Time-consistent optimal fiscal policy over the business cycle

Zhigang Feng
University of Nebraska at Omaha, zfeng@unomaha.edu

Quantitative Economics 6 (2015), 189–221 1759-7331/20150189

Time-consistent optimal fiscal policy over the business cycle

ZHIGANG FENG
Department of Economics, University of Illinois at Urbana–Champaign

This paper examines a dynamic stochastic economy with a benevolent government that cannot commit to its future policies. I consider equilibria that are time-consistent and allow for history-dependent strategies. A new numerical algorithm is developed to solve for the set of equilibrium payoffs. For a baseline economy calibrated to the U.S. economy, the capital income tax with the highest social welfare is slightly procyclical, while the labor income tax is countercyclical. Compared with the data, this equilibrium provides a better account of the cyclical properties of U.S. tax policy than other solutions that abstract from history dependence. The welfare cost of no commitment is about 0.22% of aggregate consumption as compared to the Ramsey allocation with full commitment.

KEYWORDS. Optimal fiscal policy, business cycle, recursive game theory, time consistency.

JEL CLASSIFICATION. E61, E62, H21, H62, H63.

1. INTRODUCTION

There has been a burgeoning literature on time-consistent optimal fiscal policies since Kydland and Prescott (1977) pointed out the time-inconsistency problem of optimal policies in the absence of government commitment. Most existing analysis focuses on characterizing Markov perfect equilibria (MPE) government policies. The current paper contributes to the literature by providing the first quantitative assessment of business-cycle properties of time-consistent and history-dependent fiscal policies. I consider a stochastic growth model with a balanced budget constraint for the government. Instead of characterizing MPE, I allow for history-dependent government policies and introduce a reputation mechanism (essentially, punishment strategies) in the spirit of Chari and Kehoe (1990).[1] The paper also develops a new numerical method to characterize

Zhigang Feng: z.feng2@gmail.com
I thank Felix Kübler for encouragement and guidance. Special thanks go to Jinhui Bai and Begona Domínguez for some extensive discussions on MPE and SSE, and to Manuel Santos and Anne Villamil for detailed comments and editorial help. I am indebted to V. Bhaskar, Sheng Guo, Matt Hoelle, David Kelly, Rui Li, Jianjun Miao, Udara Peiris, Adrian Peralta-Alva, Walt Pohl, Karl Schmedders, David Stockman, Kai Zhao, Fabrizio Zilibotti, the editor José-Víctor Ríos-Rull, and two referees for very helpful discussions and comments. I am grateful to NCCR-FINRISK and the Swiss National Computing Center for generous support. I also wish to thank seminar participants at 2012 Midwest Macro, 2013 SAET, St. Louis Fed, UC3M, U. of Delaware, U. of Konstanz, U. of Miami, U. of Zurich, and FIU. All remaining errors are mine.

[1] Chari and Kehoe (1990) study how the reputation mechanism can help to characterize time-consistent optimal policies. I refer readers to their paper for details. I provide a brief review below.

DOI: 10.3982/QE370

the whole set of sequential sustainable equilibria (SSE). All SSE have been ranked by their corresponding social welfare, with the best SSE (BSSE) generating the highest social welfare and the worst (WSSE) providing the lowest one among all SSE. Numerical simulations highlight that BSSE provides a better account of the cyclical behavior of U.S. tax policy than equilibrium solutions abstracting from history dependence and reputation.

The economy is populated by a continuum of representative infinitely lived households that maximize lifetime expected utility by choosing a sequence of consumption, leisure, and investment, taking prices and tax sequences as given. Firms maximize profits using a neoclassical production technology. There is a benevolent government that seeks to maximize social welfare. The endogenous consumption of public goods is financed by linear taxes on capital and labor income. There is no commitment technology in the sense that the government has an option to choose its policies at every date–event.

I describe the economy as a dynamic game between the government and the representative household, with one-sided lack of commitment (from the government), in which reputation sustains equilibrium policies. I study SSE, in which the government chooses tax policy strategically and such policy is sequentially rational. That is, the government chooses a tax policy to maximize social welfare at each date–event given that the household behaves optimally. Likewise, the representative household behaves competitively and forecasts that future policies are sequentially rational for society. The reputation mechanism ensures that if the government deviates from equilibrium policies and pursues different policies, then typically a higher tax rate on capital income will be punished with a low continuation value.[2] This happens because the deviation changes the household's expectations about future policies. The model is set up so that this new expectation delivers sufficiently low continuation values to discourage deviations. Thus, the incumbent government will choose to follow the equilibrium policies as long as either the continuation value or the current return associated with deviating is sufficiently low. Indeed, a lower continuation value raises the cost of deviating and the value of cooperation, while a lower current return reduces the value of deviation. Therefore, SSE solves a dynamic programming problem of the benevolent government subject to an incentive-compatibility constraint for the government at any history. This constraint is endogenous since the continuation value of deviation depends on future decision variables and policies.[3]

In this environment, optimal capital income taxes in the presence of commitment are zero in expectation and countercyclical. Positive capital income taxes are equivalent to ever-increasing taxes on future consumptions that create permanent distortions to the economy, and that violate Atkinson and Stiglitz's (1976) result that optimal taxes should be equal across all final consumption goods. The government uses a counter-

[2]The government's incentive to raise the capital income tax rate comes from the inelastic nature of installed capital.

[3]If the punishment associated with deviation can be determined up front (i.e. Chari and Kehoe (1990)) or can be explicitly imposed (i.e. Benhabib and Rusticchini (1997) and Marcet and Marimon (2011)), the techniques of optimal control may be used to solve optimal taxation without commitment. However, this is not necessarily the case for some interesting applications.

cyclical capital income tax to smooth the consumption of public goods. It does so because the efficiency cost over the business cycle is low; more specifically, the short-run supply elasticity of capital is very small (Judd (1993)). In contrast, labor income taxes bear most of the burden of financing government spending. They become procyclical to smooth the household's after-tax labor income.

In SSE, optimal capital income taxes are positive in the steady state. Tax rates on capital income decrease when the value of the discount factor increases.[4] In the absence of commitment, the inelastic nature of installed capital creates an incentive for the government to deviate from the zero capital income tax. This will increase its current return, but at a cost of a lower continuation value. A higher value of the discount factor means that the future return carries larger weight and makes deviation more painful. Domínguez and Feng (2012) find that less-elastic labor supply reduces the benefit of raising the current capital income tax rate and hence the incentive to deviate. These findings imply that the Ramsey allocation may be sustainable when discounting is sufficiently low or the labor supply is sufficiently inelastic.

Optimal capital income taxes in the absence of commitment are slightly procyclical and labor income taxes are mildly countercyclical. When a negative total factor productivity (TFP) shock hits the economy, the government may want to raise the tax rate on capital income to smooth consumption of public goods. This is optimal in Ramsey and MPE since the short-term efficiency cost of capital income taxation is low, as explained before.[5] When reputation is at play, such a policy will alter the household's expectations about future policies and incur a lower continuation value as a punishment to deviation. Hence, the government faces a trade-off between the gain from tax smoothing and the loss from a bad reputation. This trade-off explains why a procyclical capital income tax may be optimal for the government in BSSE, but not in WSSE, in which the government loses its reputation.

Finally, the cost of no commitment is on the order of 0.22% of total consumption for a set of parameter values calibrated to the U.S. economy. The welfare cost is measured by the percentage of consumption that the representative household is willing to pay for moving from the Ramsey equilibrium to BSSE. I also find that the average tax rate on capital income is around 5% in BSSE.

Related literature. First, the paper is related to the work of Chari and Kehoe (1990), Chang (1998), Phelan and Stacchetti (2001), and Reis (2013), all of whom study optimal policies without commitment by characterizing SSE. The current contribution is different in two important aspects. It introduces aggregate shocks, and provides a numeri-

[4]This is related to Chari and Kehoe (1990), who consider a similar model without capital income tax so that it is possible to derive the payoff of the worst SSE explicitly. They show that the discount factor is the key determinant in sustaining the Ramsey outcome. Contrary to theirs, the payoff of the worst SSE in this research has to be computed as part of the equilibria due to the existence of capital income tax and capital accumulation.

[5]Papers that characterize the Ramsey equilibrium include Chari, Christiano, and Kehoe (1994, 1995), Stockman (2004), and Zhu (1992). Klein and Ríos-Rull (2003) study the business-cycle properties of taxation in MPE.

cal algorithm to approximate the set of equilibrium values.[6] Some other papers study the same class of economies but focus on MPE (e.g., Klein and Ríos-Rull (2003), Klein, Krusell, and Ríos-Rull (2008), Martin (2010), and Ortigueira (2006)). MPE strategies are simple to characterize, since government policies are not history-dependent by assumption. However, it is not clear how much insight may be lost by looking at these particular equilibria as MPE can be a special case of SSE (i.e. Martin (2010)).

Second, this paper is related to research on endogenous fiscal policies over the business cycle (e.g., Azzimonti and Talbert (2011), Bachmann and Bai (2013a, 2013b), Barseghyan, Battaglini, and Coate (2010), Chari, Christiano, and Kehoe (1994), Debortoli and Nunes (2010), Klein and Ríos-Rull (2003), and Stockman (2004)). I focus on equilibria that are supported by reputation and satisfy time consistency.

Third, this research is also related to Benhabib and Rusticchini (1997) and Marcet and Marimon (2011). These papers explicitly impose additional constraints on the standard optimal taxation problem, such as not to allow deviations from the prescribed sequence of taxes. Therefore, one can use the techniques of optimal control to solve for optimal taxation without commitment. Instead of assuming an exogenous penalty, this paper develops a numerical algorithm that recursively determines the equilibrium continuation value of government deviation.

The paper is organized as follows. Section 2 explains the economic environment. Section 3 discusses several equilibrium concepts. Section 4 sketches a method to recursively characterize the set of SSE. Section 5 provides a numerical algorithm for the computation and simulation of the economy. The economy is calibrated to match certain features of the U.S. economy in Section 6. Section 7 presents steady-state and business-cycle properties, and Section 8 concludes. Some proofs and further computational details are gathered in the Appendix. Replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/370/code_and_data.zip.

2. Economic environment

The economy is populated by a measure 1 of identical, infinitely lived households and a continuum of competitive firms. The primitive characteristics of the economy follow a stationary Markov chain with support S. Time and uncertainty are represented by a countably infinite tree Σ. For an initial shock s_0, each node of the tree, $\varsigma \in \Sigma$, is a finite history of shocks $\varsigma = s^t = (s_0, s_1, \dots, s_t)$. Let $\pi(s^t)$ be the probability of history s^t and let $\pi(s^{t+1}|s^t)$ be the conditional probability of moving from history s^t to s^{t+1}.

[6]In an incomplete working paper, Fernandez-Villaverde and Tsyvinski (2002) consider an environment similar to that in the current paper. My work complements Fernandez-Villaverde and Tsyvinski (2002) by providing a quantitative analysis of the business-cycle properties of time-consistent optimal taxation. It also complements Fernandez-Villaverde and Tsyvinski (2002) by showing that the sequential and recursive problems are equivalent when there is an aggregate shock to the economy. This is a key result to extend the analysis of Phelan and Stacchetti (2001) into a stochastic setting.

2.1 *Firms*

In each time period $t = 0, 1, 2, \ldots$, firms produce final goods $y(s^t)$, using capital $k(s^t)$ and labor $l(s^t)$, by a neoclassical production function in which TFP depends on the current shock:

$$y(s^t) = A(s_t)F(k(s^t), l(s^t)). \tag{1}$$

Every competitive firm maximizes one-period profit by setting rental rates for inputs equal to their marginal products:

$$r(s^t) = A(s_t)F_k(k(s^t), l(s^t)), \tag{2}$$

$$w(s^t) = A(s_t)F_l(k(s^t), l(s^t)). \tag{3}$$

The production of the final good can be used for private consumption $c(s^t)$, public consumption $g(s^t)$, and investment $i(s^t)$. The law of motion for capital $k(s^t)$ is given by

$$k(s^{t+1}) = i(s^t) + (1-\delta)k(s^t), \tag{4}$$

where δ is the depreciation rate on capital.

