NBER WORKING PAPER SERIES

INTERTEMPORAL DISTORTIONS IN THE SECOND BEST

Stefania Albanesi
Roc Armenter

Working Paper 13629
http://www.nber.org/papers/w13629

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2007

We would like to thank Boragan Arouba, Marco Bassetto, Sanjay Chugh, Pierre Yared, and seminar participants at the University of Maryland, Georgetown, FRB New York, the FRB Philadelphia, SED 2007, Ente Einaudi, Universidad Carlos III, and the NBER Summer Institute for helpful comments and suggestions. The views expressed here do not necessarily reflect the views of the Federal Reserve Bank of New York, the Federal Reserve System, or the National Bureau of Economic Research.

Intertemporal Distortions in the Second best
Stefania Albanesi and Roc Armenter
NBER Working Paper No. 13629
November 2007
JEL No. E6,H21

ABSTRACT

We consider a very general class of public finance problems that encompasses the Ramsey model of optimal taxation as well as economies with limited commitment, private information, and political economy frictions, and allows for incomplete markets. We identify a sufficient condition to rule out permanent intertemporal distortions at the optimum. If there exists an admissible allocation that converges to the first best steady state, then all intertemporal distortions are temporary in the second best. We analyze a series of applications to illustrate the significance of this result.

Stefania Albanesi
Columbia University
1022 International Affairs Building
420 West 118th Street
New York, NY 10027
and NBER
sa2310@columbia.edu

Roc Armenter
International Research Function
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
roc.armenter@ny.frb.org

1 Introduction

Perhaps the key lesson in public finance is that distortions should be spread across all margins. Yet, many fiscal policy problems prescribe that distortions on the intertemporal margin ought to be purely temporary, even as other distortions persist. The Ramsey taxation model is the best known example: Chamley (1986) and Judd (1985) first established that optimal capital income taxes are zero in the steady state, so that all intertemporal distortions eventually vanish. This finding has been confirmed for a variety of more general environments.[1] By contrast, permanent intertemporal distortions typically arise in private information economies, as the recent work on dynamic optimal taxation has emphasized.[2]

Why do some policy problems unequivocally rule out permanent intertemporal distortions? What is different about economies where they are optimal? We propose a unified framework to explore these questions for a very broad class of public finance problems for infinite-horizon economies. The class of environments we consider encompasses and generalizes many well known models. We capture policy problems with arbitrary fiscal instruments and environments with primitive constraints on optimal allocations. These include variants of the Ramsey model with aggregate or idiosyncratic risk, as well as economies with incentive compatibility constraints due to limited commitment, political economy frictions, or private information. We also capture settings which combine these constraints with arbitrary asset restrictions such as incomplete markets and borrowing constraints on the government or private agents. This class of problems can be represented as a choice of allocations subject to resource feasibility and a set of additional constraints.[3] We will refer to the corresponding solution as the second best allocation. The choice of allocations subject only to resource feasibility constraints will correspond to our notion of the first best.

Our main contribution is to identify a sufficient condition that rules out permanent intertemporal distortions at the optimum: If there exists an allocation that satisfies all constraints and eventually converges to the first best steady state, then all intertemporal distortions are temporary in the second best. This condition is typically straightforward to verify since there is no need to solve for the second best plan. For example, in a Ramsey model our sufficient condition is satisfied as long as the government can save enough to finance all its expenditures from asset returns. This makes it possible to eventually eliminate distortionary taxes and converge to the first best steady state.

Why does this condition rule permanent intertemporal distortions out? Suppose a candidate optimum at the limiting stationary allocation exhibits an intertemporal wedge. Clearly a reallocation of resources over time would generate first order welfare gains. If our sufficient

[1] Atkeson, Chari and Kehoe (1999) generalize this result for a broad class of deterministic economies. Zhu (1992) and Farhi (2006) show that it also holds with aggregate uncertainty for complete and incomplete markets, respectively.

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

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

condition holds, it is possible to eventually relax all the constraints on optimal policies — this is exactly what converging to the first best amounts to. Then, it is possible to design an *intertemporal* reallocation of resources that improves on the candidate optimum by redistributing *intratemporal* distortions over time without violating any future constraints. Hence, the second best allocation cannot have permanent intertemporal distortions.[4]

Our result provides immediate insight into many policy problems. For example, consider the class of economies where a benevolent government is subject to a limited commitment constraint. In these economies, the continuation value of the second best allocation must exceed the value of an outside option at every date and state. The outside option typically consists of a beneficial temporary deviation followed by a punishment phase[5]. Since by definition the first best improves on any feasible continuation, it must also improve on the outside option. Hence, limited commitment constraints satisfy our sufficient condition and they cannot alone lead to permanent intertemporal distortions.

We also contribute to the understanding of second best problems that do exhibit permanent intertemporal distortions. Economies with private information are one such example. The second best allocation generally displays a positive intertemporal wedge in these economies, as shown in Golosov, Kocherlakota and Tsyvinski (2003). Albanesi and Sleet (2006) and Kocherlakota (2005) characterize the corresponding properties of optimal income taxes. Under private information, our condition is not satisfied, since the first best continuation allocations are typically not incentive compatible: the distortions arising from private information cannot be front-loaded. Hence, the presence of intertemporal distortions is a fundamental feature of this class of economies.

Our result can provide guidance on how to pose normative questions in fiscal policy. The recent research on optimal taxation with private information has emphasized the shortcomings of an approach based on arbitrary fiscal instruments. The advantage of constraints derived from primitives is that relevant trade-offs are not unknowingly left out, leading to greater confidence in the resulting policy prescriptions.[6] For example, the Ramsey model rules out lump sum taxes and prescribes a zero capital tax in the steady state. By contrast, second best allocations under private information generally feature permanent intertemporal distortions. The literature on private information, however, does not isolate the relevant trade-off missing in the Ramsey model. Our result makes it clear that the ability to front-load all distortions in the Ramsey model generates this fundamental difference. This finding provides a deeper understanding of both approaches.[7]

We also propose a weaker version of our sufficient condition which applies only to a subset of public policy problems, namely those in which the path of aggregate variables is not constrained by the evolution of private agents' histories. In this case, in order to rule out permanent intertemporal distortions in the second best, it is sufficient to find an allocation that

[4] The absence of permanent intertemporal distortions does not necessarily imply that capital income taxes are zero as often more than one fiscal system can implement the second best allocation.

[5] "Best sustainable equilibria," first introduced by Chari and Kehoe (1990), are perhaps the best known example of this class of second best problems.

[6] See Werning (2007) for an extensive discussion on this point.

[7] Albanesi and Armenter (2007) provide an extended discussion of optimal capital taxation in Ramsey models.

satisfies all constraints and attains the first best level of *aggregate* capital. When this weaker condition is satisfied in economies with idiosyncratic risk and incomplete insurance, there will be no intertemporal distortions even if the first best continuation is not attainable. Two notable examples are Aiyagari's (1995) optimal taxation problem with borrowing constraints and a version of the political economy model in Acemoglu, Golosov and Tsyvinski (2007). The standard Ramsey model under a balanced-budget constraint also generally satisfies this condition.

The paper proceeds as follows. Section 2 formally characterizes our class of second best problems. Allocations must satisfy feasibility and a set of additional restrictions, which we refer to as *admissibility constraints*. In order to encompass a large class of problems, our formulation for the admissibility constraints is general and abstract. One or more constraints can be present at each node of the economy and they may involve a forward-looking component. They can depend on the physical state of the economy and must be time separable. We also introduce a set of auxiliary variables to capture exogenous restrictions on the asset space or the evolution of costates implied by the admissibility constraints. This allows us to capture most second best problems that admit a primal representation. While working with a general framework, we need to impose a set of regularity conditions. These rule out some economies with differences in preferences between the government and private agents,[8] as well as economies with non-convexities. Section 3 illustrates how three well-known policy problems can be easily captured within our general formulation.

The proof of our result is presented in Section 4 and can be outlined as follows. We split the second best problem into two stages. The first stage takes as given the path for the auxiliary variables and solves for the optimal allocations subject to feasibility and the admissibility constraints. We show that if eventually no future admissibility constraint is binding in the stage one problem, then the optimal allocations will not feature permanent intertemporal distortions. The proof of this result is closely related to Zhu (1992): once all binding admissibility constraints are in the past, the structure of the first order necessary conditions for optimality is the same as in a Ramsey model with complete markets. However, the forward-looking nature of the admissibility constraints implies that the optimal allocations will usually not converge to the first best.

The second stage consists of choosing the optimal path for the auxiliary variables. Our sufficient condition for zero intertemporal distortions comes into play here: if there is an admissible allocation that attains the first best steady state, there must be a corresponding path for the auxiliary variables that eventually fully relaxes all admissibility constraints. We then prove that the solution prescribes a path for the auxiliary variables such that all future admissibility constraints eventually stop binding. Our result follows.

Section 5 illustrates several applications that illustrate the significance of our result. Section 6 discusses a weaker version of our sufficient condition that applies only to a subset of second best problems in our general class. Section 7 concludes.

[8] Acemoglu, Golosov, and Tsyvinski (2007) and Yared (2007) analyze political economy models that allow for differential discounting. We can capture these models in the case in which the government and private agents have the same discount factor.

2 The Model

The economy is characterized by a finite set of exogenous states $s_t \in S$, governed by a Markov probability process $\pi(s_t|s_{t-1})$. Let $s^t = \{s_0, s_1, ..., s_t\}$ be the history of realizations of the exogenous state and S^t the corresponding support. We use $S^d|s^t$ to denote the set of date d histories that are continuation of s^t, i.e., $S^d|s^t = \left\{s^d \in S^d|s^t \subseteq s^d, d \geq 0\right\}$.

The economy is populated by a set I of households distributed according to $\mu = \{\mu_i : i \in I\}$. Let $c_{it}\left(s^t\right)$ and $l_{it}\left(s^t\right)$ denote consumption and leisure of a household of type $i \in I$ at date t after history s^t. Then:

$$\begin{aligned} c_t\left(s^t\right) &= \left\{c_{it}\left(s^t\right) : i \in I\right\}, \\ l_t\left(s^t\right) &= \left\{l_{it}\left(s^t\right) : i \in I\right\}, \end{aligned}$$

and

$$x_t\left(s^t\right) = \left\{c_t\left(s^t\right), l_t\left(s^t\right)\right\}.$$

An allocation is a plan $x = \left\{x\left(s^t\right) : s^t \in S^t, t \geq 0\right\}$.

A constant returns to scale technology combines labor and capital inputs to produce output. Let $k_t\left(s^{t-1}\right) = \left\{k_{it}\left(s^{t-1}\right) : i \in I\right\}$ denote the distribution of capital at a given date, and $k = \left\{k_{t+1}\left(s^t\right) : s^t \in S^t, t \geq 0\right\}$ the corresponding plan. The initial distribution of capital k_0 is taken as given. The resource constraint is given by

$$\int_I \mu_i \left(c_{it}\left(s^t\right) + k_{it}\left(s^t\right)\right) di + g_t\left(s_t\right) \leq F\left(k_t\left(s^{t-1}\right), l_t\left(s^t\right), s_t\right) \tag{1}$$

where $g = \left\{g_t\left(s_t\right) : s^t \in S^t, t \geq 0\right\}$ is an exogenously given plan for government consumption. We assume that g does not generate utility nor enter the production function.

Definition 1 *An allocation $\{x, k\}$ is **feasible** if for all $s^t \in S^t, t \geq 0$ it satisfies the resource constraint (1) and the non-negativity constraints*

$$\begin{aligned} c_{it}\left(s^t\right) &\geq 0, \\ k_{it}\left(s^t\right) &\geq 0, \\ l_{it}\left(s^t\right) &\geq 0, \end{aligned}$$

for all $i \in I$.

The non-negativity constraints define the allocation spaces X and K for $x\left(s^t\right)$ and $k\left(s^{t-1}\right)$ respectively. Both X and K can be taken to be convex and compact. Let $\mathcal{X}$ and $\mathcal{K}$ denote the set of plans x and k in X and K respectively.

