Incentive Compatibility as a Nonnegative Martingale

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

December 2016

Abstract

This paper considers a dynamic Mirrleesian economy and decomposes agents' lifetime incentive compatibility (IC) constraints into a sequence of temporal ones. We encode the frequency and severeness of these temporal IC constraints by their associated Lagrange multipliers, showing that the accumulation of the Lagrange multipliers on the consumption part is a nonnegative martingale. It is emphasized that, distinct from the extant literature, this martingale property is derived directly from the IC constraints and so is independent of the optimization of the social planning problem. We apply our finding to (i) the characterization of constrained efficient intertemporal wedges in the aggregate, and (ii) the extension of the so-called immiseration to an environment more general than the extant literature.

JEL Classification: E21, D82, H21

Key Words: Nonnegative martingale, Intertemporal aggregate wedges, Dynamic Mirrleesian economies

*Corresponding author, YiLi Chien. Email: yili.chien@stls.frb.org. We thank Marek Kapička, Marcelo Veracierto, and participants at various seminars and conferences for useful comments.

1 Introduction

This paper considers a dynamic neoclassical 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. We obtain a *new* martingale property under a fairly general assumption for the stochastic process of idiosyncratic labor productivity shocks. We then apply the finding to (i) the characterization of constrained efficient intertemporal wedges in the aggregate, and (ii) the extension of the so-called immiseration to an environment more general than the extant literature. Despite the vast amount of work and progress achieved in the literature, it is surprising that relatively little is known about the property of constrained efficient allocations chosen by the social planner at the *aggregate* level while dealing with information frictions at the *individual* level. A main contribution of our paper is to shed light on the constrained efficient allocations at the aggregate level via the use of the new martingale property.

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 compatibility (IC) 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 contracts (DC) and new dynamic public finance (NDPF).[1] Instead of the conventional recursive formulation in the literature, however, we use a multiplier approach to reformulate the social planning problem in a sequential fashion. Following Fernandes and Phelan (2000), we first decompose agents' lifetime IC constraints into a sequence of temporary ones for all possible states at all times. We then encode the bite of these temporal IC constraints by their associated Lagrange multipliers. The accumulation of these multipliers evolves to ensure the satisfaction of agents' incentive compatibility over time and, at the same time, it endogenously determines stochastic Pareto-Negishi weights placed on an agent in the planner's

[1] For reviews on DC, see Part V in Ljungqvist and Sargent (2012) and Pavan (2016). For reviews on NDPF, see Golosov, Tsyvinski, and Werning (2007); Kocherlakota (2010); Golosov, Troshkin, and Tsyvinski (2011); and Golosov and Tsyvinski (2015). See Bergemann and Pavan (2015) for a recent symposium on DC and mechanism design. For an overview on the theory of recursive contracts to solve for DC/NDPF problems, see Golosov, Tsyvinski, and Werquin (2016). 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.

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.

The extant literature obtains a nonnegative martingale property in the following way; see Kocherlakota (2010, chapter 3). First it derives the so-called inverse Euler equation from the optimization of the social planning problem. This equation captures the social planner's desire to smooth the marginal cost of utility production over time. Next, it uses the inverse Euler equation to show that the marginal cost of utility production is a martingale under the restriction that $\beta R = 1$, where β is a discount factor and R a gross rate of return from saving. Our martingale property is derived directly from the IC constraints and so is independent of the optimization of the social planning problem. Importantly, it does not build on the restriction $\beta R = 1$.

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 considers 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. We apply our nonnegative martingale finding to the characterization of constrained efficient in the aggregate for the dynamic Mirrleesian economy. We address the optimal allocation over time for the economy as a whole in terms of the intertemporal aggregate wedge. 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, distortion for aggregate saving is negative or positive, depending on whether the IES is elastic or inelastic.

We also apply our nonnegative martingale finding to the extension of the so-called immiseration to a more general environment. Thomas and Worrall (1990) and Atkeson and Lucas (1992) obtained the immiseration result in i.i.d. and endowment economies. Kocherlakota (2010, chapter 3) applied Doob's (1953) martingale convergence theorem to the inverse Euler equation to derive immiseration under a more general environment. However, it requires imposing the restriction $\beta R = 1$. Using our obtained nonnegative martingale result, we demonstrate that immiseration still holds under Markov or more general stochastic processes and in a neoclassical production economy without the restriction $\beta R = 1$.[2]

Related literature

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

Werning (2007) has formulated dynamic Mirrleesian taxation by encoding IC 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; moreover, the issues investigated are different.

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 but working with a continuum of agent types, Kapička (2013) adopted the first-order approach to successfully overcome the complexity caused by the persistence of shocks as well as the large numbers of shock values.[3] Our approach differs because it characterizes intertemporal aggregate wedges completely through a specific moment of the Pareto-Negishi weight distribution. Our method fits into the literature of replacing primal state variables with dual ones. Messner, Pavoni, and Sleet (2016) argued that the dual approach is applicable to a large class of

[2] In this paper, we abstract from addressing the issue of implementation. For the implementation of constrained efficient allocations in dynamic Mirrleesian economies, see Kocherlakota (2010, chapter 4). The work provides a recipe for implementing constrained efficient allocations as part of a competitive equilibrium. Albanesi and Sleet (2006) and Golosov and Tsyvinski (2006) put forth simple implementations in certain specific stochastic environments.

[3] See also Pavan, Segal, and Toikka (2014). These authors worked with a continuum of agent types in the context of DC and extended the approach pioneered by Myerson (1981) to dynamic settings.

economies and often leads to an easy characterization of constrained efficient allocations.

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) envisioned 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 the IES. Given labor supply, Farhi and Werning (2012) quantitatively explored the welfare effect of introducing optimal capital taxation prescribed by the inverse Euler equation. The evolution of agents' private information is governed by an exogenous process in our paper as in much of the DC/NDPF literature. Makris and Pavan (2016) considered an environment in which agents' privately informed productivity evolves endogenously as a result of learning by doing. They addressed dynamic optimal taxation, showing that some of the results established on the basis of exogenous processes may be overturned.

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 a dynamic Mirrleesian economy. Section 3 proves the nonnegative martingale property. We then apply this property to the characterization of constrained efficient intertemporal aggregate wedges in Section 4 and the extension of immiseration to a more general environment in Section 5. Section 6 concludes.

2 Model economy

We consider a dynamic neoclassical 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. 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 either finitely lived ($T < \infty$) or infinitely lived ($T = \infty$) agents with a unit measure.

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, 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.

The period utility of agents is given by[4]

[4]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)) or quasilinear payoffs (see Bergemann and Pavan (2015), and Pavan (2016)).

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

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 $\lim_{c\to 0} u'(c) = +\infty$. To facilitate and simplify the analysis, we make the following assumption throughout the paper:

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.

Agents are assumed to maximize their expected discounted utility

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

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 $L_t = \sum_{\theta^t} l_t(\theta^t)\pi_t(\theta^t)$ 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 $C_t = \sum_{\theta^t} c_t(\theta^t)\pi_t(\theta^t)$ 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 that offers each agent an insurance contract with commitment.[5] 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 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 when 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 the temporary IC constraints, each of which rules out one-shot deviations at a time:

[5] As time passes, information about agent types (skills) may be revealed. 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).

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

where $(\theta^s; \widehat{\theta}_t)$ denotes $(\theta_0, \theta_1, ..., \theta_{t-1}, \widehat{\theta}_t, \theta_{t+1}, ..., \theta_s)$. However, the proof of Fernandes and Phelan (2000) builds on bounded period utility and allowing for unbounded period utility is important for our long-run analysis. Appendix A.1 shows that, even if period utility is unbounded, the following condition can uphold the result of Fernandes and Phelan (2000):[6]

$$
\lim_{t\to\infty} \sup_{\delta} \beta^t w_t(\delta(\theta^t)) = 0, \qquad (5)
$$

where $w_t(\delta(\theta^t)) \equiv \sum_{s=1}^{\infty} \sum_{\theta^{t+s}|\theta^t} \beta^s \pi_{t+s}(\theta^{t+s}|\theta^t) \left[u\left(c_{t+s}(\delta(\theta^{t+s}))\right) - v\left(\frac{l_{t+s}(\delta(\theta^{t+s}))}{\theta_{t+s}}\right)\right]$ and $\delta(.)$ denotes a reporting strategy of agents.

