Intertemporal Distortions in the Second Best*

Stefania Albanesi† Roc Armenter‡
Columbia University, NBER and CEPR Federal Reserve Bank of Philadelphia

First version: November 2007
This version: August 30, 2011

Abstract

This paper studies the long run properties of intertemporal distortions in a broad class of second best economies. Our unified framework encompasses and extends many well known models, such as variants of the Ramsey taxation model with aggregate or idiosyncratic risk, and economies with incentive compatibility constraints due to limited commitment, political economy, self-enforcement or private information, or combinations of these. We identify a sufficient condition that rules out permanent intertemporal distortions: If there exists an allocation that satisfies all constraints and eventually converges to the limiting first best allocation, then intertemporal distortions are temporary in the second best. This result uncovers a common optimality principle linking the intertemporal allocation of resources with the ability to front-load distortions for this broad class of environments. A series of applications illustrates the significance of these findings.

JEL Classification:E6,H21,H3
Keywords: Intertemporal distortions; Second best; Ramsey policies; Limited commitment; Private information; Self-enforcement constraints.

*We would like to thank Fernando Alvarez, Manuel Amador, Boragan Aruoba, Marco Bassetto, Martin Bodenstein, V.V. Chari, Sanjay Chugh, Dirk Krueger, Pierre Yared, and seminar participants at the University of Maryland, the University of Pennsylvania, Georgetown, Stanford, Chicago GSB, Duke, Cornell, UCSB, the University of Cambridge, the Federal Reserve Banks of New York, Philadelphia and San Francisco, and the NBER Summer Institute for helpful comments and suggestions. We also thank the three reviewers and the editor for very useful suggestions. The views expressed here do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

†Contact: Department of Economics, Columbia University, 420 West 118th Street, Suite 1022, New York NY 10027. Email: stefania.albanesi@columbia.edu.

‡Contact: Federal Reserve Bank of Philadelphia, 10 Independence Mall, Philadelphia PA 19106. Email: roc.armenter@phil.frb.org.

1 Introduction

The intertemporal allocation of resources has implications for aggregate income, risk sharing, asset pricing, and other key areas of interest in macroeconomics. If feasibility were the only constraint, then it would be possible to attain a fully efficient, first best, allocation. In reality, a variety of additional constraints imply that the allocation cannot be fully efficient and will feature distortions. This paper studies the long run behavior of intertemporal distortions in a broad class of constrained-efficient, second best, economies.

Many possible sources of frictions can constrain the allocation of resources. Perhaps the most immediate of these is taxes. The Ramsey model provides a benchmark for optimal taxation in macroeconomics. Chamley (1986) and Judd (1985) first established that capital income taxes are zero in the steady state and all intertemporal distortions eventually vanish, a finding that has been confirmed for a variety of more general Ramsey models.[1] Another natural source of distortions is private information. The resulting incentive problems generate permanent intertemporal distortions, as the recent work of Golosov, Kocherlakota and Tsyvinski (2003) has emphasized.[2] For other frictions, such as self-enforcement, limited commitment, political economy, incomplete markets and so on, results differ on the desirability of long-run intertemporal distortions.

Why are permanent intertemporal distortions ruled out in some environments? What is different about economies where they are optimal? To explore these questions we propose a unified framework designed to encompass many sources of frictions. We allow for a general formulation of the constraints on resource allocation, which we refer to as *admissibility constraints*. We can capture variants of the Ramsey taxation model with aggregate or idiosyncratic risk, economies with incentive compatibility constraints due to limited commitment, political economy, self-enforcement or private information, as well as principal-agent models. Our framework includes also economies that combine these constraints with arbitrary restrictions, such as incomplete markets and borrowing constraints on the government or on private agents.

Our general analysis uncovers a common optimality principle linking the intertemporal allocation of resources with the ability to front-load distortions. We show that future admissibility constraints eventually stop binding when it is possible to front-load all distortions. If no future admissibility constraints are binding, then consumption can be re-allocated intertemporally without violating any admissibility constraints. Such a reallocation induces a first-order welfare gain if there is a wedge between the current and future marginal utility of consumption. Thus, an allocation with permanent intertemporal distortions cannot be optimal when it is possible to front-load all distortions. Front-loading all distortions may not be possible in some environments, and in that case intertemporal distortions may be permanent.[3]

[1] Chari and Kehoe (1999) and Atkeson, Chari and Kehoe (1999) extend the Chamley-Judd result for a broad class of deterministic economies. Zhu (1992) and Farhi (2010) show that it holds also with aggregate uncertainty for complete and incomplete markets, respectively.

[2] Diamond and Mirrlees (1978) and Rogerson (1985) were the first to show that intertemporal distortions arise in dynamic disability insurance and moral hazard models. Golosov, Kocherlakota and Tsyvinski (2003) prove that intertemporal distortions prevail in a large class of private information economies.

[3] Front-loading is a well known property of efficient allocations with self-enforcement constrains, see Ray (2002) and

Formally, we identify a sufficient condition that rules out permanent intertemporal distortions in the second best: If there exists an allocation that satisfies all admissibility constraints and that eventually converges to the limiting first best allocation, then intertemporal distortions are temporary in the second best. Since this condition is not necessary, there are economies with no intertemporal distortions where the condition does not hold.

Importantly, the sufficient condition is a property of the set of admissible allocations, and not a property of the optimum. The condition does not imply that the second best will converge to the first best limiting allocation. Intratemporal distortions depend on the history of *past* binding constraints and typically do not vanish in the long run.

For example, our sufficient condition is satisfied in the standard Ramsey taxation model. Since the government can save enough to finance all its expenditures from asset returns, eventually it becomes possible to eliminate distortionary taxes and converge to the first best steady state. Intertemporal distortions are temporary, but intratemporal distortions, such as labor income taxes, will typically be present in the steady state. By contrast, the sufficient condition generally does not hold in private information economies, since incentive compatibility constraints will never be satisfied by the limiting first best allocation.[4]

Our main contribution is to provide a unified framework that offers insight about how the nature of admissibility constraints shapes intertemporal distortions. This can prove useful for both normative and positive analysis of second best allocations. Consider intertemporal distortions from a normative standpoint. The Ramsey framework has been challenged for its arbitrary restrictions on fiscal instruments. Deriving constraints from primitive frictions, such as private information, ensures all relevant trade-offs are considered and instills greater confidence in the resulting policy prescriptions.[5] Yet, this approach does not identify the relevant trade-off missing in the Ramsey model. Our result makes clear that the only rationale for intertemporal distortions is to relax future admissibility constraints. Thus, even as the optimal allocation front-loads distortions in both environments, it is the ability to front-load all distortions in the Ramsey model that gives rise to the different prescriptions regarding intertemporal distortions. In addition, our focus on distortions rather than taxes leads to more general predictions.[6]

From a positive standpoint, our sufficient condition can also be useful for establishing robust long-run properties of the intertemporal margin for different frictions. We discuss our result in the context of several canonical models in macroeconomics. For environments with a benevolent

the discussion in Ljungqvist and Sargent (2004), Chapter 19. Our results demonstrate that this optimality principle is intimately linked with the absence of permanent intertemporal distortions for a broad class of economies.

[4]Our sufficient condition may be satisfied under private information if there is an absorbing state, agents are risk-neutral, or there are additional instruments to provide incentives. In each of these cases there are no permanent intertemporal distortions. See Golosov and Tsyvinski (2006), Clementi and Hopenhayn (2006), and Armenter and Mertens (2010), respectively.

[5]See Kocherlakota (2005a) for an extensive discussion on this point.

[6]It also clarifies some contrasting results in the literature that are driven by specific assumptions on the decentralization. As is well known, there may be many ways to implement constrained-efficient allocations with a combination of taxes and trading restrictions in a decentralized arrangement. See Kocherlakota (2005b), Albanesi and Sleet (2006), and Albanesi (2008) for a discussion.

government setting policies under limited commitment, we show that the sufficient condition always holds. Thus, limited commitment alone cannot give rise to permanent intertemporal distortions. For economies with self-enforcement constraints, we show how the assumptions on the agents' outside option determine whether the sufficient condition holds, and thus determine the long run properties of intertemporal distortions. For this class of economies, Ray (2002) analyzes a principal-agent setting and shows that constrained-efficient allocations front-load distortions and converge to the unconstrained optimum.[7]

The general model we consider encompasses many dynamic constrained-efficient economies that have become standard in macroeconomic theory, as well as principal-agent models often used in dynamic contracting. Agents are infinitely lived with standard time separable preferences defined over consumption and labor. The production function is neoclassical and exhibits constant returns to scale. We allow for ex ante heterogeneity, idiosyncratic and aggregate shocks. The second best problem is formulated as a choice of feasible allocations subject to admissibility constraints.[8] The admissibility constraints are forward-looking and display limited history dependence. Implementability constraints in Ramsey taxation models, self-enforcement constraints and incentive compatibility constraints in private information economies typically take on this form.

The level of generality required to provide such a unified treatment comes at a cost. The assumptions required for tractability are naturally more demanding the larger is the space of problems considered. Specifically, we assume the second best allocations are interior and converge. Since the set of admissible allocations is often not convex, we also introduce a condition that rules out local maxima and inflection points.[9] These restrictions are commonplace in the literature, though their formulation is typically specific to the particular application considered. We state these assumptions directly as restrictions on the second best allocation, since doing so in terms of primitives would significantly restrict the generality of our environment. Equivalent conditions on primitives are typically available for specific applications, and these may be more transparent.

While we emphasize that our condition is sufficient but not necessary, the logic of our results holds beyond its strict mathematical confines. Our formulation of the sufficient condition is chosen to ensure that our result holds for all the economies encompassed in the general framework. Consequently, the condition is often stronger than required for the result to hold in a specific application. In particular, the sufficient condition may not be satisfied in some environments, while the front-loading principle that drives our result still holds. The Ramsey model with a government balanced budget is one example. In that model, the sufficient condition is not satisfied, yet it is possible to front-load some, though not all, distortions, which guarantees that intertemporal distortions will be temporary.[10]

Despite these limitations, verifying the sufficient condition is often more straightforward than

[7] Ray (2002) does not discuss implications for intertemporal distortions, and his setting does not feature risk aversion, physical state variables or stochastic shocks.

[8] This formulation follows the primal approach to optimal taxation, pioneered in Atkinson and Stiglitz (1980) and Lucas and Stokey (1983).

[9] This condition adapts assumption A.2 in Ray (2002) to our general environment. See Section 3.2 for details.

[10] This example is discussed in detail in Section 5.1.

fully characterizing the optimal allocation. Our result then provides a partial characterization of the limiting optimal allocation. This can be an advantage, since many constrained-efficient economies are rich and present considerable analytical and computational challenges.

The paper is organized as follows. Section 2 presents a number of examples of well known second best economies, which motivate the general second best framework presented in Section 3. Section 4 formally proves the main result. Section 5 discusses several applications to illustrate the significance and applicability of our result. Section 6 concludes by briefly discussing what form a necessary condition for the absence of permanent intertemporal distortions would take, and how the logic embedded in our results can be applied more generally.

2 Examples

We now present a number of canonical examples of second best economies that motivate our general formulation for admissibility constraints. We also include a short discussion noting whether our sufficient condition is satisfied in these models.

2.1 Ramsey taxation

Taxes are perhaps the most immediate source of distortions in the allocation of resources. The Ramsey model has been the benchmark for the analysis of optimal taxation in macroeconomics.[11] In this framework, a benevolent government sets proportional labor and capital income taxes to finance an exogenous stream of government consumption. The government has commitment and can issue debt of different maturities. Crucially, lump-sum taxes are ruled out, which implies that the first best allocation cannot be attained.

The constraints on fiscal instruments that define the Ramsey equilibrium generally can be expressed in terms of allocations, an approach that greatly simplifies the analysis. For example, admissible allocations in a Ramsey model with a representative agent and no uncertainty are fully characterized by the following *implementability* constraint:

$$\sum_{t=0}^{\infty} \beta^t \left(u_t^c c_t + u_t^l l_t \right) \geq u_0^c \left\{ \left[\left(1 - \tau_0^k \right) r_0 + 1 - \delta \right] k_0 + b_0 \right\}. \tag{1}$$

This constraint is derived by combining the government intertemporal budget constraint with the competitive equilibrium conditions for households and firms. Since consumption and labor decisions in the future affect the present value of tax revenues, the implementability constraint is forward-looking.

In stochastic economies, admissible allocations can be characterized also in terms of a single implementability constraint as long as markets are complete. However, if the government can issue only risk-free debt, it faces a sequence of implementability constraints, one for each possible history

[11] Chari and Kehoe (1999) provide an excellent review of this approach.

of shocks:

$$E_t \left\{ (u_t^c)^{-1} \sum_{d=t}^{\infty} \beta^{d-t} \left(u_d^c c_d + u_d^l l_d \right) - R_t k_t \right\} = b_{t-1} + V_t, \tag{2}$$

where b_{t-1} is the return on debt carried into period t, V_t is the present value of non-negative transfers, and R_t is the after-tax return on capital.[12] Constraint (2) is forward-looking but it cannot be expressed solely in terms of allocations. The outstanding debt and the present value of transfers parametrize the constraints on the continuation allocation in each period, even if they do not directly enter preferences or technology. Since government debt cannot be state contingent and transfers must be non-negative, there is an additional set of constraints on the evolution of these endogenous states. To accommodate this possibility, the general formulation we develop in Section 3 allows for auxiliary variables comprising the endogenous state of the economy, and a corresponding set of admissibility constraints on their law of motion.

Ramsey models generally satisfy the sufficient condition. As long as the government can save, it is possible to design a policy in which tax revenues exceed government spending until the income from government assets grows enough to finance government expenditures. At this point, there is no further need to raise tax revenues, and since taxes are the only source of distortions. Such an allocation converges to the first-best steady state that would prevail under lump sum taxes can be attained. We discuss this property more extensively in Section 5.1.

2.2 Risk Sharing

Incentive-constrained risk sharing is a classic example of a second best problem. Ex ante identical agents face idiosyncratic shocks to productivity or preferences. The first best allocation provides agents with full insurance against these shocks. However, additional frictions such as limited commitment or asymmetric information imply that full insurance typically is not attainable.

2.2.1 Self-enforcement constraints

We start with a simple example of risk sharing with self-enforcement constraints.[13] The economy is populated by ex ante identical, infinitely lived agents, with preferences defined over sequences of consumption, c, and labor, l. Agents face idiosyncratic labor productivity shocks, θ_t, for $t = 0, 1, 2, 3...$ In each period t, they are characterized by their history of productivity shocks, denoted with $\theta^t = \{\theta_0, \theta_1, ..., \theta_t\}$. Thus, an allocation is a mapping that specifies consumption as a function of the history: $\{c(\theta^t), l(\theta^t)\}_{t \geq 0}$.

The social welfare function is utilitarian:

$$E_0 \left\{ \sum_{t=0}^{\infty} \beta^t u(c(\theta^t), l(\theta^t)/\theta_t) \right\}$$

[12] See Aiyagari, Marcet, Sargent and Seppala (2003) and Farhi (2010) for an analysis of the Ramsey model under incomplete markets. For the ease of exposition we have omitted the dependence on each variable on the history of shocks.

[13] This class of environments have been analyzed by Kocherlakota (1996) and Kehoe and Levine (1993).

where the expectation is taking over all histories θ^t. Resource feasibility requires:

$$\sum_{\theta^t} \pi(\theta^t)c(\theta^t) + k_{t+1} \leq F\left(k_t, \sum_{\theta^t} \pi(\theta^t)l(\theta^t)\right),$$

where $\pi(\theta^t)$ denotes the probability of the history of labor productivity realizations, θ^t.

The first best allocation fully insures agents from their idiosyncratic shocks. The ratio of the marginal utility of consumption across agents is constant over time, and for each agent the marginal rate of substitution between consumption and labor is equal to their labor productivity. Thus, high productivity agents supply more labor for standard preferences.

Absent an enforcement technology, an agent may have an incentive to deviate from the risk-sharing arrangement in a state when she has high productivity. If she deviates, she has access to an outside option, with value $V_{out}(k_t;\theta_t)$.[14] Any second best allocation must satisfy a *self-enforcement* constraint that imposes for each date that the present value of utility from continuing with the allocation is greater that the utility associated with the outside option. Formally:

$$E_t\left\{\sum_{j=0}^{\infty} \beta^j u(c(\theta^{t+j}), l(\theta^{t+j})/\theta_{t+j})\right\} \geq V_{out}(k_t;\theta_t), \tag{3}$$

for all θ^t. The self-enforcement constraint is forward looking, and restricts the continuation allocation in each period. The value of the outside option is a function of the exogenous and endogenous state variables. We will allow for this possibility in our formulation of a general admissibility constraint.

Let $\{c^{fb}_{t+j}, l^{fb}_{t+j}\}_{j\geq 0}(k_t)$ be a resource-feasible allocation that maximizes social welfare from time t onwards with aggregate capital given by k_t. We will refer to this object as a continuation first best allocation from k_t. The sufficient condition holds for this model if the first-best continuation allocation satisfies the self-enforcement constraint, that is:

$$E_t\left\{\sum_{j=0}^{\infty} \beta^j u(c^{fb}(\theta^{t+j}), l^{fb}(\theta^{t+j})/\theta_{t+j})\right\} \geq V_{out}(k_t,\theta_t) \tag{4}$$

for all θ_t. Whether condition (4) is satisfied depends on the parameters of the economy. For example, assuming the outside option is the anarchic allocation, condition (4) holds for a high enough value of β. We return to these economies in Section 5.

