\tau_r(s^{t-1}))(1-\tau_\ell(s^{t-1}))}\right]\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{A(s_t)}. \tag{134}$$

Furthermore, combining (133) with condition (132) and rearranging, we obtain:

$$\begin{aligned}0 = \sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\Bigg\{&\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} \\ &+ \left(\frac{\rho-1}{\rho}\right)^{-1}\left[\frac{1+\tau_c(s^{t-1})}{(1-\tau_r(s^{t-1}))(1-\tau_\ell(s^{t-1}))}\right]\frac{U_L^m(s^t)}{U_C^m(s^t)}\frac{1}{A(s_t)}\Bigg\}.\end{aligned} \tag{135}$$

Next, we set the real wage as follows:

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

and therefore satisfy the household's intratemporal condition in (12). Substituting the above expression for the real wage into (134) and (135), we obtain:

$$0 = \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} - \frac{W(s^t)}{P(s^t)}\left[(1-\tau_r(s^{t-1}))\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)}.$$

$$0 = \sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})U_C^m(s^t)y^s(s^t)\left\{\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} - \frac{W(s^t)}{P(s^t)}\left[(1-\tau_r(s^{t-1}))\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{1}{A(s_t)}\right\}$$

Combining these with the CES demand functions in (71), and with some rearrangement, we derive the following two conditions:

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

and

$$\sum_{s^t|s^{t-1}} \mu(s^t|s^{t-1})\frac{U_C^m(s^t)}{P(s^t)}Y(s^t)P(s^t)^\rho\left\{p_t^s(s^{t-1}) - \left[(1-\tau_r(s^{t-1}))\left(\frac{\rho-1}{\rho}\right)\right]^{-1}\frac{W(s^t)}{A(s_t)}\right\} = 0.$$

Therefore both the flexible-price and the sticky-price firm's optimality conditions are satisfied. The remainder of the proof of sufficiency follows the same steps as in the proof of Proposition 2 [Appendix A.6].

D.2 Proof of Proposition 9

We write the planner's Lagrangian as follows:

$$\begin{aligned}
\mathcal{L} = &\sum_t\sum_{s^t}\beta^t\mu(s^t)\mathcal{W}(C(s^t),L(s^t);\varphi,\nu,\lambda) \\
&+\sum_t\sum_{s^t}\beta^t\mu(s^t)\varsigma^Y(s^t)\left\{\left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}}+(1-\kappa)y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}-Y(s^t)\right\} \\
&+\sum_t\sum_{s^t}\beta^t\mu(s^t)\varsigma^L(s^t)\left\{\kappa\frac{y^s(s^t)}{A(s_t)}+(1-\kappa)\frac{y^f(s^t)}{A(s_t)}-L(s^t)\right\} \\
&+\sum_t\sum_{s^t}\beta^t\mu(s^t)\varsigma^C(s^t)\left\{Y(s^t)-C(s^t)\right\} \\
&+\sum_t\sum_{s^t}\beta^t\mu(s^t)\xi(s^t)\left\{\chi(s^{t-1})U_C^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}+U_L^m(s^t)\frac{1}{A(s_t)}\right\}
\end{aligned}$$

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

$$0 = \kappa\varsigma^Y(s^t)\left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}}+(1-\kappa)y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}-1}y^s(s^t)^{\frac{\rho-1}{\rho}-1}+\kappa\varsigma^L(s^t)\frac{1}{A(s_t)}, \tag{136}$$

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

$$\begin{aligned}
0 =&(1-\kappa)\varsigma^Y(s^t)\left[\kappa y^s(s^t)^{\frac{\rho-1}{\rho}}+(1-\kappa)y^f(s^t)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}-1}y^f(s^t)^{\frac{\rho-1}{\rho}-1}+(1-\kappa)\varsigma^L(s^t)\frac{1}{A(s_t)} \\
&-\frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{1}{y^f(s^t)}.
\end{aligned} \tag{137}$$

Note that we can rewrite (136) as

$$0 = \kappa\varsigma^Y(s^t)Y(s^t)^{1/\rho}y^s(s^t)^{\frac{\rho-1}{\rho}}+\kappa\varsigma^L(s^t)\frac{y^s(s^t)}{A(s_t)} \tag{138}$$

We can also rewrite (137) as:

$$0 = (1-\kappa)\varsigma^{Y}(s^t)Y(s^t)^{1/\rho}y^f(s^t)^{\frac{\rho-1}{\rho}} + (1-\kappa)\varsigma^{L}(s^t)\frac{y^f(s^t)}{A(s_t)} - \frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} \quad (139)$$

Summing (138) and (139) yields:

$$0 = \varsigma^{Y}(s^t)Y(s^t) + \varsigma^{L}(s^t)L(s^t) - \frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}, \quad (140)$$

which can be rewritten as follows:

$$-\frac{\varsigma^{L}(s^t)}{\varsigma^{Y}(s^t) - \frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{1}{Y(s^t)}} = \frac{Y(s^t)}{L(s^t)}. \quad (141)$$

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

$$0 = \mathcal{W}_C(s^t) - \varsigma^{C}(s^t) + \xi(s^t)\chi(s^{t-1})U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}, \quad (142)$$

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

$$0 = -\varsigma^{Y}(s^t) + \varsigma^{C}(s^t) + \frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}\frac{1}{Y(s^t)}. \quad (143)$$

Combining (142) and (143) yields:

$$\varsigma^{Y}(s^t) = \mathcal{W}_C(s^t) + \xi(s^t)\chi(s^{t-1})\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}\left[U_{CC}^m(s^t) + \frac{1}{\rho}U_C^m(s^t)\frac{1}{Y(s^t)}\right] \quad (144)$$

The FOC with respect to $L(s^t)$ implies:

$$\varsigma^{L}(s^t) = \mathcal{W}_L(s^t) + \xi(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)}. \quad (145)$$

Finally, we use (144) and (145) to substitute for $\varsigma^{Y}(s^t)$ and $\varsigma^{L}(s^t)$ in (141) and obtain:

$$-\frac{\mathcal{W}_L(s^t) + \xi(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)}}{\mathcal{W}_C(s^t) + \xi(s^t)\chi(s^{t-1})U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}} = \frac{Y(s^t)}{L(s^t)}$$

as in (129).