While $u(.)$ may be unbounded, we consider only bounded $v(.)$ under Assmption 1 (which is a reasonable imposition, say, all agents are endowed with a unit of time each time period). Thus, if T is finite or $T = \infty$ with bounded $u(.)$, it is evident that condition (5) will be satisfied. If $T = \infty$ but with unbounded $u(.)$, Golosov, Tsyvinski, and Werquin (2016, section 2.4.2) showed that condition (5) will still be satisfied if $u(.)$ takes the constant relative risk aversion (CRRA) form (including a logarithmic utility function).[7] In our applications, we do assume that $u(.)$ takes the CRRA form; see Assumption 2 in Section 4.

The replacement of the lifetime IC constraints (3) by a sequence of the temporal IC constraints (4) is important to our analysis, which will be clear later on.

[6]Sleet and Yeltekin (2010) gave the same condition but without its formal proof.

[7]Golosov, Tsyvinski, and Werquin (2016) addressed the question of whether the solution from the Bellman equation achieves the supremum of the primal maximization problem in dynamic Mirrleesian economies. It turns out that condition (5) can ensure a positive answer; see Golosov, Tsyvinski, and Werquin (2016, propositon 4).

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 sequence 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 temporal IC constraints (4) and $\{\mu_t\}$ on the resource constraints (2). Incorporating these constraints into the objective (1) yields the following Lagrangian:

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

If $T < \infty$, the Lagrangian L is well defined; however, this may not be true if $T \to \infty$. When $T \to \infty$, the Lagrange multipliers in the Lagrangian L belong to an infinite dimensional space since the set of constraints is infinity as $T \to \infty$. A question arises: Can these multipliers be represented as a summable sequence of real numbers in the infinite dimensional space? Papers including Dechert (1982), Rustichini (1998), and Le Van and Cagri Saglam (2004) have addressed the question. They extended the Lagrangian from finite to infinite dimensional spaces and provided sufficient conditions for a positive answer. Utilizing results of these papers, Appendix A.2 presents a justification for the Lagrangian L in which multipliers can be represented as a summable sequence of real numbers.

With the imposition of Assumption 1, Appendix A.3 shows that one can rewrite the "raw" Lagrangian L and express it as

$$
\begin{aligned}
\mathcal{L} \;=\; & \sum_{t=0}^{T}\sum_{\theta^t} \beta^t \pi_t(\theta^t)\left[\phi_c(\theta^t) u\left(c_t(\theta^t)\right) - \phi_l(\theta^t) v\left(\frac{l_t(\theta^t)}{\theta_t}\right)\right] \\
& + \sum_{t=0}^{T} \mu_t \left[F(K_t, L_t) + (1-\delta)K_t - C_t - K_{t+1}\right],
\end{aligned} \tag{6}
$$

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

$$\phi_c(\theta^t) = 1 + \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \left[\psi(\theta^{s-1}, \theta_s, \widehat{\theta}_s) - \frac{\pi_t(\theta^t; \widehat{\theta}_s)}{\pi_t(\theta^t)} \psi(\theta^{s-1}, \widehat{\theta}_s, \theta_s) \right], \tag{7}$$

$$\begin{aligned} \phi_l(\theta^t) &= 1 + \sum_{s=0}^{t} \sum_{\widehat{\theta}_s} \psi(\theta^{s-1}, \theta_s, \widehat{\theta}_s) - \sum_{s=0}^{t-1} \sum_{\widehat{\theta}_s} \frac{\pi_t(\theta^t; \widehat{\theta}_s)}{\pi_t(\theta^t)} \psi(\theta^{s-1}, \widehat{\theta}_s, \theta_s) \\ &\quad - \sum_{\widehat{\theta}_t} \frac{\pi(\widehat{\theta}_t | \theta_{t-1})}{\pi(\theta_t | \theta_{t-1})} \omega(\theta_t, \widehat{\theta}_t) \psi(\theta^{t-1}, \widehat{\theta}_t, \theta_t), \end{aligned} \tag{8}$$

where $\omega(\theta_t, \widehat{\theta}_t) = \left(\frac{\theta_t}{\widehat{\theta}_t}\right)^\gamma$. One can construct a saddle-point functional equation (an analog of the Bellman equation) to solve for the problem represented by the Lagrangian $\mathcal{L}$ with its associated motions (7)-(8). This approach is known as the "recursive Lagrangian" and was proposed and developed by Marcet and Marimon (1999, 2016) for solving dynamic incentive problems. As they explained, given backward state variables and shocks at time t, the accumulative multiplier on the forward-looking constraints at time t such as $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ in (7)-(8) is all that needs to be remembered from the past by the planner at time t according to the approach. Marcet and Marimon (2016) provided a sufficiency condition guaranteeing that solutions from the recursive Lagrangian are solutions to the original planning problem. This condition is satisfied as long as the value function of the saddle-point functional equation is differentiable in the accumulative multiplier, a condition that is commonly met or imposed in most economic applications.[8] It is worth noting that our planning problem represented by (1), (2), and (4) fits well into the structure of Example 1 (intertemporal participation constraints) considered by Marimon and Werner (2016). For more on the approach, see remarks in Section 3.2.

2.2.3 Interpretation of (7) and (8)

Consider $\phi_c(\theta^t)$ in (7). 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

[8] If the condition fails to hold, Lagrange multipliers associated with binding constraints may not be unique. This feature of non-uniqueness is a manifestation of a more general problem, which is not confined to the recursive Lagrangian method. See Marimon and Werner (2016) for discussions and their resolution of the problem.

$\widehat{\theta}_s$, 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 (7) with a plus sign, whereas the multiplier $\psi(\theta^{s-1}, \widehat{\theta}_s, \theta_s)$ enters (7) 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 in $\phi_c(\theta^t)$ are to enforce IC constraints in response to the evolution of skill shocks over time. 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 the objective (6).[9] A similar interpretation applies to $\phi_l(\theta^t)$ in (8).

As is clear from (7) and (8), the Pareto-Negishi weight at time t is a function of all past realized values of the multipliers $\psi(.)$ and so it encodes the frequency and severeness of the binding of temporal IC constraints across states and over time. It is worth mentioning that the evolution of $\{\phi_c(\theta^t), \phi_l(\theta^t)\}$ in (7) and (8) represents a generalization of Eq. (36) on Mirrleesian taxation 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 $\phi_c(\theta^{-1}) = 1$, (7) can be expressed recursively as

$$\phi_c(\theta^t) = \phi_c(\theta^{t-1}) + \varepsilon(\theta^t), \tag{9}$$

where

$$\begin{aligned} \varepsilon(\theta^t) &= \sum_{\widehat{\theta}_t} \left[\psi(\theta^{t-1}, \theta_t, \widehat{\theta}_t) - \psi(\theta^{t-1}, \widehat{\theta}_t, \theta_t) \frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})} \right] \\ &\quad - \sum_{\widehat{\theta}_{t-1}} \frac{\pi(\widehat{\theta}_{t-1}|\theta_{t-2})}{\pi(\theta_{t-1}|\theta_{t-2})} \psi(\theta^{t-2}, \widehat{\theta}_{t-1}, \theta_{t-1}) \left[\frac{\pi(\theta_t|\widehat{\theta}_{t-1})}{\pi(\theta_t|\theta_{t-1})} - 1 \right]. \end{aligned} \tag{10}$$

The first summation in (10) is anticipated in view of (7). 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

[9] Multiplying both sides of (7) by $\pi_t(\theta^t)$, we see that the term $\pi_t(\theta^t)\phi_c(\theta^t)$ given in the objective (6) stems from (i) the first term of L (the original objective) and (ii) the second term of L (the IC constraints).

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 (10) then serves as a statistical reference for the planner. 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 (9). 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 be 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 (9). This kind of rewards and punishments is ex post in nature and due to the across-time link of Markov idiosyncratic shocks in agent types. 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 (today) for any agent who deviated at time $t-1$ (yesterday). As explained by Fernandes and Phelan (2000), such threat keeping would be unnecessary for the planner if idiosyncratic shocks were i.i.d. so that there were no links in agent types across time periods.

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 to our interpretation for the second summation of (10).

This completes the description of our model.

