Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Individual and Aggregate Constrained Efficient Intertemporal Wedges in Dynamic Mirrleesian Economies

Yunmin Chen
YiLi Chien
and
C.C. Yang

Working Paper 2015-043B
http://research.stlouisfed.org/wp/2015/2015-043.pdf

April 2016

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Individual and Aggregate Constrained Efficient Intertemporal Wedges in Dynamic Mirrleesian Economies

Yunmin Chen
Institute of Economics, Academia Sinica

YiLi Chien*
Federal Reserve Bank of St. Louis

C.C. Yang
Institute of Economics, Academia Sinica
Department of Public Finance, National Cheng-Chi University
Department of Public Finance, Feng-Chia University

April 2016

Abstract

Assuming a neoclassical production technology, this paper characterizes constrained efficient intertemporal wedges for the *macro aggregate* as well as the *micro individual* allocation of dynamic Mirrleesian economies. We first construct "Pareto-Negishi weights" from the multipliers on a sequence of temporary incentive constraints. For a fairly general stochastic process of idiosyncratic labor productivity shocks, we show that the evolution of the Pareto-Negishi weight associated with agents' consumption is a *nonnegative martingale*. This powerful property enables us to deliver three contributions to the literature. First, we demonstrate that several celebrated results on the intertemporal wedge of the individual allocation can all be attributed to their origin in the nonnegative martingale property of consumption Pareto-Negishi weights. Second, we address optimal distortions for the aggregate allocation and establish that the constrained efficient intertemporal wedge for aggregate savings is negative or positive, depending on whether the intertemporal elasticity of substitution is elastic or inelastic. This result is another application of the nonnegative martingale property. Third, we show how the nonnegative martingale property will be modified in the presence of additional frictions and shed new light on incorporating different types of constraints documented in the literature.

JEL Classification: E21, D82, H21

Key Words: Intertemporal wedges, Nonnegative martingale, Inverse Euler equation, Immiseration

*Corresponding author, YiLi Chien. Email: yili.chien@stls.frb.org. We thank participants in the St. Louis Fed Research Seminar for comments and questions. We also benefited from comments on an earlier version of this paper by Marek Kapicka. The views expressed are those of the individual authors and do not necessarily reflect the official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

1 Introduction

This paper considers a dynamic economy in which privately informed idiosyncratic labor productivity shocks evolve stochastically over time. The stochastic nature of these productivity shocks accounts for a variety of uninsurable risks to the value of human capital over the course of people's lives. Assuming a standard neoclassical production technology, we characterize constrained efficient intertemporal wedges for the *macro aggregate* as well as the *micro individual* allocation of the economy.

We study a social planning problem in which the planner intends to provide an efficient insurance contract to each agent of the economy, subject to agents' incentive constraints caused by information frictions in a dynamic stochastic framework. This class of problems seeks the optimal trade-off between insurance and incentives and has been studied intensively in two related strands of the literature: Dynamic Contracting (DC) and New Dynamic Public Finance (NDPF).[1] Instead of the conventional recursive formulation in the literature, however, we use a multiplier approach that reformulates the social planning problem in a sequential fashion. Following Fernandes and Phelan (2000), we first decompose agents' lifetime incentive constraints into a sequence of temporary incentive constraints for all possible states at all times. We then use the multipliers on these sequential temporary constraints to construct two variables known as "Pareto-Negishi weights," one for consumption and the other for labor input. Both evolve over time to ensure the satisfaction of agents' incentive constraints and, at the same time, they determine the evolution of the welfare weights placed on an agent in the planner's objective. Under a fairly general assumption for the stochastic process of idiosyncratic labor productivity shocks, we show that the evolution of the Pareto-Negishi weight associated with agents' consumption is a *nonnegative martingale*. This property is built on the separability of period utility between consumption and labor, but it is robust to whatever the binding patterns of agents' incentive constraints may be. Using the nonnegative martingale property, we deliver three main contributions to the literature.

First, we show that several important results on the intertemporal wedge of the micro individual allocation in the DC/NDPF literature, including the inverse Euler equation, zero expected wealth

[1] For a review on DC, see Part V in Ljungqvist and Sargent (2012). For reviews on NDPF, see Golosov, Tsyvinski, and Werning (2007); Kocherlakota (2010); Golosov, Troshkin and Tsyvinski (2011); and Golosov and Tsyvinski (2015). Kocherlakota (2010, p. 3) explicitly noted that the optimal tax problem in NDPF is isomorphic to a dynamic contracting problem between a risk-neutral principal and a risk-averse agent who is privately informed about her type.

taxes and immiseration, can all be attributed to their origin in the nonnegative martingale property of consumption Pareto-Negishi weights. This provides a deep insight into why these results hold in the first place and how their underlying mechanisms work in an intuitive and novel perspective. We consider this a nice complement to the existing literature. Given the generality of the nonnegative martingale property, we essentially extend these findings from mainly endowment economies with i.i.d. shocks to neoclassical production economies with non-i.i.d. shocks.

Second, the focus of the DC/NDPF literature has been on the micro individual rather than the macro aggregate allocation. This is true despite most of the literature considering the scenario where "the social planner does all the saving in the economy by choosing the amount of aggregate savings." (Golosov, Troshkin, and Tsyvinski (2011, p. 157)). Moreover, it is important to bring micro individual behavior to bear on macro economic aggregates, since the economy-wide aggregate variables of an economy remain the central focus of macroeconomics. Atkeson and Lucas (1992, 1995) are exceptions. They envisioned an economy with many of the so-called "component planner," whose job is to transfer aggregate resources over time for the agents in her "jurisdiction" and, simultaneously, to meet these agents' incentive constraints in distributing the aggregate resources. An important departure from the literature in their analysis is that they explicitly made a distinction between the aggregate allocation and the distribution of the aggregate allocation for the component planner. In the same vein, we address the constrained efficient allocation over time in terms of aggregate as well as individual wedges. Chien, Cole, and Lustig (2011) adopted the multiplier approach and worked with an isoelastic utility over consumption in a positive analysis of asset pricing. They showed that the intertemporal prices for the macro aggregate of an economy depend on a specific moment of the distribution of accumulative multipliers. We obtain a similar result in our normative analysis: the constrained efficient intertemporal wedges for the macro aggregate of our dynamic Mirrleesian economy depend on a specific $1/\sigma$ moment of the Pareto-Negishi weights placed on consumption, where $1/\sigma$ denotes the intertemporal elasticity of substitution (IES). In particular, the constrained efficient intertemporal wedge for aggregate saving is negative or positive, depending on whether IES is elastic or inelastic. If there exists an asymptotic steady state for the economy, we show that $\beta R \gtreqless 1$ at constrained efficient allocations as IES $\gtreqless 1$, where β is the discount factor and R is the gross rate of return from saving. This finding is another application of the nonnegative martingale property of consumption Pareto-Negishi weights, since the constrained

efficient intertemporal wedge in the aggregate is shown to be exactly affected by the $1/\sigma$ moment of the consumption Pareto-Negishi weights.

Third, we extend the benchmark model to incorporate different or additional types of frictions/constraints documented in the DC/NDPF literature. Specifically, we consider three cases: (i) Atkeson and Lucas (1995), who put a lower bound on agents' utility; (ii) Phelan (2006) and Farhi and Werning (2007, 2010), who perceived social discounting in excess of private discounting; and (iii) Armenter and Mertens (2013), who made a verification technology on agents' states available to the social planner. In all of these cases, the nonnegative martingale in the benchmark model becomes a martingale and, at the same time, there exists a force that makes consumption Pareto-Negishi weights increase monotonically over time. These properties imply that immiseration in the benchmark model is no longer applicable, which is the common result in all three cases. We are also able to extend the Atkeson and Lucas (1995) result that $\beta R < 1$ in steady state at constrained efficient allocations to a more general environment for $\sigma \leq 1$. Integrating cases (i)-(iii) into a single framework and studying them in light of the modification of the nonnegative martingale sheds new light on these different frictions/constraints introduced in the DC/NDPF literature.

It should be noted from the beginning that the focus of this paper is the characterization rather than the implementation of constrained efficient allocations. Our focus has a good excuse: Kocherlakota (2010, chapter 4) has provided a recipe for implementing constrained efficient allocations of dynamic Mirrleesian economies as part of a competitive equilibrium.

Related literature

Our paper is related to several strands of the literature. We briefly discuss them below.

Espino, Kozlowski, and Sánchez (2015) studied the partnership interaction with private information and capital accumulation in a two-agent production economy. They found that the incentive of misreporting vanishes for the agent with dominating ownership and hence the immiseration result applies only to the dominated agent. Espino (2005) considered a private information economy with heterogeneous agents in terms of their discount factor and showed that the immiseration result is avoided if the economy does not collapse in the long run. In terms of quantitative studies, Veracierto (2014) considered a similar private information environment with aggregate uncertainty and assumed log utility functions in both consumption and leisure. His result indicated

that the idiosyncratic information friction has no effect on aggregate fluctuations quantitatively. The study is consistent with our finding that the intertemporal prices are not distorted with a unitary value of IES. Given labor supply, Farhi and Werning (2012) quantitatively explored the welfare effect of introducing optimal capital taxation prescribed by the inverse Euler equation.

Werning (2007) has formulated dynamic Mirrleesian taxation by encoding incentive constraints through multipliers on these constraints. However, there is de facto no variation of multipliers over time in the Werning framework. This is because he assumed that agents' heterogeneity in their characteristics remains invariant over time. Park (2014) extended the Werning model to allow for stochastic agent types but within a Ramsey rather than Mirrleesian framework.

The use of the multiplier approach to solving macroeconomic models was pioneered by Kehoe and Perri (2002), building on an earlier work by Marcet and Marimon (1999). Chien, Cole, and Lustig (2011) used stochastic accumulative multipliers to characterize allocations and aggregate prices, providing a tractable and computationally efficient algorithm for computing equilibria in the context of asset pricing. In repeated moral hazard environments, Mele (2014) demonstrated a more efficient computational algorithm with the recursive multiplier and with the first-order approach. In an environment similar to ours, Kapička (2013) adopts the first-order approach to successfully overcome the complexity caused by the persistence of shocks as well as the large numbers of shock values. Our approach differs because it characterizes intertemporal wedges and aggregate variables completely through a specific moment of the Pareto-Negishi weight distribution. Our method fits into the literature of replacing "primal" states variables with "dual" ones. Messner, Pavoni, and Sleet (2013) argued that the dual approach is applicable to a large class of economies and often leads to an easy characterization of constrained efficient allocations.

There are other papers related to our work. We comment on and make comparisons with them as we proceed.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 proves the nonnegative martingale property. Using the property, Section 4 characterizes constrained efficient intertemporal wedges for the micro individual allocation and Section 5 for the macro aggregate allocation. Section 6 considers extensions and Section 7 concludes.

2 Benchmark model

We consider a dynamic Mirrleesian economy in which privately informed idiosyncratic skill (labor productivity) shocks evolve stochastically over time. The stochastic nature of these skill shocks accounts for a variety of uninsurable risks to the value of human capital over the course of people's lives. The environment is similar to that in Kocherlakota (2010). For simplicity, our exposition focuses on the first-order Markov process for skill shocks throughout the paper. We show in Lemma 2 that our results remain valid for a fairly general stochastic process of skill shocks.

2.1 Economic environment

2.1.1 Preferences and skill shocks

Time is discrete, indexed by $t = 0, 1, ..., T$, where $T = \infty$ is allowed. The economy is populated by a continuum of infinitely-lived agents with a unit measure. The period utility of agents is given by[2]

$$u(c_t) - v(n_t),$$

where $u(.)$ is strictly concave and strictly increasing in consumption c_t, while $v(.)$ is strictly convex and strictly increasing in labor supply n_t. Both $u(.)$ and $v(.)$ are continuously differentiable. We impose standard Inada conditions to ensure that both consumption and labor supply are positive. In particular, we impose $\lim_{c \to 0} u'(c) = +\infty$.

Agents are ex ante identical but ex post heterogeneous. Ex post heterogeneity arises because agents have different histories with regard to their realization of idiosyncratic skill shocks over time. This is the only heterogeneity across agents. Realized skill shocks are the private information of agents and so are unknown to the social planner. An agent's skill shock at time t, denoted by θ_t, takes positive values in a finite set Θ and is realized at the beginning of the time period. We attach positive probabilities to all elements of Θ. The skills at $t = 0$ are drawn from a distribution over Θ,

[2] Additive separability between consumption and labor in the instantaneous utility is standard in NDPF; see, for example, Kocherlakota (2005); Golosov, Tsyvinski, and Werning (2007); Werning (2007); and Farhi and Werning (2012). It is a critical assumption for the derivation of the so-called "Inverse Euler Equation" in NDPF; see Kocherlakota (2010, p. 54). Most papers in DC consider endowment economies; see Ljungqvist and Sargent (2012, Part V).

and then θ_t follows a first-order Markov process independently and identically across agents at each point in time. We let $\theta^t = (\theta_0, \theta_1, ..., \theta_t)$ designate the history of events for an agent up through and until time t and let $\pi_t(\theta^t)$ denote its unconditional probability as of time 0. Because of the independence of skill shocks across agents, a law of large numbers applies so that the probability $\pi_t(\theta^t)$ also represents the fraction of the population that experiences θ^t at time t. We call an agent with history θ^t simply "the agent θ^t." For $s > t$, the probability of having θ^s conditional on the realization of θ^t is denoted by $\pi_s(\theta^s|\theta^t) = \prod_{j=t}^{s-1} \pi(\theta_{j+1}|\theta_j)$, where $\pi(\theta_{j+1}|\theta_j)$ is the conditional probability of θ_{j+1} given θ_j. We set the initial skill to 1 for all agents before they draw θ_0. There are no aggregate shocks.