D.3 Proof of Theorem 4

We substitute the expressions for $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ from (101) and (102) into (129) and obtain the following Ramsey optimality condition:

$$-\frac{U_L^m(s^t)}{U_C^m(s^t)}\left\{\frac{\sum_{i\in I}\pi^i\omega_L^i(\varphi,s_t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]+\xi(s^t)\frac{U_{LL}^m(s^t)}{U_L^m(s^t)}\frac{1}{A(s_t)}}{\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]+\xi(s^t)\chi(s^{t-1})\frac{U_{CC}^m(s^t)}{U_C^m(s^t)}\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}}\right\}=\frac{Y(s^t)}{L(s^t)}.$$

Therefore the optimal monetary wedge, as defined in (37), satisfies:

$$1-\tau_M^*(s^t)=\frac{\mathcal{H}(s^{t-1})+\xi(s^t)\frac{U_{CC}^m(s^t)}{U_C^m(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}}{\mathcal{I}(s_t)+\xi(s^t)\frac{U_{LL}^m(s^t)}{U_L^m(s^t)}\frac{1}{A(s_t)}}. \tag{146}$$

where $\mathcal{I}(s_t)$ is defined in (103) and we let $\mathcal{H}(s^{t-1})\equiv\chi(s^{t-1})^{-1}\Omega_C>0$.

First, note that when $\xi(s^t)=0$, the constraint is slack. Therefore, it is clear that $\tau_M^*(s^t)=0$ if and only if

$$\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})\equiv\mathcal{H}(s^{t-1}).$$

Next we use the representation of the monetary tax in (110) and repeated here:

$$A(s_t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}=(1-\tau_M^*(s^t))\frac{Y(s^t)}{L(s^t)} \tag{147}$$

Substituting the optimal monetary wedge from (146) into (147) we obtain:

$$A(s_t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}=\left\{\frac{\mathcal{H}(s^{t-1})+\xi(s^t)\frac{U_{CC}^m(s^t)}{U_C^m(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}}{\mathcal{I}(s_t)+\xi(s^t)\frac{U_{LL}^m(s^t)}{U_L^m(s^t)}\frac{1}{A(s_t)}}\right\}\frac{Y(s^t)}{L(s^t)}.$$

Rearrangement, yields:

$$1=\frac{\mathcal{H}(s^{t-1})Y(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}+\xi(s^t)\frac{U_{CC}^m(s^t)C(s^t)}{U_C^m(s^t)}}{\mathcal{I}(s_t)A(s_t)L(s^t)+\xi(s^t)\frac{U_{LL}^m(s^t)L(s^t)}{U_L^m(s^t)}},$$

which reduces to:

$$\mathcal{I}(s_t)+(\eta+\gamma)\frac{\xi(s^t)}{A(s_t)L(s^t)}-\mathcal{H}(s^{t-1})\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}=0$$

We define

$$\hat{\xi}(s^t)\equiv\frac{\xi(s^t)}{A(s_t)L(s^t)}\mathcal{H}(s^{t-1})^{-1} \tag{148}$$

We have that:

$$\mathcal{I}(s_t) + (\eta+\gamma)\mathcal{H}(s^{t-1})\hat{\xi}(s^t) - \mathcal{H}(s^{t-1})\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho} = 0$$

Next, using condition (147) we have the following optimality condition:

$$\mathcal{I}(s_t) + (\eta+\gamma)\mathcal{H}(s^{t-1})\hat{\xi}(s^t) - \mathcal{H}(s^{t-1})(1-\tau_M(s^t))^{-1} = 0$$

We let g be the function defined by:

$$g(\mathcal{I}(s_t), \tau_M(s^t)) \equiv \mathcal{I}(s_t) + \mathcal{H}(s^{t-1})\left[(\eta+\gamma)\hat{\xi}(s^t) - (1-\tau_M(s^t))^{-1}\right].$$

The optimal monetary tax satisfies: $g(\mathcal{I}(s_t), \tau_M^*(s^t)) = 0$. By the implicit function theorem:

$$\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)} = -\frac{dg/d\mathcal{I}(s_t)}{dg/d\tau_M^*(s^t)} = -\frac{1}{\mathcal{H}(s^{t-1})\left\{(\eta+\gamma)\frac{d\hat{\xi}(s^t)}{d\tau_M(s^t)} - (1-\tau_M^*(s^t))^{-2}\right\}}$$

Therefore derivative of the optimal monetary tax satisfies:

$$\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)} = \frac{1}{\mathcal{H}(s^{t-1})}\left\{(1-\tau_M^*(s^t))^{-2} - (\eta+\gamma)\frac{d\hat{\xi}(s^t)}{d\tau_M^*(s^t)}\right\}^{-1} \tag{149}$$

An expression for $\hat{\xi}(s^t)$. The planner's optimality condition in (138) implies:

$$\varsigma^L(s^t) = -\varsigma^Y(s^t)A(s_t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}$$

Substituting this into (140) we obtain:

$$0 = \varsigma^Y(s^t)Y(s^t) - \varsigma^Y(s^t)\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} A(s_t)L(s^t) - \frac{1}{\rho}\xi(s^t)\chi(s^{t-1})U_C^m(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho},$$

which can be rewritten as:

$$0 = \frac{\varsigma^Y(s^t)}{\chi(s^{t-1})U_C^m(s^t)}\left[\frac{Y(s^t)}{A(s_t)L(s^t)} - \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\right] - \frac{1}{\rho}\frac{\xi(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}$$

Rearranging, we get that:

$$\frac{\xi(s^t)}{A(s_t)L(s^t)} = \rho\frac{\varsigma^Y(s^t)}{\chi(s^{t-1})U_C^m(s^t)}\left[\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho} - \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\right] \tag{150}$$