3 Nonnegative martingale

This section proves a powerful property of $\phi_c(\theta^t)$ over time. It is worth emphasizing that, distinct from the extant literature, this property is derived directly from the IC constraints and so independent of the optimization of the social planning problem.[10]

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

[10]It should be noted that our proof of Lemma 1 did use the optimal condition (13) to show that $\phi_c(\theta^t) \geq 0$ for all θ^t and t. However, this property is weak in that any Pareto social welfare function will satisfy it. For comparisons with the extant literature, see Section 3.2.

its evolution in (7) satisfies

$$\sum_{\theta_t} \phi_c(\theta^t)\pi(\theta_t|\theta_{t-1}) = \phi_c(\theta^{t-1}). \tag{11}$$

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

Proof. See Appendix A.4. □

By the property of (11), the term $\varepsilon(\theta^t)$ in (9) is an innovation orthogonal to $\phi_c(\theta^{t-1})$. We now explain why $\varepsilon(\theta^t)$ is an innovation or, equivalently, why (11) 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 (7) and (8), we can express $\phi_l(\theta^t)$ as

$$\phi_l(\theta^t) = \phi_c(\theta^t) + \sum_{\widehat{\theta}_t} \frac{\pi(\widehat{\theta}_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})} \psi(\theta^{t-1}, \widehat{\theta}_t, \theta_t) \left(1 - \omega(\theta_t, \widehat{\theta}_t)\right), \tag{12}$$

which describes the relationship between ϕ_c and ϕ_l. Since the second term of (12) 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 $\pi(\theta_t|x(\theta^{t-1})) > 0$ 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.5. □

The stochastic process described in Lemma 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 $\pi(\theta_t|x(\theta^{t-1})) = \pi(\theta_t)$.
- Case 2: a first-order Markov process, where $x(\theta^{t-1}) = \theta_{t-1}$ and $\pi(\theta_t|x(\theta^{t-1})) = \pi(\theta_t|\theta_{t-1})$; 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 $\pi(\theta_t|x(\theta^{t-1})) = \pi(\theta_t|\theta_{t-1}, \theta_{t-2})$.

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

3.1 Inverse Euler equation

A central finding on capital taxation in NDPF is the so-called IEE.[11] Here we show that there exists an intimate relationship between the nonnegative martingale of ϕ_c and the IEE.

Maximizing the Lagrangian (6) subject to (7)-(8) yields the first-order conditions[12]

$$\beta^t \phi_c(\theta^t) u'\left(c_t(\theta^t)\right) = \mu_t, \tag{13}$$

$$\beta^t \phi_l(\theta^t) v'\left(\frac{l_t(\theta^t)}{\theta_t}\right) \frac{1}{\theta_t} = \mu_t F_{Lt}, \tag{14}$$

$$\mu_t = \mu_{t+1}\left(F_{K,t+1} + 1 - \delta\right). \tag{15}$$

The optimal conditions (13)-(15) plus the resource constraints (2) characterize the constrained efficient individual allocation $\{c_t(\theta^t), l_t(\theta^t)\}$. The optimal conditions (13)-(15) show that the constrained efficient individual allocation depends on the Pareto weights $\{\phi_c, \phi_l\}$, which are determined by (7) and (8). If the friction of private information were absent, $\phi_c(\theta^t) = \phi_l(\theta^t) = 1$ would hold all the time and (13)-(15) would reduce to the familiar first-order conditions.

Combining Eqs. (13) and (15) gives

$$\frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)} = \frac{\beta\left(F_{K,t+1} + 1 - \delta\right)}{u'\left(c_t(\theta^t)\right)} \frac{\phi_c(\theta^{t+1})}{\phi_c(\theta^t)}, \tag{16}$$

which in turn gives

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

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:

$$\sum_{\theta_{t+1}} \frac{1}{u'\left(c_{t+1}(\theta^{t+1})\right)} \pi(\theta_{t+1}|\theta_t) = \frac{\beta\left(F_{K,t+1} + 1 - \delta\right)}{u'\left(c_t(\theta^t)\right)}. \tag{17}$$

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

[12] 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. For more details, see Golosov, Tsyvinski, and Werquin (2016).

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. 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.[13] This method is powerful, but it may be blind about the underlying reason for the result. We instead rely on Lemma 1, which makes clearer its origin. If $\phi_c(\theta^t) = 1$ all the time (i.e., there is no information friction), we see from (16) that EE rather than IEE would hold.

3.2 Some remarks

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. However, their extension builds on the restriction that $\beta R = 1$.[14] Kocherlakota (2010, chapter 3) derived a similar result under a more general environment, but it requires the imposition of the restriction $\beta R = 1$ as well. Intuitively, the planner's marginal cost of the delivery is given by $1/u'(.)$ and her desire to smooth the marginal cost of utility production over time implies the IEE. We can see from (17) that the IEE will lead to a martingale process of $1/u'(.)$ only if the restriction $\beta R = 1$ is imposed. By contrast, our martingale result (Lemmas 1 and 2) does not build on the restriction $\beta R = 1$. Furthermore, it holds under Markov or more general stochastic processes and in a neoclassical production economy.

Golosov, Tsyvinski, and Werquin (2016) deemed that the feature of a martingale is a manifestation of the same general principle that underlies cost smoothing over time *at the optimum*. Other

[13]See Rogerson (1985); Golosov, Kocherlakota, and Tsyvinski (2003); Golosov, Tsyvinski, and Werning (2007); and Kocherlakota (2010).

[14]Without the restriction, it becomes submartingale or supermartingale, depending on whether $\beta R \geq 1$ or $\beta R \leq 1$; see Farhi and Werning (2013, p. 616, Proposition 1 and the proof of Proposition 5).

famous examples include consumption smoothing in the permanent income hypothesis (Hall, 1978) and tax smoothing in public finance (Barro, 1979). However, it should be emphasized that the nonnegative martingale of our Lemma 1 is a characterization of IC constraints, independent of the optimal conditions given by (13)-(15).

Fernandes and Phelan (2000) extended privately informed idiosyncratic shocks from i.i.d. to Markov. They solved for the resulting Mirrleesian problem by the recursive primal approach or the so-called promised utility approach, emphasizing threat keeping as well as promise keeping in the enforcement of agents' IC constraints in the Markovian environment. Messner, Pavoni, and Sleet (2016) envisioned a generalized version of the recursive Lagrangian and they called the approach the "recursive dual" as a contrast to the recursive primal used by Fernandes and Phelan (2000) and others. Both Messner, Pavoni, and Sleet (2016) and Marcet and Marimon (2016) emphasized that the dual approach is particularly useful relative to the primal approach if constraints involve forward-looking state variables such as in our case. A reason is that it involves the complication of pinning down the initial forward-looking state in the recursive primal problem, while the complication is often absent in the recursive dual problem. For example, consider our planning problem. By the model specification, we have $\phi_c(\theta^t) = \phi_l(\theta^t) = 1$ if $t = 0$. However, it is not so easy to identify the optimal initial value of promised utility at $t = 0$ if adopting the recursive primal approach. More generally, the complication of the primal approach arises from the fact that its set of the feasible state space is defined implicitly and need be recovered as part of the solution to the problem.[15] The recursive Lagrangian sidesteps this complication.

Finally, in the continuous time framework of incentive problems, papers such as Sannikov (2008, 2014) start the moral hazard analysis by assuming that the stochastic process of output is a martingale, for example, a Brownian motion. Golosov, Tsyvinski, and Werquin (2016) called the approach the "martingale method." As they exposed, the martingale assumption allows for exploiting some fundamental theorems such as the Martingale representation theorem to facilitate the analysis of the problems. It should be noted that we do not assume a martingale to begin with, and that the nonnegative martingale of our Lemma 1 is not imposed from the beginning, but a property derived from agents' temporal IC constraints.

For the rest of the paper, we apply the nonnegative martingale finding of Lemma 1 together with

[15] See Proposition 8 in Golosov, Tsyvinski, and Werquin (2016).

the first-order conditions (13)-(15) to (i) the characterization of constrained efficient intertemporal wedges in the aggregate, and (ii) the extension of immiseration to a more general environment.

4 Intertemporal wedges in the aggregate

The focus of the DC/NDPF literature has been on the micro individual allocation[16] and, therefore, it is not surprising that relatively little is known about the general property of macro aggregate variables at constrained efficient allocations. 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 incentives at the micro individual level. In this section, we apply Lemma 1 to shed light on this question.