Agents are assumed to maximize their expected discounted utility

$$\sum_{t=0}^{T} \sum_{\theta^t} \beta^t \pi_t(\theta^t) \left[u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) \right], \tag{1}$$

where $\beta \in (0, 1)$ denotes the discount factor and $l_t = \theta_t n_t$ is the effective labor supply at time t.

2.1.2 Production technology

The production technology is neoclassical, denoted by $F(K_t, L_t)$, where K_t and L_t are aggregate capital and aggregate effective labor, respectively. The resource constraints of the economy are given by

$$C_t + K_{t+1} \leq F(K_t, L_t) + (1 - \delta)K_t, \quad \forall t, \tag{2}$$

where C_t denotes aggregate consumption and δ is the depreciation rate for capital. The initial capital stock K_0 is exogenously given; all agents are initially endowed with K_0 units of capital at time $t = 0$. We exclude government expenditures to focus on the insurance issue.

2.2 Planning problem

We turn to the planning problem facing the society.

2.2.1 Temporary incentive-compatibility constraints

Although realized skill shocks and their history are the private information of agents, following the seminal work of Mirrlees (1971) and most of the DC/NDPF literature, effective labor supply $l_t = \theta_t n_t$ and consumption c_t (and so saving) are publicly observable. We consider a social planner (a benevolent government) that offers each agent an insurance contract with commitment.[3] Applying the Revelation Principle allows us to restrict attention to contracts with a direct mechanism whereby agents are induced to tell the truth about their skill types. Thus, each agent reports her skill types over time and receives an allocation as a function of this report so that the allocation is required to satisfy the incentive-compatibility (IC) constraints:

$$\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]\geq\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\widehat{\theta}^t)\right)-v\left(\frac{l_t(\widehat{\theta}^t)}{\theta_t}\right)\right],\forall\theta^t,\widehat{\theta}^t. \tag{3}$$

Kocherlakota (2010) shows that the set of allocations $\{c_t(\theta^t), l_t(\theta^t)\}$ achievable by the society are exactly the ones that satisfy the IC constraints (3) and the resource constraints (2).

The IC constraints above are in terms of lifetime utility as of $t = 0$. Fernandes and Phelan (2000) showed that as long as idiosyncratic shocks are discrete and the probability of each shock is strictly positive (both conditions are met in our setting), the lifetime IC constraints are equivalent to a sequence of temporary IC constraints, each of which rules out one-period deviations at a time. Therefore, the lifetime IC constraints (3) can be equivalently expressed as

$$\begin{aligned}
& u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)+\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi_s(\theta^s|\theta^t)\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right] \\
\geq\ & u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right)-v\left(\frac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right) \\
& +\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi_s(\theta^s|\theta^t)\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right)-v\left(\frac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right],\quad\forall t,\theta^t,\widehat{\theta}_t,
\end{aligned} \tag{4}$$

[3] As time passes, information about agent types (skills) may be revealed. A benevolent government (the social planner) is assumed to commit itself to the contract without exploiting information revelation as time passes. It is known that the society is better off with such a commitment than without it; see, for example, Laffont and Tirole (1988).

where $(\theta^s; \widehat{\theta}_t)$ denotes $(\theta_0, \theta_1, ..., \theta_{t-1}, \widehat{\theta}_t, \theta_{t+1}, ..., \theta_s)$.

The replacement of the lifetime IC constraints (3) by a sequence of temporal IC constraints (4) is crucial to our analysis, which will be clear later. To facilitate and simplify the analysis, we make the following assumption:

Assumption 1. *The disutility function of labor takes the form* $v(n) = \frac{n^\gamma}{\gamma}$*, where* $\gamma > 1$*.*

This is a popular functional form for $v(.)$, and the reciprocal of $\gamma - 1$ is the Frisch elasticity of labor supply. Lemma 2 later shows that we can dispense with this assumption for a key result of our paper.

2.2.2 Social planner's problem

The social planner chooses allocations $\{c_t(\theta^t), l_t(\theta^t)\}$ to maximize the objective (1), subject to the sequences of the temporal IC constraints (4) and the resource constraints of the economy (2).

Let $\{\beta^t \pi_t(\theta^t)\psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t)\}$ be the Lagrange multipliers on the IC constraints (4) and $\{\mu_t\}$ be the Lagrange multipliers on the resource constraints (2). After incorporating the multipliers on the constraints into the objective (1), the resulting Lagrangian can be rewritten as follows (see Appendix A.1 for the derivation):

$$\begin{aligned} L &= \min_{\{\psi,\mu\}} \max_{\{c,l,K\}} \sum_{t=0}^{T} \sum_{\theta^t} \beta^t \pi_t(\theta^t) \left[\phi_c(\theta^t) u\left(c_t(\theta^t)\right) - \phi_l(\theta^t) v\left(\frac{l_t(\theta^t)}{\theta_t}\right) \right] \\ &+ \sum_{t=0}^{T} \mu_t \left[F(K_t, L_t) + (1-\delta)K_t - C_t - K_{t+1} \right], \end{aligned} \tag{5}$$

where $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ are the Pareto-Negishi weights placed on the agent θ^t, and their evolutions are governed by

$$\phi_c(\theta^t) = 1 + \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \left[\psi(\theta^{s-1}, \theta_s, \widehat{\theta}_s) - \frac{\pi_t(\theta^t; \widehat{\theta}_s)}{\pi_t(\theta^t)} \psi(\theta^{s-1}, \widehat{\theta}_s, \theta_s) \right], \tag{6}$$

$$\begin{aligned}\phi_l(\theta^t) &= 1+\sum_{s=0}^{t}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\frac{\pi_t(\theta^t;\widehat{\theta}_s)}{\pi_t(\theta^t)}\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\\ &\quad -\sum_{\widehat{\theta}_t}\frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})}\omega(\theta_t,\widehat{\theta}_t)\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t), \qquad (7)\end{aligned}$$

where $\omega(\theta_t,\widehat{\theta}_t)=\left(\frac{\theta_t}{\widehat{\theta}_t}\right)^{\gamma}$.

Consider $\phi_c(\theta^t)$ in (6). Suppose that an agent's history is θ^{s-1} and her type realized at time s is θ_s. The multiplier $\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)\geq 0$ is associated with the temporal IC constraint that she mimics $\widehat{\theta}_s$, while the multiplier $\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\geq 0$ is associated with the temporal IC constraint that she is mimicked by $\widehat{\theta}_s$. Note that the multiplier $\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)$ enters (6) with a plus sign, whereas the multiplier $\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)$ enters with a minus sign. Thus, the assigned Pareto-Negishi weight $\phi_c(\theta^t)$ for an agent will increase over time if she is capable of mimicking others, whereas it will decrease over time if others are capable of mimicking her. These adjustments are to enforce IC constraints in response to the evolution of skill shocks over time. Multiplying both sides of (6) by $\pi_t(\theta^t)$, we see that $\pi_t(\theta^t)$ represents the fraction of the agent θ^t in the population at time t and $\pi_t(\theta^t;\widehat{\theta}_s)$ the fraction of agents who mimic the agent θ^t in the population at time t. It is clear that $\phi_c(\theta^t)=1$ would hold all the time if the friction of private information were absent. The evolution of $\phi_c(\theta^t)$ simultaneously determines the welfare weights that the planner places on an agent's utility $u(.)$ over time in her objective (5). A similar interpretation applies to $\phi_l(\theta^t)$ in (7).

As is clear from (6) and (7), the Pareto-Negishi weight at time t is a function of all past realized values of the multipliers $\psi(.)$ and so it summarizes and encodes the frequency and severeness of the binding of temporal IC constraints across states and over time. The summation and encoding is robust to whatever the binding patterns of agents' incentive constraints may be. It is worth mentioning that the evolution of $\{\phi_c(\theta^t),\phi_l(\theta^t)\}$ in (6) and (7) represents a generalization of Eq. (36) in Werning (2007), in that while agent types are permanently fixed in the Werning framework, they are stochastically evolving in our framework.

The evolution of the consumption Pareto-Negishi weight $\phi_c(\theta^t)$ is key to the results of this paper. We elaborate on it a bit more. Given the initial value $\phi_c(\theta^{-1})=1$, (6) can be expressed recursively as

$$\phi_c(\theta^t) = \phi_c(\theta^{t-1}) + \varepsilon(\theta^t), \tag{8}$$

where

$$\begin{aligned}\varepsilon(\theta^t) &= \sum_{\widehat{\theta}_t} \left[\psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) - \psi(\theta^{t-1}, \widehat{\theta}_t, \theta_t) \frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})} \right] \\ &\quad - \sum_{\widehat{\theta}_{t-1}} \frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})} \psi(\theta^{t-2}, \widehat{\theta}_{t-1}, \theta_{t-1}) \left[\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} - 1 \right]. \end{aligned} \tag{9}$$

The first summation in (9) is anticipated in view of (6). As to the second summation, its rise is somewhat subtle. Note first that if idiosyncratic shocks were to evolve as an i.i.d. process, $\pi(\theta_t|\widehat{\theta}_{t-1}) = \pi(\theta_t|\theta_{t-1})$ would hold and hence the second summation would vanish for sure. In our Markov environment, $\pi(\theta_t|\widehat{\theta}_{t-1}) \neq \pi(\theta_t|\theta_{t-1})$ in general and so this second summation is not zero in general. The likelihood ratio $\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})}$ in (9) then serves as a statistical reference. If $\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} < 1$, the ratio signals that θ_t is more likely to be linked to θ_{t-1} (true type at time $t-1$) than $\widehat{\theta}_{t-1}$ (other types at time $t-1$); as such, other things being equal, the agent θ^t is rewarded by the planner with a higher $\phi_c(\theta^t)$ according to (8). By contrast, if $\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} > 1$, the ratio signals that θ_t is more likely to br linked to $\widehat{\theta}_{t-1}$ than θ_{t-1}; as such, other things being equal, the agent θ^t is punished by the planner with a lower $\phi_c(\theta^t)$ according to (8). This type of rewards and punishments is ex post in nature and builds on the Markov link of idiosyncratic shocks. It invokes the planner's "threat keeping," a terminology used by Fernandes and Phelan (2000), to ensure that there are no gains at time t for any agent who deviated at time $t-1$.[4]

This completes the description of our model.

3 Nonnegative martingale

This section proves a powerful property of $\phi_c(\theta^t)$ over time, which is stated as follows:

[4]It has long been known from the agency problem that the principal uses the likelihood ratio of high against low output to infer whether agents shirk or not. Hart and Holmström (1987, p. 80) remarked: "The agency problem is not an inference problem in a strict statistical sense; conceptually, the principal is not inferring anything about the agent's action ..., because he already knows what action is being implemented. Yet, the optimal sharing rule reflects precisely the principles of inference." The essence of their remark equally applies here.

Lemma 1. *The Pareto-Negishi weight $\phi_c(\theta^t)$ is nonnegative, has a first moment equal to 1, and its evolution in (6) satisfies*

$$\sum_{\theta_t} \phi_c(\theta^t)\pi(\theta_t|\theta_{t-1}) = \phi_c(\theta^{t-1}). \tag{10}$$

That is, $\{\phi_c(\theta^t)\}$ is a nonnegative martingale.

Proof. See Appendix A.2. □

By the property of (10), the term $\varepsilon(\theta^t)$ in (8) is an innovation orthogonal to $\phi_c(\theta^{t-1})$. We now explain why $\varepsilon(\theta^t)$ is an innovation or, equivalently, why (10) holds.

Let us calculate $E_{t-1}(\varepsilon(\theta^t))$, that is, the expectation of $\varepsilon(\theta^t)$ conditional on information at time $t-1$:

$$\begin{aligned}\sum_{\theta_t} \varepsilon(\theta^t)\pi(\theta_t|\theta_{t-1}) &= \sum_{\theta_t}\sum_{\widehat{\theta}_t}\left[\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t) - \psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})}\right]\pi(\theta_t|\theta_{t-1}) \\ &\quad -\sum_{\theta_t}\sum_{\widehat{\theta}_{t-1}}\frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})}\psi(\theta^{t-2},\widehat{\theta}_{t-1},\theta_{t-1})\left[\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} - 1\right]\pi(\theta_t|\theta_{t-1}),\end{aligned}$$

which gives rise to

$$\begin{aligned}E_{t-1}(\varepsilon_t(\theta^t)) &= \sum_{\theta_t}\sum_{\widehat{\theta}_t}\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\pi(\theta_t|\theta_{t-1}) - \sum_{\theta_t}\sum_{\widehat{\theta}_t}\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\pi(\widehat{\theta}_t|\theta_{t-1}) \\ &\quad -\sum_{\widehat{\theta}_{t-1}}\frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})}\psi(\theta^{t-2},\widehat{\theta}_{t-1},\theta_{t-1})\sum_{\theta_t}\left[\pi(\theta_t|\widehat{\theta}_{t-1}) - \pi(\theta_t|\theta_{t-1})\right].\end{aligned}$$

First, note that the first and the second components of $E_{t-1}(\varepsilon_t)$ cancel each other out because we can interchange the indices θ_t and $\widehat{\theta}_t$. This result merely reflects the fact that when θ_t mimics $\widehat{\theta}_t$, it must be true that $\widehat{\theta}_t$ is mimicked by θ_t at the same time and that both the mimicking and the mimicked share the same multiplier $\psi(.)$. It should be noted that this symmetry between θ_t and $\widehat{\theta}_t$ holds for $u(.)$ but fails to hold for $v(.)$; see (4). As such, the separability of utility between consumption and labor is critical. Next, note that the last component of $E_{t-1}(\varepsilon_t)$ is equal to zero

as well because $\sum_{\theta_t}\left[\pi(\theta_t|\widehat{\theta}_{t-1})-\pi(\theta_t|\theta_{t-1})\right]=0$. As explained earlier, $\pi(\theta_t|\widehat{\theta}_{t-1})\neq\pi(\theta_t|\theta_{t-1})$ in general in our Markov-process environment. However, the term $\frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})}\psi(\theta^{t-2},\widehat{\theta}_{t-1},\theta_{t-1})$ in the last component of $E_{t-1}(\varepsilon_t)$, which is associated with the threat keeping at time t, is given from the viewpoint of time $t-1$. As a consequence, the ex post punishments and rewards associated with the threat keeping at time t average out.