Therefore

$$\hat{\xi}(s^t) = \rho\frac{\mathcal{H}(s^{t-1})^{-1}}{\chi(s^{t-1})U_C^m(s^t)}\varsigma^Y(s^t)\left[\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho} - \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\right] \tag{151}$$

In what follows, we turn back to the equivalent equilibrium representation used in Section (B) of the Appendix. Recall the forecast errors $\epsilon(s^t)$ defined in (28). From equation (123) we have that the monetary tax satisfies:

$$1-\tau_M(s^t)=\frac{\kappa\epsilon(s^t)^{-\rho}+(1-\kappa)}{\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)}.$$

Combining this with (147), we obtain:

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

Furthermore, using (117), the following is also true:

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

Therefore, we may rewrite (151) as follows:

$$\hat{\xi}(s^t)=\rho\frac{\mathcal{H}(s^{t-1})^{-1}}{\chi(s^{t-1})U_C^m(s^t)}\varsigma^Y(s^t)\left[\frac{(1-\kappa)(1-\epsilon(s^t))}{\kappa\epsilon(s^t)^{-\rho}+(1-\kappa)}\right] \tag{152}$$

Next we use the planner's optimality condition in (144). Substituting in for $\xi(s^t)$ from (150) into (144) we have that:

$$\varsigma^Y(s^t)=\mathcal{W}_C(s^t)+\varsigma^Y(s^t)(1-\gamma\rho)\left[1-\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{A(s_t)L(s^t)}{Y(s^t)}\right]$$

Therefore

$$\varsigma^Y(s^t)=\mathcal{W}_C(s^t)\left[1-(1-\gamma\rho)\left[1-\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho}\frac{A(s_t)L(s^t)}{Y(s^t)}\right]\right]^{-1}$$

We have that:

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

Therefore

$$\varsigma^Y(s^t)=\mathcal{W}_C(s^t)\left[1-(1-\gamma\rho)\left[1-\epsilon(s^t)\frac{\kappa\epsilon(s^t)^{-\rho}+(1-\kappa)}{\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)}\right]\right]^{-1}$$

which reduces to

$$\varsigma^Y(s^t)=\mathcal{W}_C(s^t)\left[1-(1-\gamma\rho)(1-\kappa)\left[\frac{1-\epsilon(s^t)}{\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)}\right]\right]^{-1}$$

Therefore

$$\varsigma^Y(s^t)=\mathcal{W}_C(s^t)\frac{\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)}{\kappa\epsilon(s^t)^{1-\rho}+(1-\gamma\rho)(1-\kappa)\epsilon(s^t)+\gamma\rho(1-\kappa)}$$

Substituting this expression for $\varsigma^Y(s^t)$ into (152), we obtain the following expression for $\hat{\xi}(s^t)$ as a function of $\epsilon(s^t)$:

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

Next, for shorthand we let

$$\Sigma(\epsilon) \equiv \frac{(1-\kappa)(1-\epsilon)}{\kappa\epsilon^{1-\rho} + (1-\gamma\rho)(1-\kappa)\epsilon + \gamma\rho(1-\kappa)} \tag{153}$$

Therefore:

$$\hat{\xi}(s^t) = \rho\Sigma(\epsilon(s^t))(1-\tau_M(s^t))^{-1} \tag{154}$$

Derivative of $\tau_M(s^t)$. The derivative of the optimal monetary "tax" satisfies (149). Evaluating this derivative at the benchmark in which $\tau_M(s^t) = 0$, we have:

$$\left.\frac{d\tau_M(s^t)}{d\mathcal{I}(s_t)}\right|_{\tau_M(s^t)=0} = \frac{1}{\mathcal{H}(s^{t-1})} \left\{ 1 - (\eta+\gamma) \left.\frac{d\hat{\xi}(s^t)}{d\tau_M(s^t)}\right|_{\tau_M(s^t)=0} \right\}^{-1} \tag{155}$$

where $\hat{\xi}(s^t)$ satisfies (154). Taking the first derivative of the expression in (154), we get:

$$\frac{d\hat{\xi}}{d\tau_M} = \rho \left\{ \Sigma(\epsilon)(1-\tau_M)^{-2} + (1-\tau_M)^{-1} \frac{d\Sigma}{d\epsilon}\frac{d\epsilon}{d\tau_M} \right\}. \tag{156}$$

The derivative $d\Sigma/d\epsilon$ satisfies:

$$\frac{d\Sigma(\epsilon)}{d\epsilon} = -(1-\kappa)\frac{\kappa\epsilon^{1-\rho} + (1-\rho)\kappa\epsilon^{-\rho}(1-\epsilon) + (1-\kappa)}{(\kappa\epsilon^{1-\rho} + (1-\gamma\rho)(1-\kappa)\epsilon + \gamma\rho(1-\kappa))^2}$$

Next, we obtain $d\epsilon/d\tau_M$ as follows. The monetary tax satisfies equation (123):

$$1 - \tau_M(\epsilon) = \frac{\kappa\epsilon^{-\rho} + (1-\kappa)}{\kappa\epsilon^{1-\rho} + (1-\kappa)}$$

Rearranging, we obtain the following expression:

$$\kappa\epsilon^{1-\rho} - \kappa\epsilon^{-\rho} - \tau_M\kappa\epsilon^{1-\rho} - \tau_M(1-\kappa) = 0$$

We let ϱ be the function defined by:

$$\varrho(\tau_M, \epsilon) \equiv (1-\tau_M)\kappa\epsilon^{1-\rho} - \kappa\epsilon^{-\rho} - \tau_M(1-\kappa),$$

with $\varrho(\tau_M, \epsilon) = 0$. By the implicit function theorem:

$$\frac{d\epsilon}{d\tau_M} = -\frac{d\varrho/d\tau_M}{d\varrho/d\epsilon} = \frac{\kappa\epsilon^{1-\rho} + (1-\kappa)}{(1-\rho)(1-\tau_M)\kappa\epsilon^{-\rho} + \rho\kappa\epsilon^{-\rho-1}}.$$