Aggregate welfare at node s^t for any allocation plan $x \in \mathcal{X}$ is given by

$$U\left(x, s^t\right) = \sum_{d=t}^{\infty} \sum_{s^d \in S^d|s^t} \beta^{d-t} \pi\left(s^d|s_t\right) \int_I u_i\left(x_{id}\left(s^d\right), s_d\right) di. \tag{2}$$

The above formulation allows us to encompass an arbitrary utilitarian social welfare function as well as settings with ex-post heterogeneity.[9]

With the physical environment given by feasibility and payoffs by (2) we are set to define the first best or unconstrained optimum.

Definition 2 *A feasible allocation* $\left\{x^{fb}, k^{fb}\right\}$ *is* ***first best*** *if* $U\left(x^{fb}, s_0\right) \geq U\left(x, s_0\right)$ *for any feasible allocation* $\{x, k\}$.

It is straightforward to show that the state of the economy $\left\{k_t\left(s^{t-1}\right), s_t\right\}$ is sufficient to characterize first best allocations. We can then define a first best continuation of $\left\{k_t\left(s^{t-1}\right), s_t\right\}$.

Definition 3 *A feasible continuation allocation* $\left\{\tilde{x}, \tilde{k}\right\}$ *is* ***first best continuation of*** $\left\{k_t\left(s^{t-1}\right), s_t\right\}$ *if*

$$\tilde{k}_t\left(s^{t-1}\right) = k_t\left(s^{t-1}\right)$$

and

$$U\left(\tilde{x}, s^t\right) \geq U\left(x', s^t\right)$$

for any feasible continuation allocation $\{x', k'\}$ *with* $k'_t\left(s^{t-1}\right) = k_t\left(s^{t-1}\right)$.

Interesting policy problems typically involve additional constraints on the choice of allocations beyond feasibility. For example, the set of fiscal instruments may be limited or private information may give rise to incentive compatibility constraints. Also, the space of assets that private agents or the government can trade may be restricted. We refer to this class of restrictions as *admissibility constraints.* The second best problem consists of finding feasible allocations that maximize a social objective function subject to the admissibility constraints.

We introduce a general formulation for the admissibility constraints that makes it possible to analyze a large variety of policy problems. We first define an auxiliary variable $a = \left\{a_t\left(s^t\right) \subset A : s^t \in S^t, t \geq 0\right\}$ where A is a convex and compact subset of $\Re^m$. This variable allows us to capture arbitrary constraints on the asset space as well as restrictions on the evolution of costates that endogenously arise within the problem. The admissibility constraints can then be expressed as:

$$H\left(x, k_t\left(s^{t-1}\right), s_t\right) \leq a_t\left(s^t\right), \tag{3}$$

$$a_{t+1}\left(s^{t+1}\right) \in \Gamma\left(a_t\left(s^t\right), s_t\right), \tag{4}$$

for all $s^t \in S^t, t \geq 0$, and

$$a_0\left(s_0\right) \in A_0.$$

Here H is a twice differentiable, bounded function unto $\Re^m$, $\Gamma\left(a_t\left(s^t\right), s_t\right)$ is a correspondence unto $\Re^m$, and A_0 is a subset of $\Re^m$.

We can now formally define the notion of admissible allocations and second best.

[9] We have excluded economies with history dependence in preferences for simplicity. The environment can be extended to some non time-separable preferences, such as habit formation or Kreps-Portheus utility.

Definition 4 *A feasible allocation $\{x, k\}$ is **admissible** if there exists a plan for the auxiliary variable a such that*

1. *The allocation $\{x, k\}$ satisfies (3) for all $s^t \in S^t, t \geq 0$,*
2. *The auxiliary variable a satisfies (4) for all $s^{t+1} \in S^{t+1}$, $t \geq 0$, and $a_0(s_0) \in A_0$.*

Definition 5 *An admissible allocation $\{x^*, k^*\}$ is **second best** if $U(x^*, s_0) \geq U\left(x, s^0\right)$ for any admissible allocation $\{x, k\}$.*

The class of second best problems we consider is restricted by two defining properties of the admissibility constraints. These properties are stated in terms of functions $\{H,\ \Gamma\}$.

First, the admissibility constraints must display a limited degree of history dependence. Formally, allocations from nodes arbitrarily far in the past or outside the set of continuation histories do not affect the admissibility constraints. So, while we allow a finite number of past or non-continuation allocations to be restricted, the admissibility constraint must be eventually forward-looking. In addition, admissibility constraints must be time-separable with the same intertemporal discount rate as in (2), again after a finite number of periods.

Condition 6 (*Limited History Dependence*) *There exists a finite $d \geq 0$ such that for any admissible allocation $\{x, k\}$, any node s^j, and any date $t \geq j + d$, each admissibility constraint $n \leq m$ satisfies:*

1. ***Forward-looking:*** *if $s^t \notin S^t|s^j$,*

$$D_{x_t(s^t)} H_n\left(x, k_j\left(s^{j-1}\right), s_j\right) = \vec{0}. \tag{5}$$

2. ***Time separability:*** *if $s^t \in S^t|s^j$,*

$$D_{x_t(s^t)} H_n\left(x, k_j\left(s^{j-1}\right), s_j\right) = \beta^{t-j} \pi\left(s^t|s_j\right) r_n\left(x_j\left(s^j\right), s_j\right) h_n\left(x_t\left(s^t\right), s_t\right) \tag{6}$$

for some functions r_n and h_n.

Condition 6 restricts the class of problems to which we can apply our results. For example, some second best problems impose unlimited history dependence. A trivial case is a Ramsey model in which taxes are restricted to be constant. The decision at the initial date constraints allocations arbitrarily far in the future and the problem does not satisfy (5). Limited commitment constraints under heterogeneous discount rates or non-time separable preferences often fail to satisfy time separability in the form of (6). Condition 6 *does not*, however, restrict the second best plan to exhibit limited history dependence.

The second defining restriction on the class of admissibility constraints concerns the law of motion of the auxiliary variable, leading to conditions on the correspondence Γ.

Condition 7 (*Convexity of* Γ) *The correspondence $\Gamma\left(a_t\left(s^t\right), s_t\right) : A \times S \rightrightarrows A$ is continuous, convex, and its image is a convex subset of $\Re^m$ including $a_t\left(s^t\right)$ for all $s_t \in S$ and $a_t\left(s^t\right) \in A$.*

The class of second best problem defined by conditions 6 and 7 is very broad. It includes well known public finance problems with and without permanent intertemporal distortions, as well as interesting new environments stemming from combinations of the typical constraints. In Section 3, we illustrate how several benchmark second best problems can be adapted to our formulation.

2.1 Regularity Conditions

We now impose a number of regularity conditions on the problem in order to guarantee tractability. These conditions allow us to use Lagrangian methods to characterize optimal allocation and are standard in the literature.

We begin by stating a set of conditions on primitives.

Condition 8 *Functions $\{U, F\}$ are bounded, twice continuously differentiable with bounded first-order derivatives in $\mathcal{X}$ and $\mathcal{K}$. Moreover,*

1. *U is strictly increasing in consumption and leisure, and strictly concave.*
2. *F is strictly increasing in labor and capital, homogeneous of degree 1, and concave.*

We state the remaining regularity conditions as properties of the second best allocation. Ideally we would like to express these conditions in terms of the primitives of the problem. The generality of our environment, though, makes this a very cumbersome task. These regularity conditions can be restated in terms of primitives for each specific applications.

First, we impose two conditions that enable us to apply Lagrangian methods. The first is an interiority restriction.

Condition 9 *There exist a second best allocation $\{x^*, k^*\}$ such that for all nodes $s^t \in S^t$, $t \geq 0$, it lies in the interior of $X \times K$.*

The second is a non-degenerate constraint qualification.

Condition 10 *For a second best allocation $\{x^*, k^*\}$, if $\{1, 2, ..., \tilde{m}\}$ admissibility constraints are binding at node s^t, then the Jacobian*

$$D_{[x_t(s^t), k_t(s^{t-1})]} \tilde{H}\left(x^*, k_t^*\left(s^{t-1}\right), s_t\right)$$

has full rank, where $\tilde{H}$ is given by the $\{1, 2, ..., \tilde{m}\}$ constraints.

Next, we impose a condition that guarantees the admissibility constraints are well-behaved.

Condition 11 *There exists a second best allocation $\{x^*, k^*\}$ such that for any node $s^t \in S^t, t \geq 0$, any admissibility constraint $n \leq m$, and any first-best continuation allocation $\left\{x^{fb}, k^{fb}\right\}$ of $\left\{k_t^*\left(s^{t-1}\right), s_t\right\}$, if*

$$H_n\left(x^*, k_j^*\left(s^{j-1}\right), s_j\right) \geq H_n\left(x^{fb}, k_j^{fb}\left(s^{j-1}\right), s_j\right),$$

then admissibility constraint n is not binding at $s^j \in S^j | s^t$, $j \geq t$.

This condition imposes that, if the value of a given admissibility constraint reaches past the first best, then the constraint will not be binding. Jointly with our other regularity conditions, it ensures that first order necessary conditions are sufficient. Given that we allow the admissibility constraints to be non-convex, this restriction is quite weak and rules out only very special cases.

Finally, we require that the second best allocations approach a stationary limiting probability distribution, P_x^∞. The limiting distribution need *not* be unique and, indeed, many policy problems in our class feature more than one stationary distribution. We let $\mathcal{P}_x^\infty = \{P_{x1}^\infty, P_{x2}^\infty, ...\}$ denote the collection of stationary distributions for plan x.

Condition 12 *A second best allocation $\{x^*, k^*\}$ converges to a stationary distribution with probability one.*

Appendix A.1 provides a formal definition of the stationary probability distribution P_x^∞, as well as a more formal restatement of Condition 12.

Condition 12 jointly with the regularity conditions on the primitives of our second best problems ensure that the Lagrangian multipliers on the resource constraint and on the admissibility constraints also converge to a stationary distribution with probability one.

3 Examples

This section illustrates how three benchmark optimal policy problems fit into the general formulation of the second best problem developed above.

3.1 Simple Ramsey Model

We first consider a dynamic version of the Ramsey (1927) model of optimal taxation. Chari and Kehoe (2001) review the macroeconomic applications of this paradigm. Here, we describe the simplest version with a representative agent and no uncertainty.

The government chooses taxes to finance an exogenously given sequence of government consumption. The main assumption is that lump sum taxes are not available. The government can set proportional taxes on labor income τ_t^l and on capital income τ_t^k in each period. It can also make non-negative transfers T_t and issue bonds, b_{t+1}, paying off one unit of consumption at time $t+1$ and traded at price q_t. The government's flow budget constraint is:

$$\tau_t^l w_t l_t + \tau_t^k r_t k_t + q_t b_{t+1} \geq b_t + g_t + T_t,$$

where w_t and r_t denote the rental rate of labor and capital, respectively. Iterating leads to the following *intertemporal* budget constraint:

$$\sum_{t=0}^{\infty} q_t^0 \left(\tau_t^l w_t l_t + \tau_t^k r_t k_t - g_t - T_t \right) \geq b_0, \tag{7}$$

where q_t^0 denotes the price of date t consumption and b_0 is the initial stock of debt, which is taken as given.

We can define a competitive equilibrium as a policy $\left\{b_t, \tau_t^l, \tau_t^k, T_t\right\}_{t=0}^{\infty}$, prices $\{w_t, r_t, q_t\}_{t=0}^{\infty}$ and an allocation $\{c_t, l_t, k_t, g_t\}_{t=0}^{\infty}$, such that the allocation is optimal for the households and the firms given the policies and prices, all markets clear, and the government budget constraint (7) and the resource constraint (1) are satisfied. A Ramsey equilibrium for this economy is simply the competitive equilibrium that maximizes the representative agent's welfare from the standpoint of time 0.

The first step in the analysis consists in restating the problem in its primal form, that is as a choice of allocations rather than taxes. Substituting the equilibrium optimality conditions:

$$\begin{aligned} q_t &= \beta^{t+1} \frac{u_{t+1}^c}{u_0^c}, \\ \frac{-u_t^l}{u_t^c} &= \left(1 - \tau_t^l\right) w_t, \\ \frac{u_t^c}{\beta u_{t-1}^c} &= \left[r_t + (1-\delta)\right] \left(1 - \tau_t^k\right), \\ r_t &= F_t^k, \qquad w_t = F_t^l, \end{aligned}$$

into the intertemporal budget constraint yields:

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

This inequality is known as *implementability constraint* and it captures all the restrictions on the government's choice of allocations in addition to feasibility for this model and hence defines the set of admissible allocations. The forward looking nature of this constraint reflects the government's ability to allocate taxes over time by borrowing or saving.