From (6) and (7), we can express $\phi_l(\theta^t)$ as

$$\phi_l(\theta^t)=\phi_c(\theta^t)+\sum_{\widehat{\theta}_t}\frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})}\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\left(1-\omega(\theta_t,\widehat{\theta}_t)\right), \tag{11}$$

which describes the relationship between ϕ_c and ϕ_l. Since the second term of (11) is not zero in general, the labor Pareto-Negishi weight $\phi_l(\theta^t)$ fails to have a property similar to the consumption Pareto-Negishi weight $\phi_c(\theta^t)$ as described in Lemma 1.

Lemma 1 is proved under the imposition of Assumption 1 and the first-order Markov process for skill shocks. The next lemma shows that the nonnegative martingale property of $\phi_c(\theta^t)$ holds more generally.

Lemma 2. *Lemma 1 holds without the imposition of Assumption 1 and, moreover, it holds under the stochastic process that the conditional probability of θ_t is given by $\pi(\theta_t|x(\theta^{t-1});t)>0$ for all θ_t, where $x(\theta^{t-1})$ represents partial elements of $\theta^{t-1}=(\theta_0,\theta_1,...,\theta_{t-1})$.*

Proof. See Appendix A.3. □

The stochastic process described in 2 is more general than the first-order Markov process. For concreteness, let us consider three possible cases of idiosyncratic skill shocks for the specification of the conditional probability $\pi(\theta_t|.)$ described in the above lemma.

- Case 1: i.i.d. shocks, where $x(\theta^{t-1})=\varnothing$ and $\pi(\theta_t|x(\theta^{t-1});t)=\pi(\theta_t)$.
- Case 2: a first-order Markov process, where $x(\theta^{t-1})=\theta_{t-1}$ and $\pi(\theta_t|x(\theta^{t-1});t)=\pi(\theta_t|\theta_{t-1})$; this is the case we expose in the text.
- Case 3: a second-order Markov process, where $x(\theta^{t-1})=(\theta_{t-1},\theta_{t-2})$ and $\pi(\theta_t|x(\theta^{t-1});t)=\pi(\theta_t|\theta_{t-1},\theta_{t-2})$.

In all these cases, the results stated in Lemma 1 remain valid.

4 Intertemporal wedges in the individual

The Lagrangian (5) yields the first-order conditions[5]

$$\beta^t \phi_c(\theta^t) u'\left(c_t(\theta^t)\right) = \mu_t, \tag{12}$$

$$\beta^t \phi_l(\theta^t) v'\left(\frac{l_t(\theta^t)}{\theta_t}\right)\frac{1}{\theta_t} = \mu_t F_{Lt}, \tag{13}$$

$$\mu_t = \mu_{t+1}\left(F_{K,t+1} + 1 - \delta\right). \tag{14}$$

The optimal conditions (12)-(14) plus the resource constraints (2) characterize the constrained efficient individual allocation $\{c_t(\theta^t), l_t(\theta^t)\}$. The optimal conditions (12)-(14) show that the constrained efficient individual allocation depends on the Pareto weights $\{\phi_c, \phi_l\}$, which are determined by (6) and (7). If the friction of private information were absent, $\phi_c(\theta^t) = \phi_l(\theta^t) = 1$ would hold all the time and we would be back to the familiar first-order conditions.

Following Golosov, Tsyvinski, and Werning (2007), we define the inter- and the intra-temporal wedge as

$$\tau_{kt}(\theta^t) \equiv 1 - \frac{u'(c_t(\theta^t))}{\beta R_{t+1}\sum_{\theta_{t+1}} u'(c_{t+1}(\theta^{t+1}))\pi(\theta_{t+1}|\theta_t)}, \tag{15}$$

$$\tau_{lt}(\theta^t) \equiv 1 - \frac{v'\left(\frac{l_t(\theta^t)}{\theta_t}\right)}{\theta_t w_t u'(c_t(\theta^t))}, \tag{16}$$

where $w_t = F_{Lt}$ and $R_{t+1} = 1 + F_{K,t+1} - \delta$ are the wage rate and the gross rate of return from saving in competitive equilibrium. The wedges are defined in terms of the deviation between the marginal rate of substitution and the marginal rate of transformation. If there are no distortionary government interventions, $\tau_{kt}(\theta^t) = \tau_{lt}(\theta^t) = 0$ will hold in competitive equilibrium.

For the rest of the section, we use Lemma 1 plus the optimal conditions (12)-(14) to characterize

[5] Werning (2007, footnote 7) noted that the social maximization in the Mirrleesian economy is a convex problem. This can be seen after changes in variables in (3) from (c, l) to (u, v) to make the IC constraints become linear. As such, one can consider the social maximization in the Mirrleesian economy by simply using it to derive the first-order conditions. This is in contrast to the Ramsey problem, which is not convex in general; see Werning (2007, footnote 20).

constrained efficient intertemporal wedges facing individuals from several aspects. Given the non-martingale result of $\phi_l(\theta^t)$, we fail to derive useful properties to characterize constrained efficient intratemporal wedges. We thus concentrate our analysis on intertemporal wedges.

4.1 Inverse Euler equation

A central finding on capital taxation in NDPF is the so-called "inverse Euler equation" (IEE).[6] Here we show that there exists an intimate relationship between the nonnegative martingale of ϕ_c and the IEE.

Combining Eqs. (12) and (14) gives

$$\frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)} = \frac{\beta\left(F_{K,t+1}+1-\delta\right)}{u'\left(c_t(\theta^t)\right)}\frac{\phi_c(\theta^{t+1})}{\phi_c(\theta^t)}, \tag{17}$$

which in turn implies

$$\sum_{\theta_{t+1}} \frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)}\pi(\theta_{t+1}|\theta_t) = \frac{\beta\left(F_{K,t+1}+1-\delta\right)}{u'\left(c_t(\theta^t)\right)}\sum_{\theta_{t+1}}\frac{\phi_c(\theta^{t+1})}{\phi_c(\theta^t)}\pi(\theta_{t+1}|\theta_t).$$

By Lemma 1, $\sum_{\theta_{t+1}} \frac{\phi_c(\theta^{t+1})}{\phi_c(\theta^t)}\pi(\theta_{t+1}|\theta_t) = 1$. The above equation then leads to the IEE:

$$\sum_{\theta_{t+1}} \frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)}\pi(\theta_{t+1}|\theta_t) = \frac{\beta\left(F_{K,t+1}+1-\delta\right)}{u'\left(c_t(\theta^t)\right)}. \tag{18}$$

The IEE result was originally derived by Diamond and Mirrlees (1978) and Rogerson (1985). Golosov, Kocherlakota, and Tsyvinski (2003) extended the result to a general class of models and, in particular, a general law of motion for skill shocks. The IEE plus Jensen's inequality implies a positive capital wedge. Kocherlakota (2010, pp. 56-57) gave an explanation for why it is socially optimal: (i) starting from an Euler equation, "the planner is able to provide better risk-sharing by front-loading consumption" and (ii) moving to the IEE, the planner "pays the second-order cost of reducing smoothing to get the first-order benefit of improving insurance."

The standard way of deriving the IEE is by means of the so-called "perturbation method" – a small perturbation around the optimality.[7] This method is powerful, but it may be blind about

[6] On IEE, see Golosov, Tsyvinski, and Werning (2007); Kocherlakota (2010); Farhi and Werning (2012); and Golosov and Tsyvinski (2015).

[7] See Rogerson (1985); Golosov, Kocherlakota, and Tsyvinski (2003); Golosov, Tsyvinski, and Werning (2007);

the underlying reason for the result. We instead rely on Lemma 1, which makes clear its origin.[8] If $\phi_c(\theta^t) = 1$ all the time (i.e., there is no information friction), we see from (17) that EE rather than IEE would hold. In this sense, we state:

Proposition 1. *IEE is built on the nonnegative martingale of* $\{\phi_c(\theta^t)\}$.

The result of a positive capital wedge can be derived directly from the nonnegative martingale of $\{\phi_c(\theta^t)\}$. Let us denote the intertemporal individual wedge evaluated at constrained efficient allocations by $\tau^*_{kt}(\theta^t)$. Using the optimal conditions (12) and (13), (15) gives

$$\frac{1}{1-\tau^*_{kt}(\theta^t)} = \sum_{\theta_{t+1}} \frac{\phi_c(\theta^t)}{\phi_c(\theta^{t+1})}\pi(\theta^{t+1}|\theta^t). \tag{19}$$

Applying Jensen's inequality to the right-hand side of the above equation and then using the nonnegative martingale property of ϕ_c, we have

$$\sum_{\theta_{t+1}} \frac{\phi_c(\theta^t)}{\phi_c(\theta^{t+1})}\pi(\theta_{t+1}|\theta_t) > \frac{\phi_c(\theta^t)}{\sum_{\theta_{t+1}} \phi_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t)} = 1,$$

which implies $\tau^*_{kt}(\theta^t) > 0$ from (19).

4.2 Zero expected wealth taxes

Kocherlakota (2005) considered wealth taxes instead of capital taxes to implement constrained efficient allocations.[9] The imposition of the wealth tax proposed by Kocherlakota (2005) gives rise to

$$1 = \beta \frac{u'\left(c_t(\theta^{t+1})\right)}{u'\left(c_t(\theta^t)\right)}(1-\tau^*_{W,t+1}(\theta^{t+1}))\left(F_{K,t+1}+1-\delta\right),$$

and Kocherlakota (2010).

[8] In an i.i.d. endowment economy with private information, Thomas and Worrall (1990) showed that the planner's marginal cost of delivering one more unit of promised value to agents is a martingale. Farhi and Werning (2013) extended this result to a labor production economy in which idiosyncratic productivity evolves as a Markov process and is private information. They showed that the result implies the IEE. Intuitively, the planner's marginal cost of the delivery has to do with $1/u'(.)$ and the planner's desire to smooth the marginal cost of utility production over time implies the IEE; see Section 5.3.3 for more details. It should be noted that there is no capital but a linear storage technology in Farhi and Werning (2013) and the focus of the paper is on intratemporal labor wedges.

[9] Kocherlakota (2005) allowed aggregate as well as idiosyncratic shocks. We consider a version without aggregate shocks.

where $\tau^*_{W,t+1}(\theta^{t+1}))$ is the optimal wealth tax rate. This equation together with the optimal conditions (12) and (14) yields

$$1 - \tau^*_{W,t+1}(\theta^{t+1}) = \frac{\phi_c(\theta^{t+1})}{\phi_c(\theta^t)}, \tag{20}$$

which shows that the wealth tax proposed by Kocherlakota (2005) is to simply track the evolution of consumption Pareto-Negishi weights in two adjacent time periods. According to (20), $\tau^*_{W,t+1}(\theta^{t+1}) < 0$ if $\phi_c(\theta^{t+1}) > \phi_c(\theta^t)$ while $\tau^*_{W,t+1}(\theta^{t+1}) > 0$ if $\phi_c(\theta^{t+1}) < \phi_c(\theta^t)$. The optimal wealth tax thus prescribes a subsidy for agents who receive high shocks and mimic others at time $t+1$ but a tax for agents who receive low shocks and are mimicked by others at time $t+1$. Kocherlakota (2005) emphasized the importance of this wealth tax/subsidy characteristic in implementation, in that it makes savings a bad hedge against idiosyncratic shocks and so deters the joint deviation of saving too much and supplying labor too little. Taking the conditional expectation on both sides of (20) and using the nonnegative martingale of $\{\phi_c(\theta^t)\}$ leads to $E_t(\tau^*_{W,t+1}(\theta^{t+1})) = 0$, which is a main result of Kocherlakota (2005): zero expected wealth taxes at the optimum. If $\phi_c(\theta^t) = 1$ all the time (i.e., there is no information friction), we see from (20) that $\tau^*_{W,t+1}(\theta^{t+1}) = 0$ would hold all the time. In this sense, we state:

Proposition 2. *The result of zero expected wealth taxes at constrained efficient allocations is built on the nonnegative martingale of* $\{\phi_c(\theta^t)\}$.

4.3 Immiseration

A celebrated result in the DC/NDPF literature is the so-called "immiseration," in which almost every agent's consumption converges to zero in the limit at constrained efficient allocations even though aggregate consumption remains high above zero. This result implies: (i) an increasingly large fraction of agents consume a decreasingly small amount of consumption over time and their individual consumption goes to zero in the limit, and (ii) a decreasingly small fraction of agents consume an increasingly large amount of consumption over time and their individual consumption goes to infinity in the limit.[10] Thomas and Worrall (1990) and Atkeson and Lucas (1992) obtained the "immiseration" result in i.i.d. and endowment economies. Using Lemmas 1 and 2, we demon-

[10] Kocherlakota (2010, p. 68) provided an example to illustrate the co-existence of (i) and (ii). His Example 1 (p. 64) puts a lower and an upper bound on consumption to stop the force of (i) and of (ii).

strate that "immiseration" still holds under Markov or more general stochastic processes and in a neoclassical production economy.