Therefore the last term in (156) satisfies:

$$\Sigma(\epsilon)(1-\tau_M)^{-2}+(1-\tau_M)^{-1}\frac{d\Sigma}{d\epsilon}\frac{d\epsilon}{d\tau_M} = \frac{(1-\kappa)(1-\epsilon)}{\kappa\epsilon^{1-\rho}+(1-\gamma\rho)(1-\kappa)\epsilon+\gamma\rho(1-\kappa)}(1-\tau_M)^{-2} \tag{157}$$
$$-\left\{\frac{1-\kappa}{1-\tau_M}\left[\frac{\kappa\epsilon^{1-\rho}+(1-\rho)\kappa\epsilon^{-\rho}(1-\epsilon)+(1-\kappa)}{(\kappa\epsilon^{1-\rho}+(1-\gamma\rho)(1-\kappa)\epsilon+\gamma\rho(1-\kappa))^2}\right]\right.$$
$$\left.\times\left[\frac{\kappa\epsilon^{1-\rho}+(1-\kappa)}{(1-\rho)(1-\tau_M)\kappa\epsilon^{-\rho}+\rho\kappa\epsilon^{-\rho-1}}\right]\right\}$$

Evaluating this term at $\tau_M=0$, or equivalently at $\epsilon=1$, we have:

$$\left[\Sigma(\epsilon)(1-\tau_M)^{-2}+(1-\tau_M)^{-1}\frac{d\Sigma}{d\epsilon}\frac{d\epsilon}{d\tau_M}\right]_{\tau_M=0} = -\left(\frac{1-\kappa}{\kappa}\right)$$

And furthermore evaluating (156) at $\tau_M=0$, we get:

$$\left.\frac{d\hat{\xi}}{d\tau_M}\right|_{\tau_M=0} = -\rho\frac{1-\kappa}{\kappa}$$

Substituting this into (155), we obtain:

$$\left.\frac{d\tau_M(s^t)}{d\mathcal{I}(s_t)}\right|_{\tau_M(s^t)=0} = \frac{1}{\bar{\mathcal{I}}(s^{t-1})\left\{1+\rho(\eta+\gamma)\frac{1-\kappa}{\kappa}\right\}}>0.$$

where $\bar{\mathcal{I}}(s^{t-1})\equiv\mathcal{H}(s^{t-1})$. Therefore, taking a first-order Taylor approximation of $\tau_M^*(s^t)$ around the point $\mathcal{I}(s_t)=\bar{\mathcal{I}}(s^{t-1})$, we have:

$$\tau_M^*(s^t)\approx 0+\frac{1}{\bar{\mathcal{I}}(s^{t-1})\left\{1+\rho(\eta+\gamma)\frac{1-\kappa}{\kappa}\right\}}[\mathcal{I}(s_t)-\bar{\mathcal{I}}(s^{t-1})],$$

which coincides with the expression in (42).

E Constrained Profit Taxation

In this section of the appendix we provide the proofs for the economy with constrained profit taxation presented in Section 6.

E.1 Proof of Proposition 5

Necessity. Necessity of parts (i) and (ii) of the proposition follow from the same steps as those used to prove Proposition 2. What remains to be shown is necessity of the budget implementability conditions in (44).

First, we derive an expression for real profits, $\Pi(s^t)/P(s^t)$, in terms of allocations alone. We can write aggregate profits 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. Profits of these firms are given by:

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

Combining these expressions with the demand functions,

$$\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho} = \frac{p_t^f(s^t)}{P(s^t)} \quad \text{and} \quad \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{-1/\rho} = \frac{p_t^s(s^{t-1})}{P(s^t)}, \tag{158}$$

we find that real profits of these firms are given by:

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

Together, these imply that aggregate real profits can be written as:

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

Finally, the resource constraints in (35) imply that real profits are given by:

$$\frac{\Pi(s^t)}{P(s^t)} = (1-\tau_r)Y(s^t) - \frac{W(s^t)}{P(s^t)}L(s^t).$$

Next, we replace the real wage $W(s^t)/P(s^t)$ in the above expression using the representative household's intratemporal condition, (12). This gives us the following expression for real profits:

$$\frac{\Pi(s^t)}{P(s^t)} = (1-\tau_r)C(s^t) + \frac{1+\tau_c}{1-\tau_\ell}\frac{U_L^m(s^t)}{U_C^m(s^t)}L(s^t) \tag{159}$$

Multiplying both sides of this by $U_C^m(s^t)/1+\tau_c$, we get:

$$\frac{U_C^m(s^t)}{1+\tau_c}\frac{\Pi(s^t)}{P(s^t)} = (1-\tau_r)\frac{U_C^m(s^t)}{1+\tau_c}C(s^t) + \frac{U_L^m(s^t)}{1-\tau_\ell}L(s^t),$$

Substituting this expression into the implementability conditions in (18), we obtain:

$$\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi,s_t)L(s^t)\right]$$
$$= U_C^m(s_0)\bar{T} + \sigma_0^i\frac{1-\tau_\Pi}{1-\tau_\ell}\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[\frac{(1-\tau_\ell)(1-\tau_r)}{1+\tau_c}U_C^m(s^t)C(s^t) + U_L^m(s^t)L(s^t)\right].$$

Using the definition of χ in (24), these conditions can be written as:

$$\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[U_C^m(s^t)\omega_C^i(\varphi)C(s^t) + U_L^m(s^t)\omega_L^i(\varphi,s_t)L(s^t)\right]$$
$$= U_C^m(s_0)\bar{T} + \sigma_0^i\frac{1-\tau_\Pi}{1-\tau_\ell}\sum_t\sum_{s^t}\beta^t\mu(s^t)\left[\chi\frac{\rho}{\rho-1}U_C^m(s^t)C(s^t) + U_L^m(s^t)L(s^t)\right]$$

Finally, we define $\vartheta \equiv \frac{1-\tau_\Pi}{1-\tau_\ell}$ and obtain the condition in (44).

