Time Inconsistency with Heterogeneous Agents

Eduardo Dávila* Andreas Schaab†

Yale UC Berkeley

June 2025

Abstract

This paper revisits the question of time consistency in heterogeneous agent (HA) environments. HA models feature three sources of time inconsistency: dynamically inconsistent individual preferences (Strotz, 1956), forward-looking behavior (Kydland and Prescott, 1977), and interpersonal welfare comparisons. When individual preferences are symmetric, only the utilitarian social welfare function implies time consistent aggregate welfare assessments. The attribution of welfare gains to efficiency and redistribution, on the other hand, invariably changes over time when markets are incomplete and is therefore time inconsistent, irrespective of the social welfare function and the extent of heterogeneity in individual preferences. Therefore, even when optimal policy is time consistent, its justification on grounds of efficiency and redistribution changes as time passes and uncertainty is realized. These results offer a useful new perspective on time consistency problems in the design of anticipated relief policies, social insurance, labor taxation, investment policies, and social contracts.

JEL codes: D52, D60, D70, E60, H21, H23

Keywords: time inconsistency, heterogeneous agents, welfare, efficiency, redistribution, risk-sharing

---

*Email: eduardo.davila@yale.edu. Website: https://eduardodavila.com.

†Email: schaab@berkeley.edu. Website: https://andreasschaab.com.

We are grateful to Joel Flynn, Joao Guerreiro, Marina Halac, Ricardo Reis, Karthik Sastry, Jón Steinsson, Ludwig Straub, Alireza Tabaz-Salehi, and Pierre Yared for helpful conversations.

1 Introduction

Time consistency is a central concern in dynamic policy analysis. A policy that maximizes social welfare from today's perspective is called *time consistent* if it remains optimal when revisited in the future. Because only time consistent policies can be implemented credibly, understanding the sources of time *in*consistency is indispensable for normative macroeconomics. In representative agent (RA) environments, two canonical forces can undermine consistency. First, Strotz (1956) shows that non-exponential discounting implies inconsistent intertemporal preferences. When the social planner inherits such preferences, policy may become time inconsistent. Second, Kydland and Prescott (1977) show that forward-looking behavior can give rise to time consistency problems. When the planner can influence individual behavior today by making promises about policy in the future, she may have an incentive to renege on these promises ex post once individual decisions are sunk.

A rapidly expanding literature now uses heterogeneous agent (HA) models to analyze business cycles fluctuations, policy transmission, and the determinants of cross-sectional heterogeneity. HA models are quickly supplanting the RA benchmark as the workhorse of macroeconomic policy analysis. Yet the normative implications of these models for optimal policy design remain comparatively underexplored. This paper revisits the question of time consistency in environments with heterogeneous agents and incomplete markets.

In HA environments, there is a third source of time inconsistency: aggregate welfare assessments. When using a social welfare function (SWF) to judge the desirability of a policy, the planner aggregates welfare gains and losses across heterogeneous individuals. These interpersonal welfare comparisons can emerge as a new source of time inconsistency, in addition to those identified by Strotz (1956) and Kydland and Prescott (1977). Intuitively, when the planner's marginal rate of substitution between two individuals in utility units changes over time, so too can the perceived desirability of the policy. Only the utilitarian SWF is immune to this form of time inconsistency because individuals' Pareto weights are constant across dates and histories.[1] This observation was first formalized by Zuber (2011) but its intuitive importance had already been acknowledged — albeit only informally — by earlier work dating to at least Harsanyi (1955).

In this paper, we revisit the question of time consistency in heterogeneous agent environments and study the role of incomplete markets. Our main contribution is to unpack the aggregate welfare assessment and characterize the time consistency of the distinct normative considerations that make up the overall welfare assessment. Our main result shows that, in HA incomplete markets economies, assessments of efficiency and redistribution gains and losses invariably change over time and are therefore time inconsistent. Even when the overall welfare assessment is consistent in the special case of a linear SWF and with homogeneous discount factors and beliefs, the attribution of welfare gains and losses to efficiency and redistribution invariably changes over time.

[1] Heterogeneity in discount factors and beliefs can lead to time inconsistent aggregate welfare assessments even when the SWF is utilitarian, as shown by Jackson and Yariv (2014, 2015) and Mongin (1995, 1998), respectively.

We begin in Section 2 with an illustrative example. The world is inhabited by a continuum of households and there are two periods, today and tomorrow. Households work and consume in both periods, but there are no financial markets so they cannot save. Today, households are identical. Tomorrow, they face symmetric labor productivity risk. We consider a tax-and-transfer policy that raises revenue from a distortionary labor income tax to provide a uniform, lump-sum transfer to households. And we compare the assessments of this policy from the perspectives of today and tomorrow under different SWFs. This example allows us to highlight four simple observations.

First, the attribution of gains and losses from the optimal policy to efficiency and redistribution changes with the perspective of the assessment and is therefore time inconsistent. Ex ante, households are identical. They would prefer smoothing their consumption across states of the world tomorrow but are unable to do so. The social transfer helps insure households against idiosyncratic risk because the burden of taxation falls disproportionately on the more productive while the transfer is uniform. From the perspective of today, the promise of labor income taxation tomorrow achieves welfare gains due to risk-sharing. On the other hand, the tax is distortionary because it discourages households from supplying labor, generating a welfare loss from production efficiency. Suppose time passes and tomorrow's state of the world is realized. From this ex post perspective, the planner no longer perceives any scope to insure households against idiosyncratic risk. What the planner perceived as risk-sharing gains ex ante now appears as pure redistribution. Therefore, the planner initially attributes both gains and losses from labor taxation to efficiency; whereas ex post, the planner trades off redistribution gains against production efficiency losses. This is a concrete example of the time inconsistency of efficiency and redistribution assessments.

Second, the perceived production efficiency loss due to distortionary taxation is the same ex ante and ex post. While risk-sharing and redistribution assessments change over time, assessments of production efficiency gains and losses are time consistent. This highlights that some components of the overall efficiency and welfare assessment are time consistent, while others are not.

Third, the overall welfare assessment of tax policy is time consistent under the utilitarian SWF, highlighting the result of Zuber (2011). Even so, the underlying assessments of efficiency and redistribution gains and losses are not. Even though the optimal policy is time consistent, therefore, its justification varies with the perspective of the assessment.

Fourth, optimal policy is time inconsistent for all other SWFs. Only the utilitarian SWF values ex post redistribution and ex ante risk-sharing equivalently. Non-utilitarian SWFs over- or under-weight redistribution relative to the utilitarian benchmark. Non-utilitarian planners therefore perceive the redistribution gains from the social transfer ex post as more or less valuable than the risk-sharing gains ex ante. This leads to a time consistency problem.

We generalize these four observations to a large class of dynamic stochastic environments with heterogeneous agents and incomplete markets. The framework we introduce in Section 3 nests most HA models currently found in the applied literature, leveraging a general sequence-space representation in the spirit of Auclert et al. (2021). Our only restrictive assumption is that preferences take the separable expected utility form with homogeneous discount factors and beliefs. This allows us to clarify our contribution relative to the literature that studies time inconsistency resulting from heterogeneity in individual preferences and beliefs.

We set the stage in Section 4 by presenting a systematic characterization of the sources of time inconsistency for RA and HA welfare assessments, revisiting the results of Kydland and Prescott (1977) and Zuber (2011). Relative to Zuber (2011) and the subsequent social choice literature, we analytically characterize the relationship between ex ante and ex post assessments. We show that the time consistency problem from interpersonal welfare comparisons is governed by a particular cross-sectional covariance. Only when this covariance takes the opposite sign from the welfare assessment and is sufficiently large can time inconsistency arise. Our result also clarifies the role of incomplete markets. Indeed, when markets are complete, even non-utilitarian social welfare functions lead to time consistent assessments as long as individual preferences are symmetric.

We then present our main result in Section 5: The attribution of welfare gains and losses to efficiency and redistribution invariably changes over time and is therefore time inconsistent when markets are incomplete. Even in the special case of a linear SWF and homogeneous discount factors and beliefs, the justification of optimal policy on grounds of efficiency or redistribution varies with the assessment perspective. The source of time inconsistency in efficiency assessments is a change in the reference unit, or welfare numeraire, used to express and compare individual welfare gains and losses. We refer to this as *numeraire inconsistency* and again associate the resulting time consistency problem with a particular cross-sectional covariance term. When markets are complete, numeraire inconsistency can never arise because individuals' relative valuations of bundles of goods are equalized. The time inconsistency in efficiency and redistribution assessments we identify is therefore a direct result of incomplete markets.

Applications. Our goal in this paper is to show that the time inconsistency of efficiency and redistribution assessments provides a useful new perspective on applied policy questions. We present four applications and show that our approach yields new insights on classic questions in normative macroeconomics.

Our first application identifies a particular form of time inconsistency in the design of anticipated relief policies — policies intended to support individuals who experience a temporary spell of low consumption followed by a gradual, anticipated recovery. Relief policies of this sort are common, including the COVID-19 assistance programs enacted through the CARES Act, childcare support and parental leave, unemployment insurance, retraining programs, and affirmative action. In each case, the policy is designed ex ante to offset a foreseeable period of hardship and is intended to phase out once recovery is underway. We show that a particular form of time inconsistency arises in such settings: Ex post, the planner finds it optimal to extend the relief policy, even though it was originally intended to be temporary.

Our second application studies social insurance policies. When markets are incomplete, individuals cannot fully insure against idiosyncratic or differentially distributed aggregate risk. A policy that improves consumption smoothing across states of the world can therefore generate efficiency gains from an ex ante perspective. Social insurance is a central motivation behind a wide range of policies, including progressive income taxation, unemployment insurance, health care provision, disaster relief, and certain forms of macroeconomic stabilization. Such policies initially appear to improve risk-sharing and allocative efficiency but may ex post be perceived as pure redistribution: Once uncertainty is resolved and individual outcomes are realized, they require redistributive transfers from the lucky to the unlucky ex post: today's efficiency becomes tomorrow's redistribution. A social insurance policy that is considered optimal on efficiency grounds ex ante may therefore be time inconsistent.

Our third application studies investment policies that impose a resource cost in the short run for the promise of resource gains in the long run. We show that a particular form of time inconsistency arises in the optimal design of such policies in heterogeneous agent environments when markets are incomplete: production efficiency assessments are present biased when inequality is rising over time. This application identifies a new source of present bias in aggregate welfare assessments that is distinct from discount factor heterogeneity, complementing the results of Jackson and Yariv (2014, 2015).

Related literature. Our paper contributes to two literatures in macroeconomics, namely those on time inconsistency and on heterogeneous agent incomplete markets economies. Following the foundational papers of Strotz (1956) and Kydland and Prescott (1977), a vast body of work has analyzed the time consistency of government policies, especially fiscal and monetary policy.[2] Much of this tradition has worked with RA models. Notable exceptions include Werning (2007), Chang (2022), Bilbiie (2024) and Dávila and Schaab (2023) who study time consistency in the presence of heterogeneity.[3] All of these papers use linear SWFs to make interpersonal welfare comparisons. Our contribution to this body of work is three-fold. First, we show that interpersonal welfare comparisons emerge as a new source of time inconsistency when the SWF is non-linear and we characterize a specific cross-sectional covariance that governs the resulting time consistency problem. Second, we establish that efficiency and redistribution assessments are invariably time inconsistent when markets are incomplete, even for the special case of a linear SWF. Third, we

[2] See Kydland and Prescott (1980), Barro and Gordon (1983), Lucas and Stokey (1983), Rogoff (1985), Chari and Kehoe (1990), Aiyagari et al. (2002), Athey et al. (2005), Klein et al. (2008), Halac and Yared (2014), Marcet and Marimon (2019), Dovis and Kirpalani (2020, 2021), Sublet (2023), and Clayton and Schaab (2025) among many others.

[3] Werning (2007) discusses the time consistency of capital taxation. The usual time consistency problem in the RA environment arises because the planner lacks lump-sum taxation and expropriating capital ex post mimics a non-distortionary tax since accumulation decisions are already sunk. Werning (2007) shows that in the presence of heterogeneity, optimal capital taxation may be time inconsistent even in the presence of a lump-sum tax because it helps the planner redistribute across agents, which is otherwise only possible with distortionary labor taxes. This time consistency problem is therefore in the spirit of Kydland and Prescott (1977). Indeed, Werning (2007) exclusively focuses on welfare assessments under a linear SWF with common discount factor and beliefs; aggregate welfare assessments consequently do not emerge as a distinct source of time inconsistency in his environment.

present a unified characterization of the sources of time inconsistency in heterogeneous agent environments — distinguishing between time consistency problems á la Strotz (1956), Kydland and Prescott (1977) and due to interpersonal welfare comparisons — using a general sequence-space representation that nests most models that are currently employed in applied work. This allows us in particular to clarify the role of incomplete markets for time consistency as distinct from heterogeneity in individual preferences.

Our paper is therefore also part of the rapidly growing literature in macroeconomics that uses sequence-space methods to study heterogeneous agent models with incomplete markets. See Auclert et al. (2025) for a recent survey. We derive our results for a general class of HA models leveraging a representation of individual consumption functions that builds on the approach developed in Auclert et al. (2021). While most recent work in this literature has been positive, our focus is instead on the normative implications of heterogeneity.

The main result of this paper is a characterization of the time consistency of efficiency and redistribution assessments. Distinguishing notions like efficiency, risk-sharing and redistribution in simple environments like our example of Section 2 is straightforward. But this becomes non-trivial in general incomplete markets economies. Our results therefore build on prior work that has developed formal characterizations of efficiency, risk-sharing and redistribution in heterogeneous agent environments: Kaldor (1939), Hicks (1939), Benabou (2002), Bhandari et al. (2024), and in particular Dávila and Schaab (2024).

Outside of macroeconomics, there is a long tradition in social choice theory of studying the possibility of aggregating individual preferences. Arrow (1950)'s impossibility result presents an early benchmark, and subsequent work has explored alternative formulations and relaxations of Arrow's criteria. Zuber (2011) shows that time consistency generically requires a linear social welfare function. A strand of this literature studies the aggregation of heterogeneous discount factors and shows that time inconsistency in social choice is an unavoidable consequence of individual heterogeneity in time preferences (Marglin, 1963; Feldstein, 1964; Jackson and Yariv, 2014, 2015; Adams et al., 2014; Chambers and Echenique, 2018). Halevy (2015) and Millner and Heal (2018) distinguish between time invariance and time consistency. They clarify that while time invariant and stationary social welfare functions necessarily imply time consistent welfare assessments, time invariance is not required for time consistency. Relatedly, Mongin (1995, 1998) shows that belief heterogeneity also implies time inconsistent aggregate welfare assessments. In this paper, we focus on time invariant SWFs but allow for non-stationarity. We assume homogeneous discount factors and beliefs from the outset to clearly distinguish our results from this body of work. Our contribution to this literature is three-fold. First, our marginal approach based on welfare assessments is complementary to Zuber (2011)'s result as it allows us to associate time inconsistency in interpersonal welfare comparisons with a specific cross-sectional covariance term. Second, we clarify the role incomplete markets play for the time inconsistency of welfare assessments, as distinct from heterogeneity in individual preferences. Third, we show that efficiency and redistribution assessments are invariably inconsistent even when the SWF is stationary and discount factors and beliefs are homogeneous.

The tension between efficiency or risk-sharing and redistribution plays an important role in our paper and has been an enduring theme of economic thought. The early work of Marshall (1920), Pigou (1920) and Knight (1921) already recognizes that schemes designed to provide risk-sharing ex ante necessarily imply redistributive transfers ex post once uncertainty has realized. Arrow (1963) is often credited as providing an early formal statement that the pooling of risk can raise all individuals' expected utility ex ante but requires redistribution ex post:

> *[T]he preference for redistribution expressed in government taxation... can be reinterpreted as desire for insurance... Thus, optimality, in a context which includes risk-bearing, includes much that appears to be motivated by distributional value judgments.*

In a sense, our results provide a novel and practical formalization of these ideas.

Our paper also shares this theme with subsequent work in the public finance literature. The simple example we develop in Section 2 builds on Varian (1980) and Piketty and Saez (2013b). Much of the early work on optimal taxation focuses on static environments and appeals to an equivalence between ex post redistribution and ex ante risk-sharing (behind the veil of ignorance).[4] Our results emphasize an important qualification to this reasoning: Only the utilitarian SWF values risk-sharing and redistribution equivalently, whereas non-utilitarian SWFs lead to a time consistency problem. Much applied work in public finance has employed non-utilitarian SWFs (Atkinson, 1970; Atkinson and Stiglitz, 1976; Stern, 1976; Boskin and Sheshinski, 1978; Diamond, 1998; Saez, 2001; Piketty and Saez, 2012, 2013a; Saez and Stantcheva, 2016). The implications of dynamics and uncertainty have been the focus of a new dynamic public finance literature. And the interplay between ex ante risk-sharing and ex post redistribution has been explored, for instance, by Farhi and Werning (2013) and Golosov et al. (2016). Farhi and Werning (2013) observe that a utilitarian SWF implies an equivalence between ex ante risk-sharing and ex post redistribution, which breaks down for non-utilitarian SWFs. In contrast to this literature, our focus is on characterizing time inconsistency in not only welfare but also efficiency and redistribution assessments.

Lastly, there is long tradition in political economy that studies time consistency and shares our focus on the tension between efficiency and redistribution. Important examples include Alesina and Rodrik (1994), Persson and Tabellini (1994), Krusell et al. (1997), and Bisin et al. (2015). Farhi et al. (2012) study a dynamic Mirrleesian model where policy is chosen sequentially without commitment. Time inconsistency arises in their setting due to the interplay between dynamic incentive provision and redistribution; but this time consistency problem is due to forward-looking behavior á la Kydland and Prescott (1977) rather than interpersonal welfare comparisons: Capital accumulation decisions are sunk ex post, and so a utilitarian planner with redistribution concern but without commitment is tempted to expropriate capital ex post. Papers in this literature often study voting equilibria or coalition shifts rather than the welfarist perspective we focus on.

[4] See for example Mirrlees (1971), Diamond (1998) and Saez (2001).

2 A Simple Example

We start with a simple illustrative example.

2.1 Environment

There are two periods indexed by $t \in \{0,1\}$. The economy is populated by a unit measure of households indexed by $i \in [0,1]$. Households are identical in period 0 but face uninsurable idiosyncratic earnings risk in period 1. There is no aggregate uncertainty. We study a social transfer policy financed by a distortionary labor income tax.

Households. Household i's lifetime utility from the perspective of period 0 is defined by

$$V_0^i = u(c_0^i, \ell_0^i) + \beta\, \mathbb{E}_0\, u(c_1^i, \ell_1^i) \tag{1}$$

where $u(\cdot)$ denotes instantaneous flow utility from consumption and hours worked and β is the discount factor. We denote i's lifetime utility from the perspective of period 1 after uncertainty is realized by $V_1^i = u(c_1^i, \ell_1^i)$. Households face the budget constraints

$$c_0^i = w_0 \ell_0^i \tag{2}$$

$$c_1^i = (1-\tau) w_1 z_1^i \ell_1^i + T. \tag{3}$$

In period 0, household i supplies ℓ_0^i hours of labor and earns the wage w_0. Households are symmetric from the perspective of period 0 but face idiosyncratic labor productivity risk in period 1, encoded in the random variable z_1^i that is drawn from a density $g(z)$. We normalize its mean to $\int_0^1 z_1^i di = \int z g(z) dz = 1$. Finally, τ denotes a labor income tax and T is a lump-sum government transfer. Household i maximizes (1) subject to (2) and (3).

Firms. There is a representative firm that produces the final consumption good using labor according to the production function

$$Y_t = L_t, \tag{4}$$

where L_t denotes aggregate use of effective labor in production. Firms are perfectly competitive and maximize profits in each period.

Tax and transfer policy. Fiscal policy at date 1 must satisfy a balanced budget constraint. Total transfer payments $\int_0^1 T\, di$ must be financed by tax revenue $\int_0^1 \tau z_t^i \ell_t^i\, di$, which implies

$$T = \tau \int_0^1 z_1^i \ell_1^i\, di. \tag{5}$$

Equilibrium. Taking as given a tax policy τ, competitive equilibrium comprises an allocation $\{Y_t, L_t, c_t^i, \ell_t^i\}_{i,t}$ and a transfer T such that (i) households optimize, (ii) firms maximize profits, (iii) the government's balanced budget constraint is satisfied, and the (iv) markets for goods and effective labor clear in each period,

$$Y_t = \int_0^1 c_t^i \, di \tag{6}$$

$$L_t = \int_0^1 z_t^i \ell_t^i \, di. \tag{7}$$

2.2 Optimal Policy

We define social welfare from the perspective of period t as

$$W_t = \mathcal{W}(\{V_t^i\}_i) = \left(\int_0^1 (\nu^i V_t^i)^{1-\phi} \, di \right)^{\frac{1}{1-\phi}}, \tag{8}$$

with $\nu^i > 0$ for all i. This non-linear social welfare function (SWF) is often referred to as isoelastic after Atkinson (1970). Welfare criteria such as (8) are used in macroeconomics and public finance to allow for the explicit "parameterization of the planner's concern for redistribution" (Benabou, 2002; Heathcote et al., 2017). In particular, (8) nests the utilitarian, Nash, and Rawlsian SWFs for the parameter values $\phi \in \{0, 1, \infty\}$, respectively.[5]

Ex ante optimal policy. The date 0 Ramsey problem is therefore to choose an income tax τ to maximize welfare W_0 subject to the conditions of competitive equilibrium. What are the sources of welfare gains from such a policy? From the perspective of period 0, all households are identical since they are initially equally productive and face the same earnings risk in period 1. Due to diminishing marginal utility, $u''(\cdot) < 0$, households would prefer smoothing their consumption across different realizations of z_1^i, but they are unable to do so due to incomplete markets. They would happily trade state-contingent claims with each other if such financial markets existed. The government policy (τ, T) mimics such claims because the burden of taxation τ disproportionately falls on the ex post lucky while the social transfer T is uniform across households. From the perspective of period 0, therefore, the promise of labor income taxation in period 1 insures households against idiosyncratic risk and achieves welfare gains due to *risk-sharing*. On the other hand, the labor income tax τ used to finance this social transfer also discourages households from supplying labor and is therefore distortionary. The optimal income tax from the perspective of period 0 trades off

[5] The parameter ϕ denotes the inverse of the elasticity of substitution across individual lifetime utilities. As $\phi \to 0$, utilities become perfect substitutes, with $W_t = \int_0^1 \nu^i V_t^i \, di$, which is the linear utilitarian SWF. As $\phi \to 1$, we obtain the Nash or Cobb-Douglas SWF $W_t = \exp(\int_0^1 \nu^i \log V_t^i \, di)$. And as $\phi \to \infty$, utilities become perfect complements, with $W_t \to \min_i\{\nu^i V_t^i\}$. Setting $\nu^i = 1/(V_1^i)^{-\phi}$ after the assessment has been computed corresponds to the no-redistribution planner of Dávila and Schaab (2024) who sets optimal policy to maximize Kaldor-Hicks efficiency at the margin.

these two considerations.