2.2.2 Private information

We now consider the same economy under the assumption that the idiosyncratic productivity shocks, θ, are private information. This is a version of the environment considered in Golosov, Kocherlakota and Tsyvinski (2003), without aggregate shocks. The first best allocation is the same as in the

[14] There are many possible formulations of the outside option, which we discuss in detail in Section 5.2.1.

previous example.

Private information implies that allocations must satisfy a truth-telling requirement known as *incentive compatibility constraint*:

$$E_t\left\{\sum_{t=0}^{\infty}\beta^t u(c_i(\theta^t), l_i(\theta^t)/\theta_t)\right\} \geq E_t\left\{\sum_{t=0}^{\infty}\beta^t u(c_i(\tilde{\theta}^t), l_i(\tilde{\theta}^t)/\theta_t)\right\}, \tag{5}$$

$\forall\tilde{\theta}^t, \theta^t$ and for $i = 1, 2$. This constraint is forward looking as in the previous example. However, the incentive-compatibility constraint now involves restrictions across allocations awarded to different ex post types, θ^t. The general formulation of admissibility constraints allows for this possibility. Capital does not enter this constraint directly, in contrast to (3).

Incentive compatibility requires individual consumption to be correlated with realized productivity in each period. Absent this correlation, agents will report a low value of θ. The first best allocation does not satisfy (5) if utility is strictly concave and the endowment process does not exhibit any absorbing states. Thus, the sufficient condition typically does not hold.

3 General Framework

We start by describing the economy and defining the best feasible allocation, which corresponds to our notion of *first best*. We then introduce our general formulation for the admissibility constraints and the corresponding concept of *second best* allocation.

3.1 The Economy

We consider an infinite-horizon economy where time is discrete $t = 0, 1, ...$ The economy is populated by a continuum of agents who may be ex-ante heterogeneous. Let I be the finite set of ex-ante types, each denoted by subscript $i \in I$, and $\pi_i > 0$ the measure of agents of ex-ante type i, with $\sum_I \pi_i = 1$.

Aggregate shocks are denoted by $z_t \in Z$, where Z is a finite set. Let $z^t = \{z_0, z_1, ..., z_t\}$ be the history of aggregate shocks from date 0 up to date t, and $z^\infty \in Z^\infty$ denotes an infinite history. The set of continuation histories from node $\tilde{z}^d$ at a later date $t \geq d$ is denoted with:

$$Z^t|\tilde{z}^d \equiv \left\{z^t \in Z^t | z^d = \tilde{z}^d\right\}.$$

We adopt this notational convention for all variables.

Aggregate shocks are governed by a first-order Markov process π_z. The probability measure of a node z^t given past history $\tilde{z}^d$ is denoted $\pi_z\left(z^t|\tilde{z}_d\right)$. Since Z is a finite set the probability measure is well defined for all subsets of Z^t and Z^∞.

We also allow for idiosyncratic shocks, so agents may be ex-post heterogeneous. Let $\theta_t \in \Theta$ denote the idiosyncratic shock at date t, where Θ is a finite set. The agent type at any date t is given by the ex-ante type i and the history of idiosyncratic shocks $\theta^t \in \Theta^t$.

Idiosyncratic shocks are governed by the first-order Markov process π_θ and are independent from the aggregate shock z_t and the agent ex-ante type. We also assume a law of large numbers holds, so $\pi_\theta\left(\theta^t\right)$ is also the measure of agents with type θ^t at date t.

Let $s_t = \{z_t, \theta_t\}$, with $s_t \in S = Z \times \Theta$, summarize the aggregate state and idiosyncratic state. Since aggregate and idiosyncratic shocks are independent, state s_t evolves stochastically according to a first-order Markov process with probability distribution $\pi\left(s^t\right) = \pi_z\left(z^t\right)\pi_\theta\left(\theta^t\right)$.

We assume that the Markov transition matrices for both aggregate and idiosyncratic shocks are strictly positive component-wise. This guarantees that there is a unique ergodic distribution associated with the exogenous state s_t.[15]

Assumption 1 *For all $s, s' \in S$, $\pi\left(s'|s\right) > 0$.*

Let $c_i\left(z^t, \theta^t\right)$ be the consumption of ex-ante type $i \in I$ after a sequence of idiosyncratic shocks $\theta^t \in \Theta^t$ and aggregate shocks $z^t \in Z^t$. To economize on notation, we will write $c_i\left(s^t\right)$ whenever there is no confusion possible. Consumption is non-negative and bounded above by $\bar{c} > 0$ at all dates,

$$c_i\left(z^t, \theta^t\right) \in [0, \bar{c}]\,.$$

Let $C \equiv [0, \bar{c}]^{|I \times \Theta|}$ where $|I \times \Theta|$ denotes the cardinality of the set. The distribution of consumption across ex-ante types at a given node s^t is denoted with:

$$c\left(s^t\right) = \left\{c_i\left(s^t\right) \in [0, \bar{c}] : i \in I\right\}.$$

A consumption allocation, c, is a plan specifying consumption for every ex-ante type at every node s^t:

$$c = \left\{c\left(s^t\right) : s^t \in S^t, t \geq 0\right\}.$$

Hence, $c \in C^\infty$ and since C is compact, $c \in l_\infty$ with the sup norm.

A continuation consumption allocation from node s^t or $s^t \in S^t$ and $t > 0$, denoted with $c|s^t$, is a plan specifying consumption for every ex ante type at each continuation node from s^t:

$$c|s^t = \left\{c(s^d) : s^d \in S^d|s^t, d \geq t\right\}.$$

Labor in efficiency units is denoted $l_i\left(z^t, \theta^t\right) \in \left[0, \bar{l}\right]$. Following the notational convention for consumption, similarly we can define $l\left(s^t\right)$, L, l, and $l|s^t$. It will be useful to collect the consumption and labor allocation in one variable, which we will refer to as flow allocation:

$$x_i\left(z^t, \theta^t\right) = \left\{c_i\left(z^t, \theta^t\right), l_i\left(z^t, \theta^t\right)\right\}.$$

We let $X = C \times L$ and follow the notational convention to define $x\left(s^t\right)$, x, and $x|s^t$.

[15] See Theorem 11.4 in Stokey and Lucas (1996). Our results also hold for Markov processes with absorbing states and multiple ergodic distributions. We adopt this more restrictive formulation for ease of exposition.

Let $k\left(z^{t-1}\right) \in \left[0, \bar{k}\right] \equiv K$ denote the amount of aggregate capital available at node $z^t \in Z^t | z^{t-1}$. An allocation for aggregate capital is then given by:

$$k = \left\{k\left(z^t\right) : z^t \in Z^t, t \geq 0\right\},$$

for given initial value, k_{-1}. The allocation for aggregate capital, k, belongs to $K^\infty \in l_\infty$. A continuation allocation for aggregate capital will be denoted with $k|z^t$.

The allocation for individual capital is denoted with y. Specifically, let $y_i(s^{t-1}) \in [0, \bar{y}]$ be the amount of capital held by ex-ante type i at node s^{t-1}, and

$$y(s^{t-1}) = \left\{y_i(s^{t-1}) \in [0, \bar{y}] : i \in I\right\},$$

be the distribution of capital across ex-ante agents at node s^{t-1}. The initial distribution of individual capital, y_{-1}, is taken as given. The allocation of individual capital is:

$$y = \left\{y\left(z^t\right) : z^t \in Z^t, t \geq 0\right\}.$$

Definition 1 *An allocation ψ is a triple $\{x, y, k\} \in X^\infty \times Y^\infty \times K^\infty$. Let Ψ be the set of all such allocations.*

Given an allocation ψ, we can construct a corresponding sequence of probability measures $\{\mu_t\}_{t=0}^\infty$ over the Borel sets of $X^\infty \times Y^\infty \times K^\infty$. We will use the notation $\mu_t \to \mu_\infty$ to indicate that the sequence $\{\mu_t\}_{t=0}^\infty$ converges weakly to the probability measure μ_∞, and we will refer to μ_∞ as the limiting measure.[16]

Output is produced by combining labor and capital according to a constant-returns-to-scale production function F. The resource constraint at node z^t is

$$\mathbf{c}\left(z^t\right) + k\left(z^t\right) + g\left(z_t\right) \leq F\left(k\left(z^{t-1}\right), \mathbf{l}\left(z^t\right); z_t\right) \tag{6}$$

where

$$\begin{aligned} \mathbf{c}\left(z^t\right) &= \sum_{i \in I} \sum_{\theta^t \in \Theta^t} \pi_i \pi_\theta\left(\theta^t\right) c_i\left(s^t\right), \\ \mathbf{l}\left(z^t\right) &= \sum_{i \in I} \sum_{\theta^t \in \Theta^t} \pi_i \pi_\theta\left(\theta^t\right) l_i\left(s^t\right), \end{aligned}$$

are aggregate consumption and labor respectively, and $g : Z \to [0, \bar{g}]$ is an exogenous process for government expenditures. The production function includes capital depreciation and satisfies standard properties, as detailed in Assumption 2.

[16] A sequence $\{\mu_t\}_{t=0}^\infty$ converges weakly to a probability measure μ_∞ if $\lim_{t\to\infty} \mu_t(B) = \mu_\infty(B)$ for all Borel sets B of $K \times X \times Y$ with $\mu_\infty(\partial B) = 0$. Equivalently, $\lim_{t\to\infty} \int f d\mu_t = \int f d\mu_\infty$ for all continuous, bounded functions $f : K \times X \times Y \to \Re$.

Assumption 2 *The production function F is homogeneous of degree one, twice differentiable, concave and strictly increasing in labor and capital.*

Feasibility also requires that the distribution of capital across agents be consistent with aggregate capital:

$$\sum_{i\in I}\pi_i\left\{\sum_{\theta^t\in\Theta^t}\pi_\theta\left(\theta^t\right)y_i\left(s^t\right)\right\}\leq k\left(z^t\right). \tag{7}$$

Constraint (7) can be slack, to allow for a fraction of aggregate capital to be unassigned to any individual agent.

Definition 2 *An allocation $\psi\in\Psi$ is feasible if the resource constraint (6) and the capital allocation constraint (7) are satisfied for all $z^t\in Z^t$, $t\geq 0$. Let $\Psi^F\subset\Psi$ be the set of feasible allocations.*

For continuation allocations, feasibility must be defined from a particular state $\{s^t, k(z^{t-1})\}$. A continuation allocation $\psi|s^t$ is feasible from state $\{s^t, k(z^{t-1})\}$, if the resource constraint (6) and the capital allocation constraint (7) are satisfied for all $z^d\in Z^d|z^t$, $d\geq t$, for given $k(z^{t-1})\in K$.

Individual preferences over flow allocations are represented by :

$$U_i\left(x;z_0\right)=\sum_{t=0}^{\infty}\sum_{s^t\in S^t}\beta^t\pi\left(s^t\right)u_i\left(x_i\left(s^t\right);\theta_t\right)$$

where $\beta\in(0,1)$ is the intertemporal discount factor and per-period utility function u_i satisfies the standard properties. While this formulation allows for heterogeneity in preferences, all agents are restricted to share the same discount factor.

Assumption 3 *For all $i\in I$ the utility function $u_i:\left[0,\bar{c}\right]\times\left[0,\bar{l}\right]\times\Theta\rightarrow\Re$ is concave, twice differentiable, strictly increasing in consumption, and decreasing in labor.*

The social welfare function is represented by:

$$U\left(x;z_0\right)=\sum_{i\in I}\pi_iU_i\left(x;z_0\right),$$

with $U:X^\infty\rightarrow\Re$. This formulation can accommodate arbitrary Pareto weights over ex-ante types I, which could be captured in the definition of $U_i(\cdot)$.

We can now provide a formal definition of a first best allocation.

Definition 3 *A feasible allocation $\psi^{fb}\in\Psi^F$ is* first best *given initial conditions k_0 and z_0, if for all feasible allocations $\psi\in\Psi^F$:*

$$U\left(x^{fb};z_0\right)\geq U\left(x;z_0\right).$$

A continuation first best allocation from state $\{s^t, k(z^{t-1})\}$ for $s^t \in S^t$ and $k(z^{t-1}) \in K$, denoted with $\psi^{fb}|\{s^t, k(z^{t-1}\}$, is a feasible continuation allocation from $\left\{s^t, k\left(z^{t-1}\right)\right\}$ that satisfies:

$$U(x^{fb}|\left\{s^t, k\left(z^{t-1}\right)\right\}, z_t) \geq U(x|\left\{s^t, k\left(z^{t-1}\right)\right\}, z_t), \tag{8}$$

for all feasible continuation allocations $x|\left\{s^t, k\left(z^{t-1}\right)\right\}$.

The distribution of capital across agents is indeterminate in the first best, since it is inconsequential for welfare or production efficiency under our assumptions. We can thus define the sequence of probability measures associated with the first best allocation, $\left\{\mu_t^{fb}\right\}$, and its limit μ_∞^{fb} over $X \times K$, with the understanding that for any first best allocation, ψ^{fb}, the associated sequence of probability measures over $X \times K$ converges weakly to μ_∞^{fb}. This also implies that for any first best continuation allocation, $\psi^{fb}|s^t$, the associated sequence of probability measures converges to μ_∞^{fb}.

Our setup is limited to infinite-lived households, leaving overlapping generation economies out. These economies often display dynamic inefficiencies, possibly providing a rationale for permanent intertemporal distortions.[17] In addition, shifting distortions over time entails a transfer across generations and thus have pure redistributive effects. An analysis of intertemporal distortions in overlapping generation economies would be very interesting to pursue, but it is beyond the scope of this paper.

3.2 Admissibility constraints

For the first best allocation, feasibility is the only constraint on the allocation of resources. Most interesting economic environments feature additional constraints, as discussed in Section 2, which we refer to as *admissibility constraints*. We now introduce a general formulation for the admissibility constraints designed to capture a large class of second best economies that have been examined in the literature.

Our general formulation comprises two classes of constraints. First, there are constraints directly on the allocation, possibly at each date and state and for each ex ante type of agent. These are parametrized by an auxiliary variable, which does not enter preferences or production but which may be necessary to formulate the second best problem, like debt in a Ramsey economy with incomplete markets. The auxiliary variable is endogenous and may be subject to initial conditions as well as constraints on its law of motion. The latter make up the second set of constraints covered by our general formulation. We proceed by presenting the formal description of each class of admissibility constraints first, followed by a more detailed discussion of the resulting structure of the second best problem.

Specifically, at each date $t \geq 0$, for each aggregate shock history $z^t \in Z^t$, each ex-post type $\theta^t \in \Theta^t$, and each ex-ante type $i \in I$, there are $\{1, 2, ..., M\}$ admissibility constraints on the

[17] See Erosa and Gervais (2002), Garriga (2003), and Conesa, Kitao, and Krueger (2009) for recent work on overlapping generation Ramsey models where it may be optimal to tax capital.

allocation, where M is a finite number. We represent constraint $\{m, i\}$ at node s^t with:

$$H_{m,i}\left(\psi, s^t\right) \leq a_{m,i}\left(s^t\right), \tag{9}$$

where $a_{m,i}\left(s^t\right)$ is the value of the auxiliary variable, a, that parametrizes the constraint m, i at node s^t.

The function $H_{m,i}\left(\psi, s^t\right)$ is given by:

$$H_{m,i}\left(\psi, s^t\right) = \mathbf{b}_{m,i}\left(\psi, s^t\right) + \mathbf{h}^0_{m,i}\left(\psi, s^t\right) + \sum_{\tilde{s}^t \in S^t | s^{t-1}} \mathbf{h}^1_{m,i}\left(\psi, s^t; \tilde{s}^t\right), \tag{10}$$

where

$$\begin{aligned}
\mathbf{b}_{m,i}\left(\psi, s^t\right) &= b_{m,i}\left(x\left(s^t\right), y\left(s^{t-1}\right); k\left(z^{t-1}\right), s_t\right) \\
\mathbf{h}^0_{m,i}\left(\psi, s^t\right) &= d^0_{m,i}\left(x\left(s^t\right); s_t\right) \sum_{j=t+1}^{\infty} \sum_{s^j \in S^j | s^t} \beta^{j-t} \pi\left(s^j | s_t\right) h^0_{m,i}\left(x_i\left(s^j\right); s_j\right), \\
\mathbf{h}^1_{m,i}\left(\psi, s^t; \tilde{s}^t\right) &= d^1_{m,i}\left(x\left(s^t\right); s_t\right) \sum_{j=t+1}^{\infty} \sum_{\tilde{s}^j \in S^j | \tilde{s}^t} \beta^{j-t} \pi\left(s^j | s_t\right) h^1_{m,i}\left(x_i\left(\tilde{s}^j\right); s_j\right),
\end{aligned}$$

with

$$\begin{aligned}
b_{m,i} &: [0, \bar{x}]^I \times [0, \bar{y}]^I \times K \times S \to \Re, \\
h^j_{m,i} &: [0, \bar{x}] \times S \to \Re, \\
d^j_{m,i} &: [0, \bar{x}] \times S \to \Re,
\end{aligned}$$

for $j \in \{0, 1\}$. For convenience, we stack each value of $H_{m,i}\left(\psi; s^t\right)$ into $H_i\left(\psi; s^t\right)$ and then into $H\left(\psi; s^t\right)$. We assume that each component function of H is twice differentiable.