Sufficiency. Follows the same argument as in the sufficiency proof of Proposition 2.

E.2 The Ramsey Problem

In this section of the appendix, we state and solve the Ramsey problem. We assume the same utilitarian social welfare function (29) as before. Again we let $\pi^i \nu^i$ denote the Lagrange multiplier on the implementability condition (44) of type $i \in I$ and subsume these into the maximand. Given ϑ, we can define a new pseudo-welfare function $\hat{\mathcal{W}}(\cdot)$ as follows:

$$\begin{aligned}\hat{\mathcal{W}}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda, \sigma, \chi, \vartheta) \equiv \quad & \mathcal{W}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda) \\ & - \left(\vartheta \sum_{i \in I} \pi^i \nu^i \sigma_0^i\right) \left[\chi \frac{\rho}{\rho - 1} U_C^m(s^t) C(s^t) + U_L^m(s^t) L(s^t)\right]\end{aligned} \tag{160}$$

where $\mathcal{W}(\cdot)$ is defined in (30). With this, we recast the Ramsey planning problem as follows.

Ramsey Planner's Problem. *The Ramsey planner chooses an allocation* $x \equiv \{y^s(s^t), y^f(s^t), C(s^t), Y(s^t), L(s^t)\}_{t \geq 0, s^t \in S^t}$*, market weights* $\varphi \equiv (\varphi^i)$*, and constants* $\bar{T} \in \mathbb{R}$ *and* $\chi \in \mathbb{R}_+$*, in order to maximize:*

$$\sum_t \sum_{s^t} \beta^t \mu(s^t) \hat{\mathcal{W}}(C(s^t), L(s^t), s_t; \varphi, \nu, \lambda, \sigma, \chi, \vartheta) - U_C^m(s_0) \bar{T} \sum_{i \in I} \pi^i \nu^i \tag{161}$$

subject to (35)*,* (25)*, and* (26)*.*

We let $\beta^t \mu(s^t)(1 - \kappa)\xi(s^t)$ and $\beta^t \mu(s^{t-1}) \kappa \upsilon(s^{t-1})$ denote the Lagrange multipliers on the implementability conditions (25) and (26), respectively. We obtain the following Ramsey optimality condition.

Proposition 10. *A Ramsey optimum* x^* *satisfies, for all* $s^t \in S^t$*,*

$$-\frac{\hat{\mathcal{W}}_L(s^t) + (U_L^m(s^t) + U_{LL}^m(s^t) L(s^t)) \left[\kappa \upsilon(s^{t-1}) \frac{y^s(s^t)}{A(s_t) L(s^t)} + (1 - \kappa)\xi(s^t) \frac{y^f(s^t)}{A(s_t) L(s^t)}\right]}{\hat{\mathcal{W}}_C(s^t) + \chi(U_C^m(s^t) + U_{CC}^m(s^t) C(s^t)) \left[\kappa \upsilon(s^{t-1}) \left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1 - 1/\rho} + (1 - \kappa)\xi(s^t) \left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1 - 1/\rho}\right]} = \frac{Y(s^t)}{L(s^t)}. \tag{162}$$

where

$$\hat{\mathcal{W}}_C(s^t) = \mathcal{W}_C(s^t) - \left(\vartheta \sum_{i \in I} \pi^i \nu^i \sigma_0^i\right) \chi \frac{\rho}{\rho - 1} [U_C^m(s^t) + U_{CC}^m(s^t) C(s^t)] \tag{163}$$

$$\hat{\mathcal{W}}_L(s^t) = \mathcal{W}_L(s^t) - \left(\vartheta \sum_{i \in I} \pi^i \nu^i \sigma_0^i\right) [U_L^m(s^t) + U_{LL}^m(s^t) L(s^t)] \tag{164}$$

Proof. The planner's problem is exactly the same as in the proof of Proposition 6, with the exception of maximizing $\hat{\mathcal{W}}(s^t)$ instead of $\mathcal{W}(s^t)$. Therefore, following the same procedure as the proof of Proposition 6, we obtain the expression in (162). Furthermore, taking the first derivatives of $\hat{\mathcal{W}}(\cdot)$, as defined in (160), with respect to $C(s^t)$ and $L(s^t)$, we obtain the expressions in stated in (163) and (164). □

E.3 Proof of Theorem 5

The Ramsey optimum satisfies (162). Substituting into (162) our expressions for $\hat{\mathcal{W}}_C(s^t)$ and $\hat{\mathcal{W}}_L(s^t)$ from (163) and (164), as well as our expressions for $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ from (101) and (102), and solving for the implicit optimal monetary wedge, we get:

$$1-\tau_M^*(s^t)=\Bigg\{(\chi^*)^{-1}\sum_{i\in I}\pi^i\omega_C^i(\varphi)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1-\gamma)\right]$$
$$+(1-\gamma)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho}+(1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1-1/\rho}-\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i\right]\Bigg\}$$
$$\times\Bigg\{\sum_{i\in I}\pi^i\omega_L^i(\varphi,s_t)\left[\frac{\lambda^i}{\varphi^i}+\nu^i(1+\eta)\right]$$
$$+(1+\eta)\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{A(s_t)L(s^t)}+(1-\kappa)\xi(s^t)\frac{y^f(s^t)}{A(s_t)L(s^t)}-\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i\right]\Bigg\}^{-1}$$

Therefore the optimal monetary wedge satisfies:

$$1-\tau_M^*(s^t)=\frac{\mathcal{H}+(1-\gamma)\left[\kappa\upsilon(s^{t-1})\left[\frac{y^s(s^t)}{Y(s^t)}\right]^{1-1/\rho}+(1-\kappa)\xi(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1-1/\rho}-\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}{\mathcal{I}(s_t)+(1+\eta)\left[\kappa\upsilon(s^{t-1})\frac{y^s(s^t)}{A(s_t)L(s^t)}+(1-\kappa)\xi(s^t)\frac{y^f(s^t)}{A(s_t)L(s^t)}-\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}. \tag{165}$$

where $\mathcal{I}(s_t)$ and $\mathcal{H}$ are defined in (103).