We show in Appendix A that the optimality condition associated with the date 0 Ramsey problem can be written as

$$0 = \underbrace{\tau \frac{dL_1}{d\tau}}_{\substack{\text{Production}\\ \text{Efficiency}}} + \underbrace{\mathbb{C}ov_i\left(\frac{u^i_{c,1}}{\mathbb{E}_0 u^i_{c,1}}, -z^i_1 \ell^i_1\right)}_{\text{Risk-Sharing}}, \tag{9}$$

where $u^i_{c,1} = u'(c^i_1)$ denotes marginal utility. From the ex ante perspective of period 0, the planner trades off the risk-sharing gains from providing social insurance against the production efficiency losses of distortionary taxation. A marginal tax increase $d\tau$ has a direct effect on household i's income in period 1 of $-z^i_1\ell^i_1 + \frac{dT}{d\tau}$. The transfer benefit $\frac{dT}{d\tau}$ is uniform across households while the burden of taxation falls relatively on those with large z^i_1. Households value these gains in proportion to marginal utility $u^i_{c,1}$. The tax increase $d\tau$ improves risk-sharing since these gains $-z^i_1\ell^i_1 + \frac{dT}{d\tau}$ accrue to households with relatively high marginal utility, that is, $\mathbb{C}ov_i(u^i_{c,1}, -z^i_1\ell^i_1 + \frac{dT}{d\tau}) = \mathbb{C}ov_i(u^i_{c,1}, -z^i_1\ell^i_1) > 0$. At the same time, the tax increase discourages labor supply, lowers aggregate consumption, and generates a production efficiency loss of $\tau\frac{dL_1}{d\tau} < 0$.

Notice that the ex ante assessment of production efficiency and risk-sharing is invariant to the choice of social welfare function. In particular, (9) does not depend on ϕ. All social welfare functions therefore agree on the optimal labor tax from the ex ante perspective, irrespective of their concern for redistribution.

Ex post optimal policy. Suppose that time passes and uncertainty realizes at the beginning of period 1. If the planner could reoptimize at this point, would she still choose the same tax policy? From this ex post perspective, the optimal policy problem is to choose an income tax τ to maximize welfare W_1, subject to the conditions of competitive equilibrium. Ex post, all uncertainty has realized and there is no longer any scope to insure households against idiosyncratic earnings risk. What the planner perceived as risk-sharing gains ex ante now appears as pure *redistribution*. In Appendix A, we show that the condition for ex post optimal fiscal policy can be written as

$$0 = \underbrace{\tau \frac{dL_1}{d\tau}}_{\substack{\text{Production}\\ \text{Efficiency}}} + \underbrace{\mathbb{C}ov_i\left(\frac{(\nu^i)^{1-\phi}(V^i_1)^{-\phi}u^i_{c,1}}{\int_0^1 (\nu^i)^{1-\phi}(V^i_1)^{-\phi}u^i_{c,1}\,di}, -z^i_1\ell^i_1\right)}_{\text{Redistribution}}. \tag{10}$$

The optimal labor income tax now trades off redistribution gains against production efficiency losses due to distortionary labor taxation. These perceived redistribution gains now critically depend on the underlying social welfare function through the inequality-aversion parameter ϕ. The production efficiency loss is still invariant to the choice of SWF, however.

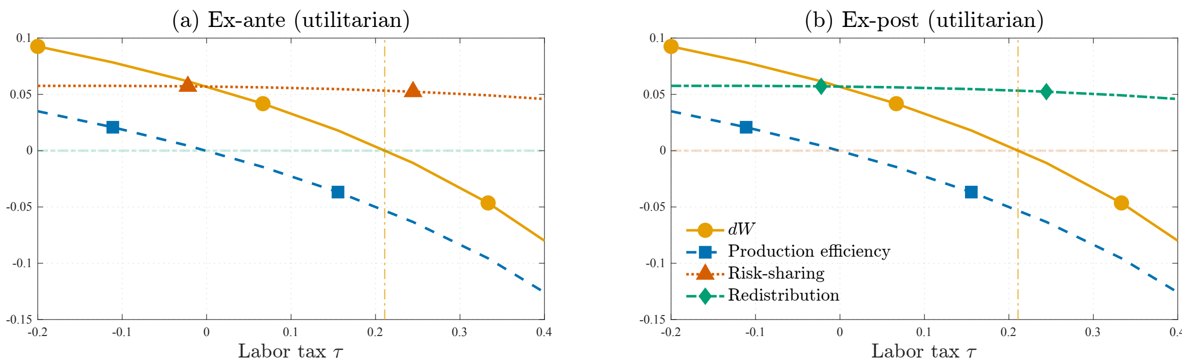

Figure 1. Utilitarian Welfare Assessments

Time inconsistency. Our simple example introduces four key ideas that we develop in the rest of the paper. Figure 1 illustrates three of these ideas.[6] It plots the ex ante (Panel a) and ex post (Panel b) welfare assessments of a marginal tax increase $d\tau$ under a utilitarian SWF ($\phi = 0$). The x-axes display the different levels of tax rates τ at which we assess the marginal perturbation $d\tau$. And the four colored lines in each panel correspond to the overall welfare assessments (yellow) and their decompositions into gains and losses due to production efficiency (blue), risk-sharing (red) and redistribution (green). The ex ante and ex post optimal tax rates correspond to the τ at which the yellow lines cross 0, that is $\frac{dW_0}{d\tau} = 0$ in Panel (a) and $\frac{dW_1}{d\tau} = 0$ in Panel (b) — illustrated by the dashed vertical lines. Whenever a line lies above (below) 0 at a given τ, the utilitarian planner associates gains (losses) with a marginal tax increase $d\tau$ due to that particular motive. Figure 1 illustrates three results:

1. Assessments of efficiency and redistribution gains and losses change over time and are therefore time inconsistent. In this concrete setting, the planner initially perceives gains from fiscal policy due to risk-sharing. In Panel (a), the red line (risk-sharing) lies above 0 for tax rates up to and exceeding 40% while the green line (redistribution) is always 0. Ex post, however, risk-sharing no longer plays any role and all gains from labor taxation are considered pure redistribution. In Panel (b), the green line is positive while the red line is 0. Ex ante, the planner therefore attributes both the gains and losses from fiscal policy to efficiency; whereas ex post, the planner trades off an efficiency loss against a redistribution gain. The main result of this paper shows that this time inconsistency of efficiency and redistribution assessments is a general feature of heterogeneous agent incomplete markets economies. What the planner perceives as the sources of welfare gains and losses invariably changes over time in these environments.

2. The planner perceives the same production efficiency loss $\tau \frac{dL_1}{d\tau}$ from taxation ex ante and ex

[6] Figures 1 and 2 use the following illustrative parametrization. We set $\beta = 0.96$ and assume CRRA preferences, $u(c, \ell) = \frac{1}{1-\gamma} c^{1-\gamma} - \frac{1}{1+\eta} \ell^{1+\eta} + \bar{u}$, with $\gamma = \eta = 2$. We set $\bar{u} = 3$ to ensure that $V_1(z) > 0$ for all z. We draw z_1^i from a log-normal distribution with mean 1 and standard deviation 0.4. Finally, we focus on equal-weighted SWFs with $\nu^i = 1$ for all i.

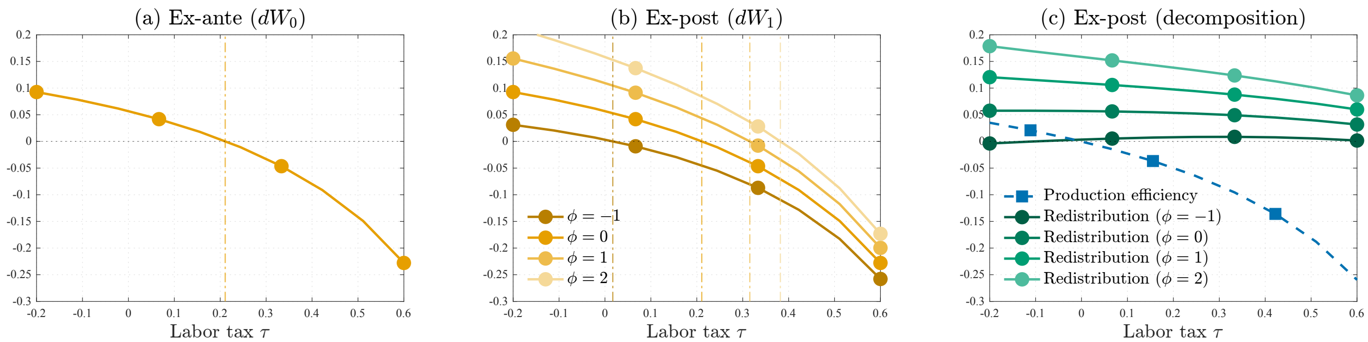

Figure 2. Time Inconsistency of Welfare Assessments

post. From both perspectives, production efficiency considerations suggest an optimal tax of 0. The blue lines in Panels (a) and (b) are identical and cross 0 at $\tau = 0$. We show in Section 5 that this insight generalizes: While risk-sharing and redistribution assessments change over time, assessments of static production efficiency gains and losses are time consistent. This result is not special to the utilitarian SWF but holds for all ϕ.

3. The overall welfare assessment $\frac{dW_0}{d\tau}$ is time consistent under the utilitarian SWF, despite the inconsistency of the risk-sharing and redistribution assessments. The yellow lines in Panels (a) and (b) are identical and cross 0 at $\tau = 0.22$. While the justification for taxation varies starkly between the ex ante and ex post perspectives, optimal policy therefore remains time consistent under the utilitarian SWF. Indeed, notice that the covariance term in (10) becomes identical to that in (9) for $\phi = 0$ and $\nu^i = 1$ for all i. This consistency is special to the utilitarian SWF and does not hold for any $\phi \neq 0$ as we show next.

Figure 2 contrasts welfare assessments across different social welfare functions. It plots the ex ante and ex post welfare assessments in Panels (a) and (b), and the decomposition of the ex post assessment in Panel (c). Each line corresponds to a particular SWF: inequality-loving ($\phi = -1$), utilitarian ($\phi = 0$), Nash ($\phi = 1$), and inequality-averse ($\phi = 2$). All SWFs agree on the ex ante welfare assessment $\frac{dW_0}{d\tau}$ as well as on the ex post production efficiency assessment: There is a single yellow line in Panel (a) and a single blue line in Panel (c). Different SWFs disagree on the ex post welfare assessment in Panel (b), however, because they disagree on the ex post redistribution in Panel (c). This implies our fourth and final observation:

4. Optimal policy is time inconsistent for all $\phi \neq 0$. Only the utilitarian SWF values ex post redistribution the same as ex ante risk-sharing. Non-utilitarian SWFs over- or under-weight redistribution relative to efficiency considerations. As ϕ increases, the planner perceives the redistribution gains from the social transfer ex post as more valuable than the risk-sharing gains ex ante (Panel c). The ex post optimal tax increases with ϕ (Panel b) and only coincides with the ex ante optimal policy for $\phi = 0$. This leads to a time consistency problem.

3 General Environment and Welfare

Our notation closely follows Chapter 8 of Ljungqvist and Sargent (2018). Time is discrete and indexed by $t \in \{0, 1, \ldots\}$. There are $I \leq \infty$ individuals indexed by $i \in \{1, \ldots, I\}$. And at the beginning of each period t a stochastic event s_t realizes. We denote histories of events by $s^t = (s_0, s_1, \ldots, s_t)$ and the unconditional probability of a particular history occuring by $\pi(s^t)$. We use the notation $s^t > s^k$ to indicate that $\pi(s^t \mid s^k) > 0$, i.e., that history s^t can be reached from s^k.

Preferences. Individuals consume a single consumption good. The consumption preferences of individual i from the perspective of date 0 are given by

$$\sum_t \beta^t \sum_{s^t} \pi(s^t)\, u^i(c^i_t(s^t))$$

where $c^i_t(s^t)$ is i's consumption in history s^t at date t. We allow the utility function $u^i(\cdot)$ to vary across individuals but assume that the discount factor β and beliefs $\pi(\cdot)$ do not. We also assume that $u^i(c^i_t(s^t))$ is well behaved and Inada conditions apply. Our assumption that β and $u^i(\cdot)$ do not vary with time rules out from the start time consistency problems due to inconsistent individual preferences in the spirit of Strotz (1956).

Consumption function and perturbation. We assume that individual consumption allocations depend on the parameters $\boldsymbol{\theta} = \{\theta_t(s^t)\}_{t,s^t}$, where each $\theta_t(s^t)$ is a scalar.[7] Concretely, we assume that there exists a smooth *consumption function* that defines i's consumption in history s^t at date t from the perspective of some other history s^k as

$$\mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^k). \tag{11}$$

This consumption function defines the mapping from parameters $\boldsymbol{\theta}$ to individuals' consumption, and we take this mapping as given throughout the paper.[8] Appendix C presents a general derivation and micro-foundation of the consumption function (11) using a sequence-space representation of heterogeneous agent models (Auclert et al., 2021).

Our focus in this paper is on comparing the welfare assessments of marginal perturbations $d\theta$, for some $\theta \in \boldsymbol{\theta}$, from the perspective of different points in time. This notion of *perspective* is critical to our analysis and at the heart of the question of time consistency. We use the term perspective

[7] We can think of $\boldsymbol{\theta}$ equivalently as a collection of parameters or as a stochastic process.

[8] This mapping can emerge, for instance, from assumptions about competitive equilibrium in which individuals optimize subject to budget constraints. But the specific micro-foundation of $\mathcal{C}^i_t(\cdot)$ is not relevant for our main results and so we take the dependence of consumption on perturbation parameters $\boldsymbol{\theta}$ as given. It is most natural in our context to think of $\boldsymbol{\theta}$ as a policy instrument such as nominal interest rates (monetary policy) or taxes (fiscal policy). It may also represent other exogenous primitives such as productivity or endowments. In the optimal policy context, taking the mapping from $\boldsymbol{\theta}$ to the allocation as given corresponds to the *dual* approach — see e.g. Dávila and Schaab (2023) or Auclert et al. (2024).

to refer to the history at which we consider and assess the perturbation $d\theta$. The consumption function (11) depends explicitly on this perspective — here history s^k — because consumption at a particular history may respond differently to the perturbation depending on the perspective of the assessment, that is, the point in time at which the perturbation is announced. When we consider the response of the allocation to a perturbation $d\theta$ from the perspective of history s^k, we take as given the past. In other words, we take as given and hold fixed the allocation prior to date k or in histories that can no longer be reached from s^k. Intuitively, an assessment at history s^k takes as given certain initial conditions that are invariant to perturbations $d\theta$. On the other hand, an assessment of the same perturbation from the earlier perspective of history s^0, for instance, would take into account how the announcement of $d\theta$ impacts the allocation between dates 0 and k. We make this intuition formal in Appendix C.[9]

Welfare assessments. We assess the welfare consequences of the perturbation $d\theta$ through the lens of a social welfare function (SWF). From the perspective of history s^k, social welfare is defined as

$$W_{s^k} = \mathcal{W}\big(\{V^i_{s^k}\}_i\big), \tag{12}$$

where $V^i_{s^k}$ is the lifetime utility of individual i from the perspective of history s^k. Using the consumption function (11), we define lifetime utility as

$$V^i_{s^k} = \sum_{t \geq k} \beta^{t-k} \sum_{s^t \succeq s^k} \pi(s^t \mid s^k)\, u^i(\mathcal{C}^i_t(s^t \mid s^k)), \tag{13}$$

where $\sum_{s^t \succeq s^k}$ sums over all histories that follow from s^k and $\pi(s^t \mid s^k)$ denotes conditional probabilities, with $\pi(s^k \mid s^k) = 1$. We suppress the dependence of $\mathcal{C}^i_t(\cdot)$ and therefore of $V^i_{s^k}$ and W_{s^k} on the perturbation parameters $\boldsymbol{\theta}$. In both definitions (12) and (13), the $_{s^k}$ subscript notation is intended to emphasize the perspective from which the welfare assessment is made and therefore the time at which the perturbation is announced. We refer to the units of $V^i_{s^k}$ as individual i (history s^k) utility units, or utils for short. The units of W_{s^k} are social (history s^k) utils.

The key assumption encoded in (12) is that $\mathcal{W}(\cdot)$ is a time-invariant function of individual lifetime utilities as perceived at the time of the assessment, but of nothing else.[10] This assumption

[9] Concretely, the derivation of the consumption function in Appendix C shows that $\partial\mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^k)/\partial\theta = 0$ for $\theta \in \{\theta_\ell(s^\ell)\}_{\ell<k, s^\ell}$ or $\theta \in \{\theta_\ell(s^\ell)\}_{\ell \geq k, s^\ell \nsucceq s^k}$. From perspective s^k, a perturbation $d\theta$ takes as given the allocation prior to date k and at histories that can no longer be reached from s^k. The same perturbation may therefore affect the allocation differently when considered from different perspectives, $\partial\mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^k)/\partial\theta \neq \partial\mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^\ell)/\partial\theta$.

[10] Two important properties of social welfare functions that are often discussed in the social choice literature are time invariance and stationarity (Koopmans, 1960; Halevy, 2015; Millner and Heal, 2018). A SWF is time invariant if it does not directly depend on the perspective of the assessment s^k. A SWF is stationary if it implies constant relative valuations between individuals over time. In this paper, we study welfare criteria that are time invariant but not stationary. The functional form of $\mathcal{W}(\cdot)$ does not vary with the perspective of the assessment s^k. But the relative valuation of individuals may not remain constant across dates and histories. This is precisely encoded in the dependence of the Pareto weights $\alpha^i_{s^k}$ on the assessment perspective s^k. This violation of stationarity is what opens the door to time inconsistency. Indeed, Halevy (2015) showed that time invariance and stationarity would necessarily imply time consistency.

is at the heart of the welfarist approach.[11] We denote by

$$\alpha^i_{s^k} = \frac{\partial \mathcal{W}(\{V^i_{s^k}\}_i)}{\partial V^i_{s^k}} > 0$$

the marginal contribution of individual i's lifetime utility to social welfare from the perspective of history s^k. And we assume that the partial derivatives of the social welfare function $\mathcal{W}$ are positive for all i. Notice that $\alpha^i_{s^k}$ depends on the time of the assessment — as emphasized again by the $_{s^k}$ notation — not because the SWF $\mathcal{W}(\cdot)$ varies with time or across histories but because it is evaluated at individuals' lifetime utilities $V^i_{s^k}$ as perceived from the perspective of s^k.

It is useful to distinguish between linear and non-linear social welfare functions. We refer to linear SWFs as utilitarian. When $\mathcal{W}(\cdot)$ is linear, the Pareto weights $\alpha^i_{s^k} = \alpha^i$ are constant across time and histories for each i, and we have $W_{s^k} = \sum_i \alpha^i V^i_{s^k}$. This is no longer the case when $\mathcal{W}(\cdot)$ is non-linear. Important examples of non-linear SWFs include isoelastic (Atkinson, 1970), Nash (Nash, 1950; Kaneko and Nakamura, 1979), and Rawlsian (Rawls, 1971). Non-linear social welfare functions are often used in applied work in macroeconomics and public finance (Atkinson, 1970; Atkinson and Stiglitz, 1976; Stern, 1976; Boskin and Sheshinski, 1978; Diamond, 1998; Saez, 2001; Benabou, 2002; Piketty and Saez, 2012, 2013a; Saez and Stantcheva, 2016; Heathcote et al., 2017).[12]

4 The Inconsistency of Welfare Assessments

We start our analysis by asking under what conditions the overall welfare assessment dW_{s^0} is time consistent. We define time consistency as follows.

Definition 1 (Time Consistency of Welfare Assessments)**.** *Consider a perturbation $d\theta_t(s^t)$. We say that the welfare assessment $\frac{dW_{s^0}}{d\theta_t(s^t)}$ from the perspective of s^0 is* time consistent *if later assessments $\frac{dW_{s^k}}{d\theta_t(s^t)}$ share the same sign for all histories $s^0 < s^k \leq s^t$. We say that the welfare assessment is time* in*consistent if there is a history s^k for which this does not hold.*

A welfare assessment is therefore time consistent if all later assessments of the same perturbation have the same sign. In that case, all later assessments agree on whether or not the perturbation is desirable. If $\boldsymbol{\theta}$ is a policy set by a planner, then consistency in the sign of the welfare assessment is sufficient to ensure the time consistency of the optimal policy.[13]

[11] We refer to the use of social welfare functions as the *welfarist approach* as in Bergson (1938), Samuelson (1947) or Boadway and Bruce (1984). As explained in Kaplow (2011), the critical restriction of a social welfare function is that $\mathcal{W}(\cdot)$ may not depend on any model outcomes other than individual lifetime utilities.