4.1 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 IES. As will be seen, whether the 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), and Golosov, Troshkin, and Tsyvinski (2016).

Under Assumptions 1 and 2, we obtain the following from (13) and (14):

$$\frac{c_t(\theta^t)}{C_t} = \frac{\phi_c(\theta^t)^{\frac{1}{\sigma}}}{H_t}, \tag{18}$$

[16] The IEE is a leading example.

$$\frac{l_t(\theta^t)}{L_t} = \frac{\phi_l(\theta^t)^{\frac{1}{1-\gamma}} \theta_t^{\frac{-\gamma}{1-\gamma}}}{J_t}, \tag{19}$$

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. (18) and (19) 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.

4.2 Intertemporal aggregate wedges

Substituting the derived c_t and l_t of the consumption and labor sharing rules (18)-(19) into the individual optimal conditions, (13)-(15), gives rise to

$$C_t^{-\sigma} = \beta C_{t+1}^{-\sigma} \left(\frac{H_{t+1}}{H_t} \right)^{\sigma} (F_{K,t+1} + 1 - \delta), \tag{20}$$

$$\frac{L_t^{\gamma-1}}{C_t^{-\sigma}} = \left(\frac{H_t^{\sigma}}{J_t^{1-\gamma}} \right) F_{Lt}. \tag{21}$$

The constrained efficient macro aggregate allocation is characterized by (20) and (21) together with the resource constraints (2). Contrasting (20)-(21) with (13)-(14) 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 wedge associated with (20).

Let us explain the result of (20). Kocherlakota (2010, p. 49) provided an intuition for the IEE (17) as follows. 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 (17) since $\rho'(u) = 1/u'\left(\rho(u)\right)$. We can understand the result of (20) 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 (18), $c_t(\theta^t) = \frac{\phi_c(\theta^t)^{\frac{1}{\sigma}}}{H_t} C_t$ 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 (20). Like the IEE, the condition (20) 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, (20) describes her intertemporal behavior over the macro aggregate allocation.

In the first-best economy without frictions, it is straightforward to see that $H_{t+1}/H_t = 1$ for all t and, therefore, (20) would reduce to the familiar Euler equation. We can thus define the intertemporal aggregate wedge associated with (20) by its deviation from the first-best:

$$\text{Intertemporal aggregate wedge} : \tau_{Kt} \equiv 1 - \left(\frac{H_{t+1}}{H_t}\right)^\sigma. \tag{22}$$

Note that while there does not exist a representative agent in the aggregate for our friction-loaded economy in competitive equilibrium, there does exist one in the first-best (a frictionless or complete markets) economy; see Guvenen (2011). Later, we expose the relation between the intertemporal *individual* wedge implied by the IEE and the optimal intertemporal *aggregate* wedge according to (22).

In the next subsection, we study the constrained efficient intertemporal aggregate allocation characterized by (20) through the lens of the aggregate wedge defined by (22).

4.3 Constrained efficient intertemporal aggregate wedges

Applying Lemma 1, this subsection characterizes the property of the constrained efficient intertemporal aggregate wedge, τ_{Kt}^{*}, for all $t < \infty$.

It is clear from the definition of the intertemporal aggregate wedge (22) that the ratio H_{t+1}/H_t is crucial. The next lemma characterizes this ratio in the short run.

Lemma 3. *$H_{t+1}/H_t \lesseqqgtr 1$ for all $t < \infty$ if $\sigma \gtreqqless 1$.*

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[17]

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

Combining the above results gives

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

which in turn gives

[17]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.

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

□

From (22), Lemma 3 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 1. $\tau^*_{Kt} \gtreqless 0$ *for all* $t < \infty$ *if* $\sigma \gtreqless 1$.

The quantity of the 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 the IES to be well above 1, other works including those of Hall (1988), Campbell and Mankiw (1989) and, Campbell (1999) find the IES to be close to 0. Proposition 1 prescribes $\tau^*_{Kt} \gtrless 0$ according to whether the IES is inelastic or elastic.

As shown earlier, the marginal cost of utility production for all agents at time t is given by $\sum_{\theta_t} c_t(\theta^t)^\sigma \pi_t(\theta_t) = C_t^\sigma H_t^{-\sigma}$. The term $H_t^{-\sigma}$ (in addition to the standard term C_t^σ) is related to the cost of spreading consumption across agents over time to satisfy the IC constraints. With $\sigma > 1$ ($\sigma < 1$), the cost of spreading consumption across agents is a convex (concave) function, meaning that the benefit of having one extra aggregate consumption is increasing (decreasing). As a result, the planner would like to "front-load" ("back-load") C_t and so $\tau^*_{Kt} \gtrless 0$. If $\sigma = 1$, the cost is linear and hence neither front-loading nor back-loading is desirable.

4.4 Relation with intertemporal individual wedges

Following Golosov, Tsyvinski, and Werning (2007), we define the intertemporal individual wedge as

$$\tau_{kt}(\theta^t) \equiv 1 - \frac{u'(c_t(\theta^t))}{\beta R_{t+1} \sum_{\theta_{t+1}} u'(c_{t+1}(\theta^{t+1}))\pi(\theta_{t+1}|\theta_t)}, \tag{23}$$

where $R_{t+1} = 1 + F_{K,t+1} - \delta$ is the gross rate of return from saving in competitive equilibrium. This wedge is 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, $\tau_{kt}(\theta^t) = 0$ will hold in competitive equilibrium.

From (13) and (15), (23) gives

$$\frac{1}{1 - \tau_{kt}^*(\theta^t)} = \sum_{\theta_{t+1}} \frac{\phi_c(\theta^t)}{\phi_c(\theta^{t+1})}\pi(\theta_{t+1}|\theta_t). \tag{24}$$

Using (20), (24) can be re-expressed as:

$$\frac{1}{1 - \tau_{kt}^*(\theta^t)} = \beta\left(F_{K,t+1} + 1 - \delta\right)\left(\frac{H_{t+1}}{H_t}\right)^{\sigma}\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma}\sum_{\theta_{t+1}} \frac{\phi_c(\theta^t)}{\phi_c(\theta^{t+1})}\pi(\theta_{t+1}|\theta_t),$$

which leads to

$$\frac{1}{1 - \tau_{kt}^*(\theta^t)} = \frac{1}{1 - \tau_{Kt}^*}\sum_{\theta_{t+1}}\left(\frac{c_t(\theta^t)/C_t}{c_{t+1}(\theta^{t+1})/C_{t+1}}\right)^{\sigma}\pi(\theta_{t+1}|\theta_t), \tag{25}$$

where we have used (18), (20) and (22). The above equation shows that the constrained efficient intertemporal wedge in the individual $\tau_{kt}^*(\theta^t)$ can be decomposed into two parts: (i) the optimal distortion from the first-best in the aggregate and (ii) a spreading of the aggregate consumption across agents over time. The first part is precisely captured by τ_{Kt}^*. Note that if $\tau_{Kt}^* = 0$, then $\tau_{kt}^*(\theta^t)$ is completely dictated by the second spreading part. This decomposition establishes a relationship between the intertemporal wedge in the aggregate and that in the individual.

In the case of logarithmic utility with fixed labor supply, Farhi and Werning (2012) showed that the social planning problem in their setup can be decomposed into an idiosyncratic planning problem and an aggregate planning problem, where the aggregate one is just a standard neoclassical growth model. Although somewhat similar, their decomposition differs from ours in two dimensions. First, while they consider labor supply as fixed and given, we do not. Second, there is no friction in their aggregate planning problem, whereas there is in ours. Atkeson and Lucas (1992, 1995) envisioned an economy having many of the so-called component planners whose job is to transfer aggregate resources across time for the agents in their "jurisdiction" and, at the same time, to meet these agents' incentive constraints in distributing the aggregate resources among agents. An important departure from the literature in their analysis is to explicitly make a distinction between the aggregate allocation and the distribution of the aggregate allocation. The decomposition of Eq. (25) makes a similar distinction for the constrained efficient intertemporal wedges.

5 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 the followings: (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.[18] 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 demonstrate that immiseration still holds under Markov or more general stochastic processes and in a neoclassical production economy.