Threshold. We first consider the conditions under which $\tau_M^*(s^t)=0$. In this state: $y^s(s^t)=y^f(s^t)=Y(s^t)=A(s_t)L(s^t)$. Condition (165) reduces to:

$$1=\frac{\mathcal{H}+(1-\gamma)\left[\kappa\upsilon(s^{t-1})+(1-\kappa)\xi(s^t)-\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}{\mathcal{I}(s_t)+(1+\eta)\left[\kappa\upsilon(s^{t-1})+(1-\kappa)\xi(s^t)-\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}$$

Furthermore, conditions (83) and (85) imply that $\xi(s^t)=\upsilon(s^{t-1})$ in this state. Therefore:

$$1=\frac{\mathcal{H}+(1-\gamma)\upsilon(s^{t-1})-(1-\gamma)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i}{\mathcal{I}(s_t)+(1+\eta)\upsilon(s^{t-1})-(1+\eta)\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i}$$

Solving this for $\mathcal{I}(s_t)$ we obtain the following threshold:

$$\bar{\mathcal{I}}_\vartheta(s^{t-1})=\mathcal{H}-(\eta+\gamma)\upsilon(s^{t-1})+\left[(1+\eta)-(1-\gamma)\frac{\rho}{\rho-1}\right]\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i.$$

Therefore if $\mathcal{I}(s_t)=\bar{\mathcal{I}}_\vartheta(s^{t-1})$, the optimal monetary tax is equal to zero: $\tau_M^*(s^t)=0$. Finally, letting$\bar{\mathcal{I}}_0(s^{t-1})=\mathcal{H}-(\eta+\gamma)\upsilon(s^{t-1})$, we obtain the following expression:

$$\bar{\mathcal{I}}_\vartheta(s^{t-1})=\bar{\mathcal{I}}_0(s^{t-1})+\left[(1+\eta)+\frac{\rho}{\rho-1}(\gamma-1)\right]\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i. \tag{166}$$

The fictitious tax wedge. We next define a fictitious tax wedge as follows:

$$1 - \hat{\tau}_\vartheta(s^t) \equiv \frac{\mathcal{H} + (1-\gamma)\upsilon(s^{t-1}) - (1-\gamma)\frac{\rho}{\rho-1}\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i}{\mathcal{I}(s_t) + (1+\eta)\upsilon(s^{t-1}) - (1+\eta)\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i} \tag{167}$$

The wedge $1-\hat{\tau}(s^t)$ is continuous and strictly decreasing in $\mathcal{I}(s_t)$, as all other terms are constants (conditional on s^{t-1}). Furthermore, note that $\hat{\tau}(s^t) = 0$ if and only if $\mathcal{I}(s_t) = \bar{\mathcal{I}}_\vartheta(s^{t-1})$. As a result, the fictitious tax $\hat{\tau}(s^t)$ trivially satisfies:

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

The remainder of the proof follows the exact same steps as in the proof of Theorem 3.

E.4 One-Period-Ahead Tax Rates

In this section of the appendix, we consider the economy with constrained profit taxation *and* one-period-ahead tax rates. Specifically, we let τ_c and τ_r at time t be contingent on s^{t-1}. We begin with our characterization of the set of sticky price allocations.

Lemma 8. *A feasible allocation $x \in \mathcal{X}$ can be implemented as a sticky-price equilibrium with one-period-ahead taxes if and only if there exist market weights $\varphi \equiv (\varphi^i)$, a scalar $\bar{T} \in \mathbb{R}$, and a weakly positive scalar $\vartheta \in \mathbb{R}_{\geq 0}$, such that parts (i)-(ii) of Proposition 8 are satisfied, and condition* (44) *holds for every $i \in I$.*

Proof. The proof is analogous to the proof of Proposition 8. □

Next, we can state the planner's problem as follows.

Ramsey Planner's Problem. *The Ramsey planner chooses an allocation, $x \equiv \{y^s(s^t), y^f(s^t), C(s^t), Y(s^t), L(s^t)\}_{t\geq 0, s^t \in S^t}$, market weights $\varphi \equiv (\varphi^i)$, and scalar $\bar{T} \in \mathbb{R}$, in order to maximize* (161) *subject to* (128) *and* (127).

We let $\beta^t \mu(s^t)(1-\kappa)\xi(s^t)$ denote the Lagrange multiplier on the implementability condition (127). The Ramsey optimum can be characterized as follows.

Lemma 9. *A Ramsey optimum x^* satisfies*

$$-\frac{\hat{\mathcal{W}}_L(s^t) + \xi(s^t) U^m_{LL}(s^t) \frac{1}{A(s_t)}}{\hat{\mathcal{W}}_C(s^t) + \xi(s^t)\chi(s^{t-1}) U^m_{CC}(s^t) \left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}} = \frac{Y(s^t)}{L(s^t)}, \qquad \forall s^t \in S^t. \tag{168}$$

Proof. The planner's problem is the same as in the proof of Proposition 9, with the exception of maximizing $\hat{\mathcal{W}}(s^t)$ instead of $\mathcal{W}(s^t)$. Therefore, we follow the same procedure as in the proof of Proposition 9, and we obtain the expression in (168). This is identical to the Ramsey optimality condition (129) but with $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ replaced by $\hat{\mathcal{W}}_C(s^t)$ and $\hat{\mathcal{W}}_L(s^t)$. □

