FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Optimal Fiscal Policy with Recursive Preferences

Anastasios G. Karantounias

Working Paper 2013-7
September 2013
Revised April 2015

Abstract: I study the implications of recursive utility for the design of optimal fiscal policy. I find that the standard policy prescriptions of the dynamic Ramsey literature are dramatically altered. Labor tax volatility is optimal and can be quantitatively substantial. Furthermore, labor taxes are countercyclical, display persistence independent of the stochastic properties of exogenous shocks and increase on average over time. At the intertemporal margin, there is a novel incentive for introducing distortions that can lead to an ex-ante capital subsidy. Ignoring the distinction between smoothing over time and smoothing over states is not an innocuous assumption for optimal policy.

JEL classification: D80; E62; H21; H63

Key words: Ramsey plan, Epstein-Zin, recursive utility, risk-sensitive preferences, labor tax, capital tax, martingale

Future versions of the paper can be found at http://www.atl-res.com/~Tasos/.

First draft: April 2010; Current version: October 2014. I am thankful to David Backus, Pierpaolo Benigno, R. Anton Braun, Vasco Carvalho, Lawrence Christiano, Kristopher S. Gerardi, Jonathan Halket, Lars Peter Hansen, Karen Kopecky, Juan Pablo Nicolini, Demian Pouzo, Victor Rios-Rull, Richard Rogerson, Thomas J. Sargent, Stanley E. Zin, to Christopher Sleet for his discussion, to seminar participants at the University of California at Davis, Carnegie Mellon University, Universitat Pompeu Fabra, the European University Institute, LUISS Guido Carli University and the University of Hong Kong and to conference participants at the 1st NYU Alumni Conference, the CRETE 2011 Conference, the 2012 SED Meetings, the 2012 EEA Annual Congress and the 2013 AEA Meetings. The views expressed here are the author's and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the author's responsibility.

Please address questions regarding content to Anastasios Karantounias, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8825, anastasios.karantounias@atl.frb.org.

1 Introduction

This paper studies the implications of the recursive preferences of Epstein and Zin (1989) and Weil (1990) for normative fiscal policy. Recursive preferences allow the disentanglement of the intertemporal elasticity of substitution from risk aversion. I consider standard dynamic Ramsey setups with complete markets and a representative household. Lump-sum taxes are not available. Linear taxes are used in order to finance exogenous stochastic government expenditures. Distortionary taxes give rise to an optimal policy problem, where a benevolent planner chooses taxes and debt under commitment in order to maximize the utility of the representative household.

The coefficients of intertemporal elasticity of substitution and risk aversion are two parameters that are a priori important in shaping dynamic policy. They control the desirability of taxing in the current versus future periods and the aversion of the policymaker towards shocks that hit the government budget constraint. Standard time-additive expected utility renders the analysis of the implications of these two parameters on tax-smoothing impossible. Moreover, by blurring the distinction between smoothing over dates and smoothing over states, standard time-additive expected utility ignores the temporal dimension of risk. Hence, questions about the implications of long-run fiscal risks on current tax and debt policies can be answered only in a limited way.

Several studies have shown that the use of recursive preferences produces more realistic asset prices, without altering significantly the desirable quantity predictions of standard time-additive expected utility models. This feature has popularized recursive preferences in a fastly growing literature that tries to merge macroeconomics and finance.[1] More realistic asset prices though raise natural questions about standard tax-smoothing prescriptions. Asset prices are central for the canonical optimal taxation problem, since they inform the policymaker about the desirability of debt and therefore the extent to which he should resort to distortionary taxation. However, little is known about recursive utility and optimal fiscal policy even in the simplest Ramsey setup. This is the task of the current paper.

Consider first an economy without capital and complete markets as in Lucas and Stokey (1983). The basic lessons of dynamic Ramsey taxation in the time-additive expected utility economy are two. First, the labor tax should be *constant* if period utility features constant elasticities. Even when elasticities are not constant, the volatility of the labor tax is small. These are the classic tax-smoothing results. Second, whenever the labor tax varies, it inherits the stochastic properties of the exogenous shocks. Thus, optimal labor taxes do not constitute a distinct source of persistence in the economy.

These classic results are overturned in the same economy with recursive preferences. Tax volatility is optimal even when period elasticities are constant and can be quantitatively substantial. Moreover, labor taxes persist independently of the stochastic properties of the exogenous shocks.

[1] The literature is vast. See indicatively Tallarini (2000), Bansal and Yaron (2004), Hansen et al. (2008), Gourio (2012) and Petrosky-Nadeau et al. (2013) among others.

In order to understand the mechanism behind this result, consider a fiscal shock that hits the government budget constraint. The policy prescription in the time-additive economy is to insure against this shock by issuing state-contingent debt. The planner issues ex-ante a high amount of debt contingent on low government expenditures that will be paid back by running a surplus and a low amount of debt contingent on high government expenditures, in order to be able to run a deficit. In both cases, the optimal amount of state-contingent debt is such that the tax distortion stays constant. Hence, state-contingent debt acts as a shock absorber.

Consider now the same situation with recursive preferences. The distinguishing feature of recursive utility is that the "long-run," as captured by future utilities, matters, i.e. the entire intertemporal profile of consumption and leisure is taken into account in the pricing of state-contingent claims. Typical parameterizations imply that the household is *averse* to volatility in future utility. From an optimal taxation perspective, the household's utility is reduced when the government is issuing debt, and this reduction is priced. Hence, equilibrium prices of state-contingent claims become *debt-elastic*. Since they provide one unit of consumption under adverse utility conditions, claims on states of the world at which the planner issues high debt command a high equilibrium price, and therefore, a low state-contingent return. In contrast, claims on states of the world at which the planner issues less debt command a high-state contingent return. So there is a novel *negative* covariance between debt and state-contingent returns with recursive utility that is absent in the time-additive case. As a result, instead of keeping the tax rate constant, the planner issues more debt (since it is cheaper) and increases taxes for low shocks and buys more assets (that have higher return) and taxes less for high shocks.

Hence, the fiscal insurance effects of state-contingent debt are *amplified* in comparison to a time-additive economy. The planner is essentially "over-insuring" against fiscal shocks and runs larger surpluses and deficits in order to use the novel valuation effects that emerge with recursive utility. Taxes decrease for high shocks and increase for low shocks, which leads to a negative correlation of taxes with government expenditures and therefore countercyclicality of the tax rate.

In a dynamic setup with recursive utility a tax rate at a future period affects the entire sequence of state-contingent claims up to that period due to the forward-looking nature of asset prices. As a result, the planner does not choose future tax rates independently from the past. Optimal tax rates exhibit *persistence* independent of the stochastic properties of exogenous shocks. Furthermore, the planner finds it optimal to delay taxes for the future, so there is a *back-loading* of distortions.

These theoretical results have also important quantitative implications. In a series of numerical exercises with a utility function that delivers perfect tax-smoothing in the time-additive economy I show the following: First, the fiscal insurance efforts of the government result into changes in tax rates that are strongly negatively correlated with government expenditures and therefore strongly countercyclical. Second, the government's "over-insurance" can be quite large. Third, the autocorrelation of the tax rate is very high, even when shocks are identically and independently distributed. Fourth, the mean and the standard deviation of the tax rate are increasing over time,

so there is on average a back-loading of tax distortions and a "fanning-out" of the distribution of the tax rate. Depending on the time horizon, the volatility of the tax rate can be substantial. A similar pattern of a positive drift and an increasing volatility over time arises for the debt-to-output ratio. Finally, in the very long-run, the stationary distributions of the tax rate and the debt-to-output ratio exhibit a high mean and substantial volatility.

Consider now an economy with capital and time-additive expected utility as in Chari et al. (1994) and Zhu (1992). These authors have shown that the basic insights about labor tax-smoothing and inherited persistence from exogenous shocks of Lucas and Stokey (1983) hold also in an economy with capital. Furthermore, they show that the ex-ante tax rate on capital income is zero for a particular constant elasticity utility function and essentially zero quantitatively otherwise.

I show how the analysis of the allocation of distortions with recursive utility extends in this economy with capital and derive further implications for capital taxes. Essentially the same mechanism about a non-constant and persistent allocation of distortions that I highlighted before in economy without capital holds also in an economy with capital. The relevant variable that captures the incentives of the planner is wealth instead of just debt. The planner uses the negative covariance of wealth with state-contingent returns that emerges with recursive utility and allocates more distortions on high-wealth states, and less distortions on low-wealth states, with similar implications for labor taxes as in the economy without capital.

Furthermore, I show how this mechanism makes the ex-ante capital tax non-zero, even for a constant elasticity period utility function. The ex-ante capital tax represents a weighted average of state-contingent intertemporal distortions. In an economy with recursive utility, low-wealth states command a lower labor tax and essentially a state-contingent capital subsidy whereas high-wealth states a higher labor tax and a state-contingent capital tax. I show that when the government hedges fiscal shocks by taking a low wealth position contingent on high government spending and a high wealth position contingent on low government spending, the state-contingent intertemporal subsidies are weighted more than the state-contingent intertemporal taxes, leading to an *ex-ante* capital *subsidy*.

In conclusion, when the attitude towards time is different than the attitude towards risk, the classic tax-smoothing results of the dynamic Ramsey literature do not hold. It is important to note that this is not an outcome of any kind of frictions. Instead, it is a natural implication of the way the planner is using state-contingent returns in order to minimize welfare distortions. Policy prescriptions favor tax volatility, persistence and postponing taxes for the future, with important quantitative repercussions.

Related literature. The main reference on optimal taxation with time-additive expected utility for an economy without capital is Lucas and Stokey (1983). The respective references for an economy with capital are Chari et al. (1994) and Zhu (1992). The models I examine reduce to

the models analyzed in these studies, if I equate the risk aversion parameter to the inverse of the intertemporal elasticity of substitution parameter. Furthermore, the economy with capital reduces to the deterministic economy of Chamley (1986), if I shut off uncertainty.[2]

Related studies include Farhi and Werning (2008), who analyze the implications of recursive preferences for private information setups and Karantounias (2013), who analyzes optimal taxation in an economy without capital, in a setup where the household entertains fears of misspecification but the fiscal authority does not. Of interest is also the work of Gottardi et al. (2014), who study optimal taxation of human and physical capital with uninsurable idiosyncratic shocks and recursive preferences.

Other studies have analyzed the interaction of fiscal policies and asset prices with recursive preferences from a positive angle. Gomes et al. (2013) build a quantitative model and analyze the implications of fiscal policies on asset prices and the wealth distribution. Croce et al. (2012a) show that corporate taxes can create sizeable risk premia with recursive preferences. Croce et al. (2012b) analyze the effect of exogenous fiscal rules on the endogenous growth rate of the economy. None of these studies though considers *optimal* policy.

Another relevant line of research is the analysis of optimal taxation with time-additive expected utility and restricted asset markets as in Aiyagari et al. (2002), Farhi (2010), Shin (2006), Sleet and Yeltekin (2006) or with time-additive expected utility and private information as in Sleet (2004). In studies like Aiyagari et al. (2002), the lack of insurance markets also causes the planner to allocate distortions in a time-varying and persistent way. However, the lack of markets implies that the planner tries to decrease taxes when when government spending is small and increase taxes when government spending is high. Instead, the opposite happens in the current paper.[3] More generally, with incomplete markets as in Aiyagari et al. (2002), the planner would like to allocate distortions in a constant way across states and dates but he *cannot*, whereas with complete markets and recursive preferences he could in principle follow a constant distortion policy, but does not find it optimal to do so.

The paper is organized as follows. Section 2 uses a simple static setup to highlight the novel mean-volatility trade-off that shows up with recursive utility. Section 3 lays out an economy without capital and section 4 sets up the Ramsey problem, its recursive formulation and derives the associated optimality conditions. The entire action with recursive utility is captured by the *excess burden* of distortionary taxation, a multiplier that reflects how distortions are allocated across states and dates. The excess burden of taxation is constant with time-additive expected

[2]It is worth noting that Chamley demonstrated the generality of the zero capital tax result at the deterministic steady state by using the preferences of Koopmans (1960). See Chari and Kehoe (1999) for a comprehensive survey of optimal fiscal policy.

[3]Furthermore, with incomplete markets as in Aiyagari et al. (2002), it is typically optimal to front-load distortions in order to create a buffer stock of assets, furnishing a tax rate with a negative drift. In contrast, in the current analysis the tax rate exhibits a positive drift, in order to take advantage of cheaper state-contingent debt. It is interesting to observe that Sleet (2004) also obtains a positive drift in the tax rate in a setup with private information about the government spending needs.

utility, whereas with recursive utility it exhibits a martingale-like behavior. Section 5 is devoted to its analysis. The implications for labor taxes are derived in section 6. Detailed numerical illustrations are provided in section 7. Section 8 extends the analysis in an economy with capital and derives implications for the ex-ante capital tax. Finally, section 9 concludes and an Appendix follows.

2 A mean-volatility trade-off

The non-optimality of a constant tax rate that arises with recursive utility can be sharply seen using a perturbation argument in a very simple static setup. Consider an economy without capital and government expenditures that take two values, $g_1 < g_2$, with probability $\pi_i, i = 1, 2$. There are two state-contingent claims that are traded before the realization of uncertainty at prices q_i, with the normalization $\sum_i q_i = 1$. These claims provide one unit of consumption at state g_i and zero otherwise. The household consumes c and works h hours. The resource constraint in the economy is $c_i + g_i = h_i, i = 1, 2$. Let the household have risk-sensitive preferences with parameter $\sigma < 0$,

$$\sigma^{-1} \ln \sum_i \pi_i \exp(\sigma U(c_i, h_i))$$

where $U = c - \frac{1}{2}h^2$. The household is *averse* to volatility in utility, in contrast to the expected utility household ($\sigma = 0$), that would be indifferent towards it. Equilibrium labor and consumption are $h_i = 1 - \tau_i$ and $c_i = 1 - \tau_i - g_i$ respectively, where τ_i is the tax rate on labor income. The equilibrium utility of the household as function of the tax rate at shock g is $u(\tau, g) = \frac{1}{2}(1 - \tau^2) - g$. Equilibrium government debt, paid back by surpluses, is $b(\tau, g) = \tau(1 - \tau) - g$. The aversion of the household to utility volatility is reflected in equilibrium prices. In particular, prices are equal to the *utility-adjusted* probabilities $q_i = \pi_i m_i$, where $m_i \equiv \exp(\sigma u(\tau_i, g_i))/\sum_i \pi_i \exp(\sigma u(\tau_i, g_i))$, a change of measure. Due to aversion to utility volatility, marginal rates of substitution, and therefore equilibrium prices, are increased when utility is low. In contrast to the expected utility case, where equilibrium prices would be π_i, government policy can affect prices. A high tax rate, by reducing utility, increases the price of a state-contingent claim.

The household's ex-ante equilibrium utility in terms of τ_i is

$$V(\tau_1, \tau_2) \equiv \sigma^{-1} \ln \sum_i \pi_i \exp(\sigma u(\tau_i, g_i)). \tag{1}$$

The equilibrium government budget constraint delivers the implementability constraint,

$$IC(\tau_1, \tau_2) \equiv \pi_1 \exp(\sigma u(\tau_1, g_1)) b(\tau_1, g_1) + \pi_2 \exp(\sigma u(\tau_2, g_2)) b(\tau_2, g_2) = 0. \tag{2}$$

The problem of the Ramsey planner is to choose τ_i in order to maximize (1) subject to (2). It is easy to see that a constant tax rate is optimal in the expected utility case. Assume now that we consider the same policy prescription as in the expected utility case and set a constant tax rate $\tau_i = \bar{\tau}, i = 1, 2$. The tax $\bar{\tau}$ has to satisfy $IC(\bar{\tau}, \bar{\tau}) = 0$, i.e. it has to finance the present value of government expenditures. Let $\bar{b}_i \equiv b(\bar{\tau}, g_i)$ denote the debt position of the government at a constant tax rate. Tax revenues are constant and the government is issuing debt that is paid back by running a surplus for the low shock, $\bar{b}_1 > 0$, whereas it runs a deficit that is financed by government assets, $\bar{b}_2 < 0$, for the high shock. Let also $\bar{m}_i$ and $\bar{u}_i$ denote the respective change of measure and equilibrium utility evaluated at the constant tax rate policy.

Consider now a small *perturbation* of the tax rate in the direction (d_1, d_2), $\tau_i = \bar{\tau} + \epsilon d_i, i = 1, 2, \epsilon > 0$ and let $W(\epsilon) \equiv V(\bar{\tau}+\epsilon d_1, \bar{\tau}+\epsilon d_2)$ denote the respective welfare. Feasible directions have to satisfy $IC_1(\bar{\tau}, \bar{\tau})d_1 + IC_2(\bar{\tau}, \bar{\tau})d_2 = 0$, with partial derivatives $IC_i(\bar{\tau}, \bar{\tau}) = \pi_i \exp(\sigma \bar{u}_i)[-\sigma\bar{\tau}\bar{b}_i + 1 - 2\bar{\tau}] > 0, i = 1, 2.$[4] The first term in IC_i denotes the marginal appreciation in the value of debt due to the effect of taxes on equilibrium prices times the debt position $\bar{b}_i$, $-\sigma\bar{\tau}\bar{b}_i$, whereas the second term denotes the marginal tax revenue, $1 - 2\bar{\tau}$. The change in welfare for a small perturbation is

$$W'(0) = V_1(\bar{\tau}, \bar{\tau})d_1 + V_2(\bar{\tau}, \bar{\tau})d_2 = -\bar{\tau}[\pi_1\bar{m}_1 d_1 + \pi_2\bar{m}_2 d_2],$$

which for feasible directions becomes

$$W'(0) = -\sigma\bar{\tau}^2\pi_1\bar{m}_1\frac{\bar{b}_1 - \bar{b}_2}{-\sigma\bar{\tau}\bar{b}_2 + 1 - 2\bar{\tau}}d_1 > 0,$$

for $d_1 > 0$, since $\bar{b}_1 > \bar{b}_2$. Therefore, an increase in tax rate at the low shock, which is accompanied by a decrease in tax rate at the high shock ($d_2 < 0$), *increases* welfare in comparison to a constant tax policy. The government would like to run a larger surplus at the low shock and a larger deficit at the high shock. Another way to state the same result is that the marginal rate of substitution V_1/V_2 is *smaller* than the absolute value of the slope of the implementability constraint IC_1/IC_2 at the *constant* tax rate policy, since the welfare change can be written as $W'(0) = V_2\big(V_1/V_2 - IC_1/IC_2\big)d_1$. Thus, the reduction in the tax rate at the high shock that becomes possible by an increase in the tax rate at the low shock is larger than the corresponding reduction in the tax rate that would keep welfare the same. As a result, welfare increases.

The improvement in welfare is coming from the effect of the tax rate on equilibrium prices. A higher tax at the low shock and a lower tax at the high shock increases the price of the Arrow

[4] It is assumed that σ is small enough in absolute value so that $-\sigma\bar{\tau}\bar{b}_2 + 1 - 2\bar{\tau} > 0$. Furthermore, $\bar{\tau} < 1/2$, i.e. the constant tax is at the increasing part of the Laffer curve.

claim at the lower shock and decreases the price at the high shock. This is welfare improving since at the low shock the government is issuing debt $\bar{b}_1 > 0$, so its value increases, whereas at the high shock the government buys assets $\bar{b}_2 < 0$, so they become cheaper. Thus, a non-constant tax policy $\tau_1 > \bar{\tau} > \tau_2$ decreases the state-contingent return on debt and increases the state-contingent return on assets, allowing welfare to increase.

Therefore, tax volatility can be welfare-improving with recursive utility. This may seem surprising since with the risk-sensitive preferences used above, the household is actually more averse to fluctuations in utility than the expected utility household. However, the increase in tax volatility is accompanied by a *reduction* in the mean tax rate. In particular, let $E^m\tau$ denote the "risk-sensitive" mean tax rate, $E^m\tau \equiv \bar{\tau} + \epsilon[\pi_1 m_1 d_1 + \pi_2 m_2 d_2]$, with a marginal change equal to $\frac{dE^m\tau}{d\epsilon}|_{\epsilon=0} = \pi_1 \bar{m}_1 d_1 + \pi_2 \bar{m}_2 d_2$. The risk-sensitive mean tax rate is the welfare-relevant object, since changes in welfare change are negatively related to changes in the mean tax rate, $W'(0) = -\bar{\tau}\frac{dE^m\tau}{d\epsilon}|_{\epsilon=0}$. So an increase in welfare corresponds to a decrease in the mean tax rate. In other words, the decreased mean tax rate has a positive *first-order* effect on welfare which dominates the negative *second-order* effect of the increased tax volatility, delivering a trade-off between mean and volatility of the tax rate that is absent with expected utility.

Although this simple static setup has a special structure due to quasi-linear utility (which essentially allows the use of the more intuitive "dual" approach), it conveys several messages that will be valid in the rest of the paper for more general period utility functions. In particular, tax distortions will be captured by a multiplier on the implementability constraint, the *excess burden* of taxation, which in contrast to the expected utility case, will *not* be constant anymore. Instead, the planner allocates more distortions on events where he has relatively high debt in marginal utility units (and not just debt as in the quasi-linear setup), exactly because this way debt becomes cheaper. If debt in marginal utility units is negatively correlated with government spending shocks, then tax distortions become negatively correlated with government spending as in the simple setup. The aversion to utility volatility will correspond to a recursive utility criterion where risk aversion is larger than the inverse of intertemporal rate of substitution, or in other words, a preference for early resolution of uncertainty. Finally, the entire intertemporal profile of future consumption and leisure matters in a fully dynamic setup, a fact which delivers further implications for the persistence, volatility and the back-loading of distortions.

3 Economy without capital

I start the analysis of the optimal allocation of distortions with recursive utility in an economy without capital as in Lucas and Stokey (1983). In a later section, I show how the analysis extends in an economy with capital as in Chari et al. (1994) and Zhu (1992) and derive the implications for capital taxation.

Time is discrete and the horizon is infinite. There is uncertainty in the economy stemming

from exogenous government expenditure shocks g. Shocks take values in a finite or countable set. Let $g^t = (g_0, g_1, ..., g_t)$ denote the partial history of shocks till time t and let $\pi_t(g^t)$ denote the probability of this history. The initial shock is assumed to be given, so that $\pi_0(g_0) = 1$.

The economy is populated by a representative household that is endowed with one unit of time and consumes $c_t(g^t)$, works $h_t(g^t)$, pays linear labor income taxes with rate $\tau_t(g^t)$ and trades in complete asset markets. Leisure of the household is $l_t(g^t) = 1 - h_t(g^t)$. The notation denotes that the relevant variables are measurable functions of the history g^t. Labor markets are competitive, which leads to an equilibrium wage of unity, $w_t(g^t) = 1$. The resource constraint in the economy reads

$$c_t(g^t) + g_t = h_t(g^t), \forall t, g^t. \tag{3}$$

3.1 Preferences

The representative household derives utility from random sequences of consumption $\{c\} \equiv \{c_t(g^t)\}_{t \geq 0, g^t}$ and leisure $\{l\} \equiv \{l_t(g^t)\}_{t \geq 0, g^t}$. The household ranks consumption and leisure plans following a recursive utility criterion of Kreps and Porteus (1978). In particular, let V_t denote the household's utility at time t. V_t follows the recursion

$$V_t = W(u(c_t, 1 - h_t), \mu_t(V_{t+1})). \tag{4}$$

The household derives utility from a composite good that consists of consumption and leisure, $u(c_t, 1 - h_t)$, and from the certainty equivalent of continuation utility $\mu_t \equiv \phi^{-1}(E_t \phi(V_{t+1}))$, where E_t denotes the conditional expectation operator given information at t with respect to the measure π, and $\phi(.)$ is an increasing and concave function that is capturing atemporal risk aversion. The time preference of the household between the composite good today and the certainty equivalent of continuation utility is captured by the time aggregator $W(.)$.