Assumption 4 *Functions* $\left\{b_{m,i}, h^j_{m,i}, d^j_{m,i}\right\}$ *are twice differentiable in* ψ *for all* $i \in I$, $m \in M$.

We offer a detailed discussion of each of the terms in (10) below. To complete the description of the admissibility constraints, we define the vector:

$$a\left(s^t\right) = \left\{a_{m,i}\left(s^t\right) \in [0, \bar{a}] : i \in I, m \in M\right\},$$

which collects the relevant values of the auxiliary variable for node $s^t \in S^t$, $t \geq 0$. Let $A = [0, \bar{a}]^{|I \times M|} \subset \Re^{|I \times M|}$ be the space for $a\left(s^t\right) \in A$. A plan for the auxiliary variable is given by:

$$a = \left\{a\left(s^t\right) : s^t \in S^t, t \geq 0\right\} \in A^{\infty}.$$

The auxiliary variable captures additional endogenous state variables that do not enter production or preferences but are necessary for defining the second best allocation problem. The

implementability constraints in the Ramsey model with incomplete markets in equation (2) are one example of this use. In this case, the auxiliary variable can be defined as a vector comprising values of b_{t-1} and V_t. It can also capture time, state and agent variation in the admissibility constraints (9). The path of the auxiliary variable may itself be subject to constraints, stemming from restrictions on trading, fiscal instruments, and so on. We formalize these as follows. We say that a plan for the auxiliary variable is admissible if for all $z^t \in Z^t$, $\theta^t \in \Theta^t$,

$$a\left(s^t\right) \in \Gamma\left(a\left(s^{t-1}\right); s_t\right) \tag{11}$$

with $a_0 \in A_0 \subseteq A$ at date $t = 0$.

The correspondence $\Gamma : A \times S \rightrightarrows A$ governs the law of motion for the auxiliary variable. We restrict Γ to be convex. Since for most second best economies the auxiliary variable will correspond to assets with linear laws of motion, the condition will be satisfied for most problems of interest.

Assumption 5 *The correspondence $\Gamma(\alpha, s) : A \times S \rightrightarrows A$ is continuous, convex, and its image is a convex subset of $\Re^{I \times M}$ including α for all $s \in S$ and $\alpha \in A$.*

We can now formally define the notion of a second best allocation for our economy.

Definition 4 *A feasible allocation $\psi \in \Psi^F$ is admissible if there exists a plan for the auxiliary variable $a \in A$, with $a_0 \in A_0$, such that for all dates $t \geq 0$, all $s^t \in S^t$: (i) the allocation, ψ, satisfies (9) for $m = 1, 2, ..., M$ and $i = 1, 2, ..., I$; (ii) the plan for the auxiliary variable, a, satisfies (11). Let Ψ^A be the set of admissible allocations.*

Definition 5 *A second best allocation is an admissible allocation ψ^* given initial conditions k_0 and z_0, if for all admissible allocations $\psi \in \Psi^A$:*

$$U(x^*; z_0) \geq U(x; z_0).$$

If the first best allocation ψ^{fb} is admissible, the admissibility constraints will never be binding and the second best problem is trivial. To distinguish this case, we introduce the concept of proper second best. This corresponds to a second best allocation that satisfies: $U\left(x^*; z^0\right) < U\left(x^{fb}; z^0\right)$.

We now discuss the formulation for the admissibility constraints (9).[18] They comprise three terms, designed to accommodate the various constraints found in the literature. The term $\mathbf{b}_{m,i}$ depends on the state of the economy $\left\{k\left(z^{t-1}\right), s_t\right\}$, the agent's decisions and capital holdings, $\left\{x(s^t), y\left(s^{t-1}\right)\right\}$. This term typically captures the value of an outside option, which may depend on the agent's allocation of capital. Outside options show up in participation constraints, as well as limited-commitment or self-enforcement constraints on the government or private agents. The term $\mathbf{h}^0_{m,i}$ is purely forward-looking, since it includes only the continuation allocation for consumption and

[18] The Web Appendix contains details on how to map several well-known economies into our framework. These examples can be helpful to understand the rationale behind our formulation.

labor. This term is common to many second-best economy constraints, such as the implementability constraints in Ramsey taxation models or the value of an allocation on the equilibrium path for a setting with participation conditions. Finally, the term in $\mathbf{h}^1_{m,i}$ allows us to incorporate constraints across allocations awarded to different ex post types of agents. Such restrictions arise in settings with constraints on redistribution or risk-sharing. Implementability constraints in private information economies are a typical example, as discussed in Section 2.2.2.[19] The terms $d^j_{m,i}$ allow us to capture endogenous discount factors that depend on allocations at the present date. These terms often appear in intertemporal budget constraints.

The formulation of $H\left(\psi, s^t\right)$ imposes some limits on the degree of history dependence in the constraints, a structure common to all the second best economies that we encompass. First, the admissibility constraint at each node cannot be arbitrarily backward-looking. Second, the value of endogenous state variables *at future dates* does not enter the admissibility constraint at the current date. Finally, the forward looking terms $\mathbf{h}^0_{m,i}$ and $\mathbf{h}^1_{m,i}$ are time-separable and discounted at the same rate as utility. These assumptions rule out some economies in the literature. For example, a Ramsey taxation model with the constraint that taxes must be constant over time cannot be captured in (9).[20] Some political economy models also consider different discount rates for the government and private agents

The specification for the constraints on the law of motion of the auxiliary variable, (11), is similarly motivated. The correspondence Γ can capture constraints on individual or aggregate asset positions, or on their variation over time, as well as certain information constraints, such as measurability conditions. The implementability constraint in a Ramsey model is a typical example, see (1) and (2) in Section 2.1. As before there are some limitations embedded in our formulation. First, the auxiliary variable must be multiplicatively separable with the terms in H. Second, the definition of Γ rules out arbitrary history-dependent constraints.[21]

Finally, we introduce some additional assumptions that enable us to use Lagrangian methods, rule out local maxima and inflection points, and give formal content to our assertions regarding the long run. We state these assumptions in terms of the second best allocation to preserve generality, since imposing similar assumptions on the primitives of the economy would leave us with much more restrictive environment. We briefly discuss the rationale for each assumption, and its relationship to the assumptions commonly found in the literature.

We begin by imposing a standard non-degenerate constraint qualification condition, which is

[19] The baseline formulation of $\mathbf{h}^j_{m,i}$ only allows for one period cross-agent restrictions on the allocation, namely, deviations $\tilde{s}^t \in S^t | s^{t-1}$ are followed by truth-telling, $\tilde{s}^j = \left\{\tilde{s}^t, s_{t+1}, s_{t+2}, \ldots, s_j\right\}$. However, the formulation can easily if cumbersomely be extended to include a finite number of terms $\mathbf{h}^j_{m,i}$.

[20] The restrictions on the structure of (9) do not rule out Ramsey models with incomplete factor taxation, such as Correia (1996) and Jones, Manuelli and Rossi (1997). These restrictions can be formulated with an admissibility constraint as (9), by including an additional constraint at date $t = 0$ that prevents the government from manipulating the present value of assets at date $t = 0$. See Armenter (2008) for a discussion.

[21] Both limitations are not very restrictive. First, the formulation for H is very flexible, allowing us to manipulate the original set of constraints to achieve the multiplicatively separable specification, if needed. Second, the correspondence Γ can easily be extended to include a finite number lags of the auxiliary variable.

necessary to apply Lagrangian methods.[22] We also assume that the second best allocation is interior. This excludes constrained-efficient allocations with corner solutions in the allocation space $X \times K \times Y$.[23]

Assumption 6 *Let ψ^* be a second best allocation. Then for all nodes $s^t \in S^t$, ψ^* is a regular point of $H_{m,i}\left(\psi, s^t\right)$ for all $i \in I, m \in M$, and belongs to the interior of $X \times K \times Y$.*

The next assumption is a regularity condition that rules out local maxima and inflection points. As it is well known, the set of admissible allocations is not convex for many second best problems. The strategy for resolving this issue varies by application. Often, sufficiency of the first-order conditions is verified numerically or strong conditions on primitives are imposed.[24] The latter are always specific to the application at hand, and cannot be translated into other environments. This poses a challenge for a general theory of second best problems.

The condition we impose is not overly restrictive and allows us to include common non-convex problems.

Assumption 7 *Let ψ^{fb} and ψ^* be a first-best and second-best allocation respectively. For each $i \in I$ and $m \in M$ let*

$$\begin{aligned} \kappa_{m,i}^{fb} &= \lim_{t\to\infty} \sup \left\{ H_{m,i}\left(\psi^{fb}, s^j\right) : s^j \in S^j, j \geq t \right\}, \\ \kappa_{m,i}^{*} &= \lim_{t\to\infty} \sup \left\{ H_{m,i}\left(\psi^{*}, s^j\right) : s^j \in S^j, j \geq t \right\}. \end{aligned}$$

Then:

$$\kappa_{m,i}^{*} \leq \kappa_{m,i}^{fb},$$

with strict equality only if $\mu_\infty^* = \mu_\infty^{fb}$.

Assumption 7 is very similar to the condition used in Ray (2002), who studies a class of a dynamic principal-agent models with limited commitment.[25] Ray (2002) imposes that agreements that lie within the Pareto-frontier, and are thus globally improvable, can also be improved locally. We have to modify this condition, since Ray's environment, which excludes stochastic shocks and physical state variables, is considerably more restrictive. Moreover, since we concentrate on the limiting behavior of second best allocations, we can impose a less stringent form of Ray's condition, expressed in terms of the lim sup of the admissibility constraints. Assumption 7 only requires that,

[22] See Luenberger (1997) for definition of a regular point and the associated theorems with the Lagrangian.

[23] Assuming that the second best allocation is interior allows us to maintain the analysis at its most general level. A typical approach to guarantee interiority in related applications is to impose Inada conditions on the primitive functional forms. However, Inada conditions rule out many interesting environments that we seek to encompass, for example, those with risk neutral agents. Interiority does indeed rule out some interesting cases, such as certain private information economies that display bunching. See Tuomala (1990) and Salanie (2003) for a discussion.

[24] For example, in private information models with labor choice, a *single-crossing* condition is imposed to ensure an ordering of the marginal rate of substitution between consumption and labor by types. In Ramsey models, a more typical approach is to guess-and-verify.

[25] Assumption A2, page 556 in Ray (2002).

if the second best allocation can be improved upon globally, then it must be possible to eventually improve upon it by a small amount.[26]

Our final assumption imposes that the second best allocation converges weakly. This restriction is required to provide formal content to our statements on the limiting properties of the optimal allocation. As it is well known, it is hard to identify sufficient conditions for convergence in many second best economies, and stringent restrictions are required to prove convergence, so that this property is often simply assumed. This difficulty is compounded in a general framework like ours. Thus, we follow most of the literature and assume that the second best allocation converges.[27]

Assumption 8 *Let $\{\mu_t^*\}$ be a sequence of probability measures associated with a second best allocation ψ^*. The sequence $\{\mu_t^*\}$ converges weakly to μ_∞^*.*

Clearly, it would be preferable to guarantee that the second best allocation satisfies Assumptions 6-8 by imposing restrictions on the primitives. Formulating these restrictions in terms of the second best allocation directly, however, has the advantage of preserving the generality of the framework. While Assumptions 6-8 may be hard to verify, in specific applications, it may be easier to identify conditions on the primitives that deliver the required property.[28]

4 Main Result

Our main result, derived in Theorem 1, identifies a sufficient condition that rules out permanent intertemporal distortions in the class of second best economies encompassed by our framework. The condition identifies a property of the set of the admissible allocations, thus it can be verified without solving for the second best allocation. It is also simple to state, given the generality of the environment.

Condition 1 *There exists an admissible allocation $\tilde{\psi}$ such that the associated sequence of probability of measures $\{\tilde{\mu}_t\}$ converges weakly to μ_∞^{fb} for any set of initial conditions A_0.*

Condition 1 formalizes the idea that it is possible to front-load all distortions for some second best problems. Clearly, if there exists an admissible allocation that converges to the first best limiting allocation, for such an allocation all admissibility constraints must eventually be non-binding. The existence of such an allocation is a property of the set of admissible allocations and does not imply

[26] Another point of distinction is that Ray's (2002) condition is imposed on all allocations, while we only require the restriction to hold for the second best allocation. Both formulation of the conditions can be difficult to verify in practice.

[27] For example, Judd (1985) presents his findings on optimal capital taxes in the steady state conditional on convergence of the second best allocation. Private information economies are typically not stationary. See Green (1987), Thomas and Worral (1990), Atkeson and Lucas (1992) and Phelan (1998). However, Atkeson and Lucas (1995) show that imposing a lower bound on utility guarantees convergence in a private information economy, without altering the main properties of the second-best allocation. Phelan (2006), Sleet and Yeltekin (2006), and Farhi and Werning (2006) study private information economies in which a lower bound on continuation utilities arises endogenously.

[28] The Web Appendix discusses the standard assumptions for each model class regarding interiority, non-convexity, and convergence.

that this allocation solves the second best problem. However, Theorem 1 shows that if Condition 1 holds, the second best allocation will not feature permanent intertemporal distortions.

We formalize the concept of intertemporal distortion as a wedge in the Euler equation for consumption. For a given an allocation ψ, let:

$$\Delta_i\left(s^t\right) = u_i^c\left(s^t\right) - \beta \sum \pi\left(s^{t+1}|s_t\right) u_i^c\left(s^{t+1}\right) F^k\left(s^{t+1}\right),$$

for all $i \in I$ and all $s^t \in S^t$, $t \geq 0$, where u^c and F^k denote the derivative of u and F with respect to c and k respectively. At the first best allocation, ψ^{fb}, $\Delta_i\left(s^t\right) = 0$ for all agents and all nodes. For a second best plan ψ^* we characterize only the limiting properties of the wedge $\Delta_i\left(s^t\right)$.

Abusing notation slightly, we extend the probability measure μ_∞^* over events of the type $\Delta_i\left(s^t\right) \geq 0$ as follows. Consider the function:

$$f : [0, \bar{x}] \times \left[0, \bar{k}\right] \times S \times X^{\#Z} \to \Re,$$

such that $f\left(x_i\left(s^t\right), k\left(s^t\right), ...\right) = \Delta_i\left(s^t\right)$. Let Υ_i the subset of the domain of f such that $f \geq 0$. By Assumptions 2 and 3, f is continuous so Υ_i is measurable. We then use notation $\mu_\infty\left(\Delta_i\left(s^t\right) \geq 0\right)$ for $\mu_\infty\left(\Upsilon_i\right)$.[29]

We can now state our main result.

Theorem 1 *Let ψ^* be a second best allocation and μ^* be the associated sequence of probability measures. If Condition 1 holds, then for each $i \in I$ either the intertemporal distortion is zero on the limit with probability one,*

$$\mu_\infty^*\left(\Delta_i\left(s^t\right) = 0\right) = 1, \tag{12}$$

or the intertemporal distortion fluctuates around zero,

$$\begin{aligned} \mu_\infty^*\left(\Delta_i\left(s^t\right) > 0\right) &> 0, \\ \mu_\infty^*\left(\Delta_i\left(s^t\right) < 0\right) &> 0. \end{aligned} \tag{13}$$

The statement of Theorem 1 clarifies in what sense all intertemporal distortions are temporary in the second best. While there may be a strictly positive or negative wedge along the transition path, Theorem 1 implies that the wedge either converges to zero with probability one in the limit or alternates in sign with positive probability. Thus, at the limiting second best allocation, the intertemporal wedge is never strictly positive or strictly negative with probability one. For economies without aggregate shocks, Theorem 1 implies that $\beta F_{ss}^k = 1$ at the steady state.

We prove Theorem 1 by decomposing the second best problem in two stages. In the first stage, we solve for the best admissible allocation taking as given the plan for the auxiliary variable. The

[29] Weak convergence $\mu_t \to \mu_\infty$ does not guarantee that $\lim_{t\to\infty} \mu_t\left(\Upsilon_i\right) = \mu_\infty\left(\Upsilon_i\right)$, so we formulate the proof in terms of $\left\{\Delta_i\left(s^t\right) : s^t \in S^t, t \geq 0\right\}$.

second stage characterizes the optimal path for the auxiliary variable. Formally, let $\Psi^A(a) \subset \Psi^F$ be the set of feasible allocations such that the admissibility constraints (9) are satisfied at all nodes $s^t \in S^t$, $t \geq 0$ for given $a \in A^\infty$.