To express the implementability constraint in terms of our general formulation, set:

$$\begin{aligned} H\left(x, s_0\right) &= \sum_{t=0}^{\infty} \beta^t \left(u_t^c c_t + u_t^l l_t\right), \\ a_0 &= u_0^c \left\{ \left[\left(1 - \tau_0^k\right) F_0^k \right] k_0 + b_0 \right\}. \end{aligned}$$

The variable a_0 represents the initial value of assets. Typically, its value is exogenously restricted to exclude a solution in which the government sets τ_0^k high enough to pay all outstanding debt and raise enough assets so that that no distortionary taxes need to be imposed in all future periods. A constraint on the initial value of τ_0^k can be translated into an initial condition for a_0. For this problem, there are no admissibility constraints at any future date and thus no auxiliary variables are needed.[10] The resulting formulation of the second best problem satisfies Condition 6 restricting the history dependence of admissibility constraints.

[10] Incomplete factor taxation models, like Correia (1996) and Jones, Manuelli and Rossi (1997), assume there is an additional factor of production that cannot be taxed. These models cannot be formulated with an implementability constraint of the type above if we just fix the capital tax at date $t = 0$. However, if we prevent the government from any manipulation of the present value of assets at date $t = 0$, we then recover an implementability constraint of the form above. The formulation at date $t = 0$ has implications for the long-run capital tax. See Armenter (2007) for a discussion.

3.2 Optimal Policies with Limited Commitment

The Ramsey model assumes that the policy is chosen once and for all in the initial period. This requires a commitment technology to guarantee that the government will not revise policies at a future date. This is a very strong assumption and several frameworks have been developed to study optimal policies when a commitment technology is not available.

Reis (2006) studies a variant with limited commitment. The physical environment and the set of fiscal instruments are the same as in the simple Ramsey model described in section 3.1. The choice of policies is modelled as a game where both the government and private agents make *sequential* decisions in every period. In each period, both the government and private agents can default on outstanding debt obligations. The government decides how much to punish households who defaulted in the previous period, chooses taxes and transfers for the current period and decides whether to default on outstanding debt. Households then choose consumption, labor, capital and whether to default on debt. Finally, markets meet and clear.

In a sustainable equilibrium for this game,[11] equilibrium bond repayment is supported by the threat of reversion to the worst sustainable equilibrium, in which the government is forced to run a balanced budget and applies confiscatory tax rates on capital. The best sustainable equilibrium is the one that maximizes the representative agents' lifetime utility and corresponds to our notion of second best plan.

The sustainability constraint requires the continuation value of the government's utility on the equilibrium path to exceed the continuation value of the worst sustainable equilibrium. This introduces an additional admissibility constraint in each period relative to the simple Ramsey model. The value of default only depends on the level of capital at the time of default and can be expressed as $V^{def}\left(k_t\left(s^{t-1}\right), s_t\right)$.[12] The sustainability constraint can then be written as:

$$U\left(x, s^t\right) \geq V^{def}\left(k_t\left(s^{t-1}\right), s_t\right).$$

This formulation can be adapted to our framework by setting:

$$H\left(x, k_t\left(s^{t-1}\right), s^t\right) = V^{def}\left(k_t\left(s^{t-1}\right), s_t\right) - U\left(x, s^t\right),$$

which gives rise to the admissibility constraint of the form of (3) by setting $a_t\left(s^t\right)$ to 0 at all periods. The time separability of preferences and the fact that the outside option only depends on the continuation allocation and the current state implies that Condition 6 is satisfied. In section 5.2, we return to the class of policy problems with limited commitment constraints and apply our result to derive a key property of the second best allocation.

3.3 Ramsey Model with Incomplete Markets

In the previous examples, the admissibility constraints are captured by the functional H in (3). We now consider the role of the auxiliary variables, a, and the constraint on their law of

[11] See Chari and Kehoe (1990).

[12] Reis (2006) analyzes a version of the model with no uncertainty. Here, we allow for aggregate shocks, captured by the variable s_t.

motion Γ in (4). Their primary purpose is to capture constraints stemming from market incompleteness. To illustrate, we concentrate on the version of the Ramsey model with incomplete markets and no capital analyzed by Aiyagari, Marcet, Sargent and Seppala (2003).

The economy is similar to the simple Ramsey model. There are aggregate shocks but there is no capital. Importantly, bond returns cannot be contingent on the state of the economy, so that markets are incomplete. Formally, bond repayments at time t are not measurable with respect to s^t and only depend on s^{t-1}, and will be denoted with $b_t\left(s^{t-1}\right)$. Government policy is given by $\left\{b_t\left(s^{t-1}\right), \tau_t^l\left(s^t\right), T_t\left(s^t\right)\right\}_{t\geq 0}$.

Assuming utility is quasi-linear in consumption, we can write the government's present value budget constraint for all nodes $s^t \in S^t, t \geq 0$:

$$\sum_{j=t}^{\infty} \sum_{s^j \in S^j|s^t} \beta^{j-t} \pi\left(s^j|s_t\right)\left(z\left(x_j\left(s^j\right), s_j\right) - T_j\left(s^j\right)\right) = b_t\left(s^{t-1}\right), \tag{8}$$

where $z\left(x_t\left(s^t\right), s_t\right)$ denotes labor income tax revenues net of government consumption at node s^t. The government also faces the constraint that transfers must be non-negative:

$$T_t\left(s^t\right) \geq 0. \tag{9}$$

To adapt the admissibility constraint to our general formulation, define:

$$V_t\left(s^t\right) = \sum_{j=t}^{\infty} \sum_{s^j \in S^j|s^t} \beta^{j-t} \pi\left(s^j|s_t\right) T_t\left(s^j\right). \tag{10}$$

The variable $V\left(s^t\right)$ corresponds to the present discounted value of transfers at node s^t. Constraint (9) requires:

$$V_t\left(s^t\right) \geq \beta V_t\left(s^{t+1}\right) \geq 0, \tag{11}$$

for all $s^t \in S^t, t \geq 0$.

Define:

$$a_t\left(s^t\right) = \begin{bmatrix} V_t\left(s^t\right) + b_t\left(s^{t-1}\right) \\ -V_t\left(s^t\right) - b_t\left(s^{t-1}\right) \\ b_{t+1}\left(s^t\right) \\ V_t\left(s^t\right) \end{bmatrix}, \tag{12}$$

and

$$\begin{aligned} H_1\left(x, s^t\right) &= \sum_{j=t}^{\infty} \sum_{s^j \in S^j|s^t} \beta^{j-t} \pi\left(s^j|s_t\right) z\left(x_t\left(s^j\right), s_j\right), \\ H_2\left(x, s^t\right) &= -H_1\left(x, s^t\right). \end{aligned} \tag{13}$$

Then, constraint (8) can be expressed as:

$$\begin{aligned} H_1\left(x, s^t\right) &\leq a_1\left(s^t\right), \\ H_2\left(x, s^t\right) &\leq a_2\left(s^t\right). \end{aligned} \tag{14}$$

Market incompleteness and the non-negativity of transfers (11) translate into constraints on the correspondence $\Gamma\left(a\left(s^t\right), s_t\right)$ in the first stage problem. The fact that (8) must hold with equality gives rise to the constraint:

$$a_1\left(s^{t+1}\right) = -a_2\left(s^{t+1}\right).$$

The non-measurability of bond returns requires:

$$a_1\left(s^{t+1}\right) = a_3\left(s^t\right) + a_4\left(s^{t+1}\right).$$

Finally, the non-negativity of transfers leads to the constraint:

$$a_4\left(s^{t-1}\right) \geq \beta a_4\left(s^t\right) \geq 0.$$

As in the Ramsey problem with complete markets, time separability of preferences guarantees that Condition 6 is satisfied. Moreover, since Γ is linear in a, Condition 7 also holds for this problem. If additional constraints, such as borrowing constraints or asset limits are imposed on the government's problem, these can be easily captured by augmenting the correspondence Γ.

4 Main Result

Our main result is that if there exists an allocation that converges to the first best steady state, then there are no permanent intertemporal distortions in the second best. If the sufficient condition is satisfied, all distortions can be front-loaded, that is, all admissibility constraints can eventually be relaxed. This allows distortions and resources to be reallocated intertemporally without tightening future constraints.

We start with a simple Lemma about first best allocations which enables us to formally state our condition.

Lemma 13 *A first best allocation x^{fb} converges to a unique stationary distribution P_{fb}^{∞}.*

This property is a straightforward implication of our regularity conditions on preferences and technology.

We now state the sufficient condition.

Condition 14 *For any initial conditions $\{k_{-1}, a_0\} \in K \times A$, there exists an admissible allocation $\{x, k\}$ which converges to the first best limiting distribution with probability one, i.e.,*

$$\lim_{t\to\infty} \Pr\left(x_t \in B\right) = P_{fb}^{\infty}\left(B\right)$$

for all measurable $B \subseteq X$.

The connection of Condition 14 with the ability to front-load all distortions is clear: an admissible allocation which eventually attains the first best has the property that all distortions have indeed been front-loaded.

Condition 14 is very easy to verify, since it just requires finding one admissible allocation that converges to the first best limiting distribution. This allocation does not need to satisfy any optimality conditions and can follow any arbitrary transition path provided it is admissible. We do not need, thus, to solve for the second best plan to know whether Condition 14 holds.

We now proceed to state the main result of the paper.

Theorem 15 *Let Condition 14 hold and* $\{x^*, k^*\}$ *be a second best allocation. Then for any* $P^\infty \in \mathcal{P}^\infty_{x^*}$ *and* $i \in I$ *either*

1. *There is no intertemporal distortion with probability* 1 *in the limit,*

$$P^\infty \left(u_i^c \left(s^t \right) = \beta \sum_{s^{t+1}} \pi \left(s^{t+1} | s^t \right) u_i^c \left(s^{t+1} \right) F_i^k \left(s^{t+1} \right) \right) = 1; \tag{15}$$

or

2. *The intertemporal distortion fluctuates around* 0,

$$\begin{aligned} P^\infty \left(u_i^c \left(s^t \right) > \beta \sum_{s^{t+1}} \pi \left(s^{t+1} | s^t \right) u_i^c \left(s^{t+1} \right) F_i^k \left(s^{t+1} \right) \right) &> 0, \\ P^\infty \left(u_i^c \left(s^t \right) < \beta \sum_{s^{t+1}} \pi \left(s^{t+1} | s^t \right) u_i^c \left(s^{t+1} \right) F_i^k \left(s^{t+1} \right) \right) &> 0. \end{aligned} \tag{16}$$

Proof. In Appendix ■

The formal statement of the result clarifies the sense in which there are no *permanent* intertemporal distortions.[13] First, the result holds for the stationary second best allocations. The absence of permanent intertemporal distortions does not rule out significant and prolonged distortions on the transition path. Second, Theorem 15 allows for the optimal allocations to be distorted in the limit. For one, the intertemporal distortion could be fluctuating around 0 instead of being identically 0, if case two prevails. The unconditional expectation may be quite different from zero. Theorem 15 only asserts that the distortion cannot be strictly positive with probability one or, for that matter, strictly negative. Finally, the absence of permanent intertemporal distortions does not necessarily translate into a prescription for zero capital income taxes. For example, Kocherlakota (2005) presents a private information economy with

[13] Intertemporal distortions may be defined in any economy where private agents or the government face dynamic choices, even if no physical technology for capital accumulation is available. In this case, one can define a social discount factor which equates the marginal utility of current consumption to the expected marginal utility of future consumption for all agents. The social discount factor in the first best will satisfy:

$$P^\infty_{fb} \left(u^c_{i,t} \left(c^{fb}_{i,t}, l^{fb}_{i,t} \right) = \beta R^{fb}_t E u^c_{i,t+1} \left(c^{fb}_{i,t+1}, l^{fb}_{i,t+1} \right) \right) = 1.$$

permanent intertemporal distortions where the expected marginal tax on capital income is zero. On the other hand, Aiyagari (1995) is an example of an economy with no permanent intertemporal distortions where a positive capital tax is needed in the second best.