I focus my analysis on the isoelastic preferences of Epstein and Zin (1989) and Weil (1990) (EZW henceforth), and use a constant elasticity of substitution time aggregator and a power utility certainty equivalent. In particular, EZW preferences take the form

$$V_t = [(1 - \beta) u(c_t, 1 - h_t)^{1-\rho} + \beta (E_t V_{t+1}^{1-\gamma})^{\frac{1-\rho}{1-\gamma}}]^{\frac{1}{1-\rho}}, \tag{5}$$

where u is assumed to be positive. The parameter $1/\rho$ captures the intertemporal elasticity of substitution between the composite good and the certainty equivalent of continuation utility, whereas the parameter γ represents risk aversion with respect to atemporal gambles in continuation values. These preferences reduce to standard time-additive expected utility when $\rho = \gamma$. Furthermore, the separation of risk aversion and intertemporal elasticity of substitution inherently

imposes a preference for early ($\rho < \gamma$) or late ($\rho > \gamma$) resolution of uncertainty, whereas with expected utility ($\rho = \gamma$) there is indifference to the temporal resolution of uncertainty. Throughout the paper I assume preference for *early* resolution of uncertainty ($\rho < \gamma$), which is the typical parameterization necessary in order to match asset-pricing facts.[5]

It is useful for later purposes to bear in mind the monotonic transformation $v_t \equiv \frac{V_t^{1-\rho}-1}{(1-\beta)(1-\rho)}$, which will be called the ρ-transformation.[6] The utility recursion (5) becomes in this case

$$v_t = U_t + \beta \frac{\left[E_t[1+(1-\beta)(1-\rho)v_{t+1}]^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{1-\rho}{1-\gamma}} - 1}{(1-\beta)(1-\rho)}, \tag{6}$$

where $U(c_t, 1-h_t) \equiv \frac{u_t^{1-\rho}-1}{1-\rho}$, with respective derivatives $U_i = u^{-\rho}u_i, i = c, l$. I refer to U as period utility and to $U_i, i = c, l$ as (period) marginal utility of consumption and leisure. This transformation provides a useful interpretation of $\rho < \gamma$ as a situation where the household is *averse* to volatility in continuation utilities.[7]

Of particular interest is the case when the intertemporal elasticity of substitution becomes unity, $\rho = 1$. Then (5) becomes $V_t = u_t^{1-\beta}\mu_t^{\beta}$, and applying the ρ-transformation for $\rho = 1$, $v_t = \frac{\ln V_t}{1-\beta}$, we get the recursion

$$v_t = \ln u_t + \frac{\beta}{(1-\beta)(1-\gamma)} \ln E_t \exp\left[(1-\beta)(1-\gamma)v_{t+1}\right], \tag{7}$$

which for $\gamma > 1$ has the interpretation of a *risk-sensitive* recursion with risk-sensitivity parameter $\sigma \equiv (1-\beta)(1-\gamma)$.[8]

It will be useful to define

$$m_{t+1} \equiv \frac{V_{t+1}^{1-\gamma}}{E_t V_{t+1}^{1-\gamma}}, t \geq 0, \tag{8}$$

with $m_0 \equiv 1$. For $\rho = 1$, the corresponding definition is $m_{t+1} = \frac{\exp[(1-\beta)(1-\gamma)v_{t+1}]}{E_t \exp[(1-\beta)(1-\gamma)v_{t+1}]}$. Note that

[5] See Epstein and Zin (1989) and Weil (1990) for detailed discussions of how risk aversion, intertemporal elasticity of substitution and preference for early or late resolution of uncertainty are intertwined.

[6] Applying the respective γ-transformation $f(V) \equiv \frac{V^{1-\gamma}-1}{(1-\beta)(1-\gamma)}$ on (5) delivers the representation used in Weil (1990).

[7] Define the monotonic function $H(x) \equiv \left[\left(1+(1-\beta)(1-\rho)x\right)^{\frac{1-\gamma}{1-\rho}} - 1\right]/[(1-\beta)(1-\gamma)]$. Recursion (6) can be written as $v_t = U_t + \beta H^{-1}(E_t H(v_{t+1}))$. $H(x)$ is *concave* for $\rho < \gamma$. Thus, $\rho < \gamma$ denotes aversion to volatility in v_{t+1}.

[8] More generally, in the case of risk-sensitive preferences, the period utility function is not restricted to be logarithmic and the recursion takes the form $v_t = U_t + \frac{\beta}{\sigma} \ln E_t \exp(\sigma v_{t+1})$, $\sigma < 0$. There is an intimate link between the risk-sensitive recursion and the multiplier preferences of Hansen and Sargent (2001) that capture the decision maker's fear of misspecification of the probability model π. See Strzalecki (2011) and Strzalecki (2013) for a decision-theoretic treatment of the multiplier preferences and an analysis of the relationship between ambiguity aversion and temporal resolution of uncertainty respectively.

m_{t+1} is positive since V_{t+1} is positive, and that $E_t m_{t+1} = 1$. So m_{t+1} can be interpreted as a change of measure of the conditional probability density $\pi_{t+1}(g_{t+1}|g^t)$, or, in other words, a conditional likelihood ratio. Similarly, define the product of the conditional likelihood ratios as

$$M_t(g^t) \equiv \prod_{i=1}^{t} m_i(g^i), \tag{9}$$

with the normalization $M_0 \equiv 1$. This object has the interpretation of an unconditional likelihood ratio and is a martingale with respect to measure π. I refer to $\pi_t \cdot M_t$ as the *continuation-value* adjusted measure.

3.2 Competitive equilibrium

Household's problem. The representative household is choosing $\{c_t(g^t), h_t(g^t), b_{t+1}(g^{t+1})\}_{t\geq 0, g^t}$ to maximize $V_0(\{c\}, \{h\})$ subject to

$$c_t(g^t) + \sum_{g_{t+1}} p_t(g_{t+1}, g^t) b_{t+1}(g^{t+1}) \leq (1 - \tau_t(g^t)) h_t(g^t) + b_t(g^t),$$

the non-negativity constraint for consumption $c_t(g^t) \geq 0$ and the feasibility constraint for labor $h_t(g^t) \in [0, 1]$, where initial debt b_0 is given. The variable $b_{t+1}(g^{t+1})$ stands for the holdings at history g^t of an Arrow claim that delivers one unit of consumption next period if the state is g_{t+1} and zero units otherwise. This security trades at price $p_t(g_{t+1}, g^t)$ in units of the history-contingent consumption $c_t(g^t)$.

The household is also facing a no-Ponzi-game condition that takes the form

$$\lim_{t\to\infty} \sum_{g^{t+1}} q_{t+1}(g^{t+1}) b_{t+1}(g^{t+1}) \geq 0 \tag{10}$$

where $q_t(g^t) \equiv \prod_{i=0}^{t-1} p_i(g_{i+1}, g^i)$, with the normalization $q_0 \equiv 1$. In other words, q_t stands for the price of an Arrow-Debreu contract at $t = 0$.

Government. The government taxes labor income and issues state-contingent debt in order to finance the exogenous government expenditures. The dynamic budget constraint of the government takes the form

$$b_t(g^t) + g_t = \tau_t(g^t) h_t(g^t) + \sum_{g_{t+1}} p_t(g_{t+1}, g^t) b_{t+1}(g^{t+1}),$$

When $b_t > 0$, the government borrows from the household and when $b_t < 0$, the government lends to the household.

Definition 1. *A competitive equilibrium with taxes is a stochastic process for prices $\{p\}$, an allocation $\{c, h, b\}$ and a government policy $\{g, \tau, b\}$ such that: 1) Given prices $\{p\}$ and taxes $\{\tau\}$, the allocation $\{c, h, b\}$ solves the households's problem. 2)Prices are such so that markets clear, i.e. the resource constraint (3) holds.*

3.3 Household's optimality conditions

The labor supply decision of the household is governed by

$$\frac{U_l(g^t)}{U_c(g^t)} = 1 - \tau_t(g^t), \tag{11}$$

which equates the marginal rate of substitution between consumption and leisure with the after-tax wage. The first-order condition with respect to an Arrow security equates its price to the household's intertemporal marginal rate of substitution,[9]

$$\begin{aligned} p_t(g_{t+1}, g^t) &= \beta \pi_{t+1}(g_{t+1}|g^t) \left(\frac{V_{t+1}(g^{t+1})}{\mu_t(V_{t+1})} \right)^{\rho-\gamma} \frac{U_c(g^{t+1})}{U_c(g^t)} \\ &= \beta \pi_{t+1}(g_{t+1}|g^t) m_{t+1}(g^{t+1})^{\frac{\rho-\gamma}{1-\gamma}} \frac{U_c(g^{t+1})}{U_c(g^t)}, \end{aligned} \tag{12}$$

where the second line uses the definition of the conditional likelihood ratio (8).[10] The transversality condition is

$$\lim_{t\to\infty} \sum_{g^{t+1}} \beta^{t+1} \pi_{t+1}(g^{t+1}) M_{t+1}(g^{t+1})^{\frac{\rho-\gamma}{1-\gamma}} U_c(g^{t+1}) b_{t+1}(g^{t+1}) = 0. \tag{13}$$

The stochastic discount factor S_{t+1} with EZW utility is

$$S_{t+1} \equiv \beta \left(\frac{V_{t+1}}{\mu_t} \right)^{\rho-\gamma} \frac{U_{c,t+1}}{U_{ct}} = \beta m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} \frac{U_{c,t+1}}{U_{ct}}. \tag{14}$$

The disentanglement of risk aversion and intertemporal elasticity of substitution ($\rho \neq \gamma$) introduces *continuation values*, scaled by their certainty equivalent μ_t, into the stochastic discount factor. As a result, besides caring for the short-run ($U_{c,t+1}/U_{ct}$), the household cares also for the "long-run", in the sense that the entire sequence of future consumption and leisure will directly

[9] The derivative of the utility index with respect to c_{t+i} can be calculated recursively from the relationship $\frac{\partial V_t}{\partial c_{t+i}} = \frac{\partial V_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial V_{t+1}} \frac{\partial V_{t+1}}{\partial c_{t+i}}, i \geq 1$. Similarly for labor. Therefore, we have $\frac{\partial V_0}{\partial c_t} = (1-\beta) V_0^{\rho} \beta^t \pi_t M_t^{\frac{\rho-\gamma}{1-\gamma}} U_{ct}$ and $\frac{\partial V_0}{\partial h_t} = -(1-\beta) V_0^{\rho} \beta^t \pi_t M_t^{\frac{\rho-\gamma}{1-\gamma}} U_{lt}$. For the ρ-transformation we have $\frac{\partial v_0}{\partial c_t} = \beta^t \pi_t M_t^{\frac{\rho-\gamma}{1-\gamma}} U_{ct}$ and $\frac{\partial v_0}{\partial h_t} = -\beta^t \pi_t M_t^{\frac{\rho-\gamma}{1-\gamma}} U_{lt}$.

[10] The change of measure M_t allows also a concise expression for the price of an Arrow-Debreu contract at $t = 0$, $q_t(g^t) = \beta^t \pi_t(g^t) M_t(g^t)^{\frac{\rho-\gamma}{1-\gamma}} \frac{U_c(g^t)}{U_c(g_0)}$.

affect its intertemporal marginal rate of substitution today.[11]

4 Ramsey problem

I formulate the Ramsey problem under commitment. The problem of the planner is to choose at period zero the competitive equilibrium that maximizes the utility of the representative household. I follow the primal approach of Lucas and Stokey (1983) and eliminate taxes and equilibrium prices from the competitive equilibrium conditions. As a result, the problem of the planner reduces to a problem of choosing allocations that satisfy the resource constraint (3) and *implementability* constraints, i.e. constraints that allow the optimal allocation to be implemented as a competitive equilibrium.

4.1 Implementability constraints

Use (11) and (12) to eliminate labor taxes and equilibrium prices from the household's dynamic budget constraint and multiply by the marginal utility of consumption to get

$$U_{ct}b_t = \Omega_t + \beta E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} U_{c,t+1} b_{t+1}, \tag{15}$$

where

$$\Omega_t \equiv U_{ct}c_t - U_{lt}h_t. \tag{16}$$

The variable Ω_t stands for consumption net of after-tax labor income in marginal utility of consumption units. In equilibrium it is also equal to the *primary surplus* in marginal utility units. Note that Ω_t is a function of consumption and labor only, $\Omega_t = \Omega(c_t, h_t)$. We can summarize this discussion in terms of a proposition:

Proposition 1. *The Ramsey planner faces the following implementability constraints:*

$$U_{ct}b_t = \Omega_t + \beta E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} U_{c,t+1} b_{t+1}, t \geq 0$$

where $c_t \geq 0$, $h_t \in [0,1]$ and (b_0, g_0) given. Furthermore, the transversality condition (13) has to be satisfied. The conditional likelihood ratios $m_{t+1}, t \geq 0$, defined in (8), are endogenously determined by continuation values that follow the recursion (5).

Complete markets allow the collapse of the household's dynamic budget constraint to a unique

[11] Bansal and Yaron (2004) and Hansen et al. (2008) have explored ways of making the intertemporal profile of consumption quantitatively important in order to increase the market price of risk.

intertemporal budget constraint. However, maintaining the dynamic budget constraint of the household is convenient for a recursive formulation.

Definition 2. *The Ramsey problem is to maximize at $t = 0$ the utility of the representative household subject to the implementability constraints of proposition 1 and the resource constraint (3).*

4.2 Recursive formulation

I follow the methodology of Kydland and Prescott (1980) and break the Ramsey problem in two subproblems: the problem from period one onward and the initial period problem. For that purpose, let z_t denote debt in marginal utility units, $z_t \equiv U_{ct} b_t$, and rewrite the dynamic implementability constraint (15) as

$$z_t = \Omega_t + \beta E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} z_{t+1}, t \geq 1. \tag{17}$$

It will be useful for later purposes to define $\omega_t \equiv E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} z_{t+1}$. The variable ω_t appears in the right-hand side of the dynamic implementability constraint and is instrumental in the interpretation of the Ramsey plan. Up to a proper scaling with current marginal utility and the subjective discount factor β, it can be thought of as the *market value* of the government portfolio of state-contingent debt, since $\omega_t = \frac{U_{ct}}{\beta} E_t S_{t+1} b_{t+1}$.

I represent the commitment problem from period one onward recursively by keeping track of the exogenous shock g_t *and* debt in marginal utility units z_t, that captures the commitment of the planner to his past promises. Note that debt in marginal utility units is a forward-looking variable that it not inherited from the past. This creates the need to specify $Z(g)$, the space where z lives. The set $Z(g)$ represents the values of debt in marginal utility units that can be *generated* from an implementable allocation when the initial shock is g and is defined in the Appendix. Let $V(z_1, g_1)$ denote the value function of the planner's problem from period one onward, where $z_1 \in Z(g_1)$ and assume that shocks follow a Markov process with transition probabilities $\pi(g'|g)$.

Bellman equation. The functional equation that determines the value function $V(.)$ takes the form

$$V(z, g) = \max_{c,h,z'_{g'}} \Big[(1-\beta) u(c, 1-h)^{1-\rho} + \beta \big[\sum_{g'} \pi(g'|g) V(z'_{g'}, g')^{1-\gamma} \big]^{\frac{1-\rho}{1-\gamma}} \Big]^{\frac{1}{1-\rho}}$$

subject to

$$z = \Omega(c,h) + \beta \sum_{g'} \pi(g'|g) \frac{V(z'_{g'}, g')^{\rho-\gamma}}{\left[\sum_{g'} \pi(g'|g) V(z'_{g'}, g')^{1-\gamma}\right]^{\frac{\rho-\gamma}{1-\gamma}}} z'_{g'} \tag{18}$$

$$c + g = h \tag{19}$$

$$c \geq 0, h \in [0,1] \tag{20}$$

$$z'_{g'} \in Z(g'). \tag{21}$$

The notation $z'_{g'}$ captures the fact that the planner is choosing *state-contingent* debt in marginal utility units next period at shock g'.

The nature of the Ramsey problem is fundamentally changed because, in contrast to the case of time-additive utility, the value function shows up in the dynamic implementability constraint and in particular in the determination of the market value of debt ω. This is due to the fact that continuation values, i.e. the entire profile of future consumption and leisure, determine the stochastic discount factor as we saw earlier. Hence, there is an essential *non-linearity* in z in the constraint.

Initial period problem. The initial value of the forward-looking variable z_1 that was taken as *given* in the formulation of the planner's problem from period one onward is chosen *optimally* in order to maximize the utility of the household at $t = 0$. In this sense, the variable z is a pseudo-state variable, i.e. a jump variable that is treated as a state variable in order to capture the commitment of the planner to the optimal plan devised at the initial period. Furthermore, the problem at the initial period is different from period one onward due to the presence of the initial debt b_0. As a result, the overall value of the Ramsey problem and the initial period policy functions (c_0, h_0, z_1) depend on (b_0, g_0).

4.3 Optimality conditions

It turns out that is easier to derive the optimality conditions of the problem by using the ρ-transformation of the value function, $v(z,g) \equiv \frac{V(z,g)^{1-\rho}-1}{(1-\beta)(1-\rho)}$. The transformed Bellman equation is stated in the Appendix.

Let Φ and λ be the multipliers on the dynamic implementability constraint and the resource constraint respectively of the transformed problem and let $m'_{g'}$ denote the conditional likelihood ratio, which obviously depends on the value function. Note that at the optimal solution the multipliers will be functions of the state, $\Phi = \Phi(z,g)$ and $\lambda = \lambda(z,g)$. The first-order necessary conditions for an interior solution at points of differentiability of the value function are

$$c: \quad U_c + \Phi\Omega_c = \lambda \tag{22}$$

$$h: \quad -U_l + \Phi\Omega_h = -\lambda \tag{23}$$

$$z'_{g'}: \quad v_z(z'_{g'}, g') + \Phi\big[1 + (1-\beta)(\rho-\gamma)v_z(z'_{g'}, g')\eta'_{g'}\big] = 0, \tag{24}$$

where

$$\eta'_{g'} \equiv V'^{\rho-1}_{g'} z'_{g'} - \mu^{\rho-1}\omega. \tag{25}$$

The variables $\Omega_i, i = c, h$ stand for the partial derivatives of Ω with respect to consumption and labor. I call $\eta'_{g'}$ the government's *relative debt position* in marginal utility units and analyze it in detail later. Recall that ω stands for the market value of debt, $\omega = \sum_{g'} \pi(g'|g) m'^{\frac{\rho-\gamma}{1-\gamma}}_{g'} z'_{g'}$, and μ for the certainty equivalent, whereas $V'_{g'}$ is shorthand for $V(z'_{g'}, g')$.[12]

The envelope condition takes the form $v_z(z, g) = -\Phi$. Note that $\Phi \geq 0$, so $v_z(z, g) \leq 0$.[13] The multiplier Φ is strictly positive if the first-best cannot be achieved, i.e. if (z, g) are not such that the first-best allocation can be supported. The initial period optimality conditions that determine (c_0, h_0) and the optimal value of z_1 are stated in the Appendix.

5 Excess burden of distortionary taxation

The entire action with recursive utility is coming from Φ, a multiplier which reflects the shadow cost of the constraints that the competitive equilibrium imposes in the second-best world. In particular, as the envelope condition shows, Φ captures the cost of an additional unit of debt in marginal utility units. It is a cost, because increases in debt have to be accompanied by an increase in distortionary taxation. In a first-best world with lump-sum taxes available, Φ would be *zero*. I refer to it as the *excess burden* of distortionary taxation since it essentially summarizes tax distortions.

5.1 Price effect of continuation value

It order to understand how optimal policy is altered with recursive preferences, it is crucial to understand how continuation values affect the stochastic discount factor, in other words how the

[12] In definition (25) I use the non-transformed value function V (which is equal to $[1 + (1-\beta)(1-\rho)v]^{\frac{1}{1-\rho}}$) as a matter of convenience, since it allows a more compact exposition of the first-order conditions.

[13] I am implicitly assuming that the government has access to lump-sum transfers, so that the dynamic implementability constraint takes the form $z_t \leq \Omega_t + \beta\omega_t$.

"long-run" alters equilibrium pricing. Two mechanisms are at play. Consider the derivative of the stochastic discount factor (14) with respect to V_{t+1},

$$\frac{\partial S_{t+1}}{\partial V_{t+1}} = (\rho - \gamma)\beta \frac{U_{c,t+1}}{U_{ct}} m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} V_{t+1}^{-1}[1 - \pi_{t+1}(g_{t+1}|g^t)m_{t+1}] < 0.$$

An *increase* in continuation value leads to a *decrease* in the stochastic discount factor when the desire to smooth over states is stronger than the desire to smooth over dates, so when there is aversion to volatility in continuation utility and agents prefer early resolution of uncertainty. Therefore, the equilibrium price of the Arrow security *decreases*. To state it differently, an agent who would have high utility at g_{t+1} would require a higher return in order to hold a claim to one unit of consumption at this state of the world.

Note though that states next period are *interconnected* through the certainty equivalent μ_t, which depends positively on continuation values ($\partial\mu_t/\partial V_{t+1} = \pi_{t+1}(g_{t+1}|g^t)(V_{t+1}/\mu_t)^{-\gamma} > 0$). Therefore, an increase in continuation value at state $\bar{g}_{t+1} \neq g_{t+1}$ will affect the stochastic discount factor through the certainty equivalent and therefore the price of an Arrow security at state g_{t+1}. To see that, let $\bar{V}_{t+1} \equiv V_{t+1}(\bar{g}_{t+1}, g^t)$ and compute the derivative of the stochastic discount factor with respect to $\bar{V}_{t+1}$ to get

$$\frac{\partial S_{t+1}}{\partial \bar{V}_{t+1}} = (\gamma - \rho)\beta \frac{U_{c,t+1}}{U_{ct}} m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} \bar{V}_{t+1}^{-1} \pi_{t+1}(\bar{g}_{t+1}|g^t)\bar{m}_{t+1} > 0,$$

where $\bar{m}_{t+1}$ corresponds to the likelihood ratio at $\bar{g}_{t+1}$. Thus, an increase in continuation value at states other than g_{t+1} *increases* the price of an Arrow security at g_{t+1}.

5.2 Debt appreciation and allocation of distortions

Continuation values are negatively related to debt in marginal utility units at the Ramsey problem, $v_z < 0$. Therefore, given our previous discussion, there is a *negative* relationship between the return of an Arrow claim and debt in marginal utility units. Hence, increases in debt in marginal utility units become less costly. This relationship is *absent* for time-additive expected utility.