Note that the IEE (17) will become a nonnegative martingale if (i) the term $F_{K,t+1} + 1 - \delta$ in (17) is replaced by a constant gross rate of return R, and (ii) the restriction $\beta R = 1$ is imposed. Assuming (i) and (ii), Kocherlakota (2010, Theorem 3.2) applied the martingale convergence theorem to the IEE. Our application of the martingale convergence theorem in the following proposition builds on Lemma 1 and imposes neither (i) nor (ii).

Proposition 2. *For almost surely all θ^t, $\lim_{t\to\infty} \phi_c(\theta^t) = 0$ at constrained efficient allocations in*

[18] Kocherlakota (2010, p. 68) provided an example to illustrate the coexistence of (i) and (ii).

our dynamic Mirrleesian economy.

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 (9) that $\lim_{t\to\infty}\varepsilon(\theta^t) = 0$, which in turn implies that almost surely all multipliers $\psi(.)$ in (10) are equal to 0 in the limit. Hence, $\lim_{t\to\infty}\phi_l(\theta^t) = \phi^*(\theta^t)$ as well.

From Eqs. (13)-(14), the convergence of $\phi_l(\theta^t)$ and $\phi_c(\theta^t)$ suggest that as $t\to\infty$,

$$\begin{aligned}
\left|c_t(\theta^{t-1},\theta_t) - c_t(\theta^{t-1},\hat{\theta}_t)\right| &\to 0 \text{ for any } \hat{\theta}_t \text{ and } \theta_t,\\
\left|c_s(\theta^s) - c_s(\theta^s;\widehat{\theta}_t)\right| &\to 0 \text{ for any } s>t,\ \theta^s \text{ and } \widehat{\theta}_t,\\
\left|l_s(\theta^s) - l_s(\theta^s;\widehat{\theta}_t)\right| &\to 0 \text{ for any } s>t,\ \theta^s \text{ and } \widehat{\theta}_t.
\end{aligned}$$

The convergence result suggests that all terms on the left and the right sides of the temporal IC constraint (4) converge, except for $l_t(\theta^{t-1},\widehat{\theta}_t)$ and $l_t(\theta^{t-1},\theta_t)$. Hence, the temporal IC constraints (4) of agents (θ^{t-1},θ_t) and of agents $(\theta^{t-1},\hat{\theta}_t)$ can be rewritten as

$$-v\left(\frac{l_t(\theta^{t-1},\theta_t)}{\theta_t}\right) \geq -v\left(\frac{l_t(\theta^{t-1},\hat{\theta}_t)}{\theta_t}\right) + \widetilde{\varepsilon}_{t,1},$$

$$-v\left(\frac{l_t(\theta^{t-1},\hat{\theta}_t)}{\hat{\theta}_t}\right) \geq -v\left(\frac{l_t(\theta^{t-1},\theta_t)}{\hat{\theta}_t}\right) + \widetilde{\varepsilon}_{t,2},$$

where both $\widetilde{\varepsilon}_{t,1}$ and $\widetilde{\varepsilon}_{t,2}\to 0$ as $t\to\infty$.

With Assumption 1 and Eq. (19), the inequalities above become

$$\left(\frac{\phi^*(\theta^t)^{1/(1-\gamma)}\theta_{t+1}^{-\gamma/(1-\gamma)}}{J_{t+1}}\frac{L_{t+1}}{\theta_{t+1}}\right)^{\gamma} \leq \left(\frac{\phi^*(\theta^t)^{1/(1-\gamma)}\hat{\theta}_{t+1}^{-\gamma/(1-\gamma)}}{J_{t+1}}\frac{L_{t+1}}{\theta_{t+1}}\right)^{\gamma} + \widetilde{\varepsilon}_{t,1},$$

$$\left(\frac{\phi^*(\theta^t)^{1/(1-\gamma)}\hat{\theta}_{t+1}^{-\gamma/(1-\gamma)}}{J_{t+1}}\frac{L_{t+1}}{\hat{\theta}_{t+1}}\right)^{\gamma} \leq \left(\frac{\phi^*(\theta^t)^{1/(1-\gamma)}\theta_{t+1}^{-\gamma/(1-\gamma)}}{J_{t+1}}\frac{L_{t+1}}{\hat{\theta}_{t+1}}\right)^{\gamma} + \widetilde{\varepsilon}_{t,2}.$$

Pick $\widetilde{\varepsilon}_{t,1}$ and $\widetilde{\varepsilon}_{t,2}$ sufficiently small. Then if $\phi^*(\theta^t) > 0$, the above inequalities lead to a contradiction with $\theta_{t+1} \neq \hat{\theta}_{t+1}$ since they give

$$\left(\theta_{t+1}^{-\gamma/(1-\gamma)}\right)^{\gamma} \leq \left(\hat{\theta}_{t+1}^{-\gamma/(1-\gamma)}\right)^{\gamma},$$
$$\left(\hat{\theta}_{t+1}^{-\gamma/(1-\gamma)}\right)^{\gamma} \leq \left(\theta_{t+1}^{-\gamma/(1-\gamma)}\right)^{\gamma}.$$

By contrast, if $\phi^*(\theta^t) = 0$, no contradiction arises. □

From (13), $\lim_{t\to\infty} \phi_c(\theta^t) = 0$ implies that $\lim_{c\to 0} u'(c_t(\theta^t)) = +\infty$. The intuition underlying immiseration is as follows. To provide incentives for more productive people to produce more, the planner needs to spread out consumption across heterogeneous agents over time. When $\phi_c(\theta^t)$ is converging, the planner accomplishes this task by immiseration: Let the consumption of an increasingly large fraction of agents go to 0 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.

As the proof of Proposition 2 shows, our 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.

5.1 Discussion

What are implications of immiseration for constrained efficient intertemporal wedges in the long run or steady state?

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 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 constrained efficient aggregate wedges τ^*_{Kt} or individual wedges $\tau^*_{kt}(\theta^t)$ will converge to 0 in the limit, even if they are non-zero in the short run. We briefly explore whether immiseration can answer the question.

Let us first consider the limit behavior of τ^*_{Kt}. Immiseration implies two opposite processes: (i) an increasingly large fraction of agents consume a decreasingly small amount of consumption over time, and (ii) a decreasingly small fraction of agents consume an increasingly large amount of consumption over time. From the proof of Lemma 1, we have $\sum_{\theta^t} \phi_c(\theta^t)\pi_t(\theta^t) = 1$. Given this result and Proposition 2, we then 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 $\tilde{\pi}(t)$ denotes the fraction of agents who experience the second process of immiseration, and $\tilde{\phi}_c(t)$, the average Pareto-Negishi weight associated with these agents. This result implies $\lim_{t\to\infty} \tilde{\pi}(t) = \lim_{t\to\infty} \tilde{\phi}_c(t)^{-1}$,[19] which in turn implies

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

where the second equality holds because $\phi_c(\theta^t)$ other than $\tilde{\phi}_c(t)$ all go to zero in the limit. From the inertemporal aggregate wedge defined by (22), we know that $\lim_{t\to\infty} \tau^*_{Kt} = 0$ if and only if $\lim_{t\to\infty} H_{t+1}/H_t = 1$. We have from (27)

$$\lim_{t\to\infty} \frac{H_{t+1}}{H_t} = \frac{\lim_{t\to\infty} \tilde{\phi}_c(t+1)^{\frac{1-\sigma}{\sigma}}}{\lim_{t\to\infty} \tilde{\phi}_c(t)^{\frac{1-\sigma}{\sigma}}} = \left[\lim_{t\to\infty} \frac{\tilde{\phi}_c(t+1)}{\tilde{\phi}_c(t)}\right]^{\frac{1-\sigma}{\sigma}}. \tag{28}$$

If $\sigma = 1$, it is clear from (28) that $\lim_{t\to\infty} H_{t+1}/H_t = 1$ holds. However, since $\tilde{\pi}(t) \to 0$ as $t \to \infty$ according to Proposition 2, we have $\lim_{t\to\infty} \tilde{\phi}_c(t) = \infty$ from (26). Thus, $\tilde{\phi}_c(t+1)/\tilde{\phi}_c(t)$ in the limit takes on the so-called indeterminate form with $\lim_{t\to\infty} \tilde{\phi}_c(t+1)/\tilde{\phi}_c(t) = \infty/\infty$; as such, the result of (28) by itself cannot tell us whether $\lim_{t\to\infty} H_{t+1}/H_t = 1$ holds as $\sigma \neq 1$.