Proposition 3. *For almost surely all* θ^t, $\lim_{t\to\infty}\phi_c(\theta^t)=0$ *at constrained efficient allocations.*

Proof. $\{\phi_c(\theta^t)\}$ is a nonnegative martingale according to Lemma 1 and thus we can apply Doob's (1953) martingale convergence theorem: for a sample path θ^t of skill realizations, there exists a finite $\phi^*(\theta^t)$ such that $\phi_c(\theta^t)$ converges along the sample path to $\phi^*(\theta^t)$ and this convergence is true for almost surely all sample paths; see Kocherlakota (2010, p. 62). Since $\lim_{t\to\infty}\phi_c(\theta^t)=\lim_{t\to\infty}\phi_c(\theta^{t-1})=\phi^*(\theta^t)$, it is implied from (8) that $\lim_{t\to\infty}\varepsilon(\theta^t)=0$. This in turn implies that almost surely all multipliers $\psi(.)$ in (9) are equal to zero in the limit.

Consider a generic sample path θ^t after a long passage of time with $\phi_c(\theta^t)=\phi^*(\theta^t)$. Suppose, to the contrary, $\phi^*(\theta^t)$ is non-zero. We show below that this leads to a contradiction.

The convergence of $\phi_c(\theta^t)$ implies two things: (i) $\phi_c(\theta^t)$ will stay and remain constant with its value equal to $\phi^*(\theta^t)$; (ii) $\phi_l(\theta^t)=\phi^*(\theta^t)$ too because the multipliers $\psi(.)$ in (11) are all equal to zero. Note that skill shocks are still stochastically evolving after time t. Now consider two agents who share the same history θ^t up to time t but draw two skill shocks θ_{t+1} and $\widehat{\theta}_{t+1}$ with $\theta_{t+1}\neq\widehat{\theta}_{t+1}$ at time $t+1$. From Eqs. (21)-(22) derived later, we have $c_{t+1}(\theta^{t+1})=c_{t+1}(\theta^t,\widehat{\theta}_{t+1})=\frac{\phi^*(\theta^{t+1})^{\frac{1}{\sigma}}}{H_{t+1}}C_{t+1}$; moreover, for all $s>t+1$ conditional on θ^{t+1}, $c_s(\theta^s)=c_s(\theta^s;\widehat{\theta}_{t+1})=\frac{\phi^*(\theta^{t+1})^{\frac{1}{\sigma}}}{H_s}C_s$; $l_s(\theta^s)=l_s(\theta^s;\widehat{\theta}_{t+1})=\frac{\phi^*(\theta^{t+1})^{\frac{1}{1-\gamma}}\theta_s^{\frac{-\gamma}{1-\gamma}}}{J_s}L_s$.

Applying these results, the temporary IC constraints (4) would reduce to

$$\begin{aligned}
-v\left(\frac{l_{t+1}(\theta^t,\theta_{t+1})}{\theta_{t+1}}\right) &\geq -v\left(\frac{l_{t+1}(\theta^t,\widehat{\theta}_{t+1})}{\theta_{t+1}}\right),\ \forall\theta^{t+1},\widehat{\theta}_{t+1};\\
-v\left(\frac{l_{t+1}(\theta^t,\widehat{\theta}_{t+1})}{\widehat{\theta}_{t+1}}\right) &\geq -v\left(\frac{l_{t+1}(\theta^t,\theta_{t+1})}{\widehat{\theta}_{t+1}}\right),\ \forall\theta^{t+1},\widehat{\theta}_{t+1}.
\end{aligned}$$

However, with $\phi_l(\theta^t)=\phi^*(\theta^t)$, (22) implies that $\frac{l_{t+1}(\theta^t,\theta_{t+1})}{l_{t+1}(\theta^t,\widehat{\theta}_{t+1})}=\left(\frac{\theta_{t+1}}{\widehat{\theta}_{t+1}}\right)^{\frac{-\gamma}{1-\gamma}}$, which in turn implies that the reduced temporary IC constraints above must be violated with $\theta_{t+1}\neq\widehat{\theta}_{t+1}$. As such, we have a contradiction and so $\phi^*(\theta^t)$ must be zero almost surely in the limit. □

From (12), $\lim_{t\to\infty}\phi_c(\theta^t)=0$ implies that $\lim_{c\to 0}u'(c_t(\theta^t))=+\infty$. The intuition underlying "immiseration" is as follows. Along a sample path θ^t, $\phi_c(\theta^t)$ is converging as time has been

long passed. However, this does not mean that θ^t stops evolving; on the contrary, θ^t still evolves stochastically over time even if $\phi_c(\theta^t)$ is converging toward $\phi^*(\theta^t)$. To provide incentives for more productive people to produce more, the planner needs to spread out consumption across heterogeneous agents over time even though $\phi_c(\theta^t)$ is converging. The planner accomplishes this task by "immiseration": let the consumption of an increasingly large fraction of agents go to zero in the limit so that, due to the Inada condition $\lim_{c\to 0} u'(c) = +\infty$, even a tiny spreading of consumption can be a significant differentiation for the agents.

The immiseration result depends heavily on the nonnegative martingale property stated in Lemma 1. Given that the nonnegative martingale of ϕ_c is robust to the more general shock process as shown in Lemma 2, the immiseration result holds more generally.

4.4 Permanent intertemporal distortions

The result of $\tau^*_{kt}(\theta^t) > 0$ for all θ^t found in NDPF is often in contrast with the classical result of Chamley (1986) and Judd (1985) that capital should go untaxed in the long run.[11] However, it is clear from the IEE (18) that, unlike the classical result of Chamley (1986) and Judd (1985), the implied $\tau^*_{kt}(\theta^t) > 0$ for all θ^t is not about the long-run property of capital taxation in the limit. To make an informative comparison with the Chamley-Judd result, one must know what happens to $\tau^*_{kt}(\theta^t)$ as $t \to \infty$.

Albanesi and Armenter (2012) studied the long-run properties of intertemporal distortions in a broad class of second-best economies. They identified a sufficient condition that rules out permanent intertemporal distortions in the second-best, that is, intertemporal individual distortions, even if desirable in the short run, will no longer be desirable in the long run if the condition is met. However, as Albanesi and Armenter (2012) acknowledged, their proposed sufficient condition does not hold in general in economies with private information. As such, their analysis leaves open the question of whether intertemporal distortions are permanent or temporary in dynamic Mirrleesian economies. In terms of our framework, the question to ask is whether $\tau^*_{kt}(\theta^t)$ will converge to zero in the limit. The following result provides the answer.

Proposition 4. $\lim_{t\to\infty} \tau^*_{kt}(\theta^t) > 0$ *for almost surely every* θ^t.

[11] A recent paper by Straub and Werning (2014) argued that the usual intuition that an indefinite positive tax on capital is equivalent to an ever-increasing tax on consumption is invalid and that the prescription of no capital taxation in the long run in the Ramsey framework may not be true.

Proof. Proposition 3 demonstrates the immiseration result that $\phi_c(\theta^t)$ converges almost surely to 0. This excludes the situation where $\lim_{t\to\infty} \frac{\phi_c(\theta^t)}{\phi_c(\theta^{t+1})} = 1$. From equation (19), the remaining possibility is that $\lim_{t\to\infty} \tau^*_{kt}(\theta^t) > 0$ for almost every θ^t. □

We conclude that constrained efficient intertemporal individual distortions are permanent rather than temporary in dynamic Mirrleesian economies. This result, to our knowledge, is new to the literature.

Kocherlakota (2010, pp. 115-117) explored the long-run property of his proposed wealth taxes, but failed to pin down the sign of the optimal tax rate as $c_t(\theta^t) \to 0$ (immiseration). He (p. 117) simply suggested: "it may well be positive." Utilizing our derived formula (20), we can say more. Due to immiseration, both the numerator $\phi_c(\theta^{t+1})$ and the denominator $\phi_c(\theta^t)$ of (20) converge to zero almost surely in the long run. This feature immediately leads to:

Proposition 5. $\lim_{t\to\infty} \tau^*_{Wt}(\theta^t) \neq 0$ *for almost surely every* θ^t.

From (6), we see that $\{\phi_c(\theta^t)\}$ is not a monotonic sequence. As such, $\lim_{t\to\infty} \tau^*_{Wt}(\theta^t)$ according to (20) may well be negative as well as positive.

5 Intertemporal wedges in the aggregate

The focus of the DC/NDPF literature has been on the micro individual allocation and, therefore, it is not surprising that relatively little is known about the general property of macro aggregate variables at constrained efficient allocations. This is true despite most of the literature considering the scenario where "the social planner does all the saving in the economy by choosing the amount of aggregate savings." (Golosov, Troshkin and Tsyvinski, 2011, p. 157). Moreover, even if microfounded, the economy-wide aggregate variables of an economy remain the central focus of macroeconomics. It is thus important to bring micro individual behavior to bear on macro economic aggregates. Within our context, it is interesting to know how the social planner allocates the macro aggregate resources over time when providing agents with insurance and dealing with their incentive constraints at the micro individual level. This section intends to shed light on the question.

5.1 Consumption and labor sharing rules

To proceed, we make a further assumption:

Assumption 2. *The consumption utility function takes the CRRA form:*

$$u(c_t) = \frac{c_t^{1-\sigma}}{1-\sigma},$$

where $\sigma > 0$, ensuring that $u(.)$ is strictly concave.

The reciprocal of σ represents the intertemporal elasticity of substitution (IES). As will be seen, whether IES is elastic or inelastic is of vital importance to the determination of aggregate constrained efficient intertemporal wedges. Whenever addressing aggregate wedges, we always impose Assumptions 1 and 2. It is worth noting that imposing Assumptions 1 and 2 is highly popular in many quantitative studies; see, for example, Farhi and Werning (2013); Golosov, Troshkin, and Tsyvinski (2016).

Under Assumptions 1 and 2, we obtain the following from (12) and (13):

$$\frac{c_t(\theta^t)}{C_t} = \frac{\phi_c(\theta^t)^{\frac{1}{\sigma}}}{H_t}, \tag{21}$$

$$\frac{l_t(\theta^t)}{L_t} = \frac{\phi_l(\theta^t)^{\frac{1}{1-\gamma}}\theta_t^{\frac{-\gamma}{1-\gamma}}}{J_t}, \tag{22}$$

where

$$\begin{aligned} H_t &\equiv \sum_{\theta^t} \phi_c(\theta^t)^{\frac{1}{\sigma}} \pi_t(\theta^t), \\ J_t &\equiv \sum_{\theta^t} \phi_l(\theta^t)^{\frac{1}{1-\gamma}} \theta_t^{\frac{-\gamma}{1-\gamma}} \pi_t(\theta^t). \end{aligned}$$

Eqs. (21) and (22) are the consumption and labor sharing rules: the amount of consumption and of effective labor destined for the agent θ^t will be completely determined, once the aggregates (C_t, L_t) and the weights $(\phi_c(\theta^t), \phi_l(\theta^t))$ are known.

The result that individual allocations can be solved as a function of aggregate allocations and Pareto-Negishi weights has been shown by Werning (2007) and Park (2014) in the Ramsey

framework. In both papers the authors noted that a key to the result lies in the levying of *linear* taxes so that marginal rates of substitution are equated with each other across agents. We show here that the result can hold even with the levying of *nonlinear* taxes in Mirrleesian economies. An important difference is that while Werning (2007) and Park (2014) restricted $\phi_c(\theta^t) = \phi_l(\theta^t)$ in their settings, we do not.

5.2 Constrained efficient aggregate allocations

Substituting the derived c_t and l_t of the consumption and labor sharing rules (21)-(22) into the individual optimal conditions, (12)-(14), gives rise to

$$C_t^{-\sigma} = \beta C_{t+1}^{-\sigma} \left(\frac{H_{t+1}}{H_t} \right)^{\sigma} (F_{K,t+1} + 1 - \delta), \tag{23}$$

$$\frac{L_t^{\gamma-1}}{C_t^{-\sigma}} = \left(\frac{H_t^{\sigma}}{J_t^{1-\gamma}} \right) F_{Lt}. \tag{24}$$

The constrained efficient macro aggregate allocation is characterized by (23) and (24) together with the resource constraints (2). Contrasting (23)-(24) with (12)-(13) clearly shows that while the micro individual allocation (c_t, l_t) is dictated by the individual Pareto-Negishi weights ϕ_c and ϕ_l, the macro aggregate allocation (C_t, L_t) is dictated by some aggregations of the individual Pareto-Negishi weights, that is, H_t and J_t. Since we fail to derive useful properties for J_t given the non-martingale result of ϕ_l, we concentrate our analysis on the intertemporal margin associated with (23).

How is the thrust of (23)? First, like the IEE, it characterizes the social planner's behavior over time in the presence of information frictions. The difference is that while the IEE describes the social planner's intertemporal behavior over the micro individual allocation, (23) describes her intertemporal behavior over the macro aggregate allocation. Second, there does not exist a representative agent for our friction-loaded economy in competitive equilibrium. It is thus not so meaningful to define the intertemporal wedge for the macro aggregate allocation analogous to that of (15) for the micro individual allocation. As an alternative, we use the first-best (a frictionless or complete markets) economy as the benchmark, in which there does exist a representative agent for the economy. In the first-best economy without frictions, it is straightforward to see that

$H_{t+1}/H_t = 1$ for all t. We can define the aggregate wedge associated with (23) by its deviation from the first-best:

$$\text{Intertemporal aggregate wedge}: \tau_{Kt} \equiv 1 - \left(\frac{H_{t+1}}{H_t}\right)^{\sigma}. \tag{25}$$

For the rest of this section, we study the constrained efficient intertemporal aggregate allocation characterized by (23) through the lens of the aggregate wedge defined by (25).