Definition 6 *Let $W(a) : A^\infty \to \Re$ be given by*

$$W(a) = \sup \left\{ U(x; s_0) : \psi \in \Psi^A(a) \right\}. \tag{14}$$

By Assumption 3, the utility function u_i is bounded above and below for all $i \in I$ as it is differentiable over a compact set. Thus $W(a)$ also is bounded above and below for all auxiliary variable plans.[30] Clearly, if an admissible allocation ψ^* is second best then there exists $a^* \in A^\infty$ such that

$$W(a^*) = U(x^*; s_0)$$

and $\{\psi^*, a^*\}$ satisfy the admissibility constraints.

This decomposition of the second best problem imparts an intuitive structure to the proof of Theorem 1 and makes interpreting the results easier. In the first stage, we show that if all future admissibility constraints are non-binding, then all intertemporal distortions must be temporary. This result, derived in Lemma 1 , establishes a link between permanent intertemporal distortions and the need to relax future binding admissibility constraints. Lemma 1 is very general, since the proof is based on a simple variational argument. Condition 1 is not used in this step of the proof.

The second stage characterizes the optimal path for the auxiliary variable and relies critically on Condition 1. Specifically, the sufficient condition connects the ability to front-load all distortions, which is a property of the set of admissible allocations, with the limiting behavior of the admissibility constraints at the optimum. Assumptions 5 and 7 also play an important role in the argument.

4.1 First stage: Choosing allocations

We start by analyzing the first-stage problem (14) given the plan a^* for the auxiliary variable. Assumption 6 allows us to write the Lagrangian associated with the solution to (14) and derive the set of first order necessary conditions (f.o.n.c. henceforth). If there were no admissibility constraints at all, the f.o.n.c. would include the familiar Euler equation featuring a zero intertemporal wedge, $\Delta_i(s^t) = 0$. In other words, if there are no constraints on the allocation of resources, the first order welfare effect of an additional unit of resources transferred from date t to date $t+1$ at rate $F_{k,t+1}$ will have no first-order welfare effect. If the second best allocation displays a wedge, $\Delta_i(s^t) \neq 0$, transferring resources across periods tightens or relaxes future admissibility constraints, leading to a first order effect on welfare.

A key insight from the analysis of the first-stage problem is that the presence of intertemporal distortions is driven only by future binding admissibility constraints. Instead, current and past binding admissibility constraints, instead, are responsible for distortions in intratemporal margins,

[30] Technically, $\Psi^A(a)$ may be empty and then $W(a) = \varnothing$ but this case is clearly of no interest.

such as labor supply decisions. At the limiting second best allocation, the welfare effects associated with intratemporal distortions are, to a first-order approximation, symmetric across periods and do not generate a first-order welfare effect in connection to resource transfers across periods. Thus, we can rule out permanent intertemporal distortions if the admissibility constraints are not binding at future dates.

To formalize this reasoning, we define for each node a set of continuation histories at which all admissibility constrains are not binding. Given an allocation and auxiliary variable plan $\{\psi, a\}$, let

$$\varpi_{m,i,t} = \left\{ s^t \in S^t | \sup \left\{ H_{m,i}\left(s^j\right) - a_{m,i}\left(s^j\right) : s^j \in S^j | s^t, j \geq t \right\} < 0 \right\} \tag{15}$$

and $\varpi_{m,t} = \cap_I \varpi_{i,m,t}$, $\varpi_t = \cap_M \varpi_{m,t}$ for all dates $t \geq 0$.[31] The set ϖ_t collects all the nodes of S^t such that no admissibility constraint is binding for all respective continuation nodes. The set ϖ_t is a subset of S^t with the property that $\varpi_t \subseteq \varpi_{t+1}$ at all dates. This property ensures that the probability sequence $\left\{\pi\left(s^t \in \varpi_t\right)\right\}$ is weakly increasing and converges.

We can now formally state the main result for the first stage of the proof.

Lemma 1 *Let ψ^* be a second best allocation, μ^* be the associated sequence of probability measures, and a^* an auxiliary variable plan such that $\{\psi^*, a^*\}$ is admissible. If:*

$$\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1,$$

then for each $i \in I$ either the intertemporal distortion is zero on the limit with probability one,

$$\mu^*_\infty\left(\Delta_i\left(s^t\right) = 0\right) = 1,$$

or the intertemporal distortion fluctuates around zero,

$$\begin{aligned} \mu^*_\infty\left(\Delta_i\left(s^t\right) > 0\right) &> 0, \\ \mu^*_\infty\left(\Delta_i\left(s^t\right) < 0\right) &> 0. \end{aligned}$$

Proof. In the Appendix. ■

We will use a simplified model to illustrate the reasoning behind Lemma 1. Consider a second best economy with a representative agent, no uncertainty, and a single admissibility constraint for each date $t \geq 0$, given by:

$$b(k_t) + \sum_{j=t}^{\infty} \beta^{j-t} h\left(c_j, l_j\right) \leq a_t. \tag{16}$$

For this step of the proof, as in problem (14), the path for the auxiliary variable $\{a_t\}$ is taken as given. We will analyze the choice of the plan for the auxiliary variable in Section 4.2.

[31] For ease of notation, we maintain the dependence of the set on the allocation ψ and the auxiliary variable plan a implicit.

To see why only future admissibility constraints distort the intertemporal margin, consider the first order welfare effect of transferring resources across periods. First, a marginal change in consumption at date t may tighten certain admissibility constraints. To see this, assume $h_t^c > 0$ and consider the marginal value of an extra unit of consumption at date t:

$$u_t^c = \frac{\lambda_t}{1 + \frac{h_t^c}{u_t^c}\Phi_t} \leq \lambda_t \tag{17}$$

where λ_t, λ_{t+1}, and ϕ_{t+1} are the Lagrangian multipliers associated to the resource constraint and the admissibility constraint (16) at t. The variable $\Phi_t = \sum_{j=0}^{t} \beta^{t-j}\phi_j$ collects the Lagrangian multipliers of all the admissibility constraints that feature the time t allocation $\{c_t, l_t\}$. If the current or any past admissibility constraint are binding $\Phi_t > 0$, there is a wedge between the marginal value of resources and the marginal utility of consumption. The wedge in (17) reflects the degree to which the allocation is distorted at that date. If the wedge at date $t+1$ is greater than the wedge at t,

$$\frac{h_t^c}{u_t^c}\Phi_t < \frac{h_{t+1}^c}{u_{t+1}^c}\Phi_{t+1}, \tag{18}$$

a transfer of consumption from t to $t+1$ will induce a first order welfare loss. At the steady state, this inequality will hold if $\Phi_t < \Phi_{t+1}$, that is, if the admissibility constraint at $t+1$ is binding.

In addition, changes in the capital stock may directly affect the admissibility constraint at date $t+1$. If the admissibility constraint is binding, there will be a wedge in the Euler equation for capital:

$$\lambda_t + \phi_{t+1} b_{t+1}^k = \beta \lambda_{t+1} F_{t+1}^k.$$

If $b_{t+1}^k > 0$, additional capital tightens the admissibility constraint at $t+1$ and the wedge is positive.

Now consider a plan for the auxiliary variable such that the admissibility constraints from date d onwards are *not* binding, so $\phi_t = 0$ for all $t \geq d$. Evaluated at the steady-state, the wedge (17) between the marginal utility of consumption and the marginal value of resources reduces to:

$$u_{ss}^c = \frac{\lambda_{ss}}{1 + \frac{h_{ss}^c}{u_{ss}^c}\bar{\Phi}} \leq \lambda_{ss},$$

where:

$$\bar{\Phi} = \sum_{j=0}^{d} \beta^{d-j}\phi_j.$$

Once all admissibility constraints stop binding, intratemporal distortions depend on the constant value, $\bar{\Phi}$, which captures the fact that allocations at date t and $t+1$ are subject to the same set of admissibility constraints. Clearly, intratemporal distortions will persist in the long-run provided any admissibility constraint was binding at some date, since this implies $\bar{\Phi} > 0$. However, at the steady state, equation (18) holds with equality, reflecting the fact that a shift in resources across dates has an *offsetting* first-order welfare effect on welfare. In addition, marginal changes in the

level of capital cannot affect future admissibility constraints, since the constraints are not binding and the standard Euler equation holds:

$$\lambda_t = \beta \lambda_{t+1} F_{t+1}^k.$$

The implication of this argument is that a transfer of resources across two consecutive periods has no first order welfare effects at the steady state if future admissibility constraints do not bind. It follows that the marginal value of resources is equated across periods and there are no intertemporal distortions at the steady state. This implies: $\beta F_{ss}^k = 1$.

Lemma 1 characterizes the limiting behavior of the intertemporal wedge $\Delta_i\left(s^t\right)$, but it leaves open the possibility that the second best allocation features a strictly positive or negative wedge along the transition path. Since the stationary distribution may not be reached in finite time, a strictly positive or negative intertemporal wedge at all nodes would be implied. Corollary 1 shows that if the sign of the wedge $\Delta_i\left(s^t\right)$ is constant along the transition path, then the wedge is 0 with probability one in the limit.

Corollary 1 *Let ψ^* be a second best allocation, μ^* be the associated sequence of probability measures, and a^* a plan for the auxiliary variable such that $\{\psi^*, a^*\}$ is admissible. In addition, let*

$$\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1.$$

If for any $i \in I$, either: (i) $\Delta_i\left(s^t\right) > 0$ for all $s^t \in S^t, t \geq 0$, or (ii) $\Delta_i\left(s^t\right) < 0$ for all $s^t \in S^t, t \geq 0$, then:

$$\mu_\infty^*\left(\Delta_i\left(s^t\right) = 0\right) = 1.$$

Proof. In the Appendix. ■

4.2 Second stage: Choosing the auxiliary variable

We now study the optimal choice of the plan for the auxiliary variable, which will denote with a^*. It is here that we use Condition 1 to show that the admissibility constraints will be binding only for a transitional phase. The argument proceeds in two steps. We first derive two important properties of the set of admissible plans for the auxiliary variable implied by Condition 1. We then use these properties to show that the optimal plan for the auxiliary variable a^* converges to a subset of A where the admissibility constraints (9) do not bind.

Condition 1, combined with Assumption 5 and the compactness of A, can be used to derive two important properties of the set of admissible plans for the auxiliary variable. We illustrate the argument in the simplified model introduced in Section 4.1. We assume a simple formulation for the correspondence Γ:

$$\Gamma\left(a_t\right) = \left[0, \gamma\left(a_t\right)\right] \tag{19}$$

where γ is a function on $\Re_+$.

Let $\left\{x_{ss}^{fb}, k_{ss}^{fb}\right\}$ denote the steady-state under the first-best allocation. The sufficient condition is satisfied if there exists an admissible allocation such that $\left\{\hat{x}_t, \hat{k}_t\right\} \rightarrow \left\{x_{ss}^{fb}, k_{ss}^{fb}\right\}$. The allocation $\left\{\hat{x}_t, \hat{k}_t\right\}$ will, generally, not be optimal. However, its existence allows us to derive some key properties of the correspondence Γ and of the set of admissible allocations.

Since $\left\{\hat{x}_t, \hat{k}_t\right\}$ is admissible, there must exist a path for the auxiliary variable, $\{\hat{a}_t\}$, that can support the first best steady state. This requires constraint (16) to be satisfied by the first best steady state allocation to satisfy:

$$a_{ss}^{fb} \equiv b(k_{ss}^{fb}) + \frac{h\left(c_{ss}^{fb}, l_{ss}^{fb}\right)}{1-\beta} \leq \hat{a}_t,$$

for all dates $t \geq d$ for some date d. In addition, $\{\hat{a}_t\}$ will satisfy constraint (19) on the law of motion of the auxiliary variable.

The sequence $\{\hat{a}_t\}$ thus defines a subset of A, $\left\{a \geq a_{ss}^{fb}\right\}$, that supports the first-best steady state. It also identifies a path for the auxiliary variable that supports a first best continuation from any initial condition, since the admissibility constraints on the allocation, (16), and on the auxiliary variable, (19), must be satisfied at each date.

Proposition 1 formally establishes this result for the general environment.

Proposition 1 *Let:*

$$A^{fb} = \left\{\alpha \in A : \lim_{t\rightarrow\infty} \sup_{s^j \in S^j, j \geq t} \{H_{m,i}(\psi^{fb}, s^j)\} \leq \alpha_{m,i} \, \forall \, i, m\right\}, \tag{20}$$

where φ^{fb} is the first best allocation. If Condition 1 holds, A^{fb} is non-empty.

Proof. In the Appendix ■

Since A^{fb} is non-empty, then by Assumption 5 for any $\alpha^{fb} \in A^{fb}$, the continuation plan ${}_t\alpha^{fb} = \{\alpha^{fb}, \alpha^{fb}, ...\}$ satisfies $\alpha^{fb} \in \Gamma(\alpha^{fb}, s^j)$ $\forall s^j \in S^j | s^t$, for $j \geq t$. This property implies that for any $a \in A^\infty$ such that $a(s^t) \in A^{fb}$ for some s^t, then there is an admissible continuation plan for a from s^t such that every element is in A^{fb}. Thus, if A^{fb} is non-empty, the first best stationary allocation can be supported by an admissible plan for the auxiliary variable.

We can now derive the second key property of the set of admissible plans for the auxiliary variable implied by Condition 1. We will first illustrate this result in the simplified model. Here, the sufficient condition immediately rules out the possibility that for all $\alpha < a_{ss}^{fb}$ we have $\gamma(\alpha) \leq \alpha$. If this were the case, the existence of an auxiliary variable $\{\hat{a}_t\}$ such that $\hat{a}_t \geq a_{ss}^{fb}$ would be ruled out. Thus, the sufficient condition combined with Assumption 5, which requires $\Gamma(\cdot)$ to be convex and $\gamma(\alpha)$ to be concave, immediately implies $\gamma(\alpha) > \alpha$ for all $\alpha < a_{ss}^{fb}$. In other words, it is always

possible to relax the admissibility constraints on the allocation in a one step continuation of the plan for the auxiliary variable.

This property is derived formally for the general framework in Proposition 2. It connects Condition 1, a global property of the set of admissible allocations, with a local property of the second best plan for the auxiliary variable.

Proposition 2 *Let Condition 1 hold. Then, for any $\alpha \in A$, there exists a scalar $\varrho \in (0,1]$ and $\alpha' \in A^{fb}$, such that:*

$$(1-\varrho)\,\alpha + \varrho\alpha' \in \Gamma(\alpha, s), \tag{21}$$

for some $s \in S$.

Proof. In the Appendix ■

If Proposition (21) did not hold, it would be possible to find a separating hyperplane between the set $\Gamma(a,s)$ for all $\{a,s\} \in A \times S$ and A^{fb}. Then, by Assumption 5, the convexity of Γ would imply that no point in the half-space containing $\Gamma(a,s)$ could be the starting point of a path leading to A^{fb}, contradicting Proposition 1.

We now use these properties of the set of admissible plans for the auxiliary variables implied by Condition 1 to show that the second best plan converges to A^{fb}. The premise of this result is that, if at any date t an admissibility constraint is binding, relaxing such a constraint has a first order welfare gain. For the simplified model, as long as an admissibility constraint is binding at date t, the optimal path for the auxiliary variable must feature $a_t^* = \gamma\left(a_{t-1}^*\right) > a_{t-1}^*$, that is, it must be strictly increasing. Thus, either the admissibility constraint will stop binding at some finite date d or the auxiliary variable will converge to a_{ss}^{fb}. Our regularity conditions then ensure that if $a_t^* \to a_{ss}^{fb}$, the Lagrangian multiplier on the admissibility constraint ϕ_t converges to 0.

The general framework features multiple admissibility constraints. Since there is no guarantee relaxing all of these constraints simultaneously is possible, the plan for the auxiliary variable need not be monotone. However, Assumption 7 implies that an auxiliary variable plan a can always be improved upon if it is in the interior of the image of the correspondence Γ. As in Ray (2002), this condition rules out local maxima and inflection points and plays a crucial role in the argument.

Proposition 3 establishes this result formally for the general framework.

Proposition 3 *Let Condition 1 hold and ψ^* be a second best allocation such that $\mu_\infty^* \neq \mu_\infty^{fb}$. Then, there exists an auxiliary variable plan a^* such that:*

$$\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1,$$

and $\{\psi^*, a^*\}$ are admissible and ϖ_t is defined in (15).

The requirement $\mu_\infty^* \neq \mu_\infty^{fb}$ restricts the attention to proper second best allocations, for which the limiting distribution does not converge to the first best.

4.3 Combining the two stages

Finally, we combine the results from both stages to prove Theorem 1. If the sufficient condition holds, by Proposition 3 it is possible to front load all distortions. Thus, the second best plan calls for eventually relaxing all future admissibility constraints. This implies that in the limit it is possible to transfer resources across periods without raising further distortions. By Lemma 1, it follows that permanent intertemporal distortions are not compatible with optimality.