[12] Also see Rawls (1971), Kaplow (2011) and Adler and Fleurbaey (2016) for further discussion.

[13] For a policy $\boldsymbol{\theta}^*$ to be optimal from the perspective of date 0, it is necessary that no welfare gains can be achieved from any feasible perturbation $d\theta_t(s^t)$, that is, $dW_{s^0}/d\theta_t(s^t) = 0$ for all s^t when evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Therefore, if all relevant ex post assessments agree on the sign with the ex ante assessment for all perturbations $d\theta$, then $\boldsymbol{\theta}^*$ is a time consistent optimal policy. In other words, when ex ante and ex post assessments agree on the sign of dW, they agree on

For ease of exposition, in the main text we focus on the specific case of a perturbation $d\theta_1(\bar{s}^1)$ at a particular history $\bar{s}^1$ at date 1, which we refer to simply as "$d\theta$". We refer to $\frac{dW_{s^0}}{d\theta}$ as the *ex ante* assessment and to $\frac{dW_{\bar{s}^1}}{d\theta}$ as the *ex post* assessment at history $\bar{s}^1$. Time consistency then requires that $\frac{dW_{s^0}}{d\theta}$ and $\frac{dW_{\bar{s}^1}}{d\theta}$ have the same sign since $\bar{s}^1$ is the only relevant history from the perspective of which the perturbation $d\theta_1(\bar{s}^1)$ might be reassessed.[14] Our proofs in Appendix B are for arbitrary perturbations.

Representative agent benchmark. To set the stage, we review the representative agent (RA) benchmark with $I = 1$. When individual preferences are consistent (Strotz, 1956), there are three distinct time consistency problems than can emerge in environments with a representative agent.

Proposition 1 (Kydland and Prescott, 1977)**.** *Suppose $I = 1$. For the date 0 welfare assessment of perturbation $d\theta$ to be time inconsistent, one of the following must be true:*

(i) (Forward-looking behavior) *θ affects the consumption allocation at date 0*

(ii) (Road not taken) *θ affects the consumption allocation at a history $s^1 \neq \bar{s}^1$ at date 1*

(iii) (State variables) $\frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid s^0)}{\partial\theta} \neq \frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid \bar{s}^1)}{\partial\theta}$

To interpret Proposition 1, notice that we can write the ex ante welfare assessment as

$$\frac{dW_{s^0}}{d\theta} = \sum_i \alpha^i_{s^0} \frac{dV^i_{s^0}}{d\theta} = \alpha^i_{s^0} \sum_t \beta^t \sum_{s^t} \pi(s^t)\,(u^i)'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial\theta},$$

where the second equality follows because $I = 1$, and the ex post assessment at history $\bar{s}^1$ as

$$\frac{dW_{\bar{s}^1}}{d\theta} = \alpha^i_{\bar{s}^1} \sum_{t\geq 1} \beta^{t-1} \sum_{s^t \geq \bar{s}^1} \pi(s^t \mid \bar{s}^1)\,(u^i)'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid \bar{s}^1)}{\partial\theta}.$$

Comparing these two expressions highlights three distinct source of time inconsistency. First, the ex ante assessment features terms corresponding to date 0, whereas the ex post assessment sums over periods $\sum_{t\geq 1}$ from date 1 onwards. If the perturbation $d\theta$ affects consumption at date 0 due to *forward-looking behavior*, then this will affect the ex ante but not the ex post assessment. Second, the ex ante assessment features terms corresponding to the perturbation's impact on the allocation at histories $s^1 \neq \bar{s}^1$. Ex post, on the other hand, these alternative histories did not realize and so the planner no longer considers these terms. We refer to this source of time inconsistency as *the*

the direction in which policy should be moved. In that case, they both find the allocation induced by $\boldsymbol{\theta} + d\theta_t(s^t)$ more desirable than the one associated with $\boldsymbol{\theta}$. This is the appropriate definition of consistency in marginal form. If the two assessments agree on the sign at all possible levels of $\boldsymbol{\theta}$, then they agree on the policy $\boldsymbol{\theta}^*$ at which dW is 0. This is why we focus on sign consistency in this paper.

[14] If the ex ante and ex post assessment are both 0, we say they are consistent. For example, if a perturbation only affects the allocation prior to the perspective of the initial assessment, then both ex ante and ex post assessments will be 0 and thus trivially agree.

road not taken. Third, the perturbation's effect on the consumption allocation at history $\bar{s}^1$ may differ depending on the perspective or time of the announcement. In that case, $\frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid s^0)}{\partial \theta} \neq \frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid \bar{s}^1)}{\partial \theta}$, which corresponds to condition (iii) of Proposition 1.[15] Finally, notice that individual weights $\alpha^i_{s^0}$ and $\alpha^i_{\bar{s}^1}$ must be strictly positive and can therefore not be a source of inconsistency.[16]

Welfare assessments with heterogeneous agents. Having set the stage by reviewing the RA benchmark, we now study the time consistency of welfare assessments in heterogeneous agent (HA) environments with $I > 1$. Our next result shows that, in the presence of heterogeneity, interpersonal welfare comparisons emerge as a new source of time inconsistency.

Proposition 2 (Inconsistency of Welfare Assessments)**.** *Suppose $I > 1$. For the date 0 welfare assessment of perturbation $d\theta$ to be time inconsistent, either one of (i) – (iii) in Proposition 1 must be true or:*

(iv) (Interpersonal welfare comparisons) $\alpha^i_{\bar{s}^1} / \alpha^i_{s^0} \neq \alpha^j_{\bar{s}^1} / \alpha^j_{s^0}$ *for at least two individuals i and j.*

Proposition 2 demonstrates that a new source of time inconsistency emerges in heterogeneous agent environments, in addition to those identified by Strotz (1956) and Kydland and Prescott (1977). With heterogeneous agents, interpersonal welfare comparisons can lead to time consistency problems.

To explain the intuition behind Proposition 2, it is useful to consider the case where conditions (i) – (iii) do not apply, allowing us to focus on the new condition (iv). We make the following "no-KP" assumption, ruling out time consistency problems in the spirit of Kydland and Prescott (1977).

Assumption A1. (No-Kydland-Prescott) *We assume that (i) $\frac{\partial \mathcal{C}_0^i(s^0 \mid s^0)}{\partial \theta} = 0$; (ii) $\frac{\partial \mathcal{C}_1^i(s^1 \mid s^0)}{\partial \theta} = 0$ for $s^1 \neq \bar{s}^1$; and (iii) $\frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid s^0)}{\partial \theta} = \frac{\partial \mathcal{C}_1^i(\bar{s}^1 \mid \bar{s}^1)}{\partial \theta}$ for all i.*

In environments that satisfy Assumption A1, we can write the ex ante welfare assessment of the

[15] Notice that (i) is a prerequisite of (iii). Condition (iii) requires that the initial conditions (state variables) we take as given from the perspective of history $\bar{s}^1$ are affected by an announcement of the perturbation at date 0. This can only happen when the allocation at date 0 is affected, which in turn changes the state variables taken as given from the perspective of $\bar{s}^1$. Nonetheless, we find it useful to explicitly distinguish between conditions (i) and (iii) because they describe different economic channels through which time inconsistency can emerge.

[16] Kydland and Prescott (1977) focus on environments without uncertainty. Their main analysis only features conditions (i) and (iii) of Proposition 1. Specifically, the main condition in their text can be written using our notation as $\partial \mathcal{C}_0^i(s^0 \mid s^0)/\partial \theta \neq 0$. That is, the allocation at date 0 is affected by the date 0 announcement of a perturbation $d\theta$. This can lead to (i) or (iii) because, in order for state variables at the beginning of date 1 to be affected, necessarily behavior at date 0 must be affected. They note that the case of uncertainty can be treated by relabeling periods as histories in the spirit of Arrow-Debreu. While we find it useful to treat (ii) as a distinct source of time inconsistency, their paper is to be credited with identifying time consistency problems (i) – (iii). Proposition 1 should therefore be interpreted as a restatement of Kydland and Prescott (1977).

perturbation $d\theta$ as

$$\frac{dW_{s^0}}{d\theta} = \underbrace{\beta\pi(\bar{s}^1)\left(\frac{1}{I}\sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{\bar{s}^1}}\right)}_{>0} \frac{dW_{\bar{s}^1}}{d\theta} + \beta\pi(\bar{s}^1)\mathbb{C}ov_i^{\Sigma}\left(\frac{\alpha^i_{s^0}}{\alpha^i_{\bar{s}^1}},\ \alpha^i_{\bar{s}^1}\frac{dV^i_{\bar{s}^1}}{d\theta}\right), \tag{14}$$

which follows directly from our proof in Appendix B.2. Equation (14) decomposes the ex ante welfare assessment $\frac{dW_{s^0}}{d\theta}$ into a term that is proportional to the ex post assessment $\frac{dW_{\bar{s}^1}}{d\theta}$ and a cross-sectional covariance term. When the covariance term is 0, the ex ante and ex post assessments necessarily share the same sign because the coefficient on the later is strictly positive.[17] Time inconsistency therefore requires that the covariance term is non-zero — in fact, it must have the opposite sign from $\frac{dW_{\bar{s}^1}}{d\theta}$ — and is sufficiently large.

The covariance term can only be non-zero when the ratio of Pareto weights $\alpha^i_{s^0}/\alpha^i_{\bar{s}^1}$ differs across individuals. This precisely corresponds to condition (iv) of Proposition 2. $\alpha^i_{s^t}$ represents the planner's valuation of individual i's gains and losses in utility units from the perspective of history s^t. Whenever the ratios $\frac{\alpha^i_{s^0}}{\alpha^i_{\bar{s}^1}}$ are not equalized in the cross section, the planner's relative valuation of individuals in utils changes over time. The covariance term in equation (14) is positive whenever the planner puts a smaller weight ex post on those individuals who have a relatively large welfare gain from the perturbation.

According to Proposition 2, interpersonal welfare comparisons become a new source of time inconsistency in HA environments. We now show that utilitarian is the unique social welfare function that ensures time consistency in interpersonal welfare comparisons.

Corollary 1 (Time Consistency under Utilitarian SWF)**.** *The utilitarian social welfare function is the only welfarist criterion for which interpersonal welfare comparisons do not lead to time consistency problems because the Pareto weights $\alpha^i_{s^t} = \alpha^i$ are constant across histories for all individuals i.*

When the social welfare function $\mathcal{W}(\cdot)$ is utilitarian and therefore linear, the individual Pareto weights $\alpha^i_{s^t}$ are constant across histories for each i. The covariance term in (14) therefore vanishes, so that the welfare assessment $\frac{dW_{s^0}}{d\theta}$ is time consistent under assumption A1. The utilitarian SWF is the only one for which this is the case. This result goes back to Zuber (2011) who showed that welfare assessments based on additively separable social welfare functions are time consistent as long as individuals have a common discount factor. For more general SWFs, the weight the planner assigns to a particular individual changes endogenously with the allocation because in that case $\alpha^i_{s^t}$ is a function of lifetime utilities $V^i_{s^t}$ evaluated from the perspective of s^t.[18]

[17] To see this, note that the individual Pareto weights $\alpha^i_{s^0}$ and $\alpha^i_{\bar{s}^1}$ must be strictly positive. As are the discount factor β and the probability $\pi(\bar{s}^1)$ since we drop all histories from our analysis that realize with probability 0 without loss of generality.

[18] An interesting question is whether the analysis of Kydland and Prescott (1977) already nests Proposition 2 for the

Corollary 2 (Role of Incomplete Markets)**.** *Assume that the SWF $\mathcal{W}(\cdot)$ is symmetric and individual preferences are symmetric. Then welfare assessments are time consistent even for non-linear SWFs as long as markets are complete so that $u'(c_1^i(\bar{s}^1))/u'(c_0^i(s^0)) = u'(c_1^j(\bar{s}^1))/u'(c_0^j(s^0))$.*

Corollary 2 clarifies the role incomplete markets play for the time consistency of welfare assessments, as distinct from individual preference heterogeneity. In particular, when markets are complete and individual marginal rates of substitution between consumption at different histories are equalized, then welfare assessments are time consistent even when the social welfare function is non-linear.

5 The Inconsistency of Efficiency and Redistribution

We present our main result in this section: Efficiency and redistribution assessments are invariably time inconsistent in heterogeneous agent incomplete markets economies. Even when the overall welfare assessment $\frac{dW_{s^0}}{d\theta}$ is time consistent as in the special case of Corollary 1, whether welfare gains and losses are attributed to efficiency or redistribution therefore changes over time.

5.1 Kaldor-Hicks Efficiency

The notion of efficiency we work with in this paper is that of Kaldor-Hicks efficiency (Kaldor, 1939; Hicks, 1939). It defines the efficiency gain from a policy perturbation as the sum of individual willingness to pay expressed in a common unit.

Welfare numeraire and individual willingness to pay. The welfare assessment $\frac{dW_{s^t}}{d\theta} = \sum_i \alpha^i_{s^t} \frac{dV^i_{s^t}}{d\theta}$ compares gains and losses across different individuals in utils. The units of $\frac{dV^i_{s^t}}{d\theta}$ are history s^t individual utils. Those of $\frac{dW_{s^t}}{d\theta}$ are social utils. And the weights $\alpha^i_{s^t}$ represent the planner's valuation of individual i's gains and losses in utils.

Making interpersonal welfare comparisons in utility units is not meaningful since utility is *ordinal* — its units are meaningless as any monotonic transformation implies the same preference ordering. We start by expressing individual gains and losses in a comparable unit (such as a consumption-equivalent) and we refer to this unit as the *welfare numeraire*. Individual i's welfare

heterogeneous agent case after an appropriate relabeling. Suppose we relabel their "x_t" as the vector of all individuals' consumption. Then the question becomes whether their key term $\partial X_1/\partial \pi_2$ already captures condition (iv) of Proposition 2. The answer is no. With heterogeneous agents, their term "$\partial S/\partial x_2$" would differ across the ex ante and ex post assessments, which is precisely the content of Proposition 2. Kydland and Prescott (1977) notationally force this term to be constant across assessment perspectives, which in general rules out heterogeneous agent environments. However, the utilitarian social welfare function is precisely the special case where their term "$\partial S/\partial x_2$" would have the same sign from the ex ante and ex post perspective. A useful restatement of Corollary 1 would therefore be: Only under the utilitarian social welfare function does the analysis of Kydland and Prescott (1977) extend to heterogeneous agent environments by relabeling their "x_t" as the vector of all individuals' consumption.

gain or loss from the perturbation can then be expressed in units of the welfare numeraire as

$$\frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}}, \quad \text{where} \quad \lambda^i_{s^t} > 0 \quad \text{and} \quad \dim(\lambda^i_{s^t}) = \frac{\text{history } s^t \text{ individual } i \text{ utils}}{\text{history } s^t \text{ welfare numeraire}}.$$

We use the *individual normalizing factor* $\lambda^i_{s^t}$ to express individual gains and losses in a common unit — the welfare numeraire. This welfare numeraire is chosen at the time of the assessment, history s^t. The units of the normalizing factor are $\dim(\lambda^i_{s^t}) = \frac{\text{history } s^t \text{ individual } i \text{ utils}}{\text{history } s^t \text{ welfare numeraire}}$. We have therefore expressed individual gains and losses as a *willingness to pay* in units of welfare numeraire, i.e.,

$$\dim\left(\frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}}\right) = \frac{\frac{\text{history } s^t \text{ individual } i \text{ utils}}{\text{units of } \theta}}{\frac{\text{history } s^t \text{ individual } i \text{ utils}}{\text{history } s^t \text{ welfare numeraire}}} = \frac{\text{history } s^t \text{ welfare numeraire}}{\text{units of } \theta}.$$

Dividing $\frac{dV^i_{s^t}}{d\theta}$ by the normalizing factor $\lambda^i_{s^t}$ expresses individual gains and losses as a willingness to pay in units of welfare numeraire from the perspective of the point in time at which the welfare assessment takes place.[19]

The normalizing factor $\lambda^i_{s^t}$ and therefore the welfare numeraire itself may depend on the time of the welfare assessment s^t. Why is this the case? Suppose we choose date 0 consumption as the unit in which we express the welfare assessment and the interpersonal welfare comparisons which they aggregate at that point in time. If we reconsider such a welfare assessment at a later date, for example from the perspective of history s^1, expressing individual willingness to pay in units of date 0 consumption is no longer feasible and a different welfare numeraire must be chosen, such as history s^1 consumption for example. As a consequence, the units in which we express welfare assessments depends on the perspective of the assessment via $\lambda^i_{s^t}$ and may change over time.

Kaldor-Hicks efficiency. Having expressed individual welfare gains and losses in a common and meaningfully comparable unit, we now revisit the overall welfare assessment that compares and aggregates them. To express the assessment $\frac{dW_{s^t}}{d\theta}$ in units of the welfare numeraire as well, we divide by $\frac{1}{I}\sum_i \alpha^i_{s^t}\lambda^i_{s^t}$. A *normalized* welfare assessment from the perspective of history s^t can then be expressed in units of the welfare numeraire as

$$\frac{dW^\lambda_{s^t}}{d\theta} = \frac{1}{\frac{1}{I}\sum_i \alpha^i_{s^t}\lambda^i_{s^t}}\frac{dW_{s^t}}{d\theta} = \frac{1}{\frac{1}{I}\sum_i \alpha^i_{s^t}\lambda^i_{s^t}}\sum_i \alpha^i_{s^t}\lambda^i_{s^t}\frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}} = \sum_i \omega^i_{s^t}\frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}}, \tag{15}$$

[19] To express individual willingness to pay in units of date t history s^t consumption, for example, we set $\lambda^i_{s^t} = u'(c^i_t(s^t))$. Its units $\dim(u'(c^i_t(s^t)))$ are individual i utils per unit of consumption. And therefore the units of $\frac{dV^i_{s^t}}{d\theta}/\lambda^i_{s^t}$ would be units of date t history s^t consumption per unit of θ.

where $\omega^i_{s^t} = \frac{1}{\frac{1}{I}\sum_i \alpha^i_{s^t}\lambda^i_{s^t}} \alpha^i_{s^t}\lambda^i_{s^t}$ is a normalized individual welfare weight. We use the superscript $^\lambda$ notation to emphasize that the units of $\frac{dW^\lambda_{s^t}}{d\theta}$ are in welfare numeraire. If $\frac{dW^\lambda_{s^t}}{d\theta} = 1.1$, for example, then the planner values the policy change $d\theta$ equal to distributing 1.1 units of welfare numeraire equally across individuals at history s^t.

A normalized welfare assessment admits a unique decomposition into Kaldor-Hicks efficiency and redistribution (Dávila and Schaab, 2024), given by

$$\frac{dW^\lambda_{s^t}}{d\theta} = \sum_i \omega^i_{s^t} \frac{\frac{dV^i_{s^0}}{d\theta}}{\lambda^i_{s^t}} = \underbrace{\sum_i \frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}}}_{\Xi^E_{s^t}\,(\text{Efficiency})} + \underbrace{\mathbb{C}ov^\Sigma_i\left(\omega^i_{s^t}, \frac{\frac{dV^i_{s^t}}{d\theta}}{\lambda^i_{s^t}}\right)}_{\Xi^{RD}_{s^t}\,(\text{Redistribution})}. \tag{16}$$

The efficiency component $\Xi^E_{s^0}$ corresponds to Kaldor-Hicks efficiency gains — the unweighted sum of individual willingness to pay — from the perspective of s^0. The normalized individual welfare weight $\omega^i_{s^0}$ encodes how the SWF $\mathcal{W}$ trades off welfare gains and losses across individuals in units of welfare numeraire. For instance if $\omega^i_{s^0} = 1.3$, then the planner finds the welfare gain from giving 1 unit of numeraire to individual i equivalent to distributing 1.3 units equally across all individuals. This welfare weight may depend on the time of the assessment s^0 through both the unnormalized weight $\alpha^i_{s^0}$ and the choice of numeraire via $\lambda^i_{s^0}$.

The sum of willingness to pay $\Xi^E_{s^t}$ in units of welfare numeraire is a useful measure of efficiency gains and losses for at least three reasons. First, it is invariant to preference-preserving transformations of utilities $u^i(\cdot)$, whereas the overall welfare assessment dW_{s_t} is not. Second, it is also invariant to the choice of SWF $\mathcal{W}(\cdot)$. Notice from the definition of efficiency in (16) that $\Xi^E_{s^t}$ is invariant to $\alpha^i_{s^t}$, which only affect the normalized welfare weights $\omega^i_{s^t}$ and therefore the redistribution component. Finally, Kaldor-Hicks efficiency satisfies the compensation principle. When the Kaldor-Hicks efficiency assessment of a perturbation $d\theta$ is positive, then the perturbation can be turned into a Pareto improvement by implementing a set of compensating transfers across individuals. In fact, it is the only welfare criterion that satisfies this property, which we will refer to as "Paretian with transfers".

5.2 The Inconsistency of Efficiency Assessments

We now characterize the time consistency of efficiency assessments, which we define as follows.