5.3 Properties of intertemporal aggregate wedges

This subsection characterizes the properties of the constrained efficient intertemporal aggregate wedge, τ^*_{Kt}. Issues involved include the short run (at any finite time t) and the long run (as $t \to \infty$). Lemma 1 plays a critical role for the results derived below.

5.3.1 The short run

It is clear from the definition of the intertemporal aggregate wedge (25) that the ratio H_{t+1}/H_t is crucial. The next lemma characterizes this ratio.

Lemma 3. *$H_t \gtreqless H_{t+1}$ for $t < \infty$ if $\sigma \gtreqless 1$.*

Proof. From Lemma 1, we have

$$\phi_c(\theta^{t-1})^{\frac{1}{\sigma}} = \left(\sum_{\theta_t} \phi_c(\theta^t)\pi(\theta_t|\theta_{t-1})\right)^{\frac{1}{\sigma}}.$$

By the power mean inequality, we have[12]

$$\left(\sum_{\theta_t} \phi_c(\theta^t)\pi(\theta_t|\theta_{t-1})\right)^{\frac{1}{\sigma}} \gtreqless \sum_{\theta_t} \phi_c(\theta^t)^{\frac{1}{\sigma}}\pi(\theta_t|\theta_{t-1}) \text{ if } \sigma \gtreqless 1.$$

Combining the above results leads to

[12] For $x_i \neq x_j$, the inequality is read as

$$\left(\sum_i x_i^p w_i\right)^{\frac{1}{p}} > \left(\sum_i x_i^q w_i\right)^{\frac{1}{q}} \text{ if } p > q, \text{where } w_i > 0 \text{ and } \sum_i w_i = 1.$$

Readers are referred to Bullen (2013) for the details.

$$\phi_c(\theta^{t-1})^{\frac{1}{\sigma}} \gtreqless \sum_{\theta_t} \phi_c(\theta^t)^{\frac{1}{\sigma}} \pi(\theta_t|\theta_{t-1}) \text{ if } \sigma \gtreqless 1,$$

which in turn gives

$$\sum_{\theta^{t-1}} \phi_c(\theta^{t-1})^{\frac{1}{\sigma}} \pi_{t-1}(\theta^{t-1}) \gtreqless \sum_{\theta^t} \phi_c(\theta^t)^{\frac{1}{\sigma}} \pi_t(\theta^t) \text{ if } \sigma \gtreqless 1.$$

□

The above lemma leads to three possibilities:

1. $\sigma = 1$: When $\sigma = 1$, $H_{t+1} = H_t$ holds, which implies at constrained efficient allocations that there is no intertemporal aggregate wedge at time t and so $\tau^*_{Kt} = 0$.

2. $\sigma > 1$: When $\sigma > 1$, $H_{t+1} < H_t$ holds, which implies at constrained efficient allocations that there is a positive intertemporal aggregate wedge at time t and so $\tau^*_{Kt} > 0$.

3. $\sigma < 1$: When $\sigma < 1$, $H_{t+1} > H_t$ holds, which implies at constrained efficient allocations that there is a negative intertemporal aggregate wedge at time t and so $\tau^*_{Kt} < 0$.

To sum up, we state:

Proposition 6. *$\tau^*_{Kt} \gtreqless 0$ for $t < \infty$ if $\sigma \gtreqless 1$.*

The quantity of the intertemporal elasticity of substitution (IES), i.e., the value of $1/\sigma$, is of great importance in various fields in economics: for example, labor supply (Keane, 2011), consumption and saving (Attanasio and Weber, 2010), asset pricing (Ljungqvist and Sargent, 2012, chapters 13-14), and investment and business cycles (Favilukis and Lin, 2013). However, estimations of its value vary. While works including those of Hansen and Singleton (1982), Attanasio and Weber (1989), and Vissing-Jørgensen and Attanasio (2003) find IES to be well above 1, other works including those of Hall (1988), Campbell and Mankiw (1989) and Campbell (1999) find IES to be close to zero. Proposition 6 prescribes $\tau^*_{Kt} \gtrless 0$ according to whether IES is inelastic or elastic. When IES is unitary, it prescribes $\tau^*_{Kt} = 0$ all the time.

5.3.2 The long run

Proposition 6 prescribes that $\tau^*_{Kt} = 0$ if $\sigma = 1$, but $\tau^*_{Kt} \neq 0$ if $\sigma \neq 1$. This is a short-run property, which concerns what happens for any finite time t. But what about the long run or in the limit, that is, $t \to \infty$? This question is important because if H_t asymptotically converges to a positive number, then at the optimum there will be no intertemporal aggregate distortion in the long run, even though there is in the short run. In the following, we demonstrate that Proposition 6 remains valid in the long run. The key underpinning the result is the immiseration property.

It is clear from Lemma 3 that if $\sigma = 1$, then $H_{t+1} = H_t = 1$ must be true in the long run as well. Thus, our long-run focus is on $\sigma \neq 1$. We show in Section 4.3 that immiseration must arise in our economy. It prescribes that a diminishing fraction of agents receives an ever-increasing fraction of the resources. Let $\tilde{\pi}(t)$ be the fraction of those agents and let $\tilde{\phi}_c(t)$ be the average Pareto-Negishi weight associated with them. Immiseration implies that as $t \to \infty$, $\tilde{\pi}(t) \to 0$ but $\tilde{\phi}_c(t) \to \infty$. However, by Lemma 1, $\sum_{\theta^t} \phi_c(\theta^t)\pi_t(\theta^t) = 1$. So, we have

$$\lim_{t\to\infty} \sum_{\theta^t} \phi_c(\theta^t)\pi_t(\theta^t) = \lim_{t\to\infty} \tilde{\phi}_c(t)\tilde{\pi}(t) = 1, \tag{26}$$

where the first equality holds because $\tilde{\phi}_c(t) \to \infty$ implies that agents other than the fraction $\tilde{\pi}(t)$ all receive $\phi_c(\theta^t) = 0$ in the limit. Thus,

$$\lim_{t\to\infty} H_t = \lim_{t\to\infty} \sum_{\theta^t} \phi_c(\theta^t)^{\frac{1}{\sigma}}\pi_t(\theta^t) = \lim_{t\to\infty} \tilde{\phi}_c(t)^{\frac{1}{\sigma}}\tilde{\pi}(t) = \lim_{t\to\infty} \tilde{\phi}_c(t)^{\frac{1}{\sigma}}\tilde{\phi}_c(t)^{-1} = \lim_{t\to\infty} \tilde{\phi}_c(t)^{\frac{1-\sigma}{\sigma}}, \tag{27}$$

where the third equality uses the implication of (26) that $\tilde{\pi}(t) \to \tilde{\phi}_c(t)^{-1}$ in the limit. With $\lim_{t\to\infty} H_t = \lim_{t\to\infty} \tilde{\phi}_c(t)^{\frac{1-\sigma}{\sigma}}$ from (27) and that $\tilde{\phi}_c(t) \to \infty$ as $t \to \infty$, we obtain:

Lemma 4. *As $t \to \infty$, (i) $H_t \to 0$ if $\sigma > 1$; (ii) $H_t \to \infty$ if $\sigma < 1$.*

Lemma 4 directly implies that $H_{t+1}/H_t < 1$ for all t if $\sigma > 1$ and $H_{t+1}/H_t > 1$ for all t if $\sigma < 1$. Based on this Lemma and the definition of the intertemporal aggregate wedge, we then have:

Proposition 7. *As $t \to \infty$, (i) $\lim_{t\to\infty} \tau^*_{Kt} > 0$ if $\sigma > 1$; (ii) $\lim_{t\to\infty} \tau^*_{Kt} < 0$ if $\sigma < 1$.*

This proposition dictates that the short-run result stated in Proposition 6 remains intact even in the long run as $t \to \infty$.

5.3.3 Intuitions

Both Propositions 6 and 7 build on Lemma 3, which in turn builds on Lemma 1. The proof of Lemma 3 is rather technical and not particularly intuitive. Here we provide some intuitions for what we have found.

First, let us explain (23). Kocherlakota (2010, p. 49) provided an intuition for the IEE (18) in the following way. Suppose $\rho(u)$ is the resource cost of producing u units of utility. Then the social planner's desire to smooth the marginal cost of utility production over time implies

$$\rho'(u(c_t(\theta^t))) = \frac{1}{\beta\left(F_{K,t+1} + 1 - \delta\right)} \sum_{\theta_{t+1}} \rho'(u(c_{t+1}(\theta^{t+1})))\pi(\theta_{t+1}|\theta_t),$$

which leads to the IEE (18) since $\rho'(u) = 1/u'\left(\rho(u)\right)$.

We can understand (23) in a similar way. Under Assumption 2, the marginal cost of utility production for the agent θ^t at time t is equal to $1/u'\left(c_t(\theta^t)\right) = c_t(\theta^t)^\sigma$. Thus, the marginal cost of utility production for all agents at time t is given by $\sum_{\theta_t} c_t(\theta^t)^\sigma \pi_t(\theta_t)$. From the consumption sharing rule (21), $c_t(\theta^t) = \frac{\phi_c(\theta^t)^{\frac{1}{\sigma}}}{H_t} C_t$ and so we have

$$\sum_{\theta_t} c_t(\theta^t)^\sigma \pi_t(\theta_t) = C_t^\sigma H_t^{-\sigma} \sum_{\theta_t} \phi_c(\theta^t)\pi_t(\theta_t) = C_t^\sigma H_t^{-\sigma},$$

where the last equality has used Lemma 1. The social planner's desire to smooth the marginal cost of utility production for all agents over time then implies (23).

Second, we explain Lemma 3 by more intuitive arguments. By the definition of H_t, we have $H_t = \sum_{\theta^t} \phi_c(\theta^t)^{\frac{1}{\sigma}} \pi_t(\theta^t)$. Note that H represents an expected utility if we view $\phi_c{}^{\frac{1}{\sigma}}$ as a utility function. In this way, we then have risk aversion ($\sigma > 1$), risk loving ($\sigma < 1$) and risk neutrality ($\sigma = 1$), respectively, in view of $\phi_c^{\frac{1}{\sigma}}$ being a function exhibiting concavity ($\sigma > 1$), convexity ($\sigma < 1$) and linearity ($\sigma = 1$), respectively.

Next, from Lemma 1, we have $E(\phi_c(\theta^t)) = 1$ for all t; from (8), we have $Var(\phi_c(\theta^{t+1})) > Var(\phi_c(\theta^t))$ for all t. Thus, while $\phi_c(\theta^t)$ has a mean equal to unity all the time, its variance increases over time.

Putting together the above arguments leads to the result that the "expected utility" H_t is decreasing ($\sigma > 1$), increasing ($\sigma < 1$), and remains unchanged ($\sigma = 1$) as time passes. That is, $H_t \gtreqless H_{t+1}$ if $\sigma \gtreqless 1$. By (25), we then have $\tau_{Kt}^* \gtreqless 0$ if $\sigma \gtreqless 1$.

5.4 Steady-state βR

This subsection asks the question: what will βR be at constrained efficient allocations if there is an asymptotic steady state in which C_t and K_t/L_t are constant over time?

If frictions were absent, $H_t = 1\ \forall t$ would hold and so (23) would reduce to

$$C_t^{-\sigma} = \beta C_{t+1}^{-\sigma}(F_{K,t+1} + 1 - \delta).$$

As a consequence, we would have $\beta R = 1$ all the time in steady state, regardless of whether $\sigma \gtreqless 1$.

In the presence of information frictions, (23) together with the properties of H_{t+1}/H_t in the long run gives

Proposition 8. *$\beta R \gtreqless 1$ in steady state at constrained efficient allocations if $\sigma \gtreqless 1$.*

With a fixed labor supply, Farhi and Werning (2012) provided a formal proof for the case of logarithmic utility, that is, $\beta R = 1$ in steady state at constrained efficient allocations if $\sigma = 1$. In the more general CRRA utility cases, they relied solely on their numerical results to find $\beta R \gtrless 1$ if $\sigma \gtrless 1$. Here we provide a formal proof for all cases of the CRRA utility function with a flexible labor supply. The proof is ultimately built on Lemma 1. Note that the derivation of the above proposition uses only the intertemporal condition (23), not the intratemporal condition (24). So, it should not be a surprise that our results are compatible with those of Farhi and Werning (2012), which assume a fixed rather than a flexible labor supply.

6 Extensions

This section extends the benchmark model to incorporate different or additional frictions/constraints that have been introduced in the literature. In all cases considered, the nonnegative martingale in the benchmark model reduces to a martingale and, at the same time, there exists a force that makes consumption Pareto-Negishi weights increase monotonically over time.

Specifically, we consider three cases in which the social planning problems are as follows:

1. Atkeson and Lucas (1995), denoted AL:

$$\begin{array}{ll} \max_{\{c,l,K\}} & \sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right], \\ & \sum_{s=t}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi_{s-t}(\theta^s|\theta^t)\left[u(c_s(\theta^s))-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right]\geq \underline{U}_1,\ \forall t,\theta^t, \\ s.t. & (4) \text{ and } (2), \end{array}$$

where $\underline{U}_1$ represents a lower bound that the promised utility of agents is entitled to at any time and in any state. The AL social planning problem is exactly the same as in the benchmark model, except for the additional constraints associated with $\underline{U}_1$. Incorporating this type of additional constraints is in the spirit of Atkeson and Lucas (1995), in that they studied the imposition of a lower bound promised utility in a setting where there are privately informed i.i.d. idiosyncratic shocks in the absence of capital. A possible rationale for incorporating this type of constraint in our framework is to extend the Mirrleesian economy from "closed" to "open" and thus allow for the possibility that agents can choose to migrate and renege on the contracts offered by the social planner at each moment of time. To prevent agents from breaking the contracts and walking away, the social planner faces the additional constraints associated with $\underline{U}_1$. We attach the Lagrangian multipliers $\{\beta^t\pi_t(\theta^t)\lambda(\theta^t)\}$ to these additional constraints.