The above finding may not be surprising. Proposition 2 implies that $\lim_{t\to\infty} \tilde{\pi}(t) = \lim_{t\to\infty} \tilde{\pi}(t+1) = 0$, which in turn implies from (26) that

$$\lim_{t\to\infty} [\tilde{\pi}(t) - \tilde{\pi}(t+1)] = \lim_{t\to\infty} \left[\frac{\frac{\tilde{\phi}_c(t+1)}{\tilde{\phi}_c(t)} - 1}{\tilde{\phi}_c(t+1)}\right] = 0.$$

[19] $\lim_{t\to\infty} \tilde{\phi}_c(t)\tilde{\pi}(t) = 1 \Rightarrow \lim_{t\to\infty} \tilde{\phi}_c(t) \lim_{t\to\infty} \tilde{\pi}(t) = 1 \Rightarrow$
$\lim_{t\to\infty} \tilde{\pi}(t) = \frac{1}{\lim_{t\to\infty} \tilde{\phi}_c(t)} = \lim_{t\to\infty} \left[\frac{1}{\tilde{\phi}_c(t)}\right] = \lim_{t\to\infty} \tilde{\phi}_c(t)^{-1}$.

Because $\lim_{t\to\infty} \tilde{\phi}_c(t+1) = \infty$, to uphold the result above, there is no requirement that $\tilde{\phi}_c(t+1)/\tilde{\phi}_c(t) = 1$ in the limit. In other words, the immiseration process of $\tilde{\pi}(t) \to 0$ and $\tilde{\phi}_c(t) \to \infty$ as $t \to \infty$ by itself is not sufficiently informative to enable us to determine whether $\lim_{t\to\infty} H_{t+1}/H_t = 1$ holds if $\sigma \neq 1$.

Note that if $H_{t+1}/H_t = 1$ holds in the limit, we will have $\beta R = 1$ in steady state according to (20), regardless of whether $\sigma \gtreqless 1$. Our analysis shows that we do have $\beta R = 1$ in steady state if $\sigma = 1$, but we cannot be sure whether $\beta R = 1$ holds in steady state if $\sigma \neq 1$. Starting from the IEE 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 on their numerical analysis to find that $\beta R \gtrless 1$ as $\sigma \gtrless 1$.[20] Our exposition above extends their result on the logarithmic utility case allowing for variable labor supply. As for the more general CRRA cases, their numerical finding is consistent with our Proposition 1 if aggregate allocations have converged to a large extent so that (20) becomes

$$1 = \beta R \left(\frac{H_{t+1}}{H_t} \right)^{\sigma}.$$

Since $H_{t+1}/H_t \lessgtr 1$ for all $t < \infty$ if $\sigma \gtrless 1$ according to Lemma 3, we then have $\beta R \gtrless 1$ as $\sigma \gtrless 1$. Of course, this analysis does not rule out the possibility that the result of $H_{t+1}/H_t \lessgtr 1$ if $\sigma \gtrless 1$ holds in the limit, and so the result of $\beta R \gtrless 1$ as $\sigma \gtrless 1$ could be indeed true in steady state.

We now turn to the limit behavior of $\tau^*_{kt}(\theta^t)$. Because $\lim_{t\to\infty} \phi_c(\theta^t) = \lim_{t\to\infty} \phi_c(\theta^{t+1}) = 0$ almost surely according to Proposition 2, the result of (24) is of little help in determining whether $\lim_{t\to\infty} \tau^*_{kt}(\theta^t) = 0$. Let us focus on the case of logarithmic utility with $\sigma = 1$. Since $\lim_{t\to\infty} \tau^*_{Kt} = 0$ if $\sigma = 1$, from (25) we have

$$\lim_{t\to\infty} \frac{1}{1 - \tau^*_{kt}(\theta^t)} = \lim_{t\to\infty} \sum_{\theta_{t+1}} \left(\frac{c_t(\theta^t)/C_t}{c_{t+1}(\theta^{t+1})/C_{t+1}} \right) \pi(\theta_{t+1}|\theta_t).$$

The value of $c_t(\theta^t)/C_t$ in the limit either goes to zero (if associated with the first process of immiseration) or goes to infinity (if associated with the second process of immiseration). As a result, immiseration by itself cannot tell us whether $\lim_{t\to\infty} \tau^*_{kt}(\theta^t) = 0$ holds even for the case of

[20] Does a steady state exist? We do not answer the question. However, the numerical exercise of Farhi and Werning (2012) seems to suggest it does.

$\sigma = 1$.

6 Conclusion

This paper considers a dynamic neoclassical economy in which privately informed idiosyncratic productivity shocks evolve stochastically over time. Instead of the conventional recursive approach in the DC/NDPF literature, we reformulate the social planning problem in a sequential fashion and decompose agents' lifetime IC constraints into a sequence of temporary ones. We encode the frequency and severeness of these temporal IC constraints by their associated Lagrange multipliers, showing that the accumulation of the Lagrange multipliers serves as stochastic Pareto-Negishi weights for efficient allocations and, more interestingly, its evolution on the consumption part is a nonnegative martingale. We apply our findings to the characterization of constrained efficient intertemporal wedges in the aggregate and the extension of immiseration to a more general environment.

The focus of our applications is on intertemporal rather than intratemporal aggregate 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 individual 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 in the aggregate in future work.

A Appendix

A.1 Equivalence of lifetime and temporal IC constraints

We show that an allocation over consumption and labor supply satisfies the lifetime IC constraints (3) if and only if it satisfies the sequence of temporary IC constraints (4).

That the lifetime IC constraints (3) implies the temporal IC constraints (4) is clear, since possible deviations in the lifetime include one-shot deviations in the temporal. To show the converse, we first prove a lemma.

Lemma 4. *Let an allocation $\{c_t(\theta^t), l_t(\theta^t)\}_{t,\theta^t}$ satisfy the temporary IC constraints (4). Then the following inequality holds:*

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

where $\theta_{t+1}^s = (\theta_{t+1},\theta_{t+2},...,\theta_s)$. In words, the inequality states that even if agents do not report honestly before time t (i.e., $\hat{\theta}^{t-1}\neq\theta^{t-1}$), it is still not profitable for agents to conduct a one-shot deviation from honesty at time t.

Proof. Consider agents whose types until time $t-1$ are given by $\hat{\theta}^{t-1}$ and from time t onward coincide with θ^t. Then agents' temporal IC constraints at time t are exactly given by the inequality stated in the lemma. □

Lemma 4 corresponds to Lemma 2.1 in Fernandes and Phelan (2000), who showed that if an allocation satisfies the lifetime IC constraint, then honesty from time t on is always the best policy for agents, even if agents do not report honestly before time t.

With Lemma 4 at hand, we go on to prove the converse. Our proof follows the logic in Fernandes and Phelan (2000, Theorem 2.1) and Golosov, Tsyvinski, and Werquin (2016, Lemma 3).

Multiplying both sides of the inequality in Lemma 4 by $\beta^t \pi_t(\theta^t)$ gives

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

Adding the term $\sum_{s=0}^{t-1} \sum_{\theta^t} \beta^s \pi_s(\theta^s) \left[u\left(c_s(\widehat{\theta}^s) \right) - v\left(\frac{l_s(\widehat{\theta}^s)}{\theta_s} \right) \right]$ to both sides of the above inequality yields

$$\begin{aligned}
& \sum_{s=0}^{t-1} \sum_{\theta^t} \beta^s \pi_s(\theta^s) \left[u\left(c_s(\widehat{\theta}^s) \right) - v\left(\frac{l_s(\widehat{\theta}^s)}{\theta_s} \right) \right] \\
& + \sum_{s=t}^{T} \sum_{\theta^s} \beta^s \pi_s(\theta^s) \left[u\left(c_s(\hat{\theta}^{t-1}, \theta^s_t) \right) - v\left(\frac{l_s(\hat{\theta}^{t-1}, \theta^s_t)}{\theta_s} \right) \right] \\
\geq \; & \sum_{s=0}^{t} \sum_{\theta^s} \beta^s \pi_s(\theta^s) \left[u\left(c_s(\widehat{\theta}^s) \right) - v\left(\frac{l_s(\widehat{\theta}^s)}{\theta_s} \right) \right] \\
& + \sum_{s=t+1}^{T} \sum_{\theta^s} \beta^s \pi_s(\theta^s) \left[u\left(c_s(\hat{\theta}^t, \theta^s_{t+1}) \right) - v\left(\frac{l_s(\hat{\theta}^t, \theta^s_{t+1})}{\theta_s} \right) \right], \forall \hat{\theta}^{t-1}, \widehat{\theta}_t.
\end{aligned}$$

