Intertemporal Distortions in the Second Best[*]

Stefania Albanesi
Columbia University and NBER

Roc Armenter
Federal Reserve Bank of New York

August 29, 2007

Abstract

We consider a very general class of optimal policy problems that encompasses the standard Ramsey model as well as environments with idiosyncratic risk, incomplete markets, limited commitment, private information, and arbitrary restrictions on assets. We identify a sufficient condition to rule out permanent intertemporal distortions. If there exists an admissible allocation that converges to the first best steady state, then all intertemporal distortions are temporary at the optimal allocation. We analyze a series of applications to illustrate the significance of this result.

---

[*] PRELIMINARY AND INCOMPLETE DRAFT. We would like to thank seminar participants at the FRB New York, Columbia University, the FRB Philadelphia, SED 2007, Ente Einaudi, 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 or the Federal Reserve System. Contact: Stefania.Albanesi@columbia.edu and Roc.Armenter@ny.frb.org.

1 Introduction

A key lesson of public finance is that distortions should be spread across all margins. Yet many fiscal policy problems prescribe that all distortions on the intertemporal margin ought to be purely temporary. Ramsey models are perhaps the best known example: Chamley (1986) and Judd (1985) first established that optimal capital income taxes are zero in the steady state and there are no intertemporal distortions. This finding has been confirmed for a variety of more general environments.[1] By contrast, recent work in dynamic public finance has focused on constrained-efficient economies that do feature intertemporal distortions, such as those with private information and political economy frictions.

In this paper we seek to understand why permanent distortions on the intertemporal margin are not optimal in some environments, and what is different about economies where they are. To this end we consider a very general class of public finance problems that can be represented as a choice of allocations subject to resource feasibility and a set of additional constraints.[2] 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.

We consider a class of environments that 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 complete factor taxation and aggregate or idiosyncratic risk, as well as economies with incentive compatibility constraints due to limited commitment or private information. We also encompass settings which combine these constraints with arbitrary asset restrictions like incomplete markets and borrowing constraints on the government or private agents.

We identify a sufficient condition to rule 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 allocation. This condition is typically straightforward to verify. For example, in a Ramsey model it simply amounts to the government's ability to save enough to eventually finance all expenditures from asset returns.

Why is this condition linked to zero intertemporal distortions? If it holds, then the constraints on optimal policies allow for all distortions to be front-loaded — this is exactly what converging to the first best amounts to. This guarantees that all intratemporal distortions can be reallocated intertemporally without tightening any future constraints — otherwise the allocation could not be converging to the first best.

Consider a candidate optimum with a non-zero intertemporal wedge in the steady state. Obviously, an *intertemporal* reallocation of resources would generate first order welfare gains. If our condition is satisfied, such a reallocation can be achieved by redistributing *intratemporal* distortions over time without violating any future constraints. Hence, it is indeed possible to

---

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

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

improve upon the candidate optimum. It follows that the second best allocation cannot have non-zero intertemporal distortions in the steady state.

Our result is capable of providing immediate insight into many policy problems. For example, consider the class of economies with a benevolent government subject to limited commitment constraints. Typically, the continuation value of the second best allocation must exceed the value of an outside option at every date and state. The outside option consists of a beneficial temporary deviation followed by a punishment phase. The formulation of the punishment phase varies in the literature, and results can be sensitive to this modelling choice. Any meaningful formulation of the outside option, however, cannot improve on the continuation first best. Hence, our sufficient condition is guaranteed to hold. We can then assert that limited commitment constraints alone cannot lead to permanent intertemporal distortions.

We also contribute to the understanding of second best problems with 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), Kocherlakota (2005), and Albanesi and Sleet (2006). Under private information, our condition is not satisfied since the first best continuation allocations are typically not incentive compatible. This means that, at the second best allocation, it is not possible to reallocate resources intertemporally without tightening future constraints. Hence, the presence of intertemporal distortions is a fundamental feature of this class of economies.

Our result also provides some 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.[3] For example, the standard 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 clear what generates this fundamental difference in the prediction on the intertemporal wedge. It is the absence of costs associated with front-loading distortions in the Ramsey framework. This finding provides a deeper understanding of both approaches.