2. Farhi and Werning (2007, 2010), denoted FW:[13]

$$\begin{array}{ll} \max_{\{c,l,K\}} & \sum_{t=0}^{T}\sum_{\theta^t}(\beta\kappa)^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right], \\ s.t. & (4) \text{ and } (2), \end{array}$$

where $\kappa > 1$. The FW social planning problem is exactly the same as in the benchmark model, except for the presence of the term $\kappa > 1$. Incorporating $\kappa > 1$ is in the spirit of Farhi and Werning (2007, 2010), in that they studied the situation where social discounting is in excess of private discounting β with privately informed idiosyncratic shocks. FW considered a dynastic setting in which a parent and child are linked by altruism but they are distinct individuals. The setting of $\kappa > 1$ arises in FW due to the presence of the additional constraints that the welfare of the child must exceed some lower bound.

[13] Phelan (2006) considered a special case of FW.

3. Armenter and Mertens (2013), denoted AM:

$$\begin{aligned}
\max \quad & \textstyle\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right], \\
& u\left(c_t(\theta^t)\right)-v\left(\tfrac{l_t(\theta^t)}{\theta_t}\right)+\textstyle\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right] \\
s.t. \quad & \geq(1-\varpi)\left(\begin{array}{c} u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right)-v\left(\frac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right)+ \\ \textstyle\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right)-v\left(\frac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right]\end{array}\right) \\
& \qquad +\varpi\underline{U}_3,\ \forall t,\theta^t,\widehat{\theta}_t, \\
& \text{and (2),}
\end{aligned}$$

where $\varpi>0$ denotes an exogenous state verification probability and $\underline{U}_3$ is the lowest utility the planner can assign to an agent if her mimicking is detected. The AM social planning problem is exactly the same as in the benchmark model, except for the replacement of the sequence of temporal IC constraints (4) by the constraints associated with $\underline{U}_3$. Replacing the IC constraints here is in the spirit of Armenter and Mertens (2013) since they introduced an exogenous state verification of agents' states into dynamic Mirrleesian economies. We attach the Lagrangian multipliers $\{\varphi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\}$ to the constraints associated with $\underline{U}_3$.

6.1 Martingale and monotonically increasing force

Despite incorporating different or additional frictions/constraints, we can integrate all three cases, AL, FW and AM, into a unified framework and view them as a single social planning problem:

$$\max_{\{c,l,K\}}\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[\widehat{\phi}_c(\theta^t)u\left(c_t(\theta^t)\right)-\widehat{\phi}_l(\theta^t)v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right],\ s.t.\ (2). \tag{28}$$

Appendix A.4 shows that the evolutions of $\widehat{\phi}_c$ and of $\widehat{\phi}_l$ satisfy the following form:

$$\begin{aligned}
\widehat{\phi}_c(\theta^t) &= \Omega_1\cdot\underbrace{\phi_c(\theta^t)}_{\text{Martingale}}+\Omega_2\cdot\underbrace{\Delta_t}_{\text{Non-decreasing}}, \\
\widehat{\phi}_l(\theta^t) &= \Omega_1\cdot\phi_l(\theta^t)+\Omega_2\cdot\underbrace{\Delta_t}_{\text{Non-decreasing}},
\end{aligned}$$

where ϕ_c and ϕ_l are the Pareto-Negishi weights in the benchmark model as given by (6) and (7). In the AL case, $\Omega_1=\Omega_2=1$ and $\Delta_t(\theta^t)=\sum_{s=0}^{t}\lambda(\theta^t)$, where $\{\lambda(\theta^t)\}$ are the multipliers on the

constraints of the lower bound promised utility. In the FW case, $\Omega_1 = \Omega_2 = 1$ and $\Delta_t(\theta^t) = \kappa^t - 1$, where $\kappa > 1$ is the factor leading to social discounting in excess of private discounting. In the AM case, $\Omega_1 = 1 - \varpi$, $\Omega_2 = \varpi$ and $\Delta_t(\theta^t) = \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \varphi(\theta^{s-1}, \theta_s, \widehat{\theta}_s)$, where $\{\varphi(\theta^{t-1}, \theta_t, \widehat{\theta}_t)\}$ are the multipliers on the sequence of the temporal IC constraints with state verification.

From $\widehat{\phi}_c$, we obtain

$$\widehat{\phi}_c(\theta^t) - \widehat{\phi}_c(\theta^{t-1}) = \Omega_1 \left(\phi_c(\theta^t) - \phi_c(\theta^{t-1})\right) + \Omega_2 \left(\Delta_t(\theta^t) - \Delta_{t-1}(\theta^{t-1})\right), \tag{29}$$

where $\Delta_t(\theta^t) \geq \Delta_{t-1}(\theta^{t-1})$, $\forall \theta^{t-1}, \theta^t$. That is, the evolution of $\widehat{\phi}_c$ consists of two components: the evolution of $\phi_c(\theta^t)$ and the nonnegative, non-decreasing sequence $\{\Delta_t(\theta^t)\}$. Irrespective of AL, FW or AM, the evolution of the consumption Pareto-Negishi weight $\widehat{\phi}_c$ associated with (28) can always be represented by the two components of (29).

First, note that the second component of (29) tends to make $\widehat{\phi}_c$ increase over time. Since, other things being equal, $\widehat{\phi}_c$ is inversely related to the marginal utility $u'(.)$ (see (12) by replacing ϕ_c with $\widehat{\phi}_c$), the increasing tendency of $\widehat{\phi}_c$ over time implies that, to smooth the cost of producing marginal utility (i.e., $1/u'(.)$) over time, the social planner will have a tendency to implement a back- instead of front-loading consumption plan if based on the second component of $\widehat{\phi}_c$ alone. This force is opposite to that of driving immiseration in the benchmark model.

Next, $\phi_c \geq 0$ always holds in the benchmark model. This property can be seen from the first-order condition (12). The condition (12) remains true in the extension, except for ϕ_c being replaced by $\widehat{\phi}_c$. Thus, the property of $\phi_c \geq 0$ in the benchmark is carried to the extension with $\widehat{\phi}_c \geq 0$, implying that $\phi_c \geq 0$ need not be true in the extension. However, the other two properties of $\phi_c(\theta^t)$ in Lemma 1 (i.e., $E[\phi_c(\theta^t)] = 1$ and (10)) remain intact. In other words, the process of $\{\phi_c(\theta^t)\}$ as a first component of (29) is a martingale, but no longer a nonnegative martingale. An important consequence is that Dobb's (1953) martingale convergence theorem is not applicable to $\{\phi_c(\theta^t)\}$ in the extension. Indeed, with $E[\phi_c(\theta^t)] = 1$ but in the absence of the lower bound $\phi_c \geq 0$, the very nature of the random walk of $\phi_c(\theta^t)$ described by (8) tends to yield an outcome in which, as $t \to \infty$, some fractions of agents have $\phi_c(\theta^t) \to +\infty$, while other fractions of agents have $\phi_c(\theta^t) \to -\infty$.

It is unclear how the two components of (29) together shape the evolution of $\widehat{\phi}_c$. Nevertheless, we are able to derive some results from (29).

6.2 Some results

Note first that the social planning problem (28) is identical to the social planning problem (5), except that the Pareto-Negishi weights $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ in the benchmark model are replaced with the new Pareto-Negishi weights $\widehat{\phi}_c(\theta^t)$ and $\widehat{\phi}_l(\theta^t)$. This means that the optimal conditions (12)-(14) remain intact in the extension as long as $\{\phi_c, \phi_l\}$ in (12)-(14) are replaced with $\{\widehat{\phi}_c, \widehat{\phi}_l\}$. Thus, all the equations/formulas derived in the benchmark model remain valid in the extension as long as the Pareto-Negishi weights $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ in the benchmark model are replaced with the new Pareto-Negishi weights $\widehat{\phi}_c(\theta^t)$ and $\widehat{\phi}_l(\theta^t)$. Recognizing this feature will be key to our analysis below.

6.2.1 Negative expected wealth taxes

From (20) and (29), we have

$$E_t\left(\tau^*_{W,t+1}(\theta^{t+1})\right) = 1 - E_t\left(\frac{\widehat{\phi}_c(\theta^{t+1})}{\widehat{\phi}_c(\theta^t)}\right) \leq 0, \forall t; \ < 0, \forall t \text{ if } \Delta_{t+1}(\theta^{t+1}) > \Delta_t(\theta^t) \text{ for some } \theta^t, \theta^{t+1}.$$

The result of zero expected wealth taxes derived by Kocherlakota (2005) now becomes that of negative expected wealth taxes.

6.2.2 Modified inverse Euler equation

From (17), we have

$$\sum_{\theta_{t+1}} \frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)} \pi(\theta_{t+1}|\theta_t) = \frac{\beta\left(F_{K,t+1} + 1 - \delta\right)}{u'\left(c_t(\theta^t)\right)} \sum_{\theta_{t+1}} \frac{\widehat{\phi}_c(\theta^{t+1})}{\widehat{\phi}_c(\theta^t)} \pi(\theta_{t+1}|\theta_t).$$

From (12) and (14), we have $\widehat{\phi}_c(\theta^t) = \mu_{t+1}\left(F_{K,t+1} + 1 - \delta\right)/\beta^t u'\left(c_t(\theta^t)\right)$. Substituting it into the above equation and then applying (29) plus the martingale of ϕ_c, we obtain

$$\sum_{\theta_{t+1}} \frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)} \pi(\theta_{t+1}|\theta_t) = \frac{\beta\left(F_{K,t+1} + 1 - \delta\right)}{u'\left(c_t(\theta^t)\right)} + \frac{\beta^{t+1}\Omega_2}{\mu_{t+1}} \sum_{\theta_{t+1}} \left(\Delta_{t+1}(\theta^{t+1}) - \Delta_t(\theta^t)\right) \pi(\theta_{t+1}|\theta_t),$$

which is qualitatively the same as the "modified IEE" derived by Farhi and Werning (2010, Eq. (27)). When $\Omega_2 = 0$, the modified IEE reduces to the IEE given by (18). In the presence of the last term in the modified IEE, it can be checked that the result of $\tau^*_{kt}(\theta^t) > 0$ implied by the IEE is not necessarily true any longer. This result is not surprising in light of the presence of the second component of (29) that exerts a "back-loading" force on consumption.

6.2.3 No immiseration

AL, FW, and AM have all found that the immiseration outcome of Proposition 3 is not applicable in their framework. The properties embodied in the two components of (29) imply this result. Formally, we prove:

Proposition 9. *Given that the evolution of the consumption Pareto-Negishi weight follows (29), the immiseration result fails to hold.*

Proof. We prove by contradiction that $\widehat{\phi}_c$ cannot converge to zero. Suppose that $\widehat{\phi}_c$ converges along a generic sample path θ^t as t goes to ∞. Given that, we obtain

$$\lim_{t\to\infty}\left[\widehat{\phi}_c\left(\theta^{t+1}\right) - \widehat{\phi}_c\left(\theta^t\right)\right] = \lim_{t\to\infty}\left[\Omega_1\left(\phi_c\left(\theta^{t+1}\right) - \phi_c\left(\theta^t\right)\right) + \Omega_2\left(\Delta_{t+1}\left(\theta^{t+1}\right) - \Delta_t\left(\theta^t\right)\right)\right] = 0,$$

which implies $\lim_{t\to\infty}\left[\phi_c\left(\theta^{t+1}\right) - \phi_c\left(\theta^t\right)\right] \leq 0$ since $\{\Delta_t\left(\theta^t\right)\}$ is a non-decreasing sequence. The result of $\lim_{t\to\infty}\left[\phi_c\left(\theta^{t+1}\right) - \phi_c\left(\theta^t\right)\right] \leq 0$ violates the property $\sum_{\theta_{t+1}} \phi_c\left(\theta^{t+1}\right)\pi\left(\theta_{t+1}|\theta_t\right) = \phi_c\left(\theta^t\right)$. The non-convergence of $\widehat{\phi}_c$ leads to a non-degenerate distribution of consumption and thus no immiseration. □

6.2.4 Aggregate wedges and steady-state βR

$H_t = \sum_{\theta^t} \widehat{\phi}_c(\theta^t)^{\frac{1}{\sigma}}\pi_t(\theta^t)$ in the extension. We have the following result.

Lemma 5. $H_t < H_{t+1}$, $\forall t$, *including* $t \to \infty$, *if* $\sigma \leq 1$.