5 Applications

We now discuss a series of applications to illustrate the relevance of our result. Section 5.1 concentrates on Ramsey models and shows that the sufficient condition generally holds if the government can save enough. We also briefly discuss two Ramsey examples that adhere to the basic logic behind our result, despite the fact that Condition 1 does not hold. Section 5.2 examines economies with limited risk-sharing. In models with self-enforcement constraints, we show how the outside option and parameter values determine whether the sufficient condition holds. For private information economies, we show that typically the sufficient condition does not hold and there are permanent intertemporal distortions, yet the optimal allocation still front-loads distortions. Section 5.3 analyzes economies where the government is subject to a limited commitment constraint. We show that when the government is benevolent, Condition 1 is always satisfied and there are no permanent intertemporal distortions. For a self-interested government, we discuss the conditions under which the sufficient condition holds. Our analysis is mostly motivated by applications in the area of optimal policies, which have concentrated on the properties of intertemporal distortions. However, our environment is more general, and also encompasses additional dynamic contracting applications. In Section 5.4, we discuss several second best models of firm dynamics to illustrate how our result can be applied in that context.

5.1 Ramsey Taxation

The Ramsey taxation model rules out lump sum taxes. This restriction on fiscal instruments prevents the economy from attaining the first best allocation. But even without lump sum taxes, the first best steady state allocation can be attained if the government can save. By accumulating assets, the government eventually can finance all spending with interest revenues and forgo distortionary taxes. While following this policy is typically not optimal, it does satisfy the admissibility constraints that define the Ramsey equilibrium. Hence, the sufficient condition holds and we can apply Theorem 1. We now discuss a series of specific Ramsey models in order to show the link between the government's ability to front-load distortions and permanent intertemporal distortions.

We begin with the basic Ramsey model with a representative agent and no uncertainty. The implementability constraint (1) is the only admissibility constraint, as discussed in Section 2.1. The sufficient condition holds if there exists an admissible allocation that converges to the first best

steady state. This implies in turn that there exists a path for the auxiliary variable such that a first best continuation allocation is admissible from some date onwards.

Let $\bar{a}(k_t)$ denote the minimum value of the auxiliary variable that supports a first-best continuation allocation from date t for $k_t \in K$. Evaluating constraint (1) at the first-best continuation allocation on any given date t we obtain:

$$(u_t^c)^{-1} \sum_{j=d}^{\infty} \beta^{j-t} \left(u_j^c c_j^{fb} + u_j^l l_j^{fb} \right) = \bar{a}(k_t).$$

An admissible allocation that converges to the first best steady-state is associated with a path for the auxiliary variable such that $a_t \geq \bar{a}(k_t)$ for some date t.[32] We can derive the corresponding value of government debt, $\bar{b}(k_t)$, that sustains $\bar{a}(k_t)$ from the competitive equilibrium conditions:

$$\bar{b}(k_t) = -\bar{a}(k_t) - F^k\left(k_t, l_t^{fb}\right) k_t.$$

The level $\bar{b}(k_t)$ will be negative, so a first best continuation allocation is admissible if the government can accumulate a sufficient level of assets. A policy with relatively high labor and capital income taxes for a transitional phase will lead to a level of debt such that $b_t < \bar{b}(k_t)$ at some date t, after which the first best continuation allocation from k_t is admissible. Then, there will be no permanent intertemporal distortions, based on Theorem 1. In the basic Ramsey model, this also implies that the capital income tax is zero in the steady state.

Why would a government be unable to save? Assuming a closed economy, the government can save enough only if private agents can *borrow* enough. Denote with b_t^p the debt issued by private agents. Bond market clearing requires $b_t^p + b_t = 0$ at all dates. If a borrowing constraint is imposed on private agents, $b_t^p \leq B$, this gives rise to a corresponding upper limit on government assets, $b_t \geq -B$. This restriction will generate an additional admissibility constraint.[33] If $\bar{b}(k_t) < -B$, the first-best continuation allocation cannot be supported. If $\bar{b}(k_t) < -B$ for all $k_t \in K$, then the sufficient condition is violated and the optimal allocation may feature intertemporal distortions.

Farhi (2010) analyzes a Ramsey model with a representative agent and risk-free debt, with aggregate shocks.[34] He finds that the limiting behavior of the second best allocation depends critically on the upper limit imposed on government assets. Following Aiyagari et al. (2002), debt and asset limits are labeled *natural* if they merely ensure that obligations will be paid back almost surely. More stringent limits are referred to as *ad hoc*. Under the natural asset limit, the capital tax rate is either zero or fluctuates around zero, and there are no intertemporal distortions.[35] Interpreting these findings in light of Theorem 1, the intertemporal wedge satisfies (13) and thus there are no permanent intertemporal distortions. Instead, under an ad hoc asset limit, the sufficient

[32] To derive the full path of auxiliary variables we must write the full sequence of implementability constraints.

[33] The additional constraint will shape the correspondence Γ governing the path for the auxiliary variable.

[34] We formally show how to represent the Ramsey problem in Farhi (2010) within the general framework presented in Section 3 in the Web Appendix.

[35] The properties of the intertemporal wedge vary with the assumptions on preferences and technology. Analytical results can be obtained only with quasi-linear preferences. Numerical simulations are used otherwise.

condition is not satisfied and the capital tax can be different from zero.

Taken together, these results point to a general lesson for Ramsey models. Intertemporal distortions will not arise if the government is subject to borrowing constraints, while saving constraints on the government and borrowing constraints on private agents may lead to intertemporal distortions. To further expand on this insight, we now briefly discuss two Ramsey economies in which our sufficient condition does not hold but comply with this principle.

We first consider the baseline Ramsey model under a balanced government budget constraint. Since the government cannot save, all spending must be financed solely from current tax revenues. Clearly, no admissible allocation converges to the first-best steady state if government consumption is strictly positive, and thus Condition 1 does not hold. Still, the balanced-budget Ramsey model usually does not feature permanent intertemporal distortions, provided government spending is sufficiently low. By contrast, if government spending is large enough and labor tax revenues are not sufficient to finance it, then the optimal capital tax is positive in the long run.[36]

For this environment we can relax Condition 1 while retaining the basic logic of the general result. Specifically, it can be shown that if there exists an admissible allocation such that capital converges to its first best steady state level, then there are no intertemporal distortions in the long run. As in Theorem 1, the ability to re-allocate consumption over time without violating any admissibility constraints rules out permanent intertemporal distortions at the optimum. The only difference is that for this specific environment it is sufficient to front-load some, not all, distortions. The existence of an admissible path of capital that converges to the first best steady state is only guaranteed for small enough government spending, since only in this case the admissibility constraint is satisfied at the first best level of capital.[37]

The same logic can be applied to other Ramsey models. Aiyagari (1995) provides an interesting example. The economy features only idiosyncratic risk and markets are incomplete. Agents can accumulate capital and borrow subject to the constraint that their net worth is positive, while there are no explicit borrowing or saving constraints on the government. Aiyagari (1995) shows that at the aggregate level, the condition $1 = \beta F_k^*$ holds. Thus, there are no permanent intertemporal distortions.[38] The sufficient condition is not satisfied, since the first best allocation features full consumption insurance, which is not attainable given the joint restrictions on assets and fiscal instruments. Yet, the argument behind Theorem 1 still provides an interpretation for this result. Since private agents can borrow against their capital holdings as long as they maintain a positive net worth, the government can save. Aggregate distortions can be front-loaded allowing the first

[36]See Judd (1985), Chamley (1986), as well as Stockman (2001). The exact level of government spending that triggers this result depends on the technology and preference specification. See Lansing (1999) for details.

[37]In this example, the relaxed sufficient condition is also necessary. For the stated assumptions on preferences and technology, there exists a threshold level of government spending, g^*, such that for all $g \leq g^*$, the first best steady state level of capital is admissible, and there are no intertemporal distortions. And for $g > g^*$, all admissible allocations feature intertemporal distortions in the long run. However, most models encompassed in our framework do not have a necessary and sufficient condition.

[38]In Aiyagari (1995), the Ramsey allocation does not feature permanent intertemporal distortions but is implemented with a positive capital tax. This tax offsets the distortion on the intertemporal margin arising from the precautionary savings motive.

best level of aggregate capital to be reached at the steady state, despite the incomplete insurance.

5.2 Risk-Sharing

We now discuss intertemporal distortions in economies with idiosyncratic risk in light of Theorem 1. There are many frictions leading to incomplete risk-sharing. We revisit the two paradigms introduced in Section 2, models with self-enforcement constraints and models with private information. In economies with self-enforcement constraints, assumptions on the outside option play a crucial role. Because of the variety of formulations in the literature, there are contrasting implications for intertemporal distortions. We use Theorem 1 to clarify which features of the formulation are critical for intertemporal distortions in the long run. By contrast, private information economies typically exhibit permanent intertemporal distortions. As we will discuss, even if the sufficient condition does not hold in these economies, the front-loading principle still applies.

5.2.1 Self-Enforcement Constraints

Let us return to the example of an economy with self-enforcement constraints introduced in Section 2.2.1. We now allow for a general specification of the outside option, which will involve the distribution of capital in case of default, as well as the agents' constraints after default. These assumptions shape the resulting self-enforcement constraints and determine whether the sufficient condition is satisfied.

We will consider two polar cases. The first assumes that a defaulting agent can seize a fraction $\delta \in [0,1]$ of the aggregate capital and operate it under autarchy. The resulting self-enforcement constraint is:

$$\sum_{s=t}^{\infty} \beta^{s-t} \sum_{\theta^s|\theta^t} \pi_\theta(\theta^s|\theta^t) u(c_i(\theta^s), l_i(\theta^s)/\theta_s) \geq V_{out}(\delta k_t; \theta_t), \tag{22}$$

where V_{out} satisfies:

$$V_{out}(\delta k_t; \theta_t) = \max_{\{x_s, \hat{k}_{s+1}\}_{s \geq t}} E_{\pi^s|\pi^t} \left[\sum_{s=t}^{\infty} \beta^{s-t} u(c_s, l_s/\theta_s) \right]$$

subject to

$$c_s + \hat{k}_{s+1} \leq F(\hat{k}_s, l_s)$$

with $\hat{k}_t = \delta k_t$.

The value of the outside option, V_{out}, depends only on the level of aggregate capital at the time of default. The self-enforcement constraint can be relaxed by awarding utility in future periods, which is achieved via a back-loaded path of consumption. For a utilitarian social welfare function, the value of capital on the equilibrium path is greater than in the outside option, and the self-enforcement constraint is relaxed by raising capital. If the distribution of Pareto weights is very uneven, the first best allocation may not satisfy the self-enforcement constraint, unless the discount factor β is high enough, agents' preferences exhibit a sufficient degree of risk aversion or the variance of the

idiosyncratic shock is large enough. These conditions on parameter will then determine whether Condition 1 is satisfied.[39]

The second case assumes that agents cannot accumulate capital after default or trade financial assets. Defaulting agents can still supply labor on competitive labor markets. Thus, the value of their outside option is:

$$V_{out}(\{w_s\}_{s\geq t};\theta_t) = \max_{\{x_s\}_{s\geq t}} E_{\pi^s|\pi^t}\left[\sum_{s=t+1}^{\infty} \beta^{s-t} u(c_s, l_s/\theta_s)\right] \tag{23}$$

subject to

$$c_s^i \leq w_s l_s^i,$$

where $\{w_s\}_{s\geq t}$ is the sequence of future equilibrium wages.

The value of the outside option, through its dependence on wages, now involves the entire future path of capital. The resulting admissibility constraint does not comply with the general formulation in (9). Still, the general logic behind Theorem 1continues to hold. The dependence of the value of the outside option on the future path of capital generates a pecuniary externality. Since a rise in capital increases equilibrium wages and consumption for defaulting agents, more capital in future periods need not relax the self-enforcement constraint.[40] Even so, if agents are sufficiently patient, the first best allocation may satisfy the self-enforcement constraint.[41] The impact of additional capital on incentives will depend on the degree of complementarity between labor and capital in production. This result is related to Davila, Hong, Krusell and Rios-Rull (2007).

5.2.2 Private Information

We now consider the same economy under the assumption that the idiosyncratic preference shocks, θ_t, are private information. The resulting incentive compatibility constraint requires consumption to be spread out across states, when agents are risk averse and face idiosyncratic risk in the subsequent periods. In particular, consumption at time t should be increasing in the realization θ_t if the single crossing condition holds. Thus, the first best is not incentive compatible and Condition 1 does not hold.

A well known property of this class of economies is that, if utility is separable between consump-

[39]The international risk-sharing model of Kehoe and Perri (2002, 2004) is an interesting example of a self-enforcement constraint of the form (22). The sufficient condition is satisfied in this model if countries are weighted symmetrically in the world social welfare function. Based on Theorem 1, intertemporal distortions will be temporary. Our framework also encompasses models where the agent's outside option depends on the individual capital holdings, $y(s^{t-1})$.

[40]This formulation of the outside option hinges on the implicit assumption that aggregate capital cannot be contingent on default.

[41]Chien and Lee (2008) study a Ramsey model with self-enforcement constraints that fits into this framework. They show that the limiting Ramsey equilibrium allocation does not feature intertemporal distortions, but the capital income tax is positive provided some self-enforcement constraints remain binding. As Aiyagari (1995), the capital tax is positive to equate the intertemporal marginal rate of substitution across constrained and unconstrained agents.

tion and labor, the second best allocation features a permanent positive intertemporal wedge:

$$u_c(c_i(\theta^t);\theta_t) < \beta F_k \sum_{\theta^{t+1}|\theta^t} \pi_\theta(\theta^{t+1}|\theta^t) u_c(c_i(\theta^{t+1});\theta_{t+1}), \tag{24}$$

where $u_c(c;\theta)$ denotes the marginal utility of consumption for an agent with current type θ. This inequality is strict when future incentive compatibility constraints are binding.[42]

This result can be interpreted in light of the front-loading principle that underlies the argument in Theorem 1. Consider the government's trade-off in the allocation of consumption between two consecutive periods. If the future incentive compatibility constraint is binding, there is a shadow cost of transferring consumption to future periods. This cost reflects the fact that to preserve incentives, future consumption must be spread across states and will be worth less in terms of utility. Thus, optimality requires agents' consumption to be front-loaded. The resulting intertemporal distortion will be permanent if agents are risk averse and face idiosyncratic shocks, since in this case future incentive compatibility constraints will always be binding.[43]

The front-loaded path of consumption under private information stands in contrast with the back-loaded path of consumption that arises with self-enforcement constraints.[44] Despite this difference, admissibility constraints are relaxed over time and distortions are front-loaded in both environments. In the self-enforcement economy, promising more consumption in the future relaxes the current and future admissibility constraints. By contrast, in the private information economy, awarding more consumption in the future tightens future incentive compatibility constraints, thus the optimal path of consumption is front-loaded at the optimum.

Permanent intertemporal distortions prevail more generally in private information economies. Rogerson (1985) shows that for an economy without capital, dynamic moral hazard entails a permanent intertemporal distortion. Albanesi (2011) examines optimal capital taxation in a dynamic moral hazard environment with a risk averse entrepreneur whose effort affects the distribution of capital returns. The second best allocation in this economy also front-loads distortions and the sign of the intertemporal wedge depends on whether capital tightens or relaxes the incentive compatibility constraint. While in standard private information models, the only way to provide incentives is to spread consumption across states and allocate it over time, additional instruments for providing incentives may eliminate permanent intertemporal distortions. For example, Armenter and Mertens (2010) allow for a random state-verification technology with exogenous punishment. In this case, Condition 1 holds and there are no permanent intertemporal distortions.

[42] See Golosov, Kocherlakota and Tsyvinski (2003).

[43] Werning (2007) examines a version of this economy where labor productivity is fixed. In this case, consistent with Theorem 1, the first best steady state is admissible, and intertemporal distortions are temporary in the second best.

[44] See Chapter 19 of Ljungqvist and Sargent (2004) for a discussion.

5.3 Government Policy under Limited Commitment

The government's policy choices may be constrained by a lack of commitment or by political economy considerations. These constraints often can be formulated as self-enforcement constraints by the government analyzed within our framework. A few examples of such environments are Fisher (1980), Chari and Kehoe (1990), Sleet and Yeltekin (2006b), Yared (2007), and Acemoglu, Golosov, and Tsyvinski (2008a, 2008b).

Theorem 1 can be used to derive a general result for models with a benevolent government with limited commitment. We show that Condition 1 is always satisfied when the government maximizes social welfare, so that limited commitment *alone* never leads to permanent intertemporal distortions. We then consider settings where the government is self-interested. Specifically, we analyze a simplified version of Acemoglu, Golosov, and Tsyvinski (2008a), and discuss the determinants of intertemporal distortions in light of Theorem 1.

5.3.1 Benevolent Government

We start with the case of benevolent policymaker. Under limited commitment, an allocation ψ is admissible if, for all $s^t \in S^t$ and $t \geq 0$, the continuation allocation $\psi|s^t$ is feasible from state $\left\{s^t, k\left(z^{t-1}\right)\right\}$, and is preferred to the outside option:

$$U\left(x|s^t; s^t\right) \geq V_{out}\left(x(s^t), k\left(z^{t-1}\right), s_t\right). \tag{25}$$

This admissibility constraint resembles the self-enforcement condition (22). Since the policymaker is benevolent, the continuation allocation is evaluated according to the social welfare function, U. The outside option $V_{out}(\cdot)$ may depend on the continuation allocation and the current state of the economy. A variety of specifications for the outside options have been considered. The most widely adopted is the notion of sustainable equilibrium developed by Chari and Kehoe (1990), where the outside option corresponds to a beneficial temporary deviation followed by a reversion to the worst subgame perfect equilibrium. For example, Reis (2006) analyzes a framework in which the worst sustainable equilibrium corresponds to financial autarchy for the government.