The negative relationship between debt and returns affects the optimal choice of state-contingent debt and the allocation of distortions over states and dates. In particular, consider the first-order condition with respect to $z'_{g'}$ (24) and rewrite it as follows:

$$\underbrace{-v_z(z'_{g'}, g')}_{\text{MC of increasing } z'_{g'}} = \Phi \cdot \Big[\underbrace{1}_{\text{EU term}} + \underbrace{(1-\beta)(\rho-\gamma)v_z(z'_{g'}, g')\eta'_{g'}}_{\text{EZW term: price effect of } z'_{g'} \text{ (+/-)}} \Big]. \tag{26}$$

The left-hand side of (26) denotes the marginal utility cost of increasing $z'_{g'}$. More debt at g' reduces utility because it is associated with more taxes in the future. The right-hand side of (26) denotes the utility benefit of the government's *marginal revenue* from debt issuance. It is proportional to the change in the market value of the government portfolio $\partial\omega/\partial z'_{g'}$, times its welfare importance, Φ.[14] The first-term in the right-hand side denotes the conventional marginal benefit of increased debt in marginal utility units when its price is not affected by z. Selling more debt for next period increases the revenues of the government and relaxes the government budget constraint. This allows less taxation at the current period. This is the only marginal benefit relevant in the time-additive expected utility world of Lucas and Stokey (1983) where $\rho = \gamma$. In that case, the optimality condition reduces to

$$-v_z(z'_{g'}, g') = \Phi,$$

which – by using the envelope condition – implies that $\Phi'_{g'} = \Phi$ for all values of the state (z, g). Thus, in the case of time-additive expected utility, the planner would optimally make the excess burden of distortionary taxation *constant*. This is the formal result that hides behind the *tax-smoothing* intuition: the policymaker should spread welfare distortions among states and dates in a constant way. Furthermore, this is also the source of Lucas and Stokey's celebrated *history-independence* result, since optimal allocations and tax rates can be written solely as functions of the exogenous shocks and the constant Φ.[15]

In contrast to time-additive expected utility, the price of debt becomes *debt-elastic* with recursive utility. Increases in debt increase its price, which is reflected in the second term at the right-hand side of (26), $(\rho - \gamma)v_z > 0$. The marginal revenue effects of the *appreciation* of the value of debt at g' depend on the relative debt position $\eta'_{g'}$, which has the following property:

Lemma 1. *(Innovation property)*

$$\sum_{g'} \pi(g'|g) m'_{g'} \eta'_{g'} = 0$$

Proof. See Appendix. □

[14] It is useful to think of the planner as minimizing the welfare cost of debt subject to achieving a particular revenue from debt issuance: $\min_{z'_{g'}} C$ subject to $\omega \geq \bar{\omega}$, where $C \equiv -\left[\sum_{g'} \pi(g'|g)\big(1 + (1-\beta)(1-\rho)v(z'_{g'}, g')\big)^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{1-\rho}{1-\gamma}} / (1-\beta)(1-\rho)$. The first-order condition of this problem is the same as (26) for the proper value of the multiplier Φ.

[15] The excess burden of taxation is also constant in a deterministic economy ($\eta'_{g'} \equiv 0, \forall g$). Thus, as far as Φ is concerned, there is no essential difference in the allocation of distortions between a deterministic world and a stochastic but time-additive world.

Therefore, $\eta'_{g'}$ can take both positive and negative values. A positive (negative) $\eta'_{g'}$ can be rewritten as $z'_{g'} > (<) m'^{\frac{1-\rho}{1-\gamma}}_{g'} \omega$, and therefore, corresponds to a debt position that is above (below) a multiple of the market value of the debt portfolio. When $\rho = 1$, we have the simplification $\eta'_{g'} = z'_{g'} - \omega$, where $\omega = \sum_{g'} \pi(g'|g) m'_{g'} z'_{g'}$, so $\eta'_{g'}$ captures the state-contingent debt position in marginal utility units relative to the value of the debt portfolio.[16]

It is natural to expect that the appreciation of debt leads to a higher marginal revenue of the government when the government sells debt ($z'_{g'} > 0$) and a lower one when the government buys assets ($z'_{g'} < 0$). However, due to the state non-separabilities that we analyzed in section 5.1, an increase in debt at g', by reducing the certainty equivalent, reduces equilibrium prices at all other states of the world $\bar{g}' \neq g'$. The marginal revenue effect of the depreciation of the value of debt at all $\bar{g}'$ is captured by the market value of the government portfolio, ω. This is the reason why the net revenue effect of debt is determined by the relative debt position $\eta'_{g'}$ and not just by $z'_{g'}$. Another way to think about the marginal revenue is in terms of *elasticities* of equilibrium prices with respect to debt. The EZW term in (26) captures the *own* elasticity of the price of debt in marginal utility units at g' and the *cross* elasticities of all other equilibrium prices at $\bar{g}' \neq g'$ with respect to $z'_{g'}$.

The elasticity of equilibrium prices with respect to debt modifies the optimal allocation of distortions over states and dates in a non-trivial way. It is convenient to collect terms and use the envelope condition in order to rewrite (26) in terms of the inverse of Φ (assuming that Φ is not zero) as[17]

$$\frac{1}{\Phi'_{g'}} = \frac{1}{\Phi} + (1-\beta)(\rho-\gamma)\eta'_{g'}, \tag{27}$$

or in sequence notation,

$$\frac{1}{\Phi_{t+1}} = \frac{1}{\Phi_t} + (1-\beta)(\rho-\gamma)\eta_{t+1}, t \geq 0, \tag{28}$$

where $\eta_{t+1} = V_{t+1}^{\rho-1} z_{t+1} - \mu_t^{\rho-1}\omega_t$. Consider a state g' where debt in marginal utility units is relatively large so that $\eta'_{g'} > 0$ and a state $\tilde{g}'$ where debt in marginal utility units is relatively small so that $\eta'_{\tilde{g}'} < 0$. Then (26) or, equivalently, (27) imply that $\Phi'_{g'} > \Phi > \Phi'_{\tilde{g}'}$. The reason is simple. The marginal revenue of the government is larger at g' than the marginal revenue at $\tilde{g}'$. As a result, a planner that minimizes welfare costs allocates more distortions at g' and less distortions at $\tilde{g}'$, $\Phi'_{g'} > \Phi'_{\tilde{g}'}$. In other words, relatively large debt positions become less costly, and relatively low

[16] Note that we could see the zero conditional mean property of lemma 1 by rewriting $\eta'_{g'}$ as $\eta'_{g'} = V'^{\rho-1}_{g'} z'_{g'} - \sum_{g'} \pi(g'|g) m'_{g'} V'^{\rho-1}_{g'} z'_{g'}$, so $\eta'_{g'}$ could be interpreted as the conditional innovation of $V'^{\rho-1}_{g'} z'_{g'}$ under $\pi_t \cdot M_t$.

[17] Otherwise, write the optimality condition in sequence notation as $\Phi_{t+1} = \Phi_t / \left[1 + (1-\beta)(\rho-\gamma)\eta_{t+1}\Phi_t\right]$. Thus, if $\Phi_t = 0$, then $\Phi_{t+i} = 0, i \geq 0$, so the first-best is an absorbing state.

debt positions more profitable, prompting the planner to increase distortions at high-debt states ($\eta > 0$) and decrease distortions at low-debt states ($\eta < 0$). Over time, we see that since $\eta'_{g'} > 0$, the excess burden of taxation increases with respect to the current one, $\Phi'_{g'} > \Phi$, so the planner taxes less today and postpones distortions for next period because debt is cheap. The opposite happens for $\eta'_{\tilde{g}'} < 0$.[18] It is worth summarizing the results about the excess burden of taxation in a proposition.[19]

Proposition 2. *In a time-additive expected utility world the excess burden of taxation is constant across states and dates. In contrast, in a recursive utility world, the planner allocates more distortions on high-debt states ($\eta > 0$) and less distortions on low-debt states ($\eta < 0$). Tax distortions increase or decrease over time depending on the realization of a high- or low-debt state respectively.*

Note that the mechanism behind the allocation of distortions across *states* is essentially the same as in the special setup of section 2. In section 2 it was possible to write utilities in terms of the tax rate, making prices *tax-elastic*. In a dynamic setup where marginal utility is present, this role is played by z. Distortions, that were captured by the tax rate in the static setup, are reflected in the excess burden of taxation. Furthermore, the designation of the states that require higher or less distortions is determined by debt in marginal utility units and not just debt.[20]

Fiscal hedging. Proposition 2 is instructing us that the crucial element for the allocation of distortions with recursive utility is the relative debt position in marginal utility units η_{t+1}. Besides that, it is silent on how debt in marginal utility units, i.e. the present discounted value of future surpluses (in marginal utility units), is associated with fiscal shocks. To answer this question, we have to understand the *fiscal hedging* of the government, i.e. the way the government is using state-contingent debt in order to insure against fiscal shocks. We typically expect that the planner insures against government expenditure shocks by taking small debt positions (or assets) when government expenditures are high (which allows running subsequently a deficit), and issuing debt for low expenditure shocks, that is paid back by surpluses. This type of fiscal insurance leads to state contingent-debt that is typically negatively correlated with government expenditures.[21] If

[18]The non-constant allocation of distortions has also implications for the size of the debt position z_t over time. It is tempting to deduce that the planner is not only allocating more distortions on a high-debt state next period ($\eta_{t+1} > 0$), but also takes a larger debt position next period than the current one. Formally, the deduction would be $\Phi_{t+1} = -v_z(z_{t+1}, g) > \Phi_t = -v_z(z_t, g) \Rightarrow z_{t+1} > z_t$. This is a statement about the concavity of the value function at g. This statement cannot be made in general due to the non-convexities of the Ramsey problem, but it turns out to be numerically true. See for further details section 7.

[19]The direction of the results is reversed when there is preference for late resolution of uncertainty ($\rho > \gamma$), which implies a *love* of volatility in future utility, a case which doesn't seem to be empirically relevant (at least from an asset-pricing perspective). In this case, increases in continuation values lead to increases in the prices of state-contingent claims and therefore to a reduction of state-contingent returns. Hence, there is a positive relationship of debt and state-contingent returns. As a result, relatively large debt positions become more costly with recursive utility, motivating the planner to allocate less distortions on high-debt states and more distortions on low-debt states, so $\Phi'_{g'} < \Phi < \Phi'_{\tilde{g}'}$ when $\eta'_{g'} > 0$ and $\eta'_{\tilde{g}'} < 0$.

[20]The relative debt position η was not relevant in the static setup because ω was zero.

[21]This was exactly the case in section 2.

this negative correlation with fiscal shocks carries over to state-contingent debt in *marginal utility* units, then proposition 2 implies that distortions will increase for low government expenditures shocks and decrease for high government expenditure shocks. In that case, the change in the excess burden of taxation is *negatively* correlated with government expenditures. We will analyze in detail a utility function in the numerical illustrations section that delivers this result.

Recursive versus sequential formulation. Readers accustomed to optimal taxation problems with complete markets may wonder how the excess burden of taxation can be *time-varying* when there is a *unique* intertemporal budget constraint. In the Appendix I employ a sequential formulation of the problem and show the mapping between the optimality conditions of the two formulations in order to make clear where this result is coming from. In short, the time-varying Φ_t in the recursive formulation reflects the shadow value of additional "implementability" constraints in the sequential formulation of the problem that arise even in a complete markets setup.[22] The benefit of the recursive formulation of the commitment problem, besides illuminating obviously that z is the relevant state variable, is the succinct summary of the effects of continuation values in terms of a varying cost of debt. This allows a clean comparison with the time-additive expected utility case. There are obvious similarities in spirit with the optimal risk-sharing literature with recursive preferences, which expresses risk-sharing arrangements in terms of time-varying Pareto weights.[23,24]

5.3 Dynamics of the excess burden of taxation

The relative debt position η_t captures the incentives of the planner to increase or decrease distortions, given the excess burden of taxation of the previous period and therefore becomes the determinant of the *conditional* time-variation of Φ_t. The law of motion (28) shows that the inverse of the excess burden of taxation at t depends on the sum of all past positions $\eta_i, i = 1, ..., t$, a property which is explained by the fact that *all* past prices of state-contingent claims change with a change in continuation values at time t. Hence, cumulative returns change. This is due to the forward-looking nature of continuation utilities, i.e. the household at $t - i$ is taking into account the entire future stream of consumption and leisure when it prices Arrow claims. Thus, the excess

[22]These constraints describe utility recursions and the law of motion of $M_t^{\frac{\rho-\gamma}{1-\gamma}}$. In the case of the multiplier preferences of Hansen and Sargent (2001), it is natural to think of the utility recursions as implementability constraints since they correspond to optimality conditions of the malevolent alter-ego of the household, that minimizes the household's utility subject to a penalty. See Karantounias (2013). This minimization procedure would also emerge naturally if we expressed recursive utility as the variational utility of Geoffard (1996).

[23]See for example Anderson (2005) and references therein.

[24] Note also that recursive utility adds z as a state variable to the optimal taxation problem, whereas z can be ignored in the time-additive case. The reason is that z is necessary for the determination of the Ramsey plan only though its shadow cost, Φ. When the excess burden of taxation is constant, the return function of the second-best problem can be augmented in such a way, so that z becomes redundant as a state variable. See Lucas and Stokey (1983) or Zhu (1992) and Chari et al. (1994).

burden of taxation and therefore the allocation depend on the *past*. Furthermore, we have:

Proposition 3. *(Martingale characterization and back-loading of distortions)*

The inverse of Φ_t is a martingale with respect to the continuation-value adjusted measure $\pi_t \cdot M_t$, and therefore, Φ_t is a submartingale with respect to $\pi_t \cdot M_t$.

Proof. Take conditional expectation in (28) to get

$$E_t m_{t+1} \frac{1}{\Phi_{t+1}} = \frac{1}{\Phi_t} E_t m_{t+1} + (1-\beta)(\rho-\gamma) E_t m_{t+1} \eta_{t+1} = \frac{1}{\Phi_t},$$

since $E_t m_{t+1} = 1$ and $E_t m_{t+1} \eta_{t+1} = 0$ by lemma 1. Thus $1/\Phi_t$ is a martingale with respect to $\pi_t \cdot M_t$. Furthermore, since the function $f(x) = 1/x$ is convex for $x > 0$, an application of the conditional version of Jensen's inequality leads to $E_t m_{t+1} \frac{1}{x_{t+1}} \geq \frac{1}{E_t m_{t+1} x_{t+1}}$. Set now $x_t = 1/\Phi_t$ and use the martingale result to finally get $E_t m_{t+1} \Phi_{t+1} \geq \Phi_t$. □

The martingale result about the inverse of the excess burden of taxation can be interpreted loosely as an indication of *persistence*. For example, if there is an *absorbing* state, then η_{t+1} becomes identically zero after the absorbing state is reached, and therefore Φ_t stays *permanently* at the level that it reaches when the absorbing state is hit.

More generally, the asymptotic behavior of distortions is an open issue. The fact that $1/\Phi_t$ is a martingale with respect to the continuation-value adjusted measure implies that M_t/Φ_t is a martingale with respect to the physical measure π, since $E_t(M_{t+1}/\Phi_{t+1}) = M_t E_t m_{t+1}(1/\Phi_{t+1}) = M_t/\Phi_t$, and therefore the non-negative ratio M_t/Φ_t converges almost surely to a finite random variable by the martingale convergence theorem.[25] However, we cannot make a general claim about almost sure convergence of $1/\Phi_t$, unless we restrict the analysis to the case of an absorbing state. In particular, since M_t is by construction a non-negative martingale with respect to π, it converges to the non-negative random variable M_∞ almost surely. If $M_\infty > 0$, then we can infer that $1/\Phi_t$ converges almost surely. However, the martingale M_t typically converges to zero, so we cannot make this claim.[26,27]

[25] Whenever I use almost surely, I refer to the physical measure π.

[26] The same issue shows up in the analysis of optimal taxation with incomplete markets and time-additive expected utility of Aiyagari et al. (2002). They find that when debt and asset limits do *not* bind, the excess burden of taxation (and not the *inverse* of the excess burden of taxation as in the current analysis) is a martingale with respect to the *risk-adjusted* measure, so actual convergence of the excess burden of taxation hinges on the convergence of the risk-adjusted measure to a positive random variable. In the special case of quasi-linear utility, they are able to eliminate the presence of the risk-adjusted measure and actually show convergence of Φ_t to zero, i.e. the Ramsey allocation converges to the first-best allocation. This type of simplification is not possible in the recursive utility case.

[27] Using the same logic as Aiyagari et al. (2002) did for the risk-adjusted measure, it is easy to show that if $M_\infty(\tilde{\omega}) > 0$ for a sample path $\tilde{\omega}$, then the increment has to converge to unity, $m_t(\tilde{\omega}) \to 1$, so $V_t^{1-\gamma}(\tilde{\omega}) \to E_{t-1} V_t^{1-\gamma}(\tilde{\omega})$. The logic is simple: $\ln M_t(\tilde{\omega}) = \sum_{i=1}^t \ln m_i(\tilde{\omega}) \to \ln M_\infty(\tilde{\omega}) > -\infty$ and therefore $\ln m_t(\tilde{\omega}) \to 0$. Thus, we can infer that if $Prob(\tilde{\omega} | m_t(\tilde{\omega}) \to 1) = 0$, then $M_\infty = 0$ almost surely (otherwise $m_t \to 1$ on a set of positive measure). Actually, this result can be strengthened to the following statement: if it is *not* the case that

The submartingale result of proposition 3 shows that the planner wants on "average" to *back-load* distortions, in the sense that distortions exhibit a *positive* drift with respect to the continuation-value adjusted measure. However, it is not clear if there is back-loading of distortions with respect to the physical measure. In particular, note that $\text{Cov}_t(m_{t+1}, \Phi_{t+1}) = E_t m_{t+1}\Phi_{t+1} - E_t\Phi_{t+1}$ (since $E_t m_{t+1} = 1$). Use the submartingale result $E_t m_{t+1}\Phi_{t+1} \geq \Phi_t$ to get

$$E_t\Phi_{t+1} \geq \Phi_t - \text{Cov}_t(m_{t+1}, \Phi_{t+1}).$$

The sign of the conditional covariance of the excess burden of taxation with the increment to the continuation-value adjusted measure m_{t+1} depends on the *fiscal hedging* of the government, which determines how distortions are allocated across shocks. To see that, consider without loss of generality the $\rho = 1 < \gamma$ case (or more generally risk-sensitive preferences). Conditional on Φ_t, we saw earlier that debt in marginal utility units that is negatively correlated with government expenditure shocks, leads to a respective negative correlation of Φ_{t+1} with fiscal shocks. But high government expenditure shocks, since they provide low utility, are associated with a higher conditional probability mass and therefore a higher $m_{t+1} = \exp[(1-\beta)(1-\gamma)v_{t+1}]/E_t \exp[(1-\beta)(1-\gamma)v_{t+1}]$, due to the household's aversion to utility volatility. Therefore, we may expect the conditional covariance of m_{t+1} with Φ_{t+1} to be *negative*. In that case, we have a back-loading of distortions with respect to the *physical* measure, $E_t\Phi_{t+1} \geq \Phi_t$. We will further explore the persistence, the drift and the convergence properties of distortions in the numerical illustrations section.

6 Optimal labor income taxation

The excess burden of taxation is the relevant statistic for the allocation of distortions over states and dates. In order to see its exact relationship with the labor tax, eliminate λ and combine the first-order conditions with respect to consumption and labor (22)-(23) to get the optimal wedge in labor supply

$$\frac{U_l}{U_c} \cdot \frac{1 - \Phi\frac{\Omega_h}{U_l}}{1 + \Phi\frac{\Omega_c}{U_c}} = 1. \tag{29}$$

Expressing the terms Ω_c/U_c and Ω_h/U_l in terms of elasticities delivers

$m_t \to 1$ almost surely, then $M_\infty = 0$ almost surely. The proof of this is coming from the work of Ian Martin who generalized the Kakutani theorem on multiplicative martingales. See Martin (2012, Theorem 1). As a result, as long as there is some positive probability that there is variation in continuation values at the limit so that $m_{t+1} \nrightarrow 1$, we run into the case of $M_\infty = 0$.

$$\frac{\Omega_c}{U_c} = 1 - \epsilon_{cc} - \epsilon_{ch} \tag{30}$$

$$\frac{\Omega_h}{U_l} = -1 - \epsilon_{hh} - \epsilon_{hc}, \tag{31}$$

where $\epsilon_{cc} \equiv -U_{cc}c/U_c > 0$ and $\epsilon_{ch} \equiv U_{cl}h/U_c$, i.e. the *own* and *cross* elasticity of the period marginal utility of consumption, and $\epsilon_{hh} \equiv -U_{ll}h/U_l > 0$ and $\epsilon_{hc} \equiv U_{lc}c/U_l$, the *own* and *cross* elasticity of the period marginal disutility of labor.[28] As a result, we get the following formula for the optimal labor tax,

Proposition 4. *(Labor tax) The optimal labor tax is*

$$\tau_t = \Phi_t \frac{\epsilon_{cc,t} + \epsilon_{ch,t} + \epsilon_{hh,t} + \epsilon_{hc,t}}{1 + \Phi_t\big(1 + \epsilon_{hh,t} + \epsilon_{hc,t}\big)}, \quad t \geq 1.$$

Proof. Use the labor supply condition $U_l/U_c = 1 - \tau$ in order to express (29) in terms of the labor tax as $\tau = -\Phi(\Omega_c/U_c + \Omega_h/U_l)/(1 - \Phi\Omega_h/U_l)$. Use now the elasticity formulas (30) and (31) to get the result.[29,30] □

The formula in proposition 4 expresses the optimal labor tax in terms of the excess burden of taxation Φ_t and in terms of elasticities of the period marginal utility of consumption and disutility of labor. When $U_{cl} \geq 0$, the cross elasticities become non-negative ($\epsilon_{ch}, \epsilon_{hc} \geq 0$) and the labor tax is positive.

In the time-additive case the excess burden of taxation is constant over states and dates, so there is variation in the tax rate *only* as long as there is variation in the period elasticities. Period elasticities capture the sensitivity of the marginal rate of substitution of consumption and leisure and the marginal utility channel in the stochastic discount factor. When these elasticities are constant, then the optimal tax rate is *constant* and optimal policy prescribes *perfect* tax-smoothing.

[28] The elasticities of the marginal utility of consumption are multiplied with minus unity.

[29] The optimal labor tax at $t = 0$ is different due the presence of initial debt,

$$\tau_0 = \Phi_0 \frac{\epsilon_{cc} + \epsilon_{ch} + \epsilon_{hh} + \epsilon_{hc} - (\epsilon_{cc} + \epsilon_{hc})c_0^{-1}b_0}{1 + \Phi_0\left(1 + \epsilon_{hh} + \epsilon_{hc} - \epsilon_{hc}c_0^{-1}b_0\right)}.$$

The respective elasticities are evaluated at the initial allocation (c_0, h_0).

[30] The labor tax formula holds also for the deterministic and stochastic time-additive case for any period utility U that satisfies the standard monotonicity and concavity assumptions, i.e. without being restricted to $U = (u^{1-\rho} - 1)/(1-\rho), u > 0$. It also holds for the risk-sensitive preferences with parameter $\sigma < 0$ and any standard U, i.e. without being confined to treat the risk-sensitive preferences as a subcase of EZW utility for $\rho = 1$. With risk-sensitive preferences, conditional likelihood ratios read $m_{t+1} = \exp(\sigma v_{t+1})/E_t \exp(\sigma v_{t+1})$ and the law of motion of Φ_t remains the same by replacing $(1-\beta)(\rho-\gamma)$ with σ in (28), with a relative debt position $\eta_{t+1} = z_{t+1} - E_t m_{t+1} z_{t+1}$.