In general, the limiting second best allocation will display permanent intratemporal distortions. For example, the Ramsey equilibrium with complete markets does not feature any intertemporal distortions but typically displays a positive labor income tax in the steady state which translates into a permanent wedge on the intratemporal margin.

We prove Theorem 15 in several steps described in the remainder of this Section. We start by splitting the second best problem into two stages. The first stage solves for the allocations given a path for the auxiliary variables. The second stage solves for the optimal path for the auxiliary variables and hence characterizes the second best plan. We show in the first stage that if future admissibility constraints eventually do not bind, then there are no permanent intertemporal distortions. This result corresponds to Proposition 18. In the second stage, we show that if the sufficient condition holds, then the optimal path for the auxiliary variables has the property that future admissibility constraint eventually will not bind.

We now provide a formal statement of the two stages that comprise the second best problem. Let $\mathcal{A}$ be the set of auxiliary variable plans in A.

Definition 16 *Let $W(a) : \mathcal{A} \to \Re$ be given by*

$$W(a) = \max_{\{x,k\}} U(x, s_0) \qquad \text{(Problem 1)}$$

subject to (1) and (3) for all $s^t \in S^t$, $t \geq 0$.

The existence of a bounded function $W(a)$ is granted by Condition 8 and $X \times K$ being a compact space.

The second stage problem then consists in choosing the auxiliary variables to maximize $W(a)$.

Proposition 17 *Let a^* solve*

$$\sup_{a \in \mathcal{A}} W(a) \qquad \text{(Problem 2)}$$

subject to (4) for all $s^t \in S^t$, $t \geq 0$ and $a_0(s_0) \in A_0$. Then an admissible allocation x^* is second best if and only if $W(a^*) = U(x^*, s_0)$.

Proof. Straightforward ■

4.1 First Stage: Choosing Allocations

We study Problem 1 given plan a^*. Conditions 8 and 10 allow us to write the Lagrangian to characterize $W(a)$:

$$\begin{aligned}\mathcal{L} &= U(x, s_0) \\ &\quad - \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \pi\left(s^t | s_0\right) \lambda\left(s^t\right) \left\{ \int_I \mu_i \left(c_{it}\left(s^t\right) + k_{it}\left(s^t\right)\right) di + g_t\left(s_t\right) - F\left(k_t\left(s^{t-1}\right), l_t\left(s^t\right), s_t\right) \right\} \\ &\quad - \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \pi\left(s^t | s_0\right) \phi\left(s^t\right)' \left\{ H\left(x, k_t\left(s^{t-1}\right), s_t\right) - a_t\left(s^t\right) \right\}.\end{aligned}$$

where $\lambda\left(s^t\right) \geq 0$ and $\phi\left(s^t\right) \in \Re_+^m$ are the Lagrangian multipliers, normalized by $\beta^t \pi\left(s^t | s_0\right)$, for the resource constraint (1) and the system of admissibility constraints (3) at each node s^t.

We adopt the following notation for derivatives:

$$\begin{aligned}u_i^c\left(s^t\right) &= \frac{\partial u_i\left(x_{it}\left(s^t\right), s_t\right)}{\partial c_{it}\left(s^t\right)}, \\ u_i^l\left(s^t\right) &= \frac{\partial u_i\left(x_{it}\left(s^t\right), s_t\right)}{\partial l_{it}\left(s^t\right)}, \\ F_i^l\left(s^t\right) &= \frac{\partial F\left(k_t\left(s^{t-1}\right), l_t\left(s^t\right), s_t\right)}{\partial l_{it}\left(s^t\right)}, \\ F_i^k\left(s^t\right) &= \frac{\partial F\left(k_t\left(s^{t-1}\right), l_t\left(s^t\right), s_t\right)}{\partial k_{it}\left(s^t\right)},\end{aligned}$$

and

$$H_i^c\left(s^t, s^j\right) = \begin{bmatrix} \frac{\partial H_1\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial c_{it}\left(s^j\right)} \\ \frac{\partial H_2\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial c_{it}\left(s^j\right)} \\ \cdots \\ \frac{\partial H_m\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial c_{it}\left(s^j\right)} \end{bmatrix},$$

and similarly for $H_i^l\left(s^t, s^j\right)$; and

$$H_i^k\left(s^t\right) = \begin{bmatrix} \frac{\partial H_1\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial k_{it-1}\left(s^{t-1}\right)} \\ \frac{\partial H_2\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial k_{it-1}\left(s^{t-1}\right)} \\ \cdots \\ \frac{\partial H_m\left(x, k_t\left(s^{t-1}\right), s_t\right)}{\partial k_{it-1}\left(s^{t-1}\right)} \end{bmatrix}.$$

Here, $H_i^c\left(s^t, s^j\right)$ and $H_i^l\left(s^t, s^j\right)$ are indexed by only two nodes but they may depend on allocations evaluated at other nodes.

The first order conditions with respect to $k_{it}\left(s^t\right)$, $c_{it}\left(s^t\right)$, and $l_{it}\left(s^t\right)$, respectively are:

$$\mu_i \lambda\left(s^t\right) = \sum_{s^{t+1}} \beta \pi\left(s^{t+1} | s_t\right) \left(\lambda\left(s^{t+1}\right) F_i^k\left(s^{t+1}\right) - \phi\left(s^{t+1}\right)' H_i^k\left(s^{t+1}\right) \right), \tag{17}$$

$$\beta^t \pi\left(s^t|s_0\right)\left\{u_i^c\left(s^t\right)-\mu_i\lambda\left(s^t\right)\right\}=\sum_{j=0}^{\infty}\sum_{s^j}\beta^j\pi\left(s^j|s_0\right)\phi\left(s^j\right)'H_i^c\left(s^j,s^t\right), \tag{18}$$

$$\beta^t \pi\left(s^t|s_0\right)\left\{u_i^l\left(s^t\right)-\mu_i\lambda\left(s^t\right)F_i^l\left(s^t\right)\right\}=\sum_{j=0}^{\infty}\sum_{s^j}\beta^j\pi\left(s^j|s_0\right)\phi\left(s^j\right)'H_i^l\left(s^j,s^t\right). \tag{19}$$

The first order conditions for this problem are necessary but, without further structure on the choice set, they are generally not sufficient.

The first important result is that, if future admissibility constraints stop binding, the allocation that solves the first stage problem will not feature permanent intertemporal distortions.

Proposition 18 *Let allocation $\{x,k\}$ solve Problem 1 for a given $a\in\mathcal{A}$. If for $P^\infty\in\mathcal{P}_x^\infty$*

$$P^\infty\left(\phi\left(s^t\right)=0\right)=1$$

then for all $i\in I$ either

1. *There is no intertemporal distortion with probability 1 in the limit and (15) holds;*

   or

2. *The intertemporal distortion fluctuates around 0 and (16) holds.*

Proof. In the Appendix ■

The key to the proof is that when future admissibility constraints stop binding, only the history of past binding constraints matters. Then, the system of equations that characterize the optimal allocation from that node onwards has the same structure as in a Ramsey problem with complete markets.[14] The allocations are a function of the state of the economy $\left\{k_t\left(s^{t-1}\right),s_t\right\}$ and a summary statistic for the history of binding constraints. The rest of the proof is similar to Zhu (1992), who shows that in a Ramsey model with complete markets at the limiting distribution the expected tax on capital income is zero. This corresponds to the absence of permanent intertemporal distortions. In economies with a representative agent, the occurrence of (15) or (16) solely depends on the properties of preferences. With heterogeneous agents, additional factors, such as the presence of borrowing constraints, also play a role.[15]

Before we move to the second stage problem, we derive two important properties of the value function $W(a)$. The first is weak quasi-concavity, which follows from the structure of Problem 1 without further assumptions. The second property amounts to strict quasi-concavity of $W(a)$ in a neighborhood of the first best plan.

Proposition 19 *Let $\varepsilon=\left\{\varepsilon\left(s^t\right):s^t\in S^t,t\geq 0\right\}$. Then the following is true:*

1. *If ε is non-negative everywhere, then $W(a+\varepsilon)\geq W(a)$,*

2. *If ε is strictly positive everywhere and $W(a+\varepsilon)=W(a)$, then $W(a)=U\left(x^{fb},s_0\right)$.*

Proof. In the Appendix ■

[14] The limits on history dependence imposed by Condition 6 are important for this step. Without them, past constraints could arbitrarily dictate the exact path of all future allocations and the result would not hold.

[15] See Section 6 for a discussion.

4.2 Second Stage: Choosing the Auxiliary Variable

We now study Problem 2 – the choice of the auxiliary variables a^*. Using the properties of Γ and the sufficient condition 14, we show that the auxiliary variable plan a^* converges to a subset of A where no admissibility constraints bind.

We start by characterizing the set of values for the auxiliary variables that can support a continuation first best plan. Such a set depends on the state of the economy $\left\{k_t\left(s^{t-1}\right), s_t\right\}$.

Definition 20 *Let* $\bar{a}\left[k_t\left(s^{t-1}\right), s_t\right] : K \times S \to \mathcal{A}$ *be such that*

$$\bar{a}\left[k_t\left(s^{t-1}\right), s_t\right]\left(s^j\right) = H\left(\tilde{x}, \tilde{k}_j\left(s^{j-1}\right), s_j\right)$$

for all $s^j \in S^j|s^t$ *where* $\left\{\tilde{x}, \tilde{k}\right\}$ *is a first-best continuation allocation of* $\left\{k_t\left(s^{t-1}\right), s_t\right\}$.

Any plan with $a\left[k_t\left(s^{t-1}\right), s_t\right] \geq \bar{a}\left[k_t\left(s^{t-1}\right), s_t\right]$ can sustain the first best allocations from s^{t-1} onwards. However, we also need the auxiliary plan $\bar{a}$ to be admissible itself, i.e., to satisfy (4) at all nodes.

Definition 21 *Let* $A^{fb}\left[k_t\left(s^{t-1}\right), s_t\right]$ *be the set of* $a_t\left(s^t\right) \in A$ *such that there exists a plan* a' *with*

$$a'\left(s^j\right) \geq \bar{a}\left[k_t\left(s^{t-1}\right), s_t\right]\left(s^j\right)$$

for all $s^j \in S^j|s^t$, $a'_t\left(s^t\right) = a_t\left(s^t\right)$, *and* a' *satisfies (4) for all* $s^j \in S^j|s^t$, $j \geq t$.

Note the sets $A^{fb}\left[k_t\left(s^{t-1}\right), s_t\right]$ are compact rectangles in A — recall A is compact itself.

Let K^{fb} denote the support of the capital allocation at the first best stationary distribution.[16] Condition 14 guarantees that, for all $\left\{k_t\left(s^{t-1}\right), s_t\right\} \in K^{fb} \times S$, the set $A^{fb}\left[k_t\left(s^{t-1}\right), s_t\right]$ is not empty. We can then define the subset in A that supports allocations at the first best limiting stationary distribution

$$\bar{A}^{fb} = \cup_{K^{fb} \times S} A^{fb}\left[k_t, s_t\right].$$

There is little we can say about the set $\bar{A}^{fb}$ as we allowed H to be non-convex.

The next Proposition states that if the second best path for the auxiliary variables a^* reaches the set $\bar{A}^{fb}$, then no admissibility constraint is ever binding at all continuation nodes.

Proposition 22 *Let* $a^*\left(s^t\right) \in \bar{A}^{fb}$. *Then the admissibility constraints (3) are not binding at any node* $s^j \in S^j|s^t, j \geq t$.

Proof. In the Appendix ■

To prove Proposition 22 we make use of Condition 11. Without it we could not navigate through the potentially complex set $\bar{A}^{fb}$. An alternative would be to impose enough structure on the primitives such that $\bar{A}^{fb}$ is convex and compact. The conditions needed would be considerably more restrictive than Condition 11.

[16] A formal definition is available in the Appendix.

There are only two steps left to finish the proof: to show that we can and we want to converge to $\bar{A}^{fb}$. Proposition 23 establishes a key property of Γ, akin to monotonicity. Proposition 23 implies we can always find an admissible plan for the auxiliary variables which converges to $\bar{A}^{fb}$ almost surely.