E.5 Proof of Theorem 6

The Ramsey optimum satisfies (168). Substituting in our expressions for $\hat{\mathcal{W}}_C(s^t)$ and $\hat{\mathcal{W}}_L(s^t)$ from (163) and (164):

$$-\frac{\mathcal{W}_L(s^t)-\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i\right)\left[U_L^m(s^t)+U_{LL}^m(s^t)L(s^t)\right]+\xi(s^t)U_{LL}^m(s^t)\frac{1}{A(s_t)}}{\mathcal{W}_C(s^t)-\chi(s^{t-1})\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i\right)\left[U_C^m(s^t)+U_{CC}^m(s^t)C(s^t)\right]+\xi(s^t)\chi(s^{t-1})U_{CC}^m(s^t)\left(\frac{y^f(s^t)}{Y(s^t)}\right)^{-1/\rho}}$$
$$=\frac{Y(s^t)}{L(s^t)}$$

This optimality condition, along with our expressions for $\mathcal{W}_C(s^t)$ and $\mathcal{W}_L(s^t)$ in (101) and (102), imply that the optimal monetary wedge, defined in (37), satisfies:

$$1-\tau_M^*(s^t)=\frac{\mathcal{H}_\vartheta(s^{t-1})+\xi(s^t)\frac{U_{CC}^m(s^t)}{U_C^m(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}-(1-\gamma)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i}{\mathcal{I}(s_t)+\xi(s^t)\frac{U_{LL}^m(s^t)}{U_L^m(s^t)}\frac{1}{A(s_t)}-(1+\eta)\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i}. \tag{169}$$

where $\mathcal{I}(s_t)$ is defined in (103) and we let $\mathcal{H}_\vartheta(s^{t-1})\equiv\chi(s^{t-1})^{-1}\Omega_C(\varphi)>0$.

First, note that when $\xi(s^t)=0$, the constraint is slack. Therefore, $\tau_M^*(s^t)=0$ if and only if

$$\mathcal{I}(s_t)=\bar{\mathcal{I}}_\vartheta(s^{t-1})\equiv\mathcal{H}_\vartheta(s^{t-1})+\left[(1+\eta)-(1-\gamma)\frac{\rho}{\rho-1}\right]\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i$$

Note that $\bar{\mathcal{I}}_0(s^{t-1})\equiv\mathcal{H}_0(s^{t-1})$. Substituting the optimal monetary wedge from (169) into (147) we obtain:

$$A(s_t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}=\left\{\frac{\mathcal{H}_\vartheta(s^{t-1})+\xi(s^t)\frac{U_{CC}^m(s^t)}{U_C^m(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{-1/\rho}-(1-\gamma)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i}{\mathcal{I}(s_t)+\xi(s^t)\frac{U_{LL}^m(s^t)}{U_L^m(s^t)}\frac{1}{A(s_t)}-(1+\eta)\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i}\right\}\frac{Y(s^t)}{L(s^t)}.$$

Rearrangement, yields:

$$1=\frac{\mathcal{H}_\vartheta(s^{t-1})Y(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}+\xi(s^t)\frac{U_{CC}^m(s^t)C(s^t)}{U_C^m(s^t)}-(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)Y(s^t)\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}}{\mathcal{I}(s_t)A(s_t)L(s^t)+\xi(s^t)\frac{U_{LL}^m(s^t)L(s^t)}{U_L^m(s^t)}-(1+\eta)\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)A(s_t)L(s^t)}.$$

which reduces to:

$$\begin{aligned}0=&\mathcal{I}(s_t)+(\eta+\gamma)\frac{\xi(s^t)}{A(s_t)L(s^t)}-\mathcal{H}_\vartheta(s^{t-1})\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\\&-\left\{(1+\eta)-(1-\gamma)\frac{\rho}{\rho-1}\frac{Y(s^t)}{A(s_t)L(s^t)}\left[\frac{y^f(s^t)}{Y(s^t)}\right]^{1/\rho}\right\}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)\end{aligned}$$

As before we define $\hat{\xi}(s^t)$ as in (148), and using condition (147) we obtain the following optimality condition:

$$
\begin{aligned}
0 =& \mathcal{I}(s_t) + (\eta+\gamma)\mathcal{H}_\vartheta(s^{t-1})\hat{\xi}(s^t) - \mathcal{H}_\vartheta(s^{t-1})(1-\tau_M(s^t))^{-1} \\
& - \left[(1+\eta) - (1-\gamma)\frac{\rho}{\rho-1}(1-\tau_M(s^t))^{-1}\right]\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)
\end{aligned}
$$

We let $\hat{g}$ be the function defined by:

$$
\begin{aligned}
\hat{g}(\mathcal{I}(s_t), \tau_M(s^t)) \equiv& \mathcal{I}(s_t) + \mathcal{H}_\vartheta(s^{t-1})\left[(\eta+\gamma)\hat{\xi}(s^t) - (1-\tau_M(s^t))^{-1}\right] \\
& - (1+\eta)\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right) + (1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)(1-\tau_M(s^t))^{-1}.
\end{aligned}
$$

The optimal monetary tax satisfies: $\hat{g}(\mathcal{I}(s_t), \tau_M^*(s^t)) = 0$. By the implicit function theorem:

$$
\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)} = -\frac{d\hat{g}/d\mathcal{I}(s_t)}{d\hat{g}/d\tau_M^*(s^t)}
$$

where

$$
\frac{d\hat{g}/d\mathcal{I}(s_t)}{d\hat{g}/d\tau_M^*(s^t)} = \frac{1}{\mathcal{H}_\vartheta(s^{t-1})\left\{(\eta+\gamma)\frac{d\hat{\xi}(s^t)}{d\tau_M(s^t)} - (1-\tau_M^*(s^t))^{-2}\right\} + (1-\gamma)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma^i(1-\tau_M^*(s^t))^{-2}}
$$

Therefore the derivative of the optimal monetary tax satisfies:

$$
\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)} = \frac{\mathcal{H}_\vartheta(s^{t-1})^{-1}}{\left[1-(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)\mathcal{H}_\vartheta(s^{t-1})^{-1}\right](1-\tau_M^*(s^t))^{-2} - (\eta+\gamma)\frac{d\hat{\xi}(s^t)}{d\tau_M(s^t)}} \tag{170}
$$