Even when period elasticities are not constant, the ensuing labor tax volatility is typically small.[31] Furthermore, the labor tax becomes a function only of the current shock g, due to the history-independence property of the Lucas and Stokey (1983) plan, so it inherits the stochastic properties of the exogenous shocks.

In contrast, with recursive utility, even in the constant elasticity case, the labor tax varies *monotonically* with the non-constant excess burden of taxation.[32] This affects both the volatility of the tax rate at t but also introduces dependence on the past through the law of motion (28), as analyzed in section 5.[33] As a result, when period elasticities are constant, the tax rate follows the behavior of the excess burden of taxation summarized in proposition 2, reflecting the efforts of the planner to use the negative covariance of debt and returns for a welfare-improving allocation of distortions.

6.1 Power utility in consumption and constant Frisch elasticity

Up to now, I have not taken a stance on the composite good u. Consider

$$u(c, 1-h) = \left[c^{1-\rho} - (1-\rho)a_h \frac{h^{1+\phi_h}}{1+\phi_h}\right]^{\frac{1}{1-\rho}} \tag{32}$$

that delivers a period utility function $U = \frac{c^{1-\rho}-1}{1-\rho} - a_h \frac{h^{1+\phi_h}}{1+\phi_h}$ with a Frisch elasticity $1/\phi_h$ and elasticities $\epsilon_{cc} = \rho, \epsilon_{hh} = \phi_h, \epsilon_{ch} = \epsilon_{hc} = 0$.[34] I am particularly interested in this utility function because it is an example of the constant elasticity class, which delivers perfect tax-smoothing in the time-additive case. The labor tax formula in proposition 4 specializes to

$$\tau_t = \frac{\Phi_t(\rho + \phi_h)}{1 + \Phi_t(1+\phi_h)}, t \geq 1, \tag{33}$$

The formula shows that the crucial parameter for the elasticities channel is ρ (and *not* the risk aversion parameter γ), whereas both ρ and γ affect the Ramsey outcome through Φ_t. Furthermore, we have:

[31] See for example Chari and Kehoe (1999).

[32] We have $\frac{\partial \tau}{\partial \Phi} = \frac{\epsilon_{cc}+\epsilon_{ch}+\epsilon_{hh}+\epsilon_{hc}}{\left[1+\Phi(1+\epsilon_{hh}+\epsilon_{hc})\right]^2} > 0$, as long as the numerator is positive. $U_{cl} \geq 0$ is sufficient for that.

[33] Note furthermore, that even if we restricted attention to the fluctuation of the tax rate at t, the response of the tax rate to shocks due to variation in period elasticities is typically opposite than the response due to the pricing of the long-run, captured by Φ. For example, in cases with isoelastic utility in consumption, separability between consumption and leisure and varying Frisch elasticity, the tax rate increases for high g, due to lower Frisch elasticity when labor is high. Instead, we typically expect Φ and therefore the tax rate to decrease for high g, in order to decrease the price of assets that the planner is typically buying in order to insure against adverse shocks.

[34] It is assumed that parameters are such so that $c^{1-\rho} - (1-\rho)a_h \frac{h^{1+\phi_h}}{1+\phi_h} > 0$, so that $u > 0$ is well defined. For $\rho = 1$, the utility recursion becomes $V_t = \exp\left[(1-\beta)\left(\ln c - a_h \frac{h^{1+\phi_h}}{1+\phi_h}\right) + \beta \ln \mu_t\right]$. If we want to drop the non-negativity restriction, we can just consider risk-sensitive preferences with the particular period utility U.

Proposition 5. *(Labor tax with constant Frisch elasticity)*

1. *The labor tax follows the law of motion*

$$\frac{1}{\tau_{t+1}} = \frac{1}{\tau_t} + \frac{(1-\beta)(\rho-\gamma)}{\rho+\phi_h}\eta_{t+1}, t \geq 1. \tag{34}$$

2. *(Allocation of distortions) Let $\rho < \gamma$. Then*

   - *if $\eta_{t+1} > 0$, then $\tau_{t+1} > \tau_t$ (because Φ_{t+1} increases)*
   - *if $\eta_{t+1} < 0$, then $\tau_{t+1} < \tau_t$ (because Φ_{t+1} decreases)*

3. *(Martingale characterization and back-loading) The inverse of the labor tax $1/\tau_t$ is a martingale with respect to the measure $\pi_t \cdot M_t$ and therefore, τ_t is a submartingale with respect to $\pi_t \cdot M_t$. Furthermore,*

$$E_t\tau_{t+1} \geq \tau_t - Cov_t(m_{t+1}, \tau_{t+1}),$$

   so if $Cov_t(m_{t+1}, \tau_{t+1}) < 0$, then τ_t is a submartingale with respect to the physical measure, $E_t\tau_{t+1} \geq \tau_t$.

Proof. Taking inverses in (33) delivers

$$\frac{1}{\tau_t} = \frac{1+\phi_h}{\rho+\phi_h} + \frac{1}{\rho+\phi_h}\frac{1}{\Phi_t}.$$

Note that $1/\tau_t$ is an *affine* function of $1/\Phi_t$. Use the law of motion of Φ_t in (28) to write the law of motion of the labor tax as in (34). Notice the close resemblance of the law of motion of the labor tax (34) to the law of motion of the excess burden of distortionary taxation (28), a fact that leads to similar conclusions about the allocation of distortions and martingale-like properties as in proposition 3 and the discussion thereafter. □

Since in the constant elasticity case the labor tax becomes a monotonic function of the excess burden of taxation, the interpretation of Φ as a measure of distortions becomes *exact*. The law of motion of the tax rate (34) is essentially the same as the law of motion of Φ_t. The entire analysis in section 5 about the allocation of distortions across states and dates and the discussion about fiscal hedging incentives, martingale and back-loading properties that was conducted in terms of the excess burden of taxation can be recast in terms of the tax rate and will not be repeated.

7 Numerical illustrations

The evolution of the excess burden and therefore the analytic formulas for taxes presented in earlier sections hinge on the sign and the size of the relative debt positions η_{t+1}, i.e. on the government's

fiscal insurance. Debt in marginal utility units is obviously an endogenous object and requires a numerical analysis to determine its correlation with exogenous shocks. In this section I provide various numerical exercices in order to highlight properties of the optimal plan.

A summary of the results is as follows. First, the government is insuring against fiscal shocks with relatively low debt positions in marginal utility units for adverse shocks and relatively high debt positions in marginal utility units for favorable shocks. With this type of fiscal hedging, the change in tax rates is strongly *negatively* correlated with government expenditures, with correlation of the order of -0.99. As a result, the change in tax rates is strongly *countercyclical*, with a correlation of -0.9762. Second, the debt positions of the government are in absolute value larger than in the expected utility case, a property that I call "over-insurance". Third, the martingale property of the inverse of the excess burden of taxation translates to highly *persistent* tax rates. The autocorrelation of the tax rate is of the order of 0.99, even when the shocks are i.i.d. Fourth, the tax rate has a *positive* drift with respect to the physical measure. The mean and the standard deviation of the tax rate are increasing over time, so there is on average a *back-loading* of tax distortions and a "fanning-out" of the distribution. Depending on the time horizon, the volatility of the tax rate can be substantial. A similar pattern arises for the debt-to-output ratio. Finally, in the very long-run, the stationary distributions of the tax rate and the debt-to-output ratio exhibit a high mean and substantial volatility.

7.1 Calibration

I use the utility function of proposition 5 that delivers the perfect tax-smoothing result for time-additive expected utility and the martingale-like result for recursive utility. In particular, let $\rho = 1$ and consider the utility recursion

$$v_t = \ln c_t - a_h \frac{h_t^{1+\phi_h}}{1+\phi_h} + \frac{\beta}{(1-\beta)(1-\gamma)} \ln E_t \exp\big((1-\beta)(1-\gamma)v_{t+1}\big), \tag{35}$$

where $\gamma > 1$. The endowment of time is normalized to unity. I assume an annual frequency and unitary Frisch elasticity and set $(\beta, \phi_h) = (0.96, 1)$. The atemporal risk aversion is $\gamma = 10$.[35] I assume that shocks are i.i.d. in order to focus on the persistence generated *endogenously* by optimal policy. Expenditures shocks take two values, $g_L = 0.072$ and $g_H = 0.088$, with probability $\pi = 0.5$. These values correspond to 18% and 22% of average first-best output respectively. So

[35] The range of the risk-aversion parameter varies wildly in studies that try to match asset-pricing facts. For example, Tallarini (2000) uses a risk aversion parameter above 50 in order to generate a high market price of risk, whereas Bansal and Yaron (2004) use low values of risk aversion in environments with long-run risks and stochastic volatility. Note that the plausibility of the size of atemporal risk aversion cannot be judged independently from the stochastic processes that drive uncertainty in the economy, since they jointly bear implications for the premium for early resolution of uncertainty. See Epstein et al. (2013) for a thoughtful evaluation of calibration practices in the asset-pricing literature from this angle.

the standard deviation of the share of government spending in average first-best output is small and equals 2%. The labor disutility parameter a_h is set so that the household works 40% of its available time if we are the first-best and government expenditures take their average value. Initial debt is zero and the initial realization of the government expenditure shock is low, $g_0 = g_L$.

Recursive preferences introduce several complications to the calculation of the optimal plan, making the numerical analysis non-trivial. At first, the presence of the value function in the constraint makes convergence of iterative procedures difficult. Moreover, the state space where z lives is endogenous. Furthermore, the problem is non-convex, which requires caution with the use of first-order conditions.[36] In addition, a precise calculation of the slope of the value function is necessary, since it determines the excess burden of taxation. Finally, standard perturbation methods are not helpful even for small shocks, because the excess burden of taxation becomes to first-order a random walk, introducing explosiveness to the solution. In the Appendix I provide details of the numerical method that was used.

7.2 Expected utility plan

The expected utility case of $\gamma = 1$ corresponds to the environment of Lucas and Stokey (1983). The Ramsey plan is history-independent and, as we saw earlier, the tax rate is constant for the particular utility function. The planner issues zero debt for the case of low shocks, $b_L = 0$, and insures against a high government expenditure shock by buying *assets*, $b_H < 0$. More specifically, whenever there is a low shock, the planner, who has no debt to pay to the agent since $b_L = 0$, runs a surplus $\tau h_L - g_L > 0$ and uses the surplus to buy assets for the contingency of a high government expenditure shock. The amount of assets is equal to $b_H = (\tau h_H - g_H)/(1 - \beta\pi)$. When the shock is high, the planner uses the interest income on these assets to finance the deficit $\tau h_H - g_H < 0$.[37]

7.3 Fiscal hedging and over-insurance

Turning to the recursive utility case, the left panel in figure 1 plots the difference between the policy function for debt in marginal utility units next period when the shock is low, z'_L, and the policy function when the shock is high, z'_H, as functions of current debt in marginal utility units. From the figure we see that $z'_L > z'_H$ for each point of the state space, so the government hedges fiscal shocks by issuing more debt in marginal utility units for the low shock and less debt for the high shock. As a result, the relative debt positions are $\eta'_H = z'_H - \omega < 0$ and $\eta'_L = z'_L - \omega > 0$,

[36] This is not surprising since this is typically the case in Ramsey problems. However, there is an additional layer of non-convexity that is coming from the non-linearity of the market value of government debt in z. Actually, for the particular utility function, the traditional non-convexity associated with the surplus in marginal utility units Ω is absent and the non-convexities emerge only with recursive utility.

[37] Note that if the initial shock was high, $g_0 = g_H$, we would have $b_H = 0$ and $b_L > 0$. The planner insures against adverse shocks by running a deficit when government expenditures are high, that are financed by debt contingent on a low expenditure shock. When shocks are low, the planner runs a surplus to pay back the issued debt.

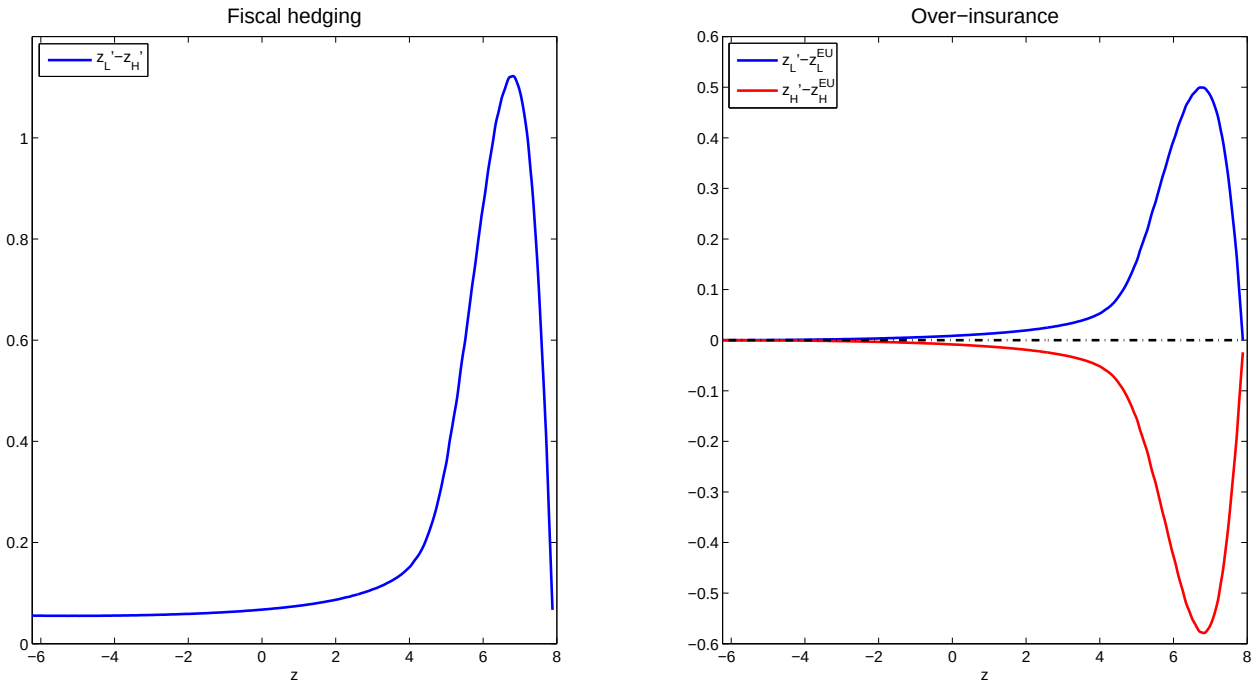

Figure 1: The left panel depicts the difference $z'_L - z'_H$. The right panel compares the position in the recursive utility case with the respective position in the expected utility case. For both graphs the current expenditure shock is low. A similar picture emerges when the current shock is high.

so the planner is allocating more distortions on the low shocks and less distortions on the high shocks, $\Phi'_L > \Phi'_H$ and $\tau'_L > \tau'_H$. Moreover, the right panel in figure 1 plots the difference in the policy functions in the recursive utility and the expected utility case, $z'_i - z_i^{\text{EU}}$, for $i = L, H$, for each value of current debt in marginal utility units. As seen from the graph, in the recursive utility case the planner is issuing more debt in marginal utility units than he would in the time-additive economy for the low shock and less debt that in the time-additive economy for the high shock. So there is a sense of "over-insurance."[38] Not only does the planner allocate more distortions on low shocks due to to the appreciation of the value of debt, but also issues more debt at these states in order to take advantage of the valuation effect. The opposite holds for high shocks. Thus, the positions become larger in absolute value than in the time-additive economy.

7.4 An instructive sample path

In order to understand the dynamics of the solution, consider at $t = 1$ a sample path of 10 low shocks, followed by a sequence of shocks that alternates between 15 high and 15 low shocks. Figure 2 plots the respective sample paths for the tax rate, consumption, the debt position in marginal

[38] A virtually identical graph would emerge if we compared the optimal policy functions z'_i with the positions that would be induced if the planner followed a sub-optimal policy of a constant Φ, i.e. if the planner ignored optimal policy prescriptions and just followed a constant-tax policy in the spirit of section 2.

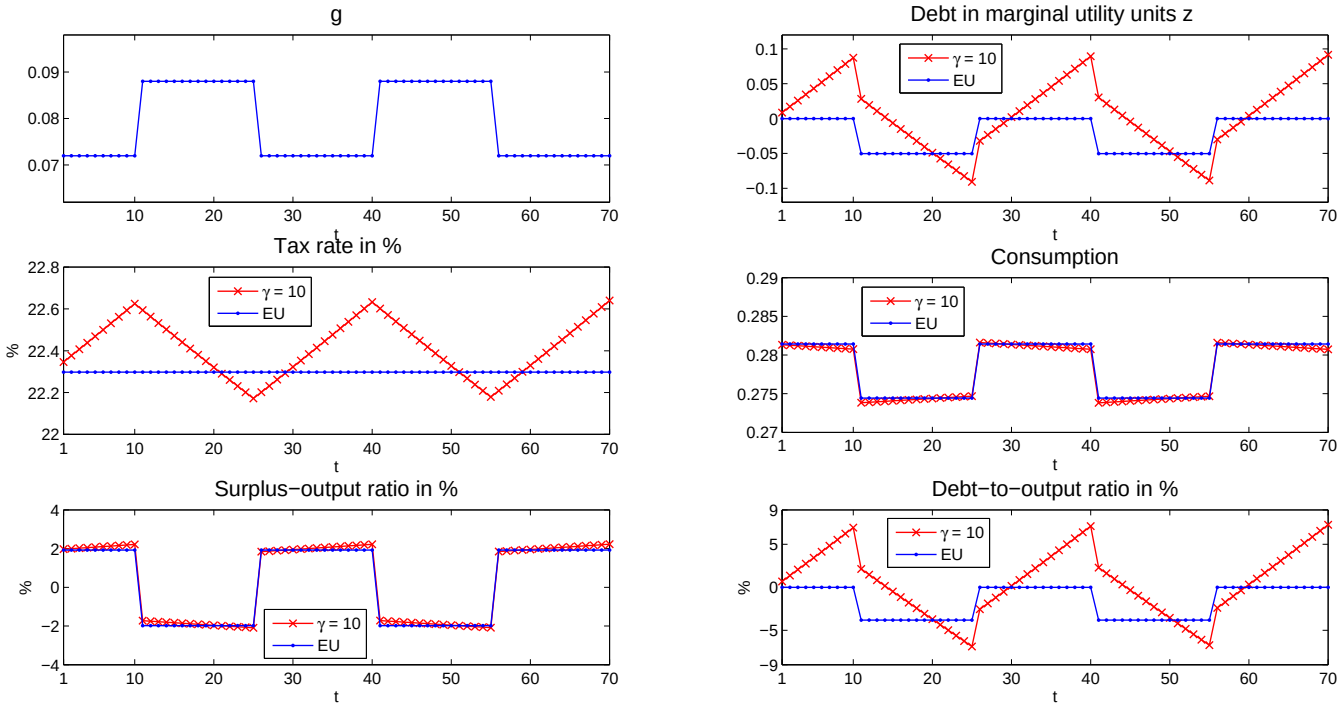

Figure 2: Sample paths for the alternating sequence of low and high shocks.

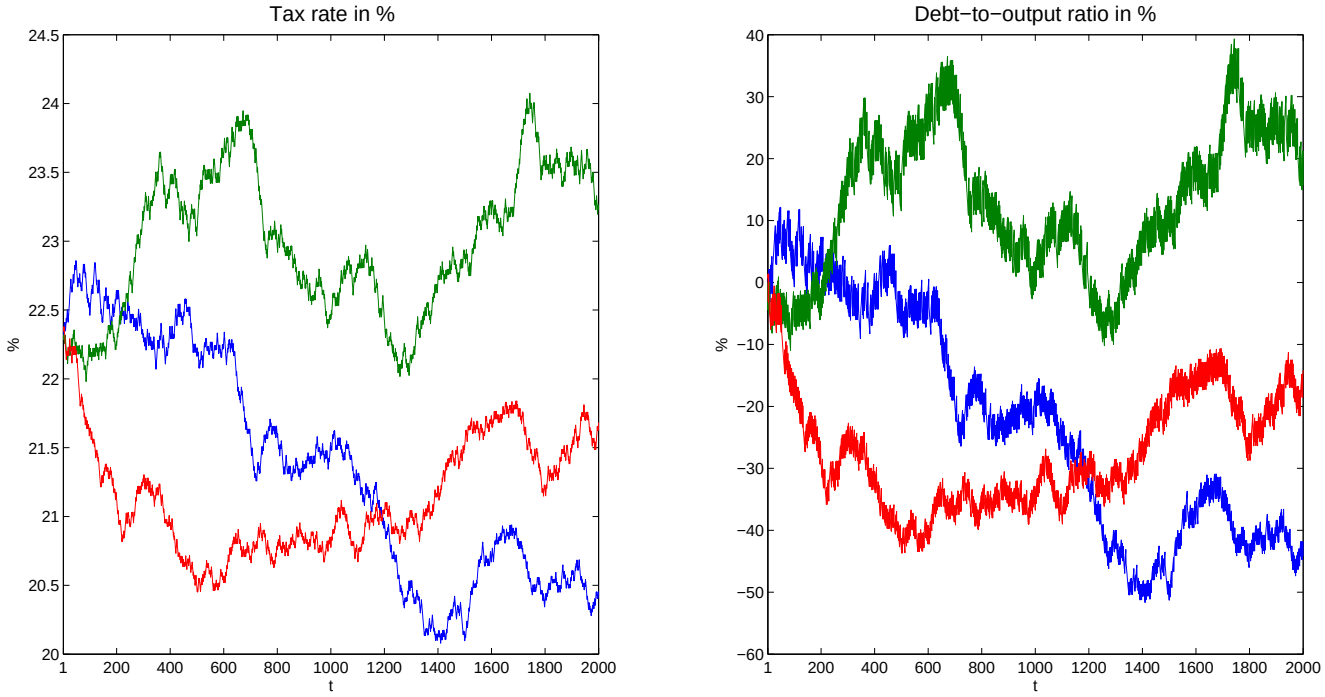

Figure 3: Random sample paths of the tax rate and the corresponding debt-to-output ratio.

utility units z, and the surplus and debt as shares of output in the expected and recursive utility case.