2.2 *Households*

The representative household derives satisfaction from consumption and leisure as given by a time-separable utility function

$$\sum_{t=0}^{\infty}\sum_{s^t}\beta^t\pi(s^t)\big[u\big(c(s^t), 1-l(s^t)\big) + G\big(g(s^t)\big)\big], \tag{5}$$

where $\beta \in (0, 1)$ is the discount factor. The instantaneous utility function $u: \mathbf{R}_+^2 \to \mathbf{R}$ and the function $G: \mathbf{R}_+ \to \mathbf{R}$ are strictly concave and differentiable, and satisfy the Inada conditions.

Each household takes the government policy $\{\tau_k(s^t), \tau_l(s^t)\}$ as exogenously given and chooses a sequence of consumptions and labor supplies $\{c(s^t), l(s^t)\}$ to maximize the objective in equation (5) subject to the budget constraint

$$c(s^t) + k(s^{t+1}) = \big(1-\tau_l(s^t)\big)w(s^t)l(s^t) + \big[1 + \big(1-\tau_k(s^t)\big)\big(r(s^t)-\delta\big)\big]k(s^t), \tag{6}$$

$$k(s_0) > 0, \tag{7}$$

where $\tau_k(s^t)$ and $\tau_l(s^t)$ are the tax rates on capital gains and labor income, respectively. These tax rates are set by the government. As explained later, the proceeds of these taxes will determine government spending. Note that $(1-\tau_k(s^t))(r(s^t)-\delta)$ means that capital depreciation is tax-deductible. Let $k(s_0)$ be the initial endowment of capital of every household.

2.3 *Government*

The government is benevolent and sets up a sequence of tax policies to maximize social welfare as given by the objective in (5). The government faces a balanced-budget constraint and finances a stream of endogenous consumption of public goods using linear taxation on capital and labor incomes. More specifically, government consumption $g(s^t)$ is defined as

$$g(s^t) = \tau_l(s^t)w(s^t)l(s^t) + \tau_k(s^t)[r(s^t) - \delta]k(s^t). \tag{8}$$

Tax rates are restricted to lie in a given interval $T = \{[\underline{\tau}, \overline{\tau}]\}$.[7]

A policy τ is a stochastic process $\tau = \{(\tau_k(s^t), \tau_l(s^t))\}_{t=0}^{\infty}$ for capital and labor income tax rates. A competitive equilibrium is defined for a given sequence of taxes.

Definition 1. Let $Y\{k_0, s_0, \tau\}$ be the economy starting with initial capital k_0, the shock s_0, and tax policy τ. A *competitive equilibrium* for $Y\{k_0, s_0, \tau\}$ is a sequence of allocations $\{l(s^t), c(s^t), k(s^{t+1})\}$ and prices $\{r(s^t), w(s^t)\}$ that satisfy the following statements.

1. Given prices and tax rates, $\{l(s^t), c(s^t), k(s^{t+1})\}$ solves the representative household's optimization problem.
2. Input prices equate marginal product.
3. The government satisfies its budget constraint period by period.
4. Good markets clear at all nodes.

3. Equilibrium policies

In this section, I characterize several equilibrium concepts emerging from the optimal taxation literature. In a later section, I present differences among these equilibria in terms of welfare implications and allocations. From now on, let x and x' represent the current and future value of variable x, while $\hat{x}$ represents an alternative to x. Let $\mathbb{E}$ be the expectation operator.

3.1 *The full-commitment economy*

The Ramsey equilibrium assumes that the government has a commitment device that ties the hands of all future governments (i.e. Chari and Kehoe (1999)). At the initial period $t = 0$, the government sets the tax policy once and for all future periods so as to maximize the social welfare given by equation (5). The Ramsey equilibrium is defined as follows.

Definition 2. A *Ramsey equilibrium* is a tax policy τ, allocations $a(\tau) = \{l(s^t; \tau), c(s^t; \tau), k(s^{t+1}; \tau)\}$, and prices $p(\tau) = \{r(s^t; \tau), w(s^t; \tau)\}$ that satisfy the following statements.

[7] An upper bound on the tax rate is needed to show the equivalence of the recursive and sequential problems for the representative household. The lower bound on the tax rate is used to bound the initial value correspondence.

1. Policy τ maximizes the representative household's utility (5) subject to government budget constraint (8), where allocations and prices are given by $a(\tau)$ and $p(\tau)$.

2. For every policy $\hat{\tau}$, allocation $a(\hat{\tau})$ solves the household's problem given $\hat{\tau}$ and $p(\hat{\tau})$.

3. For every policy $\hat{\tau}$, the price of each factor is equal to its marginal product: $r(\hat{\tau}) = F_k(\hat{\tau})$ and $w(\hat{\tau}) = F_l(\hat{\tau})$.

Without government commitment to future policies, the Ramsey policy, as defined above, suffers from a time-inconsistency problem (i.e. Kydland and Prescott (1977), Fischer (1980)). In other words, the Ramsey policy may not be implementable. In what follows, I assume away commitment and focus on time-consistent policies.

3.2 *The economy without commitment*

In an economy without commitment, in contrast to the Ramsey equilibrium, the government is now assumed to behave strategically and to set up tax rates period by period.

The timing of the model is as follows. At the beginning of every period t, shock s_t is revealed. Next, the government sets tax rates for the current period $\{\tau_k(s^t), \tau_l(s^t)\}$. Then the representative household simultaneously chooses the labor input $l(s^t)$. Given the aggregate capital chosen in the previous period and the labor input, the market determines prices for capital and labor. Finally, each household independently allocates after-tax income between consumption and investment for production in the following period. The government uses the tax revenue to finance the consumption of the public good. Figure 1 provides a graphical representation of the timing.

3.2.1 *Sequential sustainable equilibria* In the spirit of Chari and Kehoe (1990), I describe the economy as a dynamic game between the government and the representative

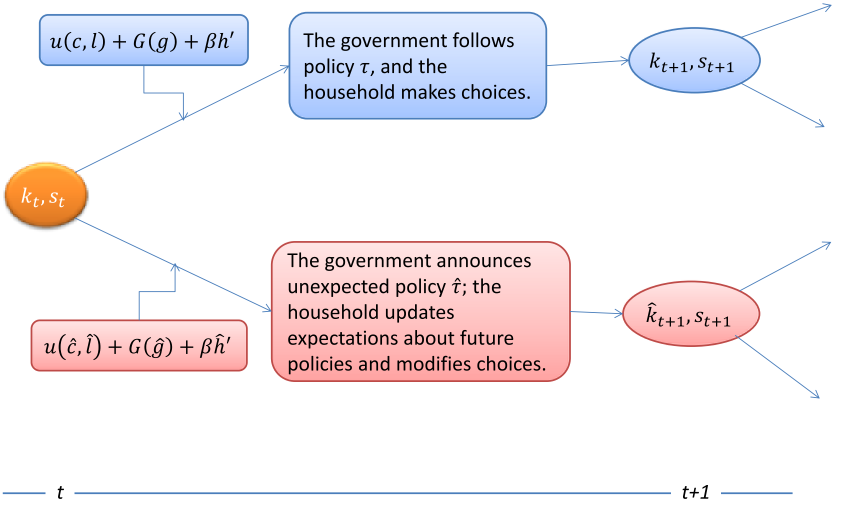

FIGURE 1. Timing of the game between the government and the household.

household. Let $\Gamma(k_0, s_0)$ denote the game in which all households are endowed with initial capital k_0 and the shock value is s_0. Let $\zeta^t = (\zeta_0, \zeta_1, \dots, \zeta_t)$ be the public history of the game, where $\zeta_t = (s_t, \tau_k(s^t), \tau_l(s^t), c(s^t), l(s^t), k(s^{t+1}))$ includes the history of shocks, government policies, and the representative household's choices. I focus on symmetric strategy profiles, as in Phelan and Stacchetti (2001), in which all households choose the same actions along the equilibrium path. A symmetric strategy profile for $\Gamma(k_0, s_0)$ is a pair of strategies $\sigma = (\sigma_C, \sigma_G)$, where σ_C is the strategy for the household and σ_G is the strategy for the government. Both are measurable functions. Strategy σ_G maps publicly observed history ζ^{t-1} and the current shock s_t into tax rates for date–event s^t, namely $(\tau_k(s^t), \tau_l(s^t)) = \sigma_G(\zeta^{t-1}, s_t)$. Similarly, strategy σ_C specifies $c(s^t)$, $l(s^t)$, and $k(s^{t+1})$ as functions of expanded history $(\zeta^{t-1}, s_t, \tau_k(s^t), \tau_l(s^t))$; that is, $(c(s^t), l(s^t), k(s^{t+1})) = \sigma_C(\zeta^{t-1}, s_t, \tau_k(s^t), \tau_l(s^t))$. Let $\sum(k_0, s_0) = \sum_C(k_0, s_0) \times \sum_G(k_0, s_0)$ denote the set of all symmetric strategy profiles for $\Gamma(k_0, s_0)$, where $\sum_C(k_0, s_0)$ represents the set of strategies for the household, and $\sum_G(k_0, s_0)$ represents the set of strategies for the government.

To facilitate the definition of SSE, let the value of a strategy σ_G (for the government) be

$$\Phi_G(k_0, s_0, \sigma) := \sum_{t=0}^{\infty} \beta^t \mathbb{E}\big[u\big(c\big(s^t, \sigma\big), 1 - l\big(s^t, \sigma\big)\big) + G\big(g\big(s^t, \sigma\big)\big)\big]. \tag{9}$$

A SSE is defined as follows.

Definition 3. A symmetric strategy profile σ of the game $\Gamma(k_0, s_0)$ is an *SSE* if for any $t \geq 0$ and history ζ^{t-1}, the following statements hold.

1. $\Phi_G(k(s^t), s_t, \sigma|_{\zeta^{t-1}}) \geq \Phi_G(k(s^t), s_t, (\sigma_c|_{\zeta^{t-1}}, \gamma))$ for any strategy $\gamma \in \sum_G(k(s^t), s_t)$ for the government.

2. $\{c(s^j), l(s^j), k(s^{j+1}), r(s^j), w(s^j), g(s^j)\}_{j=t}^{\infty}$ is a competitive equilibrium for $\Upsilon\{k_{t-1}, s_t, \tau_{s^t}\}$, where $\tau_{s^t} := \{\tau_k(s^j), \tau_l(s^j)\}_{j=t}^{\infty}$, $(\tau_k(s^t), \tau_l(s^t)) \in \sigma_G(\zeta^{t-1}, s_t)$, and $(c(s^t), l(s^t), k(s^t)) \in \sigma_C(\zeta^{t-1}, s_t, \tau_k(s^t), \tau_l(s^t))$.

In line with Phelan and Stacchetti (2001), let $\sigma|_{\zeta^{t-1}}$ denote the strategy profile in SSE with history ζ^{t-1}. Let $(\sigma_c|_{\zeta^{t-1}}, \gamma)$ represent the strategy profile in which the household plays a SSE strategy under history ζ^{t-1} while the government plays an alternative strategy. The first condition in the above definition says that the continuation payoff for the government's strategy σ_G is higher than the payoff from any deviation to a different strategy. The second condition requires that the household always responds to a government strategy with decisions that imply a competitive equilibrium, since this is the situation that is compatible with feasibility and optimality.

The SSE is characterized by strategy profiles, which are mappings from the entire history to current choices of the government and the household. The history dependence poses serious challenges in terms of solving equilibria, because one may have to keep track of the infinite histories (i.e. Phelan and Stacchetti (2001)). In the next section, I explain how to simplify the problem by summarizing the history of strategies using two additional endogenous equilibrium values.

3.2.2 *Markov perfect equilibrium* An alternative literature avoids history dependence by limiting the government's policy choices. This approach assumes that the government sets its tax policies solely conditional on some predetermined variables, such as capital stock and current value of shocks. In line with Klein and Ríos-Rull (2003), an MPE is defined as follows.

Definition 4. An *MPE* for the economy consists of tax policy functions $\tau_k = \phi^k(K, s)$ and $\tau_l = \phi^l(K, s)$ for the government, a transition function $K' = Q(K, s; \phi^k, \phi^l)$, an equilibrium value function $v(k, K, s; \phi^k, \phi^l)$ for the representative household, and individual policy functions $k' = q^k(k, K, s; \phi^k, \phi^l)$ and $l = q^l(k, K, s; \phi^k, \phi^l)$, such that the following statements hold.