This states that if an allocation satisfies the temporal IC constraints (4), then, from the perspective of time 0, a reporting strategy with honesty from time $t+1$ on (the right side of the inequality) cannot be better off than a reporting strategy with honesty from time t on (the left side of the inequality). Starting from time 0, induction with $T < \infty$ then implies

$$
\begin{aligned}
& \sum_{s=0}^{T}\sum_{\theta^s}\beta^s\pi_s(\theta^s)\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right] \\
\geq\ & \sum_{s=0}^{T}\sum_{\theta^s}\beta^s\pi_s(\theta^s)\left[u(c_s(\widehat{\theta}^s))-v\left(\frac{l_s(\widehat{\theta}^s)}{\theta_s}\right)\right],
\end{aligned} \tag{29}
$$

which states that if an allocation satisfies the temporal IC constraints (4), then, from the perspective of time 0, a reporting strategy with honesty up to time T (the left side of the inequality) cannot be worse off than any other reporting strategy up to time T (the right side of the inequality).

If $T < \infty$, the proof is done with the derivation of (29). Below we consider $T = \infty$.

We add the term $\beta^T \sum_{s=1}^{\infty}\sum_{\theta^s}\beta^s\pi_{T+s}(\theta^{T+s})\left[u\left(c_{T+s}(\theta^{T+s})\right)-v\left(\frac{l_{T+s}(\theta^{T+s})}{\theta_{T+s}}\right)\right]$ to both side of (29), and add and subtract the term $\beta^T \sum_{s=1}^{\infty}\sum_{\theta^s}\beta^s\pi_{T+s}(\theta^{T+s})\left[u(c_{T+s}(\widehat{\theta}^{T+s}))-v\left(\frac{l_{T+s}(\widehat{\theta}^{T+s})}{\theta_{T+s}}\right)\right]$ to and from the right side of (29). It then yields

$$
\begin{aligned}
& \sum_{s=0}^{\infty}\sum_{\theta^s}\beta^s\pi_s(\theta^s)\left[u\left(c_s(\theta^s)\right)-v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right] \\
\geq\ & \sum_{s=0}^{\infty}\sum_{\theta^s}\beta^s\pi_s(\theta^s)\left[u\left(c_s(\widehat{\theta}^s)\right)-v\left(\frac{l_s(\widehat{\theta}^s)}{\theta_s}\right)\right] \\
& +\beta^T\sum_{s=1}^{\infty}\sum_{\theta^{T+s}}\beta^s\pi_{T+s}(\theta^{T+s})\left[\left(u\left(c_{T+s}(\theta^{T+s})\right)-v\left(\frac{l_{T+s}(\theta^{T+s})}{\theta_{T+s}}\right)\right)-\left(u(c_{T+s}(\widehat{\theta}^{T+s}))-v\left(\frac{l_{T+s}(\widehat{\theta}^{T+s})}{\theta_{T+s}}\right)\right)\right].
\end{aligned}
$$

Note that if the last term of the above inequality vanishes as $T \to \infty$, then its right side converges to the right side of the lifetime IC constraints (3) with $T = \infty$ and, therefore, the above inequality reduces to the lifetime IC constraints (3). The condition (5) is sufficient to ensure that this last term converges to zero as $T \to \infty$.

A.2 Justifying Lagrangian L

Let ℓ^∞ be the set of sup-norm bounded sequences. Le Van and Cagri Saglam (2004) considered the following optimization problem:

$$
\min f(x) \text{ s.t. } g(x) \leq 0 \text{ with } x \in \ell^\infty,
$$

where $f : \ell^\infty \to \mathbb{R} \cup \{+\infty\}$, $g(x) = \{g_t(x)\}_{t=0}^{\infty}$ with each $g_t : \ell^\infty \to \mathbb{R} \cup \{+\infty\}$, and f and g_t are convex functions. They define $D = \{x \in \ell^\infty \mid f(x) < +\infty\}$ and $\Gamma = \{x \in \ell^\infty \mid g_t(x) < +\infty, \forall t\}$.

Let $x = \{c_t(\theta^t), l_t(\theta^t), K_{t+1}\}_{t=0}^{\infty}$ and $g(x) = \{g_{1t}(x), g_{2t}(x)\}_{t=0}^{\infty}$ with

$$g_{1t}(x) = V_t(\theta^t; \widehat{\theta}_t) - \left(u\left(c_t(\theta^t)\right) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) + \sum_{s=t+1}^{\infty} \sum_{\theta^s|\theta^t} \beta^{s-t} \pi_s(\theta^s|\theta^t) \left[u\left(c_s(\theta^s)\right) - v\left(\frac{l_s(\theta^s)}{\theta_s}\right)\right]\right),$$

$$g_{2t}(x) = C_t + K_{t+1} - F(K_t, L_t) - (1-\delta)K_t,$$

where $V_t(\theta^t; \widehat{\theta}_t)$ denotes the right-hand side of (4). To fit into the Le Van and Cagri Saglam (2004) optimization problem, we minimize $-EU$ instead of maximize EU of (1). However, because of the term $V_t(\theta^t; \widehat{\theta}_t)$, $g_{1t}(x)$ in our problem may not be a convex function as assumed in Le Van and Cagri Saglam (2004). As noted by Werning (2007, footnote 7) and exposed by Golosov, Tsyvinski, and Werquin (2016), a way of transforming our problem into a convex problem is to change variables from (c, l) into (u, v) and then minimize the resource cost associated with providing (u, v). (We do not repeat the details here.) Below we focus on the issue of presenting multipliers in the infinite dimensional space as a summable sequence of real numbers.

It is important to recognize that f and g_t are functions from ℓ^∞ to $\mathbb{R} \cup \{+\infty\}$ in the optimization problem. Dechert (1982) considered the situation where f and g_t are functions from ℓ^∞ to $\mathbb{R}$. Le Van and Cagri Saglam (2004) extended it to the situation where f and g_t are functions from ℓ^∞ to $\mathbb{R} \cup \{+\infty\}$. This extension is important for our analysis, in that f and g_t could go to $+\infty$ in the long run within our framework (namely, $-\log c$ goes to infinity as $c \to 0$ in the case of the so-called immiseration ; see Lemma 2 and its discussion).

To present multipliers in the infinite dimensional space as a summable sequence of real numbers, both Dechert (1982) and Le Van and Cagri Saglam (2004) put restrictions on the asymptotic behavior of the objective functional $f(x)$ and the constraint functions $g(x)$. For a pair $x, y \in \ell^\infty$ and $T \in \mathbb{N}$, define $x^T(x, y) = x_t$ if $t \leq T$ and $x^T(x, y) = y_t$ if $t > T$. Consider the following assumptions:

Assumption C. Continuity: $\lim_{T\to\infty} f(x^T(x, y)) = f(x)$.

Assumption S. Slater condition: $\exists x_0 \in \ell^\infty$ such that $\sup_t g_t(x_0) < 0$.

Assumption B. Uniform boundedness: $\exists M < \infty$ such that for all T large enough, $\| g(x^T(x, y)) \| \leq M$.

Assumption AI. Asymptotically insensitive: for all $\mathbb{N}$, $\lim_{t\to\infty} \left[g_t(x^{\mathbb{N}}(x, y)) - g_t(y)\right] = 0$.

Assumption ANA. Asymptotically non-anticipatory: $\lim_{T\to\infty} g_t(x^T(x,y)) = g_t(x)$, $\forall t$.

Restricting: (i) x in Assumption C such that $x \in D$, (ii) x_0 in Assumption S such that $x_0 \in D$, and (iii) x and y in Assumptions B, AI, and ANA such that $x, y \in \Gamma$ and $x^T(x,y) \in \Gamma$ for all T large enough, Le Van and Cagri Saglam (2004) showed that the above assumptions together ensure the representation of multipliers in the infinite dimensional space as a summable sequence of real numbers. They also showed that the solution to the optimization problem can be found by seeking the saddle point of the associated Lagrangian.