The planner is issuing every period claims to consumption for next period for the contingency of a low and high shock. As we saw earlier, he always takes a relatively larger position z when shocks are low and a lower position when shocks are high. Consequently, at each period that the shock remains low, the change in the tax rate is positive and the tax rate is increasing over time till the first switch. The debt position in marginal utility units is also increasing over time till the first switch, which translates to an increasing debt-to-output ratio.[39] When the shock switches to the high value the opposite pattern emerges. The government, which allocates less distortions on high shocks, starts reducing the tax rate over time. Note that the tax rate does not jump down but is slowly reduced from the highest level that it assumed at the last period when the shock was low, which is an indication of its persistence. Debt in marginal utility units drops when the shock becomes high and then starts to decrease slowly reflecting the decrease of the tax rate. The opposite pattern emerges again when we switch to the low shock. Remember that in the expected utility case the tax rate would stay constant and that debt in marginal utility units would just assume a zero value for the low shocks and a negative value (so it would stand for government assets) for the high shocks.[40]

7.5 Volatility and back-loading of distortions

Figure 3 plots some realizations of the tax rate and the corresponding debt-to-output ratios. Note the persistence in the tax rate and its volatility. To understand better the properties of the Ramsey plan, I simulate $10,000$ sample paths that are $2,000$ periods long. Figure 4 depicts the ensemble moments of interest of the tax rate and the debt-to-output ratio and table 1 reports their particular values. In the expected utility case the tax rate is 22.3% with zero standard deviation. The table depicts that there is a *positive* drift in the tax rate, which is to be expected given the discussion after proposition 3 about the role of fiscal hedging and its effect on back-loading of distortions. The mean tax rate increases over time from 22.38% at $t = 200$ to 22.93% at $t = 2000$. The increase in the mean is slow. However, the volatility of the tax rate is large and increasing over time. The standard deviation of the tax rate rises from 0.43 percentage points at $t = 200$ to 1.48 percentage points at $t = 2000$. So the distribution of the tax rate is "fanning-out" over time.[41] Another

[39] The increase in the debt position in marginal utility units over time is an outcome of the numerical finding that the value functions are concave in z for each shock, and therefore the absolute value of the slope, Φ_t, is increasing in z.

[40] Even if I used a period utility function that would imply a fluctuating tax rate in the expected utility case (for a example a utility function with time-varying Frisch elasticity), the tax rate would not change over time unless there was a switch in the shocks. This is due to the history-independence property.

[41] Although section 7 is meant only for illustrative purposes, it is worth noting that the standard deviation figures are quite large in comparison to the ones obtained in quantitative studies in time-additive economies with fluctuating tax rates. For example, Chari et al. (1994) use preferences that feature time-varying Frisch elasticity and show that the standard deviation of the tax rate typically ranges from 5 to 15 basis points across different calibrations.

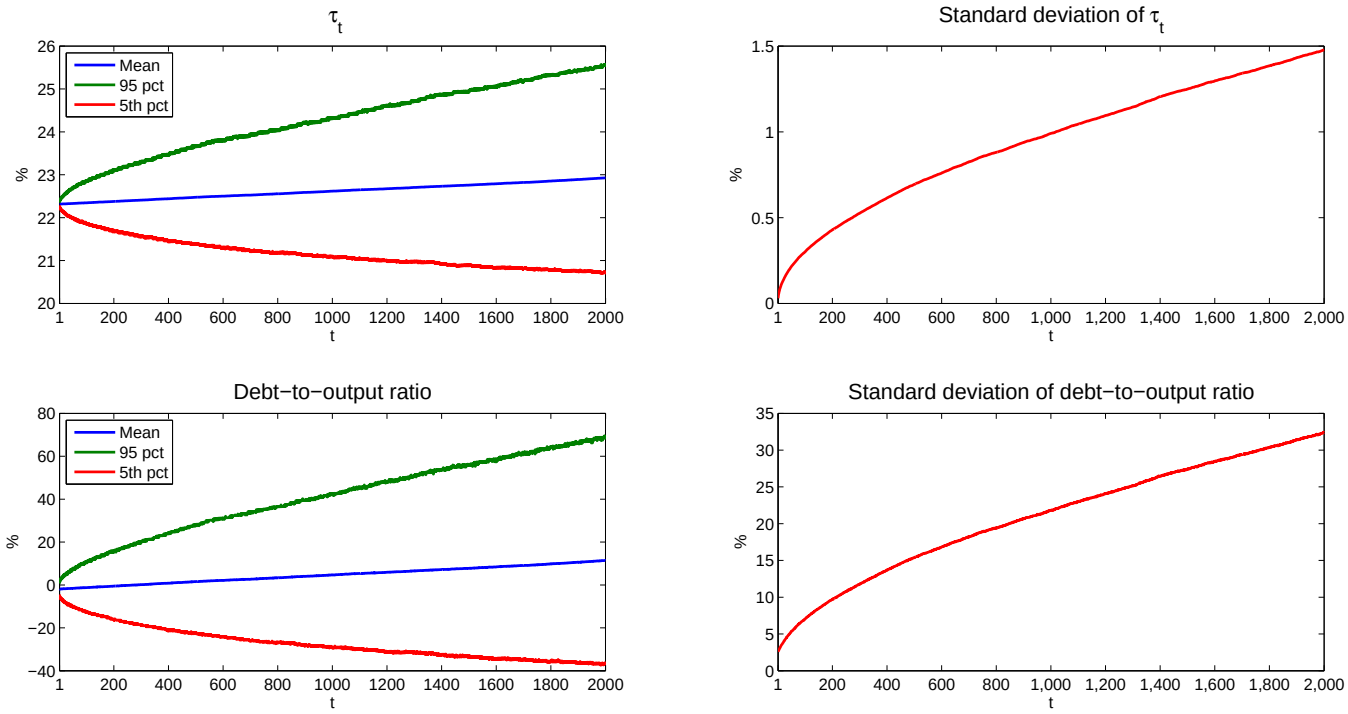

Figure 4: Ensemble moments.

Table 1: Ensemble moments.

	Expected utility	*Recursive utility*				
		t=200	t=500	t=1000	t=1500	t=2000
Tax rate in %						
Mean	22.30	22.38	22.47	22.62	22.76	22.93
Standard deviation	0	0.43	0.69	0.99	1.25	1.48
95th percentile		23.12	23.68	24.34	24.96	25.54
5th percentile		21.66	21.4	21.08	20.88	20.73
Debt-to-output ratio in %						
Mean	-1.91	-0.57	1.57	4.72	7.75	11.37
Standard deviation	1.91	9.68	15.37	21.83	27.44	32.40
95th percentile		15.17	27.51	42.32	55.98	68.39
5th percentile		-16.51	-22.21	-29.05	-33.45	-36.78

The table reports ensemble moments for the time-additive case of Lucas and Stokey ($\gamma = 1$) and the recursive utility case ($\gamma = 10$). In the expected utility case, the debt-to-output ratio takes the values 0 and -3.8139% for the low and high shock respectively.

Table 2: Statistics of tax rate sample paths.

	Recursive utility	
	short samples	long samples
Autocorrelation	0.9791	0.9980
Correlation of $\Delta\tau$ with g	-0.9999	-0.9984
Correlation of $\Delta\tau$ with output	-0.9977	-0.9762
Correlation of τ with g	-0.1098	-0.0346
Correlation of τ with output	-0.1793	-0.2418

The table reports median sample statistics across 10000 sample paths of the tax rate. For the time-additive case the respective moments are not well defined since the tax rate is constant. For the recursive utility case the median sample statistics are calculated for short samples (the first 200 periods) and long samples (2000 periods).

way to see the fanning-out is to consider the 5th and 95th percentile of the tax rate, which are decreasing and increasing respectively over time. For example, at $t = 2000$ they become 20.73% and 25.54% respectively.

The positive drift and the "fanning-out" of the distribution of the tax rate is mirrored in the debt-to-output ratio, as is clear from the lower panels in figure 4. The debt-to-output ratio has a mean of -1.91% and a standard deviation of 1.91 percentage points in the expected utility case. The mean of the debt-to-output ratio is -0.57% at $t = 200$ and rises to 11.37% at $t = 2000$. The standard deviation rises from 9.68 percentage points at $t = 200$ to 32.4 percentage points at $t = 2000$. The 5th and 95th percentile of the distribution of the debt-to-output ratio at $t = 2000$ are -36.78% and 68.39% respectively.

7.6 Persistence and countercyclicality

The martingale property of the inverse of the excess burden of taxation in proposition 3 and the inverse of the tax rate in proposition 5 translates quantitatively to a very high *persistence* of the tax rate, despite the fact that government expenditure shocks are i.i.d., which contrasts to the standard history-independence result of Lucas and Stokey (1983). Table 2 reports the median autocorrelation for short and long sample paths. The persistence in the tax rate is of the order of 0.998. The negative correlation of debt in marginal utility units with government expenditures, i.e. the fiscal hedging we highlighted earlier, leads to a strong negative correlation of the change in tax rates with government expenditure shocks. The correlation is -0.9984 for long samples. Increases in government expenditures increase output and therefore the change in tax rates is *countercyclical* with correlation equal to -0.9762. Note that the theory predicts that *changes* in tax rates are affected by the debt position, whereas the level of the tax rate depends on the

Table 3: Higher risk aversion or higher shock variance.

	Expected utility	*Recursive utility*				
		t=200	t=500	t=1000	t=1500	t=2000
Higher risk aversion						
Tax rate in %						
Mean	22.30	22.39	22.49	22.65	22.80	22.99
Standard deviation	0	0.50	0.80	1.15	1.46	1.74
95th percentile		23.25	23.90	24.66	25.39	26.1
5th percentile		21.57	21.26	20.89	20.65	20.48
Debt-to-output ratio in %						
Mean	-1.907	-0.44	1.92	5.36	8.69	12.74
Standard deviation	1.907	11.12	17.76	25.31	32.02	38.13
95th percentile		17.80	31.82	49.82	66.02	80.61
5th percentile		-18.48	-25.12	-33.14	-38.32	-42.22
Higher shock variance						
Tax rate in %						
Mean	22.29	22.46	22.67	23.00	23.35	23.76
Standard deviation	0	0.65	1.07	1.58	2.07	2.55
95th percentile		23.61	24.56	25.80	27.09	28.41
5th percentile		21.40	21.05	20.67	20.44	20.29
Debt-to-output ratio in %						
Mean	-2.83	0.05	4.65	11.92	19.44	28.26
Standard deviation	2.83	14.69	23.73	34.75	45.47	55.81
95th percentile		24.33	45.58	73.14	101.54	129.68
5th percentile		-23.57	-30.89	-39.31	-44.17	-47.51

Ensemble moments for the case of higher risk aversion ($\gamma = 11$) or the case of shocks with a standard deviation that corresponds to 3% of average first-best output. In order to avoid sample uncertainty, I use the same realizations of shocks as in table 1.

cumulative debt positions of the government. As a result, the correlation of the level of the tax rate with government expenditures or output is small (-0.0346 and -0.2418 respectively).

7.7 Higher risk aversion or more volatile shocks

Changes in tax rates depend on the deviation from expected utility (so the deviation of γ from unity in this illustration) and on the size of the debt positions that the government is taking, as is clearly seen from proposition 5. It is natural to conjecture that larger risk aversion or higher

risk in the economy in the sense of more volatile government expenditures that need to be insured against, will lead to higher effects on the tax rate. Table 3 reports the ensemble moments for two experiments of interest. At the upper part of the table risk aversion is increased to $\gamma = 11$, keeping the rest of the calibration the same. At the lower part, the standard deviation of the shocks is increased, keeping the mean value of the shocks and the rest of the parameters the same. In particular, I set $g_L = 0.068$ and $g_H = 0.092$ which correspond now to 17% and 23% of average first-best output, so the standard deviation of the share of government spending in average first-best output is 3%.

In both cases the increase over time of the mean and the standard deviation of the tax rate and the debt-to-output ratio are larger than for the baseline calibration. For the higher risk aversion case the mean tax rate is 22.39% at $t = 200$ and 22.99% at $t = 2000$, which is larger but pretty similar to the case of $\gamma = 10$, at least for the time horizons considered. The effects on the standard deviation of the tax rate are more noticeable. At $t = 200$ the standard deviation is 0.5 percentage points and rises to 1.74 percentage points at $t = 2000$. Similarly, the effects on the standard deviation of the debt-to-output ratio are larger as well. The standard deviation is 11.12 percentage points at $t = 200$ and becomes 38.13 percentage points at $t = 2000$. Turning to the case of more volatile shocks, the mean tax rate is 22.46% at $t = 200$ and rises to 23.76% at $t = 2000$. The changes in the standard deviation of the tax rate are large, reflecting the higher variance of the shocks. The standard deviation of the tax rate increases from 0.65 percentage points at $t = 200$ to 2.55 percentage points at $t = 2000$. The debt-to-output ratio has a mean of -2.83% and a standard deviation of 2.83 percentage points in the expected utility case. With recursive utility, the mean debt-to-output ratio is 28.26% and the standard deviation 55.81 percentage points at $t = 2000$. The range of the debt-to-output ratio is pretty large at $t = 2000$, with a 5th and 95th percentile that are -47.51 and 129.68 percentage points respectively.

To conclude, the martingale-like behavior of the tax rate translates quantitatively to a tax rate that behaves approximately as a random walk with drift, with a distribution that is fanning-out over time. The debt-to-output ratio follows the same behavior. The conditional volatility of the tax rate is small but the unconditional volatility is large due to high persistence. Higher deviations from expected utility or more volatile shocks make these effects more pronounced.

7.8 The very long-run

The analysis has focused on a horizon up to $2,000$ periods. It is of interest to see what happens in the long-run and explore the convergence properties (if any) of the excess burden of taxation and the tax rate.

Proposition 6. *Consider the utility function (35) and the i.i.d. shock specification. <u>If</u> the excess burden of taxation <u>does</u> converge, then it has to converge to zero.*

Proof. See Appendix. □

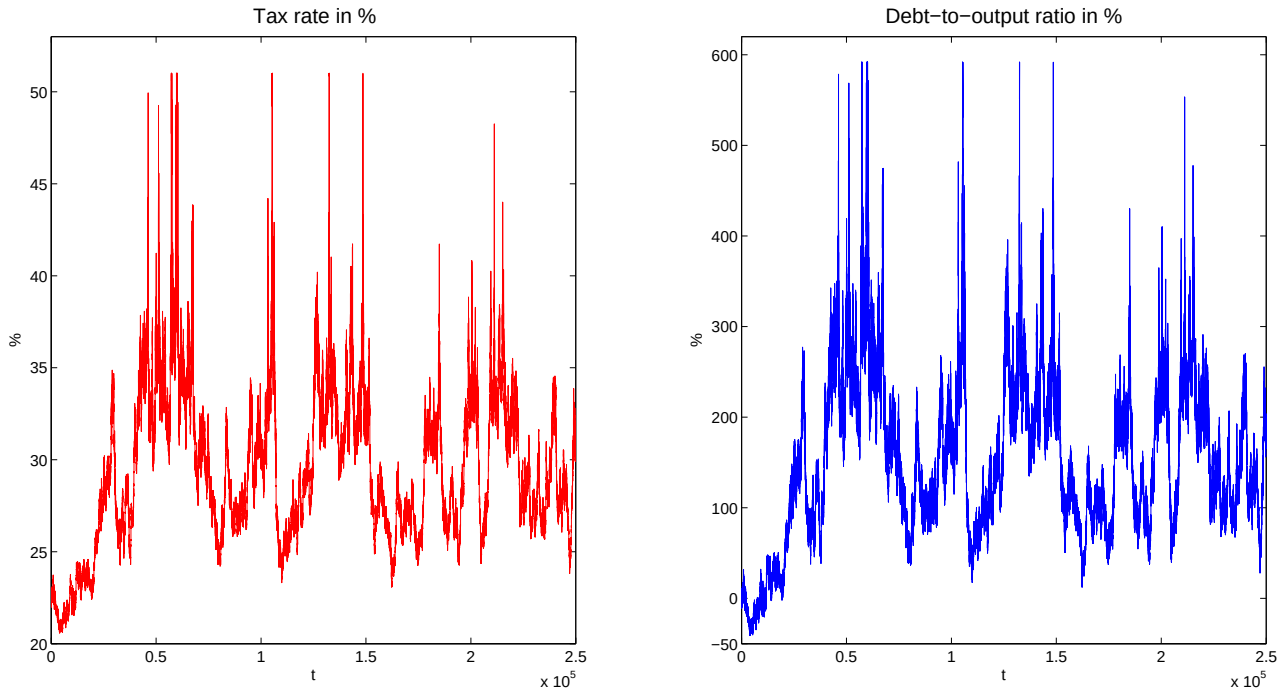

Figure 5: First 250,000 periods of the long simulation.

The intuition behind this result is simple. If the excess burden converged to a positive number (or random variable) the government would have to equalize asymptotically its state-contingent positions in marginal utility units across shocks. For the particular utility function this cannot be the case because for any constant Φ, debt in marginal utility units is always higher for low shocks, making the government vary distortions across shocks. As a result, the only candidate convergence point is the first-best case of no distortions, $\Phi = 0$. The first-best is an absorbing state, and the government uses its interest income on assets to finance government expenditures.

However, as we stressed earlier in the analysis after proposition 3, convergence is not guaranteed. And according to the numerical results, the tax rate displays a positive and *not* a negative drift. This finding contrasts also with the analysis of Aiyagari et al. (2002) in incomplete markets setups, who show that for some special cases, we have convergence to the first-best.[42]

In previous sections I have analyzed properties of the optimal plan by focusing on an interior solution of the Ramsey problem, treating the upper bounds of the state space $Z(g)$ as non-binding. At *high* values of the state variable z though, the planner may find himself wanting to increase the position for the low shock, but being constrained to do so, due to the upper bound.[43] In that case, the martingale result on the inverse of the excess burden of taxation (and the tax rate) breaks

[42]Note though that the likelihood ratio M_t does indeed converge to zero, confirming numerically the difficulty in establishing convergence results with respect to the physical measure.

[43]This is the reason why the difference $z'_L - z'_H$ in the left panel of figure 1 starts decreasing for high values of the current state.

Table 4: Moments from the stationary distribution.

	Tax rate in %	**Debt-to-output ratio in %**
Mean	30.36	171.31
Standard deviation	4.75	100.67
Autocorrelation	0.9994	0.9928

The simulation is 2 million periods long. In order to avoid the effect of initial conditions, I drop the first 50000 observations in the moment calculation.

down.[44]

The parts of the state space where the upper bound binds are visited in the very long-run due to the positive drift in the tax rate and the corresponding drift in debt in marginal utility units (which is mirrored in the drift of the debt-to-output ratio). I run a very long simulation to explore the asymptotic properties of the optimal plan. Figure 5 plots the tax rate and the corresponding debt-to-output ratio for the first $250,000$ periods. Whenever the solution is interior, the tax rate and the debt-to-output ratio exhibit on average an upward drift. The drift breaks down whenever the upper bound of the state space is hit.[45]

The possibility of a binding upper bound of the state space makes essentially the distribution of the tax rate *stationary* in the long-run. I allow the simulation to run for two million periods in order to make sure that I calculate moments from the stationary distribution. Table 4 displays the respective moments of the stationary distribution of the tax rate and the debt-to-output ratio. The mean tax rate is 30.36% with a standard deviation of 4.75 percentage points and an autocorrelation of 0.9994. The debt-to-output ratio is on average 171.31% with a standard deviation of 100.67 percentage points and a very high autocorrelation as well. In conclusion, the tax rate and the debt-to-output ratio exhibit a large mean and volatility in the long-run.[46]

[44]The optimality condition with respect to z when we take account of the upper bound of the state space is $\Phi_{t+1}(1+(1-\beta)(1-\gamma)\eta_{t+1}\Phi_t) \leq \Phi_t$. If $1+(1-\beta)(1-\gamma)\eta_{t+1}\Phi_t > 0$, we get $\frac{1}{\Phi_{t+1}} \geq \frac{1}{\Phi_t} + (1-\beta)(1-\gamma)\eta_{t+1}$, which implies that $1/\Phi_t$ is a submartingale with respect to $\pi_t \cdot M_t$. Note that we cannot infer anymore that Φ_t is a submartingale with respect to $\pi_t \cdot M_t$.

[45]So the upper bound of the state space for the low shock acts as a reflecting barrier. Another interpretation of proposition 6 is that there cannot exist an upper bound (common across the state spaces) that acts as an absorbing barrier. If it did exist, then there would exist a positive candidate point for convergence.

[46]The upper bounds of the state space correspond to a debt-to-output ratio of 593.53% and 554.56% for the low and high shock respectively. The lower bounds were set to the values of assets that would support the first-best allocation. These correspond to an *asset-to-output* ratio of −510.06% and −492.23% for the low and high shock respectively. See the Appendix for further details.

8 Extensions in an economy with capital

Consider now an economy with capital as in Zhu (1992) and Chari et al. (1994) and recursive preferences. To fix ideas, let s capture uncertainty about government expenditure shocks or technology shocks, with the probability of a partial history denoted by $\pi_t(s^t)$. The resource constraint in an economy with capital reads

$$c_t(s^t) + k_{t+1}(s^t) - (1-\delta)k_t(s^{t-1}) + g_t(s^t) = F(s_t, k_t(s^{t-1}), h_t(s^t)), \tag{36}$$

where δ denotes the depreciation rate, $k_{t+1}(s^t)$ capital measurable with respect to s^t and F a constant returns to scale production function. The representative household accumulates capital that can be rented at rental rate $r_t(s^t)$, and pays capital income taxes with rate $\tau_t^K(s^t)$. The household's budget constraint reads

$$c_t(s^t) + k_{t+1}(s^t) + \sum_{s_{t+1}} p_t(s_{t+1}, s^t) b_{t+1}(s^{t+1}) \leq (1-\tau_t(s^t))w_t(s^t)h_t(s^t) + R_t^K(s^t)k_t(s^{t-1}) + b_t(s^t),$$

where $R_t^K(s^t) \equiv (1-\tau_t^K(s^t))r_t(s^t) + 1 - \delta$, the after-tax gross return on capital.

I provide the details of the competitive equilibrium and the analysis of the Ramsey problem in the Appendix and summarize here the main results. In short, the completeness of the markets allows the recasting of the household's budget constraint in terms of wealth, $W_t \equiv b_t + R_t^K k_t$, making therefore *wealth in marginal utility units* a state variable for the optimal taxation problem. In particular, let wealth in marginal utility units be defined as $z_t \equiv U_{ct} W_t$. With this definition of z_t, the dynamic implementability constraint in an economy with capital is the same as in (17). Therefore, (z, k, s) become the relevant state variables for a recursive representation of the Ramsey problem. The excess burden of taxation Φ captures now the shadow cost of an additional unit of wealth in marginal utility units, $\Phi = -v_z(z, k, s)$, where $v(z, k, s)$ denotes the value function. The excess burden is not constant anymore as it would be in the expected utility case. In particular, we have:

Proposition 7. *The law of motion of Φ_t in an economy with capital remains (28), with an η_t that denotes the relative* <u>*wealth*</u> *position in marginal utility units. The planner allocates more distortions on high-wealth states ($\eta > 0$) and less distortions on low-wealth states ($\eta < 0$). Lemma 1 and propositions 3, 4 and 5 go through, so the same conclusions are drawn for the dynamics of the excess burden and the labor tax as in an economy without capital.*

Proposition 7 generalizes the results about the excess burden of taxation and the labor tax that we found earlier in an economy without capital. In order to interpret the allocation of distortions, recall that in an economy without capital the planner was allocating more distortions on events

where his debt positions were relatively high in order to take advantage of the negative covariance between debt in marginal utility units and the return on state-contingent debt. Essentially the same mechanism is at play in an economy where there is an additional saving instrument as capital. Due to the completeness of the markets, the relevant hedging instrument of the planner is state-contingent wealth in marginal utility units (instead of just state-contingent debt), which exhibits a negative covariance with returns due to the pricing effect of continuation values. As a result, the planner allocates more distortions on events where wealth in marginal utility units is relatively high and less distortions on events where it is relatively low.