1. For any given ϕ^k and ϕ^l, the value function and policy functions solve the representative household's problem

$$\begin{aligned} &v(k, K, s; \phi^k, \phi^l) \\ &\quad = \max_{(c,k',l)} \{u(c, 1-l) + G(g) + \beta\mathbb{E}[v(k', K', s'; \phi^k(K', s'), \phi^l(K', s'))]\} \end{aligned}$$

s.t.

$$\begin{aligned} &c + k' = (1-\tau_l)wl + [1 + (1-\tau_k)(r-\delta)]k, \\ &K' = Q(K, s; \phi^k, \phi^l). \end{aligned}$$

2. $Q(K, s; \phi^k, \phi^l) = q^k(k, K, s; \phi^k, \phi^l)$.

3. The government tax policy functions maximize social welfare on the equilibrium path:

$$\{\phi^k(K, s), \phi^l(K, s)\} \in \arg\max_{\{\tau_k, \tau_l\}} v(K, K, s; \phi^k, \phi^l).$$

The first part of the equilibrium definition says that the household's policy functions are optimal choices for any tax policies ϕ^k and ϕ^l. These policies determine the current and future tax rates on capital and labor income as functions of the capital stock and shock. The second condition reflects the consistency between the household's and aggregate behaviors. The last part of the definition specifies the optimal choice of the current government as functions of today's state variables.

One advantage of MPE over SSE is that the equilibrium is easy to characterize since it is not history-dependent. If the equilibrium is unique, one can compute the model by solving the generalized Euler equation (GEE) or by using methods based on the value function.[8]

However, this approach has its limitations. To the best of my knowledge, there is no formal existence or uniqueness proof for MPE as defined above. Numerical methods built upon the continuity and uniqueness of equilibrium may encounter additional

[8]For the deterministic economy, GEE is solved in Klein, Krusell, and Ríos-Rull (2008) using a perturbation method. Martin (2010) and Ortigueira (2006) solve similar models with a projection method. For the stochastic economy, Azzimonti and Talbert (2011) use a projection method, and Bachmann and Bai (2013b) develop a method based on value function iteration.

problems such as lack of convergence and reliance of the computed equilibrium on the initial guess (i.e. Peralta-Alva and Santos (2010)). Even though researchers can carefully choose the initial guess and achieve convergence, they still may not know whether the algorithm has converged to a true equilibrium. In contrast, I develop a method that has good convergence properties and computes all SSE. One advantage of my approach is that it allows us to pick some particular equilibria and study corresponding government behaviors; see Section 7 below.

4. Recursive formulation of equilibria

In an infinitely repeated game, equilibria are defined by strategy profiles that map the infinite history into the current optimal choices of each player. Abreu, Pearce, and Stacchetti (1990) (APS henceforth) show that it is sufficient to characterize the lifetime payoff of each player in such models instead of infinite dimensional strategy profiles. Furthermore, APS develop a recursive operator whose fixed point yields the set of equilibrium payoffs.

A naive application of APS to models in this paper will require keeping track of payoffs for a continuum of identical households, which is a formidable task. Chang (1998) and Phelan and Stacchetti (2001) make important progress in characterizing the equilibria for economies similar to mine by incorporating the insight of Kydland and Prescott (1980) into APS. The basic idea comes from the following observation. Since each household is anonymous and cannot affect aggregate prices, the household's problem can be written in a recursive way by including "marginal value of investment in capital" as an additional state variable. Henceforth, to characterize the equilibria, it is enough to keep track of the promised payoffs for the government, together with the promised marginal value of investment in capital. In what follows, I explain how to recursively formulate the dynamic problem described in Section 3.2.

4.1 *The set of equilibrium values*

In equilibrium, every household's optimal choices must satisfy the first-order conditions

$$u_c(s^t) - \beta \sum_{s^{t+1}} \pi(s^{t+1}|s^t)\{[1 + (1 - \tau_k(s^{t+1}))(r(s^{t+1}) - \delta)]u_c(s^{t+1})\} = 0, \tag{10}$$

$$u_l(s^t) - (1 - \tau_l(s^t))w(s^t)u_c(s^t) = 0, \tag{11}$$

where u_c and u_l denote the marginal utility with respect to consumption and leisure. Equation (10) is the intertemporal optimality condition, and equation (11) equates the marginal utility of consumption to the marginal utility of leisure.

I reformulate the household's problem recursively following the strategy suggested by Kydland and Prescott (1980) (see Marcet and Marimon (2011), Phelan and Stacchetti (2001), and Feng, Miao, Peralta-Alva, and Santos (2014) for recent applications). Let $m(s^t)$ be the marginal value of investment in capital:

$$m(s^t) := u_c(s^t)[1 + (1 - \tau_k(s^t))(r(s^t) - \delta)]. \tag{12}$$

Then, for any s^t, $k(s^t)$, $\{m(s^{t+1})\}$ and some tax policy τ, each household solves

$$\max_{c(s^t),l(s^t),k(s^{t+1})} u\big(c\big(s^t\big), l\big(s^t\big)\big) + \beta \sum_{s^{t+1}} \pi\big(s^{t+1}\big)\big\{m\big(s^{t+1}\big)k\big(s^{t+1}\big)\big\} \tag{13}$$

subject to the budget constraint and law of motion of capital, as described in equations (4) and (6). The recursive problem defined above is equivalent to the sequential problem, as defined in Section 2.2, provided that the transversality condition is satisfied:

$$\lim_{t\to\infty} \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \pi\big(s^t\big) m\big(s^{t+1}\big) k\big(s^{t+1}\big) = 0. \tag{14}$$

This is shown in the following proposition, which is an extension of the main result in Section 3 of Phelan and Stacchetti (2001).

Proposition 1. *Let preference and production functions take the functional forms* $u(c, l) = \frac{c^{1-\vartheta}-1}{1-\vartheta} + \nu(l)$ *and* $F(k, l) = Ak^{\alpha}l^{1-\alpha}$, *where* $\nu: [0, 1) \to \mathbf{R}$ *is a decreasing, concave function, with* $\lim_{l\to 1} \nu(l) = -\infty$. *Then the recursive and the sequential problems for the household are equivalent.*

See the Appendix for this and other proofs.

Now I can summarize the competitive equilibrium conditions for a given tax policy τ as follows.

Definition 5. Let $Y^S\{k, s, (\tau_k, \tau_l), \{m'\}\}$ be the static (one-period) economy in which each household has an initial capital stock k and the current shock s, the expected marginal value of investment is given by the vector $\{m'\}$, and the government imposes tax (τ_k, τ_l) for the current period. Then $\{c, l, k', r, w, g\}$ is a *competitive equilibrium* for $Y^S\{k, s, (\tau_k, \tau_l), \{m'\}\}$ and is denoted by $\{c, l, k', r, w, g\} \in \mathbf{CE}^S\{k, s, (\tau_k, \tau_l), \{m'\}\}$ if and only if the following conditions are satisfied.

1. Households' choices satisfy the temporal equilibrium conditions $u_c(s) = \beta\mathbb{E}\{m'\}$ and $u_l(s) = (1 - \tau_l)w(s)u_c(s)$.
2. Input prices equate their marginal products.
3. The government satisfies its budget constraint.
4. Good markets clear at each node.

The following lemma allows us to think of the original economy as a sequence of static economies with endogenously changing state variables and exogenous stochastic shocks.

Lemma 1. *Given some tax policy* $\tau = \{(\tau_{k,t}, \tau_{l,t})\}_{t=0}^{\infty}$, *an initial capital stock* k_0, *and shock* s_0, *suppose that the sequence* $\{c(s^t), l(s^t), k(s^{t+1}), r(s^t), w(s^t), g(s^t)\}_{t=0}^{\infty}$ *is such that for each* t,

$$\big\{c\big(s^t\big), l\big(s^t\big), k\big(s^{t+1}\big), r\big(s^t\big), w\big(s^t\big), g\big(s^t\big)\big\} \in \mathbf{CE}^S\big\{k\big(s^t\big), s_t, (\tau_{k,t}, \tau_{l,t}), \big\{m\big(s^{t+1}\big)\big\}\big\},$$

where

$$m(s^{t+1}) := u_c(s^{t+1})\big[1 + (1 - \tau_{k,t+1})\big(r(s^{t+1}) - \delta\big)\big].$$

Then $\{c(s^t), l(s^t), k(s^{t+1}), r(s^t), w(s^t), g(s^t)\}_{t=0}^{\infty}$ *is a competitive equilibrium for* $\Upsilon\{k_0, s_0, \tau\}$.

The lemma says that the promised marginal value of investment in capital will summarize the expectations of the household. Therefore, it is sufficient to keep track of the value of this variable. It follows that equilibria of the dynamic game for the above economy can be characterized by the set of equilibrium values $\mathbf{V}(k, s)$. More precisely, the value correspondence $\mathbf{V}$ is defined as a mapping from values of the state (k, s) into set of possible payoffs associated with a strategy profile σ that constitutes an SSE.

Definition 6. The equilibrium value correspondence is defined as

$$\mathbf{V}(k, s) := \big\{(m, h) | \sigma \text{ is a SSE for } \Gamma(k, s)\big\},$$

where m denotes the marginal value of investment in capital defined by (12) and h represents the equilibrium continuation payoff of the government Φ_G defined by (9).

4.2 *Self-generation*

In line with Phelan and Stacchetti (2001), I define an operator $\mathbb{B}$ whose fixed point is the set of equilibrium values $\mathbf{V}$. Let $\mathbf{W}(k, s): \mathbf{R}_+ \times \mathbf{S} \to \mathbf{R}^2$ denote the set of equilibrium values and let $\mathbf{A}$ represent the space of all such sets. The operator $\mathbb{B}$ maps the space $\mathbf{A}$ into itself. To facilitate the definition of $\mathbb{B}$, I explain consistency and admissibility with respect to $\mathbf{W}$ as follows.

Definition 7. A vector $\psi = (\tau_k, \tau_l, c, l, k', \{(m', h')\})$ is consistent with respect to the set $\mathbf{W}$ at (k, s) if

$$\big(c, l, k', F_k(k, l), F_l(k, l), g\big) \in \mathbf{CE}^S\big\{k, s, (\tau_k, \tau_l), \{m'\}\big\}$$

for $(m(k, s, \psi), h(k, s, \psi)) \in \mathbf{W}(k, s)$ and $(m', h') \in \mathbf{W}(k', s')$, where the values of m and h are given by

$$m(k, s, \psi) := u_c(c, 1 - l)\big[1 + (1 - \tau_k)(r - \delta)\big], \tag{15}$$

$$h(k, s, \psi) := u(c, 1 - l) + G(g) + \beta\mathbb{E}\big\{h'\big\}. \tag{16}$$

In words, consistency guarantees that the vector ψ delivers an allocation that is optimal for the household and satisfies feasibility. It also requires that the promised continuation values (m', h') belong to the same set of equilibrium values as those of (m, h).

Definition 8. The vector ψ is admissible with respect to $\mathbf{W}$ if it is consistent with respect to $\mathbf{W}$ at (k, s) and

$$h(k, s, \psi) \geq h(k, s, \hat{\psi}). \tag{17}$$

This condition says that government cannot increase its payoff by announcing any unexpected tax rate $\hat{\tau}$ other than τ. Therefore, admissibility guarantees that the government has no incentive to deviate.

Definition 9. For a given set of equilibrium values $\mathbf{W}$, operator $\mathbb{B}$ is defined as

$$\mathbb{B}(\mathbf{W})(k,s)=\{(m,h)|\exists\psi \text{ is admissible with respect to } \mathbf{W} \text{ at } (k,s)\}.$$

In the following, I adapt the results of Abreu, Pearce, and Stacchetti (1990) to the game in this paper. Note that these results are straightforward extensions of those in Phelan and Stacchetti (2001) for the environment with stochastic TFP shocks. Therefore, I only list the main results and refer readers to their paper for proofs.

1. If $\mathbf{W}\subseteq\mathbb{B}(\mathbf{W})$, then $\mathbb{B}(\mathbf{W})\subseteq\mathbf{V}$.
2. $\mathbf{V}$ is compact and the largest set of equilibrium values $\mathbf{W}$ such that $\mathbf{W}=\mathbb{B}(\mathbf{W})$.
3. $\mathbb{B}(\cdot)$ is monotone and preserves compactness.
4. If we define $\mathbf{W}_{n+1}=\mathbb{B}(\mathbf{W}_n)$ for all $n\geq 0$ and define the equilibrium value correspondence $\mathbf{V}\subset\mathbf{W}_0$, then $\lim_{n\to\infty}\mathbf{W}_n=\mathbf{V}$.