Definition 2. *Consider a perturbation $d\theta_t(s^t)$. The efficiency assessment $\Xi^E_{s^0}$ from the perspective of s^0 is* time consistent *if later assessments $\Xi^E_{s^k}$ share the same sign for all histories $s^0 < s^k \leq s^t$. The efficiency assessment is time* in*consistent if there is a history s^k for which this does not hold.*

As before, we focus on a perturbation $d\theta_1(\bar{s}^1)$ in the main text, which we simply refer to as $d\theta$. And we compare the ex ante assessment from the perspective of s^0 with the ex post assessment at $\bar{s}^1$,

$$\Xi^E_{s^0} = \sum_i \frac{\frac{dV^i_{s^0}}{d\theta}}{\lambda^i_{s^0}} \qquad \text{and} \qquad \Xi^E_{\bar{s}^1} = \sum_i \frac{\frac{dV^i_{\bar{s}^1}}{d\theta}}{\lambda^i_{\bar{s}^1}}.$$

Our proofs in Appendix B are stated for the general case. Two observations follow from comparing the ex ante and ex post efficiency assessments. First, neither $\Xi^E_{s^0}$ nor $\Xi^E_{\bar{s}^1}$ depends on the Pareto weights $\alpha^i_{s^0}$ or $\alpha^i_{\bar{s}^1}$. As discussed in Section 5.1, Kaldor-Hicks efficiency is invariant to the underlying social welfare function. It follows that condition (iv) of Proposition 2, which concerns the evolution of relative Pareto weights over time, does not apply when evaluating efficiency.

Second, comparing ex ante and ex post efficiency assessments does require taking a stance on the welfare numeraires — that is, the units in which individual willingness to pay is expressed in each assessment. The next result shows that inconsistency in individuals' valuations for these numeraires can give rise to time inconsistency in efficiency assessments.

Proposition 3 (Inconsistency of Efficiency Assessments)**.** *Suppose $I > 1$. For the date 0 efficiency assessment of perturbation $d\theta$ to be time inconsistent, either one of (i) – (iii) in Proposition 1 must be true or:*

(v) (Numeraire inconsistency) $\lambda^i_{\bar{s}^1} / \lambda^i_{s^0} \neq \lambda^j_{\bar{s}^1} / \lambda^j_{s^0}$ *for at least two individuals i and j.*

Proposition 3 establishes that efficiency assessments in heterogeneous agent environments can be time inconsistent even when Pareto weights $\alpha^i_{s^t}$ remain fixed over time. This form of inconsistency arises independently of the underlying social welfare function. In particular, even when the overall welfare assessment is time consistent — as in the special case of Corollary 1 — the attribution of welfare gains and losses to efficiency and redistribution invariably changes over time.

Proposition 3 therefore identifies a novel source of time inconsistency that emerges in heterogeneous agent environments due to the choice of welfare numeraires. While time inconsistency in the overall welfare assessment results from the evolution of Pareto weights $\alpha^i_{s^t}$ under a non-linear social welfare function, efficiency assessments remain invariant to such changes. Instead, a new condition — numeraire inconsistency — emerges as the relevant criterion. The inconsistency stems not from a change in the social valuation of individual utilities, but from a change in the reference units used to aggregate individual consumption-equivalent valuations.[20]

What then explains the inconsistency of efficiency assessments? To build intuition for the economics underlying Proposition 3, we derive a useful representation of the relationship between

[20] In any RA environment, welfare and efficiency coincide, so our discussion of time consistency in Proposition 1 also applies to efficiency assessments. In HA environments, however, we show that the conditions for time consistency of the efficiency assessment are different from those for welfare assessments.

the ex ante and ex post efficiency assessments. Under assumption A1, we can write

$$\Xi^E_{s^0} = \underbrace{\beta\pi(\bar{s}^1)\left(\frac{1}{I}\sum_i \frac{\lambda^i_{\bar{s}^1}}{\lambda^i_{s^0}}\right)}_{>0} \Xi^E_{\bar{s}^1} + \beta\pi(\bar{s}^1)\mathbb{C}ov_i\left(\frac{\lambda^i_{\bar{s}^1}}{\lambda^i_{s^0}}, \frac{\frac{dV^i_{\bar{s}^1}}{d\theta}}{\lambda^i_{\bar{s}^1}}\right). \tag{17}$$

In other words, the ex ante and ex post assessments are proportional to each other (the scalar coefficient on $\Xi^i_{\bar{s}^1}$ is always positive), except for the covariance term on the RHS.[21] If the covariance term were 0, the assessment would be time consistent since the sign would always be preserved. However, whenever the covariance term has the opposite sign from $\Xi^E_{\bar{s}^0}$ and is sufficiently large, the efficiency assessment becomes inconsistent. And the covariance term is positive when those individuals who have a large willingness to pay for the perturbation in units of ex post numeraire also have a high relative valuation for ex post numeraire relative to ex ante numeraire.

In Section 4, we studied welfare assessments that compared gains and losses across individuals in utility units. Time inconsistency emerged whenever the planner's relative valuation for individuals in utils changed over time — the ratio $\frac{\alpha^i_{\bar{s}^1}}{\alpha^i_{s^0}}$ in the covariance term of equation (14). Utilitarian welfare assessments remain time consistent precisely because the planner's ex ante and ex post valuation of gains and losses for an individual i remain constant in utils. When we make efficiency assessments, however, we compare gains and losses across individuals in a common welfare numeraire. The normalizing factor $\lambda^i_{s^t}$ captures individual i's valuation of this numeraire from the perspective of history s^t. That is, $\lambda^i_{s^t}$ captures how many utils individual i derives from one unit of the numeraire. And the ratio $\lambda^i_{\bar{s}^1}/\lambda^i_{s^0}$ captures the relative valuation of individual i for 1 unit of ex ante welfare numeraire in units of ex post welfare numeraire. Intuitively, when this ratio is large, the individual's willingness to pay for the same util gain is higher when expressed in units of ex ante welfare numeraire. The covariance term in equation (17) is therefore positive when those individuals with a large willingness to pay ex post — so $\frac{1}{\lambda^i_{\bar{s}^1}}\frac{dV^i_{\bar{s}^1}}{d\theta}$ is large — also have a higher relative valuation for welfare gains ex post — so $\lambda^i_{\bar{s}^1}/\lambda^i_{s^0}$ is large.

Individuals' valuations for the ex ante and ex post numeraires may differ independently from the underlying social welfare function. Corollary 1 no longer applies to efficiency assessments: Even though a utilitarian planner puts the same weight on individual gains and losses in utils ex ante and ex post, the individual's own valuation of the numeraires may differ regardless.

Forward- and backward-looking numeraires. Proposition 3 suggests a natural approach to ensure the time consistency of efficiency assessments: choosing ex ante and ex post numeraires such that $\lambda^i_{s^0} = \lambda^i_{\bar{s}^1}$ for all i. If each individual's valuation of the numeraire is constant across time

[21] The scalar multiplying $\Xi^E_{\bar{s}^1}$ is always positive because welfare numeraires $\lambda^i_{s^t}$ and the discount factor β are strictly positive by assumption. The probability of history $\bar{s}^1$ is weakly positive. But whenever $\pi(\bar{s}^1)$ is 0, we also have $\Xi^E_{s^0} = 0$, in which case the sign is again preserved. We assume $\pi(\bar{s}^1) > 0$ in the main text for simplicity.

and histories, then the covariance term in (17) drops out. It will be useful to distinguish between two classes of possible welfare numeraires: forward- and backward-looking numeraires.

Definition 3 (Forward-Looking Welfare Numeraire). *A welfare numeraire is* forward-looking *from the perspective of history s^k if individual i's valuation $\lambda^i_{s^k}$ is a function only of $\{c^i_t(s^t)\}_{t\geq k, s^t \geq s^k}$.*

If $\lambda^i_{s^k}$ is also a function of $\{c^i_t(s^t)\}_{t<k, s^t}$ or $\{c^i_t(s^t)\}_{t, s^t \ngeq s^k}$, on the other hand, we say that the associated numeraire is *backward-looking*. In particular, the numeraire $\lambda^i_{\bar{s}^1}$ we use for the ex post efficiency assessment is forward-looking if it is a function only of consumption at history $\bar{s}^1$ or at future histories s^k that can be reached from $\bar{s}^1$. When we use a forward-looking numeraire, we express individuals' willingness to pay in units of contemporaneous or future consumption. Differences in marginal utilities of past consumption do not influence the valuation. The two most common examples of forward-looking numeraires are contemporaneous consumption, with $\lambda^i_{\bar{s}^1} = (u^i)'(c^i_1(\bar{s}^1))$, and perpetual consumption, with $\lambda^i_{\bar{s}^1} = \sum_{t\geq 1} \sum_{s^t \geq \bar{s}^1} \beta^{t-1} \pi(s^t \mid \bar{s}^1)(u^i)'(c^i_t(s^t))$.

One natural candidate to ensure $\lambda^i_{s^0} = \lambda^i_{\bar{s}^1}$ for all i is to use the backward-looking numeraire associated with $\lambda^i_{s^0}$ when making the ex post assessment in history $\bar{s}^1$. The resulting efficiency assessment would then compare gains and losses across individuals in units of past (last period's) consumption. Efficiency assessments based on backward-looking numeraires no longer satisfy the compensation principle, however, as we show in Proposition 4 below. In other words, even if the implied sum of willingness to pay is positive in units of the backward-looking numeraire, it would not be possible to construct (hypothetical) compensating transfers to turn the perturbation into a Pareto improvement. That is because individuals have no value for history s^0 consumption from the perspective of history $\bar{s}^1$. So we cannot use this unit for compensation. In summary, while it appears straightforward to ensure the time consistency of efficiency assessments by making the ex post assessment in units of a backward-looking numeraire, such assessments no longer satisfy the compensation principle. We formalize these observations in the next Proposition.

Proposition 4 (Properties of Welfare Numeraires). *Forward-looking welfare numeraires satisfy the following two conditions:*

(i) The perturbation $d\theta$ is Paretian with transfers in units of a forward-looking numeraire if and only if $\Xi^E_{s^t} > 0$ in units of that numeraire.

(ii) It is impossible to find feasible perturbations from Pareto efficient allocations that satisfy $\Xi^E_{s^t} > 0$ for any forward-looking numeraire.

Efficiency assessments must therefore use forward-looking numeraires in order to be Paretian with transfers. But forward-looking numeraires naturally invite time consistency problems. Consider the case where we use contemporaneous consumption for both the ex ante and ex post assessments, so $\lambda^i_{s^0} = (u^i)'(c_0(s^0))$ and $\lambda^i_{\bar{s}^1} = (u^i)'(c^i_1(\bar{s}^1))$. Whenever individuals' valuations for consumption at these two histories are not equalized, the covariance term in equation (17) will be non-zero.[22]

Complete markets. There is a special class of environments in which numeraire inconsistency never arises and efficiency assessments are therefore time consistent under assumption A1. This is true whenever marginal rates of substitution between histories are equalized across all individuals — a condition satisfied in complete markets environments. Under complete markets,

$$\frac{(u^i)'(c^i_1(\bar{s}^1))}{(u^i)'(c^i_0(s^0))} = \frac{(u^j)'(c^j_1(\bar{s}^1))}{(u^j)'(c^j_0(s^0))} \tag{18}$$

for all individuals i and j. The next result shows that this condition guarantees numeraire consistency.

Proposition 5 (Consistency under Complete Markets)**.** *In environments in which all marginal rates of substitution are equalized, so equation* (18) *is satisfied, we have*

$$\frac{\lambda^i_{\bar{s}^1}}{\lambda^i_{s^0}} = \frac{\lambda^j_{\bar{s}^1}}{\lambda^j_{s^0}} \qquad \textit{for all } i \textit{ and } j. \tag{19}$$

Assumption A1 then guarantees the time consistency of efficiency assessments.

When markets are complete, individuals' relative valuation of consumption at any two histories is equalized (equation 18). Any valid welfare numeraire must correspond to a bundle of consumption at different histories and $\lambda^i_{s^t}$ represents individual i's valuation of that bundle. Therefore, if (18) holds for all pairs of histories, then valuations for any combination of histories must also be equalized, which directly implies (19). Complete markets therefore guarantee that individuals' valuations of the ex ante and ex post welfare numeraires are equalized. Condition (v) of Proposition 3 therefore never applies in complete markets environments. And under the no-KP assumption A1, which rules out conditions (i) – (iii) of Proposition 1, efficiency assessments are therefore always time consistent.

[22] One might consider using consumption at history $\bar{s}^1$ as the numeraire for the ex ante assessment at date 0. Both assessments would then be based on forward-looking numeraires and trivially satisfy $\lambda^i_{s^0} = \lambda^i_{\bar{s}^1}$. This will not be possible in all but the simplest settings, however. Consider a policy that affects the allocation not only in history $\bar{s}^1$ but also in a different history $\tilde{s}^1$ at date 1. Time consistency then requires that ex post assessments in all relevant histories at date 1 share the same sign as the ex ante assessment at date 0. If we make the ex ante assessment using history $\bar{s}^1$ consumption as the numeraire and history $\bar{s}^1$ realizes ex post, no numeraire inconsistency will arise and the assessment will be consistent. If history $\tilde{s}^1$ realizes, however, numeraire inconsistency will again be a problem as long as individuals' valuations of consumption across these two histories are not equalized.

5.3 Impossibility Result

Drawing on our discussion thus far, we now present the central result of this paper: We show that it is generally not possible to make efficiency assessments in heterogeneous agent incomplete markets environments that both (i) satisfy the compensation principle and (ii) are time consistent. Both qualifications are critical. This result only applies to heterogeneous agent economies with $I > 1$. It also requires that marginal rates of substitution across consumption in different histories are not equalized for all individuals in the cross section. We summarize these conditions in the following definition.

Definition 4. *We refer as* heterogeneous agent incomplete markets *economies to environments with $I > 1$ where there are at least two individuals i and j whose marginal rates of substitution between histories s^0 and $\bar{s}^1$ are not equalized,*

$$\frac{(u^i)'(c_1^i(\bar{s}^1))}{(u^i)'(c_0^i(s^0))} \neq \frac{(u^j)'(c_1^j(\bar{s}^1))}{(u^j)'(c_0^j(s^0))}. \tag{20}$$

We are now ready to state our main result.

Theorem 1 (Impossibility Result)**.** *In heterogeneous agent incomplete markets economies, it is not possible to make efficiency assessments that satisfy the following three properties for all $d\theta$:*

P1 *Forward-looking numeraires*

P2 *Time consistency*

In the presence of heterogeneity and market incompleteness, it is not possible to make efficiency assessments that simultaneously satisfy the desirable properties P1 – P3.[23] Theorem 1 shows that there is an intrinsic tension between making efficiency assessments that satisfy the compensation principle (are Paretian with transfers) and are time consistent.

This result is not about one of the sources of time inconsistency identified by Kydland and Prescott (1977), i.e., conditions (i) – (iii) of Proposition 1. Theorem 1 applies even to environments that additionally satisfy assumption A1, shutting down time consistency problems á la Kydland and Prescott (1977), which also apply to representative agent environments. Instead, the key forces driving Theorem 1 are unique to heterogeneous agent environments with $I > 1$.

A key step in the proof of Theorem 1 shows that property P2 implies that P3 cannot be true for all $d\theta$ whenever markets are incomplete. The argument builds on our discussion in Section 5.2. Kaldor-Hicks efficiency uniquely satisfies the compensation principle — and is therefore Paretian

[23] Theorem 1 is a statement of generic impossibility in the spirit of Arrow (1950). What this means is that, taking as given a particular environment that satisfies $I > 1$ and (20), it is impossible to make an efficiency assessment that satisfies P1 – P3 *for all $d\theta$*. In other words, we can always construct a perturbation $d\theta$ whose efficiency assessment violates one of P1 – P3. The "for all $d\theta$" condition is therefore parallel in spirit to Arrow (1950)'s condition of universal domain.

with transfers — when it is based on forward-looking numeraires. And we show in our proof in Appendix B.5 that making both the ex ante and ex post efficiency assessments using forward-looking numeraires (generically) leads to time inconsistency. That is, it is possible to construct perturbations $d\theta$ whose ex ante and ex post efficiency assessments will differ in sign.

5.4 The Origins of Efficiency Gains

We have so far characterized the time consistency of the overall welfare assessment and of the attribution of welfare gains and losses to efficiency and redistribution. We now unpack the efficiency assessment and discuss the time consistency of aggregate efficiency, risk-sharing, and intertemporal-sharing assessments.

To that end, we start by defining normalized dynamic and stochastic welfare weights. Like the normalized individual welfare weights $\omega^i_{s^t}$, these weights capture the key marginal rates of substitution that govern the planner's relative valuation of changes in consumption bundles. Given a welfare numeraire $\lambda^i_{s^t}$, we define

$$\omega^i_{t|s^k} = \frac{1}{\lambda^i_{s^k}} \beta^{t-k} \sum_{s^t \geq s^k} \pi(s^t \mid s^k) u'(c^i_t(s^t))$$

as the normalized dynamic welfare weight for consumption of individual i in period t from the perspective of history s^k. Intuitively, $\omega^i_{t|s^k}$ captures individual i's marginal rate of substitution between a unit of consumption at date t (in all histories) and one unit of welfare numeraire. The former is valued at $\beta^{t-k} \sum_{s^t \geq s^k} \pi(s^t \mid s^k) u'(c^i_t(s^t))$ and the latter at $\lambda^i_{s^k}$. Since the social welfare function $\mathcal{W}(\cdot)$ is non-paternalistic by construction, the planner's MRS between these two bundles coincides with individual i's private MRS. If $\omega^i_{t|s^k} = 0.2$, for example, then individual i (and therefore the planner) is indifferent between 1 additional unit of consumption at date t and 0.2 units of welfare numeraire.

We also define the normalized stochastic welfare weight

$$\omega^i_{t|s^k}(s^t) = \frac{\pi(s^t \mid s^k) u'(c^i_t(s^t))}{\sum_{s^t \geq s^k} \pi(s^t \mid s^k) u'(c^i_t(s^t))}$$

for consumption of individual i in history s^t from the perspective of history s^k. $\omega^i_{t|s^k}(s^t)$ captures individual i's MRS between a unit of consumption in history s^t and a unit of consumption at date t (in all histories). If $\omega^i_{t|s^k}(s^t) = 0.4$, then individual i (and therefore the planner) is indifferent between 1 additional unit of consumption in history s^t and 0.4 units of consumption in all histories at date t.

Finally, we define the cross-sectional averages of these normalized welfare weights by

$$\bar{\omega}_{t|s^k} = \frac{1}{I} \sum_i \omega^i_{t|s^k} \qquad \text{and} \qquad \bar{\omega}_{t|s^k}(s^t) = \frac{1}{I} \sum_i \omega^i_{t|s^k}(s^t).$$

We are now ready to characterize the sources of efficiency gains and losses from the perspective of history s^k.

Lemma 1 (Origins of Efficiency)**.** *The efficiency assessment $\Xi^E_{s^k}$ of a perturbation $d\theta$ from the perspective of history s^k can be decomposed into aggregate efficiency, risk-sharing, and intertemporal-sharing components,*

$$\Xi^E_{s^k} = \Xi^{AE}_{s^k} + \Xi^{RS}_{s^k} + \Xi^{IS}_{s^k},$$

where

$$\Xi^{AE}_{s^k} = \sum_{t \geq k} \bar{\omega}_{t|s^k} \sum_{s^t \geq s^k} \bar{\omega}_{t|s^k}(s^t) \sum_i \frac{\partial \mathcal{C}^i_t(s^t \mid s^k)}{\partial \theta}$$

$$\Xi^{RS}_{s^k} = \sum_{t \geq k} \bar{\omega}_{t|s^k} \sum_{s^t \geq s^k} \mathbb{C}ov^{\Sigma}_i \left(\omega^i_{t|s^k}(s^t), \frac{\partial \mathcal{C}^i_t(s^t \mid s^k)}{\partial \theta} \right)$$

$$\Xi^{IS}_{s^k} = \sum_{t \geq k} \mathbb{C}ov^{\Sigma}_i \left(\omega^i_{t|s^k}, \sum_{s^t \geq s^k} \omega^i_{t|s^k}(s^t) \sum_i \frac{\partial \mathcal{C}^i_t(s^t \mid s^k)}{\partial \theta} \right)$$

Lemma 1 characterizes the sources of efficiency gains and losses from the perspective of history s^k, building on results in Dávila and Schaab (2024).

Proposition 6 (Aggregate Efficiency)**.** *Under assumption A1, assessments of static aggregate efficiency gains and losses are time consistent.*

This result underscores that assessments of production efficiency gains and losses are special. While risk-sharing and redistribution assessments change over time, assessments of production efficiency gains and losses are time consistent. This highlights that some components of the overall efficiency and welfare assessment are time consistent, while others are not.

5.5 Minimal Example

We conclude this section with a minimal example that illustrates our results. There are two periods indexed by $t \in \{0, 1\}$ and two individuals indexed by $i \in \{a, b\}$. At the beginning of the second period, a random event occurs and one of two states indexed by $s \in \{h, \ell\}$ realizes with probability $\pi(s) = \frac{1}{2}$.

Individual i's preferences are given by $V^i_0 = u(c^i_0) + \beta \sum_s \pi(s) u(c^i_1(s))$. We assume that individuals consume their endowments in both periods. In period 0, both individuals receive and consume the same endowment of 2. In period 1, individuals' expected consumption is still 2, but they receive a higher endowment of 3 in one state and a lower endowment of 1 in the other state,

$$c_1^a(h) = 3 - \theta \qquad \text{and} \qquad c_1^b(h) = 1 + \theta$$
$$c_1^a(\ell) = 1 + \theta \qquad \text{and} \qquad c_1^b(\ell) = 3 - \theta$$

Endowment risk is idiosyncratic in the sense that aggregate consumption is 4 in both states, and the two individuals' risk exposure is symmetric. Finally, θ denotes a social transfer policy that shifts endowment from the lucky to the unlucky individual in both states.

What are the sources of welfare gains and losses from a marginal social transfer $d\theta > 0$ at an initial allocation with $\theta = 0$? From the perspective of period 0, both individuals are identical. They would prefer smoothing their consumption across states h and ℓ but are unable to do so. The social transfer θ mimics the missing insurance claims that both individuals would happily trade. The perturbation $d\theta > 0$ therefore improves risk-sharing and achieves strict efficiency gains. In fact, it is a strict Pareto improvement with $\frac{dV_0^i}{d\theta} > 0$ for all i since it allows both individuals to smooth consumption.