We now show that all the assumptions above are met by our model.

Assumption C. Continuity: $\lim_{T\to\infty} f(x^T(x,y)) = f(x)$.

Given

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

consider any $x', x'' \in D$ such that $\forall T$, $x^T(x', x'') \in D$. By the definition of $x^T(x', x'')$,

$$f\left(x^T(x', x'')\right) = -\sum_{t=0}^{T}\sum_{\theta^t} \beta^t \pi_t(\theta^t)\left[u\left(c'_t(\theta^t)\right) - v\left(\frac{l'_t(\theta^t)}{\theta_t}\right)\right] - \sum_{t=T+1}^{\infty}\sum_{\theta^t} \beta^t \pi_t(\theta^t)\left[u\left(c''_t(\theta^t)\right) - v\left(\frac{l''_t(\theta^t)}{\theta_t}\right)\right].$$

Assumption C requires that the second part of the above expression (i.e., the one involves x'') go to zero as $T \to \infty$. This requirement is satisfied under the condition (5).

Assumption ANA. Asymptotically non-anticipatory: $\lim_{T\to\infty} g_t(x^T(x,y)) = g_t(x)$, $\forall t$.

Except for applying to $g_{1t}(x^T(x,y))$ rather than $f(x^T(x,y))$, the logic of proving $\lim_{T\to\infty} g_{1t}(x^T(x,y)) = g_{1t}(x)$, $\forall t$ is the same as that of verifying Assumption C.

As to $\lim_{T\to\infty} g_{2t}(x^T(x,y)) = g_{2t}(x)$, $\forall t$, it is obvious.

Assumption AI. Asymptotically insensitive: for all $\mathbb{N}$, $\lim_{t\to\infty}\left[g_t(x^{\mathbb{N}}(x,y)) - g_t(y)\right] = 0$.

Given $\mathbb{N}$, it is clear that $g_t(x^{\mathbb{N}}(x,y)) = g_t(y)$ if t is large enough.

Assumption B. Uniform boundedness: $\exists M < \infty$ such that for all T large enough, $\| g(x^T(x,y) \| \leq M$.

By the definition of Γ, Assumption B is satisfied.

Assumption S. Slater condition: $\exists x_0 \in \ell^\infty$ such that $\sup_t g_t(x_0) < 0$.

This is a regularity condition or a maintained assumption of our model.

Finally, note that Le Van and Cagri Saglam (2004) results build on the premise that $x \in \ell^\infty$. However, because of the so-called immiseration (see Lemma 2), consumption for a zero measure of agents can go to infinity in the long run in our model; see Section 5. This violates the premise that $x \in \ell^\infty$.

Chari, Christiano, and Kehoe (1996) faced a problem similar to the one we face here. They addressed the optimality of the Friedman rule in economies with distorting taxes. In the case of money in the utility function models, they found that the Friedman rule holds exactly only if the value of real money balances in the utility function is infinity. Chari, Christiano, and Kehoe (1996) acknowledged that the infinite value of real money balances might cause a problem for the optimal allocation. To get around this technicality, they first set an upper bound for the value of real money balances and then said that the Friedman rule is optimal if the optimal allocation converges toward the Friedman rule as the upper bound is relaxed.

Following the approach idea of Chari, Christiano, and Kehoe (1996), we can modify the Lagrangian L with the inclusion of an additional constraint

$$c_t(\theta^t) \leq \bar{c}, \forall t, \theta^t,$$

where $\bar{c}$ denotes a very large but finite number. As in Chari, Christiano, and Kehoe (1996), we first work with the maximization problem without the imposition of this additional constraint. We then check if this constraint would bind for some θ^t at some t and what would happen if we relax $\bar{c}$ to, say, $2\bar{c}$ when the constraint does bind. In this way we can work within the environment of $x \in \ell^\infty$ and, at the same time, accommodate our analysis to the force of immiseration.

A.3 Deriving Lagrangian $\mathcal{L}$

This appendix gives the details of deriving the Lagrangian (6).

The part of the "raw" Lagrangian L associated with the temporary IC constraints (4) can be decomposed into three components:

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

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

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

Each component involves elements that take the form of $u(.)-v(.)$ and 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 these three components with the objective (1), we obtain the Lagrangian (6) with $\phi_c(\theta^t)$ and $\phi_l(\theta^t)$ given by (7) and (8).

A.4 Proof of Lemma 1

First, rearranging (13), 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 (9) and (10) 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 (7), 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 (11), 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.5 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.3, we obtain

$$\sum_{t=0}^{T} \sum_{\theta_t} \beta^t \pi_t(\theta^t) \left[u\left(c_t(\theta^t)\right) \phi_c(\theta^t) - v\left(\frac{l_t(\theta^t)}{\theta_t}\right) \phi_l(\theta^t) \right], \tag{30}$$

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 components of $\epsilon(\theta^t)$ be $\epsilon_1(\theta^t)$ and $\epsilon_2(\theta^t)$. Using the logic of proving Lemma 1 in Appendix A.4, it can be shown that

$$\sum_{\theta_t} \epsilon_1(\theta^t) \pi(\theta_t | x(\theta^{t-1}); t) = \sum_{\theta_t} \epsilon_2(\theta^t) \pi(\theta_t | x(\theta^{t-1}); t) = 0. \tag{31}$$

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 (30) is identical to the Lagrangian (6). Hence, following the same arguments as in Appendix A.4, $\phi_c(\theta^t) > 0$. In addition, Eq. (31) 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.

References

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

Albanesi, S., and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73(1), 1–30.

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.

Barro, R. J. (1979): "On the Determination of the Public Debt," *Journal of Political Economy*, 87(5), 940–71.

Bergemann, D., and A. Pavan (2015): "Introduction to Symposium on Dynamic Contracts and Mechanism Design," *Journal of Economic Theory*, 159, 679–701.

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.

Chari, V. V., L. J. Christiano, and P. J. Kehoe (1996): "Optimality of the Friedman rule in economies with distorting taxes," *Journal of Monetary Economics*, 37(2-3), 203–223.

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.

Dechert, W. D. (1982): "Lagrange Multipliers in Infinite Horizon Discrete Time Optimal Control Models," *Journal of Mathematical Economics*, 9(3), 285–302.

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 (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*, 106(2), 359–86.

Golosov, M., and A. Tsyvinski (2006): “Designing Optimal Disability Insurance: A Case for Asset Testing,” *Journal of Political Economy*, 114(2), 257–279.

——— (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,” *NBER Macroeconomics Annual*, 21, 317–388.

Golosov, M., A. Tsyvinski, and N. Werquin (2016): “Recursive Contracts and Endogenously Incomplete Markets,” NBER Working Paper.

Guvenen, F. (2011): “Macroeconomics with Hetereogeneity : A Practical Guide,” *Economic Quarterly*, 97(3), 255–326.

Hall, R. E. (1978): “Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,” *Journal of Political Economy*, 86(6), 971–87.

——— (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.

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.

Le Van, C., and H. Cagri Saglam (2004): "Optimal Growth Models and the Lagrange Multiplier," *Journal of Mathematical Economics*, 40(3-4), 393–410.

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

Makris, M., and A. Pavan (2016): "Incentives for Endogenous Types: Taxation under Learning-by-Doing," Discussion paper, Northwestern University.

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

——— (2016): "Recursive Contracts," Working Paper.

Marimon, R., and J. Werner (2016): "Envelope Theorem, Euler, and Bellman Equations without Differentiability," Working Paper.

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 (2016): "The Dual Approach to Recursive Optimization: Theory and Examples," Working Paper.

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

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

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

Pavan, A. (2016): "Dynamic Mechanism Design: Robustness and Endogenous Types," Discussion paper, Northwestern University.

Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82(2), 601–653.

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

Rustichini, A. (1998): "Lagrange Multipliers in Incentive-Constrained Problems," *Journal of Mathematical Economics*, 29(4), 365–380.

Sannikov, Y. (2008): "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 75(3), 957–984.

——— (2014): "Moral Hazard and Long-Run Incentives," Working Paper.

Sleet, C., and S. Yeltekin (2010): "Misery and Luxury: Long Run Outcomes with Private Information," GSIA Working Papers 2011-E19, Carnegie Mellon University, Tepper School of Business.

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," *Quarterly Journal of Economics*, 122(3), 925–967.