Result 1 has been called self-generation. From the definition of the set of equilibrium values for SSE, it is straightforward to see that $\mathbf{V}\subseteq\mathbb{B}(\mathbf{V})$. Together with result 1, it is fairly easy to reach the second result. Results 3 and 4 will be used in the next section to approximate the set of equilibrium values.

5. Computation and simulation

In this section, I first detail the numerical algorithm that is used to compute the equilibrium value correspondence. I then explain how to build strategies that support a given pair of payoffs belonging to this correspondence.

5.1 *Recursive operator*

Computing the mapping $\mathbb{B}$ amounts to finding a set $\mathbb{B}(\mathbf{W})$, the set of pairs of (m,h) that can be "enforced" today,

$$\begin{aligned}\mathbb{B}(\mathbf{W})(k,s)=\{(m,h)|\exists(\tau_k,\tau_l),\ (c,l,k',g,w,r) \text{ and } (m',h')\in\mathbf{W}(k',s'),\\ \text{for all } s'\succ s\}\end{aligned}$$

such that

$$m=u_c(c,l)\cdot\big[1+(1-\tau_k)(r-\delta)\big], \tag{18}$$

$$h=u(c,l)+G(g)+\beta\mathbb{E}\{h'\}, \tag{19}$$

$$(m,h)\in\mathbf{W}(k,s), \tag{20}$$

$$h\geq\big[u(\hat{c},\hat{l})+G(g)+\beta\mathbb{E}\{\hat{h}'\}|(\hat{m}',\hat{h}')\big]\quad\forall(\hat{m}',\hat{h}')\in\mathbf{W}(k',s'), \tag{21}$$

$$u_c(c,l) - \beta\mathbb{E}\{m'\} = 0, \tag{22}$$

$$(1-\tau_l)wu_c(c,l) - u_l(c,l) = 0, \tag{23}$$

$$\tau_k, \tau_l \in [\underline{\tau}, \overline{\tau}], \tag{24}$$

where $s' \succ s$ denotes all possible shocks that follow s. Constraints (18), (19), and (20) are called *regeneration constraints*, while (21) is an *incentive constraint*. Condition (21) states that the strategy corresponding to (m,h) generates a higher payoff to the government than any other possible ones. Equations (22) and (23) are necessary to ensure that continuation of a sustainable plan after any deviation is consistent with a competitive equilibrium.

As Chang (1998) points out, computing $\mathbb{B}(\mathbf{W})$ given $\mathbf{W}$ is complicated because of constraint (21). Chang (1998) suggests an alternative operator to circumvent this complication in the context of finding time-consistent monetary policy, while Phelan and Stacchetti (2001) develop a similar operator for a production economy like the one in the current paper.

The basic idea of this alternative operator is the following. The government needs to consider only the payoff associated with the "best" deviation, but not to evaluate the consequences of all possible actions. Also, as in APS, if the government chooses to deviate, this is then followed by the worst available punishment. Thus, I redefine the operator $\mathbb{B}:\mathbf{A}\to\mathbf{A}$ by replacing (21) with the condition

$$h \geq \tilde{h}(k,s), \tag{25}$$

where $\tilde{h}(k,s)$ is the worst possible payoff for the government when it announces unexpected tax rate $\hat{\tau}$. As in Phelan and Stacchetti (2001), I consider only extreme punishments. Therefore, I define $\tilde{h}(k,s)$ as

$$\tilde{h}(k,s) = \max_{\tau_k,\tau_l}\left\{\min_{c,l,k',(m',h')\in\mathbf{W}(k',s')}\left[u(c,l) + G(g) + \beta\mathbb{E}\{h'\}\right]\right\}$$

such that

$$[c,l,k',g,w,r] \in \mathbf{CE}^S\{(k,s,\tau_k,\tau_l,\{m'\}) \text{ for all } s' \succ s\}.$$

Chang (1998) shows that conditions (21) and (25) are equivalent in the sense that repeated application of operator $\mathbb{B}$, with either (21) or (25), yields a decreasing sequence of sets that converge to the same fixed point $\mathbf{V}$. Phelan and Stacchetti (2001) use condition (25). Note that if the value of $\tilde{h}(k,s)$ can be determined up front (i.e. Chari and Kehoe (1990)) or can be explicitly imposed (i.e. Benhabib and Rusticchini (1997) and Marcet and Marimon (2011)), the techniques of optimal control may be used to compute equilibria. However, for my application, the value of $\tilde{h}(k,s)$ depends on future decision variables and government policies, which need to be determined as part of the equilibria.

5.2 *Numerical implementation*

Even though I can simplify the computation by substituting condition (25) for condition (21), computation is not a trivial matter, as I need to operate over correspondences. Neither Chang (1998) nor Phelan and Stacchetti (2001) is explicit about the numerical implementation of their algorithms. To the best of my knowledge, Domínguez (2010) is the first paper that applies Phelan and Stacchetti's method and provides a numerical implementation by adapting the approximation technique developed by Judd, Yeltekin, and Conklin (2003) (JYC henceforth).[9] JYC's idea is to construct a convex set containing the set of equilibrium values and another convex set contained in that set. The JYC method requires the convexity of the set of equilibrium values. Phelan and Stacchetti (2001) introduce a publicly observed random variable to convexify the equilibrium value correspondence. This random variable has been interpreted as a device to synchronize the behaviors and beliefs of the government and the household. There is an obvious advantage of the JYC method: It uses a polar coordinate system to represent the position of an arbitrary point on a manifold. Thus, to approximate a convex set, one only needs to keep track of the supporting hyperplanes at each angle around the polar. This approximation scheme can reduce the computational costs substantially. However, randomization and convexification will arbitrarily enlarge the equilibrium value correspondence.

Here, I follow the approach of Feng, Miao, Peralta-Alva, and Santos (2014) (FMPS henceforth). These authors develop a numerical method to approximate upper hemicontinuous equilibrium value correspondences. Their method partitions the state space into a finite set of simplices. Compatible with this partitioning, they then consider a sequence of step correspondences, which take constant set values on each simplex. The main advantage of the FMPS method is that it does not require convexity of the equilibrium set and, thus, it is not necessary to introduce a randomization device as in Phelan and Stacchetti (2001). However, this method faces some computational challenges. It turns out that I cannot find sufficient computational resources to compute the model by directly implementing FMPS's approximation scheme.[10] To facilitate the computation, I impose two assumptions over the equilibrium value correspondence. These assumptions help to reduce the computational cost substantially and to make FMPS's approximation algorithm applicable.

Assumption 1. **W** *is convex-valued at given* $\{k, s, m\}$.

I emphasize that this assumption is weaker than the one in Phelan and Stacchetti (2001). The randomization device introduced in their paper guarantees convexity of the equilibrium value correspondence. A convex correspondence must be convex-valued, while a convex-valued correspondence does not necessarily have a convex graph. What this assumption says is that for a given (k, m), there exists a strategy that supports any

[9] Fernandez-Villaverde and Tsyvinski (2002) also discuss the same approximation technique.

[10] The current paper uses a modified approximation method based on FMPS. It has been coded in C++ with message passing interface (MPI) and takes over 20 hours with 100 central processing units (CPUs). The computational cost of directly implementing the FMPS algorithm would be over 100 times greater than the one presented here, as it needs to discretize h.

$h \in [\underline{h}, \overline{h}]$. Note that h is the payoff of the government. Given the continuity of the utility function, there is no reason that h is not continuous inside $[\underline{h}, \overline{h}]$. I cannot formally prove that the economy satisfies this assumption. However, I can verify that all computed sets of equilibrium values satisfy this assumption.

Assumption 2. *There exists, at most, one vector (c, l, k') that solves the following equation system at given (k, s) and $\{\tau_k, \tau_l, m\}$:*

$$m = u_c(c, l)\big[1 + (1 - \tau_k)(r - \delta)\big], \tag{26}$$

$$u_l(c, l) + (1 - \tau_l)wu_c(c, l) = 0, \tag{27}$$

$$\big[1 + (1 - \tau_k)(r - \delta)\big]k + (1 - \tau_l)wl - c - k' = 0. \tag{28}$$

This assumption may rule out some utility functions, especially nonseparable preferences. The following pair of utility and production functions satisfies this assumption:[11]

$$u(c, l) = \frac{c^{1-\vartheta_1} - 1}{1 - \vartheta_1} + \gamma_l \frac{(1 - l)^{1-\vartheta_2}}{1 - \vartheta_2}, \tag{29}$$

$$F(k, l) = Ak^{\alpha} l^{1-\alpha}. \tag{30}$$

It seems to me that no algorithm can get around this problem. This is because the consistency check in operator $\mathbb{B}$ requires the value of next period's capital stock to select the continuation values. Nevertheless, this assumption becomes irrelevant if there exists an algorithm that finds all real solutions to the equation system (26)–(28) (i.e. Kubler and Schmedders (2010)).

Under the above assumptions, I compute the upper and lower boundaries of $\mathbf{W}(k, s)$. In line with Phelan and Stacchetti (2001), these boundaries are represented by the functions

$$\overline{h}(k, s, m) := \sup_h \big\{h | (m, h) \in \mathbf{W}(k, s)\big\}, \tag{31}$$

$$\underline{h}(k, s, m) := \inf_h \big\{h | (m, h) \in \mathbf{W}(k, s)\big\}. \tag{32}$$

As Phelan and Stacchetti (2001) observe, the lowest value in $\mathbf{W}(k, s)$ yields the value of the worst punishment for the government $\tilde{h}(k, s) = \min_m \underline{h}(k, s, m)$, which corresponds to the equilibrium in which government decides to deviate, that is, WSSE. Note that the highest value $\max_m \overline{h}(k, s, m)$ in $\mathbf{W}(k, s)$ corresponds to equilibrium in which the government obtains the maximum payoff, that is, BSSE. This paper focuses on the boundaries of the equilibrium value correspondence, since one can learn a lot about the equilibria from looking at extremes. Understanding strategies in BSSE and WSSE conveys

[11] For the preference and production function used in this paper, the above equation system will yield the following equation in terms of l, $Ak^{\alpha} = B(1 - \delta)l^{\alpha} + [Ak^{\alpha} + B(1 - \tau_k)\alpha k^{\alpha-1}]l$, where $A = m(1 - \tau_l)(1 - \alpha)$ and $B = (1 - \beta)(1 - \alpha_p)(1 - \alpha_c)$. When $\{k, s, m, \tau_k, \tau_l\}$ are given, the right-hand side is monotone increasing in $l \in [0, 1]$. Therefore, there exists, at most, one solution $l^* \in [0, 1]$.

important information about all other strategies. To this end, I define an outer approximation of $\mathbf{W}$: $\hat{\mathbf{W}}(k, s) = \{(m, h) | h \in [\underline{h}(k, s, m), \overline{h}(k, s, m)]\}$.

I would like to remark that Assumptions 1 and 2 only simplify the computation, but not the dynamic problem per se.[12] The first assumption is used to reduce the dimension of the above problem. With this assumption, the algorithm tests only $(m, \underline{h}) \in \mathbf{W}(k)$, $(m, \overline{h}) \in \mathbf{W}(k)$, $(m', \underline{h}') \in \mathbf{W}(k')$, and $(m', \overline{h}') \in \mathbf{W}(k')$. Accordingly, the number of admissibility checks at each iteration reduces from 4×10^{24} to 1.6×10^{16} for the discretization used by this paper. The second assumption reduces the computation cost in solving for (26)–(28). Note that neither assumption changes the structure of the model economy. With these assumptions, I have the following proposition.

Proposition 2. *If the set of equilibrium values* $\mathbf{V}(k, s)$ *satisfies Assumptions* 1 *and* 2, *then for all* $(m, h) \in \mathbf{V}(k, s)$,

$$\overline{h}(k, s, m) = \max_{\tau_k, \tau_l} \{u(c, l) + G(g) + \beta \mathbb{E} \overline{h}(k', s', m')\}, \tag{33}$$

$$\underline{h}(k, s, m) = \max_{\tau_k, \tau_l} \{u(c, l) + G(g) + \beta \mathbb{E} \underline{h}(k', s', m')\}, \tag{34}$$

$$\tilde{h}(k, s) = \min_{m} \underline{h}(k, s, m) \tag{35}$$

subject to the constraint $[c, l, k', g, w, r] \in \mathbf{CE}^S\{(k, s, \tau_k, \tau_l, \{m'\})$ *for all* $s' \succ s\}$. *Here,* $\overline{h}(k, s, m)$ *and* $\underline{h}(k, s, m)$ *are* sup *and* inf *of the set* $\mathbf{V}(k, s)$ *for a given* m *as defined by* (31) *and* (32).