8.1 Capital taxation

The variation in the allocation of distortions is coming from the effect of z_t on the market value of the *wealth* portfolio, $\omega_t = E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} z_{t+1}$, through the effect of the long-run on equilibrium asset prices. Considering the optimal choice of capital, it is clear that capital has also a novel effect on ω_t, since it affects the future utility of the household, and therefore, the pricing of state-contingent wealth. This effect alters the incentives for taxation at the *intertemporal* margin. In particular, the optimality condition with respect to capital in sequence notation can be written as (details in the Appendix),

$$E_t S_{t+1}^{\star}(1-\delta+F_{K,t+1}) = 1, \quad \text{where } S_{t+1}^{\star} \equiv \beta m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} \frac{\lambda_{t+1}/\Phi_{t+1}}{\lambda_t/\Phi_t}, \tag{37}$$

where λ_t stands for the multiplier on the resource constraint (36) in the recursive formulation of the second-best problem.

I will call $S_{t+1}^{\star}$ the *planner's* stochastic discount factor. The variable $S_{t+1}^{\star}$ captures how the planner discounts the *pre-tax* gross return on capital at the *second-best* allocation. $S_{t+1}^{\star}$ contrasts to the market stochastic discount factor $S_{t+1} \equiv \beta m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} U_{c,t+1}/U_{c,t}$, which prices *after-tax* returns, $E_t S_{t+1} R_{t+1}^K = 1$. In a first-best world with lump-sum taxes available, we identically have $S_{t+1}^{\star} \equiv S_{t+1}$. The planner's discount factor $S_{t+1}^{\star}$ though can differ from S_{t+1} in the second-best world, and is useful in summarizing the optimal wedge at the intertemporal margin in the form of the *ex-ante* tax rate on capital income.

In particular, as is well known from Zhu (1992) and Chari et al. (1994), there is a *multiplicity* of state-contingent debt and capital tax policies that can implement the second-best allocation as a competitive equilibrium. However, it is well known that we can uniquely determine the ex-ante tax rate on capital income $\bar{\tau}_{t+1}^K(s^t)$, which is restricted to be *non-state contingent* and is defined as

$$\bar{\tau}_{t+1}^K \equiv \frac{E_t S_{t+1}(1-\delta+F_{K,t+1}) - 1}{E_t S_{t+1} F_{K,t+1}}.$$

Using (37), we can express the ex-ante capital tax as

$$\bar{\tau}^K_{t+1} = \frac{E_t\big[S_{t+1} - S^\star_{t+1}\big](1-\delta+F_{K,t+1})}{E_t S_{t+1} F_{K,t+1}}. \tag{38}$$

Thus, there is a positive (negative) tax rate on capital income if the numerator of (38) is positive (negative). Another way to think about the sign of the numerator is in terms of the *size* of the (non-centered) *covariances* of the planner's and the market stochastic discount factors with the pre-tax capital return $1-\delta+F_{K,t+1}$. A (non-centered) covariance of the market stochastic discount factor with the pre-tax return on capital that is larger (smaller) than the respective non-centered covariance of the planner's stochastic discount factor with the pre-tax return, leads to a positive (negative) ex-ante tax rate $\bar{\tau}^K_{t+1} > 0$ ($\bar{\tau}^K_{t+1} < 0$).

The difference in the two discount factors $S_{t+1} - S^\star_{t+1}$ can be expressed in terms of differences in the inverse of the excess burden of taxation and differences in the own and cross elasticity of the marginal utility of consumption, which leads to the following proposition about capital taxation.[47]

Proposition 8. *(Capital taxation criterion) The ex-ante tax rate on capital income $\bar{\tau}^K_{t+1}, t \geq 1$ is positive (negative) iff*

$$E_t\zeta_{t+1}\Big[\underbrace{\big(\frac{1}{\Phi_t} - \frac{1}{\Phi_{t+1}}\big)}_{\textit{change in } 1/\Phi_t} + \underbrace{\big(\epsilon_{cc,t+1} + \epsilon_{ch,t+1} - \epsilon_{cc,t} - \epsilon_{ch,t}\big)}_{\textit{change in period elasticities}}\Big] > (<) \quad 0,$$

with weights $\zeta_{t+1} \equiv S_{t+1}(1-\delta+F_{K,t+1})/E_tS_{t+1}(1-\delta+F_{K,t+1})$. If $\epsilon_{cc}+\epsilon_{ch}$ is <u>constant</u>, then the only reason for taxing the intertemporal margin comes from variation in the excess burden of taxation Φ_t.

Proof. The first-order condition with respect to consumption for $t \geq 1$ is $U_{ct} + \Phi_t\Omega_{ct} = \lambda_t$. Thus, $1/\Phi_t + \Omega_{ct}/U_{ct} = \lambda_t/(\Phi_t U_{ct}) > 0$. Write the planner's discount factor as $S^\star_{t+1} = S_{t+1}\frac{\lambda_{t+1}/(\Phi_{t+1}U_{c,t+1})}{\lambda_t/(\Phi_t U_{ct})} = S_{t+1}\frac{1/\Phi_{t+1}+\Omega_{c,t+1}/U_{c,t+1}}{1/\Phi_t+\Omega_{ct}/U_{ct}}, t \geq 1$. Use (30) to express the difference in the two discount factors as

$$\frac{S_{t+1} - S^\star_{t+1}}{S_{t+1}} = \frac{\frac{1}{\Phi_t} - \frac{1}{\Phi_{t+1}} + \epsilon_{cc,t+1} + \epsilon_{ch,t+1} - \epsilon_{cc,t} - \epsilon_{ch,t}}{\frac{1}{\Phi_t} + 1 - \epsilon_{cc,t} - \epsilon_{ch,t}}, t \geq 1. \tag{39}$$

The denominator is positive. Use the expression for the difference in the stochastic discount factors in the numerator of (38), simplify and normalize the weights so that they integrate to unity to get the criterion for capital taxation. □

[47] As it was the case with the labor tax in footnote 30, the capital tax criterion applies also for the deterministic, stochastic time-additive and risk-sensitive case for any standard U.

The fact that only the ex-ante tax rate is uniquely determined leads to a capital taxation criterion that depends on the *weighted average* of the change in the elasticity of the marginal utility of consumption (the expected utility part) and the change in the excess burden of taxation (the novel recursive utility part), with weights ζ_{t+1} that are proportional to the product of the stochastic discount factor and the pre-tax gross return on capital.

Time-additive economy. Assume that we are in either in a deterministic economy or in a stochastic but time-additive economy with $\rho = \gamma$. In both cases Φ_t is *constant* and the capital taxation criterion depends only on the change in period elasticities. For the deterministic case, capital income is taxed (subsidized) if the sum of the own and cross elasticity is increasing (decreasing). A necessary and sufficient condition for a zero capital tax at *every* period from period two onward is a sum of elasticities of the period marginal utility of consumption that is *constant*, which implies that $S^{\star}_{t+1} = S_{t+1}$. If the period utility function is such so that the elasticities are *not* constant for each period, then there is zero tax on capital income at the deterministic *steady state*, where the constancy of the consumption-labor allocation delivers constant elasticities. This delivers the zero-tax result of Chamley (1986) and Judd (1985). In the stochastic case of Chari et al. (1994) and Zhu (1992), the *ex-ante* tax is positive or negative, if the *weighted average* of the change in elasticities is positive (negative). The capital tax criterion is not as sharp as in the deterministic case, since *only* the ex-ante tax rate is uniquely determined by the second-best allocation. As a result, this notion of an "average" tax weighs the intertemporal distortions across states. Note that variation in the sum of period elasticities is a necessary condition for an intertemporal wedge, since a constant sum of period elasticities implies $S^{\star}_{t+1} = S_{t+1}$.[48]

Recursive utility, $\rho < \gamma$. For the case of recursive utility, the full version of the formula in proposition 8 applies. The change in *both* the sum of elasticities *and* the excess burden of taxation determines the intertemporal wedge. Consider the case of constant period elasticities, which would deliver a zero ex-ante capital tax in the case when the distinction between time and risk is absent. For an example in this class, consider the composite good

$$u(c, 1-h) = \left[c^{1-\rho} - (1-\rho)\mathrm{v}(h)\right]^{\frac{1}{1-\rho}}, \quad \mathrm{v}', \mathrm{v}'' > 0, \tag{40}$$

that delivers a period utility $U = (u^{1-\rho} - 1)/(1-\rho)$, that is separable between consumption and leisure and isoelastic in consumption.[49] Chari et al. (1994) and Zhu (1992) have demonstrated that these preferences deliver a *zero* ex-ante capital tax from period two onward. This is easily interpreted in terms of proposition 8, since $\epsilon_{cc} = \rho$ and $\epsilon_{ch} = 0$.

[48] It is not a sufficient condition in the stochastic case, since the weighted average can still in principle deliver a zero tax.

[49] The same comments as in footnote 34 apply. The constant Frisch elasticity case is obviously a member of this class.

With recursive preferences though, even in the constant elasticity case, there is a novel source of taxation coming from the willingness of the planner to take advantage of the pricing effects of state-contingent wealth positions. By using the law of motion of the excess burden of taxation (28) to substitute η_{t+1} for the change in $1/\Phi_t$, the criterion becomes

$$\bar{\tau}^K_{t+1} > (<) \quad 0 \quad \text{iff} \quad E_t \zeta_{t+1} \eta_{t+1} > (<) \quad 0.$$

Recall that the *level* of the excess burden of taxation captures distortions at the intratemporal margin. Proposition 8 shows that the *change* in the excess burden of taxation determines the sign of distortions at the *intertemporal* margin. States where there are positive relative wealth positions ($\eta_{t+1} > 0$), with a correspondingly beneficial appreciation of the value of wealth, make the planner increase the excess burden of taxation, $\Phi_{t+1} > \Phi_t$. This delivers a planner's discount factor that is smaller than the market discount factor, $S^\star_{t+1} < S_{t+1}$, as can be seen from (39). Intuitively, we can think of an increase in Φ_t (which implies an increased labor tax) as an incentive to introduce a *positive* state-contingent intertemporal wedge. To state it differently, a positive state-contingent wedge, by reducing the utility of the household, increases the value of wealth and therefore, the marginal revenue of the government when $\eta_{t+1} > 0$. In contrast, at states where there are negative relative wealth positions ($\eta_{t+1} < 0$), we have $\Phi_{t+1} < \Phi_t$ and therefore a decreased labor tax and an incentive for a *negative* state-contingent intertemporal wedge. The sign of the *non-state* contingent ex-ante capital tax is determined by the weighted average of the relative wealth positions, i.e. by the relative importance of positive versus negative state-contingent intertemporal distortions, which are captured by the respective fiscal hedging of the government, η_{t+1}.

8.2 Ex-ante subsidy

To gain more insight about the sign of ex-ante tax rate we have to understand the behavior of the weights ζ_{t+1}, i.e. how the stochastic discount factor and the marginal product of capital vary with shocks. Consider the separable preferences in (40) and let $\rho = 1 < \gamma$. Remember that with these preferences the stochastic discount factor becomes $S_{t+1} = \beta m_{t+1}(c_{t+1}/c_t)^{-1}$, where $m_{t+1} = \exp[(1-\beta)(1-\gamma)v_{t+1}]/E_t \exp[(1-\beta)(1-\gamma)v_{t+1}]$. The capital tax criterion simplifies to

$$\bar{\tau}^K_{t+1} > (<) \quad 0 \quad \text{iff} \quad E_t m_{t+1} c^{-1}_{t+1} \big(1 - \delta + F_{K,t+1}\big) \eta_{t+1} > (<) \quad 0.$$

Equivalently, the criterion can be expressed in terms of the sign of a conditional *covariance* (with respect to the continuation-value adjusted measure, indicated by the superscript M),

$$\bar{\tau}^K_{t+1} > (<) \quad 0 \quad \text{iff} \quad \text{Cov}^M_t\big(c^{-1}_{t+1} \cdot (1-\delta+F_{K,t+1}), \eta_{t+1}\big) > (<) \quad 0,$$

by using the fact that the average relative wealth position is *zero* by lemma 1, $E_t m_{t+1}\eta_{t+1} = 0$. In order to understand the sign of the covariance, assume that the only shocks in the economy are government expenditure shocks and that they take two values, high and low. We expect that the household will consume less at the high shock than at the low shock, and therefore marginal utility will be higher at the high shock. Furthermore, we expect that the household will work more at the high shock in order to compensate for the negative wealth effect that government expenditures impose, reducing the capital-labor ratio and increasing therefore the marginal product of capital. As a result, the product of marginal utility and the gross return on capital will be higher at the high shock than at the low shock. *If*, as in our example in an economy without capital, the government hedges expenditure shocks by taking a low wealth position in marginal utility units for the high shock and a high wealth position for the low shock, then the relative position η will be negative for high shocks and positive for low shocks. Given the assumed behavior of the weights and the fiscal hedging of the government, the capital tax criterion will be putting *more* weight on the negative positions than in the positive positions, leading therefore to $E_t m_{t+1} c^{-1}_{t+1}\big(1-\delta+F_{K,t+1}\big)\eta_{t+1} < E_t m_{t+1}\eta_{t+1} = 0$, or, in other words, to a *negative* covariance of the relative wealth positions and the product of marginal utility and capital gross returns.[50] Thus, due to the high marginal utility and capital returns at adverse fiscal shocks, the state-contingent negative wedge ("capital subsidy") at the high shock is weighted more than the state-contingent positive wedge ("capital tax") at the low shock, leading to an *ex-ante subsidy* on capital income.[51]

The capital subsidy result relies on the negative correlation of wealth in marginal utility units with government expenditure shocks. To see if this type of fiscal hedging is valid, consider an economy with the utility function (35) that was used for the numerical illustrations, and a simplified stochastic structure – deterministic except for one period.[52]

[50] To see that clearly, let $g_H > g_L$ and let subscripts denote if we are at the high or low shock. For ease of exposition I suppress the time subscripts. By assumption we have $c_H < c_L$, $F_{K,H} > F_{K,L}$, $\eta_H < 0$ and $\eta_L > 0$. Therefore, $c^{-1}_H(1-\delta+F_{K,H}) > c^{-1}_L(1-\delta+F_{K,L})$. The covariance is $\text{Cov}^M = c^{-1}_L(1-\delta+F_{K,L})\pi_L m_L\eta_L + c^{-1}_H(1-\delta+F_{K,H})\pi_H m_H\eta_H$. But $c^{-1}_H(1-\delta+F_{K,H})\pi_H m_H\eta_H < c^{-1}_L(1-\delta+F_{K,L})\pi_H m_H\eta_H$ since $\eta_H < 0$. Therefore, $\text{Cov}^M < c^{-1}_L(1-\delta+F_{K,L})\big[\pi_L m_L\eta_L + \pi_H m_H\eta_H\big] = 0$, since the relative wealth positions integrate to zero with respect to the continuation-value adjusted measure.

[51] Assume risk-sensitive preferences with parameter $\sigma < 0$ and the same isoelastic period utility as in (40). Then, it is straightforward to see that the ex-ante subsidy result holds also for $\rho \neq 1$ under the same assumptions for the weights and fiscal hedging.

[52] This stochastic setup is in the spirit of example 5 of Lucas and Stokey (1983) and is sufficiently rich for the purposes of displaying the fiscal hedging of the government and the implications for labor and capital taxes. A quantitative examination in a fully stochastic setup is a highly non-trivial task, due to the presence of the value functions in the constraint and the endogenous state spaces (for each value of the shocks) that determine where pairs of (z, k) live, and is beyond the scope of the current paper.

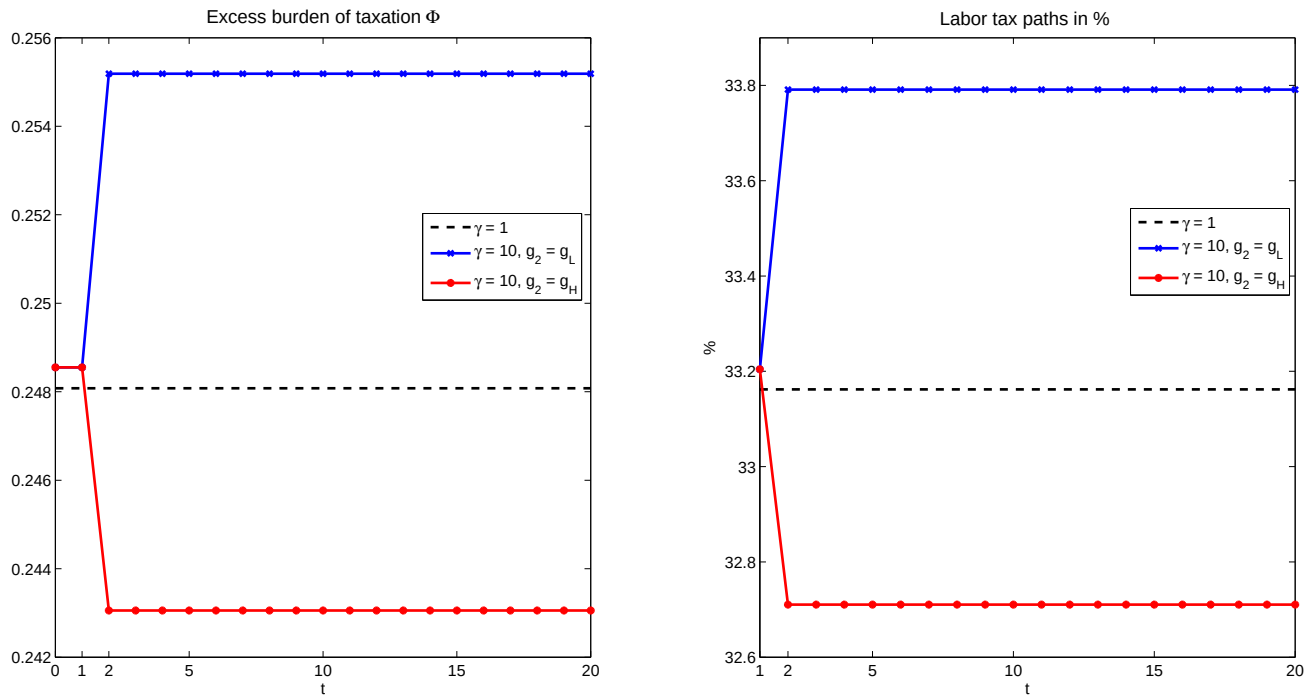

Figure 6: Paths of the excess burden of taxation and the labor tax for the high-shock and low-shock history.

The preference parameters are calibrated as in our computations in the economy without capital. The rest of the calibration and the solution method are in the Appendix. Let government expenditures take two values $g_L < g_H$. Assume that government expenditures are low with certainty except for $t = 2$. At $t = 2$ we have $g_2 = g_H$ with probability π and $g_2 = g_L$ with probability $1 - \pi$. I use superscripts for the endogenous variables in order to denote if we are at the high-shock history ($g_2 = g_H$) or at the low-shock history ($g_2 = g_L$). For example, $c_t^i, i = H, L$, denotes consumption at period $t \geq 2$ when the shock at $t = 1$ is high or low respectively.

The deterministic setup after the second period serves as an example of a case where the excess burden of taxation stays *permanently* at the values it assumes at $t = 2$. In particular, since there is no uncertainty before and after $t = 2$, we have $\Phi_1 = \Phi_0$ and $\Phi_t^i = \Phi_2^i, i = H, L, t \geq 2$. Turning to the issue of fiscal hedging, we find that the planner is taking a larger wealth position in marginal utility units at the low shock, $z_2^H < z_2^L$. As a result, he transfers distortions permanently towards the low-shock history and away from the high-shock history, so $\Phi_2^H < \Phi_2^L$. The left panel in figure 6 plots the respective paths for the excess burden of taxation and the right panel the labor tax dynamics, which obey proposition 5. Recall that this utility function implies a constant labor tax for $t \geq 1$ in the time-additive case. With recursive utility, despite the fact that government expenditures revert to a low value with certainty after $t = 2$, the labor tax becomes permanently low when there is an adverse shock at $t = 2$ and permanently high when there is favorable shock at $t = 2$.

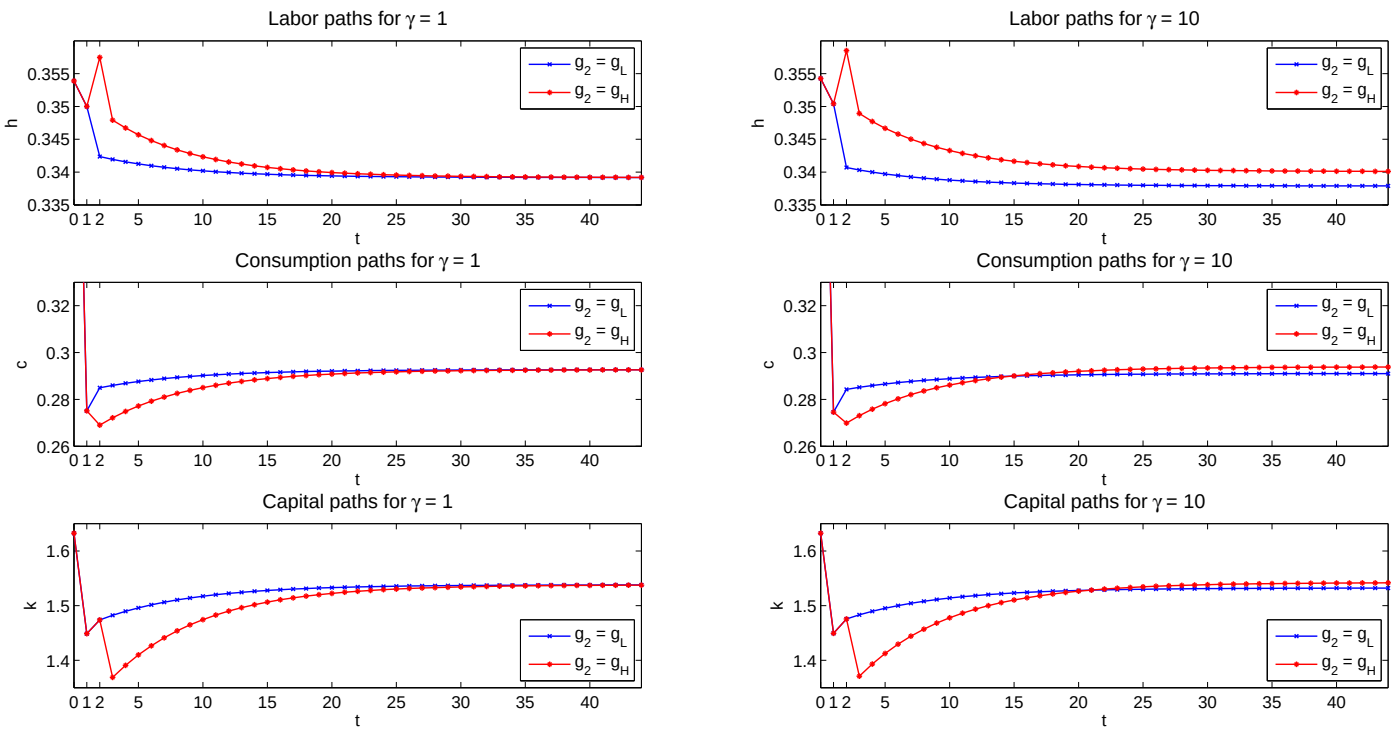

Figure 7: Left panels depict the labor, consumption and capital paths for the expected utility case ($\gamma = 1$) for the high- and low-shock history. Right panels depict the respective paths for recursive utility ($\gamma = 10$), which converge to two different steady states depending on the realization of the shock at $t = 2$.