The class of policy problems we examine does not include Ramsey models with incomplete factor taxation. In these models, the optimal capital tax is often non-zero if the supply of the untaxed factor is influenced by the level of capital. Similarly we exclude overlapping generations economies where it may be optimal to tax capital if age-specific taxes are not available.[4] These are well-understood results. Instead, the goal of this paper is to shed light on the rationale for not taxing capital. We also restrict attention to neoclassical environments with constant returns to scale. This rules out economies with non-convexities which may

---

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

[4] Correia (1996) and Jones, Manuelli and Rossi (1997) analyze incomplete factor taxation. Erosa and Gervais (2002) and Garriga (2003), Krueger, Conesa, and Kitao (2006) analyze optimal capital taxation with overlapping generations.

feature permanent intertemporal distortions. Finally, we do not consider environments in which there are differences in preferences between the government and private agents.[5]

We formally characterize our class of second best problems in section 2. Optimal allocations 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. In section 3, we illustrate 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 for 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. Even if no future admissibility constraint binds at the optimum, the forward-looking nature of the admissibility constraints implies that the optimal allocations will generally 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 of our result, and section 6 concludes.

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 $c_{it}(s^t)$ and $l_{it}(s^t)$ denote consumption and leisure of a household of type $i \in I$ at date t

---

[5] One prominent class of economies with this feature is the one analyzed by Acemoglu, Golosov, and Tsyvinski (2006) and Yared (2007). Phelan (2006), Sleet and Yeltekin (2006) and Farhi and Werning (2006) are examples with differential discounting.

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 exogenous 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 $i \in I$.

Aggregate welfare at node s^t for any allocation plan x is given by

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

where u_i is a standard utility function.

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 and satisfy 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 $a = \left\{a_t\left(s^t\right) \subset \Re^m : s^t \in S^t, t \geq 0\right\}$. This variable allows us to capture arbitrary constraints on the asset space as well as restrictions on the dynamic 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 differentiable 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.

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

We will refer to a plan $\{x, k, a\}$ such that $\{x, k\}$ is admissible, as an admissible plan.

Definition 5 *An admissible allocation* $\{x^*, k^*\}$ *is* ***second best*** *if* $U\left(x^*, s_0\right) \geq U\left(x, s^0\right)$ *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. Specifically, admissibility constraints at node s^t cannot restrict allocations arbitrarily far in the past and, after a finite number of periods, only future allocations on continuation histories of s^t can enter. In other words, each admissibility constraint must be essentially forward-looking. In addition, admissibility constraints must be eventually time-separable and intertemporal discounting must occur at the same as in (2). Formally, this restriction imposes two conditions on the function H.

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 $s^t \notin S^t | s^j$,*

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

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

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

for some functions r_n and h_n.

Condition 6 does rule out some interesting policy problems. For example, heterogeneous discounting rates do not satisfy the condition. Admissibility constraints that depend on arbitrary future realizations that are not continuations of the current state are also excluded, as well as arbitrarily non time-separable preferences.

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 $\Gamma\left(a_t\left(s^t\right), s_t\right) : \Re^m \times S \rightrightarrows \Re^m$ 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 $a_t\left(s^t\right) \in \Re^m$.*

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.

Let $X \in \Re_+^2$ be the set of $x_t\left(s^t\right)$ that belongs to an admissible plan $\{x, k, a\}$ for some $s^t \in S^t, t \geq 0$. Similarly, we can define K and A for capital and the auxiliary variable, respectively. Let $\mathcal{X}$ denote the set of plans x that belongs to an admissible plan $\{x, k, a\}$. Once again, $\mathcal{K}$ and $\mathcal{A}$ can be defined by analogy.

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 now impose conditions on the structure of the admissibility constraints, that is on the space of allocations, as well as on the functions H and Γ.

Condition 9 *The sets A, K, and X are compact and convex. The function H is bounded, twice continuously differentiable with bounded first-order derivatives in $\mathcal{X}$ and $\mathcal{K}$.*

Condition 9 allows admissibility constraints to be non-convex for a given path of the auxiliary variable, a. In particular, H can be a non-convex function of $\{x, k\}$ as long as the set of auxiliary plans $\mathcal{A}$ span a convex space $X \times K$. However, if the set $\mathcal{A}$ is small, then the first part of Condition 9 implies that H must be convex. For example, say a particular dimension of the auxiliary variable $a_{nt}\left(s^t\right)$ is constrained to be a singleton, $a_{nt}\left(s^t\right) = 0$ and the admissibility constraint n takes the simple form:

$$O_n\left(k_t\left(s^{t-1}\right), s_t\right) \leq 0.$$

Then, O must be a convex function otherwise the resulting space K would not satisfy Condition 9.