Domínguez and Feng (2012) derive a similar proposition. Note that the proposition here applies to the equilibrium value correspondence $\mathbf{V}(k, s)$, while the proposition in that paper holds for any $\mathbf{W} \supseteq \mathbf{V}$. These results suggest that one can focus on boundaries of the set of equilibrium values if the primary interest is in BSSE or WSSE. This will

[12] To better understand this point, let us think about how one would compute the model without these assumptions. For better exposition, I focus on the deterministic case. The algorithm will work as follows.

- For any k, pick any $(m, h) \in \mathbf{W}(k)$; for all possible pairs of $(\tau_k, \tau_l) \in [\underline{\tau}, \overline{\tau}] \times [\underline{\tau}, \overline{\tau}]$, find all (c, l, k') as real solutions to the equation system (26)–(28).
- Now, check whether there exists $(m', h') \in \mathbf{W}(k')$ such that the vector $(\tau_k, \tau_l, c, l, k', (m', h'))$ is admissible with respect to $\mathbf{W}$.
- Then keep (m, h) in $\mathbf{W}(k)$ if there exists $(\tau_k, \tau_l) \in [\underline{\tau}, \overline{\tau}] \times [\underline{\tau}, \overline{\tau}]$ and $(m', h') \in \mathbf{W}(k')$ that satisfy the conditions of the second step. Otherwise, drop this pair from $\mathbf{W}(k)$ and move to another pair $(\hat{m}, \hat{h})$ that belongs to the set $\mathbf{W}$.

Besides the complication of approximating the set W, which has been addressed by following the strategy of Feng et al. (2014), there are two additional challenges. First, the problem has seven state variables, namely $(k, m, h, \tau_k, \tau_l, (m', h'))$. Second, one has to solve for all real solutions for the equation system (26)–(28). In principle, I can design an algorithm to compute the model based on the above procedure. However, it turns out to be infeasible to find sufficient computing resources due to the curse of dimensionality. For the application of the current paper, the above algorithm needs to run the admissibility check over 4×10^{24} different combinations of $(k, m, h, \tau_k, \tau_l, (m', h'))$ for each iteration. Note that I set 400 grid points for k, 1000 for m and h, and 100 for τ_k and τ_l. The aggregate shocks take two values.

greatly reduce the computational cost. Based on Proposition 2, I define operator $\mathbb{F}$ as follows.

Definition 10. For any convex-valued correspondence $\hat{\mathbf{W}} = \{(m,h) | h \in [\underline{h}^0(k,s,m), \overline{h}^0(k,s,m)]\}$, define operator $\mathbb{F}$ as

$$\mathbb{F}(\hat{\mathbf{W}})(k,s) = \{(m,h) | h \in [\underline{h}^1(k,s,m), \overline{h}^1(k,s,m)]\},$$

where

$$\overline{h}^1(k,s,m) = \max_{\tau_k,\tau_l} u(c,l) + G(g) + \beta\mathbb{E}\overline{h}^0(k',s',m'), \tag{36}$$

$$\underline{h}^1(k,s,m) = \max\Big\{\max_{\tau_k,\tau_l} u(c,l) + G(g) + \beta\mathbb{E}\underline{h}^0(k',s',m'), \tilde{h}^0(k,s)\Big\}, \tag{37}$$

$$\tilde{h}^0(k,s) = \max_{\tau_k,\tau_l}\Big\{\min_{c,l,k',\{m'\}} u(c,l) + G(g) + \beta\mathbb{E}\underline{h}^0(k',s',m')\Big\} \tag{38}$$

such that vector $(\tau_k, \tau_l, c, l, k', g, w, r, \{m', h'\})$ is admissible with respect to $\hat{\mathbf{W}}$ at (k,s). Define $\underline{h}(k,s,m) = -\infty$ and $\overline{h}(k,s,m) = +\infty$ if no such vector exists.

Equation (37) is slightly different from (34). The outer maximum is used to address the case in which the incentive constraint is always violated at given (k,s). This may happen when the initial value of $\underline{h}^0$ is sufficiently low. For convenience of presentation, the numerical implementation of operator $\mathbb{F}$ is discussed in the Appendix.

The following result shows that this operator has good convergence properties. Repeated application of this operator generates a sequence of sets that converge to the equilibrium value correspondence $\mathbf{V}$.

Theorem 1. *Assume that the equilibrium value correspondence $\mathbf{V}$ satisfies Assumptions 1 and 2. Let $\hat{\mathbf{W}}_0$ be a convex-valued correspondence such that $\hat{\mathbf{W}}_0 \supset \mathbf{V}$. Let $\hat{\mathbf{W}}_n = \mathbb{F}(\hat{\mathbf{W}}_{n-1})$. Then* $\lim_{n\to\infty} \hat{\mathbf{W}}_n = \mathbf{V}$.

5.3 *Recovering strategies*

In this section, I outline how to find the strategy that corresponds to a given pair of payoffs in the equilibrium value correspondence. For better exposition, I abstract from uncertainty. I also focus on the strategy that yields the highest payoff for the government, which corresponds to BSSE. This procedure can be generalized to find strategies supporting any point belonging to the correspondence.

Step 1. $t = 0$, k_0 is given. Find the highest possible value of $h_0 = \sup\{h | (m_0, h_0) \in \mathbf{W}^*(k_0)\}$ and its corresponding m_0. Then search for the government's tax policy that can support (m_0, h_0). More specifically, choose $(\tau_{k,0}, \tau_{l,0})$ such that

$$u(c_0, l_0) + G(g_0) + \beta h_1 = h_0, \tag{39}$$

where $h_1 = \overline{h}(k_1, m_1)$ and $m_1 = \frac{u_c(c_0, l_0)}{\beta}$, $(m_1, h_1) \in \mathbf{W}^*(k_1)$. Values of (c_0, l_0, k_1) come from the solution for the equation system at given $(\tau_{k,0}, \tau_{l,0}, m_0)$:

$$m_0 - u_c(c_0, l_0) \cdot \big[1 + (1 - \tau_{k,0})(r_0 - \delta)\big] = 0, \tag{40}$$

$$u_l(c_0, l_0) - (1 - \tau_{l,0})w_0 \cdot u_c(c_0, l_0) = 0, \tag{41}$$

$$(1 - \tau_{l,0})w_0 l_0 + \big[1 + (1 - \tau_{k,0})(r_0 - \delta)\big]k_0 - (c_0 + k_1) = 0. \tag{42}$$

Therefore, the above problem is well defined in terms of $(\tau_{k,0}, \tau_{l,0}, m_0, h_0)$.

Step 2. $t = 1, k_1, m_1, h_1$ are given by the solution in Step 1. Now search for government policies $(\tau_{k,1}, \tau_{l,1})$ such that

$$u(c_1, l_1) + G(g_1) + \beta h_2 = h \tag{43}$$

as in Step 1.

Step 3. Repeat Step 2 for $t = 2, \dots, T$ for T be sufficiently large.

Note that the construction above reveals that any sustainable outcome has essentially a Markovian structure. More specifically, $k(s^t)$, $\tau_k(s^t)$, $\tau_l(s^t)$ and $m(s^t)$, $h(s^t)$ depend on history ζ^{t-1} through $m(s^{t-1})$, $h(s^{t-1})$ (i.e. Chang (1998)).

6. Calibration

The calibration follows Klein and Ríos-Rull (2003). The utility function is of the constant relative risk aversion (CRRA) form and the technology is Cobb–Douglas:

$$u(c, l) = (1 - \beta)(1 - \alpha_p)\frac{[c^{\alpha_c}(1 - l)^{1-\alpha_c}]^{1-\vartheta}}{1 - \vartheta}, \tag{44}$$

$$F(K, L) = A(s)K^{\alpha}L^{1-\alpha}. \tag{45}$$

In the baseline economy, I set $\vartheta = 1$.[13] The preference of the government for public consumption is given by

$$G(g) = (1 - \beta)\alpha_p \ln g. \tag{46}$$

I model the technology shock as a symmetric two-state Markov chain with mean 1. The parameters for process A are calibrated so as to match the variance and autocorrelation of TFP of the post-WWII U.S. economy, as reported in Prescott (1986). The rest of the parameter values are chosen such that certain moments in the stationary distribution of the Ramsey allocation are consistent with the U.S. data. More precisely, the value of α is picked to match the labor income share of national product; the value of β is picked to generate a pre-tax interest rate of around 4%.[14] The rate of capital depreciation δ is chosen to match a capital-output ratio of around 3, α_p is chosen to get a

[13] This yields a log utility: $u(c, l) = (1 - \beta)(1 - \alpha_p)[\alpha_c \log c + (1 - \alpha_c)\log(1 - l)]$.

[14] See McGrattan and Prescott (2000) for a justification of this number based on their measure of return on capital and on the risk-free rate of inflation-protected U.S. Treasury bonds.

TABLE 1. Parameter values for the baseline economy.

Parameter Values					
β	ϑ	δ	α	α_p	α_c
0.96	1	0.08	0.36	0.13	0.30
	$A \in \{0.976, 1.024\}$			$\pi^{11} = \pi^{22} = 0.946$	

share of government consumption around 20%, and α_c is chosen to get hours worked to be around 25% of total time. Finally, I set an upper limit and a lower bound on tax rates that equate to the highest and lowest tax rates observed in a simulated Ramsey economy to facilitate the comparison with the Ramsey equilibrium policy. This yields a bound for the tax rate $[0.0, 0.9]$. Phelan and Stacchetti (2001) interpret this bound as an institutional constraint. The calibrated parameter values are summarized in Table 1. Note that π^{ii} denotes the probability of remaining in the current state, where $i = 1, 2$.

7. Quantitative findings

In this section, I first present results for the economy in the steady state starting with a description of the equilibrium value correspondences, BSSE and WSSE. I then discuss the sustainability of the Ramsey equilibrium. I also compare steady-state outcomes of the economy in different equilibria. Finally, I discuss business-cycle properties of optimal taxation. I start by describing the economy with full commitment. This is followed by a quantitative characterization of optimal fiscal policy when there is no commitment device available in the economy.

7.1 *Steady-state dynamics*

7.1.1 *BSSE, WSSE, and the relationship with MPE* I approximate the equilibrium value correspondence using the algorithm described in Section 5.2. I discretize the state space with 400 equally spaced points for k and 1000 points for m. I use linear interpolation for variable values falling outside of the grid. I ran my C++ MPI code using an IBM server 1350 cluster with 50 Xeon 2.3 GHz processors. Figure 2 shows the approximated equilibrium value correspondence for the baseline economy. Given this computed correspondence, I construct strategies that support BSSE and WSSE based on the procedure explained in Section 5.3. Statistics for these equilibria are listed in Table 2.[15]

To better understand these results, I examine the behavior of the government in equilibrium. In the presence of commitment (labeled as Ramsey), the tax rate on capital income converges to zero, which is reminiscent of results of Chamley (1986) and Judd (1985). The Ramsey government sets the capital income tax rate to be zero in the deterministic steady state because a positive capital income tax creates permanent distortion against future consumption.

[15]I normalize the welfare of the Ramsey equilibrium in steady state as 100.

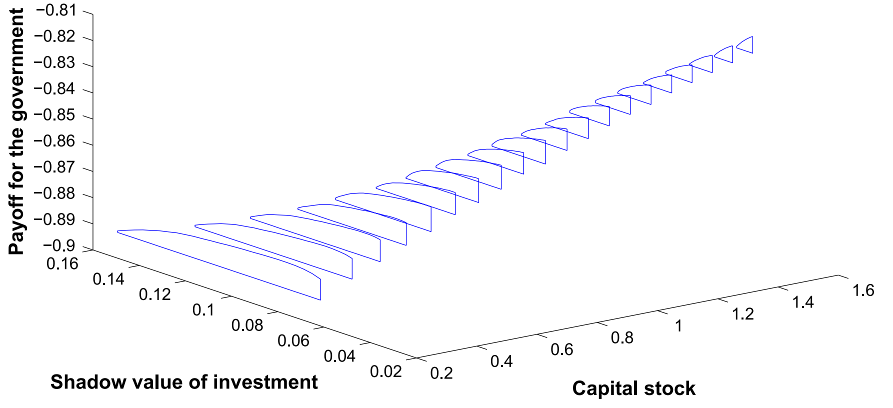

FIGURE 2. The equilibrium value correspondence for the deterministic economy.