This assessment is not time consistent. Suppose time passes and state h realizes: a has an endowment of 3 while b only has an endowment of 1. From this ex post perspective, there is no longer any scope for risk-sharing and transferring from a to b appears as pure redistribution. The perturbation is no longer a Pareto improvement since $\frac{dV_1^a(h)}{d\theta} < 0 < \frac{dV_1^b(h)}{d\theta}$.

6 Anticipated Relief Policies

This application identifies a particular form of time inconsistency that arises in the context of anticipated relief policies — transfer schemes intended to support individuals experiencing temporary spells of low consumption followed by a gradual, anticipated recovery. We consider environments in which market incompleteness prevents affected individuals from borrowing to smooth consumption, and study the design and time consistency of transfer policies aimed at mitigating the welfare costs of such shocks.

Anticipated relief policies of this sort are common. A salient example is the COVID-19 assistance programs enacted through the CARES Act, which provided temporary support to households disproportionately affected by the aggregate shock. But the time consistency problem we identify also emerges in other policy domains, including childcare support and parental leave, unemployment insurance, retraining programs, and affirmative action. In each case, the policy is designed ex ante to offset a foreseeable period of hardship and is intended to phase out once recovery is underway.

We show that in such settings, a particular form of time inconsistency arises: ex post, the planner finds it optimal to extend the relief longer than originally intended. Because markets are incomplete, individuals who experience transitory spells of low consumption cannot borrow against future income to smooth consumption. A government relief policy can therefore generate efficiency gains by replicating the consumption-smoothing individuals would choose if borrowing were feasible. Individuals prefer smoothing consumption relative to their lifetime (average) consumption level. From the ex ante perspective, lifetime consumption is relatively low, since the worst of the shock still lies ahead. The efficiency-maximizing transfer policy therefore delivers support early and phases out once consumption recovers to the ex ante lifetime consumption level.

Over time, the shock dissipates and a gradual recovery takes place. When the planner revisits the relief policy ex post, the worst of the shock now lies in the past and lifetime consumption — looking forward — now appears higher. From this ex post perspective, current consumption still falls below this updated lifetime consumption level and the individual would like to continue borrowing. The time at which consumption recovers to its lifetime level now appears farther in the future. When reassessed ex post, the efficiency-maximizing transfer is therefore extended beyond the originally optimal time of expiry.

The result is a dynamic inconsistency in the design of anticipated relief policies. From the ex ante perspective, the optimal policy phases out relatively early in the recovery. From the ex post perspective, however, the perceived lifetime consumption benchmark shifts upward over time, delaying the point at which it appears efficient to withdraw the transfer. Although the policy is initially designed to be temporary, time inconsistency leads to its gradual extension.

6.1 Environment

We consider a deterministic infinite-horizon economy with dates $t \in \{0, 1, \ldots\}$. There are two individuals indexed $i \in \{a, b\}$, and there is a single good that appears as an endowment.

Preferences. Individual i's lifetime utility from the perspective of date k is defined as

$$V_k^i = \sum_{t \geq k} \beta^{t-k} u(c_t^i).$$

Endowments. We denote individual i's endowment at date t by y_t^i. Individual a has a constant endowment of $y_t^a = y$. Individual b experiences a transitory spell of low endowments that starts at date $\underline{T}$ and lasts until $\overline{T}$, during which she gradually recovers to the endowment level y of individual a. That is,

$$y_t^b = \begin{cases} y & t < \underline{T} \\ \underline{y} + \frac{t - \underline{T}}{\overline{T} - \underline{T}}(y - \underline{y}) & \underline{T} \leq t \leq \overline{T} \\ y & t > \overline{T} \end{cases}$$

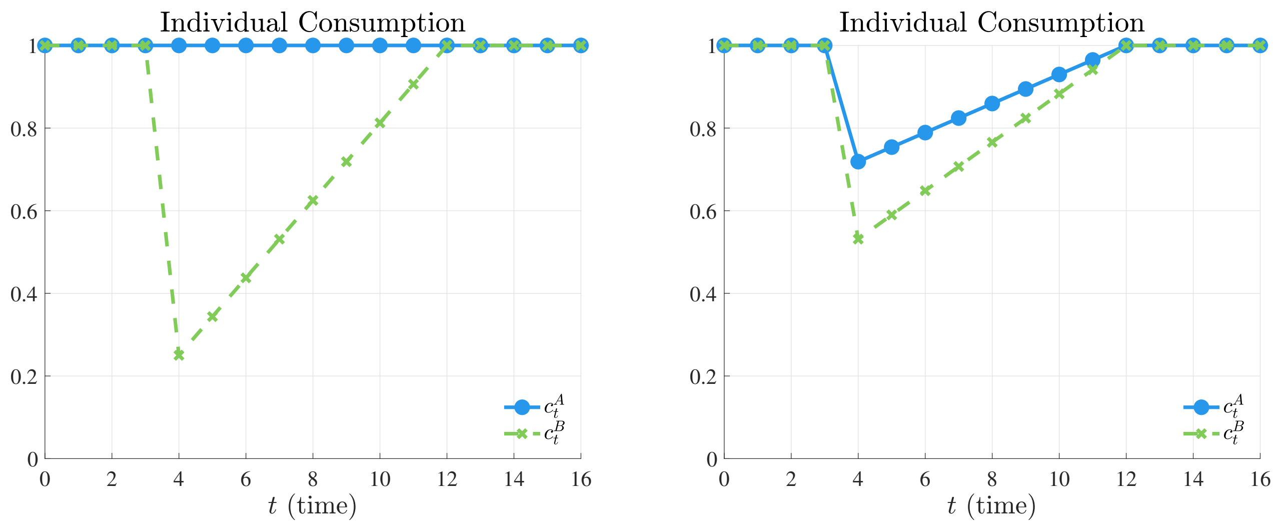

Figure 3. Consumption Profile

Relief policy. We focus on financial autarky for illustration, so individual i's consumption at date t is

$$c_t^i = y_t^i + \theta R_t^i$$

where R_t^i represents a transfer scheme that provides relief to individual b during the low-endowment spell, with $R_t^a = \frac{1}{2}(y_t^b - y_t^a)$ and $R_t^b = -R_t^a$. We plot the two individuals' consumption profiles in Figure 3 without the relief policy ($\theta = 0$) in Panel (a) and with a sizable transfer ($\theta = 0.75$) in Panel (b).

Parametrization. We assume log utility $u(c) = \log c$ and use a discount factor of $0.95^{0.25}$. We set the endowment levels to $y = 1$ and $\underline{y} = 0.25$. Finally, we assume that the low-endowment spell starts at date $\underline{T} = 4$ and lasts until $\overline{T} = 12$.

6.2 Welfare Assessment

We study welfare assessments of the relief policy R_t^i under an equal-weighted utilitarian social welfare function. Social welfare from the perspective of date k is

$$W_k = V_k^a + V_k^b.$$

We can characterize efficiency and redistribution gains and losses from the perspective of date k as $\frac{dW_k^\lambda}{d\theta} = \Xi_k^E + \Xi_k^{RD}$. Efficiency is given by

$$\Xi_k^E = \sum_i \sum_{t \geq k} \frac{\beta^{t-k} u'(c_t^i)}{\lambda_k^i} \frac{dc_t^i}{d\theta} = \mathbb{C}ov_i^\Sigma \left(\frac{\beta^{t-k} u'(c_t^i)}{\lambda_k^i}, \frac{dc_t^i}{d\theta} \right),$$

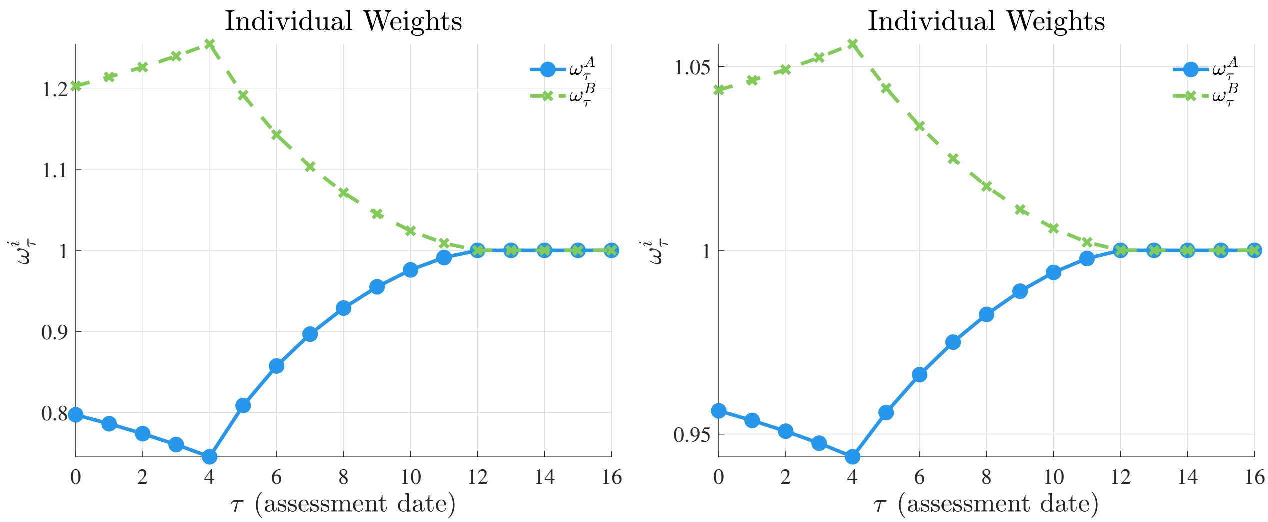

Figure 4. Normalized Individual Welfare Weights by Assessment Perspective

where the second equality follows because $\sum_i \frac{dc_t^i}{d\theta} = 0$ in this endowment economy, and redistribution is

$$\Xi_k^{RD} = \mathbb{C}ov_i^{\Sigma}\left(\omega^i, \sum_{t \geq k} \frac{\beta^{t-k} u'(c_t^i)}{\lambda_k^i} \frac{dc_t^i}{d\theta}\right).$$

We denote by $\omega_k^i = \frac{\lambda_k^i}{\frac{1}{I}\sum_i \lambda_k^i}$ the normalized individual welfare weight on individual i in units of numeraire from the perspective of date k. And we choose perpetual consumption as welfare numeraire, which implies $\lambda_k^i = \sum_{t \geq k} \beta^{t-k} u'(c_t^i)$.

Individual welfare weights. We plot the evolution of the normalized individual welfare weights ω_k^i in Figure 4, with the date of the assessment k on the x-axis. Panel (a) plots the welfare weights in the absence of a relief policy ($\theta = 0$) and Panel (b) plots them under a relief policy of $\theta = 0.75$.

The ratio ω_k^b / ω_k^a represents the planner's relative valuation of the two individuals in units of numeraire. This ratio peaks at about $\frac{1.25}{0.75} = 1.67$ at date $\underline{T} = 4$, before converging back to 1 by date $\overline{T} = 12$. At date $\underline{T}$, the planner values giving a unit of numeraire to individual b by 1.67 more than to individual a. In other words, at the onset of the low-endowment spell at date $\underline{T}$, the planner would be indifferent between giving individual b one unit of numeraire and giving individual a 1.67 units of numeraire.

The normalized individual weights govern the assessment of redistribution gains and losses from the relief policy. At the onset of the low-endowment spell, the scope for redistribution gains from the relief policy are perceived to be greatest, with the largest dispersion in ω_k^i.

Dynamic welfare weights. Figure 5 plots $\omega_{t|k}^i = \frac{1}{\lambda_k^i} \beta^{t-k} u'(c_t^i)$ for each individual i for two different assessment perspectives. Panel (a) considers an ex ante perspective at date $k = 0$ whereas Panel (b) considers an ex post perspective at date $k = 10$. Both panels plot this normalized dynamic

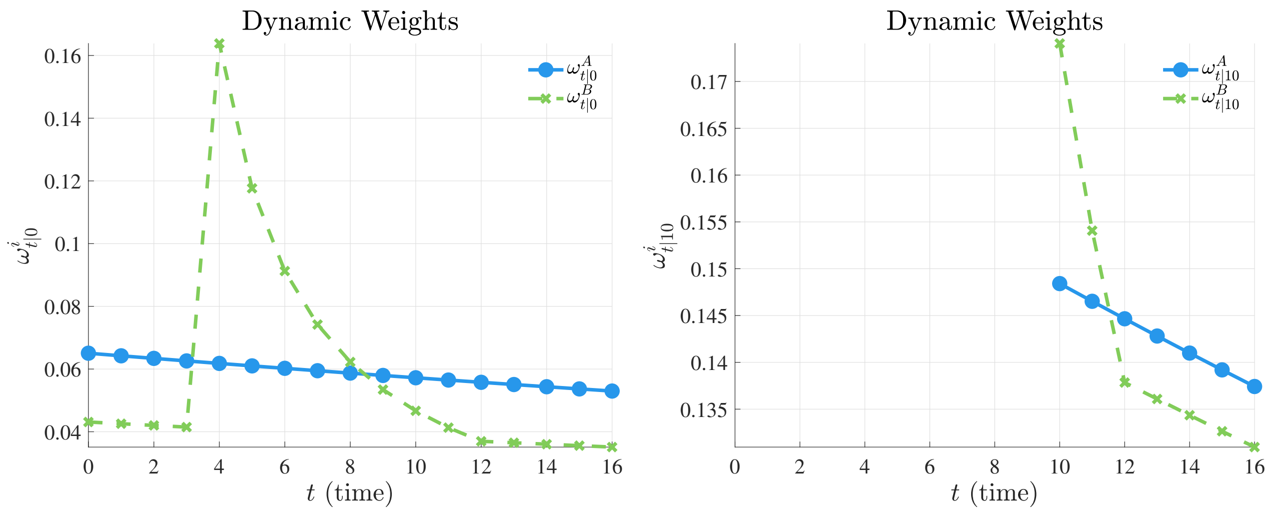

Figure 5. Normalized Dynamic Welfare Weights by Assessment Perspective

welfare weight $\omega^i_{t|k}$ against calendar time t on the x-axis.

The dynamic welfare weight for individual a from the perspective of date 0, $\omega^a_{t|0}$, captures how much the planner values one unit of consumption for individual a at date t relative to one unit of numeraire (one unit of consumption at all dates). By construction, therefore, these dynamic weights sum to 1 because giving the individual one unit of consumption at a particular date for all dates is equivalent to providing one unit of perpetual consumption. In other words, $\omega^a_{t|0}$ represents a marginal rate of substitution (MRS) between date t consumption and perpetual consumption for individual a from the perspective of date 0. If $\omega^b_{t|0} = 0.06$, as is the case for $t = 6$, the planner is indifferent between giving individual b one unit of consumption at date t and 0.06 units of perpetual consumption. Since the social welfare function we consider here is non-paternalistic, the planner and the individual agree on these marginal rates of substitution.

Cross-sectional dispersion in normalized dynamic weights $\omega^i_{t|k}$ indicate scope for efficiency gains from consumption-smoothing. Intuitively, consumption-smoothing efficiency gains can be achieved by marginally transferring consumption from the individual with a smaller dynamic weight $\omega^i_{t|k}$ at date t to the individual with the larger one.

Panel (a) of Figure 5 therefore illustrates the scope for consumption-smoothing gains from the relief policy from the ex ante perspective of date 0. The low-endowment spell starts at date $\underline{T} = 4$. At this point, $\frac{\omega^b_{4|0}}{\omega^a_{4|0}} \approx \frac{0.16}{0.06} = 2.67$, so the planner values a marginal increase of individual b's consumption 2.67 more. And since we have $\omega^b_{t|0} > \omega^a_{t|0}$ for $t \in [4, 8]$, the planner would find a relief policy desirable that starts in period 4 and expires in period 8, transferring from individual a to individual b. After period 8, the normalized dynamic weight of individual b falls below that of individual a. Intuitively, this signifies that individual b's consumption has recovered to lifetime consumption at this point.

Panel (b) repeats this exercise but from the perspective of date $k = 10$. From this ex post per-

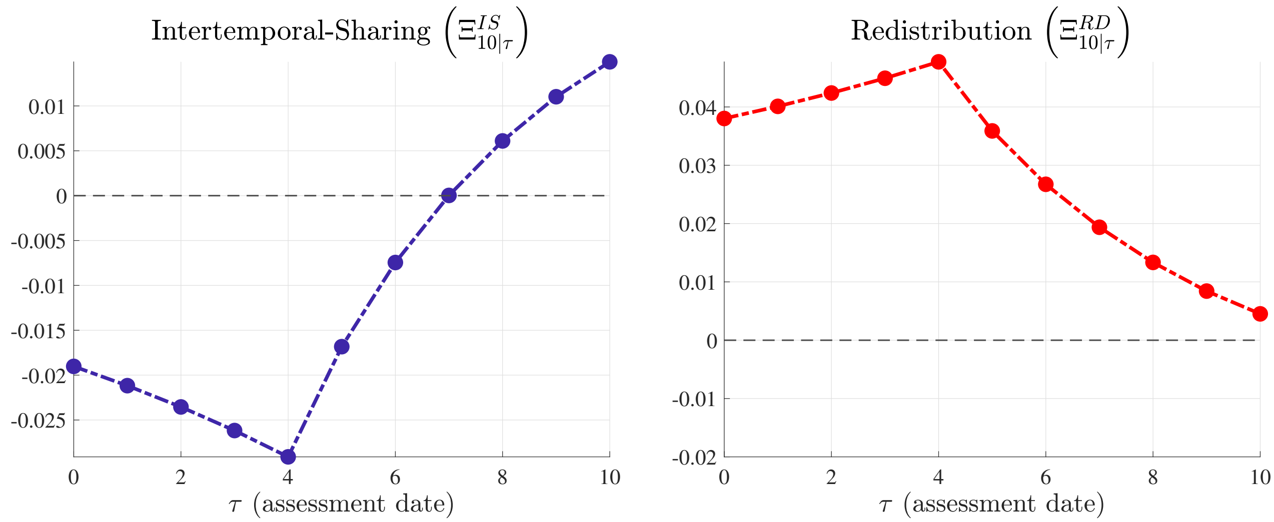

Figure 6. Welfare Assessment of Relief Policy by Assessment Perspective

spective, the worst of the low-endowment spell for individual b now lies in the past. Consequently, individual b's perceived lifetime consumption appears larger now. Her current consumption at date 10 appears to still be below lifetime consumption. As a result, individual b's normalized dynamic weight $\omega^b_{t|10}$ is still larger than individual a's in periods 10 and 11, unlike from the ex ante perspective. Ex post, the planner therefore finds it desirable to extend the relief policy and continue transferring from individual a to individual b until period 12.

Time inconsistency of welfare assessment. We now illustrate the time consistency problem that emerges in this setting. Consider extending the relief policy to period 10. That is, consider a marginal increase $d\theta_{10} > 0$ relative to the benchmark of no transfer $\theta_{10} = 0$. Figure 6 plots the welfare assessment of this extension of the relief policy from the perspective of different dates k. Panels (a) and (b) plot the efficiency and redistribution assessments, respectively, against the time of the assessment k on the x-axis.

Panel (a) demonstrates that the extending the relief policy to date 10 is perceived to result in an efficiency loss from the perspective of date 4, the onset of the low-endowment spell for individual b. However, as time goes by, the perceived efficiency loss shrinks. And by date 8, the planner finds it valuable to extend the relief policy on efficiency grounds. This illustrates the time consistency problem that may lead the planner to extend the relief policy beyond the originally intended expiration date.

7 Social Insurance

This application highlights a distinct form of time inconsistency that arises in the context of social insurance policies. When markets are incomplete, individuals cannot fully insure against idiosyncratic or differentially distributed aggregate risk. As a result, they may not be able to smooth consumption across histories at a given date. A policy that provides such consumption smoothing can therefore generate efficiency gains from an ex ante perspective. Once uncertainty is resolved and individual outcomes are realized, however, a policy that promised social insurance requires redistributive transfers from the lucky to the unlucky ex post: Today's efficiency becomes tomorrow's redistribution.

Social insurance is a central motivation behind a wide range of policies, including progressive income taxation, unemployment insurance, health care provision, disaster relief, and certain forms of macroeconomic stabilization. Such policies initially appear to improve risk-sharing and allocative efficiency but may ex post be perceived as pure redistribution. A social insurance policy that is considered optimal on efficiency grounds ex ante may therefore be time inconsistent.

7.1 Environment

In this section, we study a stochastic finite-horizon economy with terminal date T. As before, there are two individuals indexed $i \in \{a, b\}$, and there is a single good that appears as an endowment. Unlike before, individuals face idiosyncratic risk. At the beginning of each period, a state $s \in \{\ell, h\}$ is realized with Markov transition matrix

$$\Pi = \begin{pmatrix} \rho & 1-\rho \\ 1-\rho & \rho \end{pmatrix}.$$

Preferences. Individual i's preferences can be represented recursively by the non-stationary Bellman equation

$$V_t^i(s) = u(c^i(s)) + \beta \sum_{s'} \pi(s' \mid s) V_{t+1}^i(s'),$$

with terminal condition $V_T^i(s) = u(c^i(s))$. The value function $V_t^i(s)$ has a time subscript due to the non-stationary nature of this finite-horizon problem. Consumption, on the other hand, only depends on the state s and not on calendar time t as we describe next.

Endowments. We again assume financial autarky for illustration. Individual i's consumption in state s is given by

$$c^i(s) = y^i(s) \qquad \text{where} \qquad \begin{cases} y^a(s) = y + \epsilon(s)\,(1-\theta) \\ y^b(s) = y - \epsilon(s)\,(1-\theta) \end{cases}$$

There is no aggregate risk since the aggregate endowment $\sum y_i(s) = 2y$ is constant across states. All risk is idiosyncratic.