We state the remaining regularity conditions as properties of the second best allocation. While this is unappealing, it is required by the generality of our environment. These regularity conditions are often imposed in the literature and for specific applications they can be restated in terms of primitives.

First, 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 10 *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 10.

We next impose two conditions that enable us to apply Lagrangian methods. The first is an interiority restriction.

Condition 11 *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 12 *For a second best allocation* $\{x^*, k^*\}$, *if* $\{1, 2, ..., \tilde{m}\}$ *admissibility constraints are binding at node* s^t, *then the Jacobian*

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

has rank $\tilde{m}$.

Finally, we need a condition which guarantees that the admissibility constraints are well-behaved.

Condition 13 *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.

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 literature. 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 a proportional tax on labor income τ_t^l and a proportional tax 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:

$$\sum_{t=0}^{\infty} q_t^0 \left(\tau_t^l w_t l_t + \tau_t^k r_t k_t - g_t - T_t \right) \geq -b_0, \tag{5}$$

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, given the policies and prices the allocation is optimal for the households and the firms, markets clear, and the government budget constraint (5) 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:

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

The inequality captures the fact that transfers are required to be non-negative.

This restriction 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.

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,[6] 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 $V^{def}\left(k_t\left(s^{t-1}\right), s_t\right)$.[7] The sustainability constraint can then be written as:

---

[6] See Chari and Kehoe (1990).

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

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

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:

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

where $a_t\left(s^t\right)$ is identically equal to 0 in all periods. This formulation is easily extended to the case of aggregate shocks. In section 5.2, we illustrate a key property of the class of policy problems with limited commitment constraints.

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 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 $s^t \in S^t, t \geq 0$:

$$\sum_{j=t}^{\infty} \sum_{s^j \in S^j | s^t} \beta^{j-t} \pi\left(s^j | s_t\right)\left(z\left(x\left(s^j\right), s_j\right) - T\left(s^j\right)\right) = b\left(s^{t-1}\right), \tag{6}$$

where $z\left(x\left(s^t\right), s_t\right)$ 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:

$$T_t\left(s^t\right) \geq 0. \tag{7}$$

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

$$V_t\left(s^t\right) = \sum_{j=t}^{\infty} \sum_{s^j \in S^j | s^t} \beta^{j-t} \pi\left(s^j | s_t\right) T_t\left(s^j\right). \tag{8}$$

The variable $V\left(s^t\right)$ corresponds to the present discounted value of transfers at node s^t. Constraint (7) requires:

$$V_t\left(s^t\right) \geq \beta V_t\left(s^{t+1}\right) \geq 0, \tag{9}$$

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

Define:

$$a_t\left(s^t\right) = \begin{bmatrix} V_t\left(s^t\right) + b_t\left(s^{t-1}\right) \\ -V_t\left(s^t\right) - b_t\left(s^{t-1}\right) \\ b_t\left(s^t\right) \\ V_t\left(s^t\right) \end{bmatrix}, \tag{10}$$

and

$$\begin{aligned} H_1\left(x, s^t\right) &= \sum_{j=t}^{\infty} \sum_{s^j \in S^j | s^t} \beta^{j-t} \pi\left(s^j | s_t\right) z\left(x_t\left(s^j\right), s_j\right), \\ H_2\left(x, s^t\right) &= -H_1\left(x, s^t\right). \end{aligned} \tag{11}$$

Then, constraint (6) can be expressed as:

$$\begin{aligned} H_1\left(x, s^t\right) &\leq a_1\left(s^t\right), \\ H_2\left(x, s^t\right) &\leq a_2\left(s^t\right). \end{aligned} \tag{12}$$

Market incompleteness and the non-negativity of transfers (9) translate into constraints on the correspondence $\Gamma\left(a\left(s^t\right), s_t\right)$ in the first stage problem. The fact that (6) 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:

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

Often, 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 the 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, the structure of the second best problem is such that all distortions can be front-loaded, that is it is possible to reallocate resources and distortions intertemporally without tightening future admissibility constraints.

To formally express this condition, we state a simple Lemma about first best allocations.

Lemma 14 *A first best allocation x^{fb} converges to a unique stationary distribution P^{∞}_{fb}.*

This property is a straightforward implications of the neoclassical production structure and our regularity conditions.