TABLE 2. Best and worst SSE, compared with MPE and Ramsey.

Government	k/y	c/y	g/y	c/g	l	τ_k	τ_l	Welfare
Ramsey	2.959	0.509	0.254	2.005	0.245	0.000	0.397	100
BSSE	2.896	0.513	0.255	2.009	0.247	0.058	0.387	96.3
MPE	1.150	0.688	0.220	3.125	0.285	0.612	0.000	81.4
WSSE	1.003	0.681	0.215	3.188	0.286	0.789	0.000	76.8

In the absence of commitment, the government may prefer a positive capital income tax, even in the long run. This is because the supply of capital is inelastic with respect to the after-tax rate of return in the short run and the government sets tax rates for the current period only. Compared to the zero capital income tax prescribed by the initial government, a positive capital income tax helps reduce the tax rate on labor income without any impact on the current capital stock and investment. Hence, it removes some of the distortion of the current labor supply and improves the current return to the government. To this end, a prespecified zero capital income tax is not optimal for the current government or, in other words, it is not time-consistent.

In SSE, there is a reputation mechanism that helps define time-consistent policies. For any equilibrium policy, the current government may find that a deviation to a higher capital income tax will be advantageous for reasons explained above. To sustain a particular equilibrium, a reputation mechanism has been introduced to deter the government with a lower continuation value. This reputation mechanism is essentially a punishment strategy, which works as follows. If the government deviates from the equilibrium path, it will affect the representative household's expectations about future policies. To give the government enough incentives to stick to the equilibrium path, the household's best response, conditional on this new expectation, generates the lowest possible continuation value. Of course, this response satisfies feasibility and optimality for both the government and the household. In equilibrium, the government faces a trade-off between a higher current return and a lower continuation value in case of deviation. Note that

WSSE corresponds to the punishment strategy and BSSE is the best equilibrium that can be sustained by reputation.

For the baseline calibration, the steady-state capital income tax is around 6% and the labor income tax is close to 39% for BSSE. A positive capital income tax discourages capital accumulation. Consequently, the steady-state capital stock in BSSE is over 2% lower than in the Ramsey equilibrium. The detrimental effect of a positive capital income tax is more pronounced in WSSE. With a 79% capital income tax, WSSE has a steady-state capital stock less than 30% of that in the Ramsey equilibrium.

Government policy in MPE is also time-consistent. In such an equilibrium, time consistency is guaranteed by the assumption that the government sets its policy solely based on the current capital stock (and the value of the shock in the stochastic setting). In MPE, the current government considers future tax policies as given functions of future capital stocks. Hence a higher capital income tax will not create any intertemporal distortion from the perspective of the current government. Consequently, the government raises the tax rate on capital income until the labor income tax hits the lower bound.[16]

Martin (2010) suggests that MPE will coincide with WSSE if no interior solution MPE exists. Numerical exercises in this paper show that MPE generates allocations somewhere between those of BSSE and WSSE. While SSE does not impose any differentiability on strategies, most research characterizes differentiable MPE. Not requiring differentiability is less restrictive, which may yield lower payoffs to the government in WSSE than in MPE. A definitive answer to what we give up by imposing differentiability requires an algorithm to select MPE from the set of SSE. The construction of such an algorithm relies on an existence proof of MPE, something that is missing in the literature to the best of my knowledge.

7.1.2 *Is the Ramsey equilibrium sustainable?* Equilibrium policies in SSE are sustained by reputation, in essence WSSE. An interesting question is what it takes to sustain the Ramsey equilibrium. Needless to say, if the punishment to the government's deviation goes to negative infinity, Ramsey is sustainable. This is because such punishment will dwarf increased current return from any deviation. However, the value of WSSE is endogenously determined as it is an SSE and it typically does not go to minus infinity. In a model without capital accumulation, the policy game is essentially a repeated one and WSSE can be explicitly identified. It has been shown that sufficiently low discounting will sustain the Ramsey allocation (i.e. Chari and Kehoe (1990)). When there is capital accumulation, it is not straightforward to compute WSSE. On the one hand, a higher value of the discount factor makes the deviation more painful, as the future carries greater weight in determining the total payoff (current return plus discounted continuation value) to the government. At the same time, this higher value may raise the payoff of deviation, as the representative household has a stronger saving incentive in any equilibria. Thus, one must rely on numerical simulations to determine how the value of the discount factor changes the properties of BSSE. Numerical simulations show that the steady-state tax rate on capital income increases as the value of β decreases; see Table 3.

[16]The results for MPE are borrowed from Martin (2010). Although one cannot rule out the possibility that there are multiple MPE, only one equilibrium has been found and reported in Martin (2010).

Table 3. What matters—discount factor?

β	τ_k	τ_l
0.96	4.5%	34.5%
0.86	32.2%	29.8%
0.70	34.5%	27.2%
0.00	92.4%	0.00%

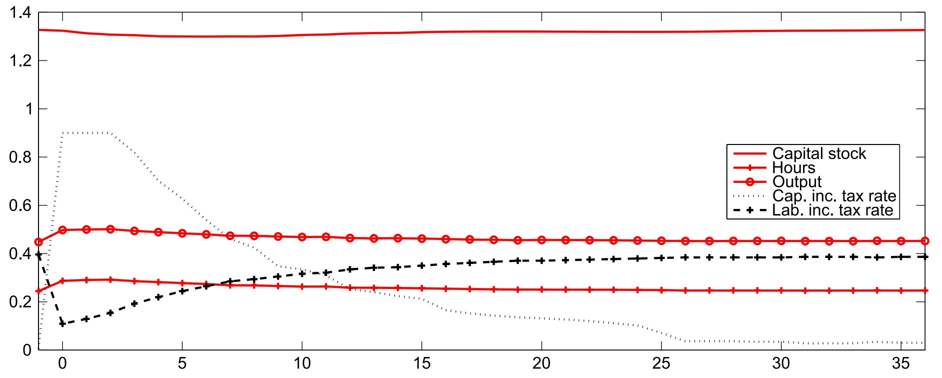

Figure 3. Simulated outcome path after the government loses commitment.

Domínguez and Feng (2012) consider a utility function that allows for variation in the labor supply elasticity: $u(c, l) = \log c - \gamma \frac{l^{1+\eta}}{1+\eta}$. They keep the value of β constant while adjusting the value of the labor supply elasticity η. Their numerical simulations suggest that the steady-state capital income tax increases with the labor supply elasticity. This is because part of the incentive for the government to raise the capital income tax comes from the benefit of reducing the labor supply distortion. If the labor supply is inelastic under the reduced tax rate, then it becomes less attractive for the government to deviate. I conclude from experiments here and those in Domínguez and Feng (2012) that the Ramsey policy with zero capital income tax may be sustainable when discounting is sufficiently low and labor supply is sufficiently inelastic.

7.1.3 *The implications of no commitment* To further demonstrate why commitment makes a difference, Figure 3 presents the dynamic adjustment of an economy in which the government switches from full commitment to no commitment. In this experiment, the economy starts from the steady state with commitment. The government loses commitment unexpectedly at period zero. A BSSE is then simulated. The figure shows the outcome path over a period of 40 years.

It is clear that the government chooses to impose a very high tax rate on capital income and to decrease the labor income tax when commitment is lost. This happens because a capital income tax is not distortionary to the current government. The capital stock starts to fall and hours worked increase. After the first few periods, the capital income tax becomes lower, and the labor income tax increases over time. These findings are consistent with Phelan and Stacchetti (2001).

7.2 *Business-cycle properties of optimal taxation*

In this section, I present the results for the stochastic economy. I start with a description of some cyclical properties of U.S. tax policy and then give a quantitative characterization of optimal fiscal policy with and without commitment, respectively. I then compare the outcomes of these two economies.

7.2.1 *Data* Table 4 presents some key facts about the U.S. tax rates over the business cycle for the period 1947–1990. Capital and labor income tax rate data are reported in Klein and Ríos-Rull (2003). Data for TFP and gross domestic product (GDP) come from the National Income and Product Account (NIPA) of the United States.

7.2.2 *Commitment* Table 5 presents the key findings of the economy with full commitment.[17] The main properties of this economy can be summarized as follows.

- The capital income tax rate is close to zero, on average.
- The labor income tax rate is high, in the sense that the lion's share of tax revenue comes from the labor income tax.
- The capital income tax rate is much more volatile than the labor income tax rate.
- The capital income tax rate is countercyclical, while the labor income tax rate is procyclical.

These results are in line with the findings for the baseline model in Chari, Christiano, and Kehoe (1994) (the case with technology shock only). The Ramsey government takes the distortion from capital income taxation into consideration. It eliminates such distortion by setting the long-run capital income tax rate at zero. In terms of business cycle properties, the optimal labor income tax is procyclical, while the optimal capital income tax is countercyclical. A procyclical labor income tax helps to smooth the representative household's after-tax wage income. The countercyclical capital income tax provides an efficient means of absorbing shocks to the government's budget, which varies with the

Table 4. Properties of the data.

Capital income tax	
Mean	0.51
Standard deviation	0.04
Autocorrelation	0.76
Correlation with output	0.17
Correlation with TFP	0.18
Labor income tax	
Mean	0.24
Standard deviation	0.03
Autocorrelation	0.95
Correlation with output	−0.08
Correlation with TFP	−0.00

[17] The model with full commitment has been solved using the primal approach, as in Stockman (2001). The author thanks David Stockman for sharing his codes and clarifying some technical questions.

Table 5. Business cycle properties of tax policies.

Government	Statistics	Mean	Std. Dev.	Autocorr.	Corr. w/ GDP	Corr. w/ TFP
Ramsey	Cap. tax	0.002	0.034	0.694	−0.665	−0.748
	Lab. tax	0.397	0.002	0.653	0.364	0.409
BSSE	Cap. tax	0.047	0.054	0.783	0.090	0.057
	Lab. tax	0.387	0.014	0.780	−0.043	−0.043
WSSE	Cap. tax	0.867	0.025	0.860	−0.140	−0.402
	Lab. tax	0.039	0.017	0.854	0.253	0.529
MPE_F	Cap. tax	0.739	0.019	0.726	−0.211	−0.386
	Lab. tax	0.000	0.000	1.000	1.000	1.000
MPE_{K-R}	Cap. tax	0.650	0.091	0.690	n/a	−0.800
	Lab. tax	0.120	0.023	0.620	n/a	0.540
Data	Cap. tax	0.510	0.040	0.760	0.170	0.180
	Lab. tax	0.240	0.030	0.950	−0.080	0.000

size of the tax base over the business cycle. Judd (1993) argues that the efficiency cost of adjusting capital income tax is low because the short-run supply elasticity of capital is very small. Hence, the Ramsey government uses capital income taxation as a shock absorber.

7.2.3 *No commitment* For the economy without commitment, I find the strategies that support the best equilibrium outcome, and include key statistics in Table 5. The main findings are listed as follows. In BSSE, I find the following elements.

- The capital income tax rate is 4.7% on average.
- The labor income tax rate is high compared with the capital income tax, at 38.7%.
- The capital income tax rate is more volatile than the labor income tax rate. Both the capital income and labor income tax rates are more volatile than in the economy with commitment.
- The capital income tax rate is slightly procyclical, while the labor income tax rate is countercyclical.

The average tax rates on capital and labor are not very different from what I found in the Ramsey equilibrium. Even though capital income is distortion-free in the short term, the government considers that a positive tax on capital will alter the representative household's expectation about future policy. Depending on the current tax rate, the household's expectations will lead to a level that yields a sufficiently low payoff for the government. This effect deters the government's incentive to exploit the inelastic aspect of capital supply. In BSSE, the government sets the average capital income tax rate at 4.7% and achieves the highest social welfare.

I also examine the aggregate behavior of the economy in WSSE, in which the equilibrium generates the lowest social welfare. It is important to look at this particular equilibrium since WSSE works as a credible threat that supports all other SSE. I summarize the findings as follows.

- The capital income tax rate is 86.7%, on average.
- The labor income tax rate is around 3.9%.
- A high capital income tax depletes the capital stock, while a low labor income tax encourages work effort.
- The capital income tax is countercyclical, while the labor income tax is procyclical.