Proposition 23 *Let 14 hold and $a_t\left(s^t\right) \in A$. Then for some non-negative scalar $\alpha < 1$ and $\bar{a}_{t+1}\left(s^{t+1}\right) \in \bar{A}^{fb}$,*

$$\alpha a_t\left(s^t\right) + (1-\alpha)\,\bar{a}_{t+1}\left(s^{t+1}\right) \in \Gamma\left(a_t\left(s^t\right), s_t\right). \tag{20}$$

Proof. In the Appendix ■

If property (20) did not hold, it would be possible to find a separating hyperplane between sets $\Gamma\left(a_t\left(s^t\right), s_t\right)$ and $\bar{A}^{fb}$. The convexity of Γ, as stated in Condition 7, would then imply that no point contained in the half-space containing $\Gamma\left(a_t\left(s^t\right), s_t\right)$ could be the starting point of a path leading to $\bar{A}^{fb}$. This would contradict Condition 14.

The final step in the argument is to show that the second best plan indeed converges to the set $\bar{A}^{fb}$. We can use Proposition 23 to show that, if a candidate second best path for a does not converge to the set $\bar{A}^{fb}$, it is possible to construct an alternative a' such that $a'_t\left(s^t\right) \geq a_t\left(s^t\right)$ for all nodes $s^t \in S^t$, $t \geq 0$. Plan a' weakly improves upon a by the quasi-concavity of value function $W(a)$ —established by Proposition 19.

Proposition 24 *Let 14 hold and $\{x^*, k^*\}$ be a second best allocation. Then there exists an auxiliary variable a^* such*

$$\lim_{t\to\infty} \Pr\left(a_t^* \in \bar{A}^{fb}\right) = 1.$$

Proof. In the Appendix ■

All the pieces for the proof of Theorem 15 are now in place. For any second best allocation plan, Proposition 24 implies that the auxiliary variables converge to the set $\bar{A}^{fb}$. By Proposition 22, eventually no future admissibility constraint is binding and the proof of Theorem 15 follows from 18. The formal proof is in the Appendix.

5 Applications

We now describe how our result can be used in a series of applications.

5.1 Ramsey Models with Asset Constraints

Condition 14 in the simple Ramsey model with a representative agent is quite simple. It amounts to the government's ability to accumulate enough assets such that it can finance government consumption via the interest revenues. In a closed economy, this is possible only if private agents can borrow so that the government can accumulate enough assets.

It is simple to characterize the level of government assets that supports the continuation first best. This level depends on the outstanding capital stock and can be calculated from the intertemporal government budget constraint, that is, the admissibility constraint, at a history k_t :

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

where all derivatives are evaluated at the first best allocations. Recall from Section 3 that $a_t = \left[\left(1 - \tau_t^k \right) r_t \right] k_t + b_t$ where b_t denotes outstanding government debt. Then, the level of government debt $\bar{b}_t (k_t)$ that sustains $\bar{a}_t (k_t)$ is implicitly defined by:

$$F_k \left(k_t, l_t^{fb} \right) k_t + \bar{b}_t (k_t) = \bar{a}_t (k_t) .$$

Denote with b_t^p the debt issued by private agents. In equilibrium bonds market clear $b_t^p + b_t = 0$ at all dates. If a borrowing constraint $\bar{b}$ is imposed on private agents, $b_t^p \leq \bar{b}$, it gives rise to an additional constraint on the stage one problem that can be captured in the correspondence Γ, specifically, $-b_t \leq \bar{b}$ in terms of government savings. If $-\bar{b}_t (k_t) > \bar{b}$ the first best cannot be sustained at k_t, since $\bar{a}_t (k_t)$ will not be admissible.

While this example is very stylized it carries a general lesson. Savings constraints on the government or borrowing constraints on private agents translate into additional admissibility constraints which may prevent the government to accumulate enough assets to pay for government consumption out of interest revenues. If this is the case, then additional admissibility constraint causes Condition 14 to fail and the optimal allocation may feature intertemporal distortions.

This observation can be used to rationalize the properties of Ramsey policies with incomplete markets. Aiyagari, Marcet, Sargent and Seppala (2002) and Farhi (2006) show that the limiting behavior of optimal allocations depend on the nature of the limits imposed on the government's assets and debt. The government is said to be subject to "natural" asset and debt limits if these bounds merely insure that obligations will be paid back almost surely. More stringent debt or asset limits are referred to as "ad hoc."

Aiyagari, Marcet, Sargent and Seppala (2002) show that with quasi-linear preferences, Ramsey policies converge almost surely to the first best if the Markov process for the aggregate shock is ergodic under the natural debt and asset limits. If instead this process exhibits an absorbing state, the optimal policies converge to a Ramsey equilibrium with complete markets that depends on the values of endogenous state variables at the time the economy hits the absorbing state. By contrast, under an ad hoc *asset* limit, the economy does not converge to the first best or to a Ramsey equilibrium with complete markets. It may not converge at all.

Farhi (2006) extends these results to an economy with capital and shows that under the natural debt and asset limits the Ramsey equilibrium converges and the capital tax rate is zero in the limiting distribution. Farhi (2006) also analyzes the properties of optimal policies for general risk-averse preferences. He shows numerically that under the natural debt and asset limits the capital income tax fluctuates around zero. Under an ad hoc asset limit, the optimal capital taxes are not zero.

We can interpret these findings in terms of Theorem 15. Under the natural debt and asset limits, Condition 14 holds. The first best allocation satisfies the natural debt and asset limits by construction. In addition, the implementability constraint (8) allows the government to

save enough to be able to finance expenditures from asset returns and thus not apply any distortionary taxes. Hence, there exists and admissible allocation that converges to the first best. By contrast, under an ad hoc asset limit on the government, there may not be an admissible allocation that attains the first best.

It turns out this reasoning can be extended to second best problems in economies with idiosyncratic shocks and borrowing constraints, such as Aiyagari (1995). In these economies it is usually assumed that the government cannot employ agent-specific transfers. This implies that the first best is never attainable as it is not possible to insure agents against the idiosyncratic shocks. Section 6 presents a weaker version of our sufficient condition, namely, that the first best level of aggregate capital is attainable, and shows how it implies that there are no permanent intertemporal distortions even if agent-specific transfers are not available.

5.2 Benevolent Governments and Limited Commitment

We now consider the general class of policy problems with limited commitment constraints and a benevolent government. An implication of our result is that the absence of permanent intertemporal distortions is a robust feature of this class of environments.

Limited commitment constraints can generally be formulated as:

$$U\left(x, s^t\right) \geq O\left(x, k_t\left(s^{t-1}\right), s_t\right),$$

where $O(\cdot)$ is the value of the outside option, which can depend on allocations and the current state of the economy $\left\{k_t\left(s^{t-1}\right), s_t\right\}$. The outside option is given by a particular allocation plan. For example, in sustainable equilibria the outside option consists of a beneficial temporary deviation followed by reversion to the worst sustainable equilibrium.

By definition the continuation first best of state $\left\{k_t\left(s^{t-1}\right), s_t\right\}$ satisfies:

$$U\left(x^{fb}, s^t\right) \geq U\left(x, s^t\right),$$

for any feasible continuation plan x. It follows that:

$$U\left(x^{fb}, s^t\right) \geq O\left(x, k_t\left(s^{t-1}\right), s_t\right),$$

as long as the specification of the outside option respects feasibility constraints. Thus, Condition 14 is satisfied.[17]

Our result can be applied to any consistent specification of the limited commitment constraint. Thus, we can conclude that limited commitment constraints alone cannot provide a rationale for permanent intertemporal distortions.

This result is related to Ray (2002), who studies the time structure of self-enforcing agreements in a principal-agent framework. He shows that the optimal structure of incentives involves backloading of payments to the agent since increasing future transfers improves incentives in the current as well as in future periods. A continuation first best allocation by

[17] For our result to apply, we also need to rule out non-convexities in $O(\cdot)$ that would render the set of admissible values of capital disconnected.

construction backloads transfers to the agent. Ray (2002) does not include capital or any additional state variables and admissibility constraints in the analysis.

The contribution of our result for limited commitment models lies in its generality. It applies with a benevolent government and with political economy frictions as we discuss in Section 6. More importantly, it does not depend on the details of the game being played by the society and the government. These details affect the value of the outside option and, as is well known, the properties of optimal allocations can be very sensitive to the choice of outside option. Since the outside option is off the equilibrium path, it is hard to discriminate between the many possible specifications. By contrast, the application of Theorem 15 makes clear that the lack of permanent intertemporal distortions is a robust property for the class of limited commitment models as a whole.

Limited commitment models can give rise to a permanent intertemporal wedge if additional admissibility constraints are present. Chari and Kehoe (1990) and Phelan and Stacchetti (2001) study sustainable equilibria in an economy where the government sets all taxes every period to satisfy a balanced-budget constraint. The limited commitment is, by itself, not enough to rule out the continuation first best: if the government had enough savings to pay for public expenditure, there would be no gain from resetting taxes. It is instead the government's inability to save that renders any continuation first best allocation unattainable.[18] Hence, when front-loading of distortions is ruled out by other frictions, sustainable equilibria can display permanent intertemporal distortions.

5.3 A Private Information Economy

Assume that the economy is populated by a continuum of ex ante identical agents with preferences given by:

$$\sum_{t=0}^{T} \beta^t \left(u\left(c_t\right) - v\left(l_t\right)\right),$$

and $T \geq 1$. Each agent produces output according to the technology:

$$y_t = \theta_t l_t,$$

where θ_t denotes idiosyncratic labor productivity at time t, where $\theta_t \in \Theta$. Assume for simplicity that θ is i.i.d. across agents. Each agent will be characterized by their realization of idiosyncratic productivity shocks, so that we can express an allocation as $\left\{c_t\left(\theta^T\right), l_t\left(\theta^T\right)\right\}_{t=0}^{T}$. In each period, the allocation $\left\{c_t\left(\theta^T\right), l_t\left(\theta^T\right)\right\}$ is measurable only with respect to θ^t for $t \geq 0$.

It is immediate to derive the first best allocation, which maximizes the agents' ex ante lifetime utility subject to the resource constraint. The first best allocation features full insurance:

$$c_t\left(\theta^T\right) = c_t\left(\tilde{\theta}^T\right), \text{ for all } \theta^T, \ \tilde{\theta}^T \text{ and for all } t \geq 0,$$

[18] It turns out that the balanced-budget constraint, by itself, usually does not lead to positive capital taxes. In Section 6 we discuss the balanced-budget constraint economy in the context of a weaker version of our sufficient condition 14.

and equates the marginal rate of substitution between consumption and labor to productivity for all agents:

$$\frac{v'\left(l_t\left(\theta^T\right)\right)}{u'\left(c_t\left(\theta^T\right)\right)} = \theta_t, \text{ for all } \theta^T, \ \tilde{\theta}^T \text{ and for all } t \geq 0.$$

This implies that labor supply is increasing in productivity:

$$l_t\left(\theta_t\right) > l_t\left(\tilde{\theta}_t\right) \text{ for } \theta_t > \tilde{\theta}_t.$$

We now relax the assumption that idiosyncratic labor productivities are observed by the government. In particular, we assume that θ_t and l_t are private information and y_t is observable, as in Mirrlees (1971). An allocation in this case is given by $\left\{c_t\left(\theta^T\right), y_t\left(\theta^T\right)\right\}_{t=0}^{T}$, where $\left\{c_t\left(\theta^T\right), y_t\left(\theta^T\right)\right\}$ is measurable only with respect to θ^t for $t \geq 0$.

The optimal allocation can be obtained as a solution to a mechanism design problem. The agents make reports on their type to the government and are assigned a consumption and labor allocation based on these reports. The informational friction implies an additional constraint on the government's problem, namely that the allocation is compatible with truthful reporting. Denoting with $U_0\left(\theta^T; \tilde{\theta}^T\right)$ the lifetime utility for an agent who reports her type to be $\tilde{\theta}^T$ when her true type is θ^T, we can write the incentive compatibility constraint as:

$$U_0\left(\theta^T; \theta^T\right) \geq U_0\left(\theta^T; \tilde{\theta}^T\right), \tag{21}$$

where θ^T is an agent's true type while $\tilde{\theta}^T \in \Theta^T$ is an agent's reported type.