We now state the sufficient condition.

Condition 15 *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 allocation limiting distribution with probability one, i.e.,*

$$\lim_{t \to \infty} \Pr(x_t \in B) = P^{\infty}_{fb}(B)$$

for all measurable $B \subseteq X$.

The connection of Condition 15 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 15 is very easy to verify, since it just requires identifying one admissible allocation that converges to the first best. This allocation does not need to satisfy any optimality conditions and can follow any arbitrary transition path provided it is admissible.

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

Theorem 16 *Let Condition 15 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,*

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

or

2. *The intertemporal distortion fluctuates around* 0,

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

Proof. In Appendix ■

The formal statement of the result clarifies the sense in which there are no *permanent* intertemporal distortions. First, the result holds for the stationary second best allocations. Second, the intertemporal wedge does not need to be identically zero with probability one. Rather, it cannot be strictly positive with probability one or, for that matter, strictly negative.

We prove Theorem 16 in several steps documented 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 that if, eventually, future admissibility constraints do not bind, then there are no permanent intertemporal distortions. This result corresponds to Proposition 19. 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.

Definition 17 *Let* $W(a) : \mathcal{A} \rightarrow \Re$ *be given by*

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

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

Condition 8 ensures that the function $W(a)$ exists and it is bounded.

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

Proposition 18 *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 12 allow us to write the Lagrangian to characterize $W(a)$:

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

where $\lambda\left(s^t\right) \geq 0$ and $\phi\left(s^t\right) \in \Re^m_+$. The Lagrangian multipliers at each node s^t have been normalized by $\beta^t \pi\left(s^t|s_0\right)$.

We adopt the following notation for derivatives:

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

and

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

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

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

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 for this problem are necessary but, without further structure on the choice set, they are generally not sufficient. 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:

$$\mu_i \lambda\left(s^t\right) = \sum_{s^{t+1}} \beta \pi\left(s^{t+1} | s_t\right) \left(\lambda\left(s^{t+1}\right) F_i^k\left(s^{t+1}\right) - \phi\left(s^{t+1}\right)' H_i^k\left(s^{t+1}\right)\right), \tag{13}$$

$$\beta^t \pi\left(s^t | s_0\right) \left\{u_i^c\left(s^t\right) - \mu_i \lambda\left(s^t\right)\right\} = \sum_{j=0}^{\infty} \sum_{s^j} \beta^j \pi\left(s^j | s_0\right) \phi\left(s^j\right)' H_i^c\left(s^j, s^t\right), \tag{14}$$

$$\beta^t \pi\left(s^t | s_0\right) \left\{u_i^l\left(s^t\right) - \mu_i \lambda\left(s^t\right) F_i^l\left(s^t\right)\right\} = \sum_{j=0}^{\infty} \sum_{s^j} \beta^j \pi\left(s^j | s_0\right) \phi\left(s^j\right)' H_i^l\left(s^j, s^t\right). \tag{15}$$

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 19 *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,*

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

or

2. *The intertemporal distortion fluctuates around* 0,

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

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.[8] In particular, 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).

Proposition 19 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. In addition, the proposition is silent about the presence of intratemporal distortions. Indeed, the Ramsey equilibrium with complete markets does not feature any intertemporal distortions but typically displays a permanent wedge on the intratemporal margin.

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 trivially follows from the structure of Problem 1. The second property amounts to strict quasi-concavity of $W(a)$ in a neighborhood of the first best plan.

Proposition 20 *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 ■

---

[8] 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 it would not be possible to assert anything about stationary allocations.

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 15, 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 21 *Let* $\bar{a}\left[k_t\left(s^{t-1}\right), s_t\right] : K \times S \rightarrow \mathcal{A}$ *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. We need the auxiliary plan $\bar{a}$ to be admissible itself, i.e., to satisfy (4) at all nodes.

Definition 22 *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* $s^j \in S^j | s^t$, $j \geq t$.

The next Proposition states that if the second best path for the auxiliary variables a^* reaches the set $A^{fb}\left[k^*_t\left(s^{t-1}\right), s_t\right]$, then no admissibility constraint is ever binding at all continuation nodes.

Proposition 23 *Let* $a^*\left(s^t\right) \in A^{fb}\left[k^*_t\left(s^{t-1}\right), s_t\right]$. *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 ■

Our sufficient condition 15 guarantees that $A^{fb}\left[k_t\left(s^{t-1}\right), s_t\right]$ is non-empty for at least a subset of K.