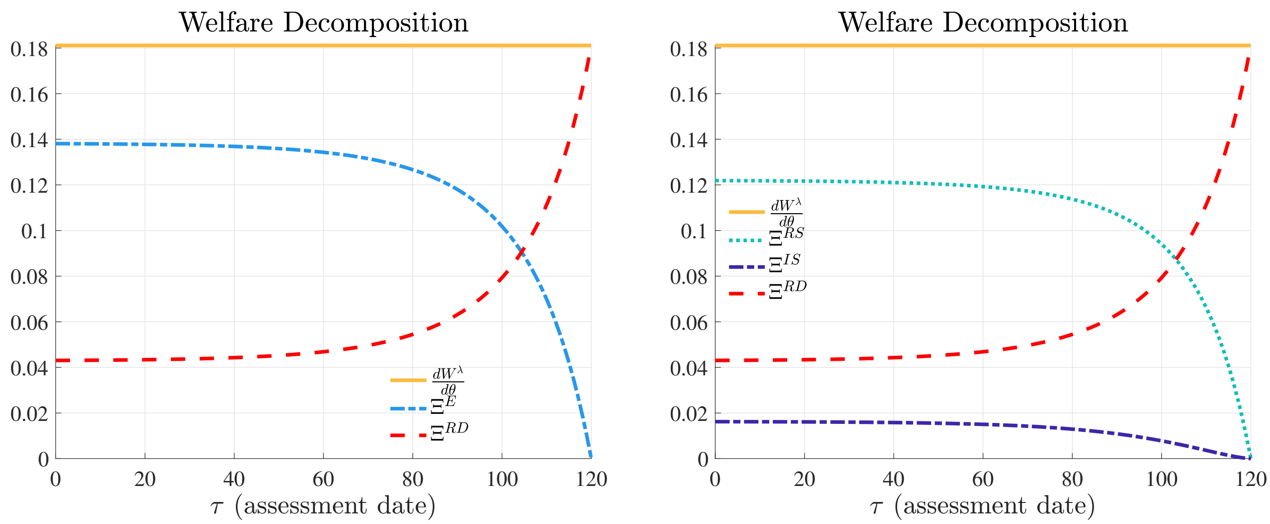

Figure 7. Welfare Assessment of Social Insurance Policy by Assessment Perspective

Social insurance policy. The social insurance policy we consider is indexed by θ. We study a marginal policy perturbation $d\theta$ around the full autarky benchmark with $\theta = 0$.

Parametrization. We assume CRRA utility $u(c) = \frac{1}{1-\gamma}c^{1-\gamma}$ with coefficient of relative risk aversion $\gamma = 2$. We set the discount factor to $\beta = 0.95$ and normalize the average endowment realization to $y = 1$. The size of the idiosyncratic earnings shock is $\epsilon(h) = -\epsilon(\ell) = 0.25$ and we set its persistence to $\rho = 0.975$. Finally, we set the terminal date to $T = 120$.

7.2 Welfare Assessment

We again study welfare assessments of the social insurance policy $d\theta$ under an equal-weighted social welfare function from the perspective of different points in time k, given by

$$W_k = V_k^a + V_k^b.$$

Time inconsistency of welfare assessment. We illustrate the main result of this application in Figure 7. Panel (a) plots the normalized welfare gain $\frac{dW_k^\lambda}{d\theta}$ (yellow) as well as its decomposition into efficiency (blue) and redistribution (red) gains. These values are plotted against the assessment date on the x-axis. We assume that state ℓ realizes each period.

From the perspective of date 0, a marginal increase $d\theta > 0$ in social insurance generates a large efficiency gain (0.14) as well as a modest redistribution gain (0.04). Since we initialize the economy in state ℓ, individuals are already unequal at date 0, which implies some redistribution from the social transfer. But the welfare assessment is dominated by the large efficiency gain. From the ex ante perspective, there is much scope for social insurance to provide consumption smoothing to the two individuals across the two states in future periods. As time passes and uncertainty is gradually realized, however, the scope for risk-sharing and efficiency gains becomes smaller. It converges to 0 in the terminal period $T = 120$ where, after the realization of the state, there is no scope for insurance anymore. When re-evaluated ex post at later dates, the gains from social insurance are increasingly attributed to redistribution. The total welfare gain from the perturbation remains constant (0.18) across assessments dates because we use a utilitarian social welfare function.

Panel (b) further decomposes efficiency gains into risk-sharing and intertemporal-sharing gains. Most of the efficiency gain is due to the policy's improvement of risk-sharing, with only a small contribution from intertemporal-sharing. Nonetheless, Panel (b) highlights that both components converge to 0 over time. The welfare gains from social insurance are initially attributed to the two sources of efficiency gains, but are increasingly perceived as pure redistribution as time passes and uncertainty is realized. Today's efficiency becomes tomorrow's redistribution.

8 Conclusion

Heterogeneous agent models are quickly supplanting the representative agent benchmark as the new workhorse of macroeconomic policy analysis. As this class of models matures, it becomes essential to clarify their normative implications. Our paper takes a step in this direction by revisiting the question of time consistency in heterogeneous agent incomplete markets environments. Specifically, three broad conclusions emerge to guide future applied work on optimal policy design with heterogeneous agents.

First, interpersonal welfare comparisons emerge as a new source of time inconsistency when markets are incomplete or individual preferences are not symmetric. In the presence of discount factor or belief heterogeneity, no time-invariant social welfare function allows for time consistent assessments. When discount factors and beliefs are homogeneous, there is a sharp distinction between linear and non-linear social welfare functions. As long as markets are incomplete, only the utilitarian criterion implies time consistent assessments. When markets are complete and individual preferences symmetric, then welfare assessments under both linear and non-linear SWFs become time consistent. In summary, heterogeneous agent models may thus feature three distinct sources of time inconsistency: dynamically inconsistent individual preferences (Strotz, 1956), forward-looking behavior (Kydland and Prescott, 1977), and interpersonal welfare comparisons.

Second, the attribution of welfare gains to efficiency and redistribution is invariably time inconsistent when markets are incomplete, irrespective of the social welfare function and the extent of heterogeneity in individual preferences. In other words, even when the overall welfare assessment is time consistent in one of the special cases discussed above, whether those gains and losses are perceived as efficiency or redistribution invariably changes over time.

Finally, assessments of production efficiency gains stand out as a special case. Unlike risk-sharing and redistribution, production efficiency assessments can be time consistent for arbitrary social welfare functions and preference heterogeneity. This happens whenever the perturbation under consideration generates a production efficiency gain or loss at a single date and history. We show that this applies to a broad class of applied policy questions and therefore presents a relevant exception to the broader prevalence of time inconsistency.

Our results suggest the following rules-of-thumb for applied work on optimal policy design with heterogeneous agents: When using a utilitarian SWF in an environment without discount factor or belief heterogeneity, optimal policy is time consistent. However, the justification of optimal policy on efficiency or redistribution grounds changes as time passes and uncertainty is resolved. When either studying environments with discount factor and belief heterogeneity or using non-utilitarian SWFs, optimal policy will generally be time inconsistent.

It may seem a natural takeaway from this discussion to restrict the study of optimal policy design to environments with symmetric individual preferences and utilitarian planning objectives, but that restriction is untenably limiting. Mounting empirical evidence documents sizable dispersion in individual rates of time preference and beliefs. And many policy questions naturally arise in the context of non-utilitarian objectives — for example, when independent institutions or agencies set policies according to a narrowly defined mandate. Whenever such a mandate is not utilitarian, the implied optimal policy may be time inconsistent as a direct consequence of the interpersonal welfare comparisons implied by the mandate. Confronting these time consistency problems in heterogeneous agent models seems a valuable avenue for future work.

References

Adams, Abi, Laurens Cherchye, Bram De Rock, and Ewout Verriest. 2014. Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review*, 104(12):4147–4183.

Adler, Matthew D and Marc Fleurbaey. 2016. *The Oxford handbook of well-being and public policy*. Oxford University Press.

Aiyagari, S Rao, Albert Marcet, Thomas J Sargent, and Juha Seppälä. 2002. Optimal taxation without state-contingent debt. *Journal of Political Economy*, 110(6):1220–1254.

Alesina, Alberto and Dani Rodrik. 1994. Distributive politics and economic growth. *The quarterly journal of economics*, 109(2):465–490.

Arrow, Kenneth J. 1950. A difficulty in the concept of social welfare. *Journal of political economy*, 58(4):328–346.

—. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5):941–973.

Athey, Susan, Andrew Atkeson, and Patrick J Kehoe. 2005. The optimal degree of discretion in monetary policy. *Econometrica*, 73(5):1431–1475.

Atkinson, Anthony B. 1970. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263.

Atkinson, Anthony Barnes and Joseph E Stiglitz. 1976. The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, 6(1-2):55–75.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Auclert, Adrien, Michael Cai, Matthew Rognlie, and Ludwig Straub. 2024. Optimal long-run fiscal policy with heterogeneous agents.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2025. Fiscal and monetary policy with heterogeneous agents. *Annual Review of Economics*, 17.

Barro, Robert J and David B Gordon. 1983. Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, 12(1):101–121.

Benabou, Roland. 2002. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517.

Bergson, Abram. 1938. A reformulation of certain aspects of welfare economics. *The Quarterly Journal of Economics*, 52(2):310–334.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas Sargent. 2024. Efficiency, insurance, and redistribution effects of government policies. Technical report, Working paper.

Bilbiie, Florin O. 2024. Monetary policy and heterogeneity: An analytical framework. *Review of Economic Studies*, p. rdae066.

Bisin, Alberto, Alessandro Lizzeri, and Leeat Yariv. 2015. Government policy with time inconsistent voters. *American Economic Review*, 105(6):1711–1737.

Boadway, Robin and Neil Bruce. 1984. A general proposition on the design of a neutral business tax. *Journal of Public Economics*, 24(2):231–239.

Boskin, Michael J and Eytan Sheshinski. 1978. Optimal redistributive taxation when individual welfare depends upon relative income. *The Quarterly Journal of Economics*, 92(4):589–601.

Chambers, Christopher P and Federico Echenique. 2018. On multiple discount rates. *Econometrica*, 86(4):1325–1346.

Chang, Roberto. 2022. Should Central Banks Have an Inequality Objective? Technical report, National Bureau of Economic Research.

Chari, Varadarajan V and Patrick J Kehoe. 1990. Sustainable plans. *Journal of political economy*, 98(4):783–802.

Clayton, Christopher and Andreas Schaab. 2025. A theory of dynamic inflation targets. *American Economic Review*, 115(2):448–490.

Dávila, Eduardo and Andreas Schaab. 2023. Optimal monetary policy with heterogeneous agents: discretion, commitment, and timeless policy.

—. 2024. Welfare Assessments with Heterogeneous Individuals.

Diamond, Peter A. 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review*, pp. 83–95.

Dovis, Alessandro and Rishabh Kirpalani. 2020. Fiscal rules, bailouts, and reputation in federal governments. *American Economic Review*, 110(3):860–888.

—. 2021. Rules without commitment: Reputation and incentives. *The Review of Economic Studies*, 88(6):2833–2856.

Farhi, Emmanuel, Christopher Sleet, Ivan Werning, and Sevin Yeltekin. 2012. Non-linear capital taxation without commitment. *Review of Economic Studies*, 79(4):1469–1493.

Farhi, Emmanuel and Iván Werning. 2013. Insurance and taxation over the life cycle. *Review of Economic Studies*, 80(2):596–635.

Feldstein, Martin S. 1964. The social time preference discount rate in cost benefit analysis. *The economic journal*, 74(294):360–379.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. Redistribution and social insurance. *American Economic Review*, 106(2):359–386.

Halac, Marina and Pierre Yared. 2014. Fiscal Rules and Discretion Under Persistent Shocks. *Econometrica*, 82(5):1557–1614.

Halevy, Yoram. 2015. Time consistency: Stationarity and time invariance. *Econometrica*, 83(1):335–352.

Harsanyi, John C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, 63(4):309–321.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2017. Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Hicks, John R. 1939. The foundations of welfare economics. *The economic journal*, 49(196):696–712.

Jackson, Matthew O and Leeat Yariv. 2014. Present bias and collective dynamic choice in the lab. *American Economic Review*, 104(12):4184–4204.

—. 2015. Collective dynamic choice: the necessity of time inconsistency. *American Economic Journal: Microeconomics*, 7(4):150–178.

Kaldor, Nicholas. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *The economic journal*, 49(195):549–552.

Kaneko, Mamoru and Kenjiro Nakamura. 1979. The Nash social welfare function. *Econometrica: Journal of the Econometric Society*, pp. 423–435.

Kaplow, Louis. 2011. *The theory of taxation and public economics*. Princeton University Press.

Klein, Paul, Per Krusell, and Jose-Victor Rios-Rull. 2008. Time-consistent public policy. *The Review of Economic Studies*, 75(3):789–808.

Knight, Frank H. 1921. *Risk, uncertainty and profit*. Houghton Mifflin.

Koopmans, Tjalling C. 1960. Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, pp. 287–309.

Krusell, Per, Vincenzo Quadrini, and José-V\'\ictor R\'\ios-Rull. 1997. Politico-economic equilibrium and economic growth. *Journal of Economic Dynamics and Control*, 21(1):243–272.

Kydland, Finn E and Edward C Prescott. 1977. Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, 85(3):473–491.

—. 1980. Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and control*, 2:79–91.

Ljungqvist, Lars and Thomas J Sargent. 2018. *Recursive macroeconomic theory*. MIT press.

Lucas, Robert E and Nancy L Stokey. 1983. Optimal fiscal and monetary policy in an economy without capital. *Journal of monetary Economics*, 12(1):55–93.

Marcet, Albert and Ramon Marimon. 2019. Recursive contracts. *Econometrica*, 87(5):1589–1631.

Marglin, Stephen A. 1963. The social rate of discount and the optimal rate of investment. *The Quarterly Journal of Economics*, 77(1):95–111.

Marshall, Alfred. 1920. *Principles of Economics, 8th edn (1920)*.

Millner, Antony and Geoffrey Heal. 2018. Time consistency and time invariance in collective intertemporal choice. *Journal of Economic Theory*, 176:158–169.

Mirrlees, James A. 1971. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Mongin, Philippe. 1995. Consistent bayesian aggregation. *Journal of Economic Theory*, 66(2):313–351.

—. 1998. The paradox of the Bayesian experts and state-dependent utility theory. *Journal of Mathematical Economics*, 29(3):331–361.

Nash, John F. 1950. The bargaining problem. *Econometrica*, 18(2):155–162.

Persson, Torsten and Guido Enrico Tabellini. 1994. Is inequality harmful for growth? *The American Economic Review*, 84(3):600–621.

Pigou, Arthur. 1920. *The economics of welfare*. Macmillan.

Piketty, Thomas and Emmanuel Saez. 2012. A theory of optimal capital taxation. Technical report, National Bureau of Economic Research.

—. 2013a. A theory of optimal inheritance taxation. *Econometrica*, 81(5):1851–1886.

—. 2013b. Optimal labor income taxation. In *Handbook of public economics*, volume 5, pp. 391–474. Elsevier.

Rawls, John. 1971. A theory of justice.

Rogoff, Kenneth. 1985. The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics*, 100(4):1169–1189.

Saez, Emmanuel. 2001. Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1):205–229.

Saez, Emmanuel and Stefanie Stantcheva. 2016. Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, 106(01):24–45.

Samuelson, Paul Anthony. 1947. *Foundations of economic analysis*.

Stern, Nicholas H. 1976. On the specification of models of optimum income taxation. *Journal of Public economics*, 6(1-2):123–162.

Strotz, R H. 1956. Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, 23(3):165–180.

Sublet, Guillaume. 2023. The Optimal Degree of Discretion in Fiscal Policy.

Varian, Hal R. 1980. Redistributive taxation as social insurance. *Journal of public Economics*, 14(1):49–68.

Werning, Ivan. 2007. Optimal fiscal policy with redistribution. *The Quarterly Journal of Economics*, 122(3):925–967.

Zuber, Stephane. 2011. The aggregation of preferences: can we ignore the past? *Theory and decision*, 70:367–384.

A Appendix for Section 2

In this Appendix, we provide additional derivations and details for Section 2. Appendix A.1 characterizes the implementability conditions for the ex ante and ex post planning problems. Appendices A.2 and A.3 present derivations of equations (9) and (10) in the main text. Finally, Appendix A.4 presents details of our numerical implementation.

A.1 Implementability Conditions

Recall the household budget constraints from the main text, $c_0^i = w_0 \ell_0^i$ and $c_1^i = (1-\tau) w_1 z_1^i \ell_1^i + T$. Maximizing preference simplies the FOCs

$$
\begin{aligned}
u_{\ell,0}^i &= -u_{c,0}^i w_0 \\
u_{\ell,1}^i &= -u_{c,1}^i (1-\tau) w_1 z_1^i.
\end{aligned}
$$

The firm's profit maximization optimality condition implies

$$
w_t = 1.
$$

We use this firm optimality condition to solve out for the real wage. The remaining equilibrium conditions can therefore be written as

$$
\begin{aligned}
u_{\ell,0}^i &= -u_{c,0}^i \\
u_{\ell,1}^i &= -u_{c,1}^i (1-\tau) z_1^i \\
c_0^i &= \ell_0^i \\
c_1^i &= (1-\tau) z_1^i \ell_1^i + T \\
Y_t &= L_t \\
T &= \tau \int_0^1 z_1^i \ell_1^i \, di \\
Y_t &= \int_0^1 c_t^i \, di \\
L_t &= \int_0^1 z_t^i \ell_t^i \, di.
\end{aligned}
$$

Notice that fiscal policy has no impact on the date 0 allocation.

A.2 Ex Ante Optimal Policy

From the perspective of date 0, the Ramsey problem is to choose a tax rate τ to maximize social welfare subject to the conditions of competitive equilibrium listed above. At an optimum, the necessary optimality condition is

$$\begin{aligned}
0 &= \frac{dW_0}{d\tau} \\
0 &= \frac{1}{\int_0^1 \lambda_0^i \alpha_0^i \, di} \frac{dW_0}{d\tau} \\
0 &= \int_0^1 \frac{1}{\lambda_0^i} \frac{dV_0^i}{d\tau} \, di + \mathbb{C}ov_i \left(\omega_0^i, \frac{dV_0^i}{d\tau} \right),
\end{aligned}$$

where we define the normalized individual welfare weight

$$\omega_0^i = \frac{\alpha_0^i \lambda_0^i}{\int_0^1 \alpha_0^i \lambda_0^i \, di}.$$

The second line represents a normalization that expresses the first-order condition in units of welfare numeraire. And the third line decomposes the welfare gain $\frac{dW_0}{d\tau}$ into an efficiency and a redistribution gain.

Next, notice that for any choice of social welfare function $\mathcal{W}(\cdot)$ and for any choice of welfare ex ante numeraire λ_0^i, the RHS of the above covariance is constant across i because individuals are identical from the perspective of date 0. Therefore, the impact of anticipated tax policy is symmetric across individuals. The covariance term thus vanishes.

Next, we plug in for lifetime utility and rewrite the Ramsey FOC as

$$\begin{aligned}
0 &= \int_0^1 \frac{1}{\lambda_0^i} \frac{dV_0^i}{d\tau} \, di \\
0 &= \int_0^1 \frac{1}{\lambda_0^i} \beta \, \mathbb{E}_0 \left[u_{c,1}^i \frac{dc_1^i}{d\tau} + u_{\ell,1}^i \frac{d\ell_1^i}{d\tau} \right] di \\
0 &= \int_0^1 \frac{1}{\lambda_0^i} \beta \, \mathbb{E}_0 \, u_{c,1}^i \left[\frac{dc_1^i}{d\tau} - (1-\tau) z_1^i \frac{d\ell_1^i}{d\tau} \right] di
\end{aligned}$$

where the second line uses the fact that tax policy has no impact on the date 0 allocation, and the third line uses household i's date 1 FOC. Now we decompose the valuation as follows

$$\begin{aligned}
0 &= \int_0^1 \mathbb{E}_0 \frac{\beta u^i_{c,1}}{\lambda^i_0} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di \\
&= \int_0^1 \mathbb{E}_0 \frac{\beta u^i_{c,1}}{\lambda^i_0} \frac{(\mathbb{E}_0 \beta u^i_{c,1})}{(\mathbb{E}_0 \beta u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di \\
&= \int_0^1 \mathbb{E}_0 \left\{ \frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0} \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] \right\} di,
\end{aligned}$$

where sometimes for convenience we use the shorthand notation

$$\omega^i_1 = \frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0} \qquad \text{and} \qquad \tilde{\omega}^i_1 = \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})}$$

where $\tilde{\omega}^i_1$ is a random variable from the perspective of date 0, whereas ω^i_1 is not.

Next, we swap the expectation operator and perform a cross-sectional covariance decomposition

$$\begin{aligned}
0 &= \mathbb{E}_0 \int_0^1 \frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0} \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di \\
&= \left(\int_0^1 \frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0} di \right) \left(\int_0^1 \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di \right) \\
&\quad + \mathbb{C}ov_i \left(\frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0}, \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] \right)
\end{aligned}$$

Notice that $\frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0}$ is constant across all i. That's because households are symmetric from the perspective of period 0: they behave exactly the same in period 0, and they are still the same in expectation in period 1. Therefore, the covariance term here vanishes. And we are left with

$$0 = \bar{\omega}_1 \int_0^1 \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di$$

where we set

$$\bar{\omega}_1 = \frac{\beta(\mathbb{E}_0 u^i_{c,1})}{\lambda^i_0},$$

which is the same for all i.