Clearly, the first best is not incentive compatible and so Condition 14 does not hold. It can be shown that the second best allocation displays the following three properties[19]. First, there is limited insurance and individual consumption is increasing in productivity. Second, there is a wedge between the marginal rate of substitution between consumption and labor and productivity for all agents except for the highest type. At the second best allocation, low productivity workers work too much relative to the first best, to make it unattractive for high productivity workers to report low productivity. Finally, the optimal allocation also features an intertemporal distortion:

$$u'\left(c_t\left(\theta^t\right)\right) \leq \beta R_t E_t u'\left(c_{t+1}\left(\theta^{t+1}\right)\right), \tag{22}$$

where R_t is the first best social discount factor[20]. The inequality is strict when agents are risk averse and face idiosyncratic risk in the subsequent periods as shown in Golosov, Kocherlakota and Tsyvinski (2003), that is when limited insurance is costly in utility terms for the agents.

This intertemporal wedge reflects the presence of a social cost of increasing an agent's expected utility in future periods in addition to the private cost, reflected in the marginal utility of current consumption. This cost stems from the adverse incentive effects of wealth. Higher wealth reduces the sensitivity of consumption to current labor supply, which tightens future incentive compatibility constraints. Another way to understand this intertemporal distortion

[19] See Albanesi and Sleet (2006).

[20] In an economy with capital, R_t equals the marginal product of capital.

is to contemplate the government's trade-off in the allocation of consumption between two consecutive periods. Consumption allocated to the future period will be worth less in terms of utility, since it must be spread across different states to preserve incentives and agents are risk averse. This induces the government to allocate consumption to the current period.

The intertemporal wedge implies that agents display a downward trend in consumption and utility over their lifetime. Hence, consumption and utility are *front loaded*. By contrast, in economies where our sufficient condition holds, distortions are front loaded so that consumption and utility are back loaded. The optimal front loading of consumption and utility under private information implies that their distribution spreads out over time. If $T \to \infty$, the fanning out of continuation utilities over time implies that there is no stationary distribution of consumption and utility in the limit. In particular, the degree of consumption inequality tends to continually increase, with all individuals in the population converging to their minimum promised lifetime utility, except for a vanishing fraction converging to their bliss point, a property known as *immiseration*.[21]

The fanning out of continuation utilities and the resulting lack of convergence stem from the need to intertemporally smooth distortions, a need that arises *only* when agents are risk averse and they face residual idiosyncratic risk. Under these conditions the intratemporal distortions, that is limited insurance and the labor wedge, cannot be removed but they can be ameliorated by front loading consumption and utility. Hence, the lack of stationarity of the optimal allocation in the second best is intrinsically linked to the impossibility of eventually eliminating all distortions, that is, the fact that the first best can never be attained.

By contrast, if agents are subject to a one time productivity shock at the beginning of their life, as in Werning (2006), there are no permanent intertemporal distortions. Similarly, if the process for idiosyncratic shocks has an absorbing state which is reached with positive probability, intertemporal distortions will be temporary.

6 A Weaker Sufficient Condition

A number of interesting second best problems fail to satisfy Condition 14 yet do not display permanent intertemporal distortions. Three notable examples include Aiyagari's (1995) optimal fiscal policy problem with idiosyncratic shocks and borrowing constraints; the class of political economy models with a self-interested ruler analyzed by Acemoglu, Golosov and Tsyvinski (2007); and the standard Ramsey model under a balanced-budget constraint.

We provide a heuristic argument for the existence of a *weaker* sufficient condition for the absence of permanent intertemporal distortions. The weaker condition is valid only in a subset of second best problems in which the path of aggregate capital is not constrained by its distribution among private agents. In this case, the existence an admissible allocation that attains the first best level of *aggregate capital* is sufficient to rule out permanent intertemporal distortions. This is a clearly weaker condition than Condition 14. For example, in economies with heterogeneous agents the admissibility constraints can rule out complete insurance, so

[21] The immizeration property is robust. It obtains in partial (Green, 1987, and Thomas and Worrall, 1990) and general (Atkeson and Lucas, 1992) equilibrium, under weak assumptions on preferences (Phelan, 1998).

the first best allocation of individual consumption and labor cannot be attained. Yet, the first best level of capital may be admissible in the long run.

The argument is quite simple. If the choice of aggregate capital is independent from the evolution of the capital distribution, the second best problem can be divided into two stages. In the first stage, aggregate variables are taken as given and individual allocations are chosen. The optimal plan for aggregate variables is then determined in the first stage. The set of constraints on each stage must be specified to ensure that the resulting solution is feasible and satisfies all the admissibility constraints in the original problem. The logic of our main argument can be then applied to the second stage problem: if there is a path for aggregate capital leading to its first best, it must possible to front-load distortions at the aggregate level and permanent intertemporal distortions will never be optimal.

This establishes that there are no *aggregate* intertemporal distortions. What about *individual* intertemporal distortions? It turns out that a stationary allocation the absence of aggregate intertemporal distortions imply there cannot be permanent individual intertemporal distortions. To see this, assume that for a subset of agents with positive measure

$$P^{\infty}\left(u_{i,t}^{c}+\epsilon<\beta E_{i,t}F^{k}\left(K^{*},L^{*}\right)u_{i,t+1}^{c}\right)=1$$

for an arbitrary $\epsilon>0$. Since there are no aggregate distortions we have $\beta F^{k}\left(K^{*},L^{*}\right)=1$. The marginal utility of consumption for agent i is then a sub martingale and thus it is unbounded. Agent i consumption necessarily converges to 0 and violates either feasibility or stationarity. A similar reasoning can be applied to the case of a negative individual intertemporal wedge.[22]

The limitation of this result is that it can only be applied to a subset of the class of economies that we are interested in. In particular, it excludes private information economies, where in general the evolution of individual histories does constrain the dynamic choice of aggregates. This property of the second best problem generally leads to permanent aggregate and individual intertemporal distortions. If a stationary limiting distribution exists, these distortions imply that the limiting path of aggregate capital is different from the first best. In economies without capital, they manifest themselves in an effective value of the social discount factor is different from the first best value of β.[23]

We now discuss three examples to illustrate the argument.

6.1 Example 1: Aiyagari (1995)

The first example is given by Aiyagari's (1995) analysis of optimal fiscal policy in an incomplete markets economy with heterogeneous agents. The government can impose linear tax rates on labor and capital as well as issue risk-free debt. Individuals experience idiosyncratic shocks

[22] For some second best problems, the solution can feature stationary aggregate variables but the individual variables may not be stationary. In this case it is only possible to prove that there are no permanent aggregate intertemporal distortions.

[23] Atkeson and Lucas (1992) and Albanesi and Sleet (2006) analyze second best allocations in private information economies with idiosyncratic taste and ability shocks, respectively. A lower bound on continuation utilities implies that there exists a stationary limiting distribution. In both cases, the social discount factor is smaller than β.

to their productivity. They can save and partially insure themselves by holding capital and government debt but their net worth must remain non-negative. There are no aggregate shocks.

The state can be summarized by aggregate capital and the distribution of net worth for this economy. The distribution of capital among agents is not pinned down in equilibrium, since agents are indifferent between holding capital or government debt. Hence, this environment belongs to the subset of our class of second best problems in which the evolution of aggregate capital can be decoupled from its distribution.

The first best is never attainable in this economy as the government lacks the agent-specific transfers needed to provide insurance agents idiosyncratic shocks. However, the government can always finance its spending exclusively from labor taxes or returns from its assets, leaving the level of capital undistorted. Hence, the weaker sufficient condition is satisfied.

Aiyagari (1995) shows that in the second best there is no permanent aggregate intertemporal wedge:

$$1 = \beta F^k \left(K^*, L^*\right).$$

To implement this outcome it is necessary to levy a positive tax on capital. The tax counteracts the precautionary demand for capital arising from the constraint on borrowing. Despite the presence of this tax, the individual intertemporal margin does not exhibit permanent distortions. The mass of agents with binding net worth constraint faces a negative intertemporal wedge, while all other agents face a positive intertemporal wedge, corresponding to the optimal capital tax. This discussion also clarifies that a positive capital tax would not be needed to implement the second best allocation absent the constraint on net worth, even with incomplete markets. In this case, all individual intertemporal wedges would equal zero in the limit.

The government's ability to accumulate debt or assets serves a dual purpose in this economy. As in our general formulation, it allows for intratemporal distortions to be reallocated over time. In addition, it decouples individual histories from individual capital holdings. It is this property of the equilibrium that ensures the weaker sufficient condition holds.

6.2 Example 2: Acemoglu, Golosov, and Tsyvinski (2007)

Acemoglu, Golosov and Tsyvinski (2007) (AGT henceforth) analyze a dynamic economy where allocations are chosen by a self-interested ruler. The ruler's utility function is defined over sequences of transfers extorted from private agents. An allocation consists of sequences of aggregate capital and transfers to the ruler as well as sequences of consumption and labor supply decisions for each private agent. The ruler cannot observe individual abilities and can seize a fraction of aggregate output. This leads to two sets of admissibility constraints in each period. The sustainability constraint ensures that the ruler will not seize output, while individual incentive compatibility constraints guarantee that private agents will not misreport their type.

Under the assumption that the ruler is weakly more patient than private agents, this model falls squarely in our framework.[24] AGT show that in this case there are no permanent

[24] The case in which the ruler is less patient than private agents generally does not converge to an interior stationary distribution and thus violates our regularity conditions.

intertemporal distortions under the additional assumption that individual histories are private. Yet, the incentive compatibility constraints on private agents imply that the first best is never attainable.

AGT prove this result by splitting the second best problem into two stages. The first stage corresponds to the choice of individual allocations subject to incentive compatibility constraints for given aggregate sequences of consumption and labor supply. The second stage consists in the provision of incentives to the ruler. That is, aggregate consumption, labor, capital and transfers to the ruler are chosen subject to the resource constraint, the sustainability constraint and an additional constraint that guarantees the existence of a solution to the first stage problem. This strategy parallels our previous discussion.

The key property of the model that guarantees this result is that the choice of individual allocations does not permanently constrain the path of aggregate variables. The second stage problem then resembles a representative agent model with a limited commitment constraint. Hence, based on our result there will be no aggregate permanent intertemporal distortions if there is a path for capital that converges to the first best level for that problem.[25] Since the ruler's continuation utility is weakly increasing in aggregate capital, based on our discussion in Section 5.2, the first best level of capital satisfies the limited commitment constraint. Some mild assumptions on the ruler's utility function guarantee that it is feasible to provide the necessary transfers in every period and maintain capital at its first best level.[26]

The assumption that individual histories of consumption, labor supply and ability are private information is key for this result. In this case, agents face a sequence of static incentive compatibility constraints. This restriction implies that individual allocations are not history dependent. In this case, there are no individual intertemporal distortions, though there will be permanent intratemporal distortions. In the case of public histories, the path of individual histories constrains the dynamic of aggregate capital, as discussed in Section 5.3, leading to both aggregate and individual intertemporal distortions.

6.3 Example 3: Balanced-Budget Constraint

Our last example examines the implications of a balanced-budget constraint on the government. In an otherwise standard Ramsey model, the government is forced to finance its expenditure solely from current tax revenues, that is, it cannot save or borrow. Both Judd (1985) and Chamley (1986) claim that the absence of permanent intertemporal distortions does not depend on assumptions about the government's ability to borrow or lend.[27]

This public policy problem has been analyzed for economies where the distribution of capital is trivial, either because there is a single representative agent or it is a version of the two-class economy of Judd (1985) where only one agent-type holds all the capital. There is thus no distinction between individual and aggregate capital in these economies.

[25] If Condition 14 holds for the stage two problem, then the weak sufficient condition holds for the overall second best problem.

[26] See Assumption 4 in AGT (2007).

[27] See also Stockman (2001) for a detailed analysis.

The first best is generally not attainable under a balanced budget constraint. However, it may be possible to finance government expenditures only through labor taxes. In this case the first best level of capital can be attained and our weaker sufficient condition applies.