The next step in the argument is to show that the auxiliary variables a^* will indeed converge to $A^{fb}\left[k^*_t\left(s^{t-1}\right), s_t\right]$. Let K^{fb} denote the support of the capital allocation at the first best stationary distribution.[9] Condition 15 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

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

Proposition 24 establishes a key property of Γ, akin to monotonicity, which allows us to prove convergence of a^* to $\bar{A}^{fb}$.

---

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

Proposition 24 *Let $a_t\left(s^t\right) \in A$. Then for some non-negative scalar $\alpha < 1$ and $\bar{a}_{t+1}\left(s^{t+1}\right) \in \bar{A}^{fb}$,*

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

Proof. In the Appendix ■

If Proposition 24 would 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 $\bar{A}^{fb}$. But this would contradict Condition 15. Therefore, if Condition 15 holds, it is possible to approach the set $\bar{A}^{fb}$.

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 24 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 20.

Proposition 25 *Let $\{x^*, k^*\}$ be a second best allocation 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 16 are now in place. For any second best allocation plan, Proposition 25 implies that the auxiliary variables converge to the set $\bar{A}^{fb}$. By Proposition 23, eventually no future admissibility constraint is binding and the proof of Theorem 16 follows from 19. 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 15 in the simple Ramsey model amounts the ability for the government to accumulate enough assets to finance government consumption via the interest payments on those assets. In a closed economy, this is possible only if private agents can borrow enough so that the government can accumulate enough assets.

It is simple to characterize the asset level 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 :

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

Recall that $a_t = u_t^c \left\{ \left[\left(1 - \tau_t^k\right) r_t \right] k_t + b_t \right\}$ 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:

$$u_t^c \left(c_t^{fb}, l_t^{fb} \right) \left\{ \left[\left(1 - \tau_t^k\right) F_k \left(k_t, l_t^{fb} \right) \right] k_t + \bar{b}_t(k_t) \right\} = \bar{a}_t(k_t).$$

Denote with b_t^p the debt issued by private agents. Then, in equilibrium: $b_t^p + b_t = 0$ at all dates. If a borrowing constraint $\bar{b}$ is imposed on private agents, $b_t^p \leq \bar{b}$ gives rise to an additional constraint on the stage one problem that can be captured in the correspondence Γ. Specifically:

$$-b_t \leq \bar{b},$$

for all t. Hence, 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. Limitations on the government's ability to accumulate assets or borrowing constraints on private agents translate into an admissibility constraint on the government that limits the growth in government assets. If this additional admissibility constraint causes Condition 15 to be violated, 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 convergence properties of optimal policies and their behavior at the limiting distribution 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 limits 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 under the natural limits, Ramsey policies converge almost surely to the first best if the Markov process for the aggregate shock is ergodic. 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. 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. Under an ad hoc asset limit, the optimal allocation typically will not converge and capital taxes are not zero.

We can interpret these findings in terms of Theorem 16. Under the natural debt and asset limits, Condition 15 holds. The first best allocation satisfies the natural debt and asset limits by construction. In addition, the implementability constraint (6) 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.

This reasoning can be extended to economies with idiosyncratic shocks and borrowing constraints, such as Huggett (1993) and Aiyagari (1994). As we discuss above, the borrowing constraints on the households translate into asset constraints on the government. If these constraints are ad hoc, then the first best steady state may not be admissible even if the government can employ agent-specific transfers. If the borrowing constraints on the household are not too stringent and agent specific transfers are available, then our sufficient condition holds and there will be no permanent intertemporal distortions.

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 optimality of a zero capital tax 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 $O\left(\cdot\right)$ 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.

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 by definition. 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 15 is satisfied.[10]

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

The contribution of this result lies in its generality. As is well known, the properties of optimal allocations under limited commitment typically depend on the formulation of the outside option. Since the outside option is off the equilibrium path, it is hard to discriminate between the many possible specifications. This drawback has undermined the applied value of optimal policy models with limited commitment. The application of Theorem 16 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. Phelan and Stacchetti (2001) study the best sustainable equilibrium in an economy where the government is forced to run a balanced budget

---

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

in every period. The government's inability to save renders any continuation first best allocation unattainable. 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:

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

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:

$$U_0\left(\theta^T; \theta^T\right) \geq U_0\left(\theta^T; \tilde{\theta}^T\right), \tag{16}$$

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 15 does not hold. It can be shown that the second best allocation displays the following three properties[11]. 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:

$$u'\left(c_t\left(\theta^t\right)\right) \leq \beta R_t E_t u'\left(c_{t+1}\left(\theta^{t+1}\right)\right). \tag{17}$$

The inequality is strict when agents are risk averse and face idiosyncratic risk in the subsequent periods, 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 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*.[12]

It is important to note that 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

---

[11] See Albanesi and Sleet (2006).