It is well known that the outside option plays a critical role in models with limited commitment, and it is hard to discriminate between the many possible specifications. However, since any outside option must be feasible and the first best continuation allocation from state $\left\{s^t, k\left(z^{t-1}\right)\right\}$ satisfies (8), it follows that:[45]

$$U\left(x^{fb}|\{s^t, k(z^{t-1})\}; s^t\right) \geq V_{out}\left(x(s^t), k\left(z^{t-1}\right), s_t\right).$$

Thus, there exists an admissible allocation that converges to the first best limiting allocation and Condition 1 is satisfied.

The application of Theorem 1 clarifies that limited commitment alone will not give rise to per-

[45] Indeed, if $U\left(x|s^t; s^t\right) < V_{out}\left(x(s^t), k\left(z^{t-1}\right), s_t\right)$ then the set of admissible allocations is empty.

manent intertemporal distortions and that this is a robust property of this class of second best economies. Limited commitment may give rise to permanent intertemporal distortions in combination with other frictions. Chari and Kehoe (1990) and Phelan and Stacchetti (2001) study sustainable equilibria in an economy where the government is subject to a balanced budget constraint. Since the government is unable to save, the limiting first best allocation is not admissible and the sufficient condition is violated, as discussed in Section 5.1. Then, sustainable equilibria can display permanent intertemporal distortions.

5.3.2 Self-Interested Policymaker

We now turn to environments where the policymaker is not benevolent. As an example, we consider a simplified version of the economy in Acemoglu, Golosov and Tsyvinski (2008a, 2008b) with no aggregate or idiosyncratic risk. The economy is populated by two types of agents. A continuum of identical private agents value streams of consumption and labor, according to standard preferences. In addition, a rent-seeking ruler values streams of transfers $\{T_t\}_{t\geq 0}$ according to the utility function:

$$\sum_{t=0}^{\infty} \beta^t v(T_t),$$

where $v(\cdot)$ is strictly increasing and concave with $v(0) = 0$ and $\beta \in (0, 1)$.

The social welfare function for this economy can be written as:

$$\sum_{t=0}^{\infty} \beta^t u(c_t, l_t) + \gamma \sum_{t=0}^{\infty} \beta^t v(T_t),$$

allowing a positive Pareto weight, $\gamma \geq 0$, for the ruler. This formulation clarifies the conflict over the distribution of resources between private agents and the ruler.

The ruler can capture a fraction of aggregate output by resorting to expropriation. Thus, any admissible allocation must satisfy:

$$\sum_{j\geq 0} \beta^j v(T_{t+j}) \geq v(\kappa F(k_t, l_t)), \tag{26}$$

where the parameter κ, intended to capture the quality of political institutions in the model, represents the fraction of aggregate output that the ruler can extract.[46] This admissibility constraint has the same structure as the self-enforcement condition in (22).

Is it possible to find an allocation that satisfies (26) and converges to the first best? In the first best steady state, the ruler is awarded a constant share of output and her consumption is constant. As in models with self-enforcement constraints, the first best steady-state is admissible if the utility loss of default is high enough in comparison to a first best continuation. This condition is satisfied if the discount factor is high enough, or if κ is low enough. In particular, if $\kappa = 0$ the admissibility

[46] Following expropriation, the ruler looses power and obtains no transfers.

constraint is trivially satisfied and the limited commitment problem disappears. Finally, the first best steady state is admissible if the ruler's weight in the social welfare function, γ, is large enough. Thus, even if the policymaker is not benevolent, limited commitment will not lead to permanent intertemporal distortions as long as the conflict over the distribution of resources is not too severe.

Acemoglu, Golosov and Tsyvinski (2008a) also allow the ruler to have a different intertemporal discount factor, δ, and show that for $\delta \geq \beta$, there are no intertemporal distortions in the limit.[47] While our framework does not allow for heterogeneity in discount factors, it is still possible to relate this result to Condition 1.[48] For $\delta \geq \beta$, the first best allocation specifies a growing consumption share for the ruler. The first best allocation must then satisfy the admissibility constraint after a finite number of periods, and it easy to see that it will always be possible to find a transition path that is admissible and converges to the first best steady state.

5.4 Firm Dynamics

Our result can also be applied to the principal-agent models used to study entrepreneurial financing and firm dynamics.[49] The contributions in this area most related to our general environment involve self-enforcement constraints or moral hazard.

Albuquerque and Hopenhayn (2004) develop a general model of lending with endogenous borrowing constraints, arising from self-enforcement frictions. This model is an application of the class of environments described in Section 5.2.1, in which it is typically possible to front-load all distortions and our sufficient condition holds. Albuquerque and Hopenhayn (2004) show that, under the optimal contract, the borrowing constraints weaken over time as the borrower acquires more equity, and eventually cease to bind. Thus, distortions are front-loaded and investment converges to the first best level. This model takes liquidation to be the outside option, though our result suggests that the properties of the optimal allocation are robust to alternative formulations.

Quadrini (2004) and Clementi and Hopenhayn (2006) study economies in which investment is financed with optimal contracts between a risk-neutral entrepreneur and an investor, with repeated moral hazard and a limited liability constraint on the entrepreneur.[50] Both environments have the property that incentive constraints become less stringent over time, as the project size and the entrepreneur's stake in the project increase. Thus, the optimal contract front-loads distortions, consistent with our findings. These economies feature intertemporal distortions as long as the incentive compatibility constraints are binding. Clementi and Hopenhayn (2006) find that after a sufficiently long sequence of goods shocks, incentive compatibility constraints cease to bind and capital increases to the first best level. This result stems from risk-neutrality of the entrepreneur,

[47] See also Sleet and Yeltekin (2006b).

[48] In addition, the constrained efficient allocation may not satisfy the interiority assumptions we have made.

[49] Our framework can accommodate principal-agent settings as long as principal and agent share the same intertemporal discount rate. See the Web Appendix for details.

[50] Albanesi (2011) shows that versions of the model in Quadrini (2004) and Clementi and Hopenhayn (2004) are equivalent to a canonical repeated moral hazard model with capital. Abraham and Pavoni (2007) derive conditions under which moral hazard models are isomorphic to the dynamic private information economy described in Section 5.2.2. The Web Appendix formally illustrates how these environments map into our general framework.

which implies that when capital is large enough, the incentive compatibility constraints cease to bind. By contrast, after a sufficiently long path of bad shocks, the project will be liquidated. In both cases, there are no intertemporal distortions in the limit, though the limiting first best allocation is not attained.

These examples of firm dynamics with incentive frictions feature a more restrictive specification for the production technology and the shocks, relative to our general environment. The second best allocation in these economies conforms to the front-loading principle embedded in Theorem 1, and our analysis suggests that the properties of optimal financing arrangements and firm dynamics derived in this context could hold for more general specifications.

6 Conclusions

Our analysis uncovers a tight link between the ability to front-load distortions and the presence of permanent intertemporal distortions. The unified approach we adopt is critical for this insight, since it allows us to identify an optimality principle that is common across a broad class of environments. However, Theorem 1 only provides a sufficient condition and thus it is silent about the long-run behavior of intertemporal distortions for environments where the sufficient condition is not satisfied. We conclude by briefly discussing what form a necessary condition would take, and how the logic embedded in our results can be applied more generally.

Since our general environment is the union of several subclasses of models, we can identify in each subclass a necessary condition and a sufficient condition for the absence of permanent intertemporal distortions. A key question is how the conditions for each subclass combine in the general framework, as a necessary condition for one model may not be necessary for another.

A necessary condition for the general environment is the union of the necessary conditions specific to each subclass, and thus it will be less restrictive than the necessary conditions for each subclass. As the generality of the environment grows, the necessary condition may be trivialized, and, indeed, this occurs in our framework. The necessary condition for the absence of permanent intertemporal distortions in our environment is simply that an admissible allocation with this property exists. This condition is also sufficient for some subclasses, but not all, therefore there exists no jointly necessary and sufficient condition for the general class.

Turning to sufficiency, the intersection of the sufficient conditions specific to each subclass will deliver a sufficient condition for the general class. Clearly then, the sufficient condition for the general model must be more restrictive than each of sufficient conditions for any subclass. Indeed, there is no guarantee that the intersection of sufficient conditions specific to each subclass is non-empty, and in that case there is no sufficient condition for the general class. Given the generality of our environment, it is actually quite surprising that a sufficient condition exists at all.

Since the sufficient condition for the general class will typically be more restrictive than needed for a specific subclass, we can obtain a less stringent sufficient condition for a particular subclass simply by extending the logic embedded in our general sufficient condition. Specifically, while for the

general class the sufficient condition for the absence of permanent intertemporal distortions requires that *all* distortions can be front-loaded, for many environments it is sufficient that *some* distortions can be front-loaded. This result rests on the ability to intertemporally re-allocate consumption without violating the admissibility constraints, which may be possible even if some distortions cannot be front loaded. Some environments encompassed in our framework, though, do require all distortions to be front-loaded. Since the sufficient condition for the general environment must apply to all underlying models, the condition will be too restrictive for many model subclasses.

Despite these limitations, we believe our result provides broader insights. As we discussed in Section 5.1, using our general result we can identify weaker sufficient conditions for the absence of permanent intertemporal distortions in specific applications. Moreover, the link between the ability to front-load distortions and permanent intertemporal distortions delivers implications beyond its predictions for the second best allocation. Based on the optimality principle that underlies our argument, if private savings are distorted, then reducing distortions in future periods can increase welfare in the class of economies we consider. This constitutes an important lesson for policy design and suggests a novel class of Pareto-improving reforms. Albanesi and Armenter (2007) illustrate this point quantitatively in the context of the Ramsey model, and show that fiscal reforms that front-load of intratemporal distortions from a sub-optimal initial allocation deliver welfare gains for a variety of parameterizations and specifications of the reform.

References

[1] Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. 2008a. Political Economy of Mechanisms. *Econometrica*, 76(3): 619-641.

[2] Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. 2008b. Market versus Governments. *Journal of Monetary Economics*, 55(1): 159-189.

[3] Aiyagari, S. Rao. 1995. Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy* 103(6): 1158-75,

[4] Aiyagari, S. Rao, Albert Marcet, Thomas J. Sargent, and Juha Seppala. 2002. Optimal Taxation without State-Contingent Debt. *Journal of Political Economy* 110 (6): 1220-1254.

[5] Albanesi, Stefania. 2011. Optimal Taxation of Entrepreneurial Capital with Private Information. Manuscript, Columbia University.

[6] Albanesi, Stefania. 2008. Discussion of "Market versus Governments." *Journal of Monetary Economics* 55 (1).

[7] Albanesi, Stefania and Roc Armenter. 2007a. Intertemporal Distortions in the Second Best. CEPR Discussion Paper 6577.

[8] Albanesi, Stefania and Roc Armenter. 2007b. Understanding Capital Taxation in Ramsey Models. Working paper, Federal Reserve Bank of New York.

[9] Albanesi, Stefania and Christopher Sleet. 2006. Dynamic Optimal Taxation with Private Information. *The Review of Economic Studies,* 73: 1-30.

[10] Albuquerque, Rui and Hugo A. Hopenhayn. 2004. Optimal Lending Contracts and Firm Dynamics. *Review of Economic Studies*, 71:285-315.

[11] Armenter, Roc. 2008. A Note on Incomplete Factor Taxation. *Journal of Public Economics* 92 (10-11): 2275-2281.

[12] Armenter, Roc and Thomas Mertens. 2010. Fraud Deterrence in Dynamic Mirrleesian Economies. *Federal Reserve Bank of Philadelphia*, working paper 10-7.

[13] Atkeson, Andrew and Robert E. Lucas, Jr. 1992. On Efficient Distribution with Private Information. *Review of Economic Studies* 59: 427-453.

[14] Atkeson, Andrew and Robert E. Lucas, Jr. 1995. Efficiency and Equality in a Simple Model of Unemployment Insurance. *Journal of Economic Theory* 66 (1): 64-88.

[15] Atkeson, Andrew, V.V. Chari, and Patrick J. Kehoe. 1999. Taxing Capital Income: A Bad Idea. Federal Reserve Bank of Minneapolis Quarterly Review 23 (3): 3–17.

[16] Atkinson, Anthony B. and Joseph E. Stiglitz. 1980. Lectures on Public Economics. McGraw-Hill, New York.

[17] Chamley, Christophe, 1986. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica* 54 (3): 607–622.

[18] Chari, V.V., and Patrick J. Kehoe. 1990. Sustainable Plans. *Journal of Political Economy,* 98 (4): 783-802.

[19] Chari, V.V., and Patrick J. Kehoe. 1999. Optimal fiscal and monetary policy. In J. B. Taylor, and M. Woodford (ed.), Handbook of Macroeconomics.

[20] Chien, Yili and Junsang Lee. 2008. Why Tax Capital? Working paper, Purdue University.

[21] Clementi, Gian Luca and Hugo A. Hopenhayn. 2006. A Theory of Financing Constraints and Firm Dynamics. *Quarterly Journal of Economics*, 229-265

[22] Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger. 2009. Taxing Capital? Not a Bad Idea After All! *American Economic Review* 99(1) pp.25-48.

[23] Correia, Isabel H. 1996. Should Capital Income Be Taxed in the Steady State? *Journal of Public Economics* 60 (1): 147-151.

[24] Davila, J., J. Hong, P. Krusell and V. Rios-Rull. 2007. Constrained-Efficient Allocations in the Growth Model with Idiosyncratic Shocks. Working paper, University of Pennsylvania.

[25] Diamond, P. and Mirrlees, J. 1978. A model of social insurance with variable retirement. *Journal of Public Economics* 10: 295–336.

[26] Erosa, Andres and Martin Gervais. 2002. Optimal Taxation in Life-Cycle Economies. *Journal of Economic Theory* 105(2): 338-369.

[27] Farhi, Emmanuel. 2010. Capital Taxation and Ownership When Markets are Incomplete. *Journal of Political Economy*, Vol. 118, No. 5, pp. 908-948

[28] Farhi, Emmanuel and Ivan Werning. 2007. Inequality and Social Discounting. *Journal of Political Economy* 115: 365-402.

[29] Fischer, Stanley. 1980. Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government. *Journal of Economic Dynamics and Control* 2: 93-107.

[30] Garriga, Carlos. 2003. Optimal Fiscal Policy in Overlapping Generation Models. Manuscript. Florida International University.

[31] Golosov, Mikhail Narayana Kocherlakota, Aleh Tsyvinski. 2003. Optimal Indirect and Capital Taxation. *The Review of Economic Studies*, Volume 70, Issue 3, Page 569.

[32] Golosov, Mikhail, and Aleh Tsyvinski. 2007. Optimal Taxation with Endogenous Insurance Markets. *Quarterly Journal of Economics* 122(2): 487-534.

[33] Golosov, Mikhail, and Aleh Tsyvinski. 2006. Designing Optimal Disability Insurance: A Case for Asset Testing. *Journal of Political Economy*, 114(2): 257-269.

[34] Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi. 1997. On the Optimal Taxation of Capital Income. *Journal of Economic Theory* 73(1): 93-117.

[35] Judd, Kenneth. 1985. Redistributive Taxation in a Perfect Foresight Model. *Journal of Public Economics* 28, 59-84.

[36] Kehoe, Timothy J. and Levine, David K. 1993. Debt-Constrained Asset Markets. *The Review of Economic Studies* 60(4): 865-88.

[37] Kehoe, Patrick and Fabrizio Perri. 2002. International Business Cycles with Endogenous Incomplete Market. *Econometrica* 70 (3): 907-928.

[38] Kehoe, Patrick and Fabrizio Perri. 2004. Competitive Equilibria with Limited Enforcement. *Journal of Economic Theory* 119 (1): 184-206.

[39] Kocherlakota, Narayana. 1996. Implications of Efficient Risk Sharing without Commitment. *Review of Economic Studies* 63: 595-610.

[40] Kocherlakota, Narayana. 2005a. Advances in Dynamic Optimal Taxation. Paper presented at the World Congress of the Econometric Society, August 2005.

[41] Kocherlakota, Narayana. 2005b. Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation. *Econometrica*.

[42] Lansing, Kevin. 1999. Optimal Redistributive capital Taxation in a Neoclassical Growth model. *Journal of Public Economics* 73, 423-453.

[43] Ljungqvist, Lars and Thomas J. Sargent. 2004. Recursive Macroeconomic Theory. The MIT Press.

[44] Lucas, Robert E., Jr., and Nancy L. Stokey. 1983. Optimal Fiscal and Monetary Policy in an Economy without Capital. *Journal of Monetary Economics* 12: 55–93.

[45] Luenberger, David G. 1997. Optimization by Vector Space Methods. John Wiley and Sons, New York.

[46] Phelan, Christopher. 1998. On the Long-Run Implications of Repeated Moral Hazard. *Journal of Economic Theory,* 79(2): 174-191.

[47] Phelan, Christopher. 2006. Opportunity and Social Mobility. *The Review of Economic Studies* 73: 487-504.