Here, I highlight an interesting finding with regard to the cyclical behavior of taxation in the economy with no commitment. The Ramsey government uses a countercyclical capital income tax to absorb the shock to government spending, while the capital income tax in BSSE is merely procyclical. As Table 5 shows, the correlation between the capital income tax and the aggregate TFP shock is 0.748 in the Ramsey equilibrium, while the correlation is only 0.057 in BSSE. I also observe quite different cyclical behavior in the labor income tax between the Ramsey and the SSE. The correlation between the labor income tax and the TFP shock is 0.409 in the Ramsey equilibrium versus −0.043 in BSSE. When compared with the U.S. data reported in Table 4, it suggests that BSSE provides a better account of the cyclical behavior of fiscal policy.

7.2.4 *Expectation and cyclicality of taxation* The steady-state analysis shows that the government faces a reputational trade-off between a higher current return and a lower continuation value in a deterministic economy. When there is uncertainty in the economy, the capital income tax as a shock absorber will interact with this reputational effect. When a negative TFP shock hits the economy, the Ramsey government will follow the contingent plan specified by the initial government and raise the current capital income tax rate so that it can smooth the public goods consumption. When there is no commitment, the government can potentially implement the same smoothing plan. However, the reputational effect may make the countercyclical capital income tax smoothing option inapplicable. A higher capital income tax rate will help to absorb the shock to government spending, but it will also alter the representative household's expectation, as argued above. When the government decides to raise the tax rate on capital income, the household only observes the announced tax rates and compares them with his/her expectation based on the past. The household cannot differentiate the government's intention between smoothing spending and deviation. If the cost due to a lower continuation value outweighs the benefit in terms of absorbing the aggregate shocks, the government may not utilize countercyclical capital income taxation. The numerical exercises suggest that this is the case in BSSE for the baseline calibration.

Next consider the behavior of WSSE and MPE over the business cycle.[18] In WSSE, the government is indifferent between deviating or not. In such a case, it becomes advantageous for the government to smooth spending by using a countercyclical capital

[18]MPE is computed using the method of value function iteration; see Martin (2010) for computational details.

income tax. Similarly, the MPE government can do the same thing since there is no reputation at all. In MPE, the government is assumed to make policy choices based on the current values of some predetermined variables. To this end, both WSSE and MPE have countercyclical capital income tax and procyclical labor income tax. Only BSSE generates co-movement between taxation and GDP that is similar to the U.S. economy. This better fit should be attributed to the history dependence and reputation mechanism considered by SSE.

Klein and Ríos-Rull (2003) were the first to characterize the business-cycle properties of optimal taxation in MPE. My model is similar to theirs as I explained in the calibration section. However, the timing in their paper is different. In Klein and Ríos-Rull (2003), the government inherits a certain capital income tax rate from the previous government and sets a vector of state contingent capital income tax rates for the next period. To differentiate the two MPEs, I label the Klein and Ríos-Rull (2003) findings as MPE_{K-R}; see Table 5. MPE_F denotes the numerical results in the current paper. These timing differences result in different long-term capital income taxes. My finding in Table 5 is a reminiscent of Klein and Ríos-Rull (2003) (see their discussion in Section 6.8 on page 1242): the capital income tax rate is higher when it is less distortionary, as in the case where the government sets the current capital income tax.

7.3 *Welfare cost of no commitment*

I also examine the welfare cost of no commitment. To measure the welfare loss/gain of a particular tax policy, compared with Ramsey, I compute the consumption equivalent variation (CEV). More specifically, I quantify the welfare effect of a given policy by asking how much consumption must be increased in each state and at each date so as to equate expected utilities in the Ramsey allocation while leaving leisure and public expenditure unchanged. For the set of parameter values I use, the welfare loss of no commitment in BSSE is merely 0.22% in consumption equivalent.

The welfare gains of the full commitment come mainly from a zero tax rate on capital income. The gain from the large initial tax on capital income is nearly zero. In Chari, Christiano, and Kehoe (1994), the government can eliminate the distortion of the capital income tax by setting a large initial tax and a zero rate thereafter. This channel only works through the issuance of government debt. For computational reasons, I assume that the government cannot issue debt. Consequently, the Ramsey government can utilize only the long-run zero tax rate on capital. Stockman (2001) investigates the effects of a balanced-budget restriction on the Ramsey allocation. There is a big welfare gain associated with switching from the Ramsey with a fixed debt, which is similar to the Ramsey government in the current study, to the Ramsey with no constraint on debt issuance. But that does not necessarily mean that there will be a bigger welfare gain when government debt is introduced. This is because the time-inconsistency problem may disappear, as the government debt can serve as the central commitment device among governments (see Domínguez (2007)).

8. Concluding remarks

This paper makes three contributions. First, it characterizes optimal time-consistent and history-dependent fiscal policies in an environment without government commitment to future policies. Furthermore, it provides the first quantitative assessment of the business-cycle properties of such policies for a baseline economy calibrated to the U.S. economy.

Second, the best sustainable capital income tax rate is shown to be slightly procyclical and positive on average, while the labor income tax is countercyclical. When compared with the data, BSSE provides a better account of the cyclical properties of the U.S. tax policy than Ramsey and MPE. Numerical exercises in this paper show that this better fit should be attributed to the history dependence and reputation mechanism considered by SSE.

Third, it makes a methodological contribution by developing a numerical algorithm to solve for SSE of economies with lack of government commitment. This approach is also applicable to a large set of environments, such as a monetary economy without commitment.

Of course, these results hold for a set of parameter values. As in most of the taxation literature, it seems difficult to state general results. Nevertheless, the specification of preferences in this paper is compatible with the existence of a balanced growth path and satisfies some properties related to uniform commodity taxation (see Chari and Kehoe (1999)).

I abstract from several important features of the data. First, for computational reasons, I maintain the assumption that the government cannot issue debt. In an environment without government default, it has been shown that public debt may serve as a substitute for commitment (see Lucas and Stokey (1983), Domínguez (2007)). If default is an option, then the government may be tempted to default on its debts and an equilibrium with positive debt may not exist (see Prescott (1977), Chari and Kehoe (1993), Stockman (2004), and Domínguez (2010)). Hence, it remains an open question as to whether or not a government's ability to issue debt can affect the properties of optimal taxation. A good understanding of these points will help us address important policy issues, such as whether the government should increase the tax rate or reconsider the debt limit when the economy is on the brink of default. This is particularly relevant in view of the recent fiscal woes and sovereign debt crisis.

I also abstract from agent heterogeneities. The presence of different types of agents may raise interesting political economy issues that deserve careful study. For computational reasons, I opt for a setup with a representative household. However, it should be noted that the algorithm developed in the current paper is built upon Feng et al. (2014), whose analysis is developed for a setting with heterogeneous agents. It should be straightforward to extend the method to an environment with different types of governments and households. Finally, the numerical method developed here faces some computational challenges—even with the help of high-performance, parallel computing. A better understanding of the construction of the equilibrium payoff set will certainly improve the efficiency of the algorithm (i.e. Abreu and Sannikov (2014)). I leave these questions for future research.

Appendix

A.1 *Proofs*

Proof of Proposition 1. I abstract from all uncertainties to simplify the exposition. If one plugs the definition of $m(s^{t+1})$ into the first order condition in terms of $k(s^t)$ of the recursive problem, then one gets the same expression as equation (10). Next I show that the transversality condition holds.

First, there exist $\underline{k} > 0$ and $\overline{k} > 0$ such that for all $k_0 \in [\underline{k}, \overline{k}]$ and feasible tax policy τ, each competitive equilibrium $Y\{k_0, \tau\}$ satisfies $k_{t+1} \in [\underline{k}, \overline{k}]$ for all $t \geq 0$ by applying an argument similar to Lemma 2 in Phelan and Stacchetti (2001).

Second, I show that there exists $\bar{m} < \infty$ such that $m_t \leq \bar{m}$ for all $t \geq 0$. Consider the situation where the capital stock of the economy hits the lower bound $\underline{k}$ and the government levies maximum tax rates $\overline{\tau} < 1$ on capital income and labor income. The Inada conditions $\lim_{c\to 0} u_c(c, l) = \infty$ and $\lim_{l\to 0} u_l(0, l) = 0$ imply that the household will be better off by spending a strictly positive amount of time $\underline{l} > 0$ in working so that she/he can obtain some income to finance a positive consumption. The household's income will be $(1 - \overline{\tau}_k)(F_k(\underline{k}, \underline{l}) - \delta)\underline{k} + (1 - \overline{\tau}_l)F_l(\underline{k}, \underline{l})\underline{l} > 0$. The first order condition (11) implies that the optimal consumption will be $[\frac{(1-\overline{\tau}_l)F_l(\underline{k},\underline{l})}{\nu_l(\underline{l})}]^{1/\sigma} > 0$, which yields a lower bound for c_t. Therefore, m_t is bounded by $\bar{m} = [1 + (1 - \overline{\tau}_k)(F_k(\underline{k}, \underline{l}) - \delta)]\underline{c}^{-\vartheta} = [1 + (1 - \overline{\tau}_k)(F_k(\underline{k}, \underline{l}) - \delta)]\frac{\nu_l(\underline{l})}{(1-\overline{\tau}_l)F_l(\underline{k},\underline{l})} > 0$.

Finally, I have $\lim_{t\to\infty} \beta^t m_t k_t \leq \lim_{t\to\infty} \beta^t \bar{m}\overline{k} = 0$. □

Proof of Proposition 2. By definition, $\overline{h}(k, s, m)$ is the maximum value of h at given (k, s, m). Therefore,

$$
\begin{aligned}
\overline{h}(k, s, m) &= \max_{\tau^k,\tau^n} \max_{m',h'} u(c, l) + G(g) + \beta\mathbb{E}\{h(k', s', m')\} \\
&= \max_{\tau^k,\tau^n} u(c, l) + G(g) + \max_{\tau^k,\tau^n} \max_{m',h'} \beta\mathbb{E}\{h(k', s', m')\} \\
&= \max_{\tau^k,\tau^n} u(c, l) + G(g) + \beta\mathbb{E}\{\overline{h}(k', s', m')\},
\end{aligned}
$$

where the first equality follows from the definition of $\overline{h}(k, s, m)$ and the second equality follows from the fact that there exists at most one pair of (c, l, k') consistent with $\{(\tau^k, \tau^n), m\}$ at given (k, s). The last equality uses the definition of $\overline{h}$.

A similar argument applies to $\underline{h}(k, s, m)$. A few comments are as follows. First, $\underline{h}(k, s, m) = \max_{\tau_k,\tau_l} \min_{m',h'} u(c, l) + G(g) + \beta\mathbb{E}\{h'\}$. Second, it should be noted that the value of $u(c, l) + G(g) + \beta\mathbb{E}\{\underline{h}(k', s', m')\}$ at given $\{\tau_k, \tau_l\}$ may be smaller than $\tilde{h}(k, s)$, which says that the incentive constraint is not satisfied when the government has the lowest continuation value. When this happens, the government needs a higher continuation value so that the incentive constraint is satisfied. However, the corresponding payoff for the present government cannot be higher than $\tilde{h}(k, s)$. This is because only the *minimization* operates when $\{\tau_k, \tau_l\}$ is given. There always exists $h' \in [\underline{h}(k', s', m'), \overline{h}(k', s', m')]$ to bind the incentive constraint when the worst continuation value violates the incentive constraint. Otherwise, m should not belong to the equilibrium value correspondence.

$\tilde{h}(k,s)$ is the payoff of WSSE, it must be in the lower boundary $\underline{h}(k,s,m)$. Because it is the worst of all, it must be equal to $\min_m \underline{h}(k,s,m)$. □

Proof of Theorem 1. First I show that the sequence of $\{\hat{\mathbf{W}}_n\}$ is decreasing and $\hat{\mathbf{W}}_n \supseteq \hat{\mathbf{W}}_{n+1}$. I claim that the upper boundary of $\{\hat{\mathbf{W}}_n\}$ is decreasing. This is because $\overline{h}^1(k,s,m)$ is defined as $\max_{\tau_k,\tau_l} u(c,l)+G(g)+\beta\mathbb{E}\{\overline{h}^0(k',s',m')\}$ such that $\psi=(\tau_k,\tau_l,c,l,k',g,w,r,\{m',h'\})$ is admissible with respect to $\hat{\mathbf{W}}_0$ at (k,s). The admissibility of the vector ψ implies that $(m,\overline{h}^1(k,s,m))\in\hat{\mathbf{W}}_0(k,s)$. Therefore, $\overline{h}^1(k,s,m)\leq\max\{h|(m,h)\in\hat{\mathbf{W}}_0(k,s)\}=\overline{h}^0(k,s,m)$. Similarly, I have $\underline{h}^1(k,s,m)\geq\underline{h}^0(k,s,m)$. The same argument holds for $\hat{\mathbf{W}}_n(k,s)$.