[12] 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).

impossibility of eventually eliminating all distortions, that is, the fact that the first best can never be attained.

6 Concluding Remarks

Our analysis leaves open several questions. The first pertains to the generality of our class of problems. It is possible that a result similar in nature to the one we derive in fact holds for a larger class of environments than the one we consider. On the other hand, we have imposed restrictions on history dependence in the admissibility constraints that are important for our result. It is easy to find policy problems with an arbitrary degree of history dependence in the constraints where our result does not hold. One example is a Ramsey model where the capital tax rate must be constant at all dates. Additional degrees of history dependence may also arise endogenously from the structure of the constraints. While our condition on the limits of history dependence is quite general, it does not identify a general class of policy problems that are not admitted in the formulation.[13]

Another question is whether it might be possible to identify a necessary condition for zero intertemporal distortions in the steady state on the class of environments that we consider. Given that our condition for the absence of intertemporal distortions is sufficient, it is possible to construct examples within our class of environments in which our condition does not hold and yet there are no intertemporal distortions. One simple example is the Ramsey model with no uncertainty and a balanced budget constraint imposed on the government. The optimal capital tax is zero in this model, as shown in Stockman (2001). This result is merely a function of the fact that a zero capital tax maximized fiscal revenues, so it is statically optimal.

A more interesting example is given by Aiyagari's (1995) incomplete markets economy. Individuals experience idiosyncratic productivity shocks and they can save by holding capital and government debt but must maintain positive net worth. Aiyagari (1995) shows that the evolution of aggregate capital at the optimal allocation follows the modified golden rule, so that there is no aggregate intertemporal wedge. To implement this outcome, it is necessary for the capital income tax to be positive, since the constraint on agents' net worth gives rise to a precautionary motive for holding capital. Moreover, the individual intertemporal margin does not exhibit permanent distortions. At the stationary distribution, the mass of agents with binding net worth constraint will feature a negative intertemporal wedge, while all other agents will feature a positive intertemporal wedge, corresponding to the optimal capital tax. Clearly, any first best continuation in this economy features full insurance and is therefore not attainable given that individual specific transfers are not available. However, the constraint on private agents allows them to borrow from the government as long as their net worth remains positive. If capital is high enough, the government could accumulate enough assets to remove the need for collecting distortionary taxes.

This observation suggest that a weaker version of our sufficient condition may hold in this

---

[13] We have also excluded for simplicity economies with history dependence in preferences. The environment can be extended to encompass a more general class of non time-separable preferences, such as habit formation or Kreps-Portheus utility. We leave this extention to future work.

economy. In our current formulation, the condition requires that a first best continuation is eventually admissible for the distribution of consumption as well as for the aggregate allocation. We conjecture that in fact it may be sufficient that only the aggregate allocation satisfies this requirement, as in Aiyagari (1995). We plan to investigate this conjecture in future work.

Finally, in many economies with permanent intertemporal distortions, the second best allocation does not converge. Private information economies constitute a particularly interesting example of this property. We plan to explore the link between convergence to a stationary distribution and the presence of intertemporal distortions in second best economies in future work.[14]

References

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

[2] Aiyagari, S Rao. 1994. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics* 109(3): 659-84.

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

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

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

[6] 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.

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

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

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

---

[14] A non-degenerate limiting distribution can be shown to exist in private infomation economies by imposing a lower bound on continuation utility, as in Atkeson and Lucas (1995). A non-degenerate limiting distribution of utility can also exist if the government discounts the future at a lower rate than private agents, as shown in Phelan (2006) and Farhi and Werning (2006). Sleet and Yeltekin (2006) show that limited commitment can provide a rationale for such differential discounting.

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

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

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

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

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

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

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

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

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

[19] Huggett, Mark. 1993. The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies. *Journal of Economic Dynamics and Control* 17: 953- 969.