[48] Phelan, Christopher and Ennio Stacchetti. 2001. Sequential equilibria in a Ramsey tax model. *Econometrica,* 69(6): 1491-1518.

[49] Quadrini, Vincenzo. 2004. Investment and Liquidation in Renegotiation-proof Contracts with Moral Hazard. *Journal of Monetary Economics*, 51(4), pp. 713-75.

[50] Ray, Debraj. 2002. The Time Structure of Self-Enforcing Agreements. *Econometrica* 70 (2), 547-582.

[51] Reis, Catarina. 2006. Taxation without Commitment. Manuscript, MIT.

[52] Rogerson, William. 1985. Repeated moral hazard. *Econometrica* 53:69-76.

[53] Salanie, Bernard. 2003. *The economics of Taxation*. The MIT Press, Cambridge, MA.

[54] Sleet, Christopher and Sevin Yeltekin. 2006a. Credibility and Endogenous Societal Discounting. *Review of Economic Dynamics*, 9 (3): 410-437.

[55] Sleet, Christopher and Sevin Yeltekin. 2006b. Optimal Taxation with Endogenously Incomplete Debt Markets. 2006. *Journal of Economic Theory*, 127: 36-73.

[56] Stokey, Nancy and Robert Lucas. 1996. *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, MA.

[57] Stockman, David. 2001. Balanced-Budget Rules: Welfare Loss and Optimal Policies. *Review of Economic Dynamics* 4, 438-459.

[58] Thomas, Jonathan and Tim Worral. 1990. Income Fluctuation and Asymmetric Information. *Journal of Economic Theory* 51 (2): 367-390.

[59] Tuomala, Matti. 1990. *Optimal Income Tax and Redistribution.* Claredon Press, Oxford, UK.

[60] Werning, Ivan. 2007. Optimal Fiscal Policy with Redistribution. *The Quarterly Journal of Economics* CXXII.

[61] Yared, Pierre. 2007. Politicians, Taxes and Debt. Manuscript, MIT.

[62] Zhu, Xiaodong. 1992. Optimal Fiscal Policy in a Stochastic Growth Model. *Journal of Economic Theory.* 58: 250–89.

A Proofs for Stage 1

A.1 Proof of Lemma 1

The set of plans Ψ is a complete vector space with the sup norm. U, F, and all terms in H are real-valued and twice differentiable functions by Assumptions 2, 3, and 4. Let ψ^* be a second best plan and a^* an auxiliary variable plan such that $\{\psi^*, a^*\}$ is admissible. By Assumption 6, ψ^* is a regular point of all the binding constraints in Problem 14 given a^*. All conditions are satisfied thus to apply the generalized Kuhn-Tucker theorem (See Luenberger (1997), page 294) and characterize the solution ψ^* with the Lagrangian

$$\begin{aligned}
\Lambda(\psi;\varphi,\phi,\lambda) &= U(x;s_0) \\
&\quad -\sum_{t=0}^{\infty}\sum_{z^t\in Z^t}\beta^t\pi\left(z^t|z_0\right)\lambda\left(z^t\right)\left\{\mathbf{c}\left(z^t\right)+k\left(z^t\right)-F\left(z^t\right)\right\} \\
&\quad -\sum_{t=0}^{\infty}\sum_{z^t\in Z^t}\beta^t\pi\left(z^t|z_0\right)\varphi\left(z^t\right)\left\{\sum_{i\in I}\sum_{\theta^t\in\Theta^t}\pi_i\pi_\theta\left(\theta^t\right)y_i\left(z^t,\theta^t\right)-k\left(z^t\right)\right\} \\
&\quad -\sum_{t=0}^{\infty}\sum_{s^t\in S^t}\beta^t\pi\left(s^t|s_0\right)\phi\left(s^t\right)'\left\{H\left(s^t\right)-a\left(s^t\right)\right\},
\end{aligned}$$

where the dependence of H and F on ψ is understood. The Lagrangian multipliers $\phi\left(s^t\right)$ are stacked vectors of length $I\times M$.

If $s^d \in \varpi_d$ then $s^t \in \varpi_t$ for all $s^t \in S^t|s^d$. By definition,

$$H_{m,i}\left(s^t\right) - a_{m,i}\left(s^t\right) \leq \sup\left\{H_{m,i}\left(s^j\right) - a_{m,i}\left(s^j\right) : s^j \in S^j|s^t, j \geq t\right\} < 0$$

and the Kuhn-Tucker conditions associated with the Lagrangian imply that $\phi_{m,i}\left(s^t\right) = 0$ for all $s^t \in S^t|s^d$ and all $\{i, m\} \in I \times M$.

Next we write the first order necessary conditions (f.o.n.c.) associated with $x\left(s^t\right)$ for any node $s^t \in S^t|s^d, t > d+1$ for some $s^d \in \varpi_d$. For $c_i\left(s^t\right)$ we have that

$$\begin{aligned}\pi_i\left(\beta^t \pi\left(s^t\right)\right)\left(u_i^c\left(s^t\right) - \lambda\left(z^t\right)\right) &= \beta^t \pi\left(s^t\right)\left(\sum_{j=0}^{d} \phi\left(s^j\right) \cdot d_i\left(s^j\right)\right)' \left[\frac{\partial h_{m,i}^0\left(x\left(s^t\right); s_t\right)}{\partial c_i\left(s^t\right)}\right]_m + \\ &\left(\beta^t \sum_{j=0}^{d} \sum_{\tilde{s}^j \in S^j|s^{t-1}} \pi\left(s^t|s^j\right) \pi\left(\tilde{s}^j|s_0\right) \phi_i\left(\tilde{s}^j\right) \cdot d_i\left(\tilde{s}^j\right)\right)' \left[\frac{\partial h_{m,i}^1\left(x\left(s^t\right); \tilde{s}_t\right)}{\partial c_i\left(s^t\right)}\right]_m\end{aligned}$$

where $\phi_i\left(s^j\right)$ is a vector collecting the M Lagrangian multipliers associated with the admissibility constraints of ex-ante type i at node s^j; $d_i\left(s^j\right)$ and $\left[\frac{\partial h_{m,i}^0\left(x\left(s^t\right); s_t\right)}{\partial c_i\left(s^t\right)}\right]_m, \left[\frac{\partial h_{m,i}^1\left(x\left(s^t\right); \tilde{s}_t\right)}{\partial c_i\left(s^t\right)}\right]_m$ are vectors similarly stacked. We have used Assumption 6 to write the f.o.n.c. with strict equality. We simplify the term

$$\beta^t \sum_{j=0}^{d} \sum_{\tilde{s}^j \in S^j|s^{t-1}} \pi\left(s^t|s^j\right) \pi\left(\tilde{s}^j|s_0\right) \phi_i\left(\tilde{s}^j\right) \cdot d_i\left(\tilde{s}^j\right) = \beta^t \pi\left(s^t\right) \sum_{j=0}^{d} \sum_{\tilde{s}^j \in S^j|s^{t-1}} \frac{\pi\left(\tilde{s}^j|s_0\right)}{\pi\left(s^j|s_0\right)} \phi_i\left(\tilde{s}^j\right) \cdot d_i\left(\tilde{s}^j\right).$$

By Assumption 1 the term is finite. Rewrite the f.o.n.c. as

$$\pi_i u_i^c\left(s^t\right) - \pi_i \lambda\left(z^t\right) = \left[\bar{\phi}_i^0\left(s^d\right)\right]' \left[\frac{\partial h_{m,i}^0\left(x\left(s^t\right); s_t\right)}{\partial c_i\left(s^t\right)}\right]_m + \left[\bar{\phi}_i^1\left(s^d\right)\right]' \left[\frac{\partial h_{m,i}^1\left(x\left(s^t\right); \tilde{s}_t\right)}{\partial c_i\left(s^t\right)}\right]_m \tag{27}$$

where

$$\begin{aligned}\bar{\phi}_i^0\left(s^d\right) &= \left(\sum_{j=0}^{d} \phi_i\left(s^j\right) \cdot d_i\left(s^j\right)\right), \\ \bar{\phi}_i^1\left(s^d\right) &= \left(\sum_{j=0}^{d} \sum_{\tilde{s}^j \in S^j|s^{t-1}} \frac{\pi\left(\tilde{s}^j|s_0\right)}{\pi\left(s^j|s_0\right)} \phi_i\left(\tilde{s}^j\right) \cdot d_i\left(\tilde{s}^j\right)\right).\end{aligned}$$

Similarly, the f.o.n.c. for labor $l_i\left(s^t\right)$ for nodes $s^t \in S^t|s^d$ can be written as

$$\pi_i u_i^l\left(s^t\right) + \pi_i \lambda\left(z^t\right) F^l\left(z^t\right) = \left[\bar{\phi}_i^0\left(s^d\right)\right]' \left[\frac{\partial h_{m,i}^0\left(x\left(s^t\right); s_t\right)}{\partial l_i\left(s^t\right)}\right]_{m \leq M} + \left[\bar{\phi}_i^1\left(s^d\right)\right]' \left[\frac{\partial h_{m,i}^1\left(x\left(s^t\right); \tilde{s}_t\right)}{\partial l_i\left(s^t\right)}\right]_{m \leq M}. \tag{28}$$

Finally the first order for aggregate capital $k\left(z^t\right)$, $z^t \in Z^t|z^d$, is

$$\lambda\left(z^t\right) = \beta \sum_{z^{t+1} \in Z^{t+1}|z^t} \pi_z\left(z_{t+1}|z_t\right) \lambda\left(z^{t+1}\right) F^k\left(z^{t+1}\right). \tag{29}$$

The system of equations given by the f.o.n.c. for consumption $c_i\left(s^t\right)$, labor $l_i\left(s^t\right)$ and capital $k\left(z^t\right)$ (27)-(29), and the resource constraint (6) is a function of the aggregate shock z_t, the realization of the idiosyncratic shock θ_t, the stock of aggregate capital $k\left(z^{t-1}\right)$, and the constants $\left\{\bar{\phi}_i^0, \bar{\phi}_i^1\right\} \equiv \left\{\bar{\phi}_i^0\left(s^d\right), \bar{\phi}_i^1\left(s^d\right)\right\}$. Thus, we can express $\left\{x_i\left(s^t\right), k\left(z^t\right)\right\}$ as:

$$\begin{aligned} x_i\left(s^t\right) &= f_x\left(\sigma; \bar{\phi}_i^0, \bar{\phi}_i^1\right), \\ k\left(z^t\right) &= f_k\left(\sigma; \bar{\phi}_i^0, \bar{\phi}_i^1\right), \end{aligned}$$

for all $s^t \in S^t|s^d$, where

$$\sigma = \left\{s^t, k\left(z^{t-1}\right)\right\} \in \Sigma \equiv S \times K$$

and

$$\begin{aligned} f_x &: \Sigma \to X, \\ f_k &: \Sigma \to K. \end{aligned}$$

The dependence on the constants $\left\{\bar{\phi}_i^0, \bar{\phi}_i^1\right\}$ will henceforth be omitted for ease of notation. The functions f_x and f_k do not describe all the allocations in the economy, but only those in continuations nodes of a given $s^d \in \varpi_d$. Under Assumptions 2, 3, and 4, both f_x and f_k are bounded.

Let $\Omega_i(\sigma) : \Sigma \to \Re$ be given by:

$$\Omega_i(\sigma) \equiv 1 - \frac{1}{u_i^c(\sigma)} \left(\left(\bar{\phi}_i^0\right)' \left[\frac{\partial h_m^0(\sigma)}{\partial c_i}\right]_{m \leq M} + \left(\bar{\phi}_i^1\right)' \left[\frac{\partial h_m^1(\sigma)}{\partial c_i}\right]_{m \leq M} \right)$$

for some $i \in I$, where we have written $u_i^c(\sigma)$ for $u_i^c\left(s^t\right)$ and so on. Moreover, Assumptions 3 and 4 implies that first derivatives of u_i, h_m^0, and h_m^1 must be bounded above and below in $[0, \bar{c}] \times \left[0, \bar{l}\right]$, and u_i^c is strictly positive. It follows that for all $\sigma \in \Sigma$, $\Omega_i(\sigma)$ is bounded above and below.

Define the operator Ξ on bounded functions as:

$$\Xi\left[\Omega_i\right](\sigma) = \frac{\sum_{s' \in S} \omega_i(\sigma, s') \Omega_i(\sigma')}{\sum_{s' \in S} \omega_i(\sigma, s')}, \tag{30}$$

where

$$\omega_i\left(\sigma, s'\right) = \beta \pi\left(s'|s\right) u_i^c\left(\sigma'\right) F^k\left(\sigma'\right)$$

and σ' is given by

$$\sigma' = \left\{s', f_k(\sigma)\right\}$$

for each $s' \in S$. By construction,

$$u_i^c\left(s^t\right) \leq (\geq)\beta \sum \pi\left(s^{t+1}|s_t\right) u_i^c\left(s^{t+1}\right) F^k\left(s^{t+1}\right)$$

if and only if:

$$\Xi\left[\Omega_i\right](\sigma) \leq (\geq)\Omega_i(\sigma).$$

By Assumptions 1 and 8, there must exist an an ergodic set $\Sigma_\infty \subseteq \Sigma$. That is, if $\sigma \in \Sigma_\infty$ then $\sigma' = \{s', f_k(\sigma)\} \in \Sigma_\infty$ for all realizations of $s' \in S$. We derive the two key properties of operator Ω in the following Lemma.

Lemma 2 *Let Σ_∞ be an ergodic set. If for all $\sigma \in \Sigma_\infty$:*

$$\Xi\left[\Omega_i\right](\sigma) \geq \Omega_i(\sigma),$$

then,

$$\Xi\left[\Omega_i\right](\sigma) = \Omega_i(\sigma), \tag{31}$$

for all $\sigma \in \Sigma_\infty$.

If , instead, for all $\sigma \in \Sigma_\infty$:

$$\Xi\left[\Omega_i\right](\sigma) \leq \Omega_i(\sigma),$$

then,

$$\Xi\left[\Omega_i\right](\sigma) = \Omega_i(\sigma), \tag{32}$$

for all $\sigma \in \Sigma_\infty$.

Proof. Let $\sigma_u \in \Sigma$ be such that $\Omega_i(\sigma_u) = \sup\{\Omega_i(\sigma) : \sigma \in \Sigma_\infty\}$. This exists since operator Ω is bounded on Σ_∞.

If $\sigma_u \in \Sigma_\infty$ then $\Xi\left[\Omega_i\right](\sigma_u) = \Omega_i(\sigma_u)$ as $\Omega_i(\sigma) \leq z$ implies $\Xi\left[\Omega_i\right](\sigma) \leq z$. By ergodicity, all continuations $\sigma' = \{s', f_k(\sigma_u)\}$ satisfy $\Omega_i(\sigma_u) \geq \Omega_i(\sigma')$. Therefore $\Xi\left[\Omega_i\right](\sigma_u) = \Omega_i(\sigma_u)$ implies $\Omega_i(\sigma_u) = \Omega_i(\sigma')$ for all continuations $\sigma' = \{s', f_k(\sigma_u)\}$. Hence for any $\sigma \in \hat{\Sigma} = \{\sigma \in \Sigma_\infty : \Omega_i(\sigma) = \Omega_i(\sigma_u)\}$, all $\sigma' = \{s', f_k(\sigma)\} \in \hat{\Sigma}$ and thus $\hat{\Sigma} = \Sigma_\infty$. It follows that $\Xi\left[\Omega_i\right](\sigma) = \Omega_i(\sigma)$ for all $\sigma \in \Sigma_\infty$.