Since the sequence is decreasing, it has a limit $\hat{\mathbf{W}}_\infty$. Proposition 2 implies that $\mathbb{F}(\mathbf{V})=\mathbf{V}$. By a simple limit argument, one has $\lim_{n\to\infty}\hat{\mathbf{W}}_\infty=\mathbf{V}$. □

A.2 *Numerical implementation of operator* $\mathbb{F}$

Let $\mathbf{K}\times\mathbf{S}\times\mathbf{M}\times\mathbf{H}$ denotes the space of all equilibrium state vectors (k,s,m,h). Let us define a grid $\hat{\mathbf{K}}=\{k^{i_1}\}_{i_1=1}^{N_k}$. I assume that $\mathbf{S}$ is finite and $\mathbf{S}=\{s^{i_2}\}_{i_2=1}^{N_s}$. After this discretization, instead of a correspondence $\mathbf{W}:\mathbf{K}\times\mathbf{S}\to\mathbf{M}\times\mathbf{H}$, I have $\hat{\mathbf{W}}:\hat{K}\times S\to M\times H$. It is equivalent to think about this correspondence as $N_k\times N_s$ sets $\hat{\mathbf{W}}(k^{i_1},s^{i_2})$, where $k^{i_1}\in\hat{\mathbf{K}}$ and $s^{i_2}\in\mathbf{S}$. Notice that $\hat{\mathbf{W}}$ approximates $\mathbf{W}$ well as N_k becomes sufficiently large.

The algorithm starts with an initial guess $\mathbf{W}^0(k,s)=\{(m,h(k,s,m))\}$ and a per-determined tolerance $\varepsilon>0$.

Step 1-1. For any given $(k,s)\in\mathbf{K}\times\mathbf{S}$, pick $(m,h)\in\mathbf{W}^0(k,s)$. Let us store the pair (m,h) in $\Omega(k,s)$ if there exist $(\tau_k,\tau_l)\in T$ and $(m',h')\in\mathbf{W}^0(k',s')$ such that

$$h=u(c,l)+G(g)+\beta\mathbb{E}\{h'\}\geq\tilde{h}^0(k,s), \tag{47}$$

$$u_c(c,l)-\beta\mathbb{E}\{m'\}=0, \tag{48}$$

where (c,l,k') are determined as solutions for the equations

$$m-u_c(c,l)\cdot\left[1+(1-\tau_k)(r-\delta)\right]=0, \tag{49}$$

$$u_l(c,l)-(1-\tau_l)w\cdot u_c(c,l)=0, \tag{50}$$

$$(1-\tau_l)wl+\left[1+(1-\tau_k)(r-\delta)\right]k-\left(c+k'\right)=0. \tag{51}$$

Note that

$$\overline{h}^0(k,s,m)=\sup_h\{h|(m,h)\in\mathbf{W}^0(k,s)\}, \tag{52}$$

$$\underline{h}^0(k,s,m):=\inf_h\{h|(m,h)\in\mathbf{W}^0(k,s)\}, \tag{53}$$

$$\tilde{h}^0(k,s)=\min_m\underline{h}^0(k,s,m). \tag{54}$$

Step 1-2. Given (k, s), and $\Omega(k, s)$, denote $\Omega^m := \{m|(m, h) \in \Omega(k, s)\}$, and define

$$\overline{h}^1(k, s, m) = \max_{\tau_k, \tau_l} \max_{(m', h') \in \mathbf{W}^0} u(c, l) + G(g) + \beta \mathbb{E}\{\overline{h}^0(k', s', m')\}, \tag{55}$$

$$\begin{aligned} \underline{h}^1(k, s, m) = \max\Big\{ & \max_{\tau_k, \tau_l} \min_{(m', h') \in \mathbf{W}^0} u(c, l) + G(g) + \beta \mathbb{E}\underline{h}^0(k', s', m'), \\ & \tilde{h}^0(k, s)\Big\} \end{aligned} \tag{56}$$

for all $m \in \Omega^m$. Otherwise, I set

$$\overline{h}^1(k, s, m) = +\infty, \tag{57}$$

$$\underline{h}^1(k, s, m) = -\infty. \tag{58}$$

Next, let

$$\tilde{h}^1(k, s) = \min_{m \in \Omega^m} \underline{h}^1(k, s, m). \tag{59}$$

Step 2. Define $\mathbf{W}^1(k, s) = \{(m, h)|m \in \Omega^m(k, s), h \in [\underline{h}^1(k, s, m), \overline{h}^1(k, s, m)]\}$.

Step 3. Set $\mathbf{W}^* = \mathbf{W}^1$ if $\| \mathbf{W}^1 - \mathbf{W}^0 \| < \varepsilon$; otherwise, set $\mathbf{W}^0 = \mathbf{W}^1$ and go back to Step 1.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica*, 58, 1041–1063. [198, 201]

Abreu, D. and Y. Sannikov (2014), "An algorithm for two player repeated games with perfect monitoring." *Theoretical Economics*, 9, 313–338. [216]

Atkinson, A. B. and J. E. Stiglitz (1976), "The design of tax structure: Direct vs. indirect taxation." *Journal of Public Economics*, 6, 55–75. [190]

Azzimonti, M. and M. Talbert (2011), "Partisan cycles and the consumption volatility puzzle." Working Paper 11-21, Federal Reserve Bank of Philadelphia. [192, 197]

Bachmann, R. and J. H. Bai (2013a), "Politico-economic inequality and the comovement of government purchases." *Review of Economic Dynamics*, 16 (4), 565–580. [192]

Bachmann, R. and J. H. Bai (2013b), "Public consumption over the business cycle." *Quantitative Economics*, 4 (3), 417–451. [192, 197]

Barseghyan, L., M. Battaglini, and S. Coate (2010), "Fiscal policy over the real business cycle: A positive theory." Report, Princeton University. [192]

Benhabib, J. and A. Rusticchini (1997), "Optimal taxes without commitment." *Journal of Economic Theory*, 77, 231–259. [190, 192, 202]

Chamley, C. (1986), "Optimal taxation of capital income in general equilibrium with infinite lives." *Econometrica*, 54, 607–622. [208]

Chang, R. (1998), "Credible monetary policy in an infinite horizon model: Recursive approaches." *Journal of Economic Theory*, 81, 431–461. [191, 198, 202, 203, 207]

Chari, V. V., L. J. Christiano, and P. J. Kehoe (1994), "Optimal fiscal policy in a business cycle model." *Journal of Political Economy*, 102, 617–652. [191, 192, 212, 215]

Chari, V. V., L. J. Christiano, and P. J. Kehoe (1995), "Policy analysis in business cycle models." In *Frontiers of Business Cycle Research*, Vol. 12 (T. F. Cooley, ed.), Princeton University Press, Princeton. [191]

Chari, V. V. and P. J. Kehoe (1990), "Sustainable plans." *Journal of Political Economy*, 98, 783–802. [189, 190, 191, 195, 202, 210]

Chari, V. V. and P. J. Kehoe (1993), "Sustainable plans and debt." *Journal of Economic Theory*, 61, 230–261. [216]

Chari, V. V. and P. J. Kehoe (1999), "Optimal fiscal and monetary policy." Staff Report 251, Federal Reserve Bank of Minneapolis. [194, 216]

Debortoli, D. and R. Nunes (2010), "Fiscal policy under loose commitment." *Journal of Economic Theory*, 145, 1005–1032. [192]

Domínguez, B. (2007), "Public debt and optimal taxes without commitment." *Journal of Economic Theory*, 135, 159–170. [215, 216]

Domínguez, B. (2010), "The time-consistency of government debt and institutional restrictions on the level of debt." Manuscript. [203, 216]

Domínguez, B. and Z. Feng (2012), "The effect of time-consistent capital taxation on capital accumulation and welfare." Manuscript. [191, 205, 210, 211]

Feng, Z., J. Miao, A. Peralta-Alva, and M. Santos (2014), "Numerical simulation of nonoptimal dynamic equilibrium models." *International Economic Review*, 55, 83–110. [198, 203, 205, 216]

Fernandez-Villaverde, J. and A. Tsyvinski (2002), "Optimal fiscal policy in a business cycle model without commitment." Manuscript. [192, 203]

Fischer, S. (1980), "Dynamic inconsistency, cooperation, and the benevolent dissembling government." *Journal of Economic Dynamics & Control*, 2, 93–107. [195]

Judd, K. (1985), "Redistributive taxation in a simple perfect foresight model." *Journal of Public Economics*, 28, 59–83. [208]

Judd, K. (1993), "Optimal taxation in dynamic stochastic economies: Theory and evidence." Manuscript, Hoover Institution. [191, 213]

Judd, K., S. Yeltekin, and J. Conklin (2003), "Computing supergame equilibria." *Econometrica*, 71, 1239–1254. [203]

Klein, P. and J.-V. Ríos-Rull (2003), "Time-consistent optimal fiscal policy." *International Economic Review*, 44, 1217–1245. [191, 192, 197, 207, 212, 215]

Klein, P., P. Krusell, and J.-V. Ríos-Rull (2008), "Time-consistent public policy." *Review of Economic Studies*, 75, 789–808. [192, 197]

Kubler, F. and K. Schmedders (2010), "Tackling multiplicity of equilibria with Gröbner bases." *Operations Research*, 58, 1037–1050. [204]

Kydland, F. E. and E. C. Prescott (1977), "Rules rather than discretion: The inconsistency of optimal plans." *Journal of Political Economy*, 85, 473–491. [189, 195]

Kydland, F. E. and E. C. Prescott (1980), "Dynamic optimal taxation, rational expectations and optimal control." *Journal of Economic Dynamics & Control*, 2, 79–91. [198]

Lucas, R. and N. L. Stokey (1983), "Optimal fiscal and monetary policy in an economy without capital." *Journal of Monetary Economics*, 12, 55–93. [216]

Marcet, A. and R. Marimon (2011), "Recursive contracts." Manuscript, Universitat Pomepeu Fabra. [190, 192, 198, 202]

Martin, F. M. (2010), "Markov-perfect capital and labor taxes." *Journal of Economic Dynamics & Control*, 34, 503–521. [192, 197, 210, 214]

McGrattan, E. R. and E. C. Prescott (2000), "Is the stock market overvalued?" *Quarterly Review Federal Reserve Bank of Minneapolis*, Fall, 20–40. [207]

Ortigueira, S. (2006), "Markov-perfect optimal taxation." *Review of Economic Dynamics*, 9, 153–178. [192, 197]

Peralta-Alva, A. and M. S. Santos (2010), "Problems in the numerical simulation of models with heterogeneous agents and economic distortions." *Journal of the European Economic Association*, 8, 617–625. [198]

Phelan, C. and E. Stacchetti (2001), "Sequential equilibria in a Ramsey tax model." *Econometrica*, 69, 1491–1518. [191, 192, 196, 198, 199, 200, 201, 202, 203, 204, 208, 211, 217]

Prescott, E. C. (1977), "Should control theory be used for economic stabilization?" *Carnegie–Rochester Conference Series on Public Policy*, 7, 13–38. [216]

Prescott, E. C. (1986), "Theory ahead of business-cycle measurement." *Carnegie–Rochester Conference Series on Public Policy*, 25, 11–44. [207]

Reis, C. (2013), "Taxation without commitment." *Economic Theory*, 52, 565–588. [191]

Stockman, D. (2001), "Balance-budget rules: Welfare loss and optimal policies." *Review of Economic Dynamics*, 4, 438–459. [212, 215]

Stockman, D. (2004), "Default, reputation and balanced-budget rules." *Review of Economic Dynamics*, 7, 382–405. [191, 192, 216]

Zhu, X. (1992), "Optimal fiscal policy in a stochastic growth model." *Journal of Economic Theory*, 58, 250–289. [191]

Submitted June, 2013. Final version accepted January, 2014.