[20] 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.

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

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

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

[24] 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.

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

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

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

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

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

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

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

[32] 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 $s^t \in S^t$, $t \geq 0$,

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

for all measurable subsets $B \subseteq X$.

We also include a formalization of Condition 10. Define the set of allocations

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

where $P_x^{\infty} \in \mathcal{P}_x^{\infty}$, i.e., P_x^{∞} a stationary distribution. Then, Condition 10 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 19. 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 (14) 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

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

so (14) can be written as

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

where

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

Similarly, the necessary first order condition for $l_{it}\left(s^t\right)$ (15) becomes

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

The necessary first order condition for $k_{it}\left(s^{t-1}\right)$ (13) is simpler, since $\phi\left(s^{t+1}\right)=0$,

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

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

$$\Omega_i^c(\sigma)=1-\frac{1}{u_i^c(\sigma)} \bar{\phi}^{\prime}\left[\begin{array}{c}
h_1^{c_i}(x(\sigma), \sigma) \\
h_2^{c_i}(x(\sigma), \sigma) \\
\ldots \\
h_n^{c_i}(x(\sigma), \sigma)
\end{array}\right]$$

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

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

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

where

$$\omega_i\left(\sigma, s^{\prime}\right)=\beta \pi^{\infty}\left(s^{\prime} \mid s\right) u_i^c\left(\sigma^{\prime}\right) F_i^k\left(\sigma^{\prime}\right)$$

and $\sigma^{\prime}$ is given by the law of motion for $k_t\left(s^{t-1}\right)$. Since $\omega_i\left(\sigma, s^{\prime}\right)>0$, 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

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

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

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

Proof of Proposition 20. If $\{x, k\}$ is admissible for plan a, it is admissible for any plan $a^{\prime}$ such that $a_t^{\prime}\left(s^t\right) \geq a_t\left(s^t\right)$ so the first point follows.

For the second point, let feasible allocation $\{x, k\}$ be such that $U\left(x, s_0\right)=W(a)$ and $U\left(x, s_0\right)=W(a+\varepsilon)$ 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 $\phi\left(s^t\right)$ equal to zero.

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

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

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 23. Let $\left\{\tilde{k}, \tilde{x}\right\}$ be a first-best continuation allocation of $\left\{k_t^*\left(s^{t-1}\right), s_t\right\}$ and $\tilde{a}$ satisfy that $\tilde{a} \geq \bar{a}\left[k_t^*\left(s^{t-1}\right), s_t\right]$ 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 $a_t^*\left(s^t\right) \in A^{fb}\left[k_t^*\left(s^{t-1}\right), s_t\right]$. 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

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

Otherwise, Condition 13 would not be satisfied. Then,

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

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 $W\left(\tilde{a}\right) > W\left(a^*\right)$ would contradict $\{x^*, k^*\}$ being a second best plan ■

Proof of Proposition 24. If $a_t\left(s^t\right) \in int\left(\Gamma\left(a_t\left(s^t\right), s_t\right)\right)$ the proof is straightforward. If $a_t\left(s^t\right) \notin int\left(\Gamma\left(a_t\left(s^t\right), s_t\right)\right)$, then $a_t\left(s^t\right)$ is an adjacent point to $\Gamma\left(a_t\left(s^t\right), s_t\right)$ as $a_t\left(s^t\right) \in \Gamma\left(a_t\left(s^t\right), s_t\right)$ 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 15, there must exist some $x \in \chi_p$ with $y \in \Gamma\left(x, s_t\right)$, $y \notin \chi_p$ 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 $z \in \Gamma\left(z, s_t\right)$, 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 25 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 26 *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 15. If $a_t^*\left(s^t\right) \in \bar{A}^{fb}$ then Proposition 23 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 13 implies admissibility constraint n cannot be binding ■

Proof of Proposition 25. 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 26 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 24, 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 27 *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 20 establishes that, if $a' \geq a$, then $W\left(a'\right) \geq W\left(a\right)$, we can use Lemma 27 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 $W\left(a\right) = W\left(a^*\right)$, otherwise we would contradict a^* being a second best plan.

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

Lemma 28 *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 27 ■

Finally, we close the argument here. Lemma 28 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 16

Proof. Let $\{x^*, k^*\}$ be a second best allocation plan. By Proposition 25 and Condition 15, 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 23 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 19 ■