This result may be surprising in light of the crucial role of the ability to front-load distortions for our result. How can the government reallocate taxes across periods without being able to save or borrow? It turns out that the government usually can independently manipulate the path of consumption and capital even if it has no access to debt. It is actually helpful to consider an economy where it *cannot*. Lansing (1999) shows that a zero capital tax cannot be implemented in a two-class economy with logarithmic preferences and a balanced-budget constraint. The key observation is that the present value of future consumption only depends on current consumption under logarithmic preferences. As a result the government cannot induce a change in the intertemporal profile of consumption and an allocation with no permanent intertemporal distortions cannot generically be implemented in equilibrium.[28]

7 Conclusions

Our result clarifies that the ability to reallocate distortions over time plays a key role for the presence of permanent intertemporal distortions in the second best. A natural question is to what extent the logic of the argument can be applied to environments outside our framework.

Of particular interest are second best problems in monetary economies. Many of these models do not admit a primal representation and therefore do not fit in our framework[29]. However, in many cases the dynamics of capital are not constrained by the monetary frictions. Following the reasoning in Section 6, we conjecture that as long as it is possible to attain the first best level of capital, there should be no permanent distortions in the investment decision.

Our argument requires restrictions on the degree of history dependence in the admissibility constraints. In policy problems with an arbitrary degree of history dependence our result does not hold. One trivial example is a Ramsey model where the capital tax rate must be constant at all dates. In other settings, the unlimited history-dependence of the constraints may be a primitive feature of the problem. Our condition on the limits of history dependence is quite general, but it does not identify a general class of second best environments that are not admitted in the formulation.

Our analysis presumes the existence of a second best plan that converges to a limiting stationary allocation. Interestingly, in many economies the second best allocation does not converge when it features permanent intertemporal distortions. Private information economies

[28] We have already briefly mentioned another example where the balanced-budget constraint leads to positive capital taxes: when it is combined with a limited commitment constraint. Note in this case the first best level of capital may not be admissible, as there is always the temptation to unexpectedly raise taxes.

[29] In general, monetary models with nominal rigidities may exhibit unlimited history dependence in the admissibility constraints violating our formulation. Moreover, money may enter the problem as an additional state variable. Siu (2004) provides a primal representation for a monetary Ramsey model with sticky prices that can be mapped in our framework.

Second best conomies where money is essential also do not admit a conventional primal representation. See Aruoba and Chugh (2007) for an example.

are a notable example. It is well-known that imposing a lower bound on continuation utility, as in Atkeson and Lucas (1995), or allowing the government to discount the future at a lower rate than private agents[30], as in Phelan (2006) and Farhi and Werning (2006), guarantees the existence of a non-degenerate limiting distribution. Yet, in both cases permanent intertemporal distortions are a feature of the limiting optimal allocation. Hence, the presence of intertemporal distortions does not depend on the convergence properties of the second best.

Finally, we restrict our attention to infinite-horizon economies. Erosa and Gervais (2002), Garriga (2003), and Krueger, Conesa, and Kitao (2006) show that it may be optimal to tax capital in overlapping generations economies if age-specific taxes are not available. One fundamental difference between infinitely lived and overlapping generations economies is that the latter can often display dynamic inefficiencies. Importantly, with overlapping generations it is not possible to shift distortions across time without an implicit transfer between generations. A reconciliation of the findings on intertemporal distortions for overlapping generations and infinite-horizon economies would obviously be of interest. We leave this topic for future work.

References

[1] Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. 2006. Political Economy of Mechanisms. MIT Economics Department Working Paper.

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

[3] 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.

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

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

[6] Armenter, Roc. 2007. A Note on Incomplete Factor Taxation. Federal Reserve Bank of New York Working Paper.

[7] Aruoba, Boragan, and Sanjay Chugh. 2007. Optimal Fiscal and Monetary Policy when Money is Essential. Manuscript, University of Maryland

[8] 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.

[30] Sleet and Yeltekin (2006) show that limited commitment can provide a rationale for such differential discounting.

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

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

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

[12] Conesa, Juan Carlos, Sagiri Kitao and Dirk Krueger. 2006. Taxing Capital? Not a Bad Idea After All! NBER WP 12880.

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

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

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

[16] Farhi, Emmanuel. 2006. Capital Taxation and Ownership When Markets are Incomplete. Manuscript, MIT.

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

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

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

[20] Golosov, Mikhail, and Aleh Tsyvinski. 2006. Optimal Taxation with Endogenous Insurance Markets. Forthcoming, *Quarterly Journal of Economics.*

[21] Golosov, Mikhail, and Aleh Tsyvinski. 2006. Designing Optimal Disability Insurance: A Case for Asset Testing. *Journal of Political Economy.*

[22] 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.

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

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

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

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

[27] 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.

[28] Mirrlees, James. 1971. An exploration in the theory of optimum income taxation. *The Review of Economic Studies* 38: 175-208.

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

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

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

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

[33] Siu, Henry. 2004. Optimal fiscal and monetary policy with sticky prices. *Journal of Monetary Economics* 51 (3).

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

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

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

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

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

A Appendix

A.1 Limiting Probability Distributions

Formally, we define **a probability measure** for any given allocation x as

$$\Pr\left(x_t \in B\right) = \sum_{s^t \in S^t} \pi\left(s^t | s_0\right) \chi_B\left(x_t\left(s^t\right)\right),$$

where $\chi_B\left(x_t\left(s^t\right)\right)$ is the indicator function and B is a subset of X, i.e., $\chi_B\left(x_t\left(s^t\right)\right) = 1$ if $x_t\left(s^t\right) \in B$, zero otherwise. The definition can be trivially extended to k and a. The corresponding conditional probability measure is given by:

$$\Pr\left(x_t \in B|s^d\right) = \sum_{s^t \in S^t|s^d} \pi\left(s^t|s_d\right) \chi_B\left(x_t\left(s^t\right)\right).$$

A **stationary distribution** P_x^∞ for plan x is a probability distribution over measurable subsets of X such that, for some $s^t \in S^t$, $t \geq 0$,

$$\lim_{j\to\infty} \Pr\left(x_j \in B|s^t\right) = P_x^\infty(B)$$

for all measurable subsets $B \subseteq X$.

We also include a formalization of Condition 12. Define the set of allocations

$$Z\left(s^t\right) = \left\{x \in \mathcal{X} : \forall B \subset X,\ \lim_{j\to\infty}\left\{\Pr\left(x_j \in B|s^t\right)\right\} = P_x^\infty(B)\right\},$$

where $P_x^\infty \in \mathcal{P}_x^\infty$, i.e., P_x^∞ a stationary distribution. Then, Condition 12 is equivalent to assume that a second best allocation x^* satisfies

$$\lim_{t\to\infty} \Pr\left(x^* \in Z\left(s^t\right)\right) = 1.$$

Finally, we define the support K^{fb} for the limiting distribution for first best allocations as the smallest subset of K such that

$$\lim_{t\to\infty} \Pr\left(k^{fb}\left(s^t\right) \in K^{fb}\right) = 1,$$

where k^{fb} is part of a first best plan.

A.2 Proofs in Section 4.1

Proof of Proposition 18. Let $\{x, k\}$ solve Problem 2 for a given $a \in \mathcal{A}$, and

$$P^\infty\left(\phi\left(s^t\right) = 0\right) = 1$$

for $P^\infty \in \mathcal{P}_x^\infty$. There exists then a node s^{t^*} such that for all $s^t \in S^t|s^{t^*}$, $\phi\left(s^t\right) = 0$ with probability one. Without loss of generality, we look at allocations along the stationary distribution at dates $t \geq t^* + d$ for d large enough. Condition 6 implies that allocations $x_t\left(s^t\right)$ only have an impact on admissibility constraints (3) for nodes $s^j \subseteq s^{t^*}$. The necessary first order condition for consumption (18) becomes

$$\beta^t \pi\left(s^t|s_0\right)\left\{u_i^c\left(s^t\right) - \mu_i \lambda\left(s^t\right)\right\} = \sum_{s^j \subseteq s^{t^*}} \beta^j \pi\left(s^j|s_0\right) \phi\left(s^j\right)' H_i^c\left(s^j, s^t\right).$$

Condition 6 also implies that

$$H_i^c\left(s^j, s^t\right) = \beta^{t-j}\pi\left(s^t|s_j\right) \begin{bmatrix} r_1\left(x_j\left(s^j\right), s_j\right) h_1^{c_i}\left(x_t\left(s^t\right), s_t\right) \\ r_2\left(x_j\left(s^j\right), s_j\right) h_2^{c_i}\left(x_t\left(s^t\right), s_t\right) \\ \cdots \\ r_m\left(x_j\left(s^j\right), s_j\right) h_m^{c_i}\left(x_t\left(s^t\right), s_t\right) \end{bmatrix}$$

so (18) can be written as

$$u_i^c\left(s^t\right)-\mu_i\lambda\left(s^t\right)=\bar{\phi}'\begin{bmatrix} h_1^{c_i}\left(x_t\left(s^t\right),s_t\right)\\ h_2^{c_i}\left(x_t\left(s^t\right),s_t\right)\\ ...\\ h_n^{c_i}\left(x_t\left(s^t\right),s_t\right)\end{bmatrix}$$

where

$$\bar{\phi}=\sum_{s^j\subseteq s^{t^\star}}\phi\left(s^j\right)'\begin{bmatrix} r_1\left(x_j\left(s^j\right),s_j\right) & 0 & ... & 0\\ 0 & r_2\left(x_j\left(s^j\right),s_j\right) & & \\ ... & & ... & \\ 0 & & & r_m\left(x_j\left(s^j\right),s_j\right)\end{bmatrix}.$$

Similarly, the necessary first order condition for $l_{it}\left(s^t\right)$ (19) becomes

$$u_i^l\left(s^t\right)-\mu_i\lambda\left(s^t\right)F_i^l\left(s^t\right)=\bar{\phi}'\begin{bmatrix} h_1^{l_i}\left(x_t\left(s^t\right),s_t\right)\\ h_2^{l_i}\left(x_t\left(s^t\right),s_t\right)\\ ...\\ h_n^{l_i}\left(x_t\left(s^t\right),s_t\right)\end{bmatrix}.$$

The necessary first order condition for $k_{it}\left(s^{t-1}\right)$ (17) is simpler, since $\phi\left(s^{t+1}\right)=0$,

$$\mu_i\lambda\left(s^t\right)=\sum_{s^{t+1}}\beta\pi\left(s^{t+1}|s_t\right)\lambda\left(s^{t+1}\right)F_i^k\left(s^{t+1}\right).$$

Hence all allocations from s^t onwards can be characterize as function of $\left\{k_t\left(s^{t-1}\right),s_t\right\}$ and the constant vector $\bar{\phi}$.

Let $\sigma\in\Sigma=K\times S\times\Re^m$ denote the complete state of the economy. Let

$$\Omega_i^c\left(\sigma\right)=1-\frac{1}{u_i^c\left(\sigma\right)}\bar{\phi}'\begin{bmatrix} h_1^{c_i}\left(x\left(\sigma\right),\sigma\right)\\ h_2^{c_i}\left(x\left(\sigma\right),\sigma\right)\\ ...\\ h_n^{c_i}\left(x\left(\sigma\right),\sigma\right)\end{bmatrix}$$

for some $i\in I$. There is no need for a time subscript now. Our regularity Condition 8 implies that $\Omega_i^c\left(\sigma\right)$ is continuous and bounded above and below.

Define the operator Θ on the space of continuous and bounded functions B as

$$\Theta\left[\Omega\right]\left(\sigma\right)=\frac{\sum_{s'}\omega_i\left(\sigma,s'\right)\Omega\left(\sigma',s'\right)}{\sum_{s'}\omega_i\left(\sigma,s'\right)}$$

where

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

and σ' is given by the law of motion for $k_t\left(s^{t-1}\right)$. Since $\omega_i\left(\sigma,s'\right)>0$, Θ maps B unto itself.