Proof. From (29), using the martingale of $\{\phi_c(\theta^t)\}$ and applying the power mean inequality gives

$$\widehat{\phi}_c(\theta^t) < \sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t) \Rightarrow \widehat{\phi}_c(\theta^t)^{\frac{1}{\sigma}} < \left(\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t)\right)^{\frac{1}{\sigma}}. \tag{30}$$

$$\left(\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t)\right) \leq \left(\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})^{\frac{1}{\sigma}}\pi(\theta_{t+1}|\theta_t)\right)^{\sigma}, \text{ if } \sigma \leq 1$$

$$\Rightarrow \left(\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t)\right)^{\frac{1}{\sigma}} \leq \sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})^{\frac{1}{\sigma}}\pi(\theta_{t+1}|\theta_t), \quad \text{if } \sigma \leq 1. \tag{31}$$

Putting (30) and (31) together yields

$$\widehat{\phi}_c(\theta^t)^{\frac{1}{\sigma}} < \left(\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})\pi(\theta_{t+1}|\theta_t)\right)^{\frac{1}{\sigma}} \leq \sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})^{\frac{1}{\sigma}}\pi(\theta_{t+1}|\theta_t) \quad \text{if } \sigma \leq 1.$$

Summing the first and the last term of the above inequality over θ^t gives

$$\begin{aligned} \sum_{\theta^t} \widehat{\phi}_c(\theta^t)^{\frac{1}{\sigma}}\pi_t(\theta^t) \quad &< \quad \sum_{\theta^t}\sum_{\theta_{t+1}} \widehat{\phi}_c(\theta^{t+1})^{\frac{1}{\sigma}}\pi(\theta_{t+1}|\theta_t)\pi_t(\theta^t) \quad \text{if } \sigma \leq 1. \\ &\Rightarrow \quad H_t < H_{t+1} \quad \text{if } \sigma \leq 1. \end{aligned}$$

Since $\widehat{\phi}_c$ does not converge, (30) and (31) hold as $t \to \infty$. That being so, $H_t < H_{t+1}$, $\forall t$, including $t \to \infty$, if $\sigma \leq 1$. □

From (25), this lemma immediately leads to

Proposition 10. $\tau^*_{Kt} < 0$, $\forall t$, *including* $t \to \infty$, *if* $\sigma \leq 1$.

Suppose there exists a steady state in which C_t and K_t/L_t are constant over time. Using Lemma 5, we obtain from (23) the following result:

Proposition 11. $\beta R < 1$ *in steady state at constrained efficient allocations if* $\sigma \leq 1$.

Imposing a lower bound for the promised utility, Atkeson and Lucas (1995) showed that $\beta R < 1$ in steady state at constrained efficient allocations in a setting where idiosyncratic shocks are i.i.d. and there is no capital. Proposition 11 extends their result to a more general environment for the cases where $\sigma \leq 1$.

7 Conclusion

This paper considers a dynamic economy in which privately informed idiosyncratic productivity shocks evolve stochastically. The stochastic nature of these productivity shocks accounts for a variety of uninsurable risks to the value of human capital over the course of people's lives. Assuming a standard neoclassical production technology, we characterize constrained efficient intertemporal wedges for the macro aggregate as well as the micro individual allocation of the economy.

Instead of the conventional recursive approach in the DC/NDPF literature, we reformulate the social planning problem in a sequential fashion. We first construct "Pareto-Negishi weights" from the multipliers on a sequence of temporary incentive constraints. For a fairly general stochastic process of idiosyncratic productivity shocks, we show that the evolution of the Pareto-Negishi weight associated with agents' consumption is a nonnegative martingale. Using this powerful property, we expose existing results in the literature and derive new results on intertemporal wedges. We also consider extensions in which the nonnegative martingale is modified.

The focus of this paper is on inter- rather than intra-temporal wedges. This is mainly due to the fact that relatively few useful properties are known about the stochastic process of the Pareto-Negishi weight associated with agents' labor input. Recent work by Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016) has addressed optimal labor distortions in dynamic Mirrleesian economies. Following their line of investigation, it is hoped that there will be a corresponding characterization of constrained efficient intratemporal wedges both in the individual and in the aggregate in future work.

A Appendix

A.1 Social planner's problem

This appendix derives the details of the social planner's problem.

The Lagrangian associated with (1), (2) and (4) is given by

$$
\begin{aligned}
L \;=\; & \min_{\{\psi,\mu\}} \max_{\{c,l,K\}} \sum_{t=0}^{T} \sum_{\theta^t} \beta^t \pi_t(\theta^t) \left[u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) \right] + \sum_{t=0}^{T} \sum_{\theta^t, \widehat{\theta}^t} \beta^t \pi_t(\theta^t) \psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) \\
& \times \left\{ \begin{array}{c} u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) + \sum_{s=t+1}^{T} \sum_{\theta^s|\theta^t} \beta^{s-t} \pi(\theta^s|\theta^t) \left[u\left(c_s(\theta^s)\right) - v\left(\frac{l_s(\theta^s)}{\theta_s}\right) \right] \\ -u\left(c_t(\theta^{t-1}, \widehat{\theta}_t)\right) + v\left(\frac{l_t(\theta^{t-1}, \widehat{\theta}_t)}{\theta_t}\right) - \sum_{s=t+1}^{T} \sum_{\theta^s|\theta^t} \beta^{s-t} \pi(\theta^s|\theta^t) \left[u\left(c_s(\theta^s; \widehat{\theta}_t)\right) - v\left(\frac{l_s(\theta^s; \widehat{\theta}_t)}{\theta_s}\right) \right] \end{array} \right\} \\
& + \sum_{t=0}^{T} \mu_t \left[F(K_t, L_t) + (1-\delta) K_t - C_t - K_{t+1} \right].
\end{aligned}
$$

The part associated with the temporary IC constraints (4) can be decomposed into three components:

$$
\begin{aligned}
& \underbrace{\sum_{t=0}^{T} \sum_{\theta_t} \sum_{\widehat{\theta}_t} \beta^t \psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) \pi_t(\theta^t) \left[u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) - u\left(c_t(\theta^{t-1}, \widehat{\theta}_t)\right) + v\left(\frac{l_t(\theta^{t-1}, \widehat{\theta}_t)}{\theta_t}\right) \right]}_{\text{Component A}} \\
& + \underbrace{\sum_{t=0}^{T} \sum_{\theta_t} \sum_{\widehat{\theta}_t} \beta^t \psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) \pi_t(\theta^t) \sum_{s=t+1}^{T} \sum_{\theta^s|\theta^t} \beta^{s-t} \frac{\pi_s(\theta^s)}{\pi_t(\theta^t)} \left[u\left(c_s(\theta^s)\right) - v\left(\frac{l_s(\theta^s)}{\theta_s}\right) \right]}_{\text{Component B}} \\
& - \underbrace{\sum_{t=0}^{T} \sum_{\theta_t} \sum_{\widehat{\theta}_t} \beta^t \psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) \pi_t(\theta^t) \sum_{s=t+1}^{T} \sum_{\theta^s|\theta^t} \beta^{s-t} \frac{\pi_s(\theta^s)}{\pi_t(\theta^t)} \left[u\left(c_s(\theta^s; \widehat{\theta}_t)\right) - v\left(\frac{l_s(\theta^s; \widehat{\theta}_t)}{\theta_s}\right) \right]}_{\text{Component C}}.
\end{aligned}
$$

Each component can be further rearranged as follows:

1. Component A:

$$\sum_{t=0}^{T}\sum_{\theta_t}\sum_{\widehat{\theta}_t}\beta^t\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)-u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right)+v\left(\frac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right)\right]$$

$$=\sum_{t=0}^{T}\sum_{\theta_t}\sum_{\widehat{\theta}_t}\beta^t\pi_t(\theta^t)\left[\begin{array}{c} u\left(c_t(\theta^t)\right)\left(\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)-\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi_t(\theta^{t-1},\widehat{\theta}_t)}{\pi_t(\theta^t)}\right) \\ -v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\left(\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)-\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})}\left(\frac{\theta_t}{\widehat{\theta}_t}\right)^{\gamma}\right)\end{array}\right].$$

2. Component B:

$$\sum_{t=0}^{T}\sum_{\theta_t}\sum_{\widehat{\theta}_t}\beta^t\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\pi_t(\theta^t)\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\frac{\pi_s(\theta^s)}{\pi_t(\theta^t)}\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right]$$

$$=\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s).$$

3. Component C:

$$\sum_{t=0}^{T}\sum_{\theta_t}\sum_{\widehat{\theta}_t}\beta^t\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\pi_t(\theta^t)\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\frac{\pi_s(\theta^s)}{\pi_t(\theta^t)}\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right)-v\left(\frac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right]$$

$$=\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_t(\theta^t;\widehat{\theta}_s)}{\pi_t(\theta^t)}.$$

Putting together the three components, we obtain the Lagrangian (5) with $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ given by (6) and (7).

A.2 Proof of Lemma 1

Here is the proof of Lemma 1.

Proof. First, rearranging (12), we have $\phi_c(\theta^t)=\frac{u'\left(c_t^*(\theta^t)\right)^{-1}}{\sum_{\theta^t}u'(c_t^*(\theta^t))^{-1}\pi_t(\theta^t)}$, where $\frac{1}{u'}$ is nonnegative; so $\phi_c(\theta^t)$ is nonnegative. Second, consider epsilon defined in equations (8) and (9) and calculate

$E_{t-1}(\varepsilon(\theta^t))$, i.e., the expectation of ε conditional on information at date $t-1$:

$$\begin{aligned}
E_{t-1}(\varepsilon(\theta^t)) &= \sum_{\theta_t} \varepsilon(\theta^t)\pi(\theta_t|\theta_{t-1}) \\
&= \sum_{\theta_t}\sum_{\widehat{\theta}_t} \left[\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t) - \psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})}\right]\pi(\theta_t|\theta_{t-1}) \\
&\quad -\sum_{\theta_t}\sum_{\widehat{\theta}_{t-1}} \frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})}\psi(\theta^{t-2},\widehat{\theta}_{t-1},\theta_{t-1})\left[\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} - 1\right]\pi(\theta_t|\theta_{t-1}) \\
&= \sum_{\theta_t}\sum_{\widehat{\theta}_t}\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\pi(\theta_t|\theta_{t-1}) - \sum_{\theta_t}\sum_{\widehat{\theta}_t}\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\pi(\widehat{\theta}_t|\theta_{t-1}) \\
&\quad -\sum_{\widehat{\theta}_{t-1}} \frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{s-1}|\theta_{t-2})}\psi(\theta^{t-2},\widehat{\theta}_{t-1},\theta_{t-1})\sum_{\theta_t}\left[\pi(\theta_t|\widehat{\theta}_{t-1}) - \pi(\theta_t|\theta_{t-1})\right] \\
&= 0.
\end{aligned}$$

We provide an intuition for the last equality (i.e., equal to zero) in the main text. Applying the zero conditional expectation result of $\varepsilon(\theta^t)$ to equation (6), we obtain

$$\sum_{\theta_t}\phi_c(\theta^t)\pi(\theta_t|\theta_{t-1}) = \phi_c(\theta^{t-1}).$$

Finally, applying the law of iterated expectation to (10), we obtain

$$\begin{aligned}
E[\phi_c(\theta^t)] &= E[E_{t-1}(\phi_c(\theta^t))] = E[\phi_c(\theta^{t-1})] \\
&= E[E_{t-2}(\phi_c(\theta^{t-1}))] = E[\phi_c(\theta^{t-2})] \\
&= ... = E[\phi_c(\theta^0)] \\
&= \phi_c(\theta^0) = 1
\end{aligned}$$

Therefore, $\phi_c(\theta^t)$ has a finite first moment. □

A.3 Proof of Lemma 2

The proof of part (1) is obvious. Under the period utility function $u(c_t) - v(n_t)$, the nonnegative martingale of $\{\phi_c(\theta^t)\}$ has to do with the consumption part of the utility only. We prove part (2)

as follows:

Following the derivation in Appendix A.1, we obtain

$$\sum_{t=0}^{T}\sum_{\theta_t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)\phi_c(\theta^t)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\phi_l(\theta^t)\right], \tag{32}$$

where

$$\phi_c(\theta^t)=1+\sum_{s=0}^{t}\sum_{\widehat{\theta}_s}\left(\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_t(\theta^t;\widehat{\theta}_s)}{\pi_t(\theta^t)}\right)$$

and

$$\phi_l(\theta^t)=1+\sum_{s=0}^{t}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_t(\theta^t;\widehat{\theta}_s)}{\pi_t(\theta^t)}-\sum_{\widehat{\theta}_t}\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|x(\theta^{t-1});t)}{\pi(\theta_t|x(\theta^{t-1});t)}\left(\frac{\theta_t}{\widehat{\theta}_t}\right)^{\gamma}.$$

We then have

$$\begin{aligned}\phi_c(\theta^t)-\phi_c(\theta^{t-1}) &= \sum_{s=0}^{t}\sum_{\widehat{\theta}_s}\left(\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_t(\theta^t;\widehat{\theta}_s)}{\pi_t(\theta^t)}\right)\\ &\quad-\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\left(\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_{t-1}(\theta^{t-1};\widehat{\theta}_s)}{\pi_{t-1}(\theta^{t-1})}\right),\end{aligned}$$

where the first term on the right-hand side can be re-expressed as

$$\begin{aligned}&\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\left(\psi(\theta^{s-1},\theta_s,\widehat{\theta}_s)-\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)\pi_{t-1}(\theta^{t-1};\widehat{\theta}_s)}{\pi(\theta_t|x(\theta^{t-1});t)\pi_{t-1}(\theta^{t-1})}\right)\\ &+\sum_{\widehat{\theta}_t}\left(\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)-\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|x(\theta^{t-1});t)}{\pi(\theta_t|x(\theta^{t-1});t)}\right).\end{aligned}$$

Thus, we obtain

$$\phi_c(\theta^t)=\phi_c(\theta^{t-1})+\epsilon(\theta^t),$$

where

$$\begin{aligned}\epsilon(\theta^t) &= \sum_{\widehat{\theta}_t}\left(\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)-\psi(\theta^{t-1},\widehat{\theta}_t,\theta_t)\frac{\pi(\widehat{\theta}_t|x(\theta^{t-1});t)}{\pi(\theta_t|x(\theta^{t-1});t)}\right)\\ &\quad -\sum_{s=0}^{t-1}\sum_{\widehat{\theta}_s}\psi(\theta^{s-1},\widehat{\theta}_s,\theta_s)\frac{\pi_{t-1}(\theta^{t-1};\widehat{\theta}_s)}{\pi_{t-1}(\theta^{t-1})}\left(\frac{\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)}{\pi(\theta_t|x(\theta^{t-1});t)}-1\right).\end{aligned}$$

For concreteness, we consider three possible cases of idiosyncratic shocks for the specification of the conditional probability $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)$, where the notation $(\theta^{t-1};\widehat{\theta}_s)$ denotes that the elements of θ^{t-1} corresponding to $\widehat{\theta}_s$ are replaced by $\widehat{\theta}_s$.