One final cross-sectional covariance decomposition yields

$$0 = \bar{\omega}_1 \left(\int_0^1 \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \, di \right) \int_0^1 \left[\frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right] di + \bar{\omega}_1 \mathbb{C}ov_i \left(\frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})}, \frac{dc^i_1}{d\tau} - (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} \right).$$

Notice that

$$\int_0^1 \frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})} \, di = 1$$

due to a law of large numbers.

Now notice that the individual budget constraint implies

$$\frac{dc^i_1}{d\tau} = (1-\tau) z^i_1 \frac{d\ell^i_1}{d\tau} - z^i_1 \ell^i_1 + \frac{dT}{d\tau}$$

where $\frac{dT}{d\tau} = L_1 + \tau \frac{dL_1}{d\tau}$. Therefore, we have

$$\begin{aligned} 0 &= \bar{\omega}_1 \int_0^1 \left[-z^i_1 \ell^i_1 + \frac{dT}{d\tau} \right] di + \bar{\omega}_1 \mathbb{C}ov_i \left(\frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})}, -z^i_1 \ell^i_1 \right) \\ &= \bar{\omega}_1 \tau \frac{dL_1}{d\tau} + \bar{\omega}_1 \mathbb{C}ov_i \left(\frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})}, -z^i_1 \ell^i_1 \right) \end{aligned}$$

where the second term in the first equality follows because dT is constant across individuals and therefore drops out from the covariance. We can now write this as

$$0 = \tau \frac{dL_1}{d\tau} - \mathbb{C}ov_i \left(\frac{u^i_{c,1}}{(\mathbb{E}_0 u^i_{c,1})}, z^i_1 \ell^i_1 \right),$$

which corresponds to equation (9) in the main text.

A.3 Ex Post Optimal Policy

We now characterize the optimality condition for optimal tax policy from the ex post perspective of period 1. The relevant equilibrium conditions are still given by

$$\begin{aligned}
u^i_{\ell,1} &= -u^i_{c,1}(1-\tau)z^i_1 \\
c^i_1 &= (1-\tau)z^i_1\ell^i_1 + T \\
Y_1 &= L_1 \\
T &= \tau\int_0^1 z^i_1\ell^i_1\, di \\
Y_1 &= \int_0^1 c^i_1\, di \\
L_1 &= \int_0^1 z^i_1\ell^i_1\, di.
\end{aligned}$$

But now the welfare decomposition works differently. We have

$$\begin{aligned}
0 &= \frac{dW_1}{d\tau} \\
0 &= \frac{1}{\int_0^1 \alpha^i_1\lambda^i_1\, di}\frac{dW_1}{d\tau} \\
0 &= \int_0^1 \frac{1}{\lambda^i_1}\frac{dV^i_1}{d\tau}\, di + \mathbb{C}ov_i\left(\omega^i_1, \frac{1}{\lambda^i_1}\frac{dV^i_1}{d\tau}\right),
\end{aligned}$$

where we define

$$\omega^i_1 = \frac{\alpha^i_1\lambda^i_1}{\int_0^1 \alpha^i_1\lambda^i_1\, di}$$

from the perspective of date 1. Also notice that the only sensible numeraire in this case is period 1 consumption, so we have

$$\lambda^i_1 = u^i_{c,1}.$$

Let's now unpack the efficiency term. We have

$$\begin{aligned}
\int_0^1 \frac{1}{\lambda^i_1}\frac{dV^i_1}{d\tau}\, di &= \int_0^1 \frac{1}{\lambda^i_1}\left[u^i_{c,1}\frac{dc^i_1}{d\tau} + u^i_{\ell,1}\frac{d\ell^i_1}{d\tau}\right] di \\
&= \int_0^1 \frac{1}{\lambda^i_1}u^i_{c,1}\left[\frac{dc^i_1}{d\tau} - (1-\tau)z^i_1\frac{d\ell^i_1}{d\tau}\right] di \\
&= \int_0^1 \left[\frac{dc^i_1}{d\tau} - (1-\tau)z^i_1\frac{d\ell^i_1}{d\tau}\right] di
\end{aligned}$$

where the last line follows due to $\lambda_1^i = u_{c,1}^i$. We now use the individual budget constraint, then the government budget constraint, and aggregate, yielding

$$\begin{aligned}
\int_0^1 \left[\frac{dc_1^i}{d\tau} - (1-\tau) z_1^i \frac{d\ell_1^i}{d\tau}\right] di &= \int_0^1 \left[-z_1^i \ell_1^i + \frac{dT}{d\tau}\right] di \\
&= \int_0^1 \left[-z_1^i \ell_1^i + L_1 + \tau \frac{dL_1}{d\tau}\right] di \\
&= \tau \frac{dL_1}{d\tau}
\end{aligned}$$

Therefore, we have the ex post optimality condition

$$\begin{aligned}
0 &= \tau \frac{dL_1}{d\tau} + \mathbb{C}ov_i\left(\frac{\alpha_1^i u_{c,1}^i}{\int_0^1 \alpha_1^i u_{c,1}^i \, di}, \frac{1}{\lambda_1^i} \frac{dV_1^i}{d\tau}\right) \\
&= \tau \frac{dL_1}{d\tau} + \mathbb{C}ov_i\left(\frac{\alpha_1^i u_{c,1}^i}{\int_0^1 \alpha_1^i u_{c,1}^i \, di}, \frac{dc_1^i}{d\tau} - (1-\tau) z_1^i \frac{d\ell_1^i}{d\tau}\right) \\
&= \tau \frac{dL_1}{d\tau} + \mathbb{C}ov_i\left(\frac{\alpha_1^i u_{c,1}^i}{\int_0^1 \alpha_1^i u_{c,1}^i \, di}, -z_1^i \ell_1^i\right).
\end{aligned}$$

And finally, we have

$$\alpha_1^i = \frac{\partial \mathcal{W}(\{V_1^i\}_i)}{\partial V_1^i} = W_1^{\phi} \nu^i (\nu^i V_1^i)^{-\phi}.$$

This leaves us with

$$0 = \tau \frac{dL_1}{d\tau} + \mathbb{C}ov_i\left(\frac{(\nu^i)^{1-\phi} (V_1^i)^{-\phi} u_{c,1}^i}{\int_0^1 (\nu^i)^{1-\phi} (V_1^i)^{-\phi} u_{c,1}^i \, di}, -z_1^i \ell_1^i\right)$$

which corresponds to equation (10) in the main text.

A.4 Numerics

We start with the competitive equilibrium conditions

$$
\begin{aligned}
u^i_{\ell,1} &= -u^i_{c,1}(1-\tau)z^i_1 \\
c^i_1 &= (1-\tau)z^i_1\ell^i_1 + T \\
Y_1 &= L_1 \\
T &= \tau \int_0^1 z^i_1\ell^i_1\, di \\
Y_1 &= \int_0^1 c^i_1\, di \\
L_1 &= \int_0^1 z^i_1\ell^i_1\, di.
\end{aligned}
$$

Notice that the only thing that changes between the ex ante and ex post assessments are the valuations, not the allocation. So we only have to work this out once, and then compute valuations from the two perspectives.

We start by solving out for Y_1 and T, yielding

$$
\begin{aligned}
u^i_{\ell,1} &= -u^i_{c,1}(1-\tau)z^i_1 \\
c^i_1 &= (1-\tau)z^i_1\ell^i_1 + \tau L_1 \\
L_1 &= \int_0^1 c^i_1\, di \\
L_1 &= \int_0^1 z^i_1\ell^i_1\, di.
\end{aligned}
$$

Now we drop one of the market clearing conditions by Walras' law. Say the goods market clearing condition, leaving us with

$$
\begin{aligned}
u^i_{\ell,1} &= -u^i_{c,1}(1-\tau)z^i_1 \\
c^i_1 &= (1-\tau)z^i_1\ell^i_1 + \tau L_1 \\
L_1 &= \int_0^1 z^i_1\ell^i_1\, di.
\end{aligned}
$$

So taking as given τ, we have 3 equations in the 3 unknowns L_1 and $\{c^i_1, \ell^i_1\}$.

Algorithm to solve in levels. Fix τ. Guess L_1. Then use a non-linear solver to solve the first two conditions for c^i_1 and ℓ^i_1, then use the last equation as the gap in a Newton. Also, we now rewrite things in terms of the realized z, which yields

$$u_{\ell,1}(z) = -u_{c,1}(z)(1-\tau)z$$
$$c_1(z) = (1-\tau)z\ell_1(z) + \tau L_1$$
$$L_1 = \int z\ell_1(z)g(z)\,dz$$

Now we use isoelastic preferences with

$$u(c,\ell) = \frac{1}{1-\gamma}c^{1-\gamma} - \frac{1}{1+\eta}\ell^{1+\eta}.$$

This lets us rewrite the first condition as

$$-\ell_1(z)^{\eta} = -c_1(z)^{-\gamma}(1-\tau)z$$

So if $\gamma = \eta = 2$, we have

$$\ell_1(z) = \frac{1}{c_1(z)}\sqrt{(1-\tau)z}$$

And using the budget constraint,

$$(1-\tau)z\ell_1(z)^2 + \tau L_1\ell_1(z) = \sqrt{(1-\tau)z}$$

Algorithm to solve derivatives. Next, we fully differentiate to solve for $\frac{dL_1}{d\tau}$. We have

$$u_{\ell\ell,1}(z)\frac{d\ell_1(z)}{d\tau} = u_{c,1}(z)z - (1-\tau)zu_{cc,1}(z)\frac{dc_1(z)}{d\tau}$$
$$\frac{dc_1(z)}{d\tau} = -z\ell_1(z) + (1-\tau)z\frac{d\ell_1(z)}{d\tau} + \tau\frac{dL_1}{d\tau} + L_1$$
$$\frac{dL_1}{d\tau} = \int z\frac{d\ell_1(z)}{d\tau}g(z)\,dz$$

We can rewrite the first equation as

$$\frac{u_{\ell\ell,1}(z)}{u_{\ell,1}(z)}\frac{u_{\ell,1}(z)}{u_{c,1}(z)}\frac{d\ell_1(z)}{d\tau} = z - (1-\tau)z\frac{u_{cc,1}(z)}{u_{c,1}(z)}\frac{dc_1(z)}{d\tau}$$
$$-\eta\frac{1}{\ell_1(z)}(1-\tau)z\frac{d\ell_1(z)}{d\tau} = z + (1-\tau)z\gamma\frac{1}{c_1(z)}\frac{dc_1(z)}{d\tau}$$
$$-\eta\frac{1}{\ell_1(z)}\frac{d\ell_1(z)}{d\tau} = \frac{1}{1-\tau} + \gamma\frac{1}{c_1(z)}\frac{dc_1(z)}{d\tau}$$

So since this is linear, we can actually solve for it explicitly. We get

$$\frac{d\ell_1(z)}{d\tau} = -\frac{1}{\eta}\ell_1(z)\left[\frac{1}{1-\tau} + \gamma\frac{1}{c_1(z)}\frac{dc_1(z)}{d\tau}\right]$$

and therefore

$$\frac{dc_1(z)}{d\tau} = -z\ell_1(z) - \frac{1}{\eta}\ell_1(z)\left[\frac{1}{1-\tau} + \gamma\frac{1}{c_1(z)}\frac{dc_1(z)}{d\tau}\right](1-\tau)z + \tau\frac{dL_1}{d\tau} + L_1$$

or simply

$$\frac{dc_1(z)}{d\tau} = -z\ell_1(z) - \frac{1}{\eta}\ell_1(z)z - \frac{\gamma}{\eta}\ell_1(z)\frac{1}{c_1(z)}\frac{dc_1(z)}{d\tau}(1-\tau)z + \tau\frac{dL_1}{d\tau} + L_1$$
$$\frac{dc_1(z)}{d\tau} = -\frac{1+\eta}{\eta}z\ell_1(z) - \frac{\gamma}{\eta}(1-\tau)z\frac{\ell_1(z)}{c_1(z)}\frac{dc_1(z)}{d\tau} + \tau\frac{dL_1}{d\tau} + L_1$$

and therefore

$$\frac{dc_1(z)}{d\tau} = \frac{1}{1+\frac{\gamma}{\eta}(1-\tau)z\frac{\ell_1(z)}{c_1(z)}}\left[-\frac{1+\eta}{\eta}z\ell_1(z) + \tau\frac{dL_1}{d\tau} + L_1\right]$$

This gives us

$$\frac{d\ell_1(z)}{d\tau} = -\frac{1}{\eta}\ell_1(z)\frac{1}{1-\tau} - \frac{\gamma}{\eta}\frac{\ell_1(z)}{c_1(z)}\frac{1}{1+\frac{\gamma}{\eta}(1-\tau)z\frac{\ell_1(z)}{c_1(z)}}\left[-\frac{1+\eta}{\eta}z\ell_1(z) + \tau\frac{dL_1}{d\tau} + L_1\right]$$

or simply

$$\frac{d\ell_1(z)}{d\tau} = -\frac{1}{\eta}\ell_1(z)\frac{1}{1-\tau} - \frac{1}{(1-\tau)z+\frac{\eta}{\gamma}\frac{c_1(z)}{\ell_1(z)}}\left[-\frac{1+\eta}{\eta}z\ell_1(z) + \tau\frac{dL_1}{d\tau} + L_1\right]$$

And so finally we arrive at

$$\begin{aligned}
\frac{dL_1}{d\tau} &= \int z\frac{d\ell_1(z)}{d\tau}g(z)\,dz \\
&= \int z\Big\{-\frac{1}{\eta}\ell_1(z)\frac{1}{1-\tau} - \frac{1}{(1-\tau)z+\frac{\eta}{\gamma}\frac{c_1(z)}{\ell_1(z)}}\left[-\frac{1+\eta}{\eta}z\ell_1(z) + \tau\frac{dL_1}{d\tau} + L_1\right]\Big\}g(z)\,dz \\
&= \int z\Big\{-\frac{1}{\eta}\ell_1(z)\frac{1}{1-\tau}\Big\}g(z)\,dz - \int z\Big\{\frac{1}{(1-\tau)z+\frac{\eta}{\gamma}\frac{c_1(z)}{\ell_1(z)}}\left[-\frac{1+\eta}{\eta}z\ell_1(z) + \tau\frac{dL_1}{d\tau} + L_1\right]\Big\}g(z)\,dz
\end{aligned}$$

where the first term becomes

$$\int z\Big\{-\frac{1}{\eta}\ell_1(z)\frac{1}{1-\tau}\Big\}g(z)\,dz = -\frac{1}{\eta}\frac{1}{1-\tau}L_1$$

And the second term becomes

$$+\frac{1+\eta}{\eta}\int\left\{\frac{1}{(1-\tau)+\frac{\eta}{\gamma}\frac{c_1(z)}{z\ell_1(z)}}z\ell_1(z)\right\}g(z)\,dz-\left[\tau\frac{dL_1}{d\tau}+L_1\right]\int\frac{1}{(1-\tau)+\frac{\eta}{\gamma}\frac{c_1(z)}{z\ell_1(z)}}g(z)\,dz$$

Now we denote

$$x(z)=\frac{1}{(1-\tau)+\frac{\eta}{\gamma}\frac{c_1(z)}{z\ell_1(z)}}$$

and can therefore write

$$\frac{dL_1}{d\tau}=-\frac{1}{\eta}\frac{1}{1-\tau}L_1+\frac{1+\eta}{\eta}\int x(z)z\ell_1(z)g(z)\,dz-\tau\frac{dL_1}{d\tau}\int x(z)g(z)\,dz-L_1\int x(z)g(z)\,dz$$

And so therefore

$$\frac{dL_1}{d\tau}=\frac{1}{1+\tau X}\left(-\frac{1}{\eta}\frac{1}{1-\tau}L_1+\frac{1+\eta}{\eta}\int x(z)z\ell_1(z)g(z)\,dz-L_1\int x(z)g(z)\,dz\right)$$

where $X=\int x(z)g(z)\,dz$.

B Proofs

B.1 Proof of Proposition 1

Proof. The result is a special case of Proposition 2. See Appendix B.2 below for the proof. With a representative agent and $I = 1$, we have $\alpha^i_{s^t} = 1$ without loss for all s^t. Therefore, we can decompose the ex ante assessment as

$$
\begin{aligned}
\frac{dW_{s^0}}{d\theta} &= \beta^k \pi(s^k) \frac{dW_{s^k}}{d\theta} \\
&+ \beta^k \pi(s^k) \left(\frac{dV_{s^0}(s^k)}{d\theta} - \frac{dV_{s^k}}{d\theta} \right) \\
&+ \sum_{t \geq k} \sum_{s^t \not\geq s^k} \beta^t \pi(s^t) u'(c_t(s^t)) \frac{\partial \mathcal{C}_t(s^t \mid s^0)}{\partial \theta} \\
&+ \sum_{0 \leq t < k} \sum_{s^t} \beta^t \pi(s^t) u'(c_t(s^t)) \frac{\partial \mathcal{C}_t(s^t \mid s^0)}{\partial \theta}
\end{aligned}
$$

Therefore, the only way in which $\frac{dW_{s^0}}{d\theta} \neq \beta^t \pi(s^t) \frac{dW_{s^t}}{d\theta}$ is for the second, third or fourth lines to be non-zero. This requires that one of the conditions (i) – (iii) in Proppposition 1 is satisfied.

■

B.2 Proof of Proposition 2

Proof. We prove the result for welfare assessments from the perspective of date 0 (without loss of generality) to make the notation easier. Consider three dates $0 < k \leq T$. We compare welfare assessments of the perturbation $d\theta_T(s^T)$ for a particular history s^T. We simply refer to this perturbation as "$d\theta$", again to make the notation easier. From the perspective of date 0 and history s^0, we have

$$
\frac{dW_{s^0}}{d\theta} = \sum_i \alpha^i_{s^0} \frac{dV^i_{s^0}}{d\theta} = \sum_i \alpha^i_{s^0} \sum_t \sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^0)}{\partial \theta}
$$

From the perspective of some history s^k at the later date k, we have

$$
\frac{dW_{s^k}}{d\theta} = \sum_i \alpha^i_{s^k} \frac{dV^i_{s^k}}{d\theta} = \sum_i \alpha^i_{s^k} \sum_{t \geq k} \sum_{s^t \geq s^k} (\beta^i)^{t-k} \pi(s^t \mid s^k) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t, \boldsymbol{\theta} \mid s^k)}{\partial \theta}
$$

Notice that we simply use $c^i_t(s^t)$ when writing marginal utility rather than the consumption function. When evaluated at the status quo, the consumption allocation (in levels rather than changes) does not depend on the perspective of the assessment, which lets us write $\mathcal{C}^i_t(s^t, \bar{\boldsymbol{\theta}} \mid s^0) = \mathcal{C}^i_t(s^t, \bar{\boldsymbol{\theta}} \mid s^k) = c^i_t(s^t)$, when evaluated at a particular $\bar{\boldsymbol{\theta}}$.

We now derive an expression that constructively isolates each potential source of time inconsistency.

Interpersonal welfare comparisons. Notice that we can write the ex ante assessment as

$$\begin{aligned}
\frac{dW_{s^0}}{d\theta} &= \sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{s^k}} \alpha^i_{s^k} \sum_t \sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta} \\
&= \left(\frac{1}{I} \sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{s^k}} \right) \sum_i \alpha^i_{s^k} \sum_t \sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta} + \mathbb{C}ov^{\Sigma}_i \left(\frac{\alpha^i_{s^0}}{\alpha^i_{s^k}},\ \alpha^i_{s^k} \frac{dV^i_{s^0}}{d\theta} \right)
\end{aligned}$$

Kydland-Prescott. We now split the sum $\sum_t \sum_{s^t}$ of the first term. We can write

$$\begin{aligned}
\frac{dW_{s^0}}{d\theta} = \left(\frac{1}{I} \sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{s^k}} \right) \Bigg\{ & \sum_i \alpha^i_{s^k} \sum_{t \geq k} \sum_{s^t \geq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta} \\
&+ \sum_i \alpha^i_{s^k} \sum_{t \geq k} \sum_{s^t \not\geq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta} \\
&+ \sum_i \alpha^i_{s^k} \sum_{0 \leq t < k} \sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta} \Bigg\} \\
&+ \mathbb{C}ov^{\Sigma}_i \left(\frac{\alpha^i_{s^0}}{\alpha^i_{s^k}},\ \alpha^i_{s^k} \frac{dV^i_{s^0}}{d\theta} \right)
\end{aligned}$$

Notice that the second and third lines correspond precisely to conditions (i) and (ii) of Proposition 1.

Next, we define the continuation lifetime utility from s^k onwards but from the perspective of s^0 as

$$V^i_{s^0}(s^k) = \sum_{t \geq k} \sum_{s^t \geq s^k} (\beta^i)^{t-k} \pi(s^t \mid s^k) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta}.$$

In other words, the only difference between this and the actual lifetime utility from the perspective of s^k, $V^i_{s^k}$, is that the consumption function is evaluated from different perspectives.