Because $\Omega_i(\sigma)$ is bounded above and below, either $\Omega_i(\sigma)$ equals a constant Ω_i^* with probability 1 in the stationary distribution P^∞, or

$$\begin{aligned} P^\infty(\Theta[\Omega_i](\sigma) > \Omega_i(\sigma)) &> 0, \\ P^\infty(\Theta[\Omega_i](\sigma) < \Omega_i(\sigma)) &> 0. \end{aligned}$$

Otherwise if, say, $P^\infty(\Theta[\Omega_i](\sigma) > \Omega_i(\sigma)) = 1$, either the upper bound is violated with probability 1 or $P^\infty(\Omega_i(\sigma) = \sup\{\Omega_i(\sigma)\}) = 1$

The result on Ω_i maps into the Proposition after using the necessary first order conditions derived above ■

Proof of Proposition 19. If $\{x, k\}$ is admissible for plan a, it is admissible for any plan a' such that $a'_t(s^t) \geq a_t(s^t)$ so the first point follows.

For the second point, let feasible allocation $\{x, k\}$ be such that $U(x, s_0) = W(a)$ and $U(x, s_0) = W(a + \varepsilon)$ for ε strictly positive. No admissibility constraint holds with equality under $a + \varepsilon$ for allocation plan $\{x, k\}$, so the necessary first order conditions must hold with all Lagrangian multipliers $\phi(s^t)$ equal to zero.

It is straightforward to show that the first best allocation can be characterized as the solution to

$$\max_x U(x, s_0)$$

subject to the resource constraint (1) and non-negativity conditions. Given Condition 8, the necessary first order conditions are sufficient as well and, quite trivially, coincide with the necessary first order conditions for Problem 2 when no admissibility constraint is binding. Thus $x = x^{fb}$. ■

A.3 Proofs in Section 4.2

Proof of Proposition 22. Let $\left\{\tilde{k}, \tilde{x}\right\}$ be a first-best continuation allocation of $\left\{k_t^*(s^{t-1}), s_t\right\}$ and $\tilde{a}$ satisfy that $\tilde{a} \geq \bar{a}\left[k_t^*(s^{t-1}), s_t\right]$ in $\left\{S^j | s^t, j \geq t\right\}$, $\tilde{a} = a^*$ elsewhere. Such $\tilde{a}$ exists and satisfies (4) for all $s^j \in S^j|s^t$, $j \geq t$ by $a_t^*(s^t) \in A^{fb}\left[k_t^*(s^{t-1}), s_t\right]$. To show that $\{\tilde{a}, k^*, x^*\}$ is an admissible plan, note that at any node $s^j \in S^j|s^t, j \geq t$, if admissibility constraint m is binding, then

$$H_m\left(x^*, k_j^*(s^{j-1}), s_j\right) < H_m\left(\tilde{x}, \tilde{k}_j(s^{j-1}), s_j\right).$$

Otherwise, Condition 11 would not be satisfied. Then,

$$H_m\left(x^*, k_j^*(s^{j-1}), s_j\right) < H_m\left(\tilde{x}, \tilde{k}_j(s^{j-1}), s_j\right) \leq \bar{a}_m\left[k_t(s^{t-1}), s_t\right](s^j) \leq \tilde{a}_{mj}(s^j).$$

If there is any admissibility constraint binding in $\left\{S^j | s^t, j \geq t\right\}$, it can be then relaxed by picking $\tilde{a}$ over a^*, but $W(\tilde{a}) > W(a^*)$ would contradict $\{x^*, k^*\}$ being a second best plan ■

Proof of Proposition 23. If $a_t(s^t) \in int\left(\Gamma\left(a_t(s^t), s_t\right)\right)$ the proof is straightforward. If $a_t(s^t) \notin int\left(\Gamma\left(a_t(s^t), s_t\right)\right)$, then $a_t(s^t)$ is an adjacent point to $\Gamma\left(a_t(s^t), s_t\right)$ as $a_t(s^t) \in \Gamma\left(a_t(s^t), s_t\right)$ by Condition 7. By the separating hyperplane theorem, there exists a half-space

$\chi_p = \left\{z \in A : pz \geq pa_t\left(s^t\right) \in \Re\right\}, p \neq \vec{0}$, such that $\Gamma\left(a_t\left(s^t\right), s_t\right) \subset \chi_p$ — recall that the image of $\Gamma\left(a_t\left(s^t\right), s_t\right)$ is a convex set by Condition 7.

Next we show that for any such a half-space, $\bar{A}^{fb} \cap \chi_p \neq \varnothing$. Assume there exists χ_p such that $\bar{A}^{fb} \cap \chi_p = \varnothing$. Since the set $\bar{A}^{fb}$ is attainable by Condition 14, there must exist some $x \in \chi_p$ with $y \in \Gamma(x, s_t)$, $y \notin \chi_p$ for some state $s_t \in S$ (otherwise there would be no way to "escape" the half-space χ_p). Pick point $z \in A$ such that for some $\gamma \in (0,1)$, $a_t\left(s^t\right) = \gamma z + (1-\gamma)x$. Such a point will belong to the closure of the complement of χ_p, i.e., $\left\{z \in A : pz \leq pa_t\left(s^t\right) \in \Re\right\}$. Since $z \in \Gamma(z, s_t)$, the convexity of Γ implies that $\gamma y + (1-\gamma) z \in \Gamma\left(a_t\left(s^t\right), s_t\right)$ but clearly $\gamma y + (1-\gamma) z \notin \chi_p$ – a contradiction.

Consider the set $G = \left\{\alpha a_t\left(s^t\right) + (1-\alpha)\,\bar{a}\left(s^{t+1}\right) : \alpha \in [0,1), \bar{a}\left(s^{t+1}\right) \in \bar{A}^{fb}\right\}$. This is a convex set with $\bar{A}^{fb} \subset G$. If $\Gamma\left(a_t\left(s^t\right), s_t\right) \cap G = \varnothing$, it would be possible then to find a separating hyperplane χ_p with $\Gamma\left(a_t\left(s^t\right), s_t\right) \subset \chi_p$ and $G \cap \chi_p \neq \varnothing$. But this would contradict $\bar{A}^{fb} \cap \chi_p \neq \varnothing$ ■

Before proving Proposition 24 we find useful to state a further property of the second best plan for the auxiliary variable, which allows to order at least one element of $\bar{A}^{fb}$ with respect the dimensions of the auxiliary variable that correspond to binding admissibility constraints.

Proposition 25 *Let some admissibility constraint m be binding at node s^t. Then there exists $a'_t \in \bar{A}^{fb}$ such that $a^*_{mt}\left(s^t\right) < a'_{mt}\left(s^t\right)$.*

Proof. Set $\bar{A}^{fb}$ is non-empty by Condition 14. If $a^*_t\left(s^t\right) \in \bar{A}^{fb}$ then Proposition 22 says no admissibility can be binding. Hence $a^*_t\left(s^t\right) \notin \bar{A}^{fb}$. If for any element $a_t\left(s^t\right) \in \bar{A}^{fb}$, we have that $a_t\left(s^t\right) \leq a^*_t\left(s^t\right)$, $a_t\left(s^t\right) \neq a^*_t\left(s^t\right)$, then the definition of $\bar{A}^{fb}$ is not satisfied. Finally, if for some $n \leq m$, $a^*_{nt}\left(s^t\right) > \sup\left\{a'_{nt} : a'_t \in \bar{A}^{fb}\right\}$, then Condition 11 implies admissibility constraint n cannot be binding ■

Proof of Proposition 24. If no admissibility constraint is binding, then $x^{fb} = x^*$ and the result follows trivially. If an admissibility constraint is binding at node s^t, then Proposition 25 implies that $a^*_t\left(s^t\right) \leq a_t\left(s^t\right), a^*_t\left(s^t\right) \neq a_t\left(s^t\right)$ for all $a_t\left(s^t\right) \in \bar{A}^{fb}$. Applying Proposition 23, there exists $a_{t+1}\left(s^{t+1}\right) > a^*\left(s^t\right)$ and $a_{t+1}\left(s^{t+1}\right) \in \Gamma\left(a^*_t\left(s^t\right), s_{t+1}\right)$.

The rest of the proof is structured with two Lemmas.

The following lemma says an auxiliary plan a can be improved if it is originally in the interior of the image of the correspondence Γ.

Lemma 26 *Let a be an admissible plan with $a_t\left(s^t\right) \in int\left(\Gamma\left(a_{t-1}\left(s^{t-1}\right), s_t\right)\right)$. Then there exists an admissible plan $\tilde{a}$ with $\tilde{a} \geq a$, $\tilde{a} \neq a$.*

Proof. We prove the Lemma by construction. Set $\tilde{a} = a$ everywhere but in the set $\left\{S^{t+j} | s^t : j \geq 0\right\}$. Since $a_t\left(s^t\right) \in int\left(\Gamma\left(a_{t-1}\left(s^{t-1}\right), s_{t-1}\right)\right)$, therefore $a_t\left(s^t\right) \in int\left(A\right)$ and there exists $z \in A$ such that $a_t\left(s^t\right) < z$ and $a\left(s^{t+j}\right) \leq z$ for all $s^{t+j} \in S^{t+j} | s^t$, $j \geq 1$. For sufficiently small scalar $\alpha > 0$,

$$\tilde{a}\left(s^t\right) = \alpha z + (1-\alpha)\, a_t\left(s^t\right) \in \Gamma\left(a_{t-1}\left(s^{t-1}\right), s_{t-1}\right).$$

By convexity of Γ,

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

and so on $s^{t+j} \in S^{t+j}|s^t$, $j \geq 1$ ■

Since Proposition 19 establishes that, if $a' \geq a$, then $W(a') \geq W(a)$, we can use Lemma 26 to conclude that, without loss of generality, if a^* is a second best plan, any admissible plan $a \geq a^*$ is also a second best plan. We say "without loss of generality" because it must be that $W(a) = W(a^*)$, otherwise we would contradict a^* being a second best plan.

We extend the previous Lemma to any pair of ordered points in A.

Lemma 27 *Let an admissible plan a and $\tilde{a}_t\left(s^t\right) \in A$ be such that $a_t\left(s^t\right) < \tilde{a}_t\left(s^t\right)$, $\tilde{a}_t\left(s^t\right) \in \Gamma\left(a_{t-1}\left(s^{t-1}\right), s_t\right)$. Then there exists an admissible plan $\tilde{a} \geq a$, $\tilde{a} \neq a$.*

Proof. Since the image of $\Gamma\left(a_{t-1}\left(s^{t-1}\right), s_t\right)$ is convex, it follows that $a_t^{\alpha}\left(s^t\right) = \alpha a_t\left(s^t\right) + (1-\alpha)\,\tilde{a}_t\left(s^t\right)$ belongs to $\Gamma\left(a_{t-1}\left(s^{t-1}\right), s_t\right)$ as well for any $\alpha \in [0,1]$. By picking $\alpha \in (0,1)$, $a_t^{\alpha}\left(s^t\right) \in int\left(\Gamma\left(a_{t-1}\left(s^{t-1}\right), s_{t-1}\right)\right)$, and we can use Lemma 26 ■

Finally, we close the argument here. Lemma 27 implies that we can take $a_{t+1}^*\left(s^{t+1}\right) > a_t^*\left(s^t\right)$ without loss of generality. Since this is true for any sequence with $a_t^*\left(s^t\right) \leq a_t\left(s^t\right), a_t^*\left(s^t\right) \neq a_t\left(s^t\right)$ for all $a_t\left(s^t\right) \in \bar{A}^{fb}$, it follows that one can take the sequence to converge almost surely to $\bar{A}^{fb}$ ■

A.4 Proof of Theorem 15

Proof. Let $\{x^*, k^*\}$ be a second best allocation plan. By Proposition 24 and Condition 14, there exists a second best plan $\{x^*, k^*, a^*\}$ with

$$\lim_{t\to\infty} \Pr\left(a_t^*\left(s^t\right) \in \bar{A}^{fb}\right) = 1.$$

Proposition 22 implies that eventually no admissibility constraint is binding for allocations $\{x^*, k^*\}$

$$\lim_{t\to\infty} \Pr\left(\phi\left(s^t\right) = 0\right) = 1.$$

Note that for all second best plans $\{x^*, k^*, a'\}$ the Lagrangian multipliers must be zero in the same nodes, otherwise $W(a') \neq W(a^*)$ and either a' or a^* would not constitute a second best plan.

The Theorem is then proven by Proposition 18 ■