The planner keeps the the tax rate permanently low or high, because any change in the tax rate in future periods will affect the price of claims at $t = 2$, due to the forward-looking nature of continuation utility. Consider for example the high shock-history. If the planner increased the tax rate at any period $t \geq 3$, he would decrease the utility of the agent at $t = 2$, leading therefore to a higher price of the claim contingent on $g_2 = g_H$. This is not optimal though, since the planner is hedging the bad shock with a small position, $z_2^H < z_2^L$, and therefore wants to have a low price, i.e. a high state-contingent return on his negative relative wealth position.

Turning to the capital tax, in the time-additive economy there is a *zero* ex-ante capital tax at $t = 2$ and a zero capital tax for $t \geq 3$.[53] For the recursive utility case, the capital tax will be zero for $t \geq 3$ since the economy becomes deterministic and the utility function belongs to the constant elasticity class. For $t = 2$, the ex-ante tax rate will not be zero and its sign depends on the fiscal

[53]The presence of initial wealth (which would be absent if we had zero initial debt, full depreciation and an initial tax rate on capital income of 100%) alters the taxation incentives for labor income at $t = 0$ and capital income at $t = 1$. In particular, the planner has an incentive to increase initial consumption in order to reduce initial wealth in marginal utility units. By subsidizing initial labor income and taxing capital income at $t = 1$, he is able to achieve that. The labor subsidies at the initial period are $\tau_0 = -17.69\%$ for the time-additive case and $\tau_0 = -17.76\%$ for the recursive utility case. Following Chari et al. (1994), I do not impose an upper bound on capital taxes. At $t = 1$ they take the values $\tau_1^K = 365.31\%$ and $\tau_1^K = 365.74\%$ for the time-additive and recursive utility case respectively. The desire to disentangle the effect of the initial conditions from the effect of uncertainty is the reason why I let the shock materialize at $t = 2$.

hedging of the government, as discussed in detail earlier. Figure 7 plots the time paths for labor, consumption and capital for the two histories. Consumption (labor) at $t = 2$ is lower (higher) when the expenditure shock is high, putting therefore a larger weight on the state-contingent "subsidy". As a result, we have an *ex-ante* subsidy, that takes the value of -0.5536% in this illustration. In addition, it is worth noting that, since the change in the labor tax is permanent, we have two different steady states depending on what value government expenditures took at $t = 2$. For the high-shock history, which is associated with a lower labor tax, the steady state entails higher labor, consumption and capital, whereas for the low-shock history, which is associated with a higher labor tax, the steady state involves lower labor, consumption and capital.

9 Concluding remarks

Dynamic optimal taxation entails the notions of time and risk. The analysis in this paper shows that when the attitudes towards these two notions are distinct, the tax-smoothing prescriptions of the dynamic Ramsey literature in frictionless environments are not valid. Labor tax volatility is optimal and can be quantitatively substantial. Furthermore, labor taxes are strongly counter-cyclical, display persistence independent of the stochastic properties of the exogenous shocks, and exhibit an upward drift over time. This pattern of labor taxes is reflected in debt-to-output ratios, which exhibit an increasing mean and volatility over time. At the intertemporal margin, there is a novel incentive to tax capital income, that can lead to ex-ante capital subsidies.

All these results are in stark contrast with the standard Ramsey prescriptions, and indicate that blurring the attitude towards risk with the attitude towards time is not an innocuous assumption for optimal fiscal policy. The results are rooted in the way state-contingent returns are formed with recursive utility. Equilibrium asset prices become debt-elastic, a fact that leads to a negative covariance between debt and returns. As a result, the policymaker takes larger debt positions in absolute value in order to minimize the cost of distortionary taxation. Cheaper debt or more profitable assets make the introduction of tax volatility and the running of larger surpluses and deficits less costly from a welfare perspective.

I have focused on time and risk in otherwise standard economies of the dynamic Ramsey tradition. An analysis beyond the representative agent framework as in Werning (2007) or Bassetto (1999), or an exploration of different timing protocols like lack of commitment, are worthy directions for future research.

A Further details of the Ramsey problem

A.1 State space

At first, define

$$
\begin{aligned}
A(g_1) \equiv \Big\{ & (z_1, V_1) | \exists \{c_t, h_t\}_{t\geq 1}, \{z_{t+1}, V_{t+1}\}_{t\geq 1}, \text{with } c_t \geq 0, h_t \in [0,1] \\
& \text{such that:} \\
& z_t = \Omega(c_t, h_t) + \beta E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} z_{t+1}, t \geq 1 \\
& V_t = \left[(1-\beta)u(c_t, 1-h_t)^{1-\rho} + \beta \mu_t (V_{t+1})^{1-\rho}\right]^{\frac{1}{1-\rho}}, t \geq 1 \\
& c_t + g_t = h_t, t \geq 1 \\
& \text{where } m_{t+1} \text{ defined as in (8)} \\
& \text{and the transversality condition holds, } \lim_{t\to\infty} E_1 \beta^t \left(\frac{M_{t+1}}{M_1}\right)^{\frac{\rho-\gamma}{1-\gamma}} z_{t+1} = 0. \Big\}
\end{aligned}
$$

The set $A(g_1)$ stands for the set of values of z and V at $t = 1$ that can be generated by an implementable allocation when the shock is g_1. From $A(g)$ we get the state space as $Z(g) \equiv \{z | \exists (z, V) \in A(g)\}$.

A.2 Initial period problem

The problem at $t = 0$ is

$$
\bar{V}_0(b_0, g_0) \equiv \max_{c_0, h_0, z_{1,g_1}} \left[(1-\beta)u(c_0, 1-h_0)^{1-\rho} + \beta \Big[\sum_{g_1} \pi_1(g_1|g_0) V(z_{1,g_1}, g_1)^{1-\gamma}\Big]^{\frac{1-\rho}{1-\gamma}}\right]^{\frac{1}{1-\rho}}
$$

subject to

$$
\begin{aligned}
& U_{c0} b_0 = \Omega(c_0, h_0) + \beta \sum_{g_1} \pi_1(g_1|g_0) \frac{V(z_{1,g_1}, g_1)^{\rho-\gamma}}{\left[\sum_{g_1} \pi_1(g_1|g_0) V(z_{1,g_1}, g_1)^{1-\gamma}\right]^{\frac{\rho-\gamma}{1-\gamma}}} z_{1,g_1} \\
& c_0 + g_0 = h_0 && \text{(A.1)} \\
& c_0 \geq 0, h_0 \in [0,1], && \text{(A.2)} \\
& z_{1,g_1} \in Z(g_1) && \text{(A.3)}
\end{aligned}
$$

where (b_0, g_0) given. The notation z_{1,g_1} denotes the value of the state variable z_1 at g_1. The overall value of the Ramsey problem $\bar{V}(.)$ depends on the initial conditions (b_0, g_0), which is why I use a different notation for the initial value function.

A.3 Transformed Bellman equation

Given the ρ-transformation of the value function, $v(z,g) \equiv \frac{V(z,g)^{1-\rho}-1}{(1-\beta)(1-\rho)}$, the Bellman equation takes the form

$$v(z,g) \quad = \quad \max_{c,h,z'_{g'}} U(c,1-h) + \beta \frac{\left[\sum_{g'} \pi(g'|g)\big(1+(1-\beta)(1-\rho)v(z'_{g'},g')\big)^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{1-\rho}{1-\gamma}} - 1}{(1-\beta)(1-\rho)}$$

subject to the transformed implementability constraint

$$z = \Omega(c,h) + \beta \sum_{g'} \pi(g'|g) \frac{[1+(1-\beta)(1-\rho)v(z'_{g'},g')]^{\frac{\rho-\gamma}{1-\rho}}}{\left[\sum_{g'} \pi(g'|g)[1+(1-\beta)(1-\rho)v(z'_{g'},g')]^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{\rho-\gamma}{1-\gamma}}} z'_{g'}$$

and to (19)-(21). The market value of the household's debt takes the form $\omega = \sum_{g'} \pi(g'|g) m'^{\frac{\rho-\gamma}{1-\gamma}}_{g'} z'_{g'}$, where $m'_{g'}$ stands for the conditional likelihood ratio,

$$m'_{g'} \equiv \frac{V(z'_{g'},g')^{1-\gamma}}{\sum_{g'} \pi(g'|g) V(z'_{g'},g')^{1-\gamma}} = \frac{\left[1+(1-\beta)(1-\rho)v(z'_{g'},g')\right]^{\frac{1-\gamma}{1-\rho}}}{\sum_{g'} \pi(g'|g)\left[1+(1-\beta)(1-\rho)v(z'_{g'},g')\right]^{\frac{1-\gamma}{1-\rho}}}.$$

A.4 Proof of lemma 1

Proof. Use the definition of η in (25) to get

$$\sum_{g'} \pi(g'|g) m'_{g'} \eta'_{g'} = \sum_{g'} \pi(g'|g) m'_{g'} V'^{\rho-1}_{g'} z'_{g'} - \mu^{\rho-1} \underbrace{\sum_{g'} \pi(g'|g) m'_{g'}}_{=1} \sum_{g'} \pi(g'|g) m'^{\frac{\rho-\gamma}{1-\gamma}}_{g'} z'_{g'}$$

$$= \mu^{\rho-1}\Big[\sum_{g'} \pi(g'|g) m'_{g'} \Big(\underbrace{\frac{V'_{g'}}{\mu}}_{=m'^{\frac{1}{1-\gamma}}_{g'}} \Big)^{\rho-1} z'_{g'} - \omega\Big] = \mu^{\rho-1}\Big[\sum_{g'} \pi(g'|g) m'^{\frac{\rho-\gamma}{1-\gamma}}_{g'} z'_{g'} - \omega\Big] = \mu^{\rho-1}[\omega-\omega] = 0.$$

□

B Computational details

State space. At first note that using the optimal wedge equation (29) and the resource constraint (3) we can express the optimal consumption-labor allocation as functions of the shock g and Φ, $c(g,\Phi)$ and $h(g,\Phi)$. Let $U^\star(g,\Phi) \equiv U(c(g,\Phi), 1-h(g,\Phi))$ and $\Omega^\star(g,\Phi) \equiv \Omega(c(g,\Phi), h(g,\Phi))$. $U^\star$

stands for the period utility at g when the excess burden of taxation is Φ and $\Omega^\star$ the respective government surplus in marginal utility units. For the utility function used we have $\Omega^\star = 1 - a_h h(g,\Phi)^{1+\phi_h}$.

I create values of z that can be generated by a competitive equilibrium. In particular, I use values that can be generated by a constant-Φ policy, which corresponds for the utility function in hand to a constant tax policy. Fix Φ to a particular value. Given a constant value of Φ we get a history-independent allocation which allows us to solve easily for the the utility recursion $v^\star(g,\Phi) = U^\star(g,\Phi) + \frac{\beta}{(1-\beta)(1-\gamma)} \ln \sum_{g'} \pi(g'|g) \exp((1-\beta)(1-\gamma)v^\star(g',\Phi))$. For each given Φ we get also the induced conditional likelihood ratio $m(g'|g) = \exp((1-\beta)(1-\gamma)v^\star(g',\Phi)) / \sum_{g'} \pi(g'|g) \exp((1-\beta)(1-\gamma)v^\star(g',\Phi))$.

The induced debt positions z for a given Φ are

$$\mathbf{z} = (I - \beta\tilde{\Pi})^{-1}\mathbf{\Omega}^\star,$$

where boldface variables denote column vectors and $\tilde{\Pi} \equiv \Pi \circ \mathcal{M}$, where Π the transition matrix of the shocks and $\mathcal{M}$ the matrix of $m(g'|g)$. The symbol $\circ$ denotes element by element (or else Hadamard) multiplication.

Thus, for each value of Φ we get a value of z for the low and high shock respectively and the corresponding utility of this policy. By construction, the constructed values of z can be generated by the competitive equilibrium and are a "nice" subset of the true state space. I vary Φ in the set $[0,\bar{\Phi}]$. The zero value of Φ corresponds to the first-best allocation, so the induced $z's$ are the level of government assets that would finance government expenditures without having to resort to distortionary taxation. I use $\bar{\Phi} = 0.5$. This choice is ad-hoc and corresponds to a tax rate of 50%. Let Z_i the state space for the low and high shock, $i = L, H$. For the lower and upper bounds of Z_i I use the minimum and maximum value of the debt position at i generated by a Φ in $[0,\bar{\Phi}]$ (which just correspond to $\Phi = 0$ and $\Phi = \bar{\Phi}$, because the implied z's are an increasing function of Φ). This delivers $Z_L = [-6.2355, 7.8819]$ and $Z_H = [-6.2911, 7.8391]$. The implied debt-to-output ratios are reported in the text.

Initial estimate of the value function. For each Φ I can associate the induced z to an induced $v^\star$, which provides an initial guess for the value function, $v^0(z, g_i)$, $z \in Z_i$. At first, I form a grid of points for $Z_i, i = L, H$ and perform value function iteration with grid search. There may be convergence issues because updating the value function in the constraint destroys contraction properties. To avoid that I have two loops:

- Inner Loop: *Given* the value function in the constraint, iterate on the Bellman equation till convergence (I use also policy function iteration to increase speed).

- Outer Loop: Update the value function in the constraint and repeat the inner loop.

The procedure is stopped when the value function in the constraint is approximately equal to the value function in the Bellman equation. The inner loop entails standard value function iteration and is convergent. There is no guarantee of convergence of the double loop. In the outer loop I use damping in order to improve convergence properties.

Final estimate of the value function. I used grid search in order to avoid non-convexities issues and the possibility of a local optimum. This procedure provides a first estimate of the value functions. For improved *precision*, I use the output of the two-loop procedure as an initial guess and fit the value functions at the two shocks with cubic splines. I use 167 breakpoints and 500 points for each Z_i and apply regression. More grid points are allocated at the upper half of each state space in order to capture better the curvature of the value functions. A continuous optimization routine is used, with initial guesses the policy functions that came from the grid search.

C Proof of proposition 6

Proof. Assume that the excess burden of taxation converges along a sample path to the value Φ (which may depend on the sample path). Recall that $\Omega^\star(g,\Phi) = 1 - a_h h(g,\Phi)^{1+\phi_h}$ (see Appendix B). Use the implicit function theorem in the two-equation system (29) and (3) to get $\partial h/\partial g = h/(h+\phi_h c) > 0$ and $\partial c/\partial g = -\phi_h c/(h+\phi_h c) < 0$. Thus, $\partial\Omega^\star/\partial g = -a_h(1+\phi_h)h^{\phi_h}\partial h/\partial g < 0$. Therefore, the surplus in marginal utility units is always *larger* for the smaller shock for *any* value of the excess burden of taxation. As a result, debt in marginal utility units is always higher for the lower shock, since for a constant Φ we have $z(g,\Phi) = \Omega^\star(g,\Phi) + \frac{\beta}{1-\beta}\sum_{g'}\pi(g')m(g')\Omega^\star(g',\Phi)$ ($m(g')$ stand for the conditional likelihood ratio induced by the constant Φ. It does not depend on the current g due to the i.i.d. assumption). But then for any $\Phi > 0$ the planner will always shift distortions towards low shocks, since $\Phi'_{g'} = \Phi/(1+(1-\beta)(1-\gamma)\eta'_{g'}\Phi)$, contradicting the premise of a constant Φ. Only in the event of a zero $\eta'_{g'}$ for any shock, i.e. only if there exists a $\Phi > 0$ such that debt in marginal utility units was equal across shocks, would it be possible to have a constant Φ. This cannot be the case, as proved earlier. The only option of having a constant Φ would be to have $\Phi = 0$, which implies that the second-best allocation converges to the first-best. In that case, the first-best is an absorbing state, and the government is using the interest income on the accumulated assets to finance government expenditures for each contingency. Note that the i.i.d. assumption in the proposition was used only to guarantee that debt in marginal utility units varies across shocks as $\Omega^\star$ does. Persistent shocks could also be allowed as long as the implied z's do vary across shocks. □

D Economy with capital

D.1 Competitive equilibrium

A price-taking firm operates the constant returns to scale technology. The firms rents capital and labor services and maximizes profits. Factor markets are competitive and therefore profit maximization leads to $w_t = F_H(s^t)$ and $r_t = F_K(s^t)$.

The first-order condition with respect to an Arrow security is the same as in (12). The labor supply condition and the Euler equation for capital are respectively

$$\frac{U_l(s^t)}{U_c(s^t)} = (1-\tau_t(s^t))w_t(s^t) \tag{D.1}$$

$$1 = \beta \sum_{s_{t+1}} \pi_{t+1}(s_{t+1}|s^t) \left(\frac{V_{t+1}(s^{t+1})}{\mu_t(V_{t+1})}\right)^{\rho-\gamma} \frac{U_c(s^{t+1})}{U_c(s^t)} R^K_{t+1}(s^{t+1}). \tag{D.2}$$

The Euler equation for capital together with (12), delivers the no-arbitrage condition

$$\sum_{s_{t+1}} p_t(s_{t+1}, s^t) R^K_{t+1}(s^{t+1}) = 1. \tag{D.3}$$

Furthermore, at the optimum we have two transversality conditions with respect to capital and Arrow securities

$$\lim_{t\to\infty} \sum_{s^t} \beta^t \pi_t(s^t) M_t(s^t)^{\frac{\rho-\gamma}{1-\gamma}} U_c(s^t) k_{t+1}(s^t) = 0 \tag{D.4}$$

$$\lim_{t\to\infty} \sum_{s^{t+1}} \beta^{t+1} \pi_{t+1}(s^{t+1}) M_{t+1}(s^{t+1})^{\frac{\rho-\gamma}{1-\gamma}} U_c(s^{t+1}) b_{t+1}(s^{t+1}) = 0. \tag{D.5}$$

D.2 Ramsey problem

Define the household's wealth as $W_t(s^t) \equiv b_t(s^t) + R^K_t(s^t)k_t(s^{t-1})$. We can recast the household's budget constraint in terms of wealth. In particular, note that

$$\begin{aligned}\sum_{s_{t+1}} p_t(s_{t+1}, s^t) W_{t+1}(s^{t+1}) &= \sum_{s_{t+1}} p_t(s_{t+1}, s^t)[b_{t+1}(s^{t+1}) + R^K_{t+1}(s^{t+1})k_{t+1}(s^t)] \\ &= \sum_{s_{t+1}} p_t(s_{t+1}, s^t) b_{t+1}(s^{t+1}) + k_{t+1}(s^t),\end{aligned}$$

by using the no-arbitrage condition (D.3). Therefore, the household's dynamic budget constraint becomes

$$c_t(s^t) + \sum_{s_{t+1}} p_t(s_{t+1}, s^t) W_{t+1}(s^{t+1}) = (1 - \tau_t(s^t)) w_t(s^t) h_t(s^t) + W_t(s^t).$$

Use now (11) and (12) to eliminate labor taxes and equilibrium prices, and multiply with the marginal utility of consumption to get

$$U_{ct} W_t = \Omega_t + \beta E_t m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}} U_{c,t+1} W_{t+1}, \tag{D.6}$$

where Ω as in (16). Define now $z_t \equiv U_{ct} W_t$ and get the same dynamic implementability constraint as in (17). The implementability constraint at $t = 0$ reads

$$U_{c0} W_0 = \Omega_0 + \beta E_0 m_1^{\frac{\rho-\gamma}{1-\gamma}} U_{c,1} W_1,$$

where $W_0 \equiv \left[(1 - \tau_0^K) F_K(s_0, k_0, h_0) + 1 - \delta\right] k_0 + b_0$, and $(k_0, b_0, \tau_0^K, s_0)$ given.

D.3 Transformed Bellman equation with capital

As in the economy without capital, use the ρ-transformation of the value function, $v(z, k, s) \equiv \frac{V(z,k,s)^{1-\rho} - 1}{(1-\beta)(1-\rho)}$. The Bellman equation takes the form

$$v(z, k, s) \quad = \quad \max_{c,h,k',z'_{s'}} U(c, 1-h) + \beta \frac{\left[\sum_{s'} \pi(s'|s) \left(1 + (1-\beta)(1-\rho) v(z'_{s'}, k', s')\right)^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{1-\rho}{1-\gamma}} - 1}{(1-\beta)(1-\rho)}$$

$$z = \Omega(c, h) + \beta \sum_{s'} \pi(s'|s) \frac{[1 + (1-\beta)(1-\rho) v(z'_{s'}, k', s')]^{\frac{\rho-\gamma}{1-\rho}}}{\left[\sum_{s'} \pi(s'|s) [1 + (1-\beta)(1-\rho) v(z'_{s'}, k', s')]^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{\rho-\gamma}{1-\gamma}}} z'_{s'} \tag{D.7}$$

$$c + k' - (1-\delta)k + g_s = F(s, k, h) \tag{D.8}$$

$$c, k' \geq 0, h \in [0, 1] \tag{D.9}$$

The values $(z'_{s'}, k')$ have to belong to the proper state space, i.e. it has to be possible that they can be generated by a competitive equilibrium with taxes when the shock is s.

D.4 First-order necessary conditions

$$c: \quad U_c + \Phi\Omega_c = \lambda \tag{D.10}$$

$$h: \quad -U_l + \Phi\Omega_h = -\lambda F_H \tag{D.11}$$

$$k': \quad \lambda = \beta \sum_{s'} \pi(s'|s) m_{s'}'^{\frac{\rho-\gamma}{1-\gamma}} v_k(z'_{s'}, k', s')[1 + (1-\beta)(\rho-\gamma)\eta'_{s'}\Phi] \tag{D.12}$$

$$z'_{s'}: \quad v_z(z'_{s'}, k', s') + \Phi\big[1 + (1-\beta)(\rho-\gamma)v_z(z'_{s'}, k', s')\eta'_{s'}\big] = 0. \tag{D.13}$$

The relative *wealth* position $\eta'_{s'}$ is defined as in (25) (with a value function V that also depends on capital now), so lemma 1 goes through also in the economy with capital. The envelope conditions are

$$v_z(z,k,s) = -\Phi \tag{D.14}$$

$$v_k(z,k,s) = \lambda(1-\delta+F_K). \tag{D.15}$$

The envelope condition (D.14) together with (D.13) delivers the same law of motion of Φ_t as in (28), leading to the same results as in proposition 3 and the discussion thereafter. Use the fact that $U_l/U_c = (1-\tau)F_H$ and express the optimal wedge in labor supply in terms of the labor tax to get the same results for the labor tax as in propositions 4 and 5. Turn into sequence notation, use the law of motion of Φ_t (28) to replace $1 + (1-\beta)(\rho-\gamma)\eta_{t+1}\Phi_t$ in (D.12) with the ratio Φ_t/Φ_{t+1} and the envelope condition (D.15) to eliminate v_k to finally get (37).