Also notice that for each individual i we have

$$(\beta^i)^t \pi(s^t) = (\beta^i)^k \pi(s^k) \cdot (\beta^i)^{t-k} \pi(s^t \mid s^k).$$

Using this, we can write

$$\begin{aligned}
\frac{dW_{s^0}}{d\theta} = \left(\frac{1}{I}\sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{s^k}}\right)\Bigg\{ & \sum_i \alpha^i_{s^k} \sum_{t\geq k}\sum_{s^t\geq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^k)}{\partial\theta} \\
& + \sum_i \alpha^i_{s^k} \sum_{t\geq k}\sum_{s^t\geq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \left(\frac{\partial \mathcal{C}^i_t(s^t\,|\,s^0)}{\partial\theta} - \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^k)}{\partial\theta}\right) \\
& + \sum_i \alpha^i_{s^k} \sum_{t\geq k}\sum_{s^t\ngeq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^0)}{\partial\theta} \\
& + \sum_i \alpha^i_{s^k} \sum_{0\leq t<k}\sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^0)}{\partial\theta} \Bigg\} \\
& + \mathbb{C}ov^{\Sigma}_i\left(\frac{\alpha^i_{s^0}}{\alpha^i_{s^k}},\ \alpha^i_{s^k}\frac{dV^i_{s^0}}{d\theta}\right)
\end{aligned}$$

or simply

$$\begin{aligned}
\frac{dW_{s^0}}{d\theta} = \left(\frac{1}{I}\sum_i \frac{\alpha^i_{s^0}}{\alpha^i_{s^k}}\right)\Bigg\{ & \sum_i \alpha^i_{s^k} (\beta^i)^k \pi(s^k) \frac{dV^i_{s^k}}{d\theta} \\
& + \sum_i \alpha^i_{s^k} (\beta^i)^k \pi(s^k) \left(\frac{dV^i_{s^0}(s^k)}{d\theta} - \frac{dV^i_{s^k}}{d\theta}\right) \\
& + \sum_i \alpha^i_{s^k} \sum_{t\geq k}\sum_{s^t\ngeq s^k} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^0)}{\partial\theta} \\
& + \sum_i \alpha^i_{s^k} \sum_{0\leq t<k}\sum_{s^t} (\beta^i)^t \pi(s^t) u'(c^i_t(s^t)) \frac{\partial \mathcal{C}^i_t(s^t\,|\,s^0)}{\partial\theta} \Bigg\} \\
& + \mathbb{C}ov^{\Sigma}_i\left(\frac{\alpha^i_{s^0}}{\alpha^i_{s^k}},\ \alpha^i_{s^k}\frac{dV^i_{s^0}}{d\theta}\right)
\end{aligned}$$

where the second line corresponds to condition (iii) of Proposition 1 and the last line corresponds to condition (iv) of Proposition 2.

Common discount factor. We conclude the proof by assuming a common discount factor $\beta^i = \beta$ as in the main text. In that case, we have

$$\sum_i \alpha^i_{s^k} (\beta^i)^k \pi(s^k) \frac{dV^i_{s^k}}{d\theta} = \beta^k \pi(s^k) \sum_i \alpha^i_{s^k} \frac{dV^i_{s^k}}{d\theta} = \beta^k \pi(s^k) \frac{dW_{s^k}}{d\theta}$$

Therefore, the first line above becomes

$$\underbrace{\left(\frac{1}{I}\sum_{i}\frac{\alpha^{i}_{s^0}}{\alpha^{i}_{s^k}}\right)\beta^{k}\pi(s^{k})}_{>0}\frac{dW_{s^k}}{d\theta}$$

This tells us that we can express the ex ante welfare assessment $\frac{dW_{s^0}}{d\theta}$ as a function of the ex post assessment $\frac{dW_{s^k}}{d\theta}$. The scalar coefficient here is necessarily strictly positive. So this term will never be a source of time inconsistency. This concludes the proof.

Heterogeneity in β^{i} implies that aggregation of discount factors can lead to a time consistency problem as discussed in Jackson and Yariv (2014, 2015).

■

B.3 Proof of Proposition 3

Proof. We consider the Kaldor-Hicks efficiency assessment of a perturbation $d\theta_{T}(s^{T})$, which we refer to simply as "$d\theta$" to make the notation easier. As in previous proofs, we compare the ex ante assessment at s^0 with ex post assessments at arbitrary histories s^k.

The ex ante efficiency assessment can be written as

$$\begin{aligned}
\Xi^{E}_{s^0} &= \sum_{i}\frac{1}{\lambda^{i}_{s^0}}\frac{dV^{i}_{s^0}}{d\theta} \\
&= \sum_{i}\frac{\lambda^{i}_{s^k}}{\lambda^{i}_{s^0}}\frac{1}{\lambda^{i}_{s^k}}\frac{dV^{i}_{s^0}}{d\theta} \\
&= \left(\frac{1}{I}\sum_{i}\frac{\lambda^{i}_{s^k}}{\lambda^{i}_{s^0}}\right)\sum_{i}\frac{1}{\lambda^{i}_{s^k}}\frac{dV^{i}_{s^0}}{d\theta} + \mathbb{C}ov^{\Sigma}_{i}\left(\frac{\lambda^{i}_{s^k}}{\lambda^{i}_{s^0}}, \frac{1}{\lambda^{i}_{s^k}}\frac{dV^{i}_{s^0}}{d\theta}\right)
\end{aligned}$$

The covariance term corresponds precisely to condition (v) of Proposition 3.

We can now rewrite the first term, splitting the sum, as

$$\Xi^E_{s^0} = \left(\frac{1}{I}\sum_i \frac{\lambda^i_{s^k}}{\lambda^i_{s^0}}\right)\Bigg\{\sum_i \frac{1}{\lambda^i_{s^k}}\beta^k \pi(s^k)\frac{dV^i_{s^k}}{d\theta}$$

$$+\sum_i \frac{1}{\lambda^i_{s^k}}\beta^k \pi(s^k)\left(\frac{dV^i_{s^0}(s^k)}{d\theta} - \frac{dV^i_{s^k}}{d\theta}\right)$$

$$+\sum_i \frac{1}{\lambda^i_{s^k}}\sum_{t\geq k}\sum_{s^t \not\succeq s^k}(\beta^i)^t \pi(s^t) u'(c^i_t(s^t))\frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta}$$

$$+\sum_i \frac{1}{\lambda^i_{s^k}}\sum_{0\leq t<k}\sum_{s^t}(\beta^i)^t \pi(s^t) u'(c^i_t(s^t))\frac{\partial \mathcal{C}^i_t(s^t \mid s^0)}{\partial \theta}\Bigg\}$$

$$+\mathbb{C}ov^{\Sigma}_i\left(\frac{\lambda^i_{s^k}}{\lambda^i_{s^0}}, \frac{1}{\lambda^i_{s^k}}\frac{dV^i_{s^0}}{d\theta}\right)$$

Lines two, three and four correspond precisely to conditions (i) – (iii) of Proposition 1. And the last line corresponds to the new numeraire inconsistency condition (v) in Proposition 3.

Finally, notice that Assumption A1, which rules out time inconsistency á la Kydland-Prescott, implies

$$\frac{dV^i_{s^0}}{d\theta} = \beta^k \pi(s^k)\frac{dV^i_{s^k}}{d\theta}.$$

Under A1, therefore, we have

$$\Xi^E_{s^0} = \left(\frac{1}{I}\sum_i \frac{\lambda^i_{s^k}}{\lambda^i_{s^0}}\right)\beta^k \pi(s^k)\Xi^E_{s^k} + \beta^k\pi(s^k)\mathbb{C}ov^{\Sigma}_i\left(\frac{\lambda^i_{s^k}}{\lambda^i_{s^0}}, \frac{1}{\lambda^i_{s^k}}\frac{dV^i_{s^k}}{d\theta}\right)$$

corresponding to equation (17) in the main text.

■

B.4 Proof of Proposition 4

Proof. We start with a proof of condition (i). If a perturbation $d\theta$ is a Pareto improvement, then it must be that $\frac{dV^i_{s^t}}{d\theta} \geq 0$ for all i, with at least one inequality. Therefore, the sum of individual welfare gains expressed in any valid (with $\lambda^i_{s^t} > 0$) welfare numeraire (forward or backward-looking) must be strictly positive. So we must also have $\Xi^E_{s^t} = \sum_i \frac{1}{\lambda^i_{s^t}}\frac{dV^i_{s^t}}{d\theta} > 0$.

If instead there is a feasible perturbation for which $\Xi^E_{s^t} = \sum_i \frac{1}{\lambda^i_{s^t}}\frac{dV^i_{s^t}}{d\theta} > 0$ where $\lambda^i_{s^t}$ corresponds to a forward-looking numeraire, then it must be possible to transfer resources from winners to losers in units of that welfare numeraire so that $\frac{1}{\lambda^i_{s^t}}\frac{dV^i_{s^t}}{d\theta} \geq 0$ for all i, with at least one strict inequality, while ensuring that $\Xi^E_{s^t}$ does not change. This direction is only valid for forward-looking numeraires: If we chose a backward-looking numeraire, the transfers needed to construct the Pareto-improving perturbations would not be feasible, as it is impossible to transfer resources in the past.

Now we prove condition (ii) by contradiction. Suppose that $\Xi^E_{s^t} > 0$ for some feasible perturbation of a given allocation. Then by virtue of (i), we can reallocate resources among individuals to find a Pareto improvement with transfers. But this means that the original allocation was not Pareto efficient, leading to a contradiction. It is important to note that this result only applies to the set of Pareto efficient allocations that solve the Pareto Problem (first-best allocations). It does not apply to constrained efficient allocations. If we chose a backward-looking numeraire, the transfers needed to construct the Pareto-improving perturbations would not be feasible, as it is impossible to transfer resources in the past.

■

B.5 Proof of Theorem 1

Proof. Our proof of Proposition 3 presents a systematic characterization of the sources of time inconsistency. It is sufficient to prove Theorem 1 under Assumption A1, which rules out time consistency problems á la Kydland and Prescott (1977). Under A1, time inconsistency is therefore governed by the value of

$$\mathbb{C}ov_i\left[\frac{\lambda^i_{s^1}}{\lambda^i_{s^0}}, \frac{\frac{dV^i_{s^1}(s^1)}{d\theta}}{\lambda^i_{s^1}}\right].$$

Property P1 of Theorem 1 restricts us to forward-looking numeraires, which means that $\lambda^i_{s^t}$ must be a function of $\{c^i_k(s^k)\}_{k\geq t, s^k \geq s^t}$.

Suppose that the forward-looking numeraire at date 0 includes date 0 consumption. See footnote 22 for a discussion of why this assumption is without loss since Theorem 1 is "for all $d\theta$". Under incomplete markets, we will therefore be able to find a pair of histories at which individual MRS are not equalized, which then also implies that

$$\frac{\lambda^i_{s^t}}{\lambda^i_{s^0}}$$

are not equalized in the cross section. And since different perturbations are associated with different values of $\frac{1}{\lambda^i_{s^t}}\frac{dV^i_{s^t}}{d\theta}$, we can always find a perturbations such that the second term in equation (17) is sufficiently large as to change the sign of $\Xi^E_{s^t}$ relative to $\Xi^E_{s^0}$. Hence, it is possible to find a perturbation for which the efficiency assessment based on a forward-looking numeraire is time inconsistent.

■

B.6 Proof of Proposition 6

Proof. Recall that the history s^k aggregate efficiency assessment is defined as

$$\Xi_{s^k}^{AE} = \sum_{t \geq k} \bar{\omega}_{t|s^k} \sum_{s^t \geq s^k} \bar{\omega}_{t|s^k}(s^t) \sum_i \frac{\partial \mathcal{C}_t^i(s^t \mid s^k)}{\partial \theta}$$

When the perturbation $d\theta$ generates a *static* aggregate efficiency gain or loss, this means that

$$\frac{\partial \mathcal{C}_t^i(s^t \mid s^k)}{\partial \theta} = 0$$

for all $t \neq T$ and $s^t \neq s^T$. Therefore, we can write the static aggregate efficiency assessment for such a perturbation from the perspective of history s^k as

$$\Xi_{s^k}^{AE} = \bar{\omega}_{T|s^k} \bar{\omega}_{T|s^k}(s^T) \sum_i \frac{\partial \mathcal{C}_T^i(s^T \mid s^k)}{\partial \theta}$$

When Assumption A1 is satisfied, then

$$\frac{\partial \mathcal{C}_T^i(s^T \mid s^k)}{\partial \theta} = \frac{\partial \mathcal{C}_T^i(s^T \mid s^t)}{\partial \theta} = \frac{dc_T^i(s^T)}{d\theta}$$

for any s^t and s^k. Therefore, under A1, the perceived effect of perturbation $d\theta$ on aggregate consumption at history s^T does not change with the perspective of the assessment.

The only sources of time inconsistency can therefore be the aggregate valuations $\bar{\omega}_{T|s^k}$ and $\bar{\omega}_{T|s^k}(s^T)$. But since $\lambda_{s^k}^i > 0$ and $u'(c_t^i(s^t)) > 0$ for all t, s^t and i, these two weights must also be strictly positive. Therefore, we have

$$\Xi_{s^0}^{AE} = \underbrace{\frac{\bar{\omega}_{T|s^0} \bar{\omega}_{T|s^0}(s^T)}{\bar{\omega}_{T|s^k} \bar{\omega}_{T|s^k}(s^T)}}_{>0} \Xi_{s^k}^{AE}.$$

This tells us that "static" aggregate efficiency assessments are always time consistent under Assumption A1, concluding our proof.

∎

C Microfoundation of the Consumption Function

There are I individuals indexed by i. Time is discrete and we capture uncertainty using the usual history notation, allowing for both aggregate and idiosyncratic uncertainty. We assume that each individual i makes a single decision $c_t^i(s^t) \in \mathbb{R}$ at date t and conditional on the realization of history s^t.

Behavior. We focus on environments in which behavior is characterized by a *policy function*

$$C\left(\phi, x, \vec{p}, \vec{\theta}\right).$$

Our notation is as follows:

- x denotes idiosyncratic time-varying state variables (ex-post heterogeneity)
- ϕ denotes idiosyncratic permanent state variables or types (ex-ante heterogeneity)
- X denotes aggregate state variables
- $\vec{p}$ denotes the "continuation price process", i.e., a stochastic process with a particular initialization (see below)
- $\vec{\theta}$ denotes the "continuation policy process" (see below)

Notice that we refer as "prices" to every (aggregate) macroeconomic variable that affects (the decision problems of) agents directly. We also introduce the aggregate state X at this point; while it is not a direct argument of the policy function $C(\cdot)$, it will affect the determination of prices below. Formally, we have

$$C : \mathbb{R}^{N_\phi} \times \mathbb{R}^{N_x} \times \mathcal{L}^{N_p} \times \mathcal{L} \to \mathbb{R}.$$

We denote the space of stochastic processes under consideration (initialized in a certain way) by $\mathcal{L}$. There is a single policy process θ.

Using this consumption function, we can now express individual i's consumption decision at date t in history s^t as

$$c_t^i(s^t) = C\left(\phi^i, x_t^i(s^t), \left\{p_\ell(s^\ell), \theta_\ell(s^\ell)\right\}_{\ell \geq t, s^\ell \geq s^t}\right).$$

In other words, given agent i's type and state, and given the pair of stochastic processes $\vec{p}$ and $\vec{\theta}$ initialized at $p_t(s^t)$ and $\theta_t(s^t)$, this policy function determines behavior $c_t^i(s^t)$.

Environment. The environment is defined as a description of how state variables and prices are determined. Notice that the permanent types ϕ^i never change.

First, we have idiosyncratic state variables, which evolve according to the law of motion

$$x_{t+1}^i(s^{t+1}) = h\left(\phi^i, x_t^i(s^t), c_t^i(s^t), p_t(s^t), \theta_t(s^t), s_{t+1}\right)$$

Second, we have the aggregate state variables. We assume they evolve according to the law of motion

$$X_{t+1}(s^{t+1}) = k\Big(X_t(s^t), p_t(s^t), \theta_t(s^t), \Big\{c_t^i(s^t)\Big\}_i, s_{t+1}\Big).$$

Third and finally, we have macroeconomic prices. We assume that they solve "market clearing conditions" given by

$$p_t(s^t) = m\Big(X_t(s^t), \theta_t(s^t), \Big\{c_t^i(s^t)\Big\}_i\Big).$$

We do not let $m(\cdot)$ explicitly depend on s_t. This is without loss because, as explained above, we already allow for the aggregate state $X_t(s^t)$ to depend on s_t directly, and so aggregate "shocks" can always be written as one of the aggregate states.

This concludes our description of the environment.

Dual representation. We now derive what we will refer to as a "dual representation" of behavior.

Notice first that we can iterate on the law of motion for $x_t^i(s^t)$ and arrive at

$$x_t^i(s^t) = H_t\Big(x_0^i(s^0), \phi^i, \Big\{c_k^i(s^k), p_k(s^k), \theta_k(s^k)\Big\}_{0\leq k<t, s^0\leq s^k<s^t}, s^t\Big)$$

So what matters for the determination of individual i's state in history s^t is the initial condition $x_0^i(s^0)$, her type, her behavior since the world started, all the shocks she drew, and then the prices and policies she faced along the way. Crucially, the individual state is *backward-looking*. Finally, we index $H_t(\cdot)$ by t because the size of the input arguments changes with time. We have

$$H_t : \mathbb{R}^{N_x} \times \mathbb{R}^{N_\phi} \times \mathbb{R}^t \times \mathbb{R}^{t\times N_p} \times \mathbb{R}^t \times \mathcal{S}^t \to \mathbb{R}^{N_x}$$

Second, for the aggregate state variables, we can also iterate backwards and arrive at

$$X_t(s^t) = K_t\Big(X_0(s^0), \Big\{p_k(s^k), \theta_k(s^k)\Big\}_{0\leq k<t, s^0\leq s^k<s^t}, \Big\{c_k^i(s^k)\Big\}_{i, 0\leq k<t, s^0\leq s^k<s^t}, s^t\Big)$$

Notice that in both cases the equalities are strict, i.e., $k < t$ and $s^k < s^t$. Notice also that the aggregate state depends on all agents' behavior.

We can now put together the previous equation with the market clearing condition, yielding

$$p_t(s^t) = M_t\Big(X_0(s^0), \Big\{\theta_k(s^k)\Big\}_{0\leq k\leq t,\ s^0\leq s^k\leq s^t}, \Big\{c_k^i(s^k)\Big\}_{i,\ 0\leq k\leq t,\ s^0\leq s^k\leq s^t}\Big).$$

Several observations are in order: First, the inequalities are now weak. That's because $p_t(s^t)$ depends both on contemporaneous policy $\theta_t(s^t)$ directly and on prior policy indirectly through the aggregate state. The same is the case for consumption. Second, one could in principle subsume the initial condition $X_0(s^0)$ directly into the function $M(\cdot)$ since it is policy-invariant. But since initial conditions will be critical when assessing policy implications from different perspectives, we leave the dependence explicit here. Finally, we must have the function $M_t(\cdot)$ depend on calendar time because the size of the input arguments changes over time. Notice, for example, that the second argument of $M_t(\cdot)$ is a $t \times 1$ vector. In other words,

$$M_t : \mathbb{R}^{N_X} \times \mathbb{R}^t \times \mathbb{R}^{t \times I} \to \mathbb{R}^{N_p}.$$

It is a function of vectors (rather than matrices to account for histories) because the only objects that matter are the realizations of policy and consumption *along the realized history*.

We can now use the previous expressions and rewrite the law of motion for individual state variables via $H_t(\cdot)$ as

$$\begin{aligned} x_t^i(s^t) &= H_t\Big(x_0^i(s^0), \phi^i, \Big\{c_k^i(s^k), p_k(s^k), \theta_k(s^k)\Big\}_{0 \le k < t, s^0 \le s^k < s^t}, s^t\Big) \\ &= \mathcal{H}_t\Big(\phi^i, x_0^i(s^0), X_0(s^0), \Big\{c_k^i(s^k)\Big\}_{0 \le k < t}^{s^0 \le s^k < s^t}, \Big\{\theta_k(s^k)\Big\}_{0 \le k \le t}^{s^0 \le s^k \le s^t}, \Big\{c_k^j(s^k)\Big\}_{j,\, 0 \le k \le t}^{s^0 \le s^k \le s^t}, s^t\Big), \end{aligned}$$

where we index brackets $\{\cdot\}$ by sub- and superscripts only to make the notation more compact. This notation makes explicit the dependence of $x_t^i(s^t)$ on consumption decisions through different channels. At the cost of conflating these channels, we can simplify and arrive at

$$x_t^i(s^t) = \mathcal{H}_t\Big(\phi^i, x_0^i(s^0), X_0(s^0), \Big\{\theta_k(s^k)\Big\}_{0 \le k \le t}^{s^0 \le s^k \le s^t}, \Big\{c_k^j(s^k)\Big\}_{j,\, 0 \le k \le t}^{s^0 \le s^k \le s^t}, s^t\Big).$$

Finally, we can go back to the policy function that determines behavior. Reproducing from above for convenience,

$$c_t^i(s^t) = C\Big(\phi^i, x_t^i(s^t), X_t(s^t), \Big\{p_\ell(s^\ell), \theta_\ell(s^\ell)\Big\}_{\ell \ge t, s^\ell \ge s^t}\Big).$$

Notice that using the price equation, we can rewrite the aggregate state as

$$X_t(s^t) = \mathcal{K}_t\Big(X_0(s^0), \Big\{\theta_k(s^k)\Big\}_{0 \le k < t}^{s^0 \le s^k < s^t}, \Big\{c_k^i(s^k)\Big\}_{i, 0 \le k < t}^{s^0 \le s^k < s^t}, s^t\Big)$$

Finally, we can plug into the consumption function, which gives us

$$c_t^i(s^t) = \tilde{\mathcal{C}}_t\Big(\phi^i, x_0^i(s^0), X_0(s^0), \Big\{\theta_k(s^k)\Big\}_{0 \le k \le t}^{s^0 \le s^k \le s^t}, \Big\{c_k^j(s^k)\Big\}_{j,\, 0 \le k \le t}^{s^0 \le s^k \le s^t}, \Big\{p_\ell(s^\ell), \theta_\ell(s^\ell)\Big\}_{\ell \ge t}^{s^\ell \ge s^t}, s^t\Big),$$

or simply

$$c_t^i(s^t) = \mathcal{C}_t\Big(s^0, \Big\{\theta_\ell(s^\ell)\Big\}_{\ell \ge 0}^{s^\ell \ge s^0}\Big),$$

after expressing the forward-looking price process as a function of the policy process.

This derivation illustrates constructively the various channels through which a perturbation in a policy parameter $\theta_k(s^k)$ affects consumption of different individuals at different horizons.