If $\sigma_u \notin \Sigma_\infty$ note that $\Xi\left[\Omega_i\right](\sigma_u) \geq \Omega_i(\sigma_u)$ by the continuity of Ω_i and Ξ. If $\Xi\left[\Omega_i\right](\sigma_u) > \Omega_i(\sigma_u)$ then there exist a continuation $\hat{\sigma}' = \{\hat{s}, f_k(\sigma_u)\}$ with $\Omega_i(\hat{\sigma}') > \Omega_i(\sigma_u)$. Consider a sequence $\{\sigma_n\}$ in Σ_∞ such that $\sigma_n \to \sigma_u$ and construct the sequence $\{\hat{\sigma}_n\} \in \Sigma_\infty$ given by $\hat{\sigma}_{n+1} = \{\hat{s}, f_k(\sigma_n)\}$. Since $\Omega_i(\hat{\sigma}_n) \to \Omega_i(\hat{\sigma}')$ there exists m such that $\Omega_i(\hat{\sigma}_m) > \Omega_i(\sigma_u)$ contradicting $\hat{\sigma}_m \in \Sigma_\infty$. Hence $\Xi\left[\Omega_i\right](\sigma_u) = \Omega_i(\sigma_u)$ and the reasoning for case $\sigma_u \in \Sigma_\infty$ applies as well. This proves the first statement. The proof of the second statement applies the same argument, using $\inf\{\Omega_i(\sigma) : \sigma \in \Sigma_\infty\}$ instead. ■

Finally, let

$$\Delta_i^-\left(s^t\right) = \min\left\{0, \Delta_i\left(s^t\right)\right\}$$

and

$$\Delta_i^+\left(s^t\right) = \max\left\{0, \Delta_i\left(s^t\right)\right\}.$$

Consider first

$$\alpha^- = \lim_{t\to\infty} \sum_{s^t} \pi\left(s^t\right) \Delta_i^-\left(s^t\right).$$

Since $\Delta_i^-\left(s^t\right)$ can be expressed as a continuous function of allocations on a compact support, the limit exists by Assumption 8. If

$$\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1$$

then

$$\alpha^- = \lim_{t\to\infty} \sum_{s^t\in\varpi_t} \pi\left(s^t\right) \Delta_i^-\left(s^t\right).$$

If $\alpha^- = 0$ then $\Xi\left[\Omega_i\right](\sigma) \geq \Omega_i(\sigma)$ for all $\sigma \in \Sigma_\infty$. By Lemma 2 then $\Xi\left[\Omega_i\right](\sigma) = \Omega_i(\sigma)$ for all $\sigma \in \Sigma_\infty$ and

$$\lim_{t\to\infty} \sum_{s^t} \pi\left(s^t\right) \Delta_i\left(s^t\right) = 0.$$

The same argument applies for

$$\alpha^+ = \lim_{t\to\infty} \sum_{s^t\in\varpi_t} \pi\left(s^t\right) \Delta_i^+\left(s^t\right)$$

using now the second part of Lemma 2. Hence either

$$\lim_{t\to\infty} \sum_{s^t} \pi\left(s^t\right) \Delta_i\left(s^t\right) = 0$$

or

$$\begin{aligned}
\lim_{t\to\infty} \sum_{s^t\in\varpi_t} \pi\left(s^t\right) \Delta_i^+\left(s^t\right) &> 0, \\
\lim_{t\to\infty} \sum_{s^t\in\varpi_t} \pi\left(s^t\right) \Delta_i^-\left(s^t\right) &> 0,
\end{aligned}$$

which proves the result.

A.2 Proof of Corollary 1

If $\Delta_i\left(s^t\right) > 0$ for all $s^t \in S^t$ then clearly

$$\lim_{t\to\infty} \sum_{s^t\in\varpi_t} \pi\left(s^t\right) \Delta_i^-\left(s^t\right) = 0$$

and thus

$$\lim_{t\to\infty}\sum_{s^t}\pi\left(s^t\right)\Delta_i\left(s^t\right)=0.$$

The same reasoning applies if $\Delta_i\left(s^t\right)<0$ for all $s^t\in S^t$.

B Proofs for Stage 2

B.1 Proof of Proposition 1

Condition 1 implies that there exists an allocation $\tilde{\varphi}$ and a plan for the auxiliary variable $\tilde{a}\in A^\infty$ such that:

$$H_{m,i}(\tilde{\psi},s^t)\leq\tilde{a}_{im}(s^t),$$

and:

$$\tilde{a}(s^{t+1})\in\Gamma(\tilde{a}(s^t),s_t),$$

for all $s^t\in S^t$ for $t\geq 0$, $i\in I$,$m\in M$ and all $\tilde{a}_0\in A$, with the property that the sequence of probability measures $\{\tilde{\mu}_t\}$ on $K\times X$ associated with $\tilde{\psi}$ converges weakly to μ_∞^{fb}.

Define $\bar{a}_t=\max\limits_{s^t\in S^t}\left\{H(\tilde{\psi},s^t)\right\}$ for all $t\geq 0$ and $\hat{a}_t=\max\{\bar{a}_t,\hat{a}_{t-1}\}$ for all $t\geq 1$, where the maximum is taken componentwise and $\hat{a}_0=\tilde{a_0}$. Since $\tilde{\psi}$ is admissible, each element in the sequence $\{\widehat{a}_t\}$ is in A. By construction, the sequence $\{\hat{a}_t\}$ is monotone increasing. Since A is a compact set, $\lim\limits_{t\to\infty}\hat{a}_t=\hat{a}_\infty\in A$. By weak convergence of $\{\tilde{\mu}_t\}$, and the continuity of H in Assumption 4, we have that

$$\lim_{t\to\infty}\sup_{s^j\in S^j,j\geq t}\{H_{m,i}(\tilde{\psi},s^j)\}=\lim_{t\to\infty}\sup_{s^j\in S^j,j\geq t}\{H_{m,i}(\psi^{fb},s^j)\}\leq\hat{a}_\infty,$$

for all $i\in I,m\in M$.

B.2 Proof of Proposition 2

If $\alpha\in A^{fb}$ the result is trivial. If $\alpha\in int\left(\Gamma\left(\alpha,s\right)\right)$ for some $s\in S$, then the result follows from Assumption 5. Consider thus $\alpha\notin A^{fb}$ and $\alpha\notin int\left(\Gamma\left(\alpha,s\right)\right)$ for all s. By Assumption 5, $\alpha\in\Gamma\left(\alpha,s\right)$ and thus α is an adjacent point to $\Gamma\left(\alpha,s\right)$. Since the image of $\Gamma\left(\alpha,s\right)$ is a convex set by Assumption 5, we apply the Separating Hyperplane Theorem to find a half-space $\chi_s=\{\gamma\in A:\,p\gamma\geq p\alpha\in\Re\}$ with $p\neq\vec{0}$, such that $\Gamma\left(\alpha,s\right)\subset\chi_s$ for a given $s\in S$.

We now show that for some $s\in S$, $A^{fb}\cap\chi_s\neq\varnothing$ for any half-space χ_s. We proceed by contradiction. Assume that for every $s\in S$ there exists a half-space χ_s such that $A^{fb}\cap\chi_s=\varnothing$. Condition 1 implies that the set A^{fb} is non-empty by Proposition 1 and that for some $s^*\in S$, there exists $x\in\chi_{s^*}$ with $y\in\Gamma\left(x,s^*\right)$, $y\notin\chi_{s^*}$. Otherwise there would be no admissible plan with $a_0\in\chi_{s^*}$ that would converge to A^{fb}.

Let $w\in A$ satisfy $\alpha=\varrho w+(1-\varrho)\,x$ for some $\varrho\in(0,1)$. Such a point will belong to the closure of the complement of χ_{s^*}, the set $\{\gamma\in A:p\gamma\leq p\alpha\in\Re\}$. Since $w\in\Gamma\left(w,s^*\right)$, the convexity of Γ

implies that $\varrho y + (1-\varrho) w \in \Gamma(\alpha, s^*)$ but $\varrho y + (1-\varrho) w \notin \chi_{s^*}$. Thus, we reach a contradiction and therefore $A^{fb} \cap \chi_s \neq \varnothing$ for some $s \in S$.

Finally, consider the set:

$$G = \left\{ (1-\varrho)\alpha + \varrho\alpha' \in A : \varrho \in (0,1], \alpha' \in A^{fb} \right\}.$$

This is a convex set with $A^{fb} \subset G$. If $\Gamma(\alpha, s) \cap G = \varnothing$ for all $s \in S$, it would then be possible to find a separating hyperplane χ_s for each s with $\Gamma(\alpha, s) \subset \chi_s$ and $G \cap \chi_s \neq \varnothing$. This would contradict our previous result. Thus for some $s \in S$, $\Gamma(\alpha, s) \cap G \neq \varnothing$ and the result follows.

B.3 Proof of Proposition 3

Let ψ^* be a second best plan and assume that Condition 1 holds. We prove the result by constructing an auxiliary variable plan a^* such that $\{\psi^*, a^*\}$ is admissible and $\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1$. We start with two Lemmas that will be useful to characterize the choice of a^*.

Lemma 3 *Let a be an admissible plan for the auxiliary variable with $a\left(s^t\right) \in int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$ for some $s^t \in S^t$. Then there exists an admissible plan $\tilde{a}$ with $\tilde{a}(s^t) > a(s^t)$ and $\tilde{a}(s^j) \geq a(s^j)$ for all $s^j \in S^j|s^t$; and $\tilde{a}(s^j) = a(s^j)$ elsewhere. Moreover if $a\left(s^j\right) \in int(A)$ for all $s^j \in S^j|s^t$ then $\tilde{a}$ can be found such that $\tilde{a}(s^j) > a(s^j)$ for all $s^j \in S^j|s^t$.*

Proof. We prove the result by construction. Let $a\left(s^t\right) \in int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$. Let $\tilde{a} = a$ everywhere but in the set $\left\{S^j|s^t : j \geq t\right\}$. Let $\gamma = [\bar{a}]_{m,i} \in \Re^{IM}$ be the join for A. Since $a\left(s^t\right) \in int\left(\Gamma\left(a\left(s^{t-1}\right), s_{t-1}\right)\right)$, $a\left(s^t\right) \in int(A)$ and thus $a\left(s^t\right) < \gamma$. Then, for sufficiently small scalar $\varrho > 0$:

$$\tilde{a}\left(s^t\right) = \varrho\gamma + (1-\varrho)\, a\left(s^t\right) \in \Gamma\left(a\left(s^{t-1}\right), s_t\right)$$

with $a\left(s^t\right) < \tilde{a}\left(s^t\right)$. By Assumption 5, Γ is convex and $\gamma \in \Gamma(\gamma, s)$, and thus

$$\tilde{a}\left(s^{t+1}\right) = \varrho\gamma + (1-\varrho)\, a\left(s^{t+1}\right) \in \Gamma\left(\tilde{a}\left(s^t\right), s_t\right).$$

Construct $\tilde{a}$ for all $s^j \in S^j|s^t$, $j \geq t+1$ by iteration. The plan $\tilde{a}$ is admissible and $\tilde{a}(s^j) \geq a(s^j)$ for all $s^j \in S^j|s^t$ by the definition of γ and $\varrho > 0$. The corollary follows trivially ■

The second Lemma uses the quasi-concavity of W.

Lemma 4 *Let $\{\psi, a\}$ be admissible, with $U(x, s_0) = W(a)$. Then for all admissible a' such that $a' \geq a$ (component-wise), we have $W(a') \geq W(a)$. Moreover, if $s^t \notin \varpi_t$ and $a'\left(s^j\right) > a\left(s^j\right)$ for all $s^j \in S^j|s^t$ then either $W(a') > W(a)$ or there exists an admissible pair $\{\psi', a'\}$ such that $U(x', s_0) = W(a')$ and $s^t \in \varpi'_t$.*

Proof. The first part is trivial as $\{\psi, a'\}$ is admissible so $W(a')$ cannot be worse than $W(a)$. Let $s^t \notin \varpi_t$ and $a'\left(s^t\right) > a\left(s^t\right)$. If $W(a') = W(a)$ then $U(x, s_0) = W(a')$ so setting $\psi' = \psi$ we have

$s^t \in \varpi'_t$ ■

Since ψ^* is admissible, there exists at least one auxiliary variable plan a such that $\{\psi^*, a\}$ is admissible. The sequence $\left\{\pi(s^t \in \varpi_t)\right\}_t$ is weakly increasing and every element belongs to $[0,1]$, hence it converges. If $\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1$ the result is proved by equating a with a^*.

If $\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) < 1$, let $\tilde{S}^\infty \subseteq S^\infty$ be the subset of histories such that

$$s^t \notin \varpi_t$$

for all $t \geq 0$. If $a\left(s^t\right) \in A^{fb}$ for any t and $s^\infty \in \tilde{S}^\infty$, then by Assumption 5 there exists an admissible auxiliary variable plan a' with $a'\left(s^j\right) = a\left(s^t\right)$ for all $s^j \in S^j|s^t$, $a' = a$ elsewhere. Clearly $a'\left(s^j\right) \in A^{fb}$ for all $s^j \in S^j|s^t$ and thus by Assumption 7 we have that $s^t \in \varpi'_t$ and $\{\psi^*, a'\}$ is admissible. By induction either we find an admissible a' with $\lim_{t\to\infty} \pi\left(s^t \in \varpi'_t\right) = 1$, which proves the result; or for some plan a with $\{\psi^*, a\}$ admissible we have that $a\left(s^t\right) \notin A^{fb}$ for all $s^t \subset s^\infty \in \tilde{S}^\infty$.

Consider now the possibility that for plan a we have $a\left(s^t\right) \in int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$ for some $s^t \in \tilde{S}^\infty$. If $a\left(s^j\right) \in int\left(A\right)$ for all $s^j \in S^j|s^t$ then by Lemma 3 there exists another admissible a' with $a'(s^t) > a(s^t)$ (component-wise) for all $s^j \in S^j|s^t$, $a' = a$ elsewhere. By Lemma 4 we have that $s^t \in \varpi'_t$ as $W\left(a\right) < W\left(a'\right)$ would contradict ψ^* being the second best. Lemma 3 still applies if $a\left(s^j\right) \notin int\left(A\right)$ because $a_{m,i}\left(s^j\right) = 0$ for some m, i. If $a\left(s^j\right) \notin int\left(A\right)$ because $a_{m,i}\left(s^j\right) = \bar{a}$ for some m, i then by Assumption 5 we can work with the subspace $\bar{A} = \{\alpha \in A : \alpha_{m,i} = \bar{a}\}$ for all $s^d \in S^d|s^j$. By Proposition 1, $A^{fb} \neq \varnothing$ and follows that the vector $\gamma = [\bar{a}]_{m,i}$ (the join of A) must belong to A^{fb}. Since $a\left(s^j\right) \notin A^{fb}$, $a\left(s^j\right) \neq \gamma$ and $\bar{A}$ is not a singleton. Defining $int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$ on $\bar{A}$ we can apply Lemma 3 again. By induction either we find a plan a' such that $\{\psi^*, a'\}$ is admissible with $\lim_{t\to\infty} \pi\left(s^t \in \varpi'_t\right) = 1$, which proves the result; or for some admissible plan a we have that $a\left(s^t\right) \notin A^{fb}$ and $a\left(s^t\right) \notin int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$ for all $s^t \subset s^\infty \in \tilde{S}^\infty$.

It follows that we find a such that $\{\psi^*, a\}$ is admissible and satisfies either (i) $\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1$ or (ii) $a\left(s^t\right) \notin A^{fb}$, $a\left(s^t\right) \notin int\left(\Gamma\left(a\left(s^{t-1}\right), s_t\right)\right)$ for all $s^t \subset s^\infty \in \tilde{S}^\infty$. Let a be an admissible plan that satisfies (ii) and define:

$$B\left(\alpha\right) = \left\{\alpha' \in A : \alpha' > \alpha\right\}$$

where the inequality is taken component-wise.

Lemma 5 *Let $\{\psi^*, a\}$ be an admissible plan. For all $s^{t-1} \in S^{t-1}$ such that $a\left(s^{t-1}\right) \notin A^{fb}$, there exists s_t with*

$$\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing.$$

Proof. By Proposition 2 there exists one s_t and $\varrho > 0$, such that $\varrho a\left(s^{t-1}\right) + (1-\varrho)\alpha \in int(\Gamma(a\left(s^{t-1}\right), s_t))$ for $\alpha \in A^{fb}$. By Assumption 7 and the definition of ϖ_t and A^{fb}, $a\left(s^{t-1}\right) < A^{fb}$ if $a\left(s^{t-1}\right) \notin A^{fb}$ and thus $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ ■

Let $s^t \subset s^\infty \in \tilde{S}^\infty$ for state s_t such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$. Since $a\left(s^t\right) \notin int(\Gamma(a\left(s^{t-1}\right), s_t))$, either $a\left(s^t\right) \in B\left(a\left(s^{t-1}\right)\right)$ or there exists $a'\left(s^t\right) \in B\left(a\left(s^{t-1}\right)\right)$ with $a'\left(s^t\right) > a\left(s^t\right)$ component-wise. For the latter, the steps in Lemma 3 and the previous argument show that there exists a' such that $\{\psi^*, a'\}$ is admissible and $s^t \in \varpi'_t$. Thus we continue with $a\left(s^t\right) \in B\left(a\left(s^{t-1}\right)\right)$ for all states.

Let $s^t \subset s^\infty \in \tilde{S}^\infty$ for state s_t such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$. If such event is recurrent, that is, for any d there exists $t \geq d$ such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ then a that converges to $\gamma = [\bar{a}]_{m,i}$. We use here the fact that the second best is proper, so $A^{fb} \neq \gamma$. Thus there exists a finite d such that $a\left(s^d\right) \in A^{fb}$ and we can apply the previous argument.

Finally we show that the set of histories $s^\infty \in \tilde{S}^\infty$ such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ is not recurrent has measure zero. For a given node $s^d \subset s^\infty \in \tilde{S}^\infty$, let P_j for $j \geq d$ be the probability that event $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ has occurred for some $t \leq j$ and $s^t \in S^t | s^d$. Since $s^d \subset s^\infty \in \tilde{S}^\infty$, $a\left(s^{t-1}\right) \notin A^{fb}$ and by Lemma 5 there exists at least one s_t such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ occurs. By Assumption 1, $\pi\left(s_t\right) \geq \delta > 0$, so

$$P_t \geq P_{t-1} + \delta\left(1 - P_{t-1}\right).$$

Clearly P_t converges to 1 and thus the set of histories such that $\Gamma\left(a\left(s^{t-1}\right), s_t\right) \cap B\left(a\left(s^{t-1}\right)\right) \neq \varnothing$ is not recurrent has measure zero. It follows that $\lim_{t\to\infty} \pi\left(s^t \in \varpi_t\right) = 1$.