D.5 Initial period optimality conditions

The problem at $t = 0$ is

$$\bar{V}_0(b_0, k_0, s_0, \tau_0^K) \equiv \max_{c_0, h_0, k_1, z_{1,s_1}} \Big[(1-\beta)u(c_0, 1-h_0)^{1-\rho} + \beta\big[\sum_{s_1} \pi_1(s_1|s_0)V(z_{1,s_1}, k_1, s_1)^{1-\gamma}\big]^{\frac{1-\rho}{1-\gamma}}\Big]^{\frac{1}{1-\rho}}$$

subject to

$$\begin{aligned} &U_{c0}\big[\big((1-\tau_0^K)F_K(s_0,k_0,h_0) + 1 - \delta\big)k_0 + b_0\big] = \Omega(c_0, h_0) \\ &+\beta \sum_{s_1} \pi_1(s_1|s_0) \frac{V(z_{1,s_1}, k_1, s_1)^{\rho-\gamma}}{\big[\sum_{s_1} \pi_1(s_1|s_0)V(z_{1,s_1}, k_1, s_1)^{1-\gamma}\big]^{\frac{\rho-\gamma}{1-\gamma}}} z_{1,s_1} \end{aligned} \tag{D.16}$$

$$c_0 + k_1 - (1-\delta)k_0 + g_0 = F(s_0, k_0, h_0) \tag{D.17}$$

$$c_0, k_1 \geq 0, h_0 \in [0,1], \tag{D.18}$$

where $(b_0, k_0, s_0, \tau_0^K)$ given. Use the ρ-transformation of the time zero problem and let Φ_0 and λ_0 denote the multipliers on the initial period implementability constraint and the resource constraint respectively. The initial period optimality conditions are:

$$c_0: \quad U_{c0} + \Phi_0\Big[\Omega_{c0} - U_{cc,0}W_0\Big] = \lambda_0 \tag{D.19}$$

$$h_0: \quad -U_{l0} + \Phi_0\Big[\Omega_{h0} + U_{cl,0}W_0 - U_{c0}(1-\tau_0^K)F_{KH,0}k_0\Big] = -\lambda_0 F_{H0} \tag{D.20}$$

$$k_1: \quad \lambda_0 = \beta\sum_{s_1}\pi(s_1|s_0)m_{1,s_1}^{\frac{\rho-\gamma}{1-\gamma}}v_k(z_{1,s_1}, k_1, s_1)\big[1+(1-\beta)(\rho-\gamma)\eta_{1,s_1}\Phi_0\big] \tag{D.21}$$

$$\begin{aligned} z_{1,s_1}: \quad & v_z(z_{1,s_1}, k_1, s_1) \\ & +\Phi_0\big[1+(1-\beta)(\rho-\gamma)v_z(z_{1,s_1}, k_1, s_1)\eta_{1,s_1}\big] = 0, \end{aligned} \tag{D.22}$$

where $W_0 = \big[(1-\tau_0^K)F_K(s_0, k_0, h_0) + (1-\delta)\big]k_0 + b_0$, the household's initial wealth and η_{1,s_1} defined as in (25). The initial period first-order conditions for an economy without capital for the variables (c_0, h_0, z_{1,s_1}), are (D.19), (D.20) and (D.22) with $W_0 = b_0, F_{H0} = 1, F_{KH} \equiv 0$.

E Details about the illustration with capital

The production function is $F = k^\alpha h^{1-\alpha}$. The parameters for the illustration are $(\beta, \gamma, \phi_h, \alpha, \delta, \tau_0^K, b_0) = (0.96, 10, 1, 1/3, 0.08, 0.3, 0)$ with a total endowment of time normalized to unity. The parameter a_h is set so that the household works 0.4 of its time at the first-best steady state. The size of g_L is set so that the share of government expenditures in the first-best steady state output is 0.22. The high shock is $g_H = 2 \cdot g_L$ and $\pi = 0.5$. The economy features a low shock for each period except for $t = 2$, which is the reason why I use a relatively large g_H.

For the utility function of the example we have $\Omega(c, h) = 1 - a_h h^{1+\phi_h}$ and $\tau_t = \tau(\Phi_t) = \Phi_t(1+\phi_h)/(1+\Phi_t(1+\phi_h))$ (see proposition 5), which holds only for $t \geq 1$ due to the presence of initial wealth W_0. The procedure to solve the problem involves a double loop for the determination of $\Phi_2^i, i = H, L$ and Φ_0.

- *Inner loop:* Fix Φ_0 and make a guess for (Φ_2^H, Φ_2^L). Given these two values of the excess burden of taxation, the problem from period $t = 3$ onward for both histories behaves as a deterministic Ramsey taxation problem, but with different Φ's depending on the high- or low-shock history. In order to solve it, modify the return function as Chari et al. (1994) do, by defining $\bar{U}(c, 1-h; \Phi) \equiv U(c, 1-h) + \Phi\Omega(c, h)$. For the high-shock history, for $t \geq 3$ solve the Bellman equation,

$$v^{\text{CCK}}(k) = \max_{c,h,k'} \bar{U}(c, 1-h; \Phi_2^H) + \beta v^{\text{CCK}}(k')$$

subject to $c + k' - (1-\delta)k + g_L = k^\alpha h^{1-\alpha}$, with the return function $\bar{U}(c, 1-h; \Phi_2^H) = \ln c - a_h \frac{h^{1+\phi_h}}{1+\phi_h} + \Phi_2^H(1 - a_h h^{1+\phi_h})$. For the low-shock history, for $t \geq 2$, solve the same Bellman equation but with the return function $\bar{U}(c, 1-h; \Phi_2^L)$.

To determine the wealth positions z_2^i and the respective innovations that allow the update of the guesses for Φ_2^i, proceed as follows: Fix k_3^H and consider the respective Euler equation:

$$\frac{1}{c_2^H} = \beta \frac{1}{c_3^H}\big[1 - \delta + \alpha\big(\frac{k_3^H}{h_3^H}\big)^{\alpha-1}\big]$$

Given k_3^H and the policy functions we found from solving the Bellman equation, the right-hand side is known, determining therefore c_2^H. Furthermore, use the intratemporal wedge condition for $g_2 = g_H$ to get $h_2^H = \left[\frac{(1-\tau_2^H)(1-\alpha)}{a_h c_2^H}\right]^{\frac{1}{\alpha+\phi_h}} k_2^{\frac{\alpha}{\alpha+\phi_h}}$, where $\tau_2^H = \tau(\Phi_2^H)$. Plug the expression for labor in the resource constraint at $g_2 = g_H$, $c_2^H + k_3^H - (1-\delta)k_2 + g_H = k_2^\alpha (h_2^H)^{1-\alpha}$ to get one equation in the unknown k_2 and use a non-linear solver to determine it. Furthermore, use the policy functions for $t \geq 3$ to determine v_3^H and z_3^H. Utilities are calculated with the original period utility function (and *not* with the modified $\bar{U}$). Finally, use (c_2^H, h_2^H) to get $v_2^H = U(c_2^H, 1-h_2^H) + \beta v_3^H$ and $z_2^H = \Omega(c_2^H, h_2^H) + \beta z_3^H$. Use now the policy functions for the low-shock history to determine v_2^L and z_2^L at k_2. Having the utility values and the wealth positions at $t = 2$ allows us to calculate the induced likelihood ratios $m_2^i, i = H, L$, the market value of the wealth portfolio $\omega_1 = \pi m_2^H z_2^H + (1-\pi) m_2^L z_2^L$ and therefore the relative wealth positions $\eta_2^i = z_2^i - \omega_1, i = H, L$, given the guess for Φ_2^i. Use the innovations η_2^i to update the guess for Φ_2^i, $\Phi_2^i = \frac{\Phi_0}{1+(1-\beta)(1-\gamma)\eta_2^i \Phi_0}, i = H, L$ and iterate till convergence.

- *Outer loop:* After we reach convergence for Φ_2^i, calculate the rest of the allocation for $t = 0, 1$ given the initial Φ_0. In particular, the Euler equation for k_2 is

$$\frac{1}{c_1 \Phi_0} = \beta \pi m_2^H \frac{1}{c_2^H \Phi_2^H}\big[1 - \delta + \alpha\left(\frac{k_2}{h_2^H}\right)^{\alpha-1}\big] + \beta(1-\pi) m_2^L \frac{1}{c_2^L \Phi_2^L}\big[1 - \delta + \alpha\left(\frac{k_2}{h_2^L}\right)^{\alpha-1}\big].$$

The right-hand side is known, which delivers c_1. Express now labor at $t = 1$ as $h_1 = \left[\frac{(1-\tau_1)(1-\alpha)}{a_h c_1}\right]^{\frac{1}{\alpha+\phi_h}} k_1^{\frac{\alpha}{\alpha+\phi_h}}$, $\tau_1 = \tau(\Phi_0)$ and use this expression to solve for k_1 from the resource constraint. Calculate furthermore $z_1 = \Omega(c_1, h_1) + \beta\omega_1$. The initial period requires a different treatment due to the presence of initial wealth $W_0 = b_0 + \big[(1-\tau_0^K)\alpha(k_0/h_0)^{\alpha-1} + 1 - \delta\big]k_0$. Use the Euler equation for capital to get the initial value of the multiplier λ_0, $\lambda_0 = \frac{\beta}{c_1}[1 - \delta + \alpha(k_1/h_1)^{\alpha-1}]$. Then use the first-order conditions for (c_0, h_0), (D.19)-(D.20) and the resource constraint at $t = 0$ to get a system in three unknowns (c_0, h_0, k_0) to be solved with

a non-linear solver. Update Φ_0 by calculating the residual in the initial budget constraint, $I \equiv \Omega(c_0, h_0) + \beta z_1 - \frac{1}{c_0} W_0$. If $I > (<) 0$ decrease (increase) Φ_0 and go back to the inner loop to redetermine $\Phi_2^i, i = H, L$ given the new Φ_0. Stop when the initial budget constraint holds, $I = 0$.

The solution method for the outer loop is based on a fixed value k_3^H, which delivers in the end an initial value of capital k_0. I experimented with k_3^H so that the endogenous initial capital corresponds to 0.9 of the first-best steady state capital. .

There is plethora of methods for solving the Bellman equation. I use the envelope condition method of Maliar and Maliar (2013). I approximate the value function with a 5th degree polynomial in capital and I use 100 grid points. Furthermore, since the steady-state capital depends on Φ_2^i, I re-adjust the bounds of the state space for each calculation of the value function in order to focus on the relevant part of the state space. For the high-shock history, I set the lower bound as $\underline{K} = 0.95 \cdot \min(k_3^H, k_{ss}^H)$ and the upper bound $\bar{K} = 1.05 \cdot \max(k_3^H, k_{ss}^H)$. In the same vain, for the low-shock history, I set $\underline{K} = 0.95 \cdot \min(k_2, k_{ss}^L)$ and $\bar{K} = 1.05 \cdot \max(k_2, k_{ss}^L)$. The variables $k_{ss}^i, i = H, L$ denote the respective steady states.

F Sequential formulation

I provide here the sequential formulation of the Ramsey problem. I consider an economy with capital. The specialization of the analysis to an economy without capital is obvious. Let $X_t \equiv M_t^{\frac{\rho-\gamma}{1-\gamma}}, X_0 \equiv 1$. Let v refer to the ρ-transformation of the utility criterion. The Ramsey problem is

$$\max v_0(\{c\}, \{h\})$$

subject to

$$\sum_{t=0}^{\infty} \beta^t \sum_{s^t} \pi_t(s^t) X_t(s^t) \Omega(c_t(s^t), h_t(s^t)) = U_{c0} W_0 \tag{F.1}$$

$$c_t(s^t) + k_{t+1}(s^t) - (1-\delta) k_t(s^{t-1}) + g_t(s^t) = F(s_t, k_t(s^{t-1}), h_t(s^t)) \tag{F.2}$$

$$X_{t+1}(s^{t+1}) = m_{t+1}(s^{t+1})^{\frac{\rho-\gamma}{1-\gamma}} X_t(s^t), \quad X_0 \equiv 1 \tag{F.3}$$

$$v_t(s^t) = U(c_t(s^t), 1 - h_t(s^t)) + \beta \frac{\left[\sum_{s_{t+1}} \pi_{t+1}(s_{t+1}|s^t) \left[1 + (1-\beta)(1-\rho) v_{t+1}(s^{t+1})\right]^{\frac{1-\gamma}{1-\rho}}\right]^{\frac{1-\rho}{1-\gamma}} - 1}{(1-\beta)(1-\rho)}, \quad t \geq 1 \tag{F.4}$$

where $W_0 \equiv R_0^K k_0 + b_0$, $(b_0, k_0, s_0, \tau_0^K)$ *given*, and $m_{t+1} = \frac{\left[1+(1-\beta)(1-\rho)v_{t+1}\right]^{\frac{1-\gamma}{1-\rho}}}{E_t\left[1+(1-\beta)(1-\rho)v_{t+1}\right]^{\frac{1-\gamma}{1-\rho}}}$.

Assign multipliers $\bar{\Phi}$, $\beta^t \pi_t \lambda_t$, $\beta^t \pi_t \nu_t$ and $\beta^t \pi_t \xi_t$ on (F.1), (F.2), (F.3) and (F.4) respectively. The derivatives of the utility function are $\frac{\partial v_0}{\partial c_t} = \beta^t \pi_t X_t U_{ct}$ and $\frac{\partial v_0}{\partial h_t} = -\beta^t \pi_t X_t U_{lt}$. The first-order necessary conditions are

$$c_t, t \geq 1: \quad X_t(s^t)U_c(s^t) + \bar{\Phi} X_t(s^t)\Omega_c(s^t) + \xi_t(s^t)U_c(s^t) = \lambda_t(s^t) \tag{F.5}$$

$$h_t, t \geq 1: \quad -X_t(s^t)U_l(s^t) + \bar{\Phi} X_t(s^t)\Omega_h(s^t) - \xi_t(s^t)U_l(s^t) = -\lambda_t(s^t)F_H(s^t) \tag{F.6}$$

$$k_{t+1}(s^t), t \geq 0: \quad \lambda_t(s^t) = \beta \sum_{s_{t+1}} \pi_{t+1}(s_{t+1}|s^t)\lambda_{t+1}(s^{t+1})[1 - \delta + F_K(s^{t+1})] \tag{F.7}$$

$$X_t(s^t), t \geq 1: \quad \nu_t(s^t) = \bar{\Phi}\Omega_t(s^t) + \beta \sum_{s_{t+1}} \pi_{t+1}(s_{t+1}|s^t) m_{t+1}(s^{t+1})^{\frac{\rho-\gamma}{1-\gamma}} \nu_{t+1}(s^{t+1}) \tag{F.8}$$

$$v_t(s^t), t \geq 1: \quad \xi_t(s^t) = (1-\beta)(\rho-\gamma)X_t(s^t)\phi_t(s^t) + m_t(s^t)^{\frac{\rho-\gamma}{1-\gamma}} \xi_{t-1}(s^{t-1}), \tag{F.9}$$

where

$$\phi_t(s^t) \equiv V_t(s^t)^{\rho-1}\nu_t(s^t) - \mu_t(s^t)^{\rho-1} \sum_{s_t} \pi_t(s_t|s^{t-1}) m_t(s^t)^{\frac{\rho-\gamma}{1-\gamma}} \nu_t(s^t),$$

and $\xi_0 \equiv 0$. The optimality conditions with respect to the initial consumption-labor allocation are (D.19) and (D.20).

I will show now the mapping between the sequential formulation and the recursive formulation and in particular the relationship between the time-varying Φ_t and ξ_t. Solve at first (F.8) forward to get

$$\nu_t = \bar{\Phi} E_t \sum_{i=0}^{\infty} \beta^i \frac{X_{t+i}}{X_t} \Omega_{t+i}$$

and therefore $\nu_t = \bar{\Phi} U_{ct} W_t = \bar{\Phi} z_t$, i.e. ν_t – the shadow value to the planner of an increase in X_t– is equal to wealth (in marginal utility terms) times the cost of taxation $\bar{\Phi}$. Thus, ϕ_t –the "innovation" in the multiplier ν_t– is equal to a multiple of η_t, $\phi_t = \bar{\Phi}\eta_t$. Furthermore, define the scaled multiplier $\tilde{\xi}_t \equiv \xi_t / X_t, \tilde{\xi}_0 \equiv 0$ and note that it follows the law of motion

$$
\begin{aligned}
\tilde{\xi}_t &= (1-\beta)(\rho-\gamma)\phi_t + \tilde{\xi}_{t-1} \\
&= (1-\beta)(\rho-\gamma)\sum_{i=1}^{t}\phi_i = (1-\beta)(\rho-\gamma)\sum_{i=1}^{t}\eta_i\bar{\Phi}
\end{aligned}
$$

Turn now to the multiplier in the text which, when solved backwards, delivers $\Phi_t = \Phi_0/(1+(1-\beta)(\rho-\gamma)\sum_{i=1}^{t}\eta_i\Phi_0)$, where Φ_0 is the multiplier on the initial period implementability constraint. Thus, by setting $\Phi_0 = \bar{\Phi}$ we have

$$
\Phi_t = \frac{\bar{\Phi}}{1+\tilde{\xi}_t}, \tag{F.10}
$$

or, in terms of the non-scaled ξ_t, $\Phi_t = \bar{\Phi}X_t/(X_t+\xi_t)$. Therefore, the time-varying excess burden of taxation captures the shadow value of continuation utilities that determine intertemporal marginal rates of substitution. Consider now the multipliers λ_t in the sequential formulation and their relationship to their counterparts in the recursive formulation. (F.5) can be written as $U_{ct} + \frac{\bar{\Phi}X_t}{X_t+\xi_t}\Omega_{ct} = \frac{\lambda_t}{X_t+\xi_t}$. Given (D.10) and (F.10), we get that $\lambda_t = (X_t+\xi_t)\lambda_t^R$, where λ_t^R stands for the multipliers of the recursive formulation. Thus,

$$
\begin{aligned}
\frac{\lambda_{t+1}}{\lambda_t} &= \frac{X_{t+1}+\xi_{t+1}}{X_t+\xi_t}\frac{\lambda_{t+1}^R}{\lambda_t^R} = \frac{X_{t+1}}{X_t}\frac{\frac{\bar{\Phi}X_t}{X_t+\xi_t}}{\frac{\bar{\Phi}X_{t+1}}{X_{t+1}+\xi_{t+1}}}\frac{\lambda_{t+1}^R}{\lambda_t^R} \\
&= m_{t+1}^{\frac{\rho-\gamma}{1-\gamma}}\frac{\lambda_{t+1}^R}{\lambda_t^R}\frac{\Phi_t}{\Phi_{t+1}}.
\end{aligned}
$$

Thus, (F.7) delivers the same optimality condition with respect to capital as (37).

References

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

Anderson, Evan W. 2005. The dynamics of risk-sensitive allocations. *Journal of Economic Theory* 125 (2):93–150.

Bansal, Ravi and Amir Yaron. 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* LIX (4):1481–1509.

Bassetto, Marco. 1999. Optimal Fiscal Policy with Heterogenous Agents. *Quantitative Economics* Forthcoming.

Chamley, Christophe. 1986. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54 (3):607–622.

Chari, V.V. and Patrick J. Kehoe. 1999. Optimal fiscal and monetary policy. *Handbook of Macroeconomics* 1:1671–1745.

Chari, V.V., Lawrence J. Christiano, and Patrick J. Kehoe. 1994. Optimal Fiscal Policy in a Business Cycle Model. *Journal of Political Economy* 102 (4):617–652.

Croce, Mariano M., Howard Kung, Thien T. Nguyen, and Lucas Schmid. 2012a. Fiscal Policies and Asset Prices. *The Review of Financial Studies* 25 (9):2635–2672.

Croce, Mariano M., Thien T. Nguyen, and Lukas Schmid. 2012b. The market price of fiscal uncertainty. *Journal of Monetary Economics* 59 (5):401–416.

Epstein, Larry G. and Stanley E. Zin. 1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* 57 (4):937–969.

Epstein, Larry G., Emmanuel Farhi, and Tomasz Strzalecki. 2013. How Much Would You Pay to Resolve Long-Run Risk? *American Economic Review* Forthcoming.

Farhi, Emmanuel. 2010. Capital Taxation and Ownership when Markets are Incomplete. *Journal of Political Economy* 118 (5):908–948.

Farhi, Emmanuel and Iván Werning. 2008. Optimal savings distortions with recursive preferences. *Journal of Monetary Economics* 55 (1):21–42.

Geoffard, Pierre-Yves. 1996. Discounting and optimizing: Capital accumulation problems as variational minmax problems. *Journal of Economic Theory* 69:53–70.

Gomes, Francisco, Alexander Michaelides, and Valery Polkovnichenko. 2013. Fiscal Policy and Asset Prices with Incomplete Markets. *The Review of Financial Studies* 26 (2):531–566.

Gottardi, Piero, Atsushi Kajii, and Tomoyuki Nakajima. 2014. Optimal taxation and debt with uninsurable risks to human capital accumulation. Mimeo, European University Institute.

Gourio, François. 2012. Disaster Risk and Business Cycles. *American Economic Review* 102 (6):2734–66.

Hansen, Lars Peter and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *American Economic Review* 91 (2):60–66.

Hansen, Lars Peter, John C. Heaton, and Nan Li. 2008. Consumption Strikes Back? Measuring Long-Run Risk. *Journal of Political Economy* 116 (2):260–302.

Judd, Kenneth L. 1985. Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics* 28 (1):59–83.

Karantounias, Anastasios G. 2013. Managing pessimistic expectations and fiscal policy. *Theoretical Economics* 8 (1):193–231.

Koopmans, Tjalling C. 1960. Stationary ordinary utility and impatience. *Econometrica* 28 (2):287–309.

Kreps, D. M. and E. L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46:185–200.

Kydland, Finn E. and Edward C. Prescott. 1980. Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control* 2:79–91.

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

Maliar, Lilia and Serguei Maliar. 2013. Envelope Condition Method versus Endogenous Grid Method for Solving Dynamic Programming Problems. *Economic Letters* 120:262–266.

Martin, Ian. 2012. On the Valuation of Long-Dated Assets. *Journal of Political Economy* 120 (2):346–358.

Petrosky-Nadeau, Nicolas, Lu Zhang, and Lars-Alexander Kuehn. 2013. Endogenous disasters and asset prices. Mimeo, Carnegie Mellon University.

Shin, Yongseok. 2006. Ramsey meets Bewley: Optimal Government Financing with Incomplete Markets. Mimeo, Washington University in St. Louis.

Sleet, Christopher. 2004. Optimal taxation with private government information. *Review of Economic Studies* 71 (4):1217–1239.

Sleet, Christopher and Sevin Yeltekin. 2006. Optimal taxation with endogenously incomplete debt markets. *Journal of Economic Theory* 127 (1):36–73.

Strzalecki, Tomasz. 2011. Axiomatic Foundations of Multiplier Preferences. *Econometrica* 79 (1):47–73.

———. 2013. Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion. *Econometrica* 81 (3):1039–1074.

Tallarini, Thomas D. Jr. 2000. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (3):507–532.

Weil, Philippe. 1990. Non-Expected Utility in Macroeconomics. *Quarterly Journal of Economics* CV (1):29–42.

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

Zhu, Xiaodong. 1992. Optimal fiscal policy in a stochastic growth model. *Journal of Economic Theory* 58 (2):250–289.