An expression for $\hat{\xi}(s^t)$. Following the same steps as in the proof of Theorem 4, we get the following expression for $\hat{\xi}(s^t)$:

$$
\hat{\xi}(s^t) = \rho\frac{\mathcal{H}_\vartheta(s^{t-1})^{-1}}{\chi(s^{t-1})U_C^m(s^t)}\varsigma^Y(s^t)\left[\frac{(1-\kappa)(1-\epsilon(s^t))}{\kappa\epsilon(s^t)^{-\rho}+(1-\kappa)}\right] \tag{171}
$$

Next we use the planner's optimality condition in (144) but with $\hat{\mathcal{W}}_C(s^t)$ in place of of $\mathcal{W}_C(s^t)$. Following the same steps as in the proof of Theorem 4, we find that $\varsigma^Y(s^t)$ satisfies:

$$
\varsigma^Y(s^t) = \hat{\mathcal{W}}_C(s^t)\frac{\kappa\epsilon(s^t)^{1-\rho}+(1-\kappa)}{\kappa\epsilon(s^t)^{1-\rho}+(1-\gamma\rho)(1-\kappa)\epsilon(s^t)+\gamma\rho(1-\kappa)}.
$$

Substituting this expression for $\varsigma^Y(s^t)$ into (171), we obtain the following expression for $\hat{\xi}(s^t)$ as a function of $\epsilon(s^t)$:

$$
\hat{\xi}(s^t) = \rho\frac{\mathcal{H}_\vartheta(s^{t-1})^{-1}}{\chi(s^{t-1})U_C^m(s^t)}\hat{\mathcal{W}}_C(s^t)\Sigma(\epsilon)(1-\tau_M(s^t))^{-1}.
$$

where $\Sigma(\epsilon)$ is defined in (153) and $\tau_M(s^t)$ satisfies equation (123). Furthermore, substituting in our expression for $\hat{\mathcal{W}}_C(s^t)$ from (163), we obtain:

$$\hat{\xi}(s^t) = \rho \frac{\mathcal{H}_\vartheta(s^{t-1})^{-1}}{\chi(s^{t-1})} \left[\frac{\mathcal{W}_C(s^t)}{U_C^m(s^t)} - \chi(s^{t-1})(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i\right)\right] \Sigma(\epsilon)(1-\tau_M(s^t))^{-1}$$

Using the fact that $\mathcal{W}_C(s^t) = U_C^m(s^t)\Omega_C(\varphi)$, the above reduces to:

$$\hat{\xi}(s^t) = \rho \left[1 - (1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta \sum_{i\in I} \pi^i \nu^i \sigma_0^i\right) \mathcal{H}_\vartheta(s^{t-1})^{-1}\right] \Sigma(\epsilon)(1-\tau_M(s^t))^{-1}. \tag{172}$$

Derivative of $\tau_M(s^t)$. The derivative of the optimal monetary tax satisfies (170). Evaluating this derivative at the benchmark in which $\tau_M(s^t) = 0$, we have:

$$\left.\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)}\right|_{\tau_M(s^t)=0} = \frac{\mathcal{H}_\vartheta(s^{t-1})^{-1}}{\left[1-(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)\mathcal{H}_\vartheta(s^{t-1})^{-1}\right] - (\eta+\gamma)\left.\frac{d\hat{\xi}(s^t)}{d\tau_M(s^t)}\right|_{\tau_M(s^t)=0}} \tag{173}$$

where $\hat{\xi}(s^t)$ satisfies (172). Taking the first derivative of the expression in (172), we get:

$$\frac{d\hat{\xi}}{d\tau_M} = \rho\left[1-(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)\mathcal{H}_\vartheta(s^{t-1})^{-1}\right]\left[\Sigma(\epsilon)(1-\tau_M)^{-2} + (1-\tau_M)^{-1}\frac{d\Sigma}{d\epsilon}\frac{d\epsilon}{d\tau_M}\right] \tag{174}$$

Note that the last term in (174) coincides with the last term in (156). Therefore, as in the proof of Theorem 4, we have that the last term in (174) satisfies (157). Evaluating this term at $\tau_M = 0$, or equivalently at $\epsilon = 1$, we have:

$$\left[\Sigma(\epsilon)(1-\tau_M)^{-2} + (1-\tau_M)^{-1}\frac{d\Sigma}{d\epsilon}\frac{d\epsilon}{d\tau_M}\right]_{\tau_M=0} = -\left(\frac{1-\kappa}{\kappa}\right)$$

And furthermore evaluating (174) at $\tau_M = 0$, we get:

$$\left.\frac{d\hat{\xi}}{d\tau_M}\right|_{\tau_M=0} = -\rho\frac{1-\kappa}{\kappa}\left[1-(1-\gamma)\frac{\rho}{\rho-1}\left(\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right)\mathcal{H}(s^{t-1})^{-1}\right]$$

Substituting this expression into (173), we obtain:

$$\left.\frac{d\tau_M^*(s^t)}{d\mathcal{I}(s_t)}\right|_{\tau_M(s^t)=0} = \frac{1}{\left(1+\rho(\eta+\gamma)\frac{1-\kappa}{\kappa}\right)\left[\mathcal{H}_\vartheta(s^{t-1}) + (\gamma-1)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}$$

Therefore, taking a first-order Taylor approximation of $\tau_M^*(s^t)$ around $\mathcal{I}(s_t) = \hat{\mathcal{I}}(s^{t-1})$, we have:

$$\tau_M^*(s^t) \approx 0 + \frac{1}{\left(1+\rho(\eta+\gamma)\frac{1-\kappa}{\kappa}\right)\left[\mathcal{H}_\vartheta(s^{t-1}) + (\gamma-1)\frac{\rho}{\rho-1}\vartheta\sum_{i\in I}\pi^i\nu^i\sigma_0^i\right]}[\mathcal{I}(s_t) - \bar{\mathcal{I}}_\vartheta(s^{t-1})],$$

as was to be shown.