- Case 1: i.i.d. shocks where $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t)$ for $s\leq t-1$.

- Case 2: a first-order Markov process where $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t|\widehat{\theta}_{t-1})$ for $s=t-1$, and $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t|\theta_{t-1})$ for $s<t-1$.

- Case 3: a second-order Markov process where $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t|\widehat{\theta}_{t-1},\theta_{t-2})$ for $s=t-1$, $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t|\theta_{t-1},\widehat{\theta}_{t-2})$ for $s=t-2$, and $\pi(\theta_t|x(\theta^{t-1};\widehat{\theta}_s);t)=\pi(\theta_t|\theta_{t-1},\theta_{t-2})$ for $s<t-2$.

Let the first and the second component of $\epsilon(\theta^t)$ be $\epsilon_1(\theta^t)$ and $\epsilon_2(\theta^t)$. Using the logic of proving Lemma 1 in Appendix A.2, it can be shown that

$$\sum_{\theta_t}\epsilon_1(\theta^t)\pi(\theta_t|x(\theta^{t-1});t)=\sum_{\theta_t}\epsilon_2(\theta^t)\pi(\theta_t|x(\theta^{t-1});t)=0. \tag{33}$$

We are ready to show that $\phi_c(\theta^t)$ is a nonnegative martingale. First, except for a more general shock process, the Lagrangian associated with (32) is identical to the Lagrangian (5). Hence, following the same arguments as in Appendix A.2, $\phi_c(\theta^t)>0$. In addition, Eq. (33) indicates

$$\sum_{\theta_t}\phi_c(\theta^t)\pi(\theta_t|\theta_{t-1})=\phi_c(\theta^{t-1}),$$

which implies that the unconditional mean of $\phi_c(\theta^t)$ equals 1 as well. This completes the proof.

A.4 Pareto-Negishi weights in the extension

The AL case imposes a lower bound on the promised utility. This type of additional constraints can be expressed as

$$\sum_{s=t}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi_{s-t}(\theta^s|\theta^t)\left[u(c_s(\theta^s))-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right]\geq \underline{U}_1,\ \forall t,\theta^t,$$

where $\underline{U}_1$ represents the lower bound that agents are entitled to in any state and at any time. Attach Lagrangian multipliers $\{\beta^t\pi_t(\theta)\lambda(\theta^t)\}$ to the constraints above and incorporate them into the Lagrangian (5). The new Pareto-Negishi weights become

$$\begin{aligned}\widehat{\phi}_c(\theta^t) &= \phi_c(\theta^t)+\sum_{s=0}^{t}\lambda(\theta^s),\\ \widehat{\phi}_l(\theta^t) &= \phi_l(\theta^t)+\sum_{s=0}^{t}\lambda(\theta^s),\end{aligned}$$

where $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ are given by (6) and (7).

The FW case considers the situation where social discounting is in excess of private discounting. Let the social planner's discount factor be $\beta\kappa$ with $\kappa>1$. The Lagrangian (5) is modified to be

$$\begin{aligned}L &= \min_{\{\psi,\mu\}}\max_{\{c,l,K\}}\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\kappa^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]+\sum_{t=0}^{T}\sum_{\theta^t,\widehat{\theta}^t}\beta^t\pi_t(\theta^t)\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\\ &\quad\times\left\{\begin{array}{c} u\left(c_t(\theta^t)\right)-v\left(\frac{l_t(\theta^t)}{\theta_t}\right)+\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right]\\ -u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right)+v\left(\frac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right)-\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right)-v\left(\frac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right]\end{array}\right\}\\ &\quad+\sum_{t=0}^{T}\mu_t\left[F(K_t,L_t)+(1-\delta)K_t-C_t-K_{t+1}\right],\end{aligned}$$

which can be expressed as

$$\begin{aligned}L &= \min_{\{\psi,\mu\}}\max_{\{c,l,K\}}\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[(\phi_c(\theta^t)+\kappa^t-1)u\left(c_t(\theta^t)\right)-(\phi_l(\theta^t)+\kappa^t-1)v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]\\ &\quad+\sum_{t=0}^{T}\mu_t\left[F(K_t,L_t)+(1-\delta)K_t-C_t-K_{t+1}\right]\end{aligned}$$

where $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ are given by (6) and (7). As such, the new Pareto-Negishi weight on consumption becomes

$$\begin{aligned}\widehat{\phi}_c(\theta^t) &= \phi_c(\theta^t) + \kappa^t - 1,\\ \widehat{\phi}_l(\theta^t) &= \phi_l(\theta^t) + \kappa^t - 1.\end{aligned}$$

The AM case endows the social planner with a random verification technology. With the availability of the verification technology, temporary incentive constraints are modified to be

$$\begin{gathered}u\left(c_t(\theta^t)\right) - v\left(\tfrac{l_t(\theta^t)}{\theta_t}\right) + \textstyle\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s)\right) - v\left(\tfrac{l_s(\theta^s)}{\theta_s}\right)\right]\\ \geq (1-\varpi)\left(u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right) - v\left(\tfrac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right) + \textstyle\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right) - v\left(\tfrac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right]\right)\\ +\varpi\underline{U}_3, \forall t, \theta^t, \widehat{\theta}_t,\end{gathered}$$

where $\varpi > 0$ denotes an exogenous state verification probability and $\underline{U}_3$ is the lowest utility the planner can assign to an agent if her mimicking is detected.

Let $\{\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\}$ be the multipliers associated with the temporary incentive constraints. The Lagrangian becomes

$$\begin{aligned}L &= \min_{\{\psi,\mu\}}\max_{\{c,l,K\}}\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right] + \sum_{t=0}^{T}\sum_{\theta^t,\widehat{\theta}^t}\beta^t\pi_t(\theta^t)\psi(\theta^{t-1},\theta_t,\widehat{\theta}_t)\\ &\times\left\{\begin{array}{c} u\left(c_t(\theta^t)\right) - v\left(\tfrac{l_t(\theta^t)}{\theta_t}\right) + \sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s)\right) - v\left(\tfrac{l_s(\theta^s)}{\theta_s}\right)\right]\\ -(1-\varpi)\left(\begin{array}{c} u\left(c_t(\theta^{t-1},\widehat{\theta}_t)\right) - v\left(\tfrac{l_t(\theta^{t-1},\widehat{\theta}_t)}{\theta_t}\right)\\ +\sum_{s=t+1}^{T}\sum_{\theta^s|\theta^t}\beta^{s-t}\pi(\theta^s|\theta^t)\left[u\left(c_s(\theta^s;\widehat{\theta}_t)\right) - v\left(\tfrac{l_s(\theta^s;\widehat{\theta}_t)}{\theta_s}\right)\right]\end{array}\right)\\ -\varpi\underline{U}_3\end{array}\right\}\\ &+\sum_{t=0}^{T}\mu_t\left[F(K_t,L_t) + (1-\delta)K_t - C_t - K_{t+1}\right].\end{aligned}$$

The above can be rewritten as

$$\begin{aligned}L &= \min_{\{\psi,\mu\}}\max_{\{c,l,K\}}\sum_{t=0}^{T}\sum_{\theta^t}\beta^t\pi_t(\theta^t)\left[\widehat{\phi}_c(\theta^t)u\left(c_t(\theta^t)\right) - \widehat{\phi}_l(\theta^t)v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right]\\ &\quad+\mu_t\left[F(K_t,L_t) + (1-\delta)K_t - C_t - K_{t+1}\right],\end{aligned}$$

where

$$\begin{aligned}
\widehat{\phi}_c(\theta^t) &= (1-\varpi)\phi_c(\theta^t) + \varpi \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \psi(\theta^{s-1}, \theta_s, \widehat{\theta}_s), \\
\widehat{\phi}_l(\theta^t) &= (1-\varpi)\phi_l(\theta^t) + \varpi \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \psi(\theta^{s-1}, \theta_s, \widehat{\theta}_s),
\end{aligned}$$

and $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ are given by (6) and (7).

References

Albanesi, S., and R. Armenter (2012): "Intertemporal Distortions in the Second Best," *Review of Economic Studies*, 79(4), 1271–1307.

Armenter, R., and T. M. Mertens (2013): "Fraud Deterrence in Dynamic Mirrleesian Economies," *Journal of Monetary Economics*, 60(2), 139–151.

Atkeson, A., and J. Lucas, Robert E (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59(3), 427–53.

——— (1995): "Efficiency and Equality in a Simple Model of Efficient Unemployment Insurance," *Journal of Economic Theory*, 66(1), 64–88.

Attanasio, O. P., and G. Weber (1989): "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *Economic Journal*, 99(395), 59–73.

Attanasio, O. P., and G. Weber (2010): "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy," *Journal of Economic Literature*, 48(3), 693–751.

Bullen, P. (2013): *Handbook of Means and Their Inequalities*, Mathematics and Its Applications. Springer Netherlands.

Campbell, J. Y. (1999): "Asset Prices, Consumption, and the Business Cycle," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 19, pp. 1231–1303. Elsevier.

Campbell, J. Y., and N. G. Mankiw (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pp. 185–246. National Bureau of Economic Research, Inc.

Chamley, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54(3), 607–22.

Chien, Y., H. Cole, and H. Lustig (2011): "A Multiplier Approach to Understanding the Macro Implications of Household Finance," *Review of Economic Studies*, 78(1), 199–234.

Diamond, P. A., and J. A. Mirrlees (1978): "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10(3), 295–336.

Doob, J. (1953): *Stochastic Processes*, Wiley Publications in Statistics. John Wiley & Sons.

Espino, E. (2005): "On Ramsey's Conjecture: Efficient Allocations in the Neoclassical Growth Model with Private Information," *Journal of Economic Theory*, 121(2), 192–213.

Espino, E., J. Kozlowski, and J. M. Sánchez (2015): "Too Big to Cheat: Efficiency and Investment in Partnerships," Working Papers 2013-001, Federal Reserve Bank of St. Louis.

Farhi, E., and I. Werning (2007): "Inequality and Social Discounting," *Journal of Political Economy*, 115, 365–402.

——— (2010): "Progressive Estate Taxation," *The Quarterly Journal of Economics*, 125(2), 635–673.

——— (2012): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 120(3), 398–445.

——— (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

Favilukis, J., and X. Lin (2013): "Long Run Productivity Risk and Aggregate Investment," *Journal of Monetary Economics*, 60(6), 737–751.

Fernandes, A., and C. Phelan (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 91(2), 223–247.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

Golosov, M., M. Troshkin, and A. Tsyvinski (2011): "Optimal Taxation: Merging Micro and Macro Approaches," *Journal of Money, Credit and Banking*, 43(5), 147–174.

——— (2016): "Redistribution and Social Insurance," *American Economic Review*, forthcoming.

Golosov, M., and A. Tsyvinski (2015): "Policy Implications of Dynamic Public Finance," *Annual Review of Economics*, 7, 147–171.

Golosov, M., A. Tsyvinski, and I. Werning (2007): "New Dynamic Public Finance: A User's Guide," in *NBER Macroeconomics Annual 2006, Volume 21*, NBER Chapters, pp. 317–388. National Bureau of Economic Research, Inc.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96(2), 339–57.

Hansen, L. P., and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–86.

Hart, O., and Holmström (1987): "The Theory of Contracts," in *Advances in Economic Theory, Fifth World Congress*, ed. by T. Bewley. Cambridge University Press.

Judd, K. L. (1985): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28(1), 59–83.

Kapička, M. (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *Review of Economic Studies*, 80(3), 1027–1054.

Keane, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

Kehoe, P. J., and F. Perri (2002): "International Business Cycles with Endogenous Incomplete Markets," *Econometrica*, 70(3), 907–928.

Kocherlakota, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

——— (2010): *The New Dynamic Public Finance*, The Toulouse Lectures in Economics. Princeton University Press.

Laffont, J.-J., and J. Tirole (1988): "The Dynamics of Incentive Contracts," *Econometrica*, 56(5), 1153–75.

Ljungqvist, L., and T. J. Sargent (2012): *Recursive Macroeconomic Theory, Third Edition*, MIT Press Books. The MIT Press.

Marcet, A., and R. Marimon (1999): "Recursive Contracts," Working Paper, Universitat Pompeu Fabra.

Mele, A. (2014): "Repeated Moral Hazard and Recursive Lagrangeans," *Journal of Economic Dynamics and Control*, 42(C), 69–85.

Messner, M., N. Pavoni, and C. Sleet (2013): "The Dual Approach to Recursive Optimization: Theory and Examples," Working papers, SSRN.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

Park, Y. (2014): "Optimal Taxation in a Limited Commitment Economy," *Review of Economic Studies*, 81(2), 884–918.

Phelan, C. (2006): "Opportunity and Social Mobility," *Review of Economic Studies*, 73(2), 487–504.

Rogerson, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

Straub, L., and I. Werning (2014): "Positive Long Run Capital Taxation: Chamley-Judd Revisited," NBER Working Papers 20441, National Bureau of Economic Research, Inc.

Thomas, J., and T. Worrall (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51(2), 367–390.

Veracierto, M. (2014): "Adverse Selection, Risk Sharing and Business Cycles," Working Paper Series WP-2014-10, Federal Reserve Bank of Chicago.

Vissing-Jørgensen, A., and O. P. Attanasio (2003): "Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion," *American Economic Review*, 93(2), 383–391.

Werning, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122